天翼云
State Cloud

弹性云主机

天翼云科技有限公司

1 产品动态 17
2 新手指引 36
2.1 操作导航 37
2.2 新手入门：购买并登录 Windows 弹性云主机 39
2.3 新手入门：购买并登录 Linux 弹性云主机 41
3DeepSeek 专题 42
3.1DeepSeek 专题导航 42
3.2 图解 DeepSeek 与公有云深度融合 43
3.2.1 从基础设施到智能中枢：DeepSeek 如何重塑公有云服务价值链 43
3.2.2 高性能 GPU 云主机助力 DeepSeek 深度应用 47
3.2.3 天翼云 SD-WAN 与 DeepSeek 超强联动，开启云上高效互联新时代 49
3.3 实践指南：DeepSeek 驱动高效能云生态 50
3.3.1GPU 云主机/弹性云主机：零基础搭建 DeepSeek 云端环境指南 50
3.3.1.1 在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本 50
3.3.1.2 基于 OpenWebUI+Ollama+DeepSeek 实现大模型私有知识库零代码构建的最佳实践 82
3.3.1.3 基于 DeepSeek 和 Chroma 构建个性化知识库的最佳实践 88
3.3.1.4 基于 LLaMA-Factory 微调 DeepSeek-R1-Distill-Qwen-7B 最佳实践 94
3.3.1.5 基于 Open WebUI 实现 DeepSeek API 调用及外部大模型 API 接入最佳实践 105
3.3.1.6AnythingLLM 协同 Ollama 启用 DeepSeek 大模型最佳实践 110
3.3.1.7CherryStudio 协同 Ollama 启用 DeepSeek 大模型 127
3.3.1.8 天翼云 CPU 实例部署 DeepSeek-R1 模型最佳实践 133
3.3.2GPU 物理机：物理机搭建 DeepSeek 指南 145
3.3.3SD-WAN 跨境：SD-WAN 助力 DeepSeek 模型定向加速 145
3.3.4 智算容器：云容器引擎与 DeepSeek 融合实践 145
3.3.5 函数计算：天翼云函数计算与 DeepSeek 大模型 145
3.4Q&A：典型问题解析与策略应对 145
4 产品概述 146
4.1 产品定义 146
4.2 术语解释 149
4.3 产品优势 151
4.4 弹性云主机选型 153
4.5 产品应用场景 157
4.6 功能特性 169
4.7 产品使用须知 172
4.8 产品使用限制 173
4.9 云主机与其他服务之间的关系 177
4.10 产品规格 181
4.10.1 规格说明 181
4.10.2 规格族 184
4.10.3 通用云主机 194
4.10.3.1 概述 194

4.10.3.2 通用型 197
4.10.3.3 计算型 216
4.10.3.4 内存型 234
4.10.3.5 增强型 245
4.10.4 本地盘云主机 255
4.10.5GPU 加速/AI 加速云主机 261
4.10.6 国产云主机 284
4.10.6.1 鲲鹏系列 285
4.10.6.2 海光系列 289
4.10.6.3 飞腾系列 297
4.10.7 经济型云主机 302
4.11 镜像 303
4.11.1 镜像类型 303
4.11.2Cloud-init 303
4.12 安全 306
4.12.1 数据保护技术 306
4.12.1.1Hypervisor 安全 306
4.12.1.2 用户加密 308
4.12.2 服务的访问控制 310
4.12.3 监控安全风险 311
4.12.4 故障恢复 312
4.13 网络 313
4.14 云硬盘 315
4.15 地域和可用区 318
5 购买指南 321
5.1 计费说明 321
5.1.1 计费项及其计费方式 321
5.1.2 价格与折扣 322
5.2 计费模式 345
5.2.1 按量计费模式 345
5.2.2 包年包月计费模式 347
5.3 变更配置 350
5.3.1 变更概述 350
5.3.2 支持变配的实例规格 351
5.3.3 云主机冷变配规则说明 351
5.4 续费说明 354
5.4.1 规则说明 355
5.4.2 手动续订 355
5.4.3 自动续订 358
5.5 退费说明 366
5.5.1 规则说明 366
5.5.2 退费流程 371
5.6 包周期按量互转 373
5.7 成本优化 374

5.8 计费 FAQ 375
6 快速入门 379
6.1 注册账号 379
6.2 创建弹性云主机 380
6.3 登录 Linux 弹性云主机 386
6.4 登录 Windows 弹性云主机 389
6.5 入门实践 392
6.6 修订记录 400
7 用户指南 401
7.1 云主机实例 401
7.1.1 实例生命周期 401
7.1.2 查看实例 404
7.1.2.1 查看实例状态 404
7.1.2.2 查看实例详情 404
7.1.2.3 导出实例信息 405
7.1.2.4 搜索实例 405
7.1.3 登录实例 406
7.1.3.1 登录 Windows 弹性云主机实例 406
7.1.3.1.1Windows 弹性云主机登录方式概述 406
7.1.3.1.2Windows 弹性云主机控制中心远程登录（VNC 方式） 408
7.1.3.1.3 远程桌面连接（MSTSC 方式） 409
7.1.3.1.4 在 Linux 主机上登录 Windows 云主机 410
7.1.3.1.5 在移动设备上登录 Windows 云主机 412
7.1.3.1.6Mac OS 系统登录 Windows 云主机 413
7.1.3.2 登录 Linux 弹性云主机实例 414
7.1.3.2.1Linux 弹性云主机登录方式概述 414
7.1.3.2.2Linux 弹性云主机远程登录（VNC 方式） 416
7.1.3.2.3SSH 密钥方式登录 418
7.1.3.2.4SSH 密码方式登录 420
7.1.3.2.5 在移动设备上登录 Linux 云主机 422
7.1.3.2.6Mac OS 系统登录 Linux 云主机 424
7.1.3.3 管理远程登录用户 425
7.1.4 管理实例属性 426
7.1.4.1 修改实例名称 426
7.1.4.2 修改主机名称（hostname） 428
7.1.4.3 批量设置有序的实例名称或主机名称 430
7.1.4.4 修改实例描述 432
7.1.4.5 更改实例时区 433
7.1.4.6 开启和关闭实例保护功能 435
7.1.4.7 配置主机名与 IP 地址的映射 436
7.1.4.8 实例自定义数据注入 438
7.1.4.9 实例自定义元数据 445
7.1.5 管理实例状态 456
7.1.5.1 开机实例 456

7.1.5.2 关机实例 457
7.1.5.3 重启实例 459
7.1.5.4 重装操作系统 460
7.1.5.5 变更规格 462
7.1.5.6 创建相同配置实例 464
7.1.5.7 克隆实例 464
7.1.6 管理云主机组 466
7.1.7 安装驱动和工具包（可选） 470
7.1.7.1NVIDIA 驱动安装指引 470
7.1.7.2 创建配备 GPU 驱动的 GPU 云主机（Linux） 473
7.2 云主机启动模版 477
7.2.1 创建云主机启动模版 477
7.2.2 查看云主机启动模版 478
7.2.3 修改云主机启动模版 480
7.2.4 使用云主机启动模版创建云主机 481
7.2.5 删除云主机启动模版 482
7.3 云主机镜像 483
7.3.1 镜像概述 483
7.3.2 通过镜像创建云主机 485
7.3.3 通过云主机创建 Windows 私有镜像 486
7.3.4 通过云主机创建 Linux 私有镜像 487
7.3.5 镜像更多操作 488
7.4 云硬盘 488
7.4.1 云硬盘概述 488
7.4.2 创建云硬盘 491
7.4.3 挂载数据盘 497
7.4.4 卸载数据盘 501
7.4.5 扩容云硬盘 503
7.4.6 加密云硬盘 505
7.4.6.1 支持云硬盘加密功能 505
7.4.6.2 管理加密云硬盘 508
7.4.7 设置云硬盘释放策略 513
7.5 网卡 515
7.5.1 弹性网卡概述 515
7.5.2 创建弹性网卡 517
7.5.3 管理弹性网卡 519
7.5.4 修改弹性网卡基本信息、分配 IPv6、辅助私网 IP 520
7.5.5 如何修改内网 IP、切换 VPC 521
7.5.6 开启网卡多队列功能 524
7.5.7 虚拟 IP 概述 527
7.5.8 IPv6 带宽 528
7.6 弹性 IP 531
7.6.1 弹性 IP 概述 531
7.6.2 绑定弹性 IP 531

7.6.3 解绑弹性 IP 533
7.6.4 更换弹性 IP 534
7.6.5 修改弹性 IP 带宽 535
7.6.6 使用 SNAT 访问公网 536
7.7 密码和密钥对 540
7.7.1 密码 540
7.7.1.1 密码使用场景介绍 540
7.7.1.2 在控制台重置密码 541
7.7.2 密钥对 542
7.7.2.1 密钥对使用场景介绍 542
7.7.2.2 在控制台创建密钥对 544
7.7.2.3 导入密钥对 545
7.7.2.4 为云主机绑定/更换密钥对 546
7.7.2.5 为云主机解绑密钥对 547
7.8 云主机快照 548
7.8.1 云主机快照管理 548
7.8.2 云主机快照策略 559
7.9 云主机备份 566
7.9.1 备份弹性云主机 566
7.10 文件系统 567
7.10.1 一键挂载文件系统 567
7.11 监控 569
7.11.1 监控弹性云主机 569
7.11.2 监控 Agent 概览 570
7.11.3 监控 Agent 版本说明 576
7.11.4 安装监控 Agent 576
7.11.5 云主机支持的操作系统监控指标（安装监控 Agent） 583
7.11.6 设置告警规则 596
7.11.7 查看监控指标 599
7.11.8 弹性云主机支持的基础监控指标 601
7.12 安全 605
7.12.1 提升云主机安全的方法 605
7.12.2 安全组 610
7.12.2.1 安全组概述 610
7.12.2.2 安全组配置示例 617
7.12.2.3 配置安全组规则 620
7.12.2.3.1 修改安全组规则 620
7.12.2.4 更改安全组 621
7.12.3 数据保护技术 622
7.12.3.1 用户加密 622
7.12.4 重点操作短信二次验证 623
7.12.4.1 验证规则 623
7.12.4.2 开启重点操作短信验证 624
7.12.4.3 关闭重点操作短信验证 625

7.12.5 云服务器 CTIAM 626
7.12.5.1 统一身份认证 IAM 介绍 626
7.12.5.2 通过 IAM 用户控制资源访问 635
7.12.6 云审计 636
7.13 迁移 639
7.13.1 迁移云主机实例 639
7.13.2 迁移本地盘云主机数据 646
7.14 运维 647
7.14.1 云助手 647
7.14.1.1 云助手产品介绍 647
7.14.1.2 安装云助手客户端 651
7.14.1.3 购买指南 653
7.14.1.4 创建命令 653
7.14.1.5 免登录执行命令 654
7.14.1.6 查看命令执行结果 656
7.14.1.7 查看和执行公共命令 656
7.14.1.8 上传文件 658
7.14.1.9 查看文件上传结果 660
7.15 资源与标签 661
7.15.1 标签管理 661
7.15.1.1 标签概述 661
7.15.1.2 标签设计原则及示例 662
7.15.1.3 绑定标签 664
7.15.1.4 使用标签筛选资源 666
7.15.1.5 解绑标签 667
7.15.2 云主机服务委托 668
7.15.3 配额调整 669
8 常见问题 674
8.1 高频常见问题 674
8.2 地域与可用区类 675
8.2.1 如何选择地域？ 675
8.2.2 地域之间是否存在差异？ 676
8.2.3 已购买的弹性云主机，是否支持更换地域？ 677
8.2.4 什么是可用区，怎样选择可用区？ 677
8.2.5 可用区之间的数据传输是否需要收费？ 677
8.2.6 是否可以将应用的不同组件分散部署到不同的地域？ 678
8.2.7 是否提供支持应用灾备的相关服务？ 678
8.2.8 是否可以在不同的区域之间实施应用灾备？ 678
8.2.9 如何获取弹性云主机的物理机房位置？ 678
8.3 购买与退订冻结相关问题 679
8.3.1 弹性云主机提供哪些计费方式，区别是什么？ 679
8.3.2 同一台云主机是否同时支持两种计费方式？ 679
8.3.3 是否支持更改已有云主机的计费方式？ 679
8.3.4 账户里面有余额，为什么无法创建按量付费云主机？ 679

8.3.5 弹性云主机关机后还会继续计费吗？ 680
8.3.6 一个账户最多可以创建多少按量付费的云主机？ 680
8.3.7 如何选择适合业务的云主机实例？ 680
8.3.8 购买实例时，资源已经售罄怎么办？ 680
8.3.9 什么是 ARM CPU 架构与 X86 CPU 架构？ 681
8.3.10 控制台无法看到云主机，如何处理？ 681
8.3.11 如何处理支付订单后云主机开通失败？ 682
8.3.12 申请开通弹性云主机需要多久？ 682
8.3.13 什么情况下弹性云主机会被冻结，冻结后怎么办？ 682
8.3.14 如何退订弹性云主机？ 683
8.3.15 已到期资源如果续订，续订周期如何计算？ 683
8.3.16 是否支持账户余额变动提醒？ 683
8.3.17 弹性云主机怎样停止计费？ 684
8.4 创建与删除 684
8.4.1 创建类 684
8.4.2 删除类 685
8.5 登录与连接 688
8.5.1VNC 登录类 688
8.5.1.1VNC 方式登录后，播放音频文件没有声音怎么办？ 688
8.5.1.2 如何修改 Windows 弹性云主机的分辨率 690
8.5.1.3 怎样修改远程登录的端口？ 692
8.5.1.4 修改 Linux 弹性云主机的默认 SSH 端口后，使用新端口无法登录？ 701
8.5.1.5 使用 VNC 远程登录，想要复制本地字符发送到虚机中，发现一次性无法达到描述的 3000 个字符 702
8.5.2 控制台登录类 703
8.5.2.1VNC 方式登录弹性云主机时，登录界面显示乱码怎么办？ 703
8.5.2.2 通过控制台登录弹性云主机时提示 1006 或 1000 怎么办？ 704
8.5.2.3 VNC 方式登录弹性云主机时，系统黑屏输入无反应？ 704
8.5.2.4 VNC 方式登录弹性云主机后，查看数据失败，VNC 无法正常使用？ 705
8.5.3 登录前准备类 705
8.5.3.1 使用创建时的用户名和密码无法 SSH 方式登录云主机 705
8.5.3.2 远程登录时需要输入的帐号和密码是什么 707
8.5.3.3 弹性云主机登录前准备工作有哪些 707
8.5.3.4 远程登录忘记密码怎么办 708
8.5.4 远程连接类 708
8.5.4.1MSTSC 方式登录 Windows 2012 的弹性云主机，登录失败怎么办？ 708
8.5.4.2 Windows 2012 系统卸载某些软件后无法进入系统桌面怎么办？ ..709
8.5.4.3 远程登录弹性云主机时，对浏览器版本的要求？ 711
8.5.4.4 使用私有镜像创建 Windows 弹性云主机时设置的密码不生效怎么办？ 712
8.5.4.5 弹性云主机实例出现了异地登录怎么办 713
8.5.4.6 无法远程登录 Windows 云主机怎么办 713
8.5.4.7 无法远程登录 Linux 云主机怎么办 715
8.6 使用故障类 717

8.6.1 弹性云主机使用时有什么限制 717
8.6.2 宿主机故障时，弹性云主机是否能自动恢复 717
8.6.3 如何处理管理控制台异常提示信息 717
8.6.4 如何修改弹性云主机（Windows）的 SID 719
8.6.5Windows 弹性云主机没有声音 722
8.6.6 弹性云主机配置的邮件无法正常发送 728
8.6.7 使用 free -m 命令查询弹性云主机内存大小与实际配置不符 732
8.6.8inux 虚机内关键文件路径 732
8.6.9 鲲鹏、海光、飞腾系列云主机挂载云硬盘和弹性网卡异常 736
8.7 规格变更类 737
8.7.1 弹性云主机怎样变更规格 737
8.7.2 国产化云主机变配有什么限制 738
8.7.3 主机升配后 CPU 的 ID 是否会变化 738
8.7.4Windows 弹性云主机变更规格后磁盘脱机怎么办？ 738
8.7.5Linux 弹性云主机变更规格后磁盘脱机怎么办？ 742
8.7.6 弹性云主机变配操作无法成功提交 744
8.8 操作系统管理 745
8.8.1 图形化系统安装类 745
8.8.1.1 云主机是否有图形界面 745
8.8.1.2CentOS 7 系列弹性云主机如何安装图形化界面 745
8.8.1.3CentOS 8 系列弹性云主机如何安装图形化界面 747
8.8.1.4Ubuntu 系列弹性云主机如何安装图形化界面 749
8.8.1.5Debian 系列弹性云主机安装图形化界面流程 753
8.8.2 操作系统故障类 756
8.8.2.1 针对 Intel 处理器芯片存在的 Meltdown 和 Spectre 安全漏洞，应该如何规避？ 756
8.8.2.3 怎样查看 GPU 加速型云主机的 GPU 使用率？ 760
8.8.2.4 Linux 弹性云主机如何升级内核？ 764
8.8.2.5 用户自持镜像导致的弹性云主机操作系统无法正常启动是什么原因？ 764
8.8.2.6 文件系统损坏，Linux 弹性云主机启动失败 766
8.8.2.7 文件系统 checkfs 使用 fsck 工具检查并修复文件系统 767
8.8.3 重装操作系统 768
8.8.3.1 重装弹性云主机时是否可以选择其他操作系统 768
8.8.3.2 重装操作系统需要多长时间？ 770
8.8.3.3 用户能否自己安装或者升级操作系统 770
8.8.3.4 重装/切换操作系统/变更规格对磁盘数据有影响吗 771
8.9 文件上传/数据传输 771
8.9.1 使用 FTP 上传文件时写入失败、文件传输失败 771
8.9.2 WinSCP 无法连接到 Linux 云主机 773
8.9.3 Windows 云主机搭建了 FTP，通过公网无法访问 775
8.9.4 使用 FTP 上传文件时客户端连接服务端超时 782
8.9.5 本地 Windows 主机通过 MSTSC 上传文件到 Windows 云主机 784
8.9.6 本地 Windows 主机使用对象存储上传文件到 Windows 云主机 787

8.9.7 本地 Windows 主机使用 WinSCP 上传文件到 Linux 云主机 788
8.9.8 本地 MacOS 系统主机上传文件到 Windows 云主机 790
8.9.9 本地 Linux 主机使用 SCP 上传文件到 Linux 云主机 794
8.9.10 本地 Linux 主机使用 SFTP 上传文件到 Linux 云主机 794
8.9.11 本地 Windows 主机使用 FTP 上传文件到 Windows/Linux 云主机 796
8.9.12 本地 Linux 主机使用 FTP 上传文件到 Linux 云主机、 797
8.9.13 怎样在本地主机和 Windows 云主机之间互传数据？ 798
8.9.14FTP 服务器上的文件夹因权限访问出错怎么办？ 801
8.10 迁移云服务器 802
8.10.1 弹性云主机可以迁移到其他地域/帐号吗？ 802
8.11 镜像源管理 803
8.11.1 内网 yum 源及 NTP 配置 803
8.11.2Windows 或 Linux 操作系统镜像怎么选 806
8.11.3 常见 docker 源配置指导 806
8.12 磁盘管理类 814
8.12.1 磁盘分区与虚拟内存类 814
8.12.2 磁盘扩容类 817
8.12.3 卸载云硬盘操作指南及故障排除 817
8.12.4 其它 818
8.13 密码与密钥对类 821
8.13.1 密码 821
8.13.1.1 云主机默认密码是多少？ 821
8.13.1.2 忘记天翼云账号密码怎么办？ 821
8.13.1.3 如何在控制台重置弹性云主机密码？ 822
8.13.1.4 如何在操作系统内部修改云主机密码？ 823
8.13.1.5 用户创建弹性云主机成功后远程登录提示密码错误？ 823
8.13.1.6 一键重置密码后无法使用新密码登录弹性云主机 824
8.13.1.7 用户未修改密码，正常使用弹性云主机一段时间后无法登录 824
8.13.1.8 如何关闭 Selinux？ 825
8.13.2 密钥对 825
8.13.2.1 如何创建密钥对？ 825
8.13.2.2 如何使用密钥对？ 826
8.13.2.3 密钥对无法下载 826
8.13.2.4 无法导入密钥对 826
8.13.2.5 怎样查询弹性云主机使用的密钥对是哪个？ 829
8.13.2.6 已开云主机是否还支持使用密钥对？ 830
8.13.2.7 如何在云主机上使用安装 USB 类型的 Ukey 秘钥？ 830
8.13.2.8 弹性云主机中使用 cloudbase-init 是什么？ 830
8.13.2.9Python 升级导致 Cloud-init 不工作 832
8.13.2.10 使用密钥文件无法正常登录 Linux 弹性云主机 832
8.13.2.11 通过 puttygen.exe 工具创建的密钥对，导入管理控制台失败 833
8.14 网络配置类 835
8.14.1 弹性 IP 类 835
8.14.2DNS 与 NTP 配置类 843

8.14.2.1 配置 DNS 843
8.14.2.2 怎样配置弹性云主机的 DNS 和 NTP 信息？ 847
8.14.3 网卡类 852
8.14.3.1 弹性云主机添加网卡后是否会自动启动？ 852
8.14.3.2 怎样修改云主机的 VPC 网段、子网的网段？ 853
8.14.4 路由类 853
8.14.5 访问网站或应用故障类 854
8.14.6iptables 如何配置、建议配置规则指导 862
8.14.7 其它类 864
8.14.7.1 云主机 mac 是否为固定的？ 864
8.14.7.2DHCP 无法正常获取内网 IP？ 864
8.14.7.3 如何查看、修改 Linux 弹性云主机的内核参数？ 866
8.14.7.4 端口映射配置 867
8.14.7.5 如何查看云主机的 mac 地址？ 868
8.14.7.6 网络性能测试方法 870
8.14.7.7 不同帐号下弹性云主机内网是否可以互通？ 876
8.14.7.8 我购买的云主机是否在同一子网？ 876
8.14.7.9 云主机网络优化方案 876
8.14.7.10 如何排查带宽超过限制 877
8.15 安全类 879
8.15.1 安全组和安全规则配额是多少 879
8.15.2 安全组加规则报错：参数错误，请检查你的参数 879
8.15.3 怎么设置云主机密码可以在一定程度上提升云主机安全？ 880
8.15.4 如何关闭重点操作短信验证 880
8.15.5 弹性云主机上是否可以搭建数据库？ 881
8.15.6 如何保护弹性云主机安全 881
8.15.7 如何防止实例被暴力破解 882
8.16 云主机卡顿 882
8.16.1Linux 云主机卡顿 882
8.16.2 Windows 云主机卡顿 889
8.17 数据库应用 894
8.17.1 弹性云主机是否支持 Oracle 数据库？ 895
8.18 云主机快照常见问题 895
8.19 产品咨询类 896
8.19.1 怎样注册天翼云 896
8.19.2 弹性云主机可以做什么 897
8.19.3 私有数据是否支持物理邮寄 897
8.19.4 升级 CPU、内存、带宽时是否需要关机 898
8.19.5Linux 虚机 product_uuid 参数是否可以修改 898
8.19.6 香港节点的云主机是否可以实现与国内云主机互通 898
8.19.7 云主机 windows 2019 操作系统对应普通电脑哪个操作系统 899
8.20 其它类 899
8.21 云助手常见问题 904
9 故障修复 906

9.1 高频故障案例 906
9.1.1 云主机带宽占用高怎么办 906
9.1.2Linux 操作系统执行 passwd 命令重置密码失败提示：Authentication token manipulation error 908
9.1.3 云主机错误状态及解决方案 910
9.1.4 如何解决云平台创建云主机状态错误 912
9.1.5 云主机端口不通怎样排查？ 914
9.1.6 云主机网络延迟和丢包，如何定位 918
9.1.7 弹性云主机启动缓慢 921
9.1.8ping 不通或丢包时如何进行链路测试？ 929
9.1.9Linux 系统云主机如何自查病毒及中病毒后处理建议 935
9.1.10Windows 系统云主机如何自查病毒及中病毒后处理建议 941
9.2 操作系统类（Windows） 946
9.2.1 云主机时间与标准时间不一致 946
9.2.2windows 操作系统下卸载云硬盘操作指南及故障排除 951
9.2.3Windows 弹性云主机蓝屏如何处理 952
9.2.4Windows 弹性云主机配置双网卡外网访问 953
9.2.5Windows 无法正常启动时怎样恢复数据？ 957
9.2.6 如何查看 Windows 云主机的登录日志？ 958
9.2.7CentOS 设置系统时区 962
9.2.8 启动 Tomcat 时报错，提示 80 端口被占用怎么办？ 963
9.2.9 怎样实现 Windows 云主机文件共享？ 964
9.2.10 输入法无法使用怎么办？ 967
9.2.11 Windows 云主机配置文件共享和网络磁盘映射方法 971
9.2.12Windows 云主机如何保持会话连接长时间不断开？ 977
9.2.13Windows 弹性云主机不能复制粘贴内容？ 979
9.2.14Windows 弹性云主机不能复制粘贴内容？ 982
9.2.15 怎样设置 Windows 弹性云主机首选语言 984
9.3 操作系统类（Linux） 989
9.3.1 弹性云主机新内核启动失败如何设置使用第二内核启动 990
9.3.2 为什么操作系统实际版本和购买时的镜像版本不一致？ 990
9.3.3 怎样配置 Linux 分析工具：atop 和 kdump 992
9.3.4Linux 操作系统云主机安装 Docker 支持的操作系统版本 1002
9.3.5Ubuntu 操作系统如何设置默认启动内核 1003
9.3.6 弹性云主机 IP 地址丢失处理方法 1006
9.3.7 内核参数 kernel.unknown_nmi_panic 配置错误导致 Linux 弹性云主机异常重启 1007
9.3.8pip 安装软件时出现错误：command 'gcc' failed with exit status 1 1009
9.3.9CentOS 云主机根目录设置成 777 权限后如何恢复？ 1010
9.3.10Linux 弹性云主机执行命令或启动服务时出现错误：Cannot allocate memory 1011
9.3.11 无法编辑 fstab 文件怎么办 1012
9.3.12 非 root 用户切 root 用户时，连接超时怎么办？ 1013
9.3.13 修改/etc/security/limits.conf 文件，重启后不生效怎么办？ 1014

9.3.14 使用 taskset 命令让进程运行在指定 CPU 上 1015
9.3.15Web 访问超时且系统日志打印：nf_conntrack: table full, dropping packet 1017
9.3.16 弹性云主机中/etc/rc.local 开机启动脚本不生效怎么办？ 1019
9.3.17emergency mode（紧急模式）问题处理方法 1020
9.4 云主机错误状态处理方式 1028
9.4.1 弹性云主机配置的邮件无法正常发送 1028
9.4.2 如何解决云平台创建云主机状态错误 1029
9.4.3 云主机错误状态及解决方案 1030
9.5 远程连接报错 1033
9.5.1 远程连接云主机出现蓝屏 1033
9.5.2 远程连接 Windows 云主机报错：出现身份验证错误，要求的函数不受支持 1034
9.5.3 天翼云 Windows 云主机远程连接时出现报错： 此计算机无法连接到远程计算机 1036
9.5.4 远程连接 Windows 云主机报错：没有远程登录的权限 1042
9.5.5 登录 Windows 云主机提示“内部错误”怎么办？ 1044
9.5.6 远程连接 Windows 云主机报错：无法验证此远程计算机的身份 1045
9.5.7Windows 云主机远程连接时出现报错：RDP 连接断开，出现内部错误 1046
9.5.8Windows 云主机远程连接时出现报错：您的连接已丢失 1049
9.5.9Windows 云主机报错：您的凭据无法工作 1052
9.6 网络配置 1060
9.6.1resolv 文件被重置怎么办？ 1060
9.6.2Linux 系统重启后/etc/hosts 自动添加主机名解析 1061
9.6.3CentOS 7 重启后 dhclient 未运行，导致无法获取 IP 1062
9.6.4Linux 系统使用多网卡时网卡名称出现漂移 1064
9.6.5 同一子网的两块网卡均绑定弹性公网 IP 1066
9.7 磁盘空间管理 1068
9.7.1Linux 操作系统云主机中 buffer 和 cache 占用内存怎么办？ 1068
9.7.2Linux 文件系统提示：Read-only file system 1069
9.7.3Linux 操作系统云主机磁盘分区提示空间不足怎么办？ 1071
9.7.4 文件已经删除，但空间未释放怎么办？ 1073
9.7.5Inode 节点耗尽导致无法创建新文件问题处理 1073
9.7.6Linux 如何创建 swap 分区/swap 文件 1075
9.8GPU 驱动故障 1078
9.8.1GPU 驱动异常怎么办？ 1078
9.8.2GPU 实例启动异常，查看系统日志发现 NVIDIA 驱动空指针访问怎么办？ 1079
9.8.3GPU 设备显示异常 1080
9.8.4T4 GPU 设备显示异常 1081
9.8.5G 系列弹性云主机 GPU 驱动故障 1082
9.8.6GPU 驱动不可用 1084
9.9SSH 连接 1085
9.9.1 怎样长时间保持 SSH 会话连接不断开？ 1085
9.9.2 /etc /passwd 文件损坏导致云主机登录失败怎么办？ 1086
9.9.3 怎样设置允许或禁止用户/IP 通过 SSH 连接云主机？ 1089

9.9.4Linux 启动 sshd 服务出现/var/empty/sshd 无法访问的解决方案 1090
9.9.5CentOS 7 修改 SSH 默认端口后无法连接怎么办？ 1092
9.9.6 开启 UseDNS 导致 SSH 连接缓慢怎么办？ 1094
9.9.7 怎样禁用 SSH 密码方式连接云主机？ 1095
9.10 多用户登录 1096
9.10.1 配置多用户登录后，普通用户登录闪屏怎么办？ 1096
9.10.2Windows 云主机如何配置多用户登录？（Windows 2008） 1097
9.10.3Windows 云主机如何配置多用户登录？（Windows 2012） 1108
9.10.4 多用户登录 Windows 主机时无法打开浏览器 1119
9.11 密码与密钥对 安装 IIS 服务 1120
9.11.1Linux 云主机怎样切换密钥登录为密码登录？ 1120
9.11.2 Linux 云主机如何进入单用户模式重置 root 密码 1122
9.11.3 密钥对方式购买的弹性云主机，使用私钥文件获取登录密码失败 1130
9.11.4 如何更换密钥对 1131
9.11.5 使用密钥文件无法正常登录 Linux 弹性云主机 1132
9.11.6 Linux 云主机重置密码 1133
9.11.7 无法导入密钥对 1135
9.12 安装 IIS 服务 1138
9.12.1 怎样做网页定向？ 1138
9.12.2 IIS 服务修改已绑定的网站域名 1142
9.12.3Windows 云主机上安装 IIS 服务 1143
10 最佳实践 1149
10.1 创建云主机最佳实践 1149
10.3 基于 Tomcat 构建 Java web 环境（CentOS 7.4） 1162
10.4 基于弹性云主机部署 WordPress 的安全防护 1172
10.5 典型网站类场景 1
10.6 镜像部署 windows 环境 1
10.7 麒麟系统云主机配置图形化界面 1
10.8 下载并安装 SQL Server 1
10.9Mysql 数据库自建，使用，问题排查最佳实践 1
10.10 线下 OpenStack 环境导入公有云环境 1
10.11 创建双栈云主机 1
10.12 弹性云主机配置 1
10.13 弹性云主机之间迁移最佳实践 1
10.14Linux 服务器 SSH 登录的安全加固 1
10.16Linux 系统进入单用户模式 1
10.17 修改云主机默认远程端口 1
10.18 设置 Windows 操作系统首选语言 1
10.19ECS 安全组实践（入方向规则） 1
10.20 搭建微信公众号后台 1
10.21 网络性能测试的最佳实践 1
10.22 网站性能测试的最佳实践 1
10.23 网站搭建指导 1
10.24 手动部署 Java Web 1

10.25 基于天翼云 ECS 搭建 Oracle RAC 集群 1
10.26 手动搭建 LNMP 环境 1
10.27 搭建 FTP 1
10.28 快速构建 FTP 站点（Windows） 1
10.29 快速构建 FTP 站点（Linux） 1
10.30 手工部署 RabbitMQ（CentOS 7.4） 1
10.31 跨账号同区域迁移云主机 1
10.32 手工搭建 Ghost 博客（Ubuntu 20.04） 1
10.33 手工安装宝塔面板（CentOS7.2） 1
10.34 快速部署 SD-WAN vCPE 最佳实践 1
10.35 通过配置监控 Agent 实现指定进程监控 1
10.36 部署 Palworld 幻兽帕鲁服务器最佳实践 1
10.36.1 部署 Palworld 幻兽帕鲁服务器最佳实践（Ubuntu） 1
10.36.2 部署 Palworld 幻兽帕鲁服务器最佳实践（Windows） 1
10.36.3Palworld 幻兽帕鲁世界参数修改最佳实践（Ubuntu） 1
10.36.4Palworld 幻兽帕鲁世界参数修改最佳实践（Windows） 1
10.36.5Palworld 幻兽帕鲁更新游戏最佳实践 1
10.36.6 已有天翼云 Ubuntu 云主机，搭建幻兽帕鲁服务器 1
10.36.7 已有天翼云 Windows 云主机，搭建幻兽帕鲁服务器 1
10.37 大模型学习机最佳实践 1
10.37.1 一键部署与登录 Llama 3/Llama 2 大模型学习机 1
10.37.2 部署与登录 Llama 3/Llama 2 大模型学习机 1
10.37.3 已有天翼云云主机，部署与登录 Llama 3/Llama 2 大模型学习机 1
10.37.4 修改大模型学习机登录密码 1
10.37.5 大模型学习机服务启停最佳实践 1
10.37.6 大模型学习机图像生成最佳实践 1
10.37.7 大模型学习机文本生成最佳实践 1
10.37.8 大模型学习机文本生成模型微调最佳实践 1
10.37.9 使用大模型学习机微调自己的文本生成模型 1
10.37.10 使用大模型学习机微调自己的图像生成模型 1
10.38AIGC 实践 1
10.38.1 在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本 1
10.38.2 在天翼云使用 vLLM 运行 DeepSeek 的最佳实践-32B 等版本 1
10.38.3 基于 DeepSeek 和 Chroma 构建个性化知识库的最佳实践 1
10.38.4 基于 OpenWebUI+Ollama+DeepSeek 实现大模型私有知识库零代码构建的最佳实践 1
10.38.5 基于 LLaMA-Factory 微调 DeepSeek-R1-Distill-Qwen-7B 最佳实践 1
10.38.6 基于 CPU 的 AI 推理加速技术在天翼云 EMR 云主机上的应用 1
10.38.7 基于 Open WebUI 实现 DeepSeek API 调用及外部大模型 API 接入最佳实践 1
10.38.8AnythingLLM 协同 Ollama 启用 DeepSeek 大模型最佳实践 1
10.38.9CherryStudio 协同 Ollama 启用 DeepSeek 大模型 1
10.38.10 天翼云 CPU 实例部署 DeepSeek-R1 模型最佳实践 1
10.38.11Open WebUI 快速入门 1

10.38.12 天翼云 GPU 云主机构建高性能 Deepseek 集群最佳实践-32B 等版本 ... 1
10.38.13 使用云主机启动模版快速部署 DeepSeek 的最佳实践 ... 1
10.38.14 弹性伸缩搭配 GPU 云主机水平扩展最佳实践 ... 1
10.38.15 基于 Anything LLM 实现云主机外挂 Embedding 模型、向量数据库的最佳实践指南 ... 1
10.38.160 大模型推理性能测试实践 ... 1
10.38.161pen WebUI 及 vLLM 版本升级最佳实践 ... 1
10.39 基于弹性云主机的视频处理应用加速实践 ... 1

1 产品动态

2024 年 2 月

时间节点	功能名称	功能描述	相关文档
2024/02/01	通用计算型云主机 s8r 上线	通用计算型云主机 s8r 正式上线。	通用计算型 s8r 规格

2023 年 12 月

时间节点	功能名称	功能描述	相关文档
2023/12/25	网络增强型云主机 c7ne 上线	网络增强型云主机 c7ne 正式上线。	网络增强型 c7ne 规格
2023/12/21	支持云主机一键挂载 NAS 文件系统	支持从控制台便捷且快速地为云主机一键挂载 NAS 文件系统。	一键挂载文件系统

时间节点	功能名称	功能描述	相关文档
2023/12/01	支持云主机热变配	热变配，即开机变配，是指云主机在运行中状态下进行变更规格的操作。	变更配置概述云主机热变配规则说明
2023/12/01	云主机支持实例描述及修改 hostname 名称	支持用户对主机 hostname 名称进行修改，并为主机添加实例描述说明	修改云主机主机名修改实例描述

2023 年 11 月

时间节点	功能名称	功能描述	相关文档
2023/11/30	云主机支持按需实例删除保护功能	如果您购买的主机计费类型为按需计费，那么您可以为您的主机开启删除保护功能，通过使用该功能，可以避免由于操作不当，导致主机被误删。	开启和关闭实例保护功能
2023/11/30	支持云主机解绑同订单弹性IP、卸载同订单云硬盘	优化了云主机解绑弹性 IP、卸载云硬盘功能，支持解绑或卸载和云主机一起购买的弹性 IP、云硬盘。	卸载云硬盘解绑弹性IP

时间节点	功能名称	功能描述	相关文档
2023/11/29	支持云主机创建后绑定/解绑密钥对	对已创建的Linux云主机实例，支持绑定/解绑密钥对功能。 绑定密钥对功能，支持对未绑定密钥对的Linux云主机实例新绑定密钥对，支持对已绑定密钥对的Linux云主机实例更换密钥对。 解绑密钥对功能，支持对已绑定密钥对的Linux云主机实例解绑密钥对。	为云主机绑定/更换密钥对 为云主机解绑密钥对
2023/11/27	云主机系统盘、数据盘加密	云硬盘加密是指在创建云主机（选择系统盘、数据盘）时或者单独创建数据盘时勾选加密选项，创建完成后云主机操作系统内数据会在云硬盘所在宿主机被自动加密，无需自建和维护密钥管理基础设施，即可轻松保护数据的隐私性和自主性。	加密云硬盘

2023 年 9 月

时间节点	功能名称	功能描述	相关文档
2023/09/14	支持 VNC 的远程发送命令功能	Linux 弹性云主机支持 VNC 的远程发送命令。	Linux 弹性云主机远程登录（VNC 方式）
2023/09/07	支持主子用户权限功能	支持用户开通、管理主子账号。	登入及开通方式

2023 年 8 月

时间节点	功能名称	功能描述	相关文档
2023/08/31	自研 8 代云主机上线	自研 8 代云主机 s8、m8、c8 正式上线。	8 代机概述

2023 年 6 月

时间节点	功能名称	功能描述	相关文档

时间节点	功能名称	功能描述	相关文档
2023/06/20	新增磁盘增强型云主机	新增实例，上线磁盘增强型云主机 d3。	本地盘云主机产品规格

2023 年 5 月

时间节点	功能名称	功能描述	相关文档

时间节点	功能名称	功能描述	相关文档
2023/05/31	支持不指定宿主机创建云主机	支持不指定宿主机创建 2 系列、6 系列、7 系列云主机。	创建云主机
2023/05/03	新增超高 IO 型云主机	新增实例，上线超高 IO 型云主机 ip3	本地盘云主机产品规格

2023 年 4 月

时间节点	功能名称	功能描述	相关文档
2023/04/15	新增鲲鹏通用型云主机 ks1	搭载鲲鹏 920 处理器，云主机实例共享 CPU。提供基础计算能力，可满足基础业务上云需求。	鲲鹏通用型 ks1 云主机规格
2023/04/15	新增海光通用型云主机 hs1	搭载 Hygon C86 处理器，云主机实例共享 CPU。提供基础计算能力，可满足基础业务上云需求。	海光通用型 hs1 云主机规格
2023/04/15	新增飞腾通用型云主机 fs1	搭载飞腾 S2500 处理器，云主机实例共享 CPU。提供基础计算能力，可满足基础业务上云需求。	飞腾通用型 fs1 云主机规格

2023 年 3 月

时间节点	功能名称	功能描述	相关文档
2023/03/30	私有镜像支持设置内存	私有镜像支持最大最小内存的设置。	通过镜像文件创建系统盘镜像
2023/03/30	公共镜像分类与排序	创建云主机界面，公共镜像分类进行了细分，且操作系统版本下拉框进行了排序。	通过镜像创建云主机

时间节点	功能名称	功能描述	相关文档
2023/03/30	支持无对象存储创建私有镜像	多 AZ 的资源池，如果没有部署对象存储，支持通过云主机创建私有镜像。	通过云主机创建系统盘镜像

2023 年 2 月

时间节点	功能名称	功能描述	相关文档

时间节点	功能名称	功能描述	相关文档
2023/02/28	GPU 云主机按量计费模式上线	GPU 云主机支持按量计费的计费模式。	GPU 云主机按量计费模式
2023/02/28	GPU 云主机变配功能上线	GPU 云主机支持同规格族内的 GPU 云主机进行升配、降配操作。	GPU 云主机变配
2023/02/18	公共镜像增加 CTyunOS	公共镜像支持 CTyunOS 2.0.1 版本。	镜像服务产品规格

时间节点	功能名称	功能描述	相关文档
2023/02/07	GPU 云主机新增规格：P8A	GPU 云主机海口 2 上线 A100 直通 GPU 云主机。	GPU 云主机产品规格

2023 年 1 月

时间节点	功能名称	功能描述	相关文档

时间节点	功能名称	功能描述	相关文档
2023/01/30	镜像服务支持内网 yum 源与 ntp 服务	内网 yum 源支持无公网 ip 的场景下访问 yum 源，ntp 服务可供用户进行时间同步。	内网 yum 源与 NTP 配置
2023/01/30	通过云主机快照创建系统盘镜像	系统盘镜像的镜像源增加云主机快照。	通过云主机快照创建系统盘镜像

时间节点	功能名称	功能描述	相关文档
2023/01/15	公共镜像增加 Fedora CoreOS	公共镜像支持 Fedora CoreOS 36 版本。	镜像服务产品规格

2022 年 11 月

时间节点	功能名称	功能描述	相关文档

时间节点	功能名称	功能描述	相关文档
2022/11/30	GPU 云主机新增规格：G7、PI7	在华东 1 上线 A10 直通、虚拟化 GPU 云主机。	GPU 云主机产品规格
2022/11/29	上线通用型 s7 云主机	搭载第三代英特尔®至强®可扩展处理器，可提供最大 200 万 PPS 网络收发包，最大 12Gbps 内网带宽。	s7 弹性云主机规格
2022/11/29	上线计算型 c7 云主机	搭载第三代英特尔®至强®可扩展处理器，可提供最大 1100 万 PPS 网络收发包，最大 40Gbps 内网带宽。	c7 弹性云主机规格

时间节点	功能名称	功能描述	相关文档
2022/11/29	上线内存型 m7 云主机	搭载第三代英特尔®至强®可扩展处理器，可提供最大 1100 万 PPS 网络收发包，最大 40Gbps 内网带宽。	m7 弹性云主机规格

2022 年 9 月

时间节点	功能名称	功能描述	相关文档

时间节点	功能名称	功能描述	相关文档
2022/09/30	上线内存型 m6 云主机	搭载第二代英特尔®至强®可扩展处理器，可提供最大 1000 万 PPS 网络收发包，最大 40Gbps 内网带宽。	m6 弹性云主机规格
2022/09/30	上线计算型 c6 云主机	搭载第二代英特尔®至强®可扩展处理器，可提供最大 1000 万 PPS 网络收发包，最大 40Gbps 内网带宽。	c6 弹性云主机规格
2022/09/30	上线通用型 s6 云主机	搭载第二代英特尔®至强®可扩展处理器，可提供最大 100 万 PPS 网络收发包，最大 6Gbps 内网带宽。	s6 弹性云主机规格

2022 年 7 月

时间节点	功能名称	功能描述	相关文档
2022/07/16	上线飞腾系列云主机	首次上线飞腾系列云主机，包括鲲鹏计算增强型 fc1、鲲鹏内存优化型 fm1。	飞腾系列云主机规格
2022/07/16	上线鲲鹏系列云主机	首次上线鲲鹏系列云主机，包括鲲鹏计算增强型 kc1、鲲鹏内存优化型 km1。	鲲鹏系列云主机规格
2022/07/16	上线海光系列云主机	首次上线海光系列云主机，包括海光计算增强型 hc1、鲲鹏内存优化型 hm1。	海光系列云主机规格

2 新手指引

2.1 操作导航

云服务概述

弹性云主机（CT-ECS，Elastic Cloud Server）是一种可随时获取、弹性可扩展的计算服务。云主机由 CPU、内存、镜像、云硬盘等组成，同时结合 VPC、安全组、数据多副本保存等能力，打造一个既高效又可靠安全的计算环境，确保服务持久稳定运行。

弹性云主机的开通和使用非常便捷、高效，您只需要指定 CPU、内存、镜像等配置信息。开通完成后，您就可以像使用本地 PC 或物理服务器一样在云上使用弹性云主机，同时您也可以按需随时调整弹性云主机的规格配置。

- 了解弹性云主机的使用限制与使用须知，请参考：使用须知、使用限制。
- 弹性云主机的资源规格限制，请参考管理配额。

怎样选择实例规格类型

- 了解弹性云主机应用场景，请参考：弹性云主机应用场景。
- 了解弹性云主机实例规格类型，请参考：规格说明。

了解云主机的计费模式

按量计费

按量计费是一种后付费模式，即先使用再付费，系统会根据云主机的实际使用情况按秒计费。

说明
自 2021 年 4 月 1 日 4:00 起，按需弹性云主机、按需云硬盘、按需独享带宽将计费单元由小时调整为秒。

例如您在 13:30:30 创建了一台按需付费主机，相关资源包括计算资源（vCPU 和内存）、镜像和云盘（系统盘），然后在 13:55:30 释放实例，则：结算周期为 13:00:00～14:00:00，在 13:30:30～13:55:30 间产生计费，该结算周期内的计费时长为 1500 秒。期间费用=实例小时价格/3600*1500。

详细信息请查看按量计费。

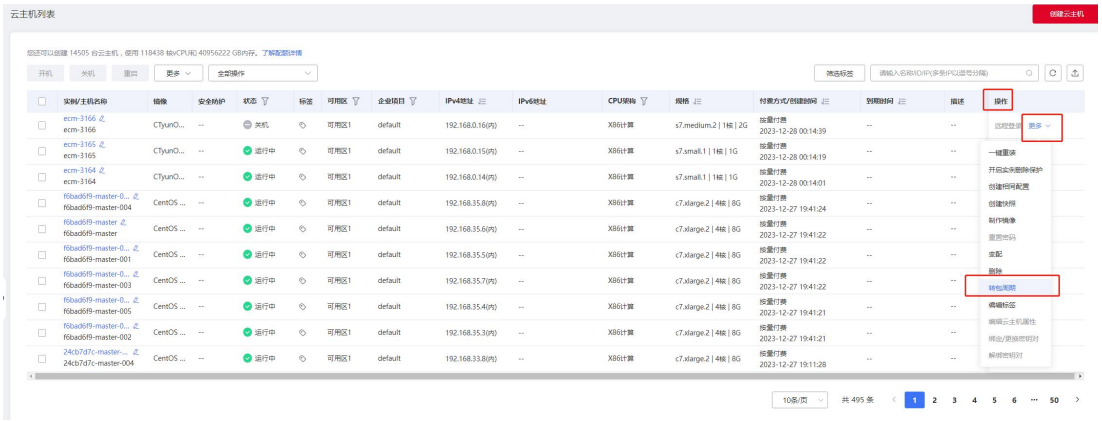
包年包月计费

包年包月计费是一种预付费购买方式，即先付费再使用。按订单的购买周期计费。用户可通过包年包月计费提前预留资源，同时享受比按量计费更大的价格优惠。

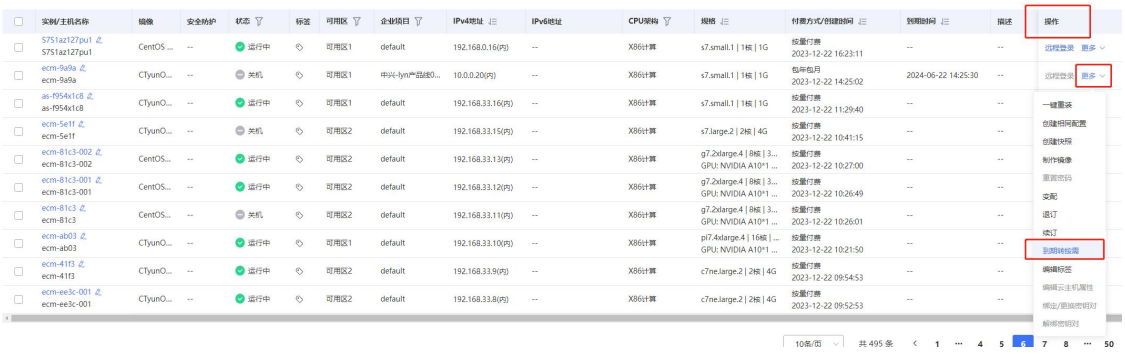
详细信息请查看包年包月计费模式。

更改计费方式

如果您需要长期使用当前弹性云服务器，可以将按需购买的云服务器转为包周期计费模式，节省开支，可在云主机控制台进行操作，如下图所示。



如果您希望更灵活的使用弹性云主机，可以将包周期计费模式转成按需付费，可在云主机控制台进行操作，如下图所示。



价格计算器

您可以使用 ECS 价格计算器快速了解弹性云服务器配置费用。

购买并登录云主机

- 在管理控制台购买 ECS
- 新手入门：购买并登录 Windows 弹性云主机
- 新手入门：购买并登录 Linux 弹性云主机

2.2 新手入门：购买并登录 Windows 弹性云主机

购买弹性云主机

1. 登录天翼云“控制中心(控制台)”，在页面上方选择用户对应购买弹性云主机的地域，例如“福州 3”。选择“计算”大类下的“弹性云主机”产品。或是在天翼云官网首页点击“立即订购”。
2. 单击“创建云主机”。

配置弹性云主机实例规格

1. 选择计费模式“包年/包月”或是“按量付费”。
2. 完成页面必填项的选择，例如地域、主机名称、规格、镜像类型、存储等。

说明

创建云主机过程，请您根据规格特性，选择符合业务需求的云主机规格。请参考：弹性云主机规格说明。

使用云主机需要您支付相应费用，详细请见：按量计费模式和包年包月计费模式。

创建云主机需要配置系统盘与数据盘，如果您对存储性能有要求，请参考：磁盘类型及性能介绍。

您需要为存储支付相应的费用，详细请见：云硬盘计费模式。

网络配置

1. 选择弹性云主机对应的网络配置选项含网卡、扩展网卡、安全组、弹性 IP 等，根据页面提示完成操作。
2. 点击“下一步：高级配置”按钮。

说明安全组是控制云主机访问策略的组件，相关概念与操作，请参考：安全组概述。

如果您有访问公网的需求，请您使用弹性 IP，相关概念与应用场景，请参考：弹性 IP 产品简介。

您需要为弹性 IP 支付相应的费用，请参考：弹性 IP 计费概述。

高级配置

1. 选择“密码”或“密钥对”登录方式。
2. 选择创建密码的方式。
3. 点击“下一步：确认配置”按钮。

确认配置并购买

1. 确认配置项正确后点击“立即购买”。
2. 进入订单创建页面，可刷新查看订单创建状态。

登录 Windows 弹性云主机

查看登录 Windows 弹性云主机。

操作教学视频

您可参考视频专区中的云主机创建视频执行操作：视频专区-弹性云主机。

2.3 新手入门：购买并登录 Linux 弹性云主机

购买弹性云主机

1.登录天翼云“控制中心(控制台)”，在页面上方选择用户对应购买弹性云主机的地域，例如“福州 3”。选择“计算”大类下的“弹性云主机”产品。或是在天翼云官网首页点击“立即订购”。
2.单击“创建云主机”。

配置弹性云主机实例规格

1.选择计费模式“包年/包月”或是“按量付费”。
2.完成页面必填项的选择，例如地域、主机名称、规格、镜像类型、存储等。

说明
创建云主机过程，请您根据规格特性，选择符合业务需求的云主机规格。请参考：弹性云主机规格说明。
使用云主机需要您支付相应费用，详细请见：按量计费模式和包年包月计费模式。
创建云主机需要配置系统盘与数据盘，如果您对存储性能有要求，请参考：磁盘类型及性能介绍。
您需要为存储支付相应的费用，详细请见：云硬盘计费模式。

网络配置

1. 选择弹性云主机对应的网络配置选项含网卡、扩展网卡、安全组、弹性 IP 等，根据页面提示完成操作。

2. 点击“下一步：高级配置”按钮。

说明

安全组是控制云主机访问策略的组件，相关概念与操作，请参考：安全组概述。

如果您有访问公网的需求，请您使用弹性 IP，相关概念与应用场景，请参考：弹性 IP 产品简介。

您需要为弹性 IP 支付相应的费用，详细请见：弹性 IP 计费概述。

高级配置

1. 选择“密码”或“密钥对”登录方式。

2. 选择创建密码的方式。

3. 点击“下一步：确认配置”按钮。

确认配置并购买

1. 确认配置项正确后点击“立即购买”。

2. 进入订单创建页面。

3. 订单创建完成后相应的云主机实例也创建完成。此时可刷新查看订单创建状态。

登录 Linux 弹性云主机

查看登录 Linux 弹性云主机。

操作教学视频

您可参考视频专区中的云主机创建视频执行操作：视频专区-弹性云主机。

3DeepSeek 专题

3.1DeepSeek 专题导航

图解：DeepSeek 与公有云深度融合

- 从基础设施到智能中枢：DeepSeek 如何重塑公有云服务价值链
- 高性能 GPU 云主机助力 DeepSeek 深度应用
- 天翼云 SD-WAN 与 DeepSeek 超强联动，开启云上高效互联新时代

实践指南：DeepSeek 驱动高效能云生态

- GPU 云主机/弹性云主机：零基础搭建 DeepSeek 云端环境指南
- GPU 物理机：物理机搭建 DeepSeek 指南
- SD-WAN 跨境：SD-WAN 助力 DeepSeek 模型定向加速
- 智算容器：云容器引擎与 DeepSeek 融合实践
- 函数计算：天翼云函数计算与 DeepSeek 大模型

Q&A：典型问题解析与策略应对

- 常见问题解答

3.2 图解 DeepSeek 与公有云深度融合

3.2.1 从基础设施到智能中枢：DeepSeek 如何重塑公有云服务价值链

天翼公有云×DeepSeek 产品架构和技术优势

天翼云 State Cloud

中国电信 CHINA TELECOM | 天翼云 State Cloud

天翼公有云 x DeepSeek



Cloud4DeepSeek & DeepSeek4Cloud

产品架构



技术优势

高性能算力底座

- 高性能网络：基于RoCE和InfiniBand协议的RDMA网络，支持万卡GPU 400Gbps接入和多租隔离。
- 高性能存储：最高提供千万IOPS、TBps 吞吐，亚毫秒级时延的高性能并行文件存储。

云原生智算套件

- 高效资源管理：支持GPU、NPU、RDMA等异构资源统一纳管，可视化构建万卡规模容器集群与智能化监控，帮助用户快速构建AI生产环境。
- 智能调度策略：支持GPU/NPU共享调度、拓扑感知调度和故障感知等调度策略，降本增效，提高资源利用率和 AI 任务效率。

高性能集合通信库

- 高可靠通信基座：智能链路容错机制，实现秒级训练中异常链路自动切换与全链路事件溯源。
- 自研流体重力算法：智能识别网络态势，动态调整数据流量，突破网络短板问题，集合通信性能提升10%+。

全链路智算底座可观测

- 监控指标：支持超300 项核心监控指标，包括计算GPU、RDMA 和网络端口等IaaS对象的单节点及集群多粒度观测能力。
- 任务级观测：基于torch profiler深度定制，实现训练过程节点、网卡对的全链路观测。

天翼云
State Cloud

天翼公有云智算底座主推产品



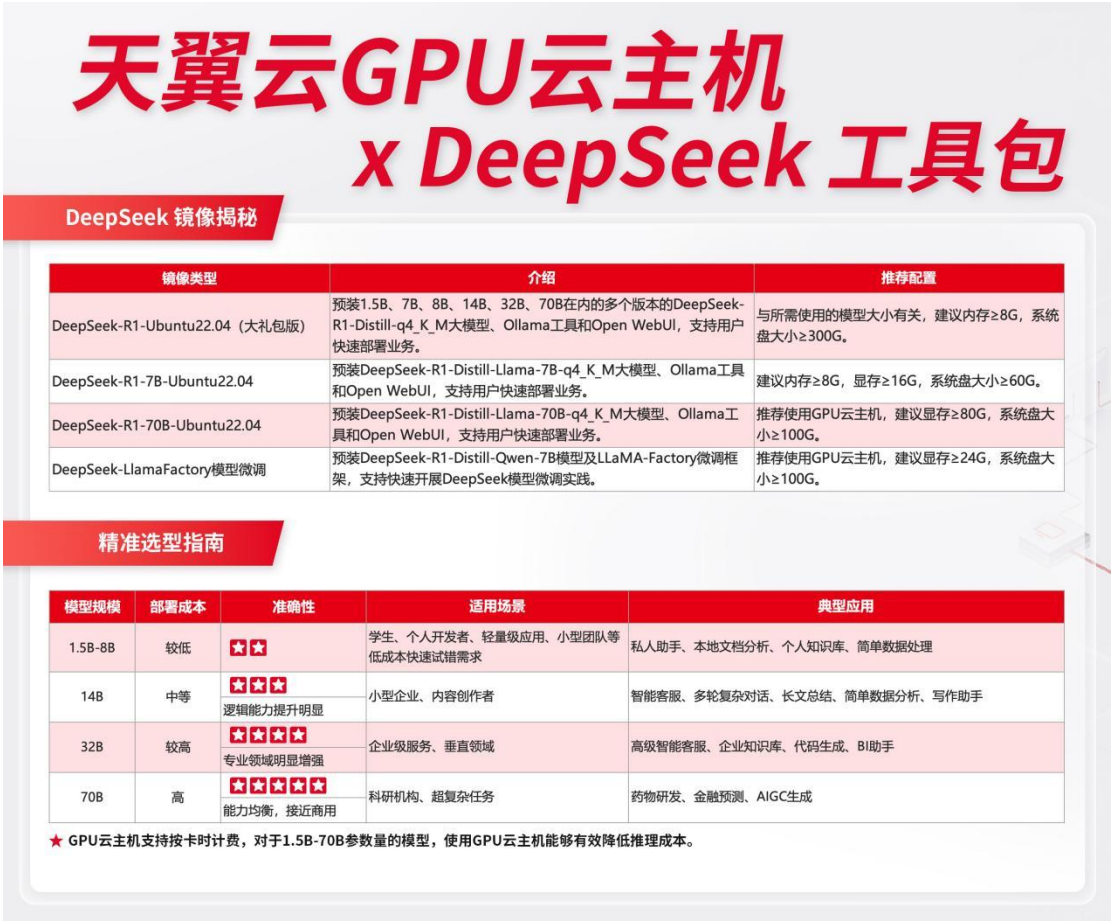
天翼公有云×DeepSeek 方案



返回 DeepSeek 专题导航。

3.2.2 高性能 GPU 云主机助力 DeepSeek 深度应用





返回 DeepSeek 专题导航。

3.2.3 天翼云 SD-WAN 与 DeepSeek 超强联动，开启云上高效互联新时代



返回 DeepSeek 专题导航。

3.3 实践指南：DeepSeek 驱动高效能云生态

3.3.1GPU 云主机/弹性云主机：零基础搭建 DeepSeek 云端环境指南

3.3.1.1 在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本

什么是 DeepSeek

DeepSeek 是一个基于 Transformer 架构的大型语言模型（LLM），由深度求索（DeepSeek）公司开发。它能够处理自然语言理解、生成、翻译、问答等多种任务。在目前大模型主流榜单中，DeepSeek-V3 在开源模型中位列榜首，与世界上最先进的闭源模型不分伯仲。在对话模型典型任务方面的评测效果如下：

	Benchmark (Metric)	DeepSeek V3	DeepSeek V2.5 0905	Qwen2.5 72B-Inst	Llama3.1 405B-Inst	Claude-3.5 Sonnet-1022	GPT-4o 0513
	Architecture	MoE	MoE	Dense	Dense	-	-
	# Activated Params	37B	21B	72B	405B	-	-
	# Total Params	671B	236B	72B	405B	-	-
English	MMLU (EM)	88.5	80.6	85.3	**88.6**	88.3	87.2
	MMLU-Redux (EM)	**89.1**	80.3	85.6	86.2	88.9	88.0
	MMLU-Pro (EM)	75.9	66.2	71.6	73.3	**78.0**	72.6
	DROP (3-shot F1)	**91.6**	87.8	76.7	88.7	88.3	83.7
	IF-Eval (Prompt Strict)	86.1	80.6	84.1	86.0	**86.5**	84.3
	GPQA-Diamond (Pass@1)	59.1	41.3	49.0	51.1	**65.0**	49.9
	SimpleQA (Correct)	24.9	10.2	9.1	17.1	28.4	**38.2**
	FRAMES (Acc.)	73.3	65.4	69.8	70.0	72.5	**80.5**
	LongBench v2 (Acc.)	**48.7**	35.4	39.4	36.1	41.0	48.1
Code	HumanEval-Mul (Pass@1)	**82.6**	77.4	77.3	77.2	81.7	80.5
	LiveCodeBench (Pass@1-COT)	**40.5**	29.2	31.1	28.4	36.3	33.4
	LiveCodeBench (Pass@1)	**37.6**	28.4	28.7	30.1	32.8	34.2
	Codeforces (Percentile)	**51.6**	35.6	24.8	25.3	20.3	23.6
	SWE Verified (Resolved)	42.0	22.6	23.8	24.5	**50.8**	38.8
	Aider-Edit (Acc.)	79.7	71.6	65.4	63.9	**84.2**	72.9
	Aider-Polyglot (Acc.)	**49.6**	18.2	7.6	5.8	45.3	16.0
Math	AIME 2024 (Pass@1)	**39.2**	16.7	23.3	23.3	16.0	9.3
	MATH-500 (EM)	**90.2**	74.7	80.0	73.8	78.3	74.6
	CNMO 2024 (Pass@1)	**43.2**	10.8	15.9	6.8	13.1	10.8
Chinese	CLUEWSC (EM)	90.9	90.4	**91.4**	84.7	85.4	87.9
	C-Eval (EM)	**86.5**	79.5	86.1	61.5	76.7	76.0
	C-SimpleQA (Correct)	**64.1**	54.1	48.4	50.4	51.3	59.3

当前 DeepSeek 模型有多个版本，参数量从 1.5B、70B 到 671B 不等，适用于不同的应用场景和计算资源。参数量越多，资源消耗越多，使用者可以根据自己实际需求选择不同模型版本。

注意

- 为维护良好的网络环境和公共秩序，禁止任何用户或机构利用天翼云平台从事违法活动、传播不良信息或实施违反社会公序良俗的行为。若发生此类行为，天翼云将依法采取必要措施，停止相关服务功能，并依法追究责任。
- 模型由 DeepSeek 公司提供，模型的结果仅供参考，并可能因不同的环境、数据或操作条件而有所不同，我们不对模型的任何结果负责。

快速体验 DeepSeek

Ollama 是一个专为本地运行大语言模型（LLMs）而设计的工具，支持多种模型格式，并提供了简单易用的命令行接口。这里，我们以 DeepSeek-r1-7b 模型为例，讲解如何在天翼云使用 Ollama 运行 DeepSeek 模型，读者可以根据自己实际需求修改云主机模型和云主机配置。

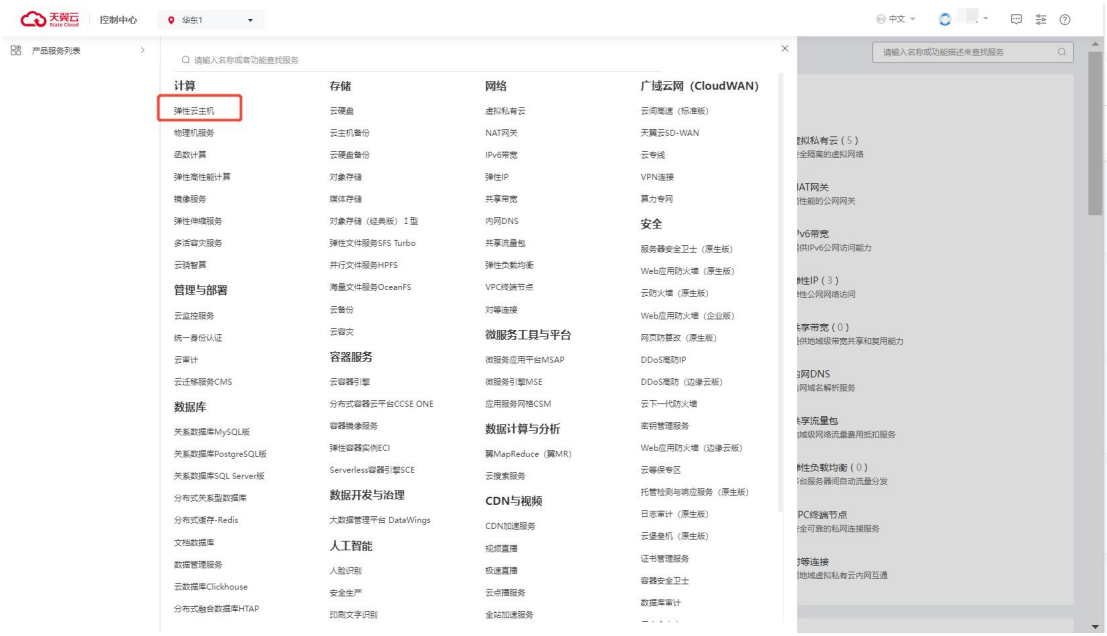
步骤一：创建 GPU 云主机

1. 进入创建云主机页面

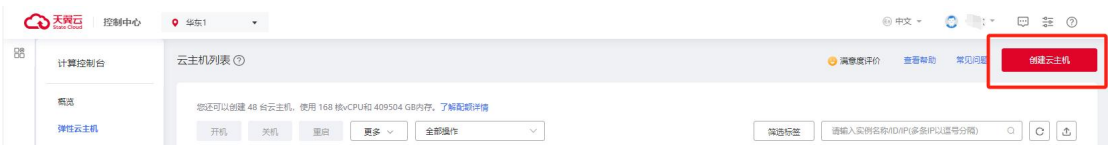
a. 点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



b. 单击“产品服务列表>弹性云主机”，进入主机列表页。



c. 单击“创建云主机”，进入弹性云主机创建页。



2. 进行基础配置

a. 根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、

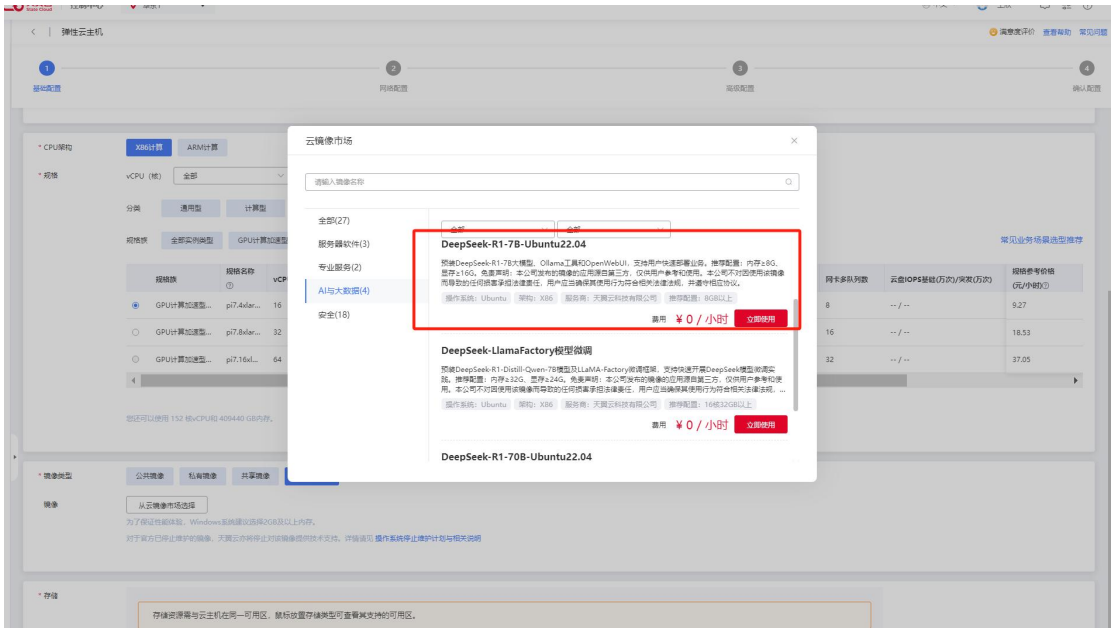
“实例名称”、“主机名称”等。

b.选择规格。此处选择"CPU 架构"为"X86"、"分类"GPU 加速/AI 加速型"、"规格族"为"GPU 计算加速型 pi7"、"规格"为"pi7.4xlarge.4"。

c.选择镜像。“镜像类型”选择“镜像市场”,在云镜像市场中选择预置了 DeepSeek R1 模型的 DeepSeek-R1-7B-Ubuntu22.04 镜像。

注意

本镜像推荐配置：内存≥8G、显存≥16G。



注意

目前也提供了预装 DeepSeek-R1-70B 模型的镜像，如您有需求，也可在云镜像市场中进行选择。

d.设置云盘类型和大小。

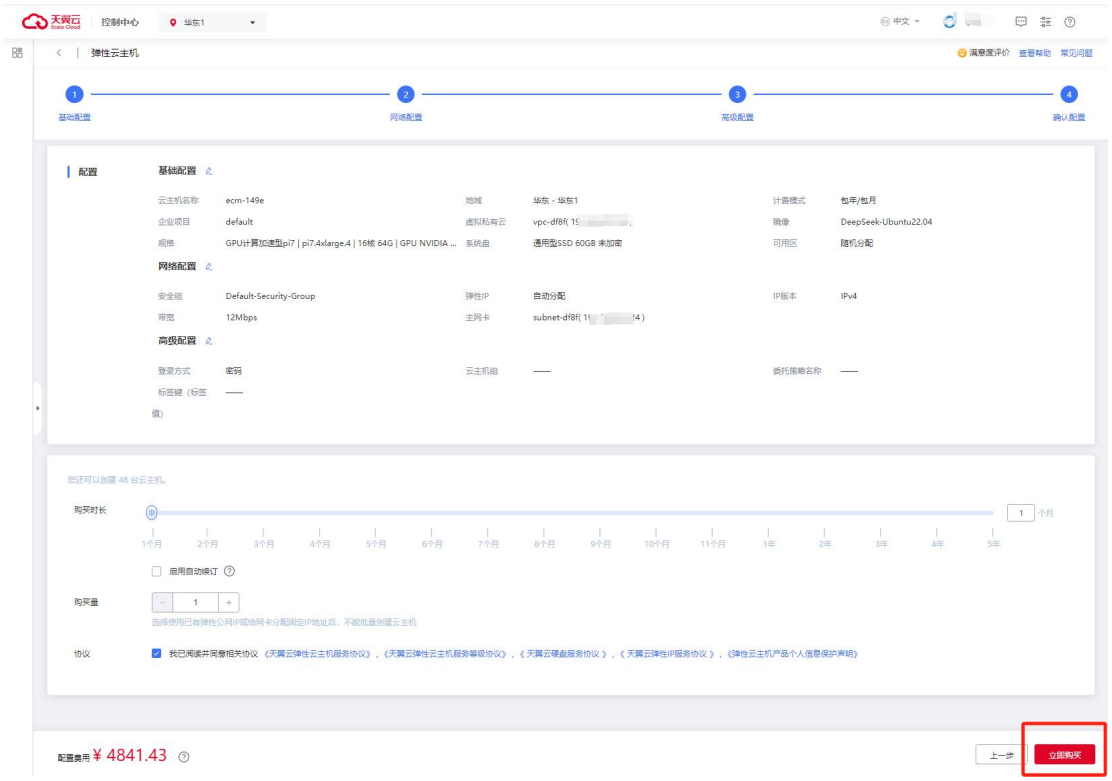


3. 网络及高级配置

设置网络，包括"网卡"、"安全组"，同时配备 "弹性 IP" 用于下载和访问模型；

设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。

4. 确认并支付



步骤二：使用 DeepSeek 模型

1. 通过 web 界面进行模型交互

注意

镜像自带的 ollama 工具监听 127.0.0.1:11434、webui 监听 0.0.0.0:3000 端口，云主机默认不对外开放任何端口访问，请按需开放端口访问规则，避免数据泄露。

a. 放行云主机安全组的 3000 端口。具体操作请参考添加安全组规则-弹性云主机-用户指南-安全-安全组-配置安全组规则 - 天翼云。



b. 访问DeepSeek模型的可视化界面。登录地址为:http://{公网ip地址}:3000。

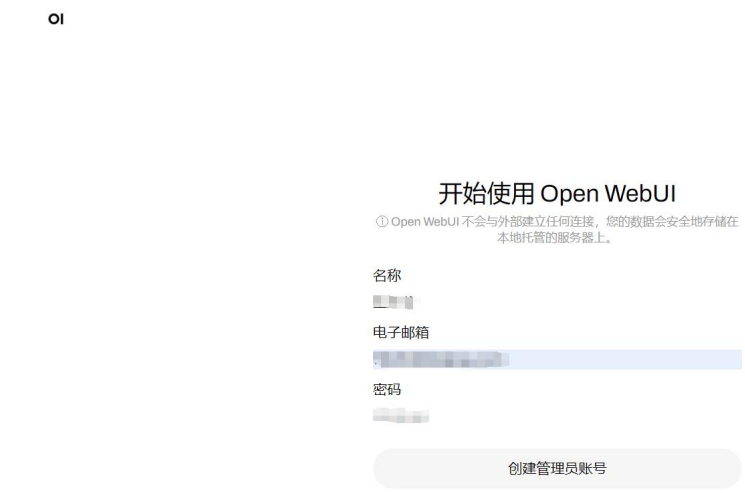
注意

云主机全自动安装 DeepSeek 模型和可视化界面，请等待云主机启动 5 分钟后，再访问登录界面。

首次登录页面如下：

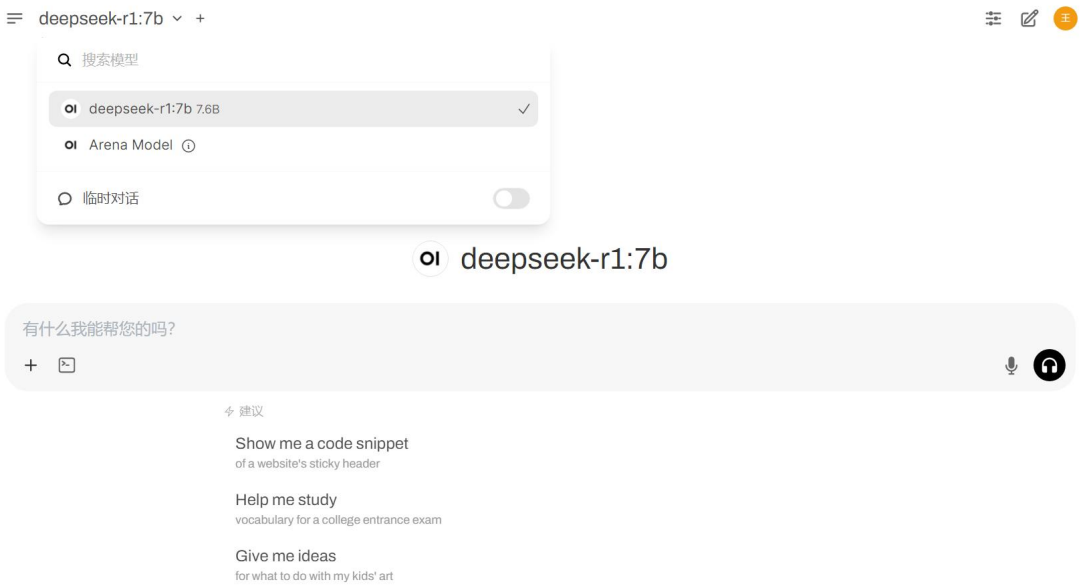


c. 注册管理员账号。

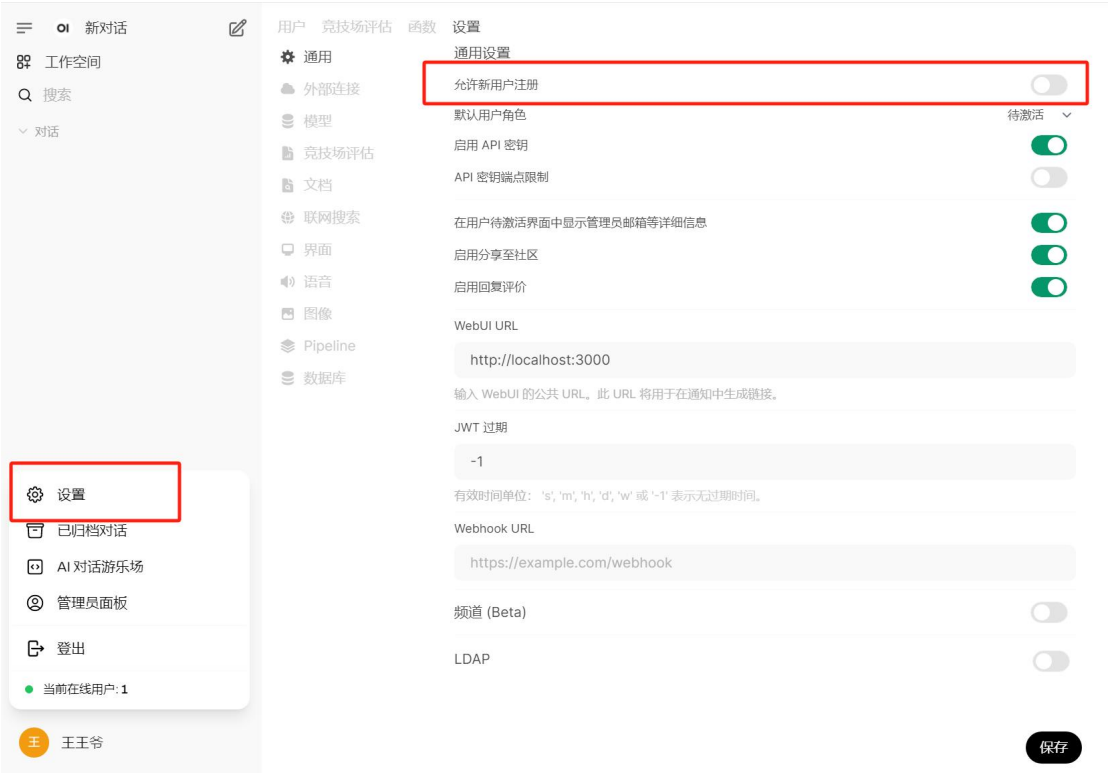


d. 使用设置。

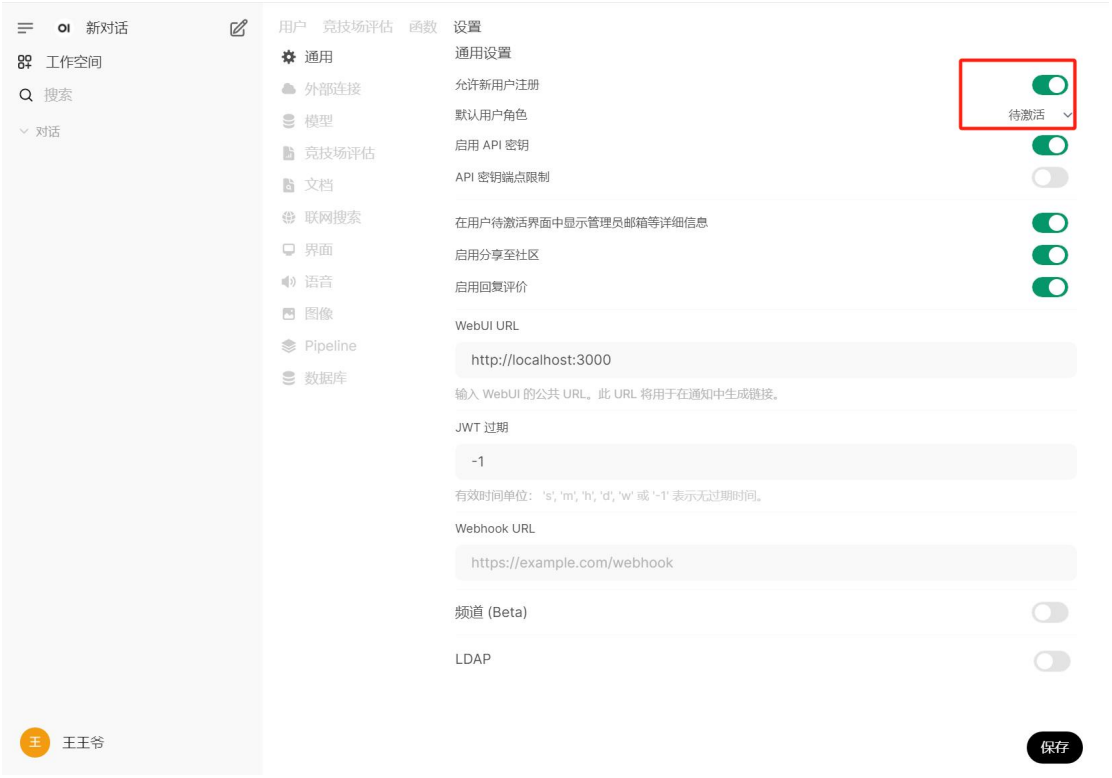
刷新进入首页，在模型下拉列表中，选择刚部署的 DeepSeek:7b 模型。



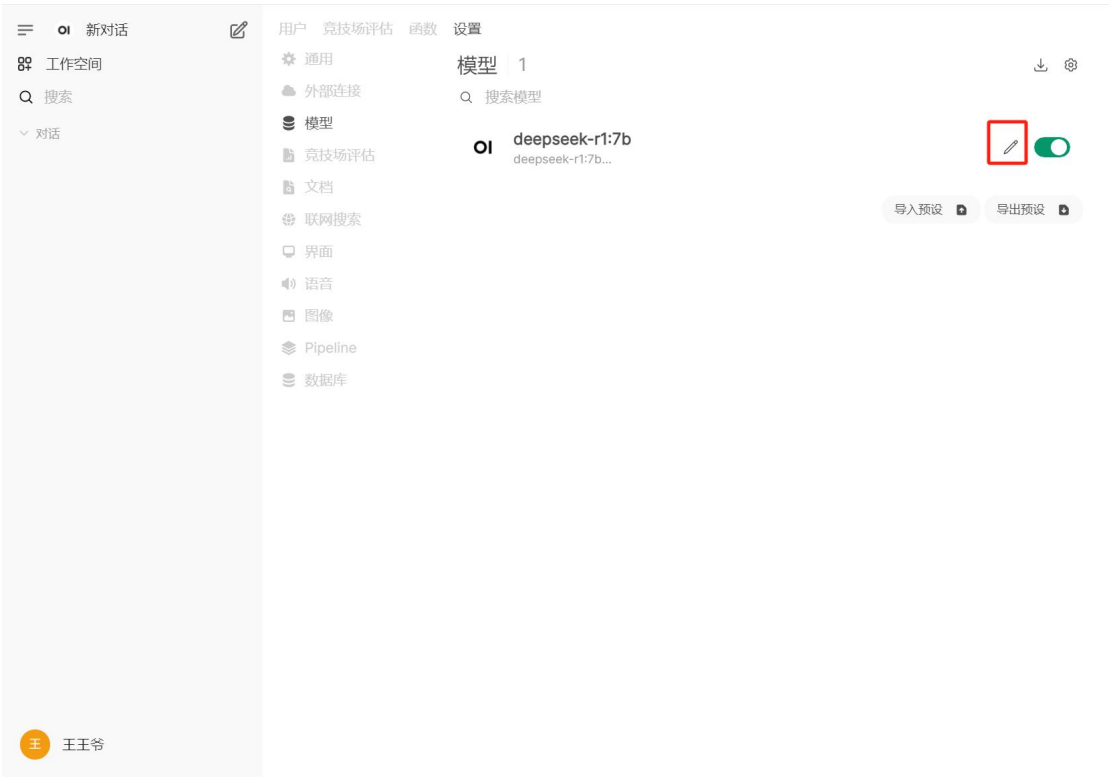
点击左下角进入设置页面，如果您不想开放其他用户注册使用，则需要关闭 “允许用户注册” 功能。

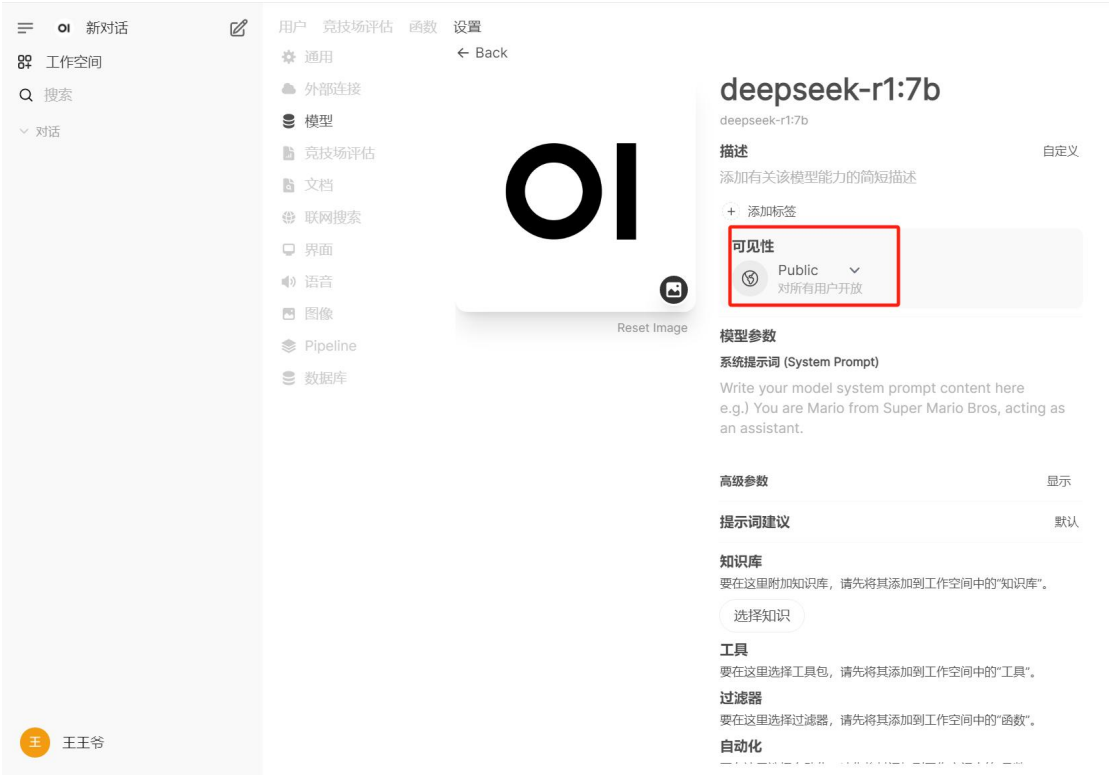


如果您允许用户注册，还可以设置用户注册之后的行为，例如选择新用户注册后默认用户角色为“用户”/“待激活” 等，需要管理员手动激活。

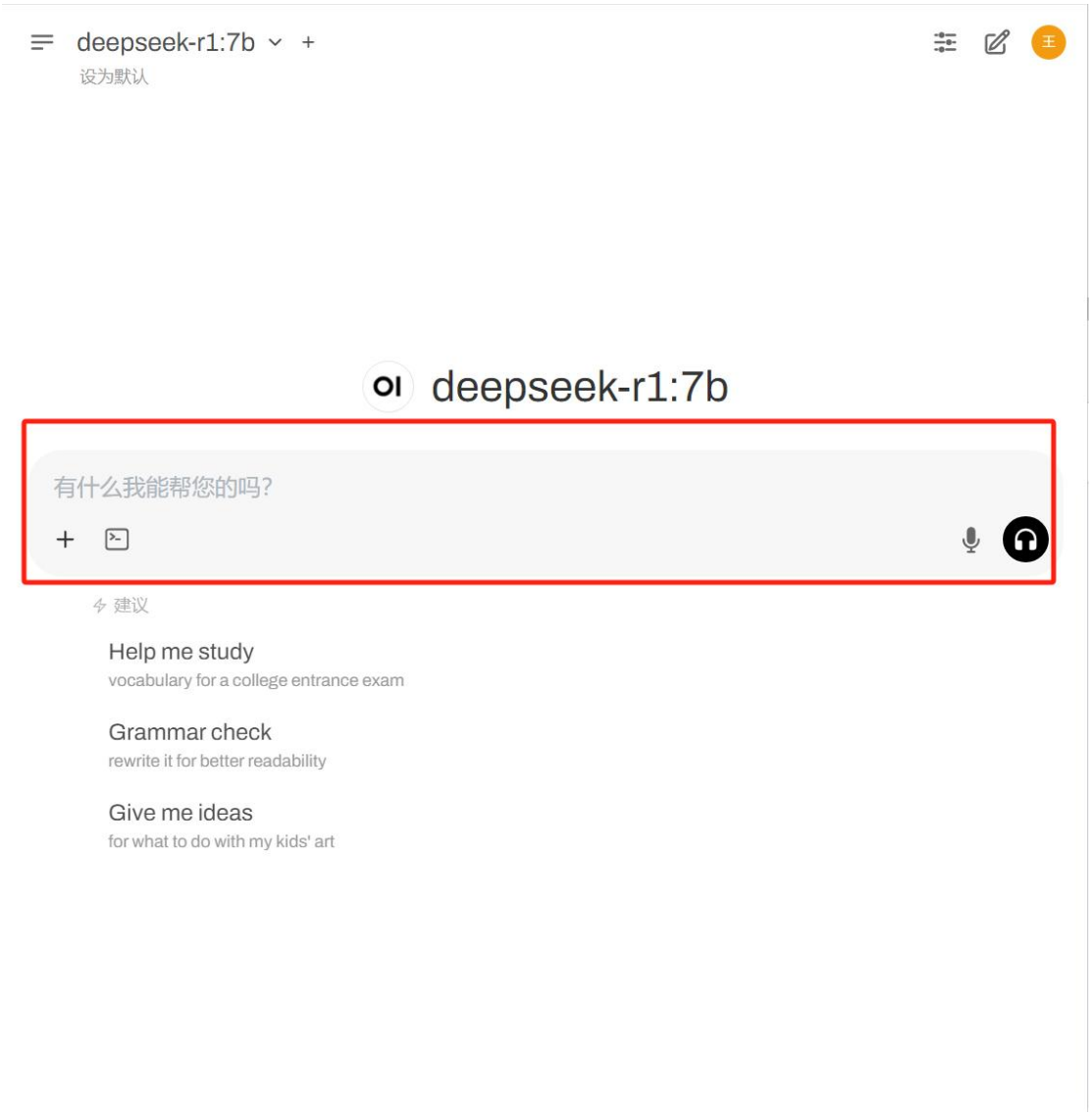


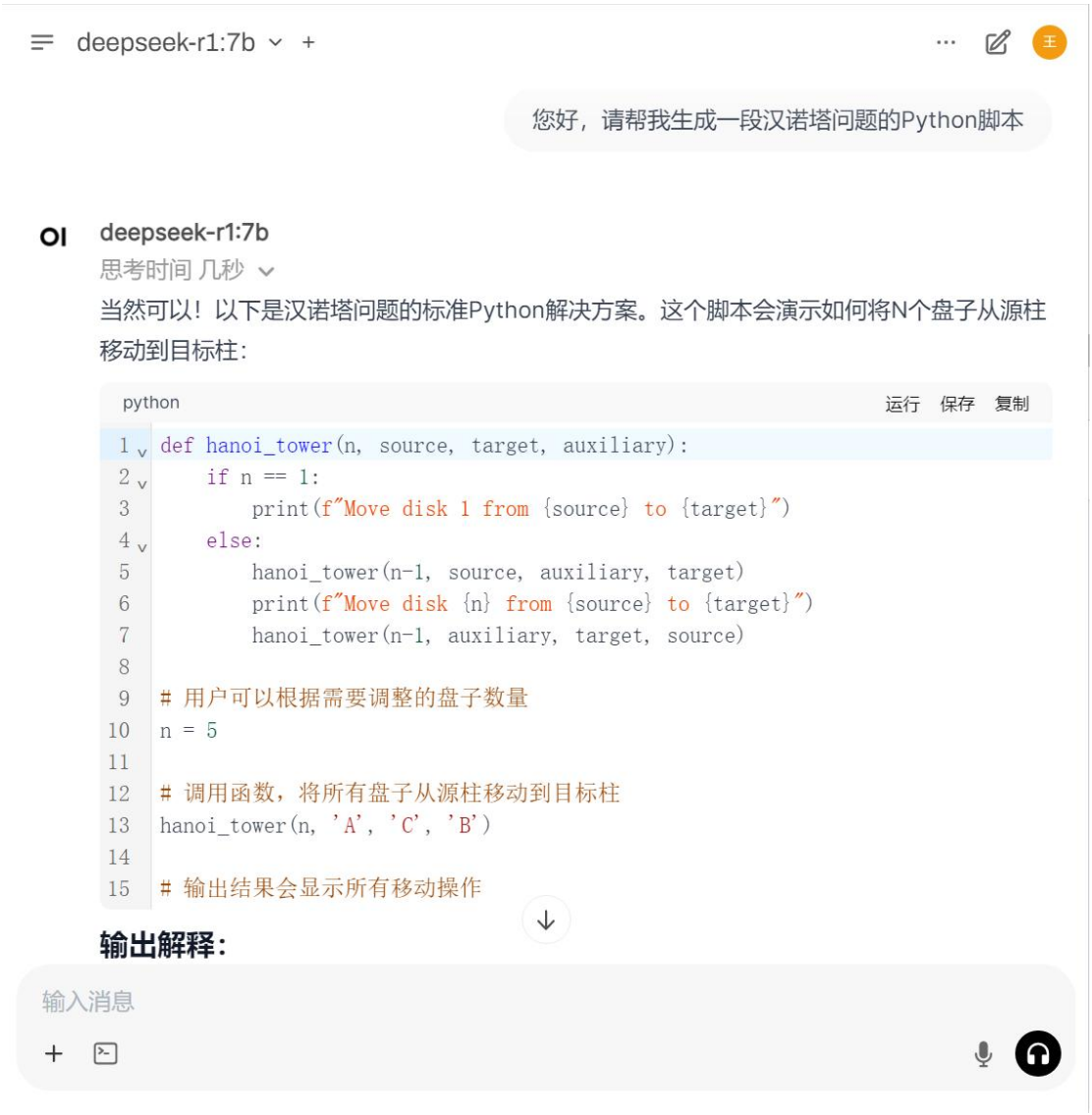
设置模型可见性。多用户模式下，建议把模型设置为"Public"。





e. 使用 DeepSeek 模型进行模型推理。


deepseek-r1:7b
设为默认
deepseek-r1:7b
有什么我能帮您的吗?
建议
Help me study
vocabulary for a college entrance exam
Grammar check
rewrite it for better readability
Give me ideas
for what to do with my kids' art



自定义部署 DeepSeek

步骤一：规格选型

不同版本的模型参数量不同，参数量越多，资源消耗越多。不同规模的企业、客户有不同诉求，客户可根据自己的实际需求，选择所需镜像和模型版本。参数较少的 DeepSeek 模型可以在 CPU 上正常运行、如 1.5B、7B 模型，更大的模型需要 GPU 才能获得良好的体验。天翼云提供了不同模型的硬件配置建议，如下：

模型	推荐显卡数				
	T4	V100	V100S	A10	A100

DeepSeek-R1-1.5B	1 卡 16G（1*16G）	1 卡 32G（1*32G）	1 卡 32G（1*32G）	1 卡 24G（1*24G）	1 卡 40G（1*40G）
DeepSeek-R1-7B	2 卡 32G（2*16G）	1 卡 32G（1*32G）	1 卡 32G（1*32G）	1 卡 24G（1*24G）	1 卡 40G（1*40G）
DeepSeek-R1-8B	2 卡 32G（2*16G）	1 卡 32G（1*32G）	1 卡 32G（1*32G）	1 卡 24G（1*24G）	1 卡 40G（1*40G）
DeepSeek-R1-14B	4 卡 64G（4*16G）	2 卡 64G（2*32G）	2 卡 64G（2*32G）	2 卡 48G（2*24G）	2 卡 80G（2*40G）
DeepSeek-R1-32B	-	4 卡 128G（4*32G）	4 卡 128G（4*32G）	4 卡 96G（4*24G）	4 卡 160G（4*40G）
DeepSeek-R1-70B	-	-	-	-	4 卡 160G（4*40G）

对应的 GPU 云主机规格请参见 NVIDIA GPU 云主机-GPU 云主机-产品简介-产品规格 - 天翼云。

步骤二：云主机开通

创建弹性云主机-弹性云主机-快速入门 - 天翼云

创建配备 GPU 驱动的 GPU 云主机（Linux）-GPU 云主机-用户指南-创建 GPU 云主机 - 天翼云

步骤三：手动部署 DeepSeek

1. 安装 ollama

a. 下载安装脚本

使用非天翼云 DeepSeek 镜像时，需要手动安装 ollama，这里以 amd64 为例，将如下脚本保存到本地，命名为 install_ollama.sh。

```
#!/bin/sh# This script installs Ollama on Linux.# It detects the current operating system architecture and installs the appropriate version of Ollama.

set -eu


red="$( (/usr/bin/tput bold || :; /usr/bin/tput setaf 1 || :) 2>&-)"

plain="$( (/usr/bin/tput sgr0 || :) 2>&-)"

status() { echo ">>> $*" >&2; }error() { echo "${red}ERROR:${plain} $*"; exit 1; }warning() { echo "${red}WARNING:${plain} $*"; }


TEMP_DIR=$(mktemp -d)cleanup() { rm -rf $TEMP_DIR; }trap cleanup EXIT

available() { command -v $1 >/dev/null; }require() {

    local MISSING=''

    for TOOL in $*; do

        if ! available $TOOL; then

            MISSING="$MISSING $TOOL"

        fi

    done


    echo $MISSING

}


[ "$(uname -s)" = "Linux" ] || error 'This script is intended to run on Linux only.'
```

```
ARCH=$(uname -m)case "$ARCH" in
    x86_64) ARCH="amd64" ;;
    aarch64|arm64) ARCH="arm64" ;;
    *) error "Unsupported architecture: $ARCH" ;;esac

IS_WSL2=false

KERN=$(uname -r)case "$KERN" in
    *icrosoft*WSL2 | *icrosoft*wsl2) IS_WSL2=true;;
    *icrosoft) error "Microsoft WSL1 is not currently supported. Please
use WSL2 with 'wsl --set-version <distro> 2'" ;;
    *) ;;esac

VER_PARAM="${OLLAMA_VERSION:+?version=$OLLAMA_VERSION}"

SUDO=if [ "$(id -u)" -ne 0 ]; then
    # Running as root, no need for sudo
    if ! available sudo; then
        error "This script requires superuser permissions. Please re-run
as root."
    fi

    SUDO="sudo"fi
```

```
NEEDS=$(require curl awk grep sed tee xargs)if [ -n "$NEEDS" ]; then
    status "ERROR: The following tools are required but missing:"
    for NEED in $NEEDS; do
        echo "  - $NEED"
    done
    exit 1fi
for BINDIR in /usr/local/bin /usr/bin /bin; do
    echo $PATH | grep -q $BINDIR && break || continuedone
OLLAMA_INSTALL_DIR=$(dirname ${BINDIR})
if [ -d "$OLLAMA_INSTALL_DIR/lib/ollama" ] ; then
    status "Cleaning up old version at $OLLAMA_INSTALL_DIR/lib/ollama"
    $SUDO rm -rf "$OLLAMA_INSTALL_DIR/lib/ollama"fi
status "Installing ollama to $OLLAMA_INSTALL_DIR"$SUDO install -o0 -g0
-m755 -d $BINDIR$SUDO install -o0 -g0 -m755 -d "$OLLAMA_INSTALL_DIR"
status "Downloading Linux ${ARCH} bundle"
curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}.tgz${VE
R_PARAM}" | \
    $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"if
[ "$OLLAMA_INSTALL_DIR/bin/ollama" != "$BINDIR/ollama" ] ; then
    status "Making ollama accessible in the PATH in $BINDIR"
    $SUDO ln -sf "$OLLAMA_INSTALL_DIR/ollama" "$BINDIR/ollama"fi
```

```
# Check for NVIDIA JetPack systems with additional downloadsif [ -f
/etc/nv_tegra_release ] ; then
    if grep R36 /etc/nv_tegra_release > /dev/null ; then
        status "Downloading JetPack 6 components"
        curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}-jetpack
6.tgz${VER_PARAM}" | \
            $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"
    elif grep R35 /etc/nv_tegra_release > /dev/null ; then
        status "Downloading JetPack 5 components"
        curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}-jetpack
5.tgz${VER_PARAM}" | \
            $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"
    else
        warning "Unsupported JetPack version detected.  GPU may not be
supported"
    fifi
install_success() {
    status 'The Ollama API is now available at 127.0.0.1:11434.'
    status 'Install complete. Run "ollama" from the command line.'
}trap install_success EXIT
# Everything from this point onwards is optional.
```

```
configure_systemd() {
    if ! id ollama >/dev/null 2>&1; then
        status "Creating ollama user..."
        $SUDO useradd -r -s /bin/false -U -m -d /usr/share/ollama ollama
    fi
    if getent group render >/dev/null 2>&1; then
        status "Adding ollama user to render group..."
        $SUDO usermod -a -G render ollama
    fi
    if getent group video >/dev/null 2>&1; then
        status "Adding ollama user to video group..."
        $SUDO usermod -a -G video ollama
    fi

    status "Adding current user to ollama group..."
    $SUDO usermod -a -G ollama $(whoami)

    status "Creating ollama systemd service..."
    cat <<EOF | $SUDO tee /etc/systemd/system/ollama.service >/dev/null
[Unit]
Description=Ollama Service
After=network-online.target

[Service]
```

```
ExecStart=$BINDIR/ollama serve

User=ollama

Group=ollama

Restart=always

RestartSec=3

Environment="PATH=$PATH"


[Install]

WantedBy=default.target

EOF

    SYSTEMCTL_RUNNING="$(systemctl is-system-running || true)"

    case $SYSTEMCTL_RUNNING in

        running|degraded)

            status "Enabling and starting ollama service..."

            $SUDO systemctl daemon-reload

            $SUDO systemctl enable ollama


            start_service() { $SUDO systemctl restart ollama; }

            trap start_service EXIT

            ;;

        *)

            warning "systemd is not running"

            if [ "$IS_WSL2" = true ]; then
```

```
                warning "see
https://learn.microsoft.com/en-us/windows/wsl/systemd#how-to-enable-
systemd to enable it"
            fi
            ;;
    esac
}
if available systemctl; then
    configure_systemdfi
# WSL2 only supports GPUs via nvidia passthrough# so check for nvidia-smi
to determine if GPU is availableif [ "$IS_WSL2" = true ]; then
    if available nvidia-smi && [ -n "$(nvidia-smi | grep -o "CUDA Version:
[0-9]*\.[0-9]*")" ]; then
        status "Nvidia GPU detected."
    fi
    install_success
    exit 0fi
# Don't attempt to install drivers on Jetson systemsif [ -f
/etc/nv_tegra_release ] ; then
    status "NVIDIA JetPack ready."
    install_success
    exit 0fi
# Install GPU dependencies on Linuxif ! available lspci && ! available
lshw; then
```

```
    warning "Unable to detect NVIDIA/AMD GPU. Install lspci or lshw to
automatically detect and install GPU dependencies."
    exit 0fi
check_gpu() {
    # Look for devices based on vendor ID for NVIDIA and AMD
    case $1 in
        lspci)
            case $2 in
                nvidia) available lspci && lspci -d '10de:' | grep -q
'NVIDIA' || return 1 ;;
                amdgpu) available lspci && lspci -d '1002:' | grep -q
'AMD' || return 1 ;;
            esac ;;
        lshw)
            case $2 in
                nvidia) available lshw && $SUDO lshw -c display -numeric
-disable network | grep -q 'vendor: .* \[10DE\]' || return 1 ;;
                amdgpu) available lshw && $SUDO lshw -c display -numeric
-disable network | grep -q 'vendor: .* \[1002\]' || return 1 ;;
            esac ;;
        nvidia-smi) available nvidia-smi || return 1 ;;
    esac
}
if check_gpu nvidia-smi; then
    status "NVIDIA GPU installed."
```

```
    exit 0fi
if ! check_gpu lspci nvidia && ! check_gpu lshw nvidia && ! check_gpu
lspci amdgpu && ! check_gpu lshw amdgpu; then
    install_success
    warning "No NVIDIA/AMD GPU detected. Ollama will run in CPU-only
mode."
    exit 0fi
if check_gpu lspci amdgpu || check_gpu lshw amdgpu; then
    status "Downloading Linux ROCm ${ARCH} bundle"
    curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}-rocm.tg
z${VER_PARAM}" | \
        $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"

    install_success
    status "AMD GPU ready."
    exit 0fi

CUDA_REPO_ERR_MSG="NVIDIA GPU detected, but your OS and Architecture are
not supported by NVIDIA.  Please install the CUDA driver manually
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/"# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#rhel-7-centos-7# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#rhel-8-rocky-8# ref:
```

```
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html#rhel-9-rocky-9# ref: https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html#fedorainstall_cuda_driver_yum() {
    status 'Installing NVIDIA repository...'

    case $PACKAGE_MANAGER in
        yum)
            $SUDO $PACKAGE_MANAGER -y install yum-utils
            if curl -I --silent --fail --location "https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(uname -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo" >/dev/null ; then
                $SUDO $PACKAGE_MANAGER-config-manager --add-repo https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(uname -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo
            else
                error $CUDA_REPO_ERR_MSG
            fi
            ;;
        dnf)
            if curl -I --silent --fail --location "https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(uname -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo" >/dev/null ; then
                $SUDO $PACKAGE_MANAGER config-manager --add-repo https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(uname -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo
```

```
            else
                error $CUDA_REPO_ERR_MSG
            fi
            ;;
    esac

    case $1 in
        rhel)
            status 'Installing EPEL repository...'
            # EPEL is required for third-party dependencies such as dkms
and libvdpau
            $SUDO $PACKAGE_MANAGER -y install
https://dl.fedoraproject.org/pub/epel/epel-release-latest-$2.noarch.
rpm || true
            ;;
    esac

    status 'Installing CUDA driver...'

    if [ "$1" = 'centos' ] || [ "$1$2" = 'rhel7' ]; then
        $SUDO $PACKAGE_MANAGER -y install nvidia-driver-latest-dkms
    fi

    $SUDO $PACKAGE_MANAGER -y install cuda-drivers
```

```
}

# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html#ubuntu# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html#debianinstall_cuda_driver_apt() {
    status 'Installing NVIDIA repository...'
    if curl -I --silent --fail --location "https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(uname -m | sed -e 's/aarch64/sbsa/')/cuda-keyring_1.1-1_all.deb" >/dev/null ; then
        curl -fsSL -o $TEMP_DIR/cuda-keyring.deb https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(uname -m | sed -e 's/aarch64/sbsa/')/cuda-keyring_1.1-1_all.deb
    else
        error $CUDA_REPO_ERR_MSG
    fi

    case $1 in
        debian)
            status 'Enabling contrib sources...'
            $SUDO sed 's/main/contrib/' < /etc/apt/sources.list | $SUDO tee /etc/apt/sources.list.d/contrib.list > /dev/null
            if [ -f "/etc/apt/sources.list.d/debian.sources" ]; then
```

```
                $SUDO sed 's/main/contrib/' <
/etc/apt/sources.list.d/debian.sources | $SUDO tee
/etc/apt/sources.list.d/contrib.sources > /dev/null
            fi
            ;;
    esac

    status 'Installing CUDA driver...'
    $SUDO dpkg -i $TEMP_DIR/cuda-keyring.deb
    $SUDO apt-get update

    [ -n "$SUDO" ] && SUDO_E="$SUDO -E" || SUDO_E=
    DEBIAN_FRONTEND=noninteractive $SUDO_E apt-get -y install
cuda-drivers -q
}
if [ ! -f "/etc/os-release" ]; then
    error "Unknown distribution. Skipping CUDA installation."fi

. /etc/os-release

OS_NAME=$ID
OS_VERSION=$VERSION_ID

PACKAGE_MANAGER=for PACKAGE_MANAGER in dnf yum apt-get; do
```

```
    if available $PACKAGE_MANAGER; then
        break
    fidone
if [ -z "$PACKAGE_MANAGER" ]; then
    error "Unknown package manager. Skipping CUDA installation."fi
if ! check_gpu nvidia-smi || [ -z "$(nvidia-smi | grep -o "CUDA Version: [0-9]*\.[0-9]*")" ]; then
    case $OS_NAME in
        centos|rhel) install_cuda_driver_yum 'rhel' $(echo $OS_VERSION | cut -d '.' -f 1) ;;
        rocky) install_cuda_driver_yum 'rhel' $(echo $OS_VERSION | cut -c1) ;;
        fedora) [ $OS_VERSION -lt '39' ] && install_cuda_driver_yum $OS_NAME $OS_VERSION || install_cuda_driver_yum $OS_NAME '39';;
        amzn) install_cuda_driver_yum 'fedora' '37' ;;
        debian) install_cuda_driver_apt $OS_NAME $OS_VERSION ;;
        ubuntu) install_cuda_driver_apt $OS_NAME $(echo $OS_VERSION | sed 's/\.//') ;;
        *) exit ;;
    esacfi
if ! lsmod | grep -q nvidia || ! lsmod | grep -q nvidia_uvm; then
    KERNEL_RELEASE="$(uname -r)"
    case $OS_NAME in
        rocky) $SUDO $PACKAGE_MANAGER -y install kernel-devel kernel-headers ;;
```

```
        centos|rhel|amzn) $SUDO $PACKAGE_MANAGER -y install
kernel-devel-$KERNEL_RELEASE kernel-headers-$KERNEL_RELEASE ;;
        fedora) $SUDO $PACKAGE_MANAGER -y install
kernel-devel-$KERNEL_RELEASE ;;
        debian|ubuntu) $SUDO apt-get -y install
linux-headers-$KERNEL_RELEASE ;;
        *) exit ;;
    esac

    NVIDIA_CUDA_VERSION=$($SUDO dkms status | awk -F: '/added/ { print
$1 }')
    if [ -n "$NVIDIA_CUDA_VERSION" ]; then
        $SUDO dkms install $NVIDIA_CUDA_VERSION
    fi

    if lsmod | grep -q nouveau; then
        status 'Reboot to complete NVIDIA CUDA driver install.'
        exit 0
    fi

    $SUDO modprobe nvidia
    $SUDO modprobe nvidia_uvmfi
# make sure the NVIDIA modules are loaded on boot with
nvidia-persistencedif available nvidia-persistenced; then
    $SUDO touch /etc/modules-load.d/nvidia.conf
```

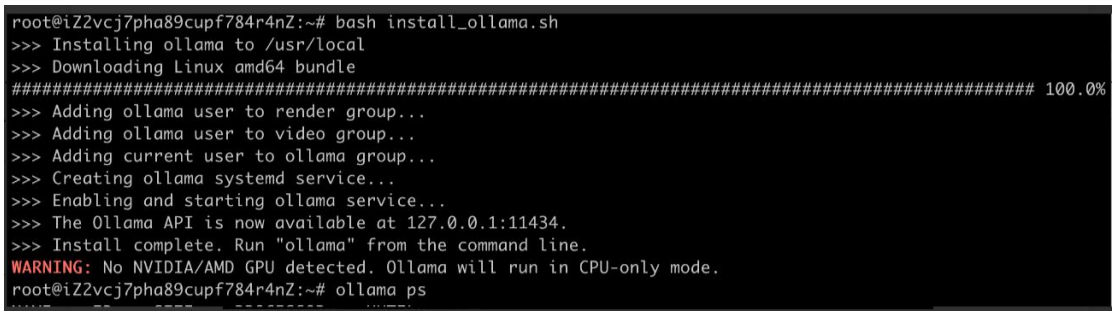
```
    MODULES="nvidia nvidia-uvm"
    for MODULE in $MODULES; do
        if ! grep -qxF "$MODULE" /etc/modules-load.d/nvidia.conf; then
            echo "$MODULE" | $SUDO tee -a
/etc/modules-load.d/nvidia.conf > /dev/null
        fi
    donefi


status "NVIDIA GPU ready."
install_success
```

b. 执行安装

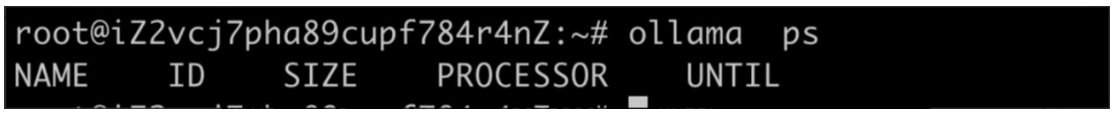
```
bash install_ollama.sh
```

如下显示则为安装成功。



c. 测试 ollama 服务安装情况

```
ollama ps
```



2. 下载模型

天翼云镜像站也为常见镜像提供了加速能力，包括 1.5B\7B\14B\70B\671B 等相关模型，可以通过如下手段体验其他模型，以 14B 模型为例：

```
mkdir deepseek-r1-14bcd deepseek-r1-14bwget
https://mirrors.ctyun.cn/ollama/models/deepseek-r1-14b/deepseek-r1-14b.ggufwget
https://mirrors.ctyun.cn/ollama/models/deepseek-r1-14b/ModelFileollama create deepseek-r1:14b -f ModelFile
```

```
root@test-deepseek-1:~# ollama create deepseek-r1:14b ModelFile
Error: accepts 1 arg(s), received 2
root@test-deepseek-1:~# ollama create deepseek-r1:14b -f ModelFile
gathering model components
copying file sha256:6e9f90f02bb3b39b59e81916e8cfce9deb45aeaeb9a54a5be4414486b907dc1e 100%
parsing GGUF
using existing layer sha256:6e9f90f02bb3b39b59e81916e8cfce9deb45aeaeb9a54a5be4414486b907dc1e
using existing layer sha256:369ca498f347f710d068cbb38bf0b8692dd3fa30f30ca2ff755e211c94768150
using existing layer sha256:f4d24e9138dd4603380add165d2b0d970bef471fac194b436ebd50e6147c6588
writing manifest
success
root@test-deepseek-1:~# ollama list
NAME                ID              SIZE      MODIFIED
deepseek-r1:14b     67711fd7c370    9.0 GB    4 seconds ago
deepseek-r1:7b      0a8c26691023    4.7 GB    3 days ago
```

如果你想体验其他模型，也可以直接通过 ollama 下载。

```
ollama pull deepseek-r1:14b
```

3. 使用 ollama 运行模型

```
ollama run deepseek-r1:14b
```

4.通过 web 界面进行交互

a.安装 open-webui

```
docker pull ghcr.io/open-webui/open-webui:main
```

b.启动容器

```
docker run -d --net=host -e PORT=3000 -e
OLLAMA_BASE_URL=http://127.0.0.1:11434 -eENABLE_SIGNUP=true -e
ENABLE_OPENAI_API=False -v open-webui:/app/backend/data --name
open-webui --restart always ghcr.io/open-webui/open-webui:main
```

后续请参考上文“快速体验 DeepSeek——步骤二：使用 deepseek 模型”。

FAQ

盘不够了，我应该怎么办？

如果您在模型部署过程中发现云盘的容量不够，可以采取如下措施：

1. 根据云硬盘扩容概述-云硬盘-用户指南-扩容云硬盘 - 天翼云对已有云盘进行扩容。

2. 新建一块数据盘并挂载，相关操作见挂载云硬盘-云硬盘-快速入门 - 天翼云、初始化数据盘-弹性云主机-快速入门 - 天翼云。

如何修改 ollama 模型的存储位置？

在 linux 环境下，ollama 默认模型存储目录是 /usr/share/ollama/.ollama/models/，我们建议您使用云硬盘独立挂载数据盘，将模型存储到数据盘中。模型存储位置是由环境变量控制的，我们需要修改 ollama 的环境变量重启服务才能修改存储目录，我们以 /data/ollama/models 目录为例：

1. 打开 ollama.service 文件

```
vi /etc/systemd/system/ollama.service
```

注意

请确保 ollama 用户组中的 ollama 用户具备访问该目录的读写权限

2. 新增相关环境变量

```
[Unit]
Description=Ollama Service
After=network-online.target

[Service]
ExecStart=/usr/local/bin/ollama serve
User=ollama
Group=ollama
Restart=always
RestartSec=3
Environment"PATH=/usr/local/cuda/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/game
s:/usr/local/games:/snap/bin"
Environment="OLLAMA_MODELS=/data/ollama/models"

[Install]
WantedBy=default.target
~
~
```

3. 重启服务

```
systemctl daemon-reloadsystemctl restart ollama
```

如何通过 vllm 部署模型？

建议使用天翼云提供的 DeepSeek 镜像，减少安装过程中可能遇到的问题。如您必须通过 vllm 部署模型，请参考如下步骤：

1.安装依赖包

```
# 安装 pipcurl https://bootstrap.pypa.io/get-pip.py -o
get-pip.pypython get-pip.py# 修改 pip 镜像源，以清华源为例 pip config set
global.index-url https://pypi.tuna.tsinghua.edu.cn/simple

# 安装依赖包 pip install modelscope==1.22.3pip install
openai==1.61.0pip install tqdm==4.67.1pip install
transformers==4.48.2pip install vllm==0.7.1pip install jinja2==3.1.0
```

2.下载模型

为了方便，我们以较小的 DeepSeek-R1-Distill-Qwen-1.5B 为例（3.5G 左右），将以下内容保存为 model_download.py 文件，参数 cache_dir 为模型的下载路径，您可以按需修改，需确保存储空间足够存放模型。

```
from modelscope import snapshot_download

model_dir =
snapshot_download('deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B',
cache_dir='/root/model-path', revision='master')
```

运行该文件，启动模型下载。

```
python model_download.py
```

3.运行模型

完成下载后，即可使用 vllm 运行模型。

```
vllm serve/root/model-path/deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B
--served-model-name DeepSeek-R1-Distill-Qwen-1.5B
--max-model-len=2048 --api-key my-token --port=8000
```

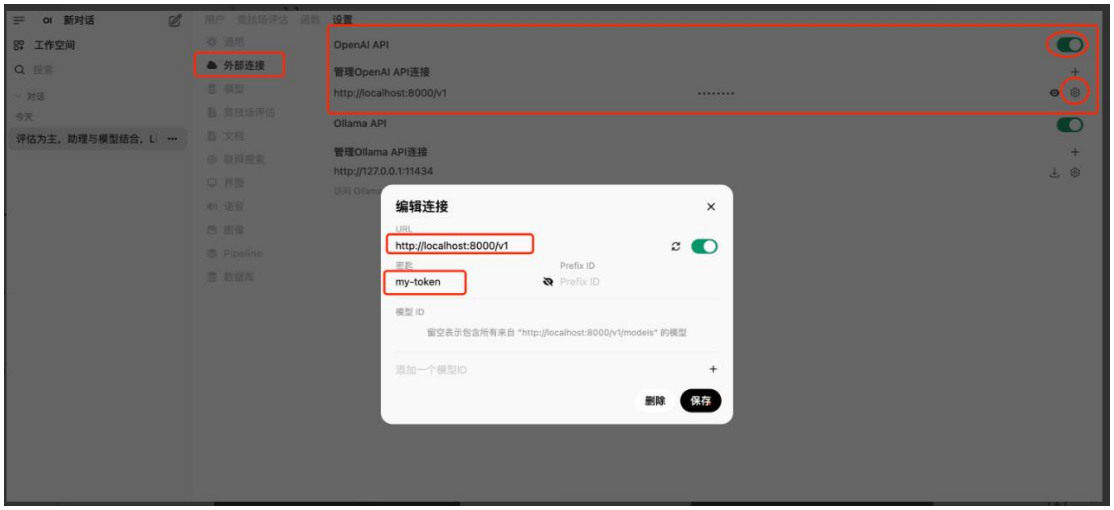
注意

其中 --api-key 参数用于模型 api 调用时认证动作，可按需修改值。

4.使用 WebUI 访问模型

基于我们已经安装、配置好的 OpenWebUI 页面，可以直接访问模型。

管理员设置中，配置 OpenAI API 访问地址，密钥就是上文中的 --api-key。



选择对应的模型，直接进行问答体验。



返回 DeepSeek 专题导航。

3.3.1.2 基于 OpenWebUI+Ollama+DeepSeek 实现大模型私有知识库零代码构建的最佳实践

前言——私有知识库作用简介

大模型私有知识库是结合大语言模型技术与企业或组织自有数据的一种知识管理和应用方案，旨在为特定用户群体提供更精准、专业且安全的知识服务。

大模型私有知识库是基于大语言模型构建的，专门服务于特定组织或个人的知识存储和检索系统。它整合组织内部的专业知识、业务数据、历史文档等，形成专属知识集合，借助大模型的强大语言理解和生成能力，为用户提供高效的知识查询和问答服务。

本教程提供一种基于开源框架的私有知识库搭建方案。如果需要构建企业级服务且对数据安全有需求，则需相关企业对该服务进行定制，以完成私有知识库的搭建。

本文介绍的方法采用了开源框架自带的数据库服务，但该方法不保证相关数据在开源框架下的安全性。请务必在上传前对敏感数据进行脱敏处理。

本教程使用的 GPU 云主机规格如下所示，用于部署 DeepSeek-r1:70b 模型，配置仅供参考。

```
cpu 32 核

内存  128GGPU:  NVIDIA  A10*2  (48GB)
```

如需体验其它模型版本，请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--自定义部署 DeepSeek--步骤二：规格选型，选择合适的云主机。

优势

知识准确性：让模型访问定制的信息，从而提高回答的准确性和可靠性。

可解释性：检索过程可以明确指出回答所依据的信息来源，增强了回答的可解释性。

减少幻觉：降低了语言模型生成无事实依据内容（即 “幻觉”）的可能性。

一、环境准备

1.参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--快速体验 DeepSeek，准备相关设备及模型资源。

2.待模型启动后，通过访问 http://{公网 ip 地址}:3000 进入 OpenWebUI 界面。按照界面提示完成账号注册，开启私有知识库搭建之旅。



二、私有数据准备

搭建私有知识库需提前准备相应的数据材料。OpenWebUI 支持多种文本格式，包括但不限于：

- Markdown：轻量级标记语言格式，便于快速书写和阅读，能清晰呈现文本结构，常用于文档撰写、笔记记录等场景。
- PDF：通用文档格式，可保留文档原始排版、字体、图像等元素，广泛应用于正式文档、书籍、报告的分发与存储。
- TXT：最基本的文本格式，仅包含文本内容，无复杂格式设置，适用于纯文本信息的存储和传输。
- HTML：超文本标记语言格式，用于创建网页，包含文本内容和网页元素标记信息，可被浏览器解析显示。
- docx：是 Microsoft Word 的文档格式，能包含丰富的文本格式、图表、样式等内容，OpenWebUI 可以借助相关的文档解析库来提取其中的文本信息，用于知识库的构建和查询。

对于结构简单的纯文本材料，可直接通过 OpenWebUI 进行解析。若需了解除纯文本外其他文件格式的支持情况，请查阅 OpenWebUI 官方文档。

三、私有知识库搭建教程

1、登录 OpenWebUI 后，点击左上角 “工作空间”。



2、在“工作空间”中找到“知识库”选项，点击右侧的“+”号新建知识库。



3、在新建知识库页面，填写知识库名称、描述，设置可见性等信息，完成后点击“创建知识库”。





4、新建知识库成功后，再次点击右侧的“+”号，选择要添加的知识文档（可一次性添加多个）。添加成功后，点击对应文档查看加载的语料内容，验证知识解析是否准确。

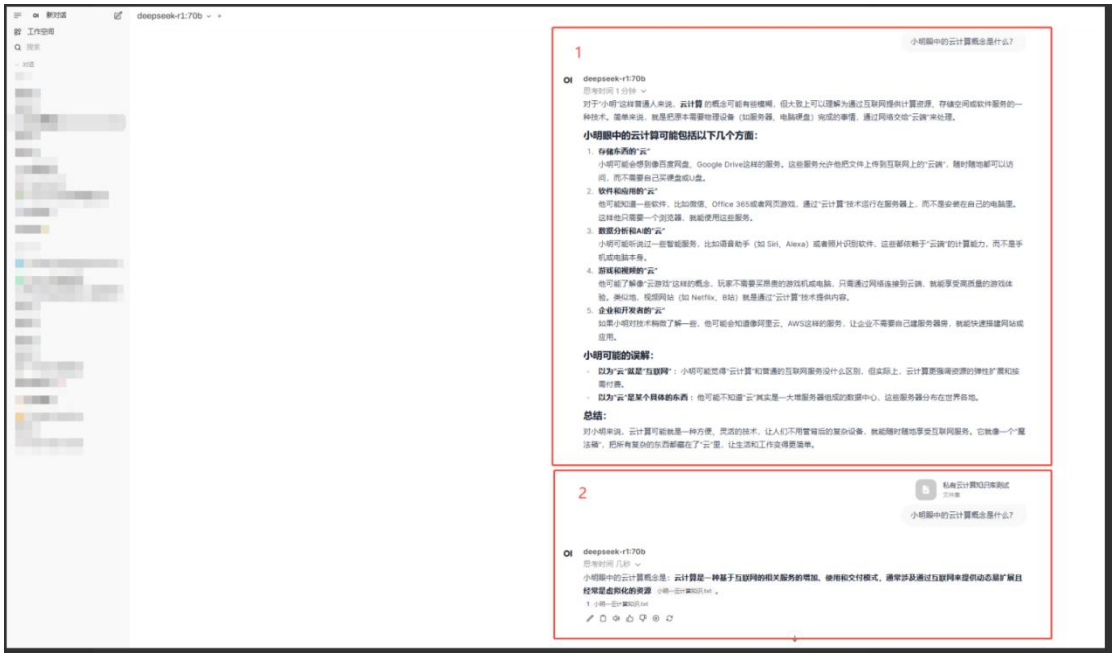




四、私有知识模型推理验证

1、点击左上角 “新对话”，选择已部署的推理模型。





2、在输入对话内容时，以 “#” 号开头，可选择对应知识库或知识文档作为推理参考。



3、若不想继续使用某个知识库，可按照特定操作流程将其移除，具体操作可参考相关图示。



通过以上步骤，您即可完成一个简单的私有知识库搭建，赶紧动手尝试吧！

五、推荐使用场景

1、智能客服训练 AI 回答公司政策、产品 FAQ，减少人工客服负担。

2、企业内部搜索让员工直接向 AI 提问，无需翻遍公司 Wiki 或文档库。

3、智能助手帮助销售整理客户信息，协助 HR 处理面试反馈，助力研发自动生成技术文档。

4、智能营销策划协助市场团队分析市场趋势和竞争对手数据，生成针对性的营销活动方案，提升营销效果。

5、财务智能分析工具帮助财务人员自动完成财务报表分析，预测财务风险，为企业决策提供数据支持。

6、生产流程优化系统利用 AI 实时监测生产设备运行状况，预测设备故障，提前

安排维护，减少生产中断。

7、培训内容生成助手，为培训部门快速生成员工培训课件、在线课程内容，满足不同岗位的培训需求。

8、供应链管理智能助手，根据销售数据、库存水平和供应商信息，辅助优化采购计划。

返回 DeepSeek 专题导航。

3.3.1.3 基于 DeepSeek 和 Chroma 构建个性化知识库的最佳实践

前言——私有知识库作用简介

大模型私有知识库，作为大语言模型技术与企业、组织自有数据深度融合的创新知识管理及应用解决方案，能够为特定用户群体提供更为精准、专业且安全的知识服务。具体来讲，它是借助大语言模型搭建而成，专门为特定组织或个人定制的知识存储与检索系统。此系统会对组织内部的专业知识、业务数据、历史文档等各类信息进行深度整合与精细化处理，从而构建出独一无二的专属知识集合。依托大模型强大的语言理解与生成能力，用户能够在此基础上实现高效的知识查询与问答交互。

在本教程中，我们将为您详细介绍一种基于开源框架的私有知识库搭建方案。利用该方案，您可以在本地便捷地搭建起相应的私有知识库，大幅提升文本检索能力。

教程使用配置说明：C7 通用型云主机

```
cpu 8 核

内存 32G
```

优势

知识准确性：让模型访问定制的信息，从而提高回答的准确性和可靠性。

可解释性：检索过程可以明确指出回答所依据的信息来源，增强了回答的可解释性。

减少幻觉：降低了语言模型生成无事实依据内容（即“幻觉”）的可能性。

一、deepseek 自部署

请参考在天翼云使用Ollama运行 DeepSeek的最佳实践-7b版-弹性云主机-最佳实践-AIGC实践 - 天翼云 --自定义部署DeepSeek。

二、私有知识仓库构建

1、 安装向量数据库

a）检查python环境是否存在，不存在可参考安装

```
# 检查是否存在pythonpython --version

# 安装python环境

apt install python3
```

```
root@vm:~/deepseek-r1-1.5b# python --version
Python 3.10.12
```

b）安装向量数据库及与 ollama 交互的组件

```
#  安装  Chroma 向量数据库，基于langchain开源框架解析文本

pip install langchain_chroma#  安装  Ollama 交互库

pip install langchain_ollama#  安装  langchain_community 库

pip install langchain langchain_community
```

```
root@vm:~/deepseek-r1-1.5b# pip install langchain_chroma
Collecting langchain_chroma
  Downloading langchain_chroma-0.2.1-py3-none-any.whl (11 kB)
Collecting numpy<2.0.0,>=1.22.4
  Downloading numpy-1.26.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.2/18.2 MB 104.7 kB/s eta 0:00:00
Collecting chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0
  Downloading chromadb-0.6.3-py3-none-any.whl (611 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 611.1/611.1 KB 63.3 kB/s eta 0:00:00
Collecting fastapi<1,>=0.95.2
  Downloading fastapi-0.115.8-py3-none-any.whl (94 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 94.8/94.8 KB 72.5 kB/s eta 0:00:00
Collecting langchain-core!=0.3.0,!=0.3.1,!=0.3.10,!=0.3.11,!=0.3.12,!=0.3.13,!=0.3.14,!=0.3.2,!=0.3.3,!=0
  Downloading langchain_core-0.3.34-py3-none-any.whl (412 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 413.0/413.0 KB 91.5 kB/s eta 0:00:00
Collecting overrides>=7.3.1
  Downloading overrides-7.7.0-py3-none-any.whl (17 kB)
Collecting typing_extensions>=4.5.0
  Downloading typing_extensions-4.12.2-py3-none-any.whl (37 kB)
Collecting opentelemetry-instrumentation-fastapi>=0.41b0
  Downloading opentelemetry_instrumentation_fastapi-0.51b0-py3-none-any.whl (12 kB)
Collecting orjson>=3.9.12
  Downloading orjson-3.10.15-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (130 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 130.3/130.3 KB 82.6 kB/s eta 0:00:00
Collecting chroma-hnswlib==0.7.6
```

```
root@vm:~/deepseek-r1-1.5b# pip install langchain_ollama
Collecting langchain_ollama
  Downloading langchain_ollama-0.2.3-py3-none-any.whl (19 kB)
Requirement already satisfied: langchain-core<0.4.0,>=0.3.33 in /usr/local/lib/python3.10/dist-packages (from langchai
Collecting ollama<1,>=0.4.4
  Downloading ollama-0.4.7-py3-none-any.whl (13 kB)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.10/dist-packages (from langchain-core<0.
Requirement already satisfied: typing-extensions>=4.7 in /usr/local/lib/python3.10/dist-packages (from langchain-core<
Requirement already satisfied: pydantic<3.0.0,>=2.5.2 in /usr/local/lib/python3.10/dist-packages (from langchain-core<
Requirement already satisfied: langsmith<0.4,>=0.1.125 in /usr/local/lib/python3.10/dist-packages (from langchain-core
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in /usr/local/lib/python3.10/dist-packages (from langch
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/dist-packages (from langchain-core<0.4.0,>=0.3
Requirement already satisfied: packaging<25,>=23.2 in /usr/local/lib/python3.10/dist-packages (from langchain-core<0.4
Requirement already satisfied: httpx<0.29,>=0.27 in /usr/local/lib/python3.10/dist-packages (from ollama<1,>=0.4.4->la
Requirement already satisfied: certifi in /usr/lib/python3/dist-packages (from httpx<0.29,>=0.27->ollama<1,>=0.4.4->la
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/dist-packages (from httpx<0.29,>=0.27->ollam
```

```
root@vm:~# pip install langchain langchain_community
Collecting langchain
  Downloading langchain-0.3.18-py3-none-any.whl (1.0 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.0/1.0 MB 41.5 kB/s eta 0:00:00
Collecting langchain_community
  Downloading langchain_community-0.3.17-py3-none-any.whl (2.5 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.5/2.5 MB 57.6 kB/s eta 0:00:00
Requirement already satisfied: langsmith<0.4,>=0.1.17 in /usr/local/lib/python3.10/dist-packages (from langchain) (0.3.8)
Collecting langchain-text-splitters<1.0.0,>=0.3.6
  Downloading langchain_text_splitters-0.3.6-py3-none-any.whl (31 kB)
Collecting aiohttp<4.0.0,>=3.8.3
  Downloading aiohttp-3.11.12-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.6 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.6/1.6 MB 413.3 kB/s eta 0:00:00
Requirement already satisfied: numpy<2,>=1.26.4 in /usr/local/lib/python3.10/dist-packages (from langchain) (1.26.4)
Collecting async-timeout<5.0.0,>=4.0.0
  Downloading async_timeout-4.0.3-py3-none-any.whl (5.7 kB)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/dist-packages (from langchain) (6.0.2)
```

2、安装文本解析插件，这里以 pdf 进行示例

```
pip  install pdfplumber
```

```
root@vm:~/hp# pip install pdfplumber
Collecting pdfplumber
  Using cached pdfplumber-0.11.5-py3-none-any.whl (59 kB)
Collecting pypdfium2>=4.18.0
  WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after
able')': /packages/be/d9/a2f1ee03d47fbeb48bcfde47ed7155772739622cfadf7135a84ba6a97824/pypdfium2
  Downloading pypdfium2-4.30.1-py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.9 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.9/2.9 MB 53.3 kB/s eta 0:00:00
Collecting Pillow>=9.1
  Downloading pillow-11.1.0-cp310-cp310-manylinux_2_28_x86_64.whl (4.5 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 4.5/4.5 MB 47.3 kB/s eta 0:00:00
Collecting pdfminer.six==20231228
  Downloading pdfminer.six-20231228-py3-none-any.whl (5.6 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.6/5.6 MB 44.2 kB/s eta 0:00:00
Collecting cryptography>=36.0.0
  Downloading cryptography-44.0.0-cp39-abi3-manylinux_2_28_x86_64.whl (4.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 4.2/4.2 MB 13.2 kB/s eta 0:00:00
```

3、基于私有数据开展模型对话

```
root@vm:~/hp# python3 ollama_private_vector_data_create.py
Number of requested results 4 is greater than number of elements in index 2, updating n_results = 2
检索到的文档摘要： <think>
嗯，我现在要总结一下这些文档中的主要主题。首先，我得仔细阅读每一段内容，找出它们之间的联系和区别。

第一段 talking about high generation能力，也就是生成文本的能力，包括快速生成高质量的中文文章、报告、故事等，强调语言流畅、逻辑连贯，并且可以根据不同的任

第二段是领域专业性，比如针对代码专门优化训练，提供准确专业的回答支持行业应用，这可能意味着DeepSeek - Coder在代码开发中特别强。

第三段讲可扩展性，模型架构设计灵活，适应数据和计算资源的变化，这说明模型能够随着使用情况不断优化，保持性能稳定。

接下来，应用场景部分分为智能写作、智能客服、代码开发和知识问答，这些都是用户可能遇到的具体问题解决方法，比如生成文案、处理客服请求、编写代码、获取信息查

DeepSeek的模型家族包括DeepSeek-Coder、DeepSeek-Llama和DeepSeek-R1。DeepSeek-Coder专注于代码领域，而DeepSeek-Llama在通用任务上表现好，DeepSeek-R1则有一定

特点部分提到强大的语言理解能力，并且具备在不同场景下的性能提升，这可能包括通过训练数据进行优化，提高模型的适应能力和扩展性。

综上所述，文档主要讲的是DeepSeek在自然语言处理和相关领域的应用，包括生成、问答、代码开发等多个方面，重点放在模型的专业性、可扩展性和多样化应用场景上。
</think>

```markdown
# 深度探索（DeepSeek）技术总结

## 1. 主要主题总结
```

图中使用的模型对话脚本如下，仅供参考：
```

```
from langchain_community.document_loaders  import
PDFPlumberLoaderfrom langchain.text_splitter  import
RecursiveCharacterTextSplitterfrom langchain_chroma  import
Chromafrom langchain_ollama  import OllamaEmbeddings, ChatOllamafrom
langchain_core.output_parsers  import StrOutputParserfrom
langchain_core.prompts  import ChatPromptTemplatefrom
langchain_core.runnables  import RunnablePassthrough
# 加载 PDF 文档
file="DeepSeek.pdf"
loader  = PDFPlumberLoader(file)
docs  = loader.load()
# 分割文档，分割参数会影响推理准确性
text_splitter  = RecursiveCharacterTextSplitter(chunk_size=500,
chunk_overlap=0)
all_splits  = text_splitter.split_documents(docs)
# 创建向量存储
local_embeddings  = OllamaEmbeddings(model="deepseek-r1:1.5b")
vectorstore  = Chroma.from_documents(documents=all_splits,
embedding=local_embeddings)
# 初始化模型
model  = ChatOllama(
    model="deepseek-r1:1.5b",
)
# 定义摘要提示模板
summary_prompt = ChatPromptTemplate.from_template(
```

```
    "总结这些检索到的文档中的主要主题：{docs}"
)
# 将传入的文档转换成字符串的形式 def format_docs(docs):
    return "\n\n".join(doc.page_content for doc in docs)
# 创建摘要链
summary_chain ={"docs": format_docs}| summary_prompt | model |
StrOutputParser()
# 定义问题
question ="DeepSeek 项目的目的是什么？"
# 检索相关文档
docs = vectorstore.similarity_search(question)
# 生成摘要
summary = summary_chain.invoke(docs)print("检索到的文档摘要：",
summary)
# 定义提示模板
RAG_TEMPLATE ="""
你是一个问答助手。请使用以下检索到的上下文信息来回答问题。如果你不知道答案，请直接说不知道。答案最多使用三句话，并保持简洁。

<context>
{context}
</context>

回答以下问题：
```

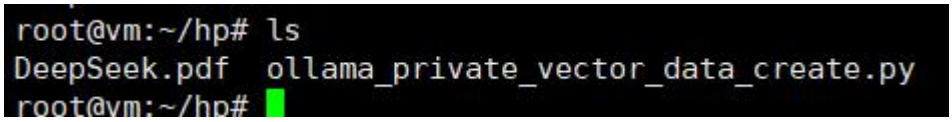
```
{question}
"""

rag_prompt  = ChatPromptTemplate.from_template(RAG_TEMPLATE)
# 获取检索器
retriever  = vectorstore.as_retriever()
# 创建问答链
qa_chain = (
    {"context": retriever | format_docs, "question":
RunnablePassthrough()}
    | rag_prompt
    | model
    | StrOutputParser()
)
# 运行问答链
answer  = qa_chain.invoke(question)print("问题的答案：", answer)
```

综上，您成功构建了一个定制化的私有知识库检索模型。

注意

对应文本数据需要与运行脚本在同一目录

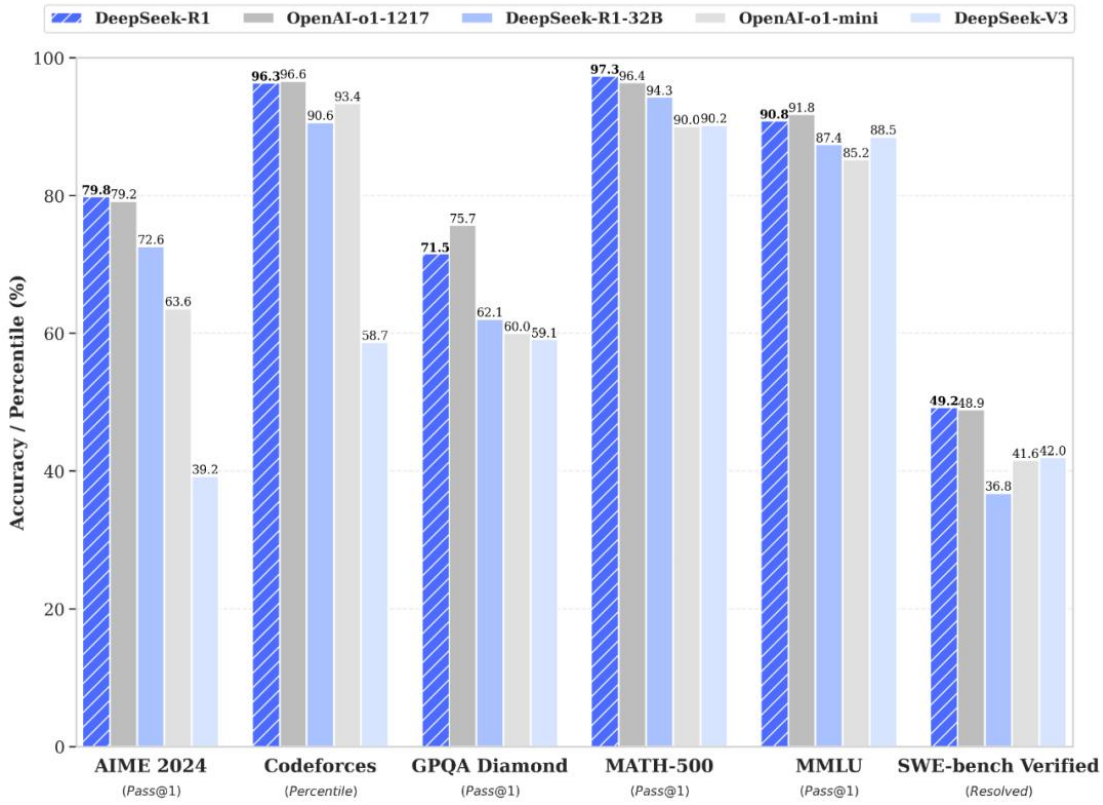


返回 DeepSeek 专题导航。

3.3.1.4 基于 LLaMA-Factory 微调 DeepSeek-R1-Distill-Qwen-7B 最佳实践

DeepSeek 简介

DeepSeek 是一个基于 Transformer 架构的大型语言模型（LLM），由深度求索（DeepSeek）公司开发。它能够处理自然语言理解、生成、翻译、问答等多种任务。DeepSeek-R1 在后训练阶段大规模使用了强化学习技术，在仅有极少标注数据的情况下，极大提升了模型推理能力。在数学、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。



DeepSeek 公司在开源 DeepSeek-R1-Zero 和 DeepSeek-R1 两个参数量超过 660B 模型的同时，通过 DeepSeek-R1 的输出，蒸馏了 6 个小模型开源给社区，其中 32B 和 70B 模型在多项能力上实现了对标 OpenAI o1-mini 的效果。

	AIME 2024 pass@1	AIME 2024 cons@64	MATH-500 pass@1	GPQA Diamond pass@1	LiveCodeBench pass@1	CodeForces rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759.0
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717.0
o1-mini	63.6	80.0	90.0	60.0	53.8	**1820.0**
QwQ-32B	44.0	60.0	90.6	54.5	41.9	1316.0
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954.0
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189.0
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481.0
DeepSeek-R1-Distill-Qwen-32B	**72.6**	83.3	94.3	62.1	57.2	1691.0
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205.0
DeepSeek-R1-Distill-Llama-70B	70.0	**86.7**	**94.5**	**65.2**	**57.5**	1633.0

模型微调简介

大语言模型微调技术通过调整预训练模型参数，使其适应特定任务需求。其核心价值在于让通用的人工智能工具变得更“专业”和“实用”，就像把一个会做各种菜的厨师培训成专精川菜的大厨。它能让企业用较低成本获得适合自身需求的 AI 能力，例如：

1. 降低成本：无需从头训练昂贵的大模型，只需在现有通用模型基础上“补充学习”企业内部的业务知识、行业术语，就能快速打造专属 AI 助手，成本大幅降低。

2. 精准解决问题：通用 AI 可能懂诗词绘画，但不懂医疗诊断或法律合同。通过微调，AI 能深度掌握特定领域的专业知识，例如快速分析病历、自动生成符合行业规范的报告。

3. 保护数据隐私：企业可将内部数据安全地用于微调，既让 AI 掌握核心业务知识，又避免敏感数据直接暴露在公共模型中。

4. 推动普惠应用：中小型企业也能通过这项技术享受 AI 红利，例如教育机构定制学习辅导 AI，工厂打造设备维护专家系统，真正让 AI 进入百行千业。

在下文中，我们将采用两个具体实践案例来说明模型微调的实际效果：

1. 基于天翼云文档数据的微调。利用企业自身的文档数据构建微调数据集，仅使用单台多 GPU 卡的云主机即可完成模型微调训练。经过微调后的模型能够在企业内部知识领域提供准确的回答。

2. 基于医学计算数据的微调。利用医学领域的专业知识数据构建微调数据集，

仅使用单台多 GPU 卡的云主机即可完成模型微调训练。经过微调后的模型能够在识别与分析病例时取得更高的计算准确度。

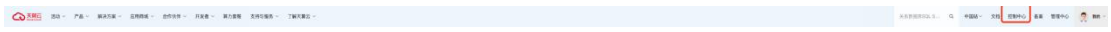
快速微调 DeepSeek

LLaMA-Factory 是一个开源项目，旨在为用户提供简便的方式来训练和部署大型语言模型（LLM）。这里，我们以蒸馏小模型 DeepSeek-R1-Distill-Qwen-7B 模型为例，讲解如何在天翼云使用 LLaMA-Factory 运行与微调 DeepSeek 模型，读者可以根据自己实际需求选择模型和云主机配置。

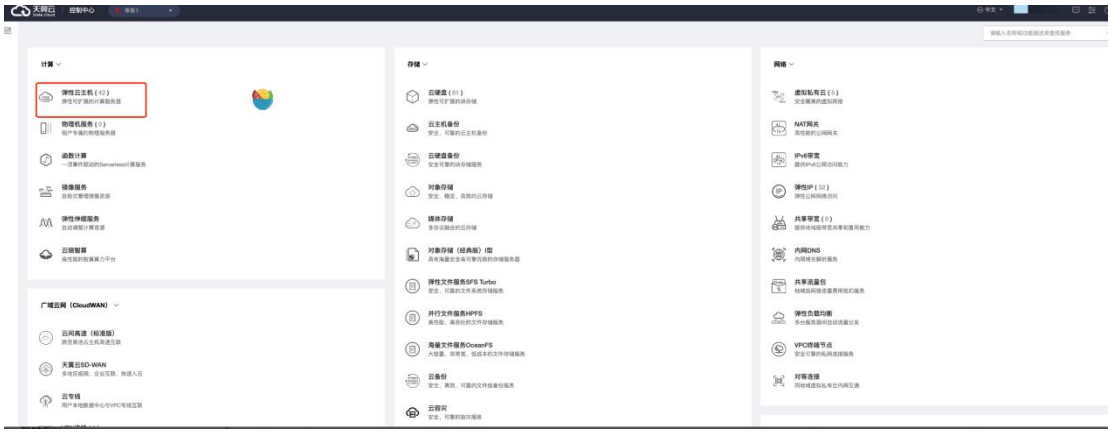
步骤一：创建 GPU 云主机

1. 进入创建云主机页面。

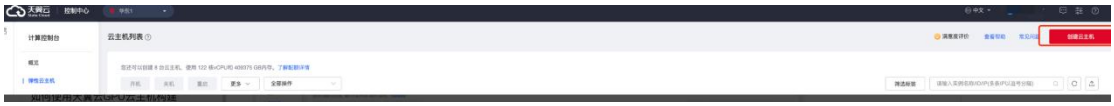
a. 点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



b. 单击“服务列表>弹性云主机”，进入主机列表页。



c. 单击“创建云主机”，进入弹性云主机创建页。



2. 进行基础配置

a. 根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、“实例名称”、“主机名称”等。

b. 选择规格。此处选择"CPU 架构"为"X86"、"分类"为"GPU 加速/AI 加速型"、"规格族"为"GPU 计算加速型 pi7"、"规格"为"pi7.4xlarge.4"。

注意

本案例仅做微调流程演示，选择的机型规格仅支持 7B 模型的少量数据微调，如果您的数据量较大，请选择显存更高的机型规格。

c.选择镜像。“镜像类型”选择“镜像市场”，在云镜像市场中选择预置了 DeepSeek-R1-Distill-Qwen-7B 模型的 DeepSeek-LlamaFactory-Ubuntu22.04 镜像。



3.网络及高级配置

设置网络，包括"网卡"、"安全组"，同时配备 "弹性 IP" 用于下载和访问模型；设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。注意安全组需要放行 22 端口(用于 ssh)和 7860 端口(用于访问 web 页面)。

4.确认并支付

步骤二：微调 DeepSeek 模型

1.启动 LLaMA-Factory 服务

在完成云主机开通后，即可通过 ssh 连接到到云主机命令行。模型及 LLaMA-Factory 服务均已放置在/root 目录下，且已完成配套环境安装，可直接运行。

```
# 进入目录 cd /root/LLaMA-Factory# 启动 webui 服务（也可通过 nohup 放到后台启动）

llamafactory-cli webui
```

在成功启动 LLaMA-Factory 服务后，您将可以通过 http://{公网 ip 地址}:7860/ 访问到相关页面。

```
# 访问服务 http://{公网 ip 地址}:7860/
```

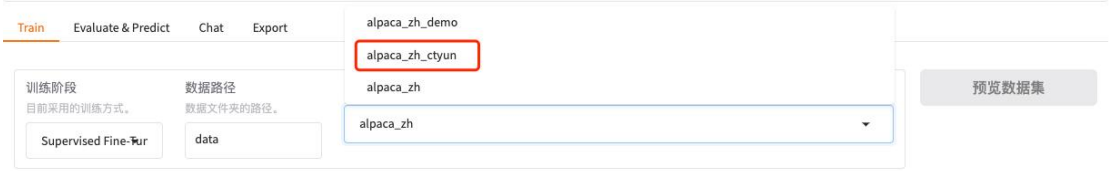
2.配置基础大模型和微调方法

在成功进入 LLaMA-Factory 页面后，您首先需要按照如图的内容进行页面语言，模型名称和模型路径的设置。注意模型预置在/root/DeepSeek-R1-Distill-Qwen-7B 目录下。



3.配置微调数据集

在 Train 页面内，您需要配置用于微调训练的数据集。镜像内预置了 alpaca 格式的天翼云示例数据，您也可参考示例数据修改/root/LLaMA-Factory/data/目录下的 alpaca_zh_ctyun.json(微调数据集)和 dataset_info.json(数据集信息)两个文件，改用您自己的私有数据。



镜像内置两份数据集，单条内容示例如下：

data/alpaca_zh_ctyun.json(天翼云文档数据集)

```
{
    "instruction": "天翼云的通用型弹性云主机有哪些规格？",
    "input": "",
    "output": "<think>用户询问通用型云主机规格，首先需要明确通用型的定位是平衡计算和内存资源。根据产品文档，通用型以 s7 系列为代表，应当列举不同后缀规格及其配置差异，最后补充适用场景说明。</think>\n 通用型弹性云主机有多个规格，例如 s7.small、s7.medium、s7.large、s7.xlarge 等，每种规格提供不同的 vCPU 和内存配置，适合多种业务场景。"
}
```

data/alpaca_en_medcalc_bench_max2k.json(医学计算数据集 MedCalc-Bench)

```
{
    "instruction": "A 16-year-old female adolescent was referred to our hospital with severe hypertension (systolic pressure 178 mmHg), which was first detected 7 months prior to presentation during a routine annual physical ...",
    "input": "What is the patient's Creatinine Clearance using the Cockroft-Gault Equation in terms of mL/min? You should use the patient's adjusted body weight in kg instead of the patient's actual body weight if the patient is ...",
    "output": "<think>The formula for computing Cockcroft-Gault is given by CrCl = ((140 - age) * adjusted weight * (gender_coefficient)) / (serum creatinine * 72), where the gender_coefficient is 1 if male, and 0.85 if female..."
}
```


4.配置相关训练参数

在 Train 页面内，还有大量可修改的微调训练参数。如果您对他们还不够了解，可暂时不进行修改。其中，训练轮数与样本量级关联性较大，如果您的微调数据量很少，则可能需要设置较大的训练轮数，才能有效果。



5.启动微调训练

点击最下方的开始按钮，即可基于上面选择的基础模型和微调数据集，启动模型微调训练。页面下方会显示实时的训练进度，训练日志和 loss 变化情况。



如果您的训练样本较多，单张 A10 显卡的 24G 显存很容易因为无法承载，而导致报错"CUDA out of memory"。此时，您需要将云主机变配到显存更大的机型规格，并开启 DeepSpeed stage3 进行模型参数分片，如下图所示。



同时，您还需要修改默认 DeepSpeed 配置中的部分参数，以保证训练正常进行。修改/root/LLaMA-Factory/cache/ds_z3_config.json 中的如下内容：

```
"zero_optimization": {
    "stage": 3,
    "overlap_comm": true,
    "contiguous_gradients": true,
    "sub_group_size": 1e9,
    "reduce_bucket_size": "auto",
```

```
    "stage3_prefetch_bucket_size": "auto",
    "stage3_param_persistence_threshold": "auto",
    "stage3_max_live_parameters": 1e4, //降低该参数以减少显存占用
    "stage3_max_reuse_distance": 1e4, //降低该参数以减少显存占用
    "stage3_gather_16bit_weights_on_model_save": true}
```

步骤三：测试模型微调效果

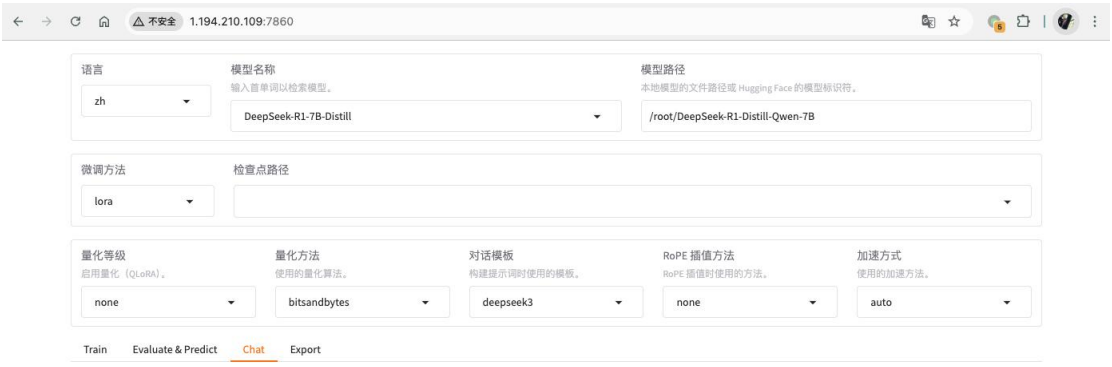
LLaMA-Factory 框架提供了模型推理的能力，可直接进入 Chat 页面加载模型进行对话。

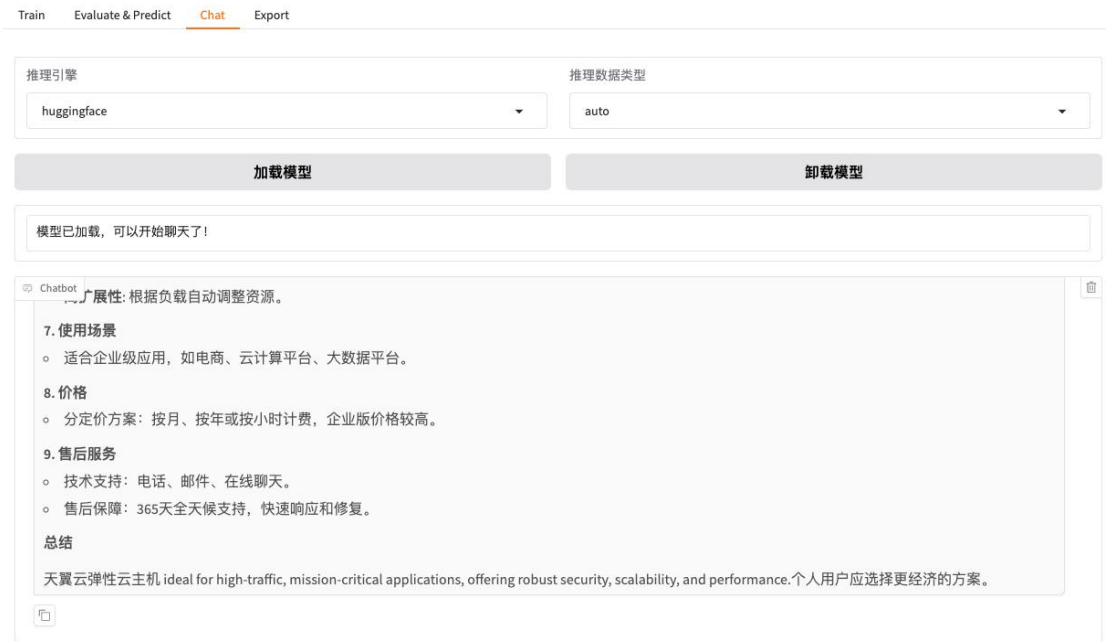
1.测试基础大模型效果

保持模型名称和模型路径不变，不设置检查点路径，在 Chat 页面直接点击加载模型，即可测试基础大模型的效果。

本案例使用 prompt="天翼云弹性云主机包含哪些规格"进行测试。

基础 DeepSeek-R1-Distill-Qwen-7B 大模型给出了通用但是错误的回答。



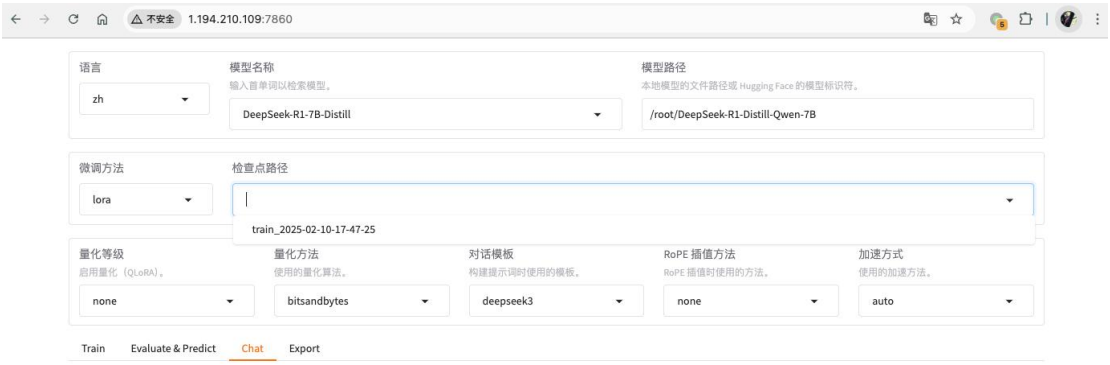


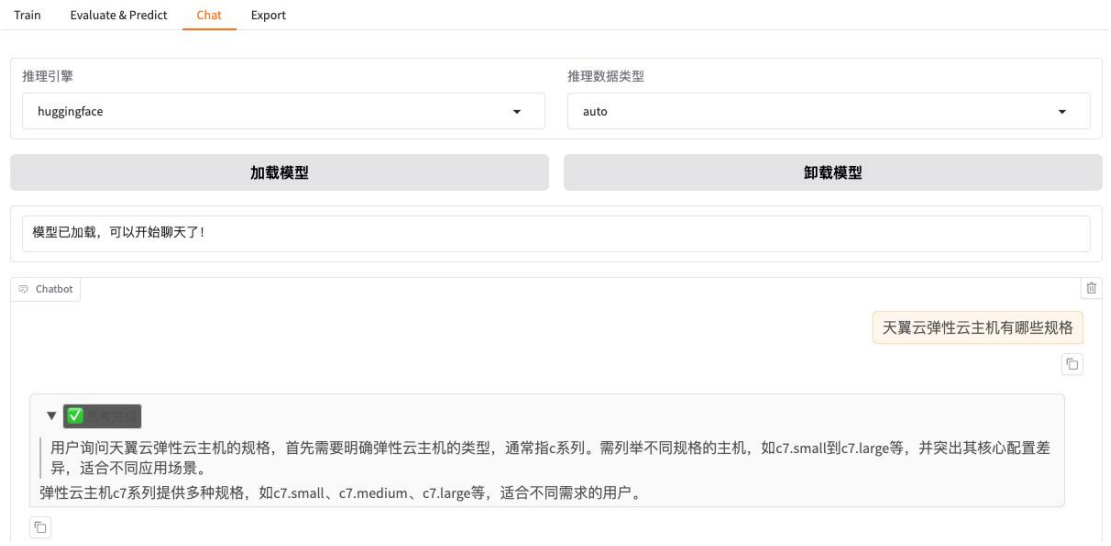
2. 测试微调大模型的效果

本案例采用了 lora 的微调方法，因此需将检查点路径设置为步骤二中完成训练的 lora 小模型。

本案例使用 prompt="天翼云弹性云主机包含哪些规格"进行测试。

微调后的大模型能够给出符合微调数据集的回答。





从纯净版 ubuntu 系统构建微调环境

除了使用预装镜像外，您也可以从任意 ubuntu 云主机开始，按以下步骤搭建出 LLaMA-Factory 微调环境，并下载任意您需要的模型。

步骤一：安装 Nvidia 显卡驱动与 CUDA 库

```
# 安装 nvidia 显卡驱动(如已选择带驱动镜像，则忽略)apt updateapt install ubuntu-drivers-common alsa-utils# 查看推荐的驱动版本 ubuntu-drivers devices# 安装推荐版本 apt install nvidia-driver-550# 重启一次云主机 reboot

# 安装 cuda-12.2wget https://developer.download.nvidia.com/compute/cuda/12.2.0/local_installers/cuda_12.2.0_535.54.03_linux.run# 注意取消对 nvidia driver 的安装 sh cuda_12.2.0_535.54.03_linux.run# 安装完成后在~/.bashrc 中写入环境变量# 添加 CUDA 可执行文件路径 export PATH=/usr/local/cuda-12.2/bin${PATH:+:${PATH}}# 添加 CUDA 库文件路径 export LD_LIBRARY_PATH=/usr/local/cuda-12.2/lib64${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}# 使环境变量生效 source ~/.bashrc
```

步骤二：安装 git 与下载模型

```
# 安装 git 和 lfs
```

```
apt update

apt install git

curl -s
https://packagecloud.io/install/repositories/github/git-lfs/script.d
eb.sh | bash

apt install git-lfs

git lfs install

# 下载模型# 国内用 modelscope 源速度最快

git clone
https://www.modelscope.cn/deepseek-ai/DeepSeek-R1-Distill-Qwen-7B.gi
t# 清理掉不用的.git 目录下 lfs 大文件版本数据(能节省一半存储空间)

git lfs prune -f
```

步骤三：安装 LLaMA-Factory 及相关训练环境

```
# 安装 miniconda

wget
https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.s
h# 下载完成后手动安装

sh Miniconda3-[版本号]-Linux-x86_64.sh# 使环境变量生效 source
~/.bashrc

# 安装 LLaMA-Factory

git clone --depth 1 https://github.com/hiyouga/LLaMA-Factory.gitcd
LLaMA-Factory

pip install -e ".[torch, metrics]" -i
https://pypi.tuna.tsinghua.edu.cn/simple

pip install deepspeed -i https://pypi.tuna.tsinghua.edu.cn/simple
```

```
pip install flash-attn --no-build-isolation -i
https://pypi.tuna.tsinghua.edu.cn/simple
```

返回 DeepSeek 专题导航。

3.3.1.5 基于 Open WebUI 实现 DeepSeek API 调用及外部大模型 API 接入最佳实践

使用场景

大模型 API 提供了接口以便用户更好地进行 DeepSeek 管理和应用开发。用户在开通天翼云 DeepSeek 云主机后，利用 Open WebUI API 即可实现对外提供 API 调用，同时也可以非常方便的实现外部大模型 API 接入使用。

本教程使用的 GPU 云主机规格如下所示，用于部署 DeepSeek-r1:7b 模型，配置仅供参考。

```
cpu 16 核
内存 64G
GPU: NVIDIA A10*1 (24GB)
```

如需体验其它模型版本，请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--自定义部署 DeepSeek--步骤二：规格选型，选择合适的云主机。

准备

参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--快速体验 DeepSeek，准备相关设备及模型资源。

认证

OpenWebUI 使用 Bearer Token 认证保证客户端请求的合法性，其 Token 为 APIKEY 中 JWT 令牌(Json Web Token)。获取 API key 方式如下：

1. 登录 Open WebUI。



2.选择设置->账户，选择获取或者新建 API KEY。



API 接口

1.列举模型

URL：GET /api/models

描述：列出可用的模型列表，并提供相关模型的基本信息。

例子:

```
curl -H "Authorization: Bearer API_KEY" http://公网 IP:3000/api/models
```

2.对话补全

URL：POST /api/chat/completions

描述：根据输入的上下文，来让模型补全对话内容。

例子：

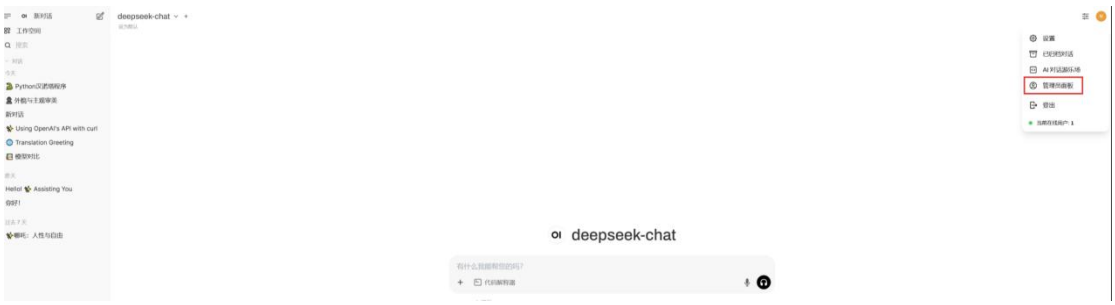
```
curl -L -X POST 'http://公网 IP:3000/api/chat/completions' \
-H 'Content-Type:application/json' \
-H 'Accept:application/json' \
-H 'Authorization:BearerAPI_KEY' \
-d '{
   "model":"deepseek-r1:7b",
    "messages": [
    {
     "content":"用英语翻译一下你好",
     "role":"user"
    }
  ],
  "stream":false
}'
```

使用外部大模型

Open WebUI 通过云主机方式部署完自己的 DeepSeek 后，可以直接嵌入使用外部大模型，嵌入后可以通过 WEB、API 进行模型使用。配置方式如下：

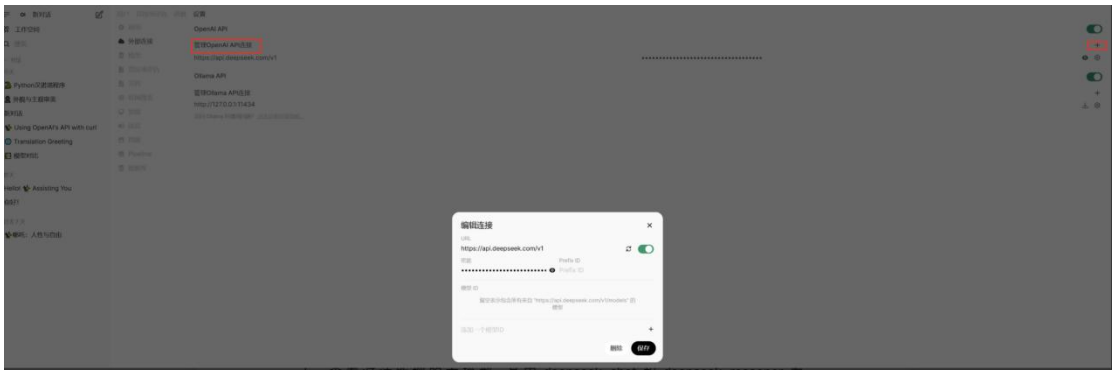
1. 登录 Open WebUI。

2. 选择管理员面板->设置->外部连接。



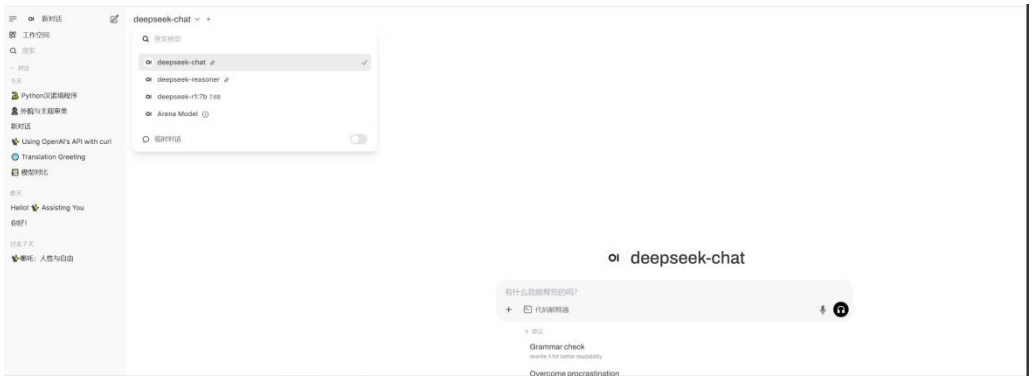


3.选择 OpenAI API 开启，点击管理 OpenAI API 连接，新增外部大模型连接，输入外部模型的 URL 和 API KEY。



4.点击保存。

5.查看对话新增的大模型，其中 DeepSeek-chat 和 DeepSeek-reasoner 为 DeepSeek 官网新增模型。

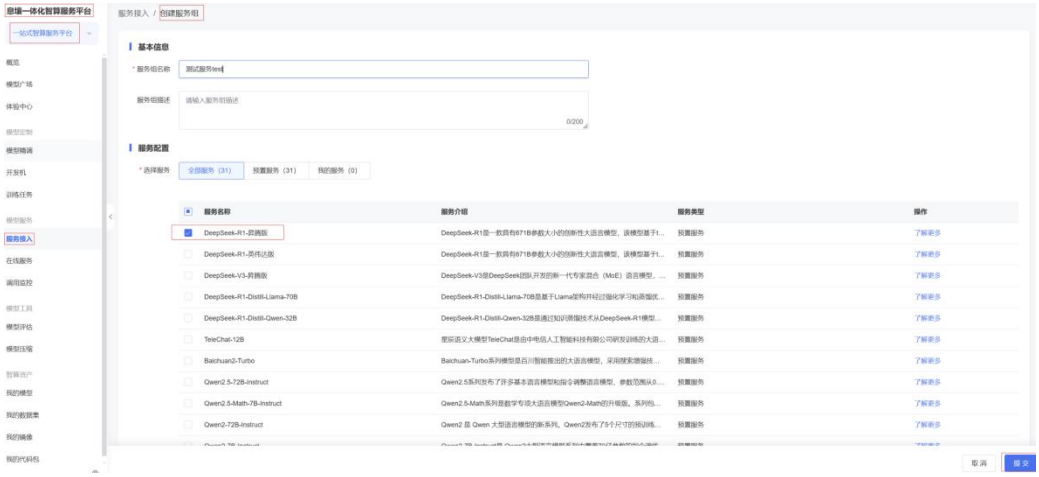


使用天翼云息壤一体化智算服务平台

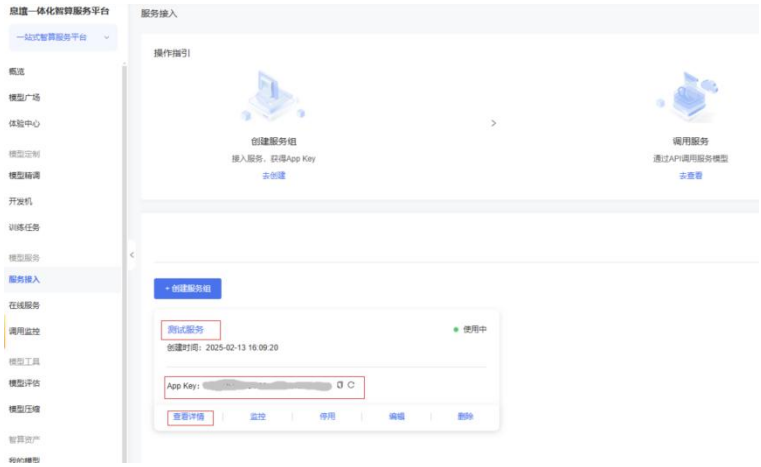
准备工作

使用天翼云息壤提供的大模型不仅需要提供 API KEY,还需要提供对应的模型 ID。其获取方式如下:

1.登录天翼云官网,选择息壤一体化智算服务平台->一站式智算服务平台->服务接入->创建服务组->DeepSeek-R1-昇腾版，点击提交。



2.选择测试服务，获 App key 作为 API KEY。选择查看详情->了解更多。



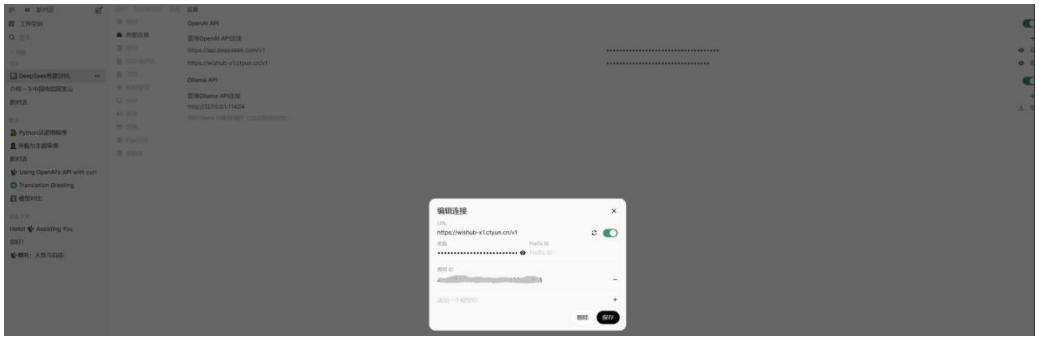
3.选择 API 文档，获取模型 ID 和终端地址。



Open WebUI 接入

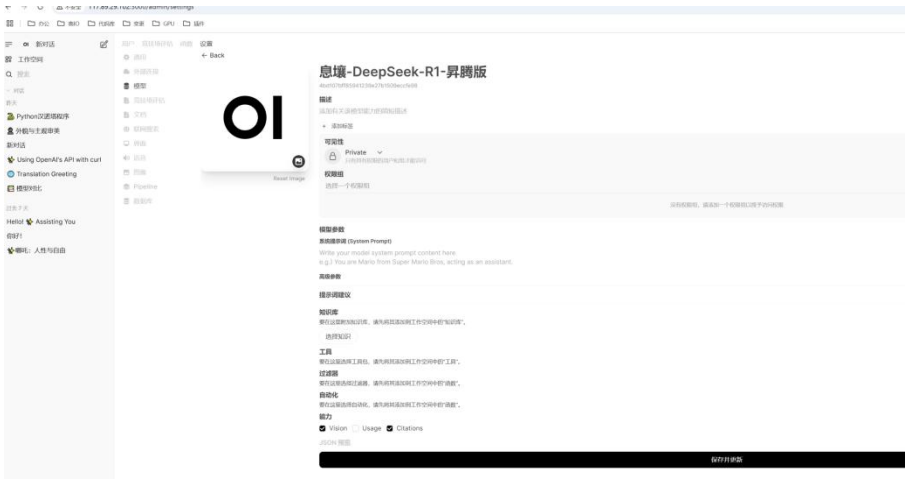
1.登录 Open WebUI 并选择管理员面板->设置->外部连接。

2.选择 OpenAI API 开启，点击管理 OpenAI API 连接，新增息壤大模型终端地址、API KEY 和模型 ID。终端地址需要增加 V1 版本。



3. 点击保存。

4. 点击模型，修改模型 ID 为易理解的模型。



5. 查看对话新增的大模型，其中息壤-DeepSeek-R1-昇腾版为新增模型。



注意

请妥善保管您的 API KEY，不要在公共场合泄露。

返回 DeepSeek 专题导航。

3.3.1.6AnythingLLM 协同 Ollama 启用 DeepSeek 大模型最佳实践

推荐使用场景

工具	推荐场景	不适用场景
AnythingLLM	企业文档问答、私有知识库管理、需严格权限控制的团队协作	快速测试模型、轻量级个人聊天需求
Open WebUI	个人或开发者快速体验模型、多模型切换测试、多模态交互(如语音/图像)	复杂文档管理、企业级安全与协作需求

注意

建议使用天翼云提供的 DeepSeek 镜像创建云主机，减少安装过程中可能遇到的问题。镜像选择请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本-弹性云主机-最佳实践-AIGC 实践 - 天翼云。

操作步骤：

一、云主机端口配置更新

1.进入云主机详情页，选择安全组选项，添加新的开放端口。

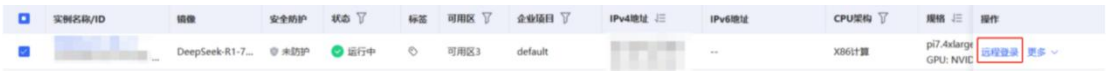
建议选择新加 3001 端口。添加完成后，可在安全组规则列表中查看新添加的 3001 端口规则，确认其状态为正常生效。







2.进入控制台，远程登陆即将安装 AnythingLLM 的云主机。



二、AnythingLLM 服务部署（ docker 方式）

1.获取 AnythingLLM docker image 文件。

由于国内无法访问 docker hub 网站，您需要修改 docker 的服务配置文件，添加国内可访问的源，以便能够顺利拉取 AnythingLLM docker image 文件。操作步骤如下：

- 向 docker 配置文件中添加国内可访问的 docker 镜像源
- 重启 docker 服务
- 拉取 AnythingLLM docker image

Docker 守护进程（ daemon ）的配置文件。添加国内可访问镜像源。sudo vi /etc/docker/daemon.json# 重启 docker 服务 sudo systemctl restart docker# 确认 docker 服务正常 sudo systemctl status docker# 拉取 anythingllm 镜像

docker pull mintplexlabs/anythingllm

```
root@ecm-aigc-test:~# sudo vi /etc/docker/daemon.json
root@ecm-aigc-test:~# sudo systemctl restart docker
root@ecm-aigc-test:~# sudo systemctl status docker
• docker.service - Docker Application Container Engine
     Loaded: loaded (/lib/systemd/system/docker.service; enabled; vendor preset: enabled)
     Active: active (running) since Tue 2025-02-18 14:49:44 CST; 11s ago
TriggeredBy: • docker.socket
       Docs: https://docs.docker.com
   Main PID: 363483 (dockerd)
      Tasks: 39
     Memory: 28.5M
        CPU: 242ms
     CGroup: /system.slice/docker.service
             ├─363483 /usr/bin/dockerd -H fd:// --containerd=/run/containerd/containerd.sock
             ├─363691 /usr/bin/docker-proxy -proto tcp -host-ip 0.0.0.0 -host-port 3000 -container-ip 172.17.0.2 -container-port 3001
             └─363702 /usr/bin/docker-proxy -proto tcp -host-ip :: -host-port 3000 -container-ip 172.17.0.2 -container-port 3001

Feb 18 14:49:42 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:42.806234391+08:00" level=info msg="Starting up"
Feb 18 14:49:42 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:42.806535786+08:00" level=info msg="OTEL tracing is not configur
Feb 18 14:49:42 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:42.853239755+08:00" level=info msg="[graphdriver] using prior st
Feb 18 14:49:42 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:42.855934725+08:00" level=info msg="Loading containers: start."
Feb 18 14:49:43 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:43.594220275+08:00" level=info msg="Default bridge (docker0) is
Feb 18 14:49:43 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:43.988487598+08:00" level=info msg="Loading containers: done."
Feb 18 14:49:44 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:44.030943927+08:00" level=info msg="Docker daemon" commit=4c9b3b
Feb 18 14:49:44 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:44.030997550+08:00" level=info msg="Daemon has completed initial
Feb 18 14:49:44 ecm-aigc-test dockerd[363483]: time="2025-02-18T14:49:44.044714842+08:00" level=info msg="API listen on /run/docker.so
Feb 18 14:49:44 ecm-aigc-test systemd[1]: Started Docker Application Container Engine.
root@ecm-aigc-test:~# sudo docker pull mintplexlabs/anythingllm
Using default tag: latest
latest: Pulling from mintplexlabs/anythingllm
3713021b0277: Pull complete
4f4fb700ef54: Pull complete
2e34c2dfb39a: Pull complete
d36c412a557a: Pull complete
4e3ea4c2625c: Pull complete
0e60e62892da: Pull complete
6dca9380efe7: Pull complete
578bda425106: Pull complete
c3d83711a1ff: Pull complete
a03b8dfb9685: Pull complete
1e822b21773f: Pull complete
2330397bbdc8: Pull complete
ac3c5f050331: Pull complete
48dc98216779: Pull complete
Digest: sha256:db45b3f9242b87e650e46081b81fb9004b88971c2e53cc85047200df7e110ee8
Status: Downloaded newer image for mintplexlabs/anythingllm:latest
docker.io/mintplexlabs/anythingllm:latest
```

daemon.json 文件添加的内容为：

```
{
    "registry-mirrors": ["http://hub.geekery.cn",
"http://hub.littlediary.cn", "http://docker.rainbond.cc"]
}
```

daemon.json 文件如下，仅供参考：

```
{
    "registry-mirrors": ["http://hub.geekery.cn", "http://hub.littlediary.cn", "http://docker.rainbond.cc"]
}
```

2.启动 AnythingLLM 服务。

- 首先查看下载的 AnythingLLM 容器镜像的 tag
- 启动 AnythingLLM 服务
- 查看 AnythingLLM 服务状态

```
# 查看当前的容器镜像列表

docker image list# 启动 AnythingLLM 服务，下述 PROT 参数需要与第一部分开放的端口相对应

docker run -d -e PORT=3001 -e OLLAMA_BASE_URL=http://127.0.0.1:11434 -e ENABLE_SIGNUP=true -e ENABLE_OPENAI_API=False --add-host=host.docker.internal:host-gateway --net=host -v anythingLLM:/app/backend/data --name anythingLLM --restart always mintplexlabs/anythingllm:latest# 查看 AnythingLLM 服务状态

docker ps -a
```

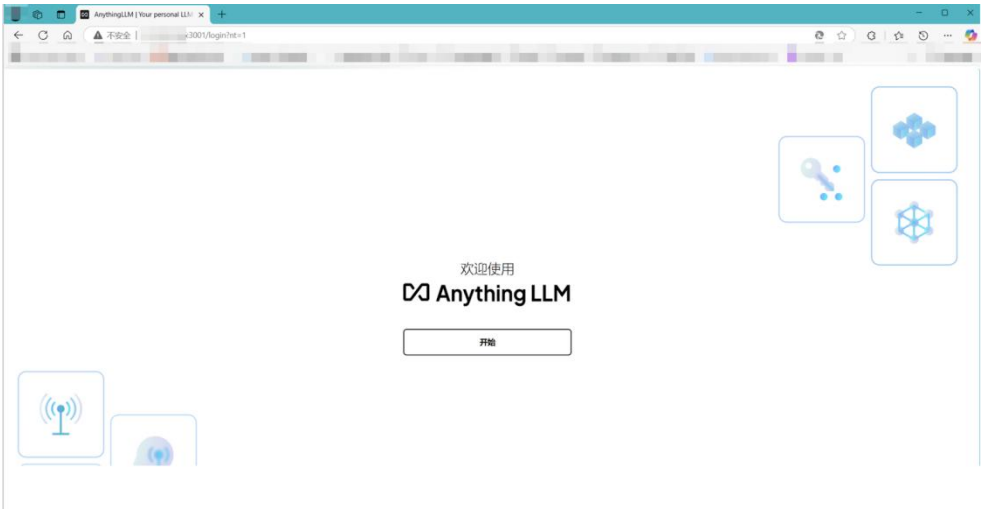




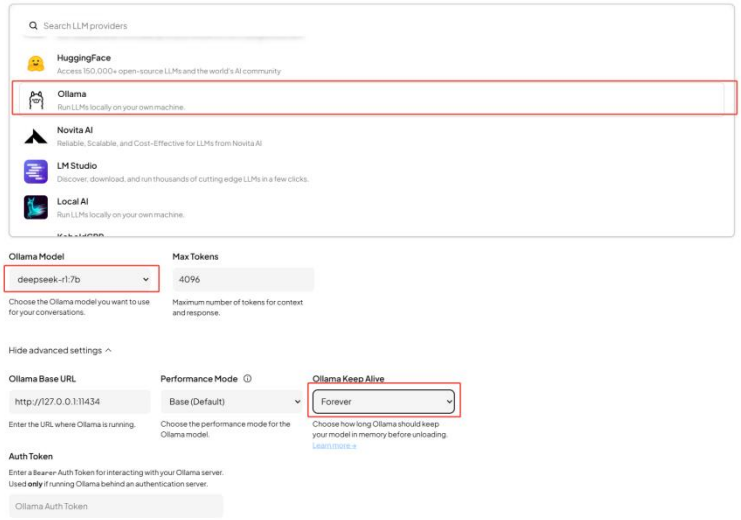
三、AnythingLLM 前端页面配置

1.公网访问 AnythingLLM。

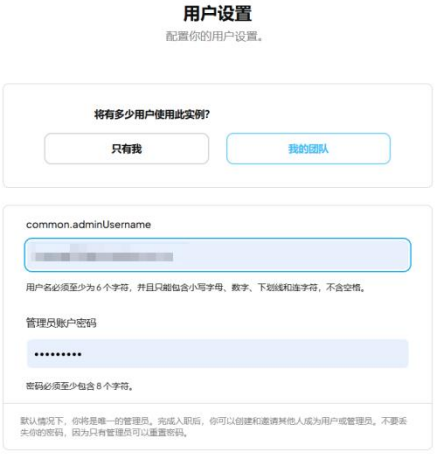
在浏览器输入 http://{your_ip}:3001 (your_ip 这台云主机的 eip)，会出现如下页面



2.选择 Ollama 作为模型加载项。



3.用户设置，按页面提示操作，注册账号，跳过用户调查。

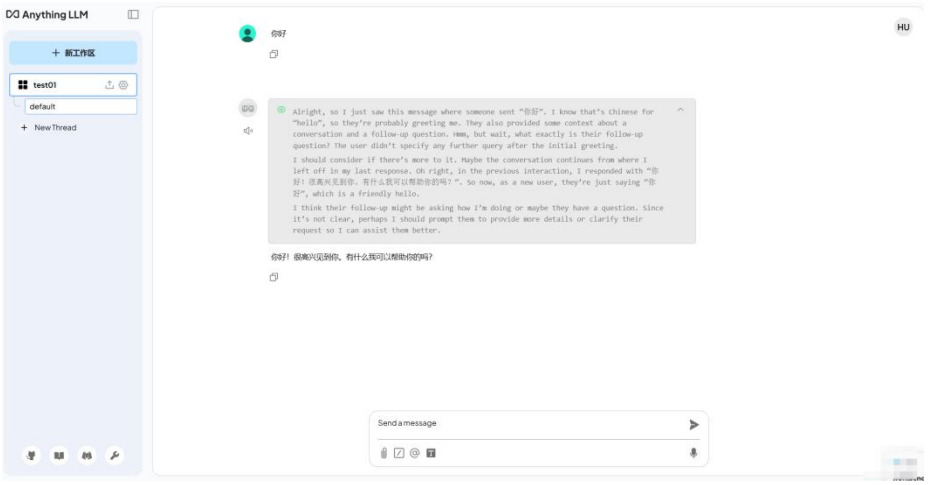




4. 创建工作区。



5. 点击工作区 test01 打开对话框，输入聊天内容。

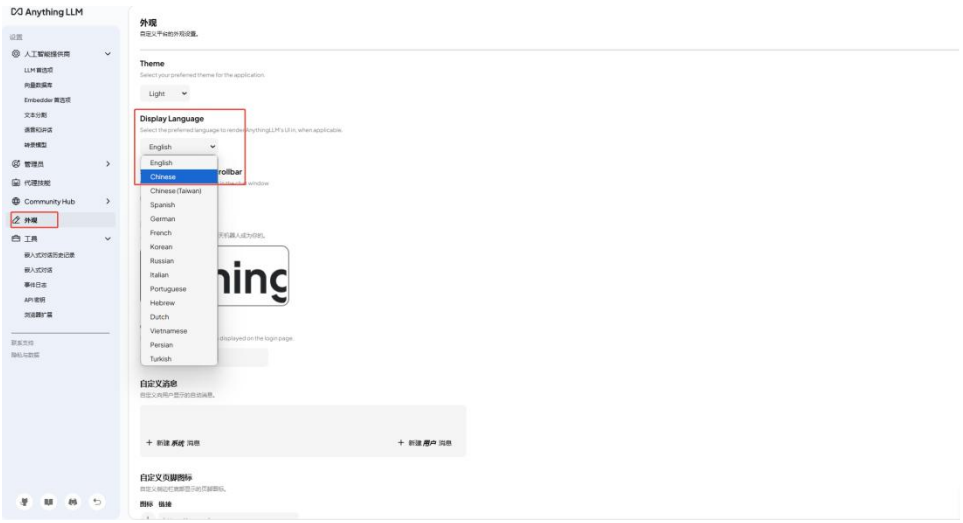


四、AnythingLLM 个性化探索

下面给出了一些个性化探索案例，您如果想要了解更多，请访问 AnythingLLM 社区。

1. 设置 AnythingLLM 默认语言格式为中文。参考下图进行设置。





2. AnythingLLM 启用联网搜索服务。

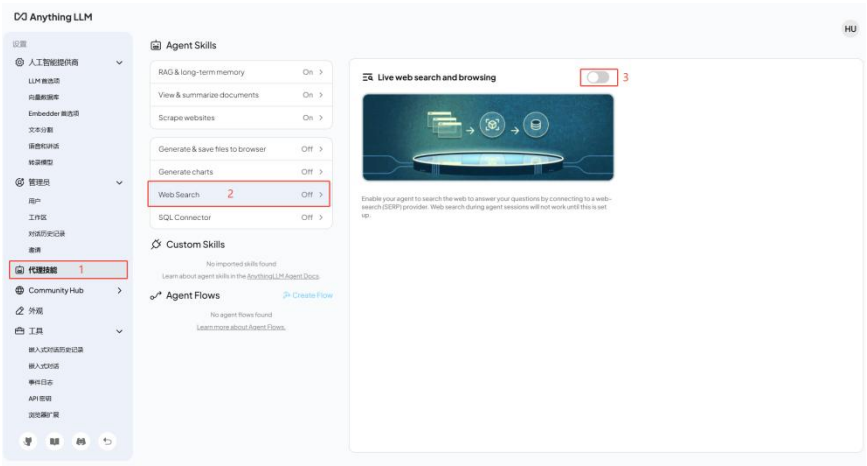
启用 AnythingLLM 的联网搜索服务，需要您提供一个联网搜索引擎 api key 来访问 Web 资源，目前 AnythingLLM 支持的搜索引擎可以参考如下步骤查看。

a. 打开 AnythingLLM 的联网搜索功能

- 首先点击 AnythingLLM 的全局设置按钮。

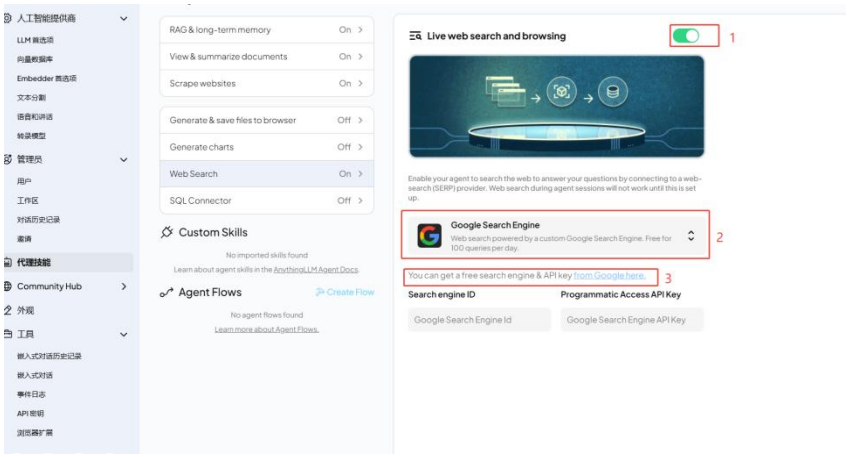


- 参考下图，找到“代理技能”，选择 agent skills 中的 Web Search，点击步骤 3，开启对应功能。



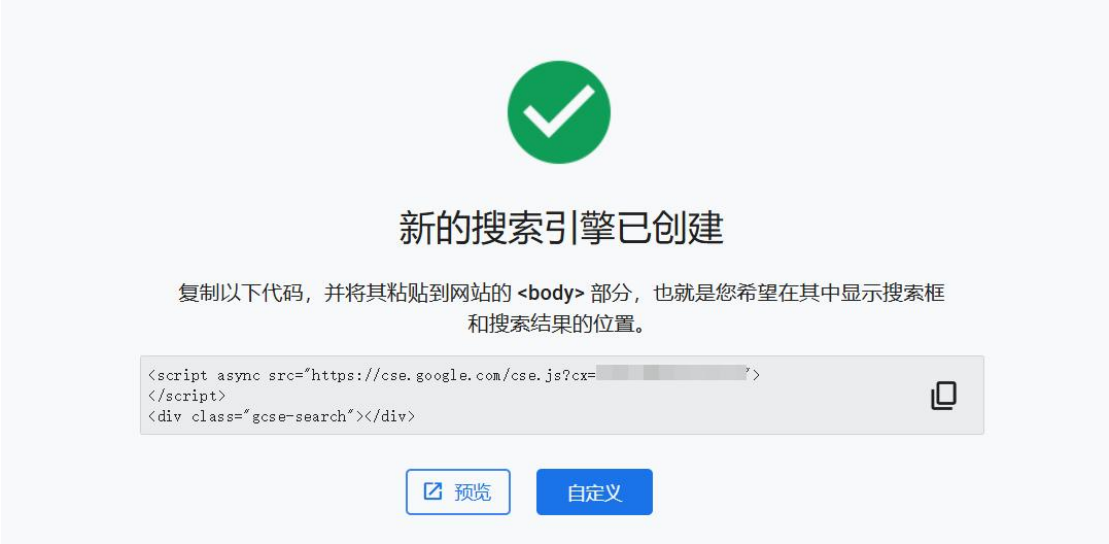
b. 获取 Google 搜索引擎的 ID 以及 API key

- 打开下图中步骤 3 对应的网页，按提示注册或登录账号、命名搜索引擎、选择搜索范围（可输入指定网页或搜索全网），以及选择是否启用图片搜索功能等操作。完成上述设置并通过人机验证后，点击创建，按提示获取搜索引擎的 ID 以及 API key。





- 创建成功后，得到如下图示内容：



- 进入自定义页面，获取搜索引擎的 ID。



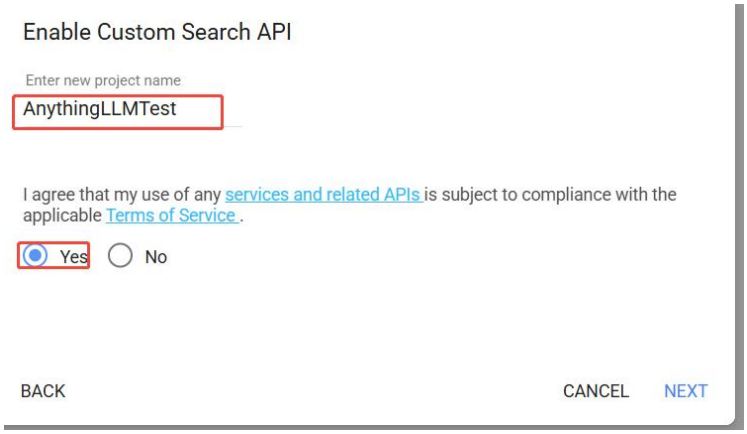
- 随后下滑到页面底部，点击程序化访问模块中的“开始使用”按钮。



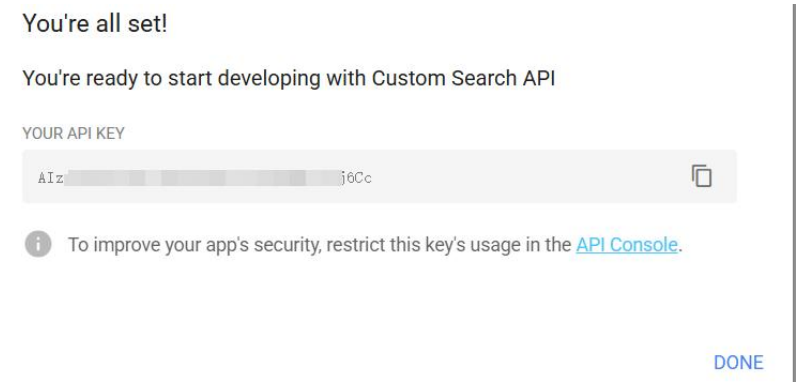
- 跳转到如下页面，点击获取密钥：



- 按照提示输入 api 名称，查看 api 使用协议并按照您的想法选择是否接受：

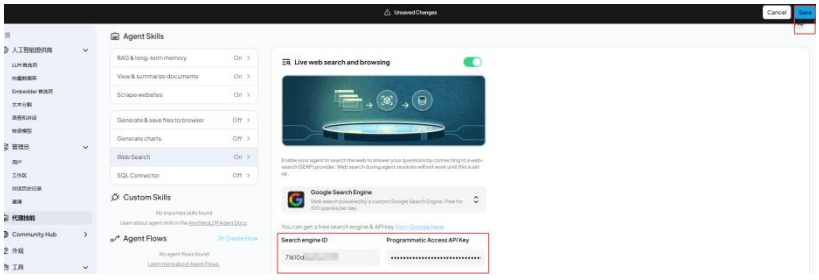


- 按照提示操作后得到 api 调用的 key。

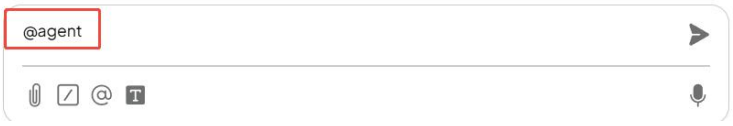


c. 启用 AnythingLLM 的联网搜索功能

- 配置 Google Search Engine ：回到 AnythingLLM 联网搜索的页面，填入上述步骤获取的 ID 和 API key，点击右上角保存。



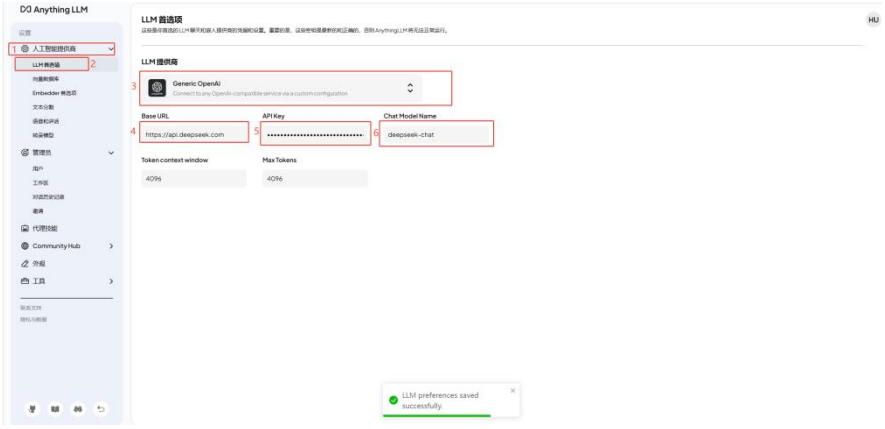
- 保存成功后得到如下提示：
- 在对话框以 @agent 启用联网功能：

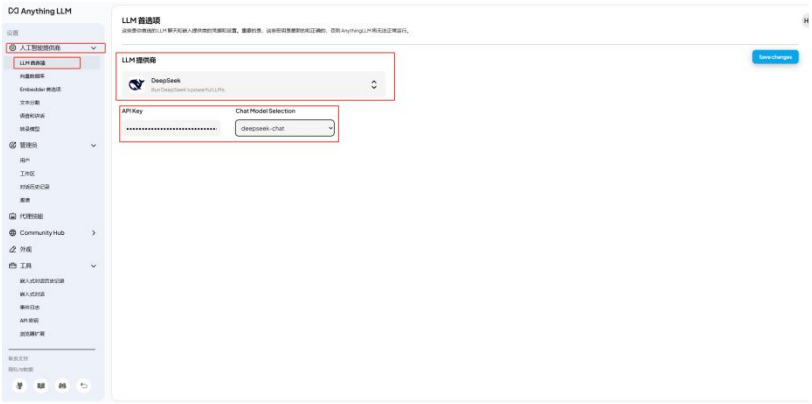


3. AnythingLLM 使用 DeepSeek API

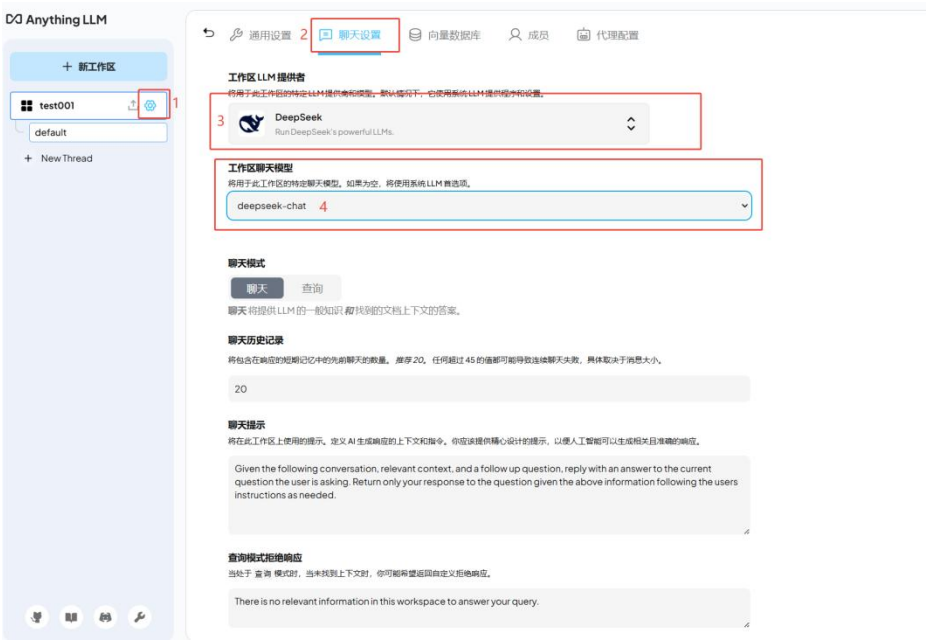
a. 前往 DeepSeek API 开放平台 注册一个账号，购买 API 使用资格。

b. 参考下述操作步骤，选择 Generic OpenAI 选项，填入 API 认证信息。或者选择 DeepSeek 选项，填入 API 认证信息。

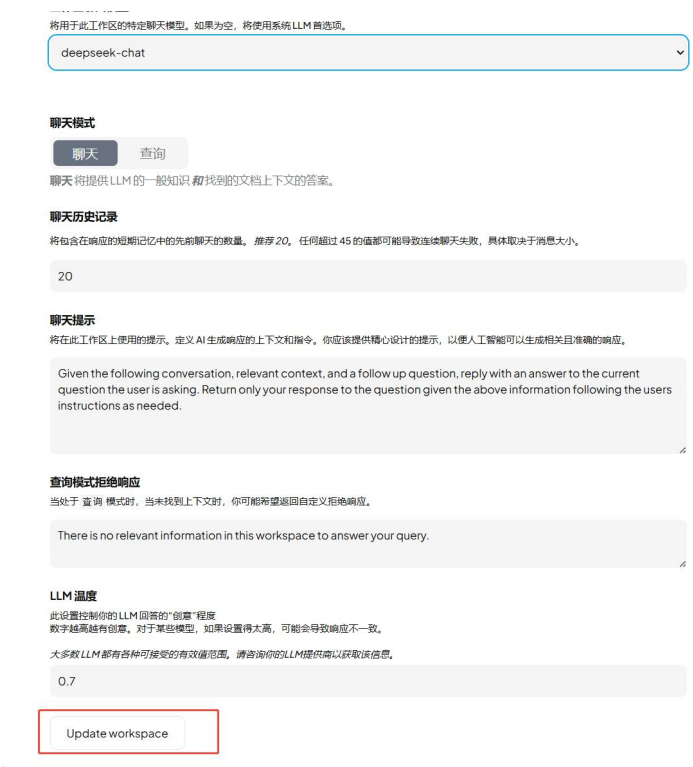


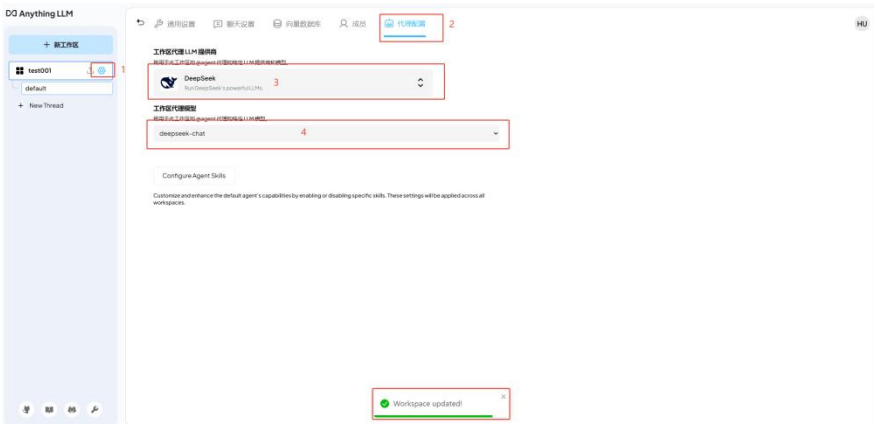


c. 到您想要使用的工作区，点击设置，更新当前聊天设置，以启用 DeepSeeK API 服务。

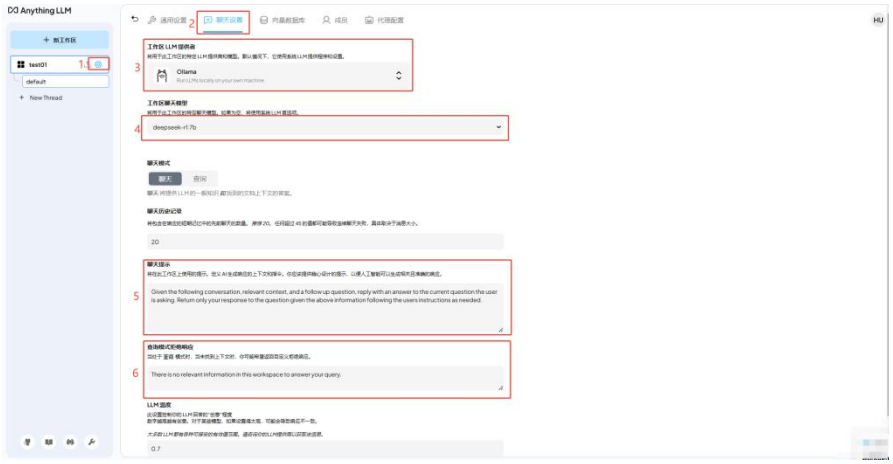


d. 需要点击下图按钮来保存配置。



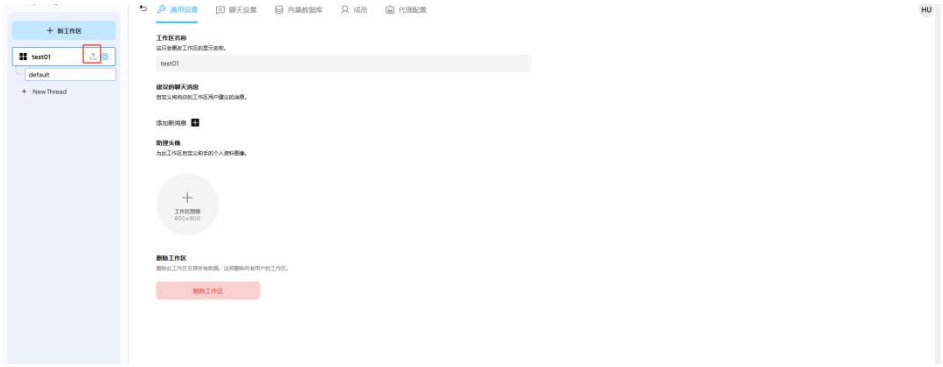


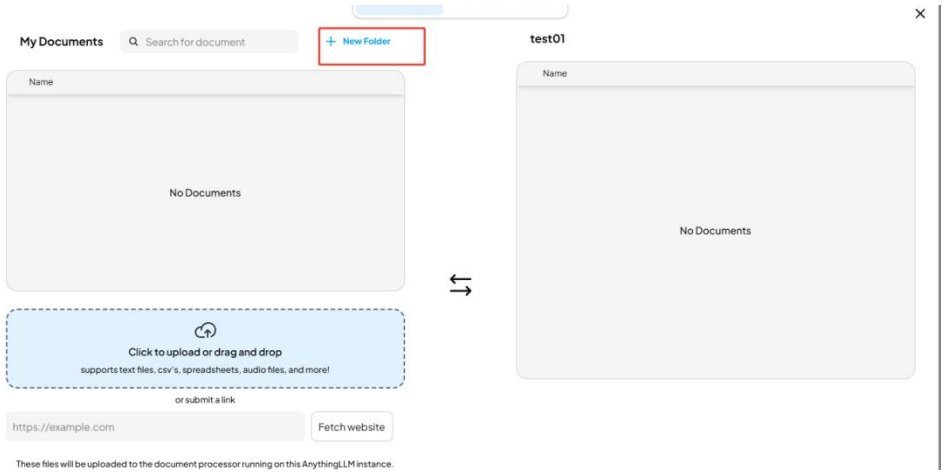
4. 调整对话模型及提示词

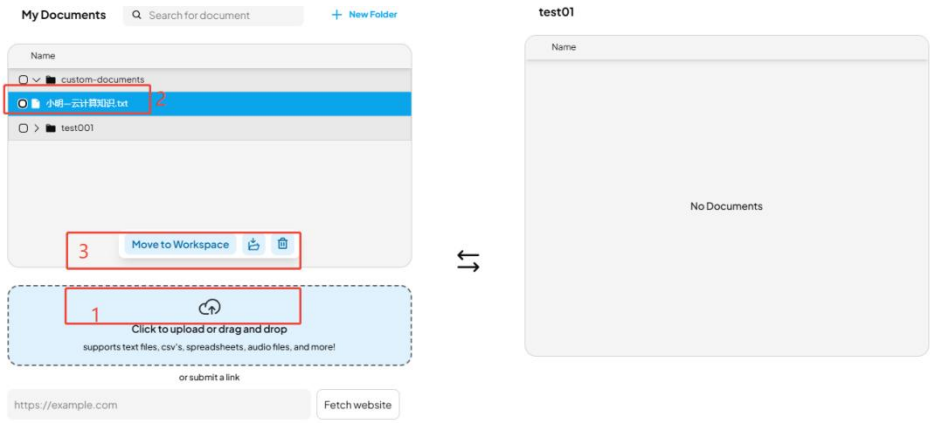


5. 知识库构建

a. 上传文件。







其中第 3 步需要第二部选择上传的文件后才会出现按钮，点击“Move to Workspace”，等待出现成功提示即可。



b. 基于知识库的对话。

首先选择上传了语料的工作区，随后直接提问即可。





返回 DeepSeek 专题导航。

3.3.1.7CherryStudio 协同 Ollama 启用 DeepSeek 大模型

应用场景

使用 CherryStudio 加载 DeepSeek 大模型的应用场景如下：

1. 学习研究 ：学生用其整理复习资料、辅助学术研究、助力论文写作；研究人员用于文献综述、分析问题等。
2. 工作办公 ：项目管理中生成报告、分析风险；内容创作时构思创意、撰写文案；代码开发里获取代码示例、解决问题。
3. 生活娱乐 ：作为智能聊天伙伴解闷陪伴，为创意活动如绘画等提供灵感启发。

4. 知识管理 ：个人可构建知识体系，团队能实现知识共享。

注意

建议使用天翼云提供的 DeepSeek 镜像创建云主机，减少安装过程中可能遇到的问题。镜像选择请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本-弹性云主机-最佳实践-AIGC 实践 - 天翼云。

操作步骤

一、 云主机端口配置更新

首先，您需要登录天翼云控制中心，对部署 DeepSeek 大模型的云主机添加新的端口开放操作。通过天翼云控制台进入部署了 DeepSeek 大模型的云主机详情页，选择安全组，添加新的开放端口，建议选择新加 114343 端口，该端口是 Ollama 服务默认监听的端口。添加完成后，可在安全组规则列表中查看新添加的 11434 端口规则，确认其状态为正常生效。







随后进入控制台，远程登陆即将安装 AnythingLLM 的云主机。



二、Cherry Studio 桌面软件安装

1、下载 Cherry Studio 桌面软件

访问 Cherry Studio 官网，选择合适的软件版本并下载，本案例中以 Cherry-Studio-0.9.27-setup.exe (88.1 MB) - Windows 标准安装包为例。



2、安装 Cherry Studio 软件

找到您下载的 Cherry Studio 安装包，按照提示进行安装。



三、Cherry Studio 实现自部署 DeepSeek 大模型对话

1、云主机 ollama 服务配置更新

对云主机上的 ollama 服务配置文件进行更新，添加第一部分中开放的端口。

如果您想同时在 OpenWebUI 以及其它服务上面启用，建议您使用 11434 端口，这样不用重启 OpenWebUI 服务。

```
1 # 编辑 ollama.service 服务参数，添加监听所有 ip 设置 2 vim /etc/systemd/system/ollama.service
```

添加内容如下：

```
1 Environment="OLLAMA_HOST=0.0.0.0"2 Environment="OLLAMA_ORIGINS=*"
```



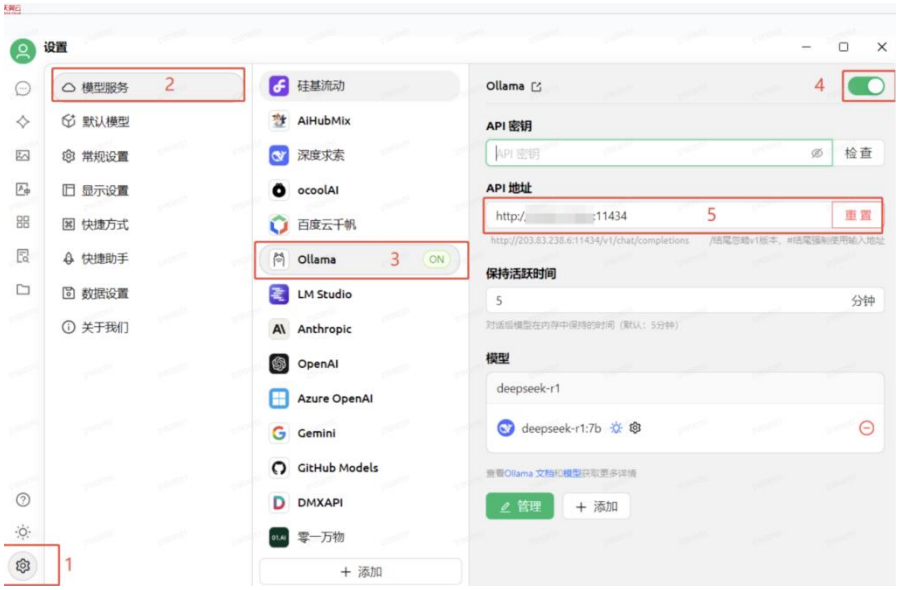
随后按照以下操作，重启 ollama 服务。

1 # 重新加载系统管理器的配置文件，使得 ollama.service 文件修改生效 2 sudo systemctl daemon-reload3 # 重启 ollama 服务 4 sudo systemctl restart ollama.service5 # 查看 ollama 服务监听端口 6 sudo netstat -tulnp | grep 11434

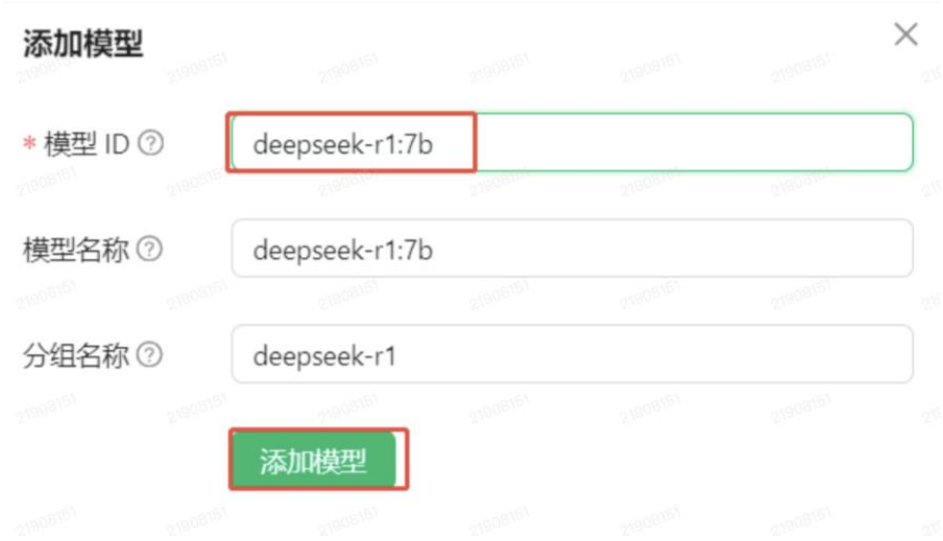
```
root@ecm-aigc-test:~# vim /etc/systemd/system/ollama.service
root@ecm-aigc-test:~# sudo systemctl daemon-reload
root@ecm-aigc-test:~# sudo systemctl restart ollama.service
root@ecm-aigc-test:~# sudo netstat -tulnp | grep 11434
tcp6       0      0 :::11434                :::*                    LISTEN      31184/ollama
root@ecm-aigc-test:~# vim /etc/systemd/system/ollama.service
```

2、Cherry Studio 中加载云主机上的 ollama 服务

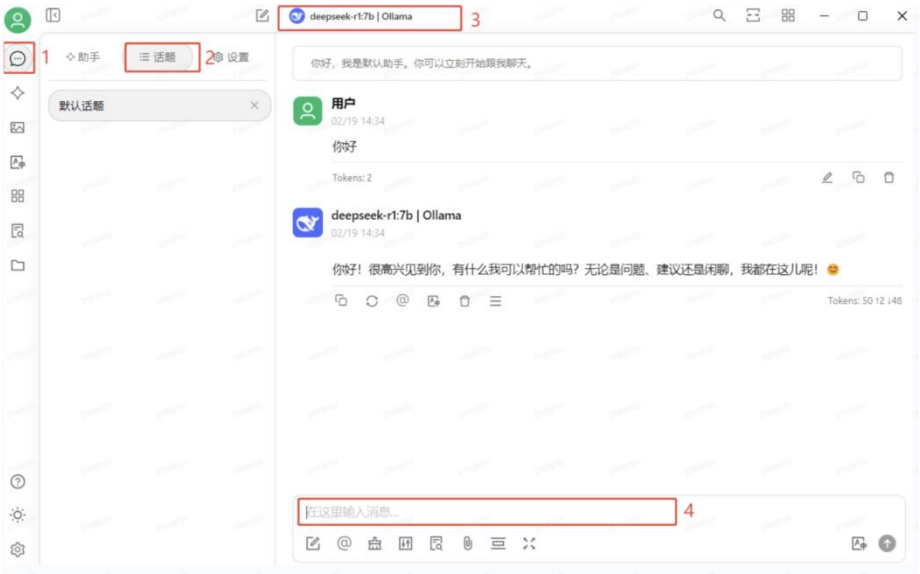
首先，打开安装好的 Cherry Studio 桌面程序，找到设置按钮，请参考下图步骤进行设置，第 5 步中的 ip 是您云主机公网的访问 ip。



随后点击下方的添加按钮，输入预装的 deepseek-r1:7b ，点击添加模型。



最后，按照下图提示进行对话吧。



FAQ

1、如何基于 Ollama 使用其它大模型？

请参考 ollama 模型站支持的大模型，在云主机上面下载您想要下载的大模型。

```
1 ollama pull $mode-name
```

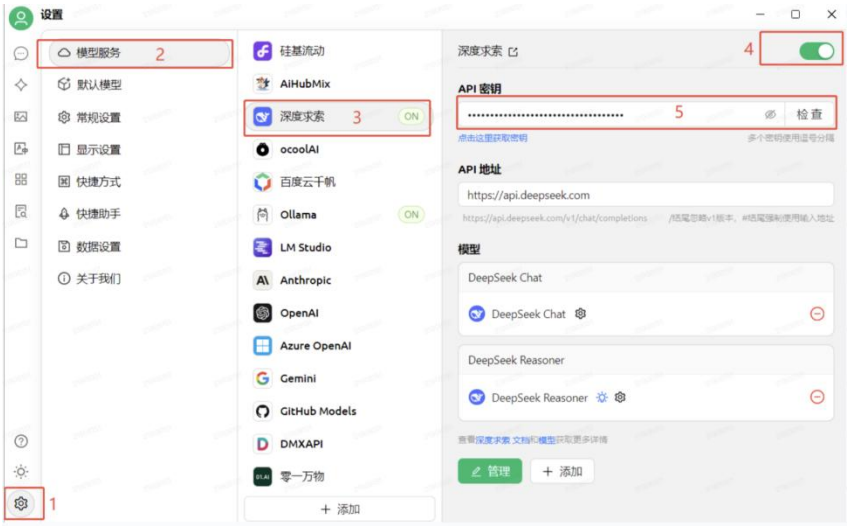
```
root@ecm-aigc-test:~# ollama pull deepseek-r1:1.5b
pulling manifest
pulling aabd4debf0c8...   0%
```

随后在 Cherry Studio 上面选择启用对应的大模型，相关操作参考下图。

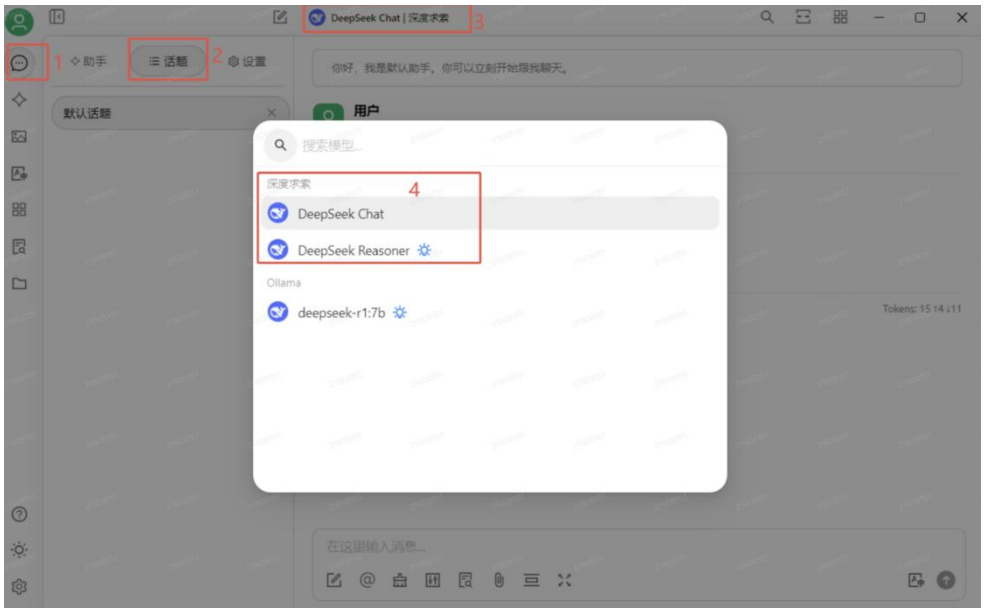


2、如何配置外部大模型？

参考下图步骤，填入您获取的其它大模型调用的 API 密钥，启用相应大模型服务。



随后，在对话框中选择对应的大模型，即可使用对应大模型服务。



3、如何启用联网搜索功能？

Cherry Studio 目前已经支持的联网模型服务商有：

- Google Gemini
- 腾讯混元
- 智谱 AI

注意

如果您想了解更多的 Cherry Studio 功能，请访问 Cherry Studio 文档。

返回 DeepSeek 专题导航。

3.3.1.8 天翼云 CPU 实例部署 DeepSeek-R1 模型最佳实践

1 背景

1.1 大模型算力需求

大模型因其参数规模庞大、结构复杂，通常需要强大的计算资源来支持其推理过程，这使得算力成为大模型应用的核心要素。随着 DeepSeek-R1 模型的问世，各行各业纷纷展开了关于如何接入大模型能力的广泛调研与探索，市场对大模型推理算力的需求呈现出爆发式增长的趋势。例如在医疗、金融、零售等领域，企业迫切希望通过接入 DeepSeek 大模型来提升决策效率和业务能力，从而推动行业的创新发展。在这一背景下，算力的供给和优化成为推动大模型落地应用的重要因素。

近年来,CPU 制程和架构的提升以及英特尔® 高级矩阵扩展 AMX（Advanced Matrix Extensions）加速器的面世带来了算力的快速提升。英特尔对大模型推理等多个 AI 领域持续深入研究，提供全方位的 AI 软件支持，兼容主流 AI 软件且提供多种软件方式提升 CPU 的 AI 性能。目前，已有充分的数据显示 CPU 完全可以用于大模型推理场景。

1.2 CPU 算力用于大模型推理的适用场景

CPU 适用于以下大模型推理场景：

场景 1：大模型推理需要的内存超过了单块 GPU 的显存容量，需要多块或更高配 GPU 卡，采用 CPU 方案，可以降低成本；

场景 2：应用请求量小，GPU 利用率低，采用 CPU 推理，资源划分的粒度更小，可有效降低起建成本；

场景 3：GPU 资源紧缺，CPU 更容易获取，且可以胜任大模型推理。

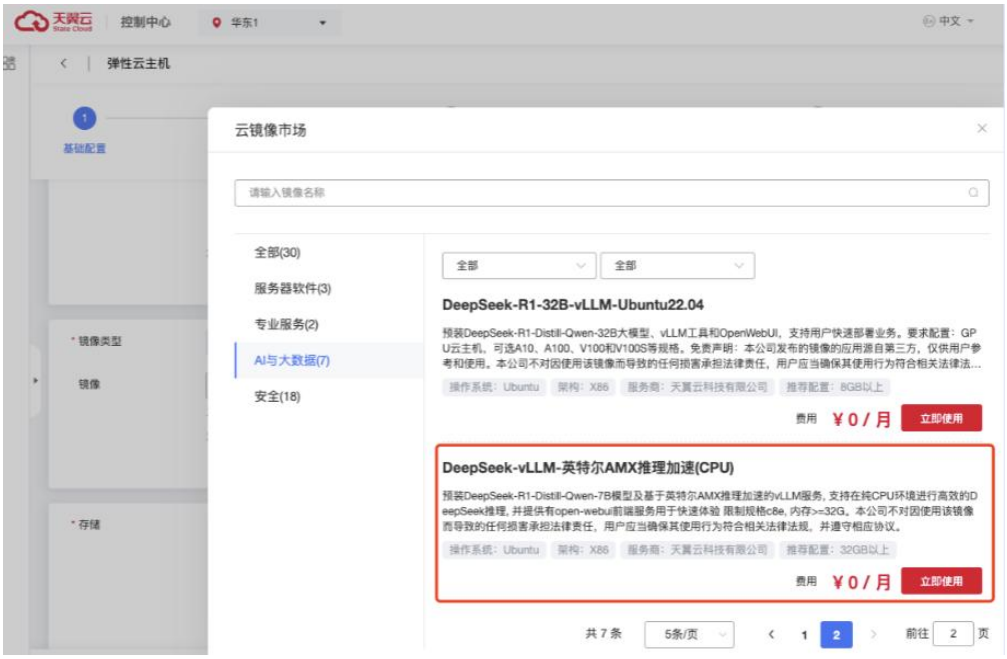
2 天翼云 EMR 实例 DeepSeek-R1-Distill-Qwen-7B 蒸馏模型部署实践

本节内容主要介绍如何在天翼云 EMR 实例上，基于 Intel® xFasterTransformer 加速库和 vllm 推理引擎完成模型部署，并展示相关性能指标。

2.1 服务部署

为了方便用户使用，天翼云联合英特尔制作了一键部署的云主机镜像，内置DeepSeek-R1-Distill-Qwen-7B模型、vLLM推理框架、xFT加速库以及open-webui前端可视环境。您可在天翼云控制台选择下列资源池和镜像，开通云主机进行体验。

类型	可用资源
可用资源池	华东-华东1-az2、华东-华东1-az3
可用规格	c8e系列（最低内存需求32G）
可用镜像	DeepSeek-vLLM-英特尔AMX推理加速(CPU)

类型	可用资源
	

完成云主机开通后，推理服务会在 5 分钟内自动启动，您无需进行任何其他操作。

注意

如需在云主机外访问服务，您需要绑定弹性 IP，并在安全组内放行 22/3000/8000 端口。

2.2 模型使用

2.2.1 open-webui 前端使用

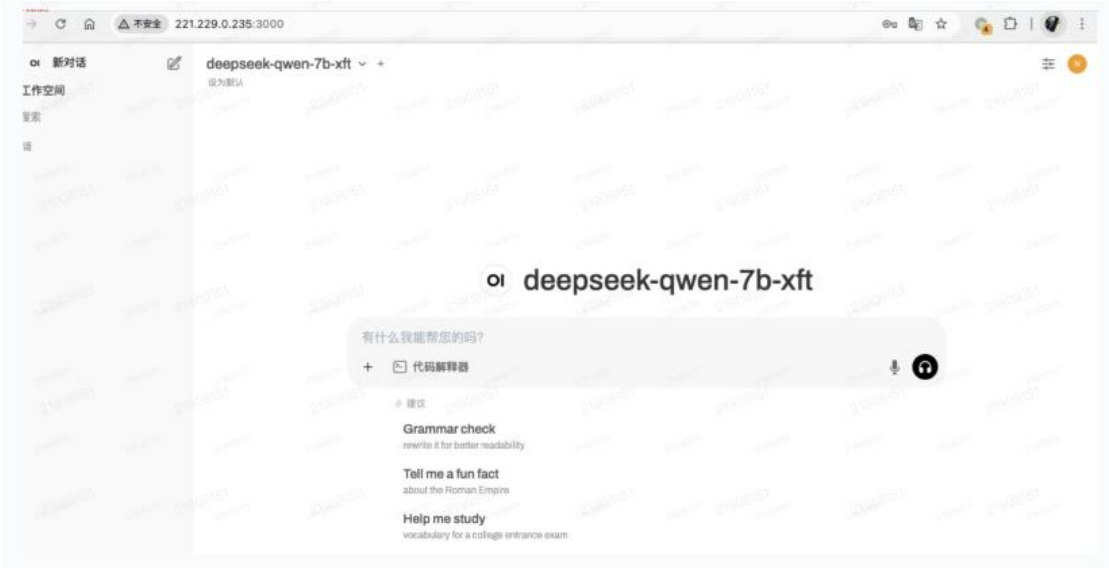
镜像已内置 open-webui，并已完成和 vllm 的连接配置，可直接通过以下地址进行访问:

```
http://[弹性 IP]:3000/
```

注意

1.首次打开页面时，您需要先完成管理员注册，以进行后续的用户管理。注册数据均保存在云实例的/root/volume/open-webui 目录下。

2.如果首次打开对话页面时没有模型可供选择,请您稍等几分钟让模型完成加载即可。



2.2.2 vllm api 调用

镜像内置 vllm 服务可直接通过如下地址访问：

```
# 根路径

http://[弹性 IP]:8000/

# 查询现有模型

http://[弹性 IP]:8000/v1/models

# 其他 api 接口参阅 vllm 文档
```

注意

vllm 服务配置有 API_KEY，您可在云实例的/root/recreate_container.sh 文件开头查看到当前值，并可进行修改以确保服务安全。

2.3 性能指标

借助于英特尔 AMX 的加速能力，本推理服务能够取得显著的性能提升，天翼云完成测试并给出参考指标如下：

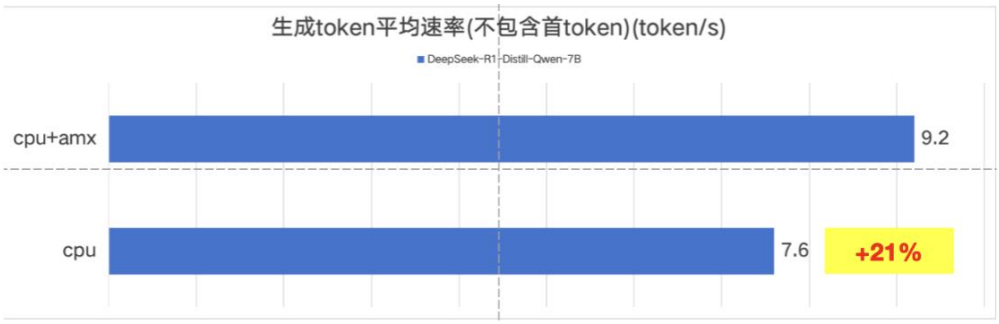
基本参数

- vcpu 数：24（物理核 12）
- 内存：64GB

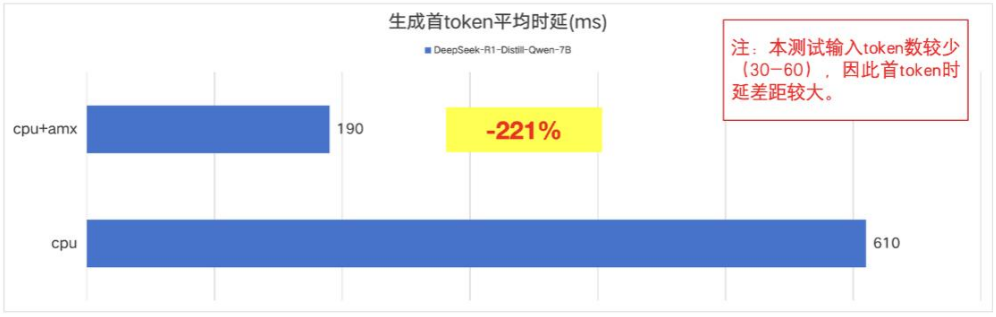
- 硬盘：60G 通用型 SSD
- 模型：DeepSeek-R1-Distill-Qwen-7B（bf16）
- batch size：1
- 输入 token 个数：30-60
- 输出 token 个数：256

性能数据

平均 token 生成速率：



首 token 时延：



在天翼云 c8e 系列 24vcpu 云主机上，启用 AMX 加速能力后，DeepSeek 7B 蒸馏模型（BF16）推理速度能够超越 9token/s，满足日常使用需求。

3 基于英特尔®至强®6 处理器部署满血版 DeepSeek-R1 671B 实践

3.1 性能指标

DeepSeek R1 671B 满血版模型以其卓越的性能，为用户带来了极致的效果体验，不过其部署成本也不容小觑。若采用传统的 GPU 部署方式，需要 8－16 张 GPU 才能提供足够的支持，这无疑大幅增加了硬件购置、能耗以及维护等方面的成本。在这样的背景下，天翼云基于英特尔®提供的至强®6 处理器服务器进行了 DeepSeek R1 671B 满血版 Q4_K_M 模型的部署尝试，测试结果如下：

- 1-instance 1-socket：

平均吞吐性能 9.7~10 token/s

- 2-instance 1-socket：

平均 7.32 token/s 和 7.38token/s， 共 14.7token/s

从上面测试数据可以看到，采用单实例单 socket 部署下，DeepSeek R1 671B 满血版模型可达到平均 9.7～10 token/s 的吞吐量，而在双实例部署模式中，总体吞吐量提升至 14.7 token/s。单颗 CPU 系统的吞吐性能可以达到普通用户正常使用的需要。

3.2 英特尔®至强®6 处理器简介

英特尔®至强®CPU 为 DeepSeek R1 671B 模型的部署提供了一个极具竞争力的方案。英特尔®至强®CPU 具备支持 T 级超大内存的能力，这使得它在权重存储与加载方面表现高效。对于像 DeepSeek R1 671B 这样的超大模型，其所需的显存容量在多卡 GPU 配置下才能满足，而英特尔®至强®CPU 能够凭借其强大的内存支持能力，为该模型提供良好的运行环境。

此外，DeepSeek R1 模型采用的 MOE（Mixture of Experts）结构，通过参数稀疏化的方式，使得在单 token 推理时仅需激活少量专家参数。这种特性显著降低了推理过程中的算力要求，与 CPU 的计算特点相契合，使得模型在 CPU 系统上的运行更加高效。这意味着在英特尔 ® 至强 ®CPU 上部署 DeepSeek R1 671B 模型，不仅能够充分发挥模型的性能优势，还能有效降低部署成本，避免了对大量 GPU 的依赖。

如需复现以上性能测试结果，请参看附录 2

总结

通过本次实践，无论是在天翼云 EMR 云实例上结合 xFasterTransformer 部署 DS R1 distill Qwen-7B 蒸馏模型，还是基于英特尔® 至强® 6 处理器部署满血版 DeepSeek-R1 671B 模型，均验证了 CPU 系统在 DeepSeek 大模型推理上的可行性和符合业界普遍要求的性能表现。CPU 系统不仅能够灵活应对不同规模的模型需求，无论是轻量化蒸馏模型还是全功能满血模型，都能高效满足用户场景需求，提供了一种低成本、经济高效的解决方案。

附录 1 英特尔® 至强® 可扩展处理器与 AI 加速技术

最新英特尔® 至强® 可扩展处理器产品

英特尔第五代® 至强® 可扩展处理器（代号 Emerald Rapids）——为 AI 加速而生

第五代英特尔® 至强® 处理器以专为 AI 工作负载量身定制的设计理念，实现了核心架构和内存系统的双重飞跃。其 64 核心设计搭配高达 320MB 的三级缓存（每核心由 1.875MB 提升至 5MB），相较上代缓存容量实现近三倍增长，为大规模并行 AI 推理提供充裕的本地数据存储空间。与此同时，处理器支持 DDR5-5600 高速内存，单路最大 4TB 的容量保证了大数据处理时的带宽和延迟优势。基于这些硬件提升，Emerald Rapids 整体性能较上一代提升 21%，AI 推理性能平均提升 42%，在大语言模型推理场景中可实现最高 1.5 倍的性能加速，同时大幅降低总拥有成本达 77%。

英特尔® 至强®6 处理器（代号 GNR Granite Rapids）——引领 CPU AI 算力革新

全新 GNR 处理器专为应对人工智能、数据分析及科学计算等计算密集型任务而设计。该产品在内核数量、内存带宽及专用 AI 加速器方面均实现重大突破：

- 核心与性能：每 CPU 配备多达 128 个性能核心，单路核心数较上一代翻倍，同时平均单核性能提升达 1.2 倍、每瓦性能提升 1.6 倍，进一步强化了 CPU 在大模型推理中的独立处理能力；
- AI 加速功能：内置英特尔® 高级矩阵扩展（AMX）新增对 FP16 数据类型的支持，使得生成式 AI 和传统深度学习推理任务均能获得显著加速；
- 内存与 I/O 突破：支持 DDR5-6400 内存及英特尔首款引入的 Multiplexed Rank DIMM（MRDIMM）技术，有效将内存带宽提升至上一代的 2.3 倍；同时，高达 504MB 的三级缓存和低延迟设计确保数据能够更快加载，为复杂模型训练和推理缩短响应时间。

英特尔® 至强® 6 处理器不仅通过更多的核心和更高的单线程性能提升了 AI 大模型推理能力，同时也能够作为机头 CPU 为 GPU 和其他加速器提供高速数据供给，进一步缩短整体模型训练时间。在满足混合工作负载需求的同时，其 TCO 平均降低 30%，大模型推理加速最高可达 2.4 倍。

无论是第五代至强还是全新的至强 6 处理器，英特尔均通过在核心架构、缓存系统、内存技术和专用 AI 加速器方面的全面革新，提供了业界领先的 AI 计算支持。

这两款产品为数据中心和高性能计算平台在 AI 推理、训练以及多样化工作负载下提供了强大而高效的算力保障

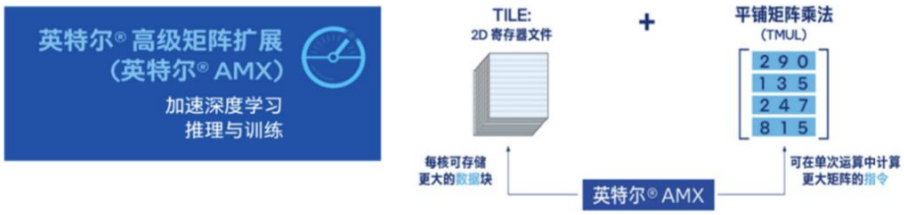


图 1 英特尔高级矩阵扩展（AMX）

英特尔全方位的 AI 软件生态支持

英特尔及其合作伙伴凭借多年 AI 积累，围绕至强®可扩展处理器打造了完善的软件生态：广泛支持主流开源框架，通过插件优化及多样化开源工具链，使用户在 x86 平台上能够轻松开发、部署通用 AI 应用，无需手动调整，同时确保从终端到云的全程安全保护。

此外，至强®处理器内置多种 AI 加速指令（如 AVX-512、AMX），使得任何兼容软件均可直接调用加速功能。开发者可免费下载英特尔分发版工具、库及开发环境，充分利用这些内置加速器应对各类 AI 管线需求。结合多样化硬件优势与开放生态，英特尔通过经济、可扩展的方案，将 AI 能力无缝延伸至云端与边缘。

其中，xFasterTransformer（xFT）是英特尔官方开源的 AI 推理框架，专为大语言模型在至强®平台上深度优化。xFT 不仅支持多种数据精度（FP16、BF16、INT8、INT4），还能利用多 CPU 节点实现分布式部署，显著提升推理性能并降低成本。其简单的安装和与主流 Serving 框架（如 vLLM、FastChat、MLServer、MindSpore Transformer、PaddlePaddle）的兼容性，帮助用户快速加速大模型应用。在 3.1 节中基于天翼云 EMR 云主机和 xFasterTransformer 加速引擎实现了对与 DeepSeek R1 蒸馏模型的高效推理部署。

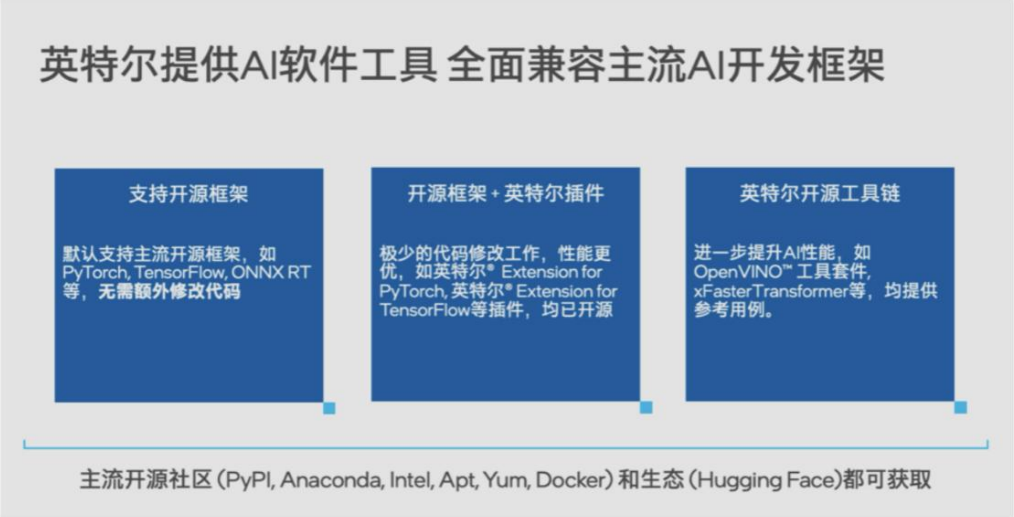


图 2 英特尔提供 AI 软件工具 全面兼容主流 AI 开发框架

附录 2 CPU 环境下部署 DeepSeek-R1 671B 模型实践

环境配置

硬件配置

- CPU：Intel® Xeon® 6980P Processor, 128core 2.00 GHz
- 内存 24*64GB DDR5-6400
- 存储 1TB NVMe SSD

软件环境

- OS: Ubuntu 22.04.5 LTS
- Kernel: 5.15.0-131-generic
- llama.cpp: github bd6e55b
- cmake: 3.22.0
- gcc/g++: 11.4.0
- Python: 3.12.8
- git: 2.34.1

BIOS里关闭 sub NUMA 相关配置。

注意

版本是指本测试中服务器上安装的版本，并非要求的最低版本。

部署步骤

1. 安装 llama.cpp

参考 llama.cpp 官网的安装说明，我们的步骤如下。

```
# 下载 llama.cpp 推理框架源码

git clone https://github.com/ggerganov/llama.cpp.git cd llama.cpp

# 预先准备 intel oneapi 库 source /opt/intel/oneapi/setvars.sh

# 基于 oneapi 库对 llama.cpp 进行编译

cmake -B build -DGGML_BLAS=ON -DGGML_BLAS_VENDOR=Intel10_64lp
-DCMAKE_C_COMPILER=icx -DCMAKE_CXX_COMPILER=icpx -DGGML_NATIVE=ON

cmake --build build --config Release -j$nproc
```

2. 下载模型文件

我们直接使用了社区制作的 DeepSeek 671B 满血模型的 Q4 量化版，您也可以下载 DeepSeek 官方 BF16 版本，并通过 llama.cpp 提供的脚本转换为 GGUF 格式。社区提供了从 1bit 到 8bit 不同版本的量化选项，具体区别可以参考社区网页。我们选择了使用最受欢迎的 Q4_K_M 版本。如果追求最佳效果，建议使用 Q8_0 版本。

Filename	Quant type	File Size	Split	Description
DeepSeek-R1-Q8_0.gguf	Q8_0	713.29GB	true	Extremely high quality, generally unneeded but max available quant.
DeepSeek-R1-Q4_K_M.gguf	Q4_K_M	404.43GB	true	Good quality, default size for most use cases, *recommended*.

```
# 下载 unsloth 制作的社区版量化模型(hf-mirror 和 modelscpoe 源都可)
```

```
git clone --no-checkout https://hf-mirror.com/unsloth/DeepSeek-R1-GGUF
cd DeepSeek-R1-GGUF/# 建议 nohup 执行，预计至少需要半天时间，同时确保磁盘容量足够 400G.
git lfs pull --include="DeepSeek-R1-Q4_K_M/*"
```

Q4_K_M 版本的文件大小为 404.43GB，下载过程可能会比较耗时。下载完成后，您可以在 DeepSeek-R1-Q4_K_M 目录下找到一系列 .gguf 文件，例如 DeepSeek-R1-Q4_K_M-00001-of-00009.gguf。

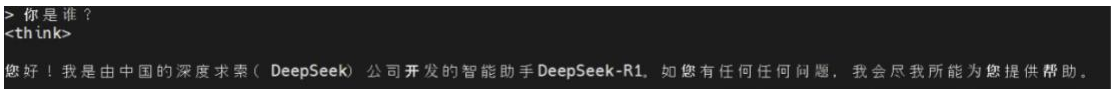
3. 模型加载和运行

使用 llama-cli，指定模型文件路径并启用交互模式，DeepSeek R1 满血版就可以在 CPU 上顺利运行了。

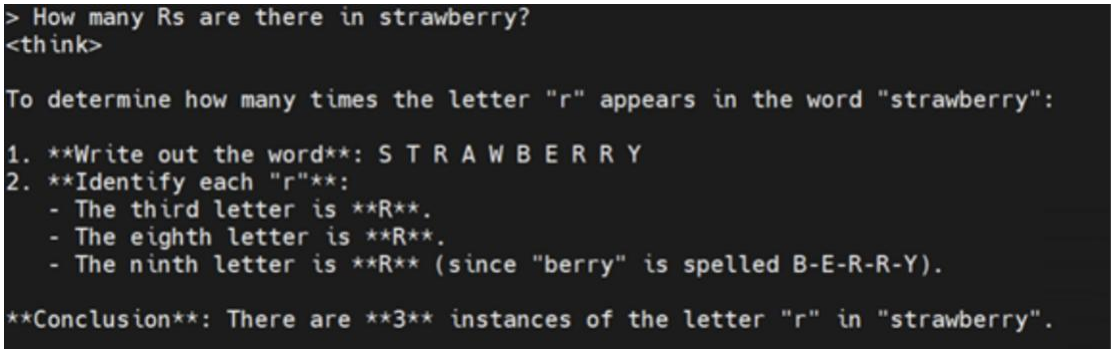
```
build/bin/llama-cli -m
/tmp/DeepSeek-R1-Q4_K_M/DeepSeek-R1-Q4_K_M-00001-of- 00009.gguf -i
```

下面用几个示例展现 DeepSeek-R1 671B 满血版强大的的 reasoning 推理能力：

测试模型自我认知：



验证推理能力的经典“草莓”问题：



“等灯等灯”的意思：



4.性能及优化

那么 CPU 运行满血版 R1 的性能怎么样呢 ?我们做了相关性能测试。对于 Q4_K_M 模型，使用如下命令进行:

```
export MODEL_PATH=/tmp/DeepSeek-R1-Q4_K_M/DeepSeek-R1-Q4_K_M-00001-of-00009.gguf
numactl -C 0-127 -m 0 ./build/bin/llama-cli -m $MODEL_PATH -t 128 --temp 0.6 -s 42 -no-cnv --no-kv-offload -ctk q4_0 -c 8192 -n 128 -p "<|User|>以孤独的夜行者为题写一篇 750 字的散文，描绘一个人在城市中夜晚漫无目的行走的心情与所见所感，以及夜的寂静给予的独特感悟。<|Assistant|>"
```

这里使用numactl 来指定使用单路 CPU（0-127，6980P 有 128 核），以及这一路 CPU 的内存节点（numa0），避免跨 numa 传输以获取最佳性能。

llama.cpp 是本地编译的，编译的时候使用Intel oneAPI 可以有效提升它的性能。英特尔尝试用了 oneAPI里的 Intel C++编译器和数学加速库 MKL，结合 jemalloc 内存管理优化，**推理速度可以达到每秒 9.7~10 词元**(TPS，tokens per second)。

上面的实验是在单路 CPU 上进行的，**我们又在两路 CPU 上各自独立启动 1 个模型实例，总速度可以达到 14.7TPS** （7.32TPS+7.38TPS)。

再进一步，英特尔观察到基于现有的 llama.cpp 软件方案，在 CPU 平台没有实现高效的专家并行和张量并行等优化，CPU 核心利用率和带宽资源没有充分发挥出

来，6980P 的 128 核心运行1 个模型还有不少性能储备。预计可以继续增加实例数来获得更好的总 TPS。

另外，通常情况下，CPU 的 SNC (Sub-NUMA Clustering)设置可以获得更高的带宽，但是受限于软件并未优化实现良好匹配，此次实验关闭了 SNC 测试。

以下方式的系统配置也有助于提升性能：

- BIOS里关闭 AMP prefetcher
- 用cpupower 打开 CPU 的 pstate 性能模式
- 提高CPU 的 uncore 频率
- 关闭超线程(Hyper-Threading)

注意

为了加快试验进度，我们限制了词元输出长度(-n 128)。经过验证，增加输出长度(例如-n 512)对于生成性能指标的影响不大。

返回 DeepSeek 专题导航。

3.3.2GPU 物理机：物理机搭建 DeepSeek 指南

GPU 物理机：物理机搭建 DeepSeek 指南

3.3.3SD-WAN 跨境：SD-WAN 助力 DeepSeek 模型定向加速

SD-WAN 跨境：SD-WAN 助力 DeepSeek 模型定向加速

3.3.4 智算容器：云容器引擎与 DeepSeek 融合实践

智算容器：云容器引擎与 DeepSeek 融合实践

3.3.5 函数计算：天翼云函数计算与 DeepSeek 大模型

函数计算：天翼云函数计算与 DeepSeek 大模型

3.4Q&A：典型问题解析与策略应对

天翼云提供了底层资源，DeepSeek 模型还是要客户自己部署的吗？

天翼云提供了预置 DeepSeek 模型的镜像，开机即用。

DeepSeek 模型、Ollama 工具下载慢，我该怎么办？

目前使用天翼云的镜像源，可加快访问速度。如果想使用其他模型，也可以自定义部署。

昇腾版本和英伟达版本有啥区别？

昇腾和英伟达版本主要区别在于硬件设备（如 A100 和昇腾），具体的参数层面区别难以量化。

开箱即用的镜像，能否更换其他参数规模的模型？

可以的，GPU 云主机提供了预置 DeepSeek R1:7B 和 DeepSeek R1:70B 两款模型镜像，提供了 DeepSeek LLamaFactory 模型微调镜像，并配备完整的 ollama、openWebUI 工具，用户可根据需要进行自定义部署。

模型部署过程中发现云盘的容量不够怎么办？

根据云硬盘扩容概述对已有云盘进行扩容或购买数据盘进行挂载。

天翼云公有云哪些资源池有 DeepSeek 预置镜像？

目前 DeepSeek R1:7B Ubuntu 云主机镜像，已上线至福州 25、郑州 5、长沙 42、上海 36、华北 2、华南 2、西南 1、华东 1，其余资源池按需加载；裸金属镜像 DeepSeek R1:7B Ubuntu 已上线上海 15 资源池，其余资源池按需加载。

有没有 vLLM 部署指导？

请参考：在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本。

返回 DeepSeek 专题导航。

4 产品概述

4.1 产品定义

弹性云主机（CT-ECS，Elastic Cloud Server）是一种可随时获取、弹性可扩展的计算服务。云主机由 CPU、内存、镜像、云硬盘等组成，同时结合 VPC、安全组、数据多副本保存等能力，打造一个既高效又可靠安全的计算环境，确保服务持久稳定运行。

弹性云主机的开通和使用非常便捷、高效，您只需要指定 CPU、内存、镜像等配置信息。开通完成后，您就可以像使用本地 PC 或物理服务器一样在云上使用弹性云主机，同时您也可以按需随时调整弹性云主机的规格配置。

产品架构

弹性云主机，整合了计算、存储与网络等资源，通过和其他产品、服务组合，为用户提供安全可靠、高性能、弹性可伸缩的虚拟计算服务，可覆盖各种不同量级的业务场景。

- 产品使用入口：您可以通过控制台、openAPI 管理和使用弹性云主机。
- 计费方式：您可以按需选择弹性云主机计费方式，可选包年包月或按量计费。
- 地域/可用区：您可以按需选择创建云主机的地域和可用区。
- 云主机组：可以在创建弹性云主机时，通过加入云主机组。通过设置反亲和策略，将云主机尽量分散地部署在不同的宿主机上，以提高业务的可靠性。
- 操作日志：开启云审计服务后，系统会记录弹性云主机相关的操作事件，您可以在控制台操作日志界面进行查看。
- 配额管理：为防止资源滥用，平台限定了各服务资源的配额，对用户的资源数量和容量做了限制。如您最多可以创建多少台弹性云主机、多少块云硬盘。如果当前资源配额限制无法满足使用需要，您可以申请扩大配额。
- 访问控制：可以进行精细的权限管理，以达到不同用户之间的权限隔离。
- 镜像服务：镜像包含了初始系统环境、应用环境、软件配置等信息，您通过公共镜像、私有镜像、共享镜像、安全产品镜像或应用镜像批量创建弹性云主机，可以实现业务的快速部署。
- 快照服务：快照指的是云硬盘数据在某个时刻的完整拷贝或镜像，是一种重要的数据容灾手段，当数据丢失时，可通过快照将数据完整的恢复到快照时间点。
- 云硬盘：可以像物理硬盘一样分区格式化，满足大部分通用业务场景下的数据存储需求。

- 云备份：为弹性云主机提供简单易用的备份服务，当发生病毒入侵、人为误删除、软硬件故障等事件时，可将数据恢复到任意备份点。
- 虚拟私有云：通过虚拟私有云建立专属的网络环境，设置子网、安全组，并通过弹性公网 IP 实现外网链接。
- 安全组：安全组是一个逻辑上的分组，为具有相同安全保护需求并相互信任的云主机提供访问策略。
- 弹性伸缩：可以灵活调整弹性云主机数量，高效匹配业务要求。
- 云迁移：支持将客户云下或其他云的计算资源迁移至天翼云。
- 云监控：通过云监控用户可以实时观测弹性云主机的使用情况。

为什么选择弹性云主机

选择弹性云主机，您可以轻松构建具有以下优势的计算资源：

- 无需自建机房，无需采购以及配置硬件设施。
- 分钟级交付，快速部署，缩短应用上线周期。
- 成本透明，按需使用，支持根据业务波动随时扩展和释放资源。
- 提供 GPU 云主机以及通用 x86/ARM 架构云主机。
- 支持通过内网访问其他天翼云服务，形成丰富的行业解决方案，降低公网流量成本。
- 提供安全组、ACL、角色权限控制、防流量攻击等多重安全方案。
- 提供性能监控框架和主动运维体系。
- 提供行业通用标准 API，提高易用性和适用性。

管理方式

天翼云提供如下方式进行云主机的配置和管理：

- 控制台：天翼云提供 Web 化的服务管理平台，即控制台。
- OpenAPI：天翼云提供基于 HTTPS 请求的 API（Application programming interface）管理方式。

4.2 术语解释

实例

实例即弹性云主机，是由 CPU、内存、操作系统、云硬盘组成的基础的计算组件。

镜像

镜像是一个包含了软件及必要配置的弹性云主机模板，至少包含操作系统，还可以包含应用软件（例如，数据库软件）和私有软件。通过镜像，您可以创建弹性云主机。

公共镜像

公共镜像为系统默认提供的镜像，常见的标准操作系统镜像，所有用户可见，包括操作系统以及预装的公共应用。公共镜像具有高度稳定性，皆为正版授权，请放心使用，您也可以根据实际需求自助配置应用环境或相关软件。

私有镜像

私有镜像为用户自己创建的镜像。包含操作系统或业务数据、预装的公共应用以及用户的私有应用的镜像，仅用户个人可见。

云硬盘

云硬盘（Elastic Volume Service，EVS）可以为弹性云主机提供高可靠、高性能、规格丰富并且可弹性扩展的块存储服务，满足不同场景的业务需求，适用于分布式文件系统、开发测试、数据仓库以及高性能计算等场景。

安全组

安全组是一个逻辑上的分组，为具有相同安全保护需求并相互信任的弹性云主机提供访问策略。安全组创建后，用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，即受到这些访问规则的保护。

Cloud-init

Cloud-init 是开源的云初始化程序，能够对新创建弹性云主机中指定的自定义信息（主机名、密钥和用户数据等）进行初始化配置。

通过 Cloud-init 进行弹性云主机的初始化配置，将对您使用弹性云主机、镜像服务和弹性伸缩产生影响。

CPU 积分

CPU 积分是一种用来衡量云主机计算、存储以及网络配置利用率的方式。云主机利用 CPU 积分机制保证云主机基准性能，解决超分云主机长期占用 CPU 资源的问题。

云备份

云备份（Cloud Backup and Recovery，CBR）可以为云主机、云硬盘提供简单易用的备份服务，当发生病毒入侵、人为误删除、软硬件故障等事件时，可将数据恢复到任意备份点。

快照

指定的单个或多个云硬盘（系统盘或数据盘），可以支持快速备份和恢复快速保存指定时刻云硬盘的数据。

虚拟私有云

虚拟私有云（Virtual Private Cloud，以下简称 VPC）为弹性云主机构建了一个逻辑上完全隔离的专有区域，用户可在自己的逻辑隔离区域中定义虚拟网络，为弹性云主机构建一个逻辑上完全隔离的专有区域。也可在 VPC 中定义安全组、VPN、IP 地址段、带宽等网络特性，方便管理、配置内部网络，进行安全、快捷的网络变更。还可以自定义安全组内与组间弹性云主机的访问规则，加强弹性云主机的安全保护。

主弹性网卡

在创建云主机实例时，随实例默认创建的弹性网卡称作主弹性网卡。主弹性网卡无法与实例进行解绑。

扩展弹性网卡

创建扩展弹性网卡，将其附加到云主机实例上，也可将其从实例上进行解绑。每台实例可附加的扩展弹性网卡数量由实例规格决定。

密钥对

密钥对即 SSH 密钥对，包含一个公钥和一个私钥，是为用户提供远程登录云主机的认证方式，是一种区别于传统的用户名和密码登录的认证方式。

标签

标签是弹性云主机的标识。为弹性云主机添加标签，可以方便用户识别和管理拥有的弹性云主机资源。

4.3 产品优势

快捷易用，动态调整计算资源

- 即开即用：无论一台还是百台，均可在几分钟内开通使用；
- 大规模开通：通过调用云主机 API，可在十分钟内开通千台实例；
- 弹性伸缩：提供弹性伸缩服务，可根据业务灵活调整云主机数量。

配置丰富，支持根据业务按需选购

- 多种实例规格：提供通用计算型、计算增强型、内存优化型、GPU 型等多种规格，同时支持规格升级，面向各类场景满足用户的实时需求；
- 多种存储类型：支持单独购买云硬盘挂载到云主机上，并提供多种存储类型，满足不同的 I/O 性能要求。

专有网络，可灵活自定义网络空间

- 通过隧道技术实现 100%二层网络隔离，满足不同行业客户的安全隔离需要；
- 基于安全组和访问 ACL 进行三层以上网络双重访问控制，满足行业用户的网络安全规范；
- 自定义子网网段划分、IP 地址及路由策略，按需配置，即开即通；
- 支持专线、VPN 等多种方式接入云端。

安全稳定，多重防护保障数据可靠性

- 高达 99.9999999%的数据持久性；
- 云硬盘采用三副本技术，支持云硬盘自动备份，进一步保障数据安全性；
- 提供云主机安全、网站安全服务等精细化专业防护能力。

多种镜像，助力业务批量快速部署

- 支持 Linux、Windows 等多种公有镜像；
- 支持用户上传私有镜像，也可以将云主机导成私有镜像，实现业务的自定义与批量快速部署；
- 支持共享镜像，方便多用户统一部署。

高效管理，多维监控保驾护航

- 多种管理方式：提供可视化管理平台及控制台、远程终端等多种管理方式，操作便捷；
- 多维度监控：支持多维度资源监控，提供全面监控及告警机制，方便数据出现异常及时处理。

高性价比，尊享优质服务

- 多种计费模式：提供包年/包月、按量付费等多种计费模式，根据需求灵活控制成本；
- 合理控制：可根据业务自动调整云主机数量，有效控制资源投入；
- 优质服务：五星级电信 IDC 环境，专业运营团队 7*24 小时服务。

多样异构芯片，更优算力选择

- 多种芯片：覆盖多款因特尔芯片、AMD 芯片以及鲲鹏、飞腾、海光等主流国产芯片。
- 多种模型：提供基于 CPU 芯片具备通用计算能力的云主机规格，也提供基于 GPU 芯片具备的异构计算能力的云主机规格，能满足不同的计算模型需求。

4.4 弹性云主机选型

弹性云主机选型背景介绍

进行弹性云主机选型之前，您需要根据资源需求、可用性和可靠性、成本等因素，做出性价比最优的决策。通常需要考虑以下几个方面：

- 预估资源需求：针对不同的业务需求，结合业务的使用场景、负载大小、流量等因素，预估需要的 CPU、内存、存储等资源，以此来选择合适的弹性云主机实例类型。
- 可用性要求：根据业务可用性要求，选择不同可用区或者不同地域的弹性云主机部署业务，提高业务系统的容灾能力。

- 可靠性要求：选择弹性云主机实例时，综合考虑业务架构整体的可靠性要求，选择与数据库、中间件等产品相匹配的弹性云主机，确保实例能够持续稳定地运行。
- 考虑成本：不同的实例类型价格不同，一般来说更大的内存、vCPU、带宽价格更高，需要权衡性价比，根据实际需求选择合适的实例类型。

规格介绍

天翼云弹性云主机提供了多种实例规格类型，不同规格类型的弹性云主机具有不同的性能、价格。根据 vCPU、内存、网络性能、存储吞吐量等配置划分，每种规格类型又包括多个实例规格可供选择。通过这样的方式，您可以根据自己的需要选择最适合自己业务场景的弹性云主机实例，以获得最佳的性价比。

按照特性可分为通用型、计算增强型、内存优化型、GPU 型等。实例规格名称格式为 x<规格类型>x<代系>.xxxx<vCPU 个数>.x<内存数/vCPU 数比>：

- <规格类型>：如 s、c、m、p、g 等，代表弹性云主机规格类型。其中 s 为通用型，c 为计算增强型，m 为内存优化型，p 为 GPU 计算加速型，g 为 GPU 图形加速基础型。
- <代系>：如 2、6 等，代表不同硬件规格，更大数字一般拥有更高的性能。
- <vCPU 个数>：如 small、medium、large、xlarge 等，代表 vCPU 核数。其中，medium 为 1 vCPU，large 为 2 vCPU，xlarge 为 4 vCPU，small 特指 1 vCPU 1G 内存的小规格。
- <内存数/vCPU 数比>：如 1、2、4、8 等，代表内存 GB 数和 vCPU 的比例。

例如，s7.2xlarge.2 表示通用型弹性云主机，具有 8 核 CPU 和 16GB 内存。更多类型的特性和具体规格可参考页面产品规格。

场景选型

常见业务选型推荐

业务场景	选型原则	推荐实例类型
Web 服务器、开发测试环境以及小型数据库应用等场景。	计算、存储、内存等资源需求平衡，无特定配置要求。	通用型，推荐最小规格为 2 核 4G 通用型云主机
计算性能高且稳定的企业级应用，如大规模数据处理、在线游戏等。	对 CPU 或内存等计算资源要求较高，具有性能稳定且资源独享的特点。	计算增强型，推荐最小规格为 2 核 4G 通用计算增强型云主机
大型数据库、高性能计算、NoSQL 等场景。	内存要求较高、需要大量数据计算。	内存优化型，推荐最小规格为 2 核 16G 内存优化型云主机
在深度学习、高性能数据库、计算流体动力学、计算金融、地震分析、分子建模、基因组学等领域场景，适用于科学计算等。	采用 GPU 直通技术，能够提供超高的通用计算能力。选型详见：NVIDIA 驱动安装指引-GPU 云主机-用户指南-安装 NVIDIA 驱动 - 天翼云 (ctyun.cn)。	GPU 计算加速型，推荐最小规格为 16 核 64G GPU 计算加速型云主机
高清视频、图形渲染、远程桌面等场景。	提供图形处理器(GPU)及较高的计算性能配置，搭配相关驱动可以适用于图形渲染要求较高的应用。选型详见:NVIDIA 驱动安装指引-GPU 云主机-用户指南-安装 NVIDIA 驱	GPU 图形加速基础型，搭配 GRID 驱动，推荐最小规格为 8 核 24G GPU 图形加速基

	动 - 天翼云（ctyun.cn）。	础型云主机

注意

选择区域：不同区域的云服务之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。具体操作请参见如何选择区域。

选择可用区：同一可用区内网互通，不同可用区之间物理隔离。如果您需要提高应用的高可用性，建议您将弹性云主机创建在不同的可用区内。具体操作请参见如何选择可用区。

验证与调整

当您完成选型并开始使用弹性云主机实例后，建议您根据一段时间的性能监控信息，验证所选实例规格是否合适。

假设您选择了 s7.2xlarge.2，通过监控发现实例 CPU 使用率一直较高，建议您登录实例检查 CPU 的具体情况，还可以根据具体情况调整处理器与内存的大小，以满足业务需求。具体操作请参见云监控服务快速入门。

相关搭配产品推荐

弹性伸缩服务

弹性伸缩服务（CT-AS，Auto Scaling）是根据您的业务需求，通过策略自动调整计算资源的管理服务。该服务能够根据预设规则自动调整伸缩组内的弹性云主机数量，在业务需求上升时自动增加弹性云主机实例，业务需求下降时自动减少弹性云主机实例。

弹性文件服务

弹性文件服务（CT-SFS，Scalable File Service）提供按需扩展的高性能文件存储，可为云上多个弹性云主机、容器、物理机提供共享访问，具备高可用性和高数据持久性，为海量的小文件、低延迟高 IOPS 型应用提供有力支持。

云硬盘备份

云硬盘备份（CT-VBS，Volume Backup Service）是针对弹性云主机的系统盘、数据盘提供的备份服务。您可对存储重要数据的磁盘进行备份，并在弹性云主机磁盘故障、误删数据、遭到黑客攻击等情况下，将备份的数据快速

恢复到源盘，最大限度保证数据的安全性。

云主机备份

云主机备份（CT-CSBS，Cloud Server Backup Service）支持基于多云硬盘一致性快照技术的备份服务，支持利用备份数据恢复弹性云主机数据，提供数据备份和操作恢复的整体方案。

云监控

云监控服务面向弹性云主机、云硬盘、弹性 IP 等产品提供监控服务，实现性能指标监控、自动告警、历史信息查询等功能。云监控服务使您可以更详细地了解云资源使用情况，并方便您及时调整。

4.5 产品应用场景

应用场景：网站应用

场景说明

个人或企业搭建门户网站、办公 OA 等压力适中的网站，需要保证业务顺畅运行，系统稳定可靠。

场景痛点

- 传统服务器成本较高，费用投入大，且容易造成资源浪费；
- 访问量不定时激增，弹性扩容不足，现有服务器无法负荷，用户体验差。

应用推荐

- 网站初始阶段访问量小，只需要一台低配置的弹性云主机实例即可运行 Apache 或 Nginx 等 Web 应用程序、数据库、存储文件等；
- 随着网站发展，您可以随时升级弹性云主机实例的配置，或者增加云主机实例数量，无需担心低配计算单元在业务突增时带来的资源不足。

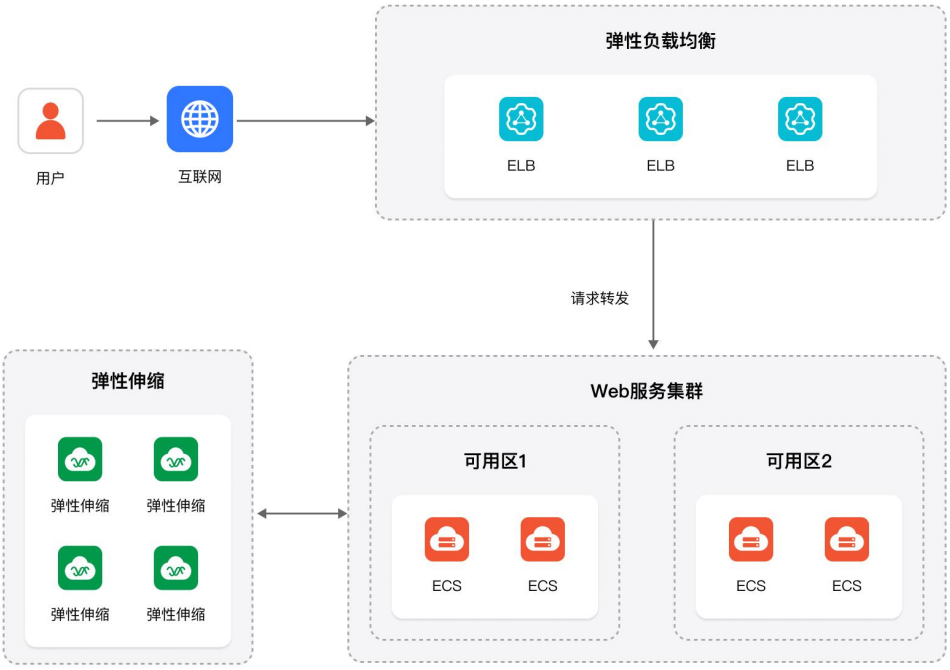
产品优势

- 天翼云弹性云主机即开即用，分钟级交付，快速接入部署在全国范围的数据中心；
- 可按需变更规格、按需扩容，不影响业务正常运行，既兼顾成本，又能保障用户体验；
- 通过负载均衡方式将业务分摊到多台云主机，提升业务处理能力；
- 通过对象存储的高并发支持能力，解决网站频繁访问时的页面崩溃问题。

推荐搭配

弹性伸缩服务、弹性负载均衡、云数据库、对象存储。

场景架构图



应用场景：电子商务

场景说明

在促销、秒杀、爆款、直播等场景下，电商网站访问量可能会在短时间内产生巨大的波动，要求服务器具备高性能、快速弹性、高稳定等特点。

场景痛点

- 直播秒杀等活动对网站性能要求高，系统响应慢会影响用户体验；
- 访问量高时页面加载慢、网络延时大，访问量低时又容易造成资源浪费。

应用推荐

可以配合使用弹性伸缩，自动化实现在请求高峰来临前增加弹性云主机实例，并在进入请求低谷时减少弹性云主机实例。满足访问量达到峰值时对资源的要求，同时降低了成本。

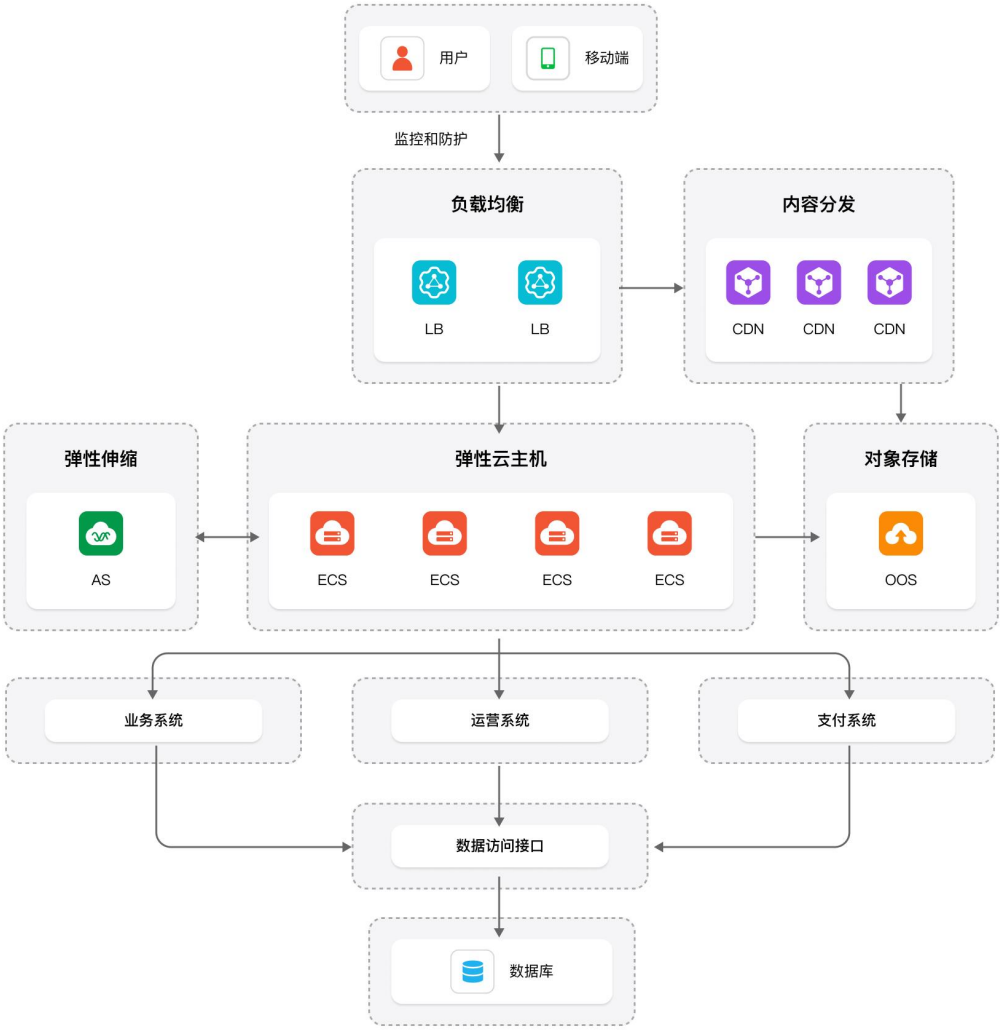
产品优势

- 自动调整弹性计算资源，随时应对业务高峰；
- 利用对象存储技术，将流式数据或文件数据存储到云端，解决高吞吐量、高并发业务压力；
- 对于访问量较大的情况，可通过负载均衡将业务分摊到多台云主机，保障用户体验。

推荐搭配

弹性伸缩服务、弹性负载均衡、对象存储。

场景架构图



应用场景：游戏部署

场景说明

游戏业务对服务器性能、可靠性、网络、弹性伸缩能力要求极高。

场景痛点

- 服务器性能不足会造成卡顿、停服等严重问题，影响玩家游戏体验，引起用户流失；

- 无法准确预估开发周期、玩家数量和增长速度，易造成资源不足或者资源浪费。

应用推荐

- 使用弹性云主机为游戏业务提供高可靠、高性能、高稳定，低时延的算力支撑，给玩家顺畅游戏体验；
- 根据游戏业务对资源的需求快速弹性扩容，合理利用计算资源，节约成本。

产品优势

- 多节点多可用区，有助于游戏业务快速实现大范围布局；
- 根据需求快速弹性扩容，合理利用资源，节约成本；
- 专业的 GPU 虚拟化技术，进程层面资源调度，每台服务器提供更大的并发，降低上云成本；
- 先进的编解码协议，结合优化的网络传输，保证更低的时延和体验效果。

推荐搭配

弹性伸缩服务、弹性负载均衡、云数据库。

场景架构图



应用场景：大数据分析

场景说明

大数据分析场景处理大容量数据，需要高 I/O 能力和快速的数据交换处理能力。例如 MapReduce 、Hadoop 计算密集型。

场景痛点

- 业务数据量不断增大，对计算资源需求较大，造成企业成本压力；
- 数据处理具有周期性，业务波动变化频繁，需要灵活调度资源。

应用推荐

使用本地盘云主机，在保证海量存储空间、高存储性能的前提下，可以为云端的 Hadoop 集群、Spark 集群提供更高的网络性能。

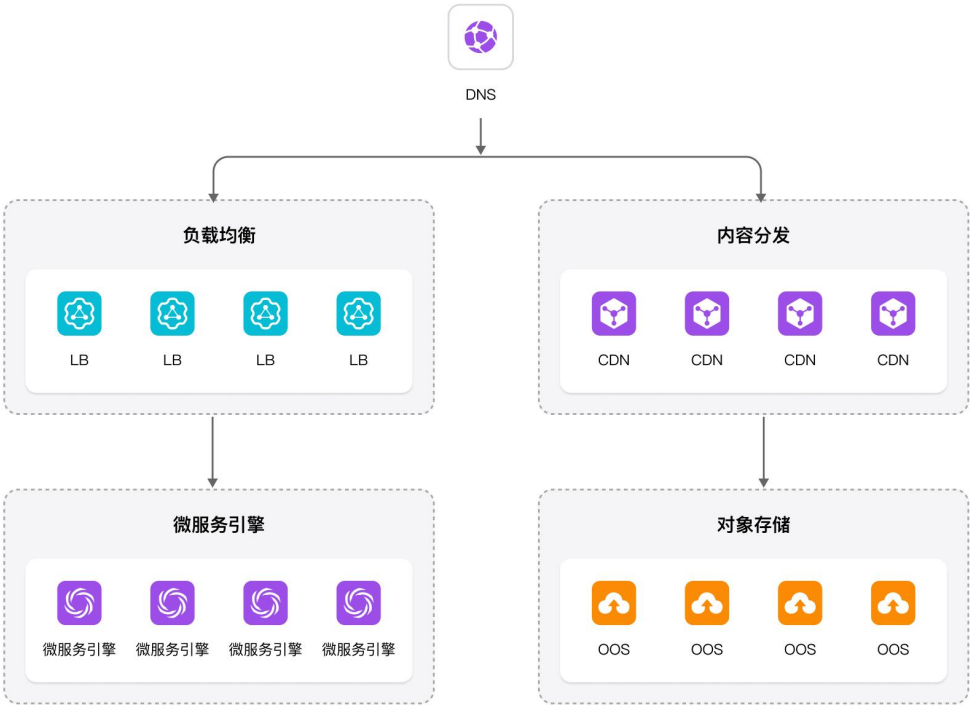
产品优势

- 使用本地盘云主机，提供高可靠存储资源，便于安全存储数据、文件和应用程序；
- 对于重点保护数据，可通过对象存储的多副本冗余功能，实现异地数据容灾；
- 可按需在线弹性扩展，不需要迁移数据，满足大容量、高并发请求的挑战。

推荐搭配

弹性伸缩服务、弹性负载均衡、对象存储。

场景架构图



应用场景：图形渲染

场景说明

专业级 CAD、视频渲染、图形处理等场景需要强大计算能力，追求极致性能体验，以及超高性价比。

场景痛点

- 渲染时间长：单个特效镜头渲染时长超过十小时；
- 渲染效率低：设备有限，大量渲染任务需排队执行；
- 渲染成本高：物理机升级和维护消耗大量运营成本。

应用推荐

异构 GPU 具有优异的 GPU 计算加速能力，可实现图片秒级实时渲染，广泛应用于图形渲染、云端图形工作站、视频转码等场景。

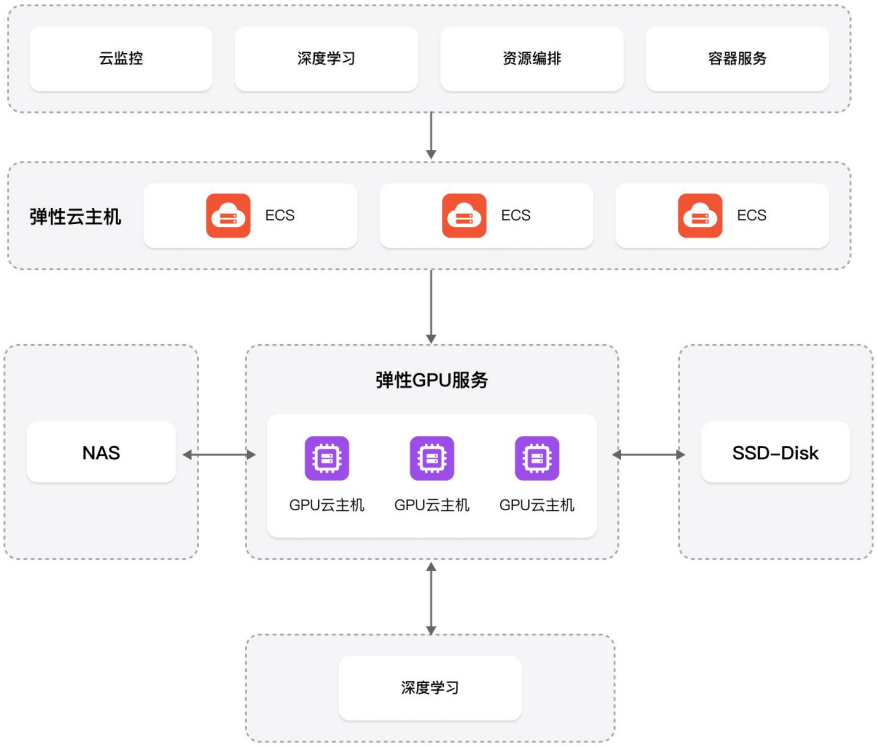
产品优势

- 采用业界先进的 GPU 硬件，同步业界前沿技术，无缝切换新款硬件；
- 功能强大，全面支持多种 GPU 应用程序、深度学习框架，如 OpenGL、DirectX；
- 方便快捷，提供和标准云主机一致的使用方式和管理功能；
- 规格丰富，提供多种显存，满足不同的图形图像场景；
- 通过云监控服务进行 GPU 资源的监控与告警。

推荐搭配

弹性伸缩服务、云监控。

场景架构图



应用场景：深度学习

场景说明

深度学习的训练或预测平台，需要消耗大量计算资源，追求卓越的性能表现。

场景痛点

- 深度学习过程产生大量临时数据，对存储的要求较高；
- 深度学习神经网络计算对网络时延有极高的要求。

应用推荐

对于持续且大量的人工神经网络计算的深度学习场景，天翼云推荐 GPU 实例，不但性能表现卓越，同时大量节省成本。

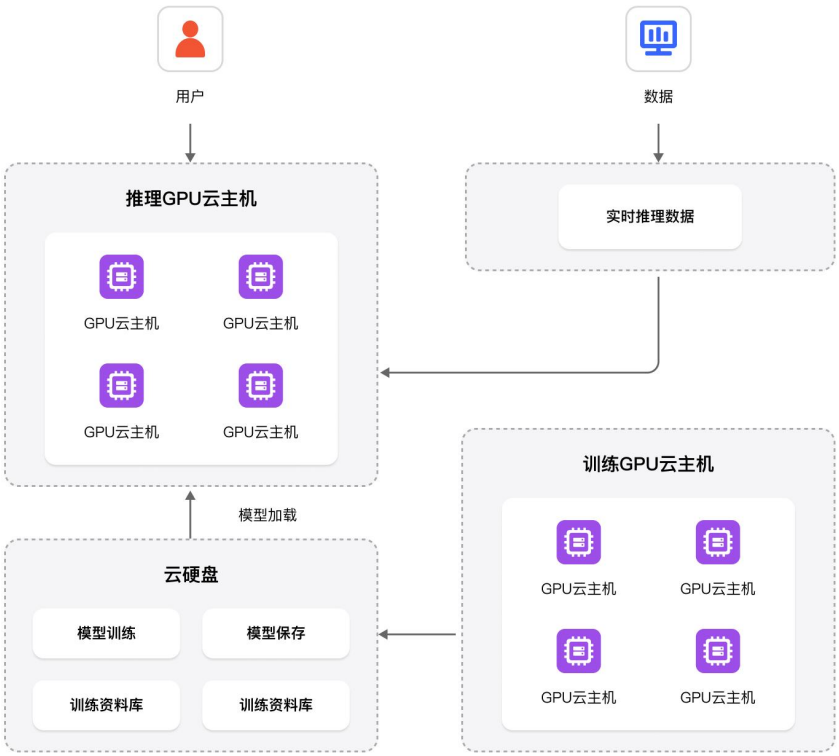
产品优势

- 搭配使用对象存储、云数据库、云监控等服务，可快速搭建功能完备的深度学习离线训练系统；
- 功能强大，全面支持多种 GPU 应用程序、深度学习框架，如 CUDA、OpenCL；
- 方便快捷，提供和标准云主机一致的使用方式和管理功能；
- 规格丰富，提供多种显存，满足不同的图形图像场景；
- 通过云监控服务进行 GPU 资源的监控与告警。

推荐搭配

GPU 云主机、对象存储、云数据库、云监控。

场景架构图



应用场景：自建数据库

场景说明

数据库场景对内存要求较高，需满足海量数据需求。

场景痛点

- 数据量大，对云主机内存要求高，设备采购价格高昂；
- 数据访问量大，要求快速的数据交换和处理，传统方式无法及时扩容和缩容。

应用推荐

- 随取随用，可根据业务需求选购不同规格内存优化型云主机，节约成本；
- 配置超高 IO 的云硬盘和合适的带宽，弹性扩容，快速处理海量数据。

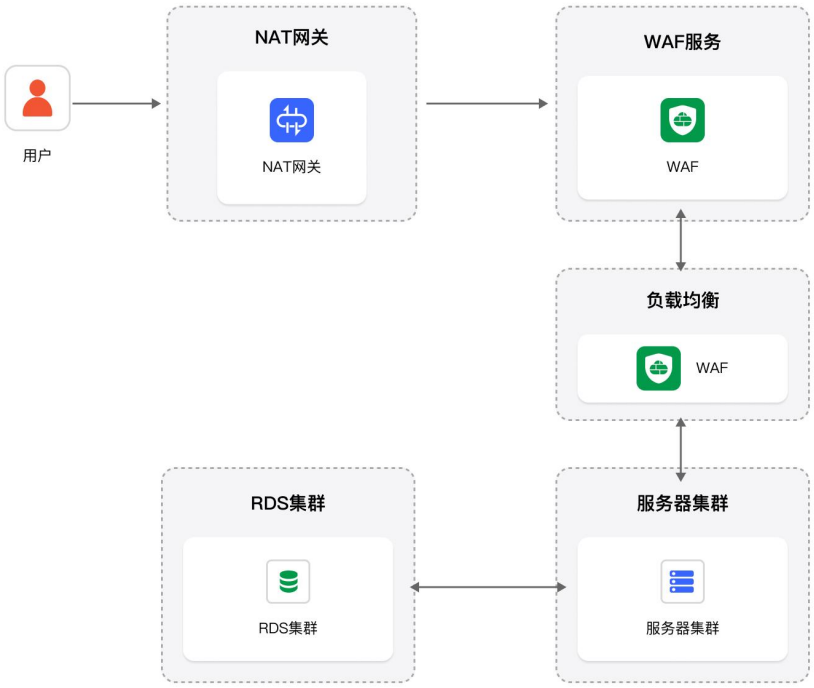
产品优势

- 超高 IO 云硬盘的读写性能强、吞吐带宽高，可满足用户各类数据库（如 MySQL、NoSQL）的部署需求；
- 可实时通知弹性伸缩策略执行结果，使用户掌握伸缩的动态情况，便于及时调整策略以保证数据库业务稳定性；
- 快速实现网络接入互联互通，多副本保障数据安全。

推荐搭配

弹性伸缩服务、云硬盘。

场景架构图



小程序应用

场景说明

对计算与网络有一定要求的应用，需要保证最优用户体验。

场景痛点

- 传统服务器无法灵活扩容，现有服务器无法负荷会导致用户体验差。
- 无法保障可用性、可靠性，需要及时备份、及时恢复等能力。

产品优势

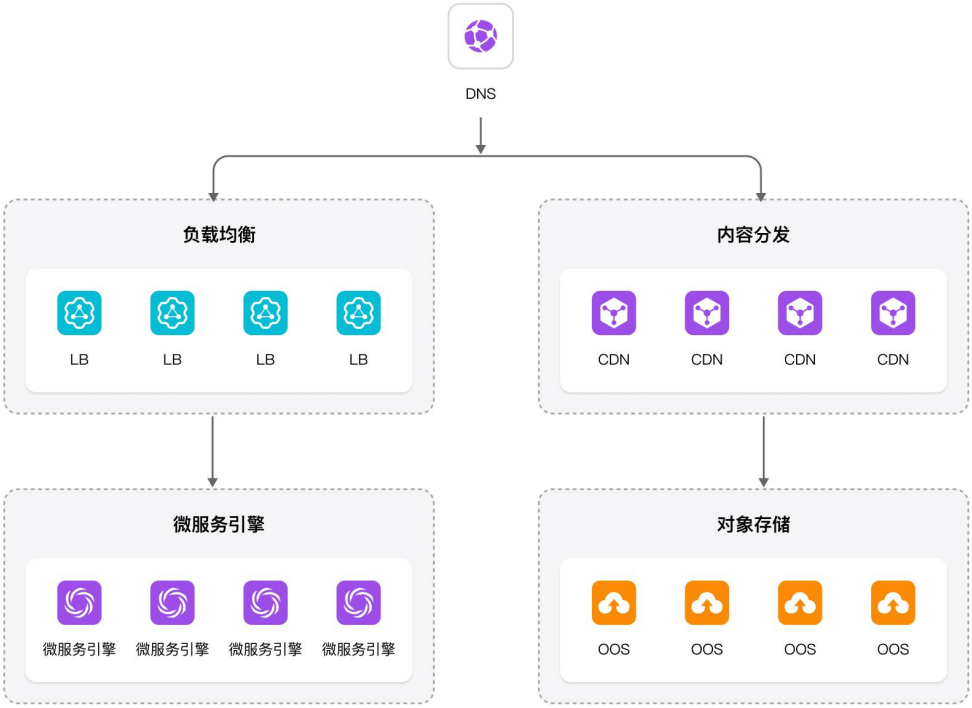
- 云主机支持按需扩容，在保障用户流畅访问小程序的同时，合理优化成本支出。

- 数据自动备份到对象存储，多种备份策略，一键配置。
- 主备实例支持手工/自动切换，秒级切换，快速升级。

推荐搭配

弹性负载均衡、云数据库 ClickHouse、弹性伸缩服务。

场景架构图



4.6 功能特性

支持多节点，多规格随需购买

- 多节点可选，多省份部署资源

- 天翼云节点覆盖国内 31 省，各省份部署多个云资源池。

- 规格丰富、随需购买、满足不同应用场景需求
- 提供了多种规格云主机供用户选择，还可根据业务需求变化进行规格变更。带宽升级不影响业务正常运行，CPU 内存升级仅数分钟，高可用性满足各种应用场景。

支持多种管理和鉴权方式

- 可视化管理平台，操作便捷
- 控制中心对云主机进行管理。可对云主机进行开机、关机、重置密码、重装系统等操作。

- 提供多种登录方式，安全高效
- 通过 VNC 方式、SSH 方式（仅适用于 Linux 云主机）和 MSTSC 方式（仅适用于 Windows 云主机）可登录云主机。其中，VNC 方式适用于未绑定弹性 IP 的云主机登录查看。支持用户远程登录云主机执行操作，并提供 Web 方式的管理控制台对云主机进行单台或批量管理操作，可进行开机、关机、重启、删除等操作。

- 支持多种鉴权方式
- 在创建云主机时，可选择通过密码和密钥方式登录。如选择密码方式登录，可在创建时设置登录密码，或在创建后通过“重置密码”操作设置密码。

支持多种操作系统选择

- 多种镜像，实现业务的快速部署
- 支持常用的 Linux、Windows 镜像，如 CentOS、Ubuntu、Debian 等。支持将云主机导成私有镜像，并可基于私有镜像创建云主机，实现业务的批量、快速部署。支持将私有镜像共享给其它用户，方便多用户统一部署。

- 支持按需重装操作系统
- 弹性云主机支持主流的 Windows Server2012、2016、2019、centos6.x、7.x、8.x 等操作系统，满足用户对不同业务的部署需求。如在安装应用过程中出现问题，还可通过控制台重装当前操作系统。

支持多种数据存储选择

- 多种存储类型随意选择，满足不同 I/O 性能要求
- 提供普通 IO、高 IO、超高 IO、通用 SSD、极速型 SSD 五种类型的云硬盘，满足不同业务对 IO 性能的不同需求。购买方式分为两种，在购买云主机同时购买云硬盘或后期根据需要单独购买云硬盘挂载至云主机上。
- 提供云主机备份、云硬盘备份
- 支持对云主机的系统盘或数据盘进行备份。基于备份数据，用户可创建新的云主机或恢复磁盘数据。

支持网络灵活规划

- 通过虚拟私有云（ VPC）实现网络的灵活规划
- VPC 可提供一个逻辑隔离的网络， 通过 VPC 可实现云中的网络规划，包括创建子网、配置安全组、连接公网，或者使用 VPN 将 VPC 与企业数据中心互联。其中子网可提供 IP 地址管理等。

- 支持多网卡
- 通过为云主机配置多块网卡，将不同的内网 IP 地址与不同网卡进行绑定，实现同一云主机划分至不同的虚拟云子网以及配置不同的安全组策略。

支持多维度监控与告警

- 全面监控及告警机制，保障云主机正常运行

- 云主机提供监控服务，实现性能指标监控、自动告警、历史信息查询等功能。可借助云监控服务，了解包括 CPU、内存、磁盘和网络使用情况等几十项云主机性能监控指标，并支持查看一个月内的云主机监控信息，帮助用户了解云主机实例的历史运行情况，还可根据用户预设规则提供及时的告警通知。

4.7 产品使用须知

弹性云主机使用须知

- 违法行为：禁止使用弹性云主机从事任何违法活动，包括但不限于网络攻击、黑客行为、非法数据获取、侵犯他人隐私、网络钓鱼、传播恶意软件等。
- 网络攻击：禁止使用弹性云主机进行任何形式的网络攻击，如分布式拒绝服务攻击（DDoS）、端口扫描、入侵他人系统等。
- 未授权访问：禁止使用弹性云主机未经授权地访问、入侵或侵犯他人的系统、网络或数据。
- 虚假行为：禁止使用弹性云主机从事虚假活动，如冒充他人身份、伪造数据、欺骗或误导他人。
- 资源滥用：禁止滥用弹性云主机资源，包括但不限于恶意占用计算资源、滥发垃圾邮件、进行大规模的非法广告活动等。
- 侵犯知识产权：禁止使用弹性云主机侵犯他人的知识产权，包括但不限于非法下载、分享、传播受版权保护的内容。
- 非法内容：禁止使用弹性云主机存储、传输或发布任何非法、淫秽、暴力、仇恨言论或其他违反道德准则的内容。
- 破坏性操作：禁止使用弹性云主机进行破坏性操作，如破坏网络、服务器、数据库或其他计算设备的正常运行。
- 违反云服务提供商政策：禁止违反所选择云服务提供商的使用政策、服务条款和合同约定。

Windows 操作系统使用须知

- 不要结束系统进程，结束系统进程可能会导致服务器发生蓝屏、重启。
- 不要重命名、删除或禁用 Windows 下的 administrator 账号，以免影响服务器使用。
- 不建议修改注册表，修改注册表可能会导致系统启动失败。如果一定要修改，请在修改前备份注册表。
- 不建议修改服务器时钟，修改服务器时钟可能会使 DHCP 租约失效进而导致 IP 丢失。
- 修改 IIS 的配置文件，务必先把文件备份，然后用能保留文件格式的软件修改，防止配置文件格式被破坏导致 IIS 无法使用。
- 不要删除 VMTool 程序，会导致云主机运行异常。

Linux 操作系统使用须知

- 不要修改/etc/issue 文件内容，否则，根据实例创建的自定义镜像的系统发行版本无法被正确识别，使用该镜像创建的实例无法正常启动。
- 不要随意更改根目录所在分区下各个目录的权限，尤其是/etc、/sbin、/bin、/boot、/dev、/usr 和/lib 等目录的权限。如果权限更改不当会导致系统出现异常。
- 不要修改系统目录的权限或名称，否则可能导致系统无法正常运行或启动。
- 不要编译 Linux 系统的内核，或对内核进行任何其他操作。不建议修改系统默认 DNS SERVER（/etc/resolv.conf），否则可能会导致软件源、NTP 等无法使用。
- 不建议修改主机默认的内网网络配置信息（IP、子网掩码及网关地址），否则可能会导致网络异常。
- 请谨慎使用 root 等管理账号进行 fio、mkfs、fsck、扩容等操作，避免误操作引起的数据受损。

4.8 产品使用限制

限制概述

使用弹性云主机有以下限制：

- 弹性云主机不支持安装虚拟化软件和二次虚拟化。
- 不支持声卡应用。
- 不支持直接加载外接硬件设备（如硬件加密狗、U 盘、外接硬盘、银行 UKey 等），您可以尝试软件加密狗或者动态口令二次验证等。
- 请勿卸载弹性云主机硬件的驱动程序。
- 请勿修改网卡的 MAC 地址。
- 部分软件或应用的许可证（License）需要与弹性云主机的硬件信息绑定。当云主机进行迁移操作时可能会引起硬件信息的变更，进而导致 License 失效。
- 由于物理机故障导致云主机发生迁移，迁移时可能会出现重启或关机现象，建议您在系统业务进程配置自动拉起和开机自启动，或者通过业务集群部署、主备部署等方式实现业务的高可用。
- 建议您为部署核心业务的云主机做好数据备份。
- 建议您为云主机上的业务指标做好监控配置。

资源限制

云主机使用限制

限制项	默认配额
创建弹性云主机的用户限制	实名认证

限制项	默认配额
创建按量付费资源的限制	账户余额不得小于 100 元
一个账号在一个地域下可创建的弹性云主机限制	50 台（提交工单可提升配额）
一个账号在一个地域下可创建实例的 vCPU 限制	200 核（提交工单可提升配额）
一个账号在一个地域下可创建实例的内存限制	409600 GB（提交工单可提升配额）
一个账号在一个地域下可创建的私有镜像限制	10（提交工单可提升配额）
单台弹性云主机挂载磁盘限制	默认 1 块系统盘、8 块数据盘（提交工单可提升配额）
单台弹性云主机的多网卡限制	默认 1 张主网卡、4 张扩展网卡（提交工单可提升配额）

云硬盘使用限制

云硬盘使用限制请参考：云硬盘使用限制。

弹性 IP 使用限制

弹性 IP 使用限制请参考：弹性 IP 使用限制。

虚拟私有云使用限制

虚拟私有云使用限制请参考：虚拟私有云使用限制。

弹性网卡使用限制

2023 年 11 月 20 日后，部分资源池，单台云主机支持绑定的弹性网卡数量将会根据主机 CPU 配置差异发生变化，具体规则数额如下：

注：如果您是在 2023 年 11 月 20 日前购买的云主机，仍然可以使用已绑定的弹性网卡。但执行解绑或主机升降配等变更操作时，将按照新的规则计算主机可绑定弹性网卡数量。

云主机规格限制网卡数量：

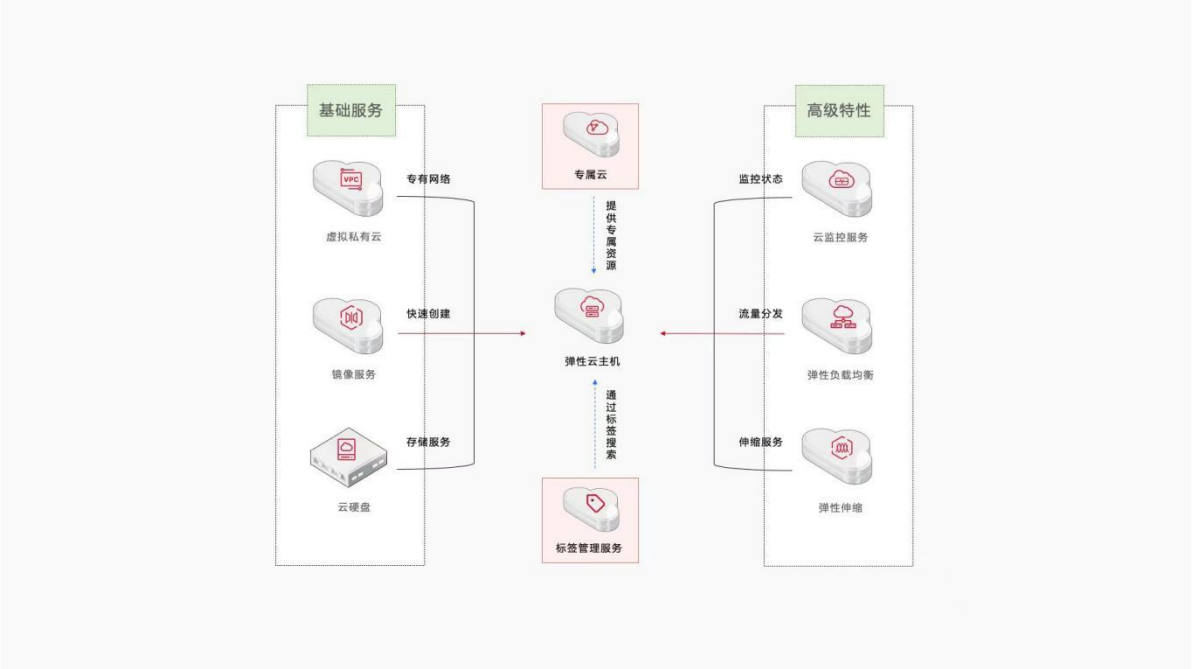
主机 cpu 核数	支持绑定的弹性网卡数量
1 - 2	3
3 - 8	4
9 - 64	8
>64	15

在云主机创建页中，单台云主机可挂载的网卡数为规格限制网卡数量、网卡配额数中的较小值，且不可超过 5 卡，即单台云主机可挂载的网卡数=min（规格限制网卡数量，配额数，5）。

在云主机详情页中，单台云主机可挂载的网卡数为规格限制网卡数量、网卡配额数中的较大值，即单台云主机可挂载的网卡数=max（规格限制网卡数量，配额数）。

4.9 云主机与其他服务之间的关系

弹性云主机与其他云服务的关系如图所示



弹性云主机与其他云服务的关系如表所示

服务名称	云主机与其他服务之间的关系	相关内容
镜像	镜像提供了运行云主机所需的信息，包括操作系统、初始化应用数据等。您可以通过镜像创建弹性云主机，提高弹性云主机的部署效率。	详细内容参见镜像文档

服务名称	云主机与其他服务之间的关系	相关内容
公共镜像	天翼云官方提供的基础镜像，均已获得正版授权，涵盖 Windows Server 系统镜像和主流的 Linux 系统镜像。云主机使用公共镜像可减少安全风险和漏洞，同时用户可通过使用公共镜像获得与实例一致的系统环境或软件，从而避免复杂的服务器配置问题，实现即开即用。	详细内容参见公共镜像文档
私有镜像	用户自行创建或导入的镜像，包含了初始系统环境、应用环境、软件配置等信息，可以节省重复配置的时间。云主机使用私有镜像可使云主机配置上具有高度的灵活性，满足用户个性化的需求。	详细内容参见私有镜像文档
共享镜像	由其他用户共享而来的私有镜像。云主机中使用共享镜像使得用户之间可以共享和重复使用的预先配置的操作系统和软件环境模板，避免了重复的操作和配置工作。当一个用户创建满足特定需求的镜像时，其他用户可以通过共享该镜像快速获取相同配置的环境，提高部署效率和保持一致性。	详细内容参见共享镜像文档
安全产品镜像	安全产品镜像用于加载部分安全产品服务。可在云主机控制台-创建云主机-镜像类型中选择安全产品镜像，满足加载 CSSP、DAS、LAS、OSM 等安全产品的安全性需求。	详细内容参见安全产品镜像文档

服务名称	云主机与其他服务之间的关系	相关内容
云硬盘	数据块级别的块存储产品，采用分布式三副本机制，为云主机提供数据可靠性保证，可以将云硬盘挂载到弹性云主机，并可以扩容云硬盘的容量。	详细内容参见云硬盘文档
快照	某一时间点云主机状态的备份文件，用于备份或者恢复整个云主机。快照能够记录某一块云硬盘在某个时刻的数据，通过回滚将云硬盘数据恢复至快照时间点，用户可以通过快照快速创建出多个具有相同数据的云硬盘用于业务部署。	详细内容参见快照文档
云主机备份	提供对弹性云主机数据的备份保护服务。支持对弹性云主机中的所有云硬盘（系统盘和数据盘）进行备份，并利用备份数据恢复弹性云主机数据。当云主机或磁盘出现故障或者人为错误导致数据误删时，可以自助快速恢复数据。	详细内容参见云主机备份文档
VPC	基于天翼云创建的自定义私有网络，为弹性云主机提供一个逻辑上完全隔离的专有网络，您还可以在 VPC 中定义安全组、IP 地址段、带宽等网络特性。 用户可以通过 VPC 方便地管理、配置内部网络，进行安全、快捷的网络变更。同时，用户可以自定义安全组内与组间弹性云主机的访问规则，加强弹性云主机的安全保护。	详细内容参见 VPC 文档

服务名称	云主机与其他服务之间的关系	相关内容
弹性网卡	一种独立的虚拟网卡，用于连接云主机与私有网络。可以在云主机上添加/删除，实现业务的灵活扩展和迁移。	详细内容参见弹性网卡文档
安全组	为云主机提供网络安全防护机制，用于防止未经授权的访问和保护计算机网络免受恶意攻击。它是一种虚拟防火墙，用于限制入向和出向网络流量通行。用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，即受到这些访问规则的保护。	详细内容参见安全组文档
弹性伸缩	弹性伸缩可以自动调整云主机数量，可按照您定义的伸缩配置和伸缩策略对弹性云主机进行伸缩，帮您节约资源和人力运维成本。	详细内容参见弹性伸缩文档
负载均衡	将访问流量自动分发到多台弹性云主机上，提高搭建在弹性云主机上应用系统对外的服务能力，提高应用程序容错能力。	详细内容参见负载均衡文档
云监控服务	云监控为云主机提供监控服务，提供性能指标监控、自动告警、历史信息查询等功能。当用户开通了弹性云主机后，无需额外安装其他插件，即可在云监控查看对应服务的实例状	详细内容参见云监控

服务名称	云主机与其他服务之间的关系	相关内容
	态。	文档
日志审计服务	日志审计服务通过云主机日志进行全面的标准化处理，及时发现各种安全威胁、异常行为事件。	详细内容参见日志审计文档

4.10 产品规格

4.10.1 规格说明

每个规格的弹性云主机由不同的硬件设备提供，具有不同的性能、价格。根据业务场景和 vCPU、内存、网络性能、存储吞吐等配置划分，天翼云云主机提供了多种实例规格类型，一种类型又包括多个实例规格。

实例规格名称格式为 x<规格类型>x<代系>.xxxx<vCPU 个数>.x<内存数/vCPU 数比>：

- <规格类型>：如 s、m、c、k、h、f、ip、d、a 等，代表云主机规格类型。其中 s 为通用型，m 为内存优化型，c 为计算增强型，k 代表鲲鹏，h 代表海光，f 代表飞腾，ip 代表超高 IO 本地盘云主机、d 代表磁盘增强云主机、a 代表 AMD。
- <代系>：如 2、6 等，代表不同硬件规格，更大数字一般拥有更高的性能。
- <vCPU 个数>：如 small、medium、large、xlarge 等，代表 vCPU 核数。其中，medium 为 1 vCPU，large 为 2 vCPU，xlarge 为 4 vCPU，small 特指 1 vCPU 1G

内存的小规格。

- <内存数/vCPU 数比>：如 1、2、4、8 等，代表内存 GB 数和 vCPU 的比例。

例如，s2.2xlarge.1 表示通用型 2 代云主机，具有 8 核 CPU 和 8GB 内存。

建议您先了解产品规格、性能和使用限制等信息，然后按照实际需要进行选择。当您完成选型并开始使用云主机后，建议您根据一段时间的性能监控信息，验证所选实例规格是否合适。例如：通过监控发现实例 CPU 使用率一直较低，建议您登录实例检查内存占用率是否较高，如果内存占用较高，您可以调整为处理器与内存资源配比更合适的规格族。

规格指标说明如下：

指标名称	说明
网络收发包	网络收发包能力指出方向和入方向相加能达到的最大能力。网络收发包 PPS 性能的测试方法，请参见网络性能测试方法。
网络带宽	网络带宽能力指出方向和入方向相加能达到的最大能力。网络带宽性能的测试方法，请参见网络性能测试方法。
基准带宽	弹性云主机在网络上能够稳定传输数据的最高速度。

最大带宽	弹性云主机某个时间内可以达到的最大内网带宽，取决于该台弹性云主机所在的物理机上其他虚机的网络资源使用情况。 最大带宽是“无”表示不支持突发，最大只能达到基准带宽。
连接数	弹性云主机可以同时容纳客户请求的最高连接数，包含主动和被动建立的连接。
网卡多队列	主网卡支持的最大网卡队列数，涵盖了收和发两个方向。通过开启网卡多队列功能，您可以将弹性云主机中的网卡中断分散给不同的 CPU 处理，以满足网卡的需求，从而提升网络 PPS 和带宽性能。
弹性网卡	弹性云主机最大支持的弹性网卡数量。弹性网卡是虚拟私有网络 VPC 中的虚拟网络接口，用于连接 ECS 与私有网络。包括主网卡和辅助网卡，主网卡无法独立创建和从实例卸载。
云盘基础 IOPS	单台云主机所挂载的 X 系列云盘每秒能处理的 I/O 个数，单位为万次，为整机总的 IOPS 能力，不受云盘数量和单盘能力的影响。 由于当前仅有部分规格支持挂载 X 系列云盘，该参数仅对这类规格生效，详情可参见 X 系列云硬盘。
云盘基	单台云主机所挂载的 X 系列云盘每秒可以成功传输的数据数量，单位为 Gbit/s，为整机总的带宽能力，不受云盘数量和单盘能力的影响。 由于当前仅有部分规格支持挂载 X 系列云盘，该参数仅对这类规格生

础带宽	效，详情可参见 X 系列云硬盘。
本地盘	本地盘数量*单块盘容量。本地盘是云主机所在物理机上的本地硬盘设备。本地盘能够为云主机提供本地存储访问能力，具有低时延、高随机 IOPS、高吞吐量和高性价比的优势。
GPU	GPU 数量*GPU 类型。GPU 即图形处理器，功能包括图形渲染、图像处理、计算加速等。GPU 相对于 CPU 而言，具有更多的处理单元和更高的并行处理能力，适合用于大规模并行计算等场景。

4.10.2 规格族

根据系统架构以及使用场景，弹性云主机规格族可以分为以下几类：

CPU 架构	类型	子类型	描述
X86 计算	通用型	通用型 s2 云主机	通用型云主机共享宿主机的 CPU 资源，主要提供基本水平的 vCPU 性能、平衡的计算、内存和网络资源，具有较高性价比，支持通用的业务运行。适用于不会经常或始终用尽 vCPU 性能的场景，如小型网站、轻量级研发测试环境、小型数据库等。
		通用型 m2 云主机	
		通用	

CPU架构	类型	子类型	描述
		型s3云主机	
		通用型s6云主机	
		通用型s7云主机	
		通用型s8云主机	
		通用型s8r云主机	

CPU架构	类型	子类型	描述
		通用型s8e云主机	
	计算型	计算型c3云主机	计算型云主机独享宿主机的CPU资源，实例间无CPU争抢，并且没有进行资源超配，同时搭载全新网络加速引擎，实现接近物理服务器的强劲稳定性能。提供更大规格的CPU和内存组合，适用于计算密集型业务等场景，如大型网站、电商营销等。
		计算型c6云主机	
		计算型c7云主机	
		计算型c8云主	

CPU架构	类型	子类型	描述
		机	
		计算型c8e云主机	
		计算型c8a云主机	
	内存型	内存型m3云主机	内存型云主机独享宿主机的CPU资源，实例间无CPU争抢，并且没有进行资源超配，同时搭载全新网络加速引擎，实现接近物理服务器的强劲稳定性能。CPU和内存配比可达1:8，适用于高内存计算应用，如大数据分析、核心数据库等。
		内存型m6云主机	
		内存型m7	

CPU架构	类型	子类型	描述
		云主机	
		内存型m8云主机	
		内存型m8e云主机	
		内存型m8a云主机	
	网络型	网络增强型c7ne云主机	配套25GE智能网卡，通过软硬结合的方式提高虚拟化性能，大幅提升实例的网络带宽能力和网络收发包能力。适用各种于网络密集型应用场景，如NFV/SD-WAN、移动互联网、视频弹幕、电信业务转发等；中小型数据库系统、缓存、搜索集群；各种类型和规模的企业级应用；大数据分析和机器学习。
		网络	

CPU架构	类型	子类型	描述
		增强型c8ne云主机	
		网络增强型m8ne云主机	
	本地盘云主机	超高IO型云主机ip3	搭载NVMe SSD本地盘，为CPU独享型的x86架构云主机，提供透传整块倍数的本地盘规格。
		超高IO型云主机ir3	搭载NVMe SSD本地盘，为CPU独享型的x86架构云主机，提供切分本地盘大小规格。
		磁盘增强型云	搭载SATA HDD本地盘，为CPU独享型的x86架构云主机。

CPU架构	类型	子类型	描述
		主机d3	
	GPU云主机	图形加速基础型G5	基于NVIDIA虚拟化GPU技术，能够提供全面的专业级的图形加速能力，同时适用于小规模推理。
		图形加速基础型G5s	
		图形加速基础型G6	
		图形加速基础型G7	
		计算加速	采用GPU直通技术，满足深度学习、科学计算、图像渲

CPU架构	类型	子类型	描述
		型P2V	染等场景下用户的性能需求。
		计算加速型P2Vs	
		计算加速型PI2	
		计算加速型PI7	
		计算加速型P8A	
		计算加速型PN8I	
		计算	

CPU架构	类型	子类型	描述
		加速型PN8S	
		计算加速型PAK1	
		计算加速型PCH1	
	海光系列	hc1型云主机	搭载海光 Hygon C86 7285 或 Hygon C86 7380 处理器
		hm1型云主机	
		hs1型云主机	
		hc3型云主机	搭载海光 Hygon C86 7493 处理器

CPU架构	类型	子类型	描述
		hm3型云主机	
		hs3型云主机	
	经济型	经济型e云主机	经济型云主机共享宿主机的CPU资源，提供基本水平的vCPU性能、平衡的计算、内存和网络资源，具有较低的价格，有可用性保障，无性能保障。支持小型的业务运行。适用于对价格敏感，性能要求低的场景，如小型网站、开发测试环境、个人云上学习环境等。
ARM计算	鲲鹏系列	kc1型云主机	搭载华为鲲鹏920处理器
		km1型云主机	
		ks1型云主机	
	飞	fc1型云主机	搭载腾云S2500处理器

CPU架构	类型	子类型	描述
	腾系列	fm1型云主机	
		fs1型云主机	

4.10.3 通用云主机

4.10.3.1 概述

通用云主机的不同代系，对应着不同的处理器型号，一般更高代系拥有更优的性能。

2 系列

提供基本水平的 vCPU 性能、平衡的计算、内存和网络资源，在主机负载较轻时，可以提供较高的计算能力，在主机负载较重时无法保证实例计算性能的稳定。

3 系列

采用第一代英特尔® 至强® 可扩展处理器 （Sky Lake）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术，配套 10GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制，提供强劲稳定的计算性能、更高网络带宽和 PPS 收发包能力。适用于对计算与网络有一定要求的场景，如小型网站、轻量级研发测试环境、中小型数据库等。

6 系列

采用第二代英特尔® 至强® 可扩展处理器 （Cascade Lake）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术， 配套 25GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制，提供强劲稳定的计算性能、更高网络带宽和 PPS 收发包能力。适用于对计算与网络有一定要求的场景，如通用数据库及缓存服务器、中重载企业应用等。

7 系列

采用第三代英特尔® 至强® 可扩展处理器 （Ice Lake）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术，配套 25GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制，提供更强劲稳定的计算性能、更高网络带宽和 PPS 收发包能力。适用于对计算与网络有更高性能要求的 Web 应用、电商平台、短视频平台、在线游戏、保险金融等各类中重载企业应用。

8 系列

采用第三代英特尔® 至强® 可扩展处理器 （Ice Lake）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术，配套 100GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制，可提供 128C1024G 的大规格实例、100Gbps 带宽、3000 万收发包能力 。适用于计算类应用场景、web 服务应用场景、媒体处理应用场景、汽车工业设计仿真、数字建模模拟、海量数据处理等场景，为高性能计算、互联网大数据、政企、金融等行业客户提供更强劲的计算性能，帮助客户降低 TCO，提高企业产品竞争力。

7ne 系列

采用第三代英特尔® 至强® 可扩展处理器 （Ice Lake）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术，配套 25GE 智能网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制，通过软硬结合的方式提高虚拟化性能，大幅提升实例的网络带宽能力和网络收发包能力。可提供 148C592G 的大规格实例、3000 万收发包能力 。适用各种于网络密集型应用场景，如

NFV/SD-WAN、移动互联网、视频弹幕、电信业务转发等；中小型数据库系统、缓存、搜索集群；各种类型和规模的企业级应用；大数据分析和机器学习。

8r 系列

采用第四代英特尔® 至强® 可扩展处理器 （Sapphire Rapids）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术， 配套 25GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制。支持高级矩阵扩展(Intel® AMX，Advanced Matrix Extension)，可提高 CPU 的深度学习训练和推理性能，16CPU 及以上规格支持 Intel 数据保护与压缩加速技术 (Intel® QAT，Quick Assist Technology ） ，可提高特定应用程序和工作负载的加密、解密和压缩性能。

8e 系列

采用第五代英特尔® 至强® 可扩展处理器 （Emerald Rapids）， 基于新一代虚拟化平台，支持 IPI 虚拟化（IPI Virtualization），使用 NUMA(Non Uniform Memory Access Architecture)绑定技术， 配套 25GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制。支持 Intel® 高级矩阵扩展（Intel®AMX）、TME（Total Memory Encryption）运行内存加密等高级特性，同时可提供 192C1536G 大规格实例，计算性能、安全性、稳定性更强大，在人工智能、媒体应用、网站应用的性能均有所提升。

8ne 系列

采用第五代英特尔® 至强® 可扩展处理器 （Emerald Rapids）， 基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术，搭载天翼云自研紫金智能网卡，通过软硬结合的方式提高虚拟化性能，大幅提升实例的网络带宽能力和网络收发包能力。可提供 192C1536G 的大规格实例、3600 万收发包能力 。适用各种于网络密集型应用场景，如 NFV/SD-WAN、移动互联网、视频弹幕、电信业务转发等；中小型数据库系统、缓存、搜索集群；各种类型和规模的企业级应用；大数据分析和机器学习。

8a 系列

搭载 AMD EPYC™ Genoa 处理器（2.6GHz），基于新一代虚拟化平台，使用 NUMA(Non Uniform Memory Access Architecture)绑定技术， 配套 100GE 网卡，搭载全新网络加速引擎以及 DPDK(Data Plane Development Kit)快速报文处理机制，可提供稳定的算力输出和更强劲的 I/O 引擎。适用于对价格敏感、性能要求高的场景，如各种类型和规模的企业级应用、大数据分析与处理、高网络收发包场景（如游戏、直播等）、AI 训练与推理，音视频转码类应用等。

4.10.3.2 通用型

通用型云主机主要提供基本水平的 vCPU 性能、平衡的计算、内存和网络资源。采用非绑定 CPU 共享调度模式，vCPU 会根据系统负载被随机分配到空闲的 CPU 超线程上。在主机负载较轻时，可以提供较高的计算能力，但是在主机负载较重时，可能由于不同实例 vCPU 争抢物理 CPU 资源而导致计算性能波动不稳定。有可用性 SLA 保证，但无性能 SLA 保证。

通用型与计算型相比，更侧重于资源性能的共享，无法保证实例计算性能的稳定，但是性价比更高，适用于对成本比较敏感、对性能抖动容忍度较高的场景。

- 在售规格：s8e、s8r、s8、s7、s6、s3、s2、m2

适用场景

- 小型网站
- 轻量级研发测试环境
- 小型数据库等

通用型弹性云主机特点

规格名称	计算	磁盘类型	网络
s8e	1.CPU/内存配比： 1:2/1:4	1.普通 IO 2.高 IO	1.支持 IPv6 2.实例网络性能与计算规格

规格名称	计算	磁盘类型	网络
	2. vCPU 数量范围：2-192 3. 处理器：第五代英特尔®至强®可扩展处理器 4. 基频/睿频：2.6GHz/3.1GHz	3. 通用型 SSD 4. 超高 IO 5. 极速型 SSD 6. XSSD-0、XSSD-1、XSSD-2	对应，规格越高网络性能越强 3. 最大网络收发包：700 万 PPS 4. 最大内网带宽：25Gbps
s8r	1. CPU/内存配比：1:2/1:4 2. vCPU 数量范围：2-128 3. 处理器：第四代英特尔®至强®可扩展处理器 4. 基频/睿频：2.8GHz/3.3GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：700 万 PPS 4. 最大内网带宽：25Gbps
s8	1. CPU/内存配比：1:2/1:4 2. vCPU 数量范围：2-96 3. 处理器：第三代英特尔®至强®可扩展处理器	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO 5. 极速型 SSD	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：700 万 PPS 4. 最大内网带宽：25Gbps

规格名称	计算	磁盘类型	网络
	4. 基频/睿频： 2. 8GHz/3. 5GHz	6. XSSD-0、 XSSD-1、 XSSD-2	
s7	1. CPU/内存配比： 1:1/1:2/1:4 2. vCPU 数量范围：1-32 3. 处理器：第三代英特尔®至强®可扩展处理器 4. 基频/睿频： 2. 8GHz/3. 5GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO 5. 极速型 SSD	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：200 万 PPS 4. 最大内网带宽：12Gbps
s6	1. CPU/内存配比： 1:1/1:2/1:4 2. vCPU 数量范围：1-16 3. 处理器：第二代英特尔®至强®可扩展处理器 4. 基频/睿频： 3. 0GHz/4. 0GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO 5. 极速型 SSD	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：100 万 PPS 4. 最大内网带宽：6Gbps
s3	1. CPU/内存配比： 1:1/1:2/1:4 2. vCPU 数量范围：1-16 3. 处理器：英特尔®至	1. 普通 IO 2. 高 IO 3. 通用型 SSD	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强

规格名称	计算	磁盘类型	网络
	强®可扩展处理器 4.基频/睿频： 2.1-2.3GHz/ 3.2-4.0GHz	4.超高 IO	3.最大网络收发包:18 万 PPS 4.最大内网带宽：2.5Gbps
s2	1.CPU/内存配比： 1:1/1:2/1:4 2.vCPU 数量范围:1-32 3.处理器：英特尔®至强®处理器 E5 家族 4.基频/睿频： 2.1-2.6GHz/ 3.1-4.0GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：无 4.最大内网带宽：5Gbps
m2	1.CPU/内存配比：1:8 2.vCPU 数量范围:1-32 3.处理器：英特尔®至强®处理器 E5 家族 4.基频/睿频： 2.1-2.6GHz/ 3.1-4.0GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：无 4.最大内网带宽：5Gbps

s8e 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
s8e.large.2	2	4	1	5	45	25	2	3	1	1
s8e.large.4	2	8	1	5	45	25	2	3	1	1
s8e.xlarge.2	4	8	1.5	5	50	25	4	4	2	1.5
s8e.xlarge.4	4	16	1.5	5	50	25	4	4	2	1.5
s8e.2xlarge.2	8	16	2.5	5	80	25	8	4	2.5	2
s8e.2xlarge.4	8	32	2.5	5	80	25	8	4	2.5	2
s8e.3xl	12	24	4	5	12	25	8	8	3	2.5

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
arge.2					0					
s8e.3xlarge.4	12	48	4	5	120	25	8	8	3	2.5
s8e.4xlarge.2	16	32	5	12	150	50	8	8	4	3
s8e.4xlarge.4	16	64	5	12	150	50	8	8	4	3
s8e.6xlarge.2	24	48	6	12	200	60	12	8	5	4
s8e.6xlarge.4	24	96	6	12	200	60	12	8	5	4
s8e.8xlarge.2	32	64	8	12	300	100	16	8	6	5
s8e.8xlarge.4	32	128	8	12	300	100	16	8	6	5

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
s8e.16xlarge.2	64	128	16	20	600	200	32	8	8	6
s8e.16xlarge.4	64	256	16	20	600	200	32	8	8	6
s8e.24xlarge.2	96	192	20	25	700	240	32	8	10	7
s8e.24xlarge.4	96	384	20	25	700	240	32	8	10	7
s8e.32xlarge.2	128	256	25	无	700	240	32	8	16	10
s8e.32xlarge.4	128	512	25	无	700	240	32	8	16	10
s8e.48xlarge.2	192	384	25	无	700	240	32	8	20	16
s8e.48x	192	76	25	无	70	24	3	8	2	16

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
large.4		8			0	0	2		0	

s8r 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	QAT数量
s8r.large.2	2	4	1	5	45	25	2	3	0
s8r.large.4	2	8	1	5	45	25	2	3	0
s8r.xlarge.2	4	8	1.5	5	50	25	4	4	0

s8r.xlarge.4	4	16	1.5	5	50	25	4	4	0
s8r.2xlarge.2	8	16	2.5	5	80	25	8	4	0
s8r.2xlarge.4	8	32	2.5	5	80	25	8	4	0
s8r.3xlarge.2	12	24	4	5	120	25	8	8	0
s8r.3xlarge.4	12	48	4	5	120	25	8	8	0
s8r.4xlarge.2	16	32	5	12	150	50	8	8	1
s8r.4xlarge.4	16	64	5	12	150	50	8	8	1
s8r.6xlarge.2	24	48	6	12	200	60	12	8	1

s8r.6xlarge.4	24	96	6	12	200	60	12	8	1
s8r.8xlarge.2	32	64	8	12	300	100	16	8	2
s8r.8xlarge.4	32	128	8	12	300	100	16	8	2
s8r.16xlarge.2	64	128	16	20	600	200	32	8	2
s8r.16xlarge.4	64	256	16	20	600	200	32	8	2
s8r.24xlarge.2	96	192	20	25	700	240	32	8	2
s8r.24xlarge.4	96	384	20	25	700	240	32	8	2
s8r.32xlarge.2	128	256	25	无	700	240	32	8	2

s8r.32xlarge.4	128	512	25	无	700	240	32	8	2

s8 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
s8.large.2	2	4	1	5	45	25	2	3	2	1.5
s8.large.4	2	8	1	5	45	25	2	3	2	1.5
s8.xlarge.2	4	8	1.5	5	50	25	4	4	4	2
s8.xlarge.4	4	16	1.5	5	50	25	4	4	4	2
s8.2xlarge.2	8	16	2.5	5	80	25	8	4	5	3

s8.2xlarge.4	8	32	2.5	5	80	25	8	4	5	3
s8.3xlarge.2	12	24	4	5	120	25	8	8	7	4
s8.3xlarge.4	12	48	4	5	120	25	8	8	7	4
s8.4xlarge.2	16	32	5	12	150	50	8	8	8	5
s8.4xlarge.4	16	64	5	12	150	50	8	8	8	5
s8.6xlarge.2	24	48	6	12	200	60	12	8	11	6
s8.6xlarge.4	24	96	6	12	200	60	12	8	11	6
s8.8xlarge.2	32	64	8	12	300	100	16	8	15	8
s8.8xlarge.4	32	128	8	12	300	100	16	8	15	8
s8.16xlarge.2	64	128	16	20	600	200	32	8	30	16
s8.16x	64	256	16	20	600	20	3	8	30	16

large.4						0	2			
s8.24xlarge.2	96	192	20	25	700	240	32	8	40	20
s8.24xlarge.4	96	384	20	25	700	240	32	8	40	20

s7 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
s7.small.1	1	1	0.8	0.1	10	1
s7.medium.2	1	2	0.8	0.1	10	1
s7.medium.4	1	4	0.8	0.1	10	1
s7.large.2	2	4	1.5	0.2	15	1
s7.la	2	8	1.5	0.2	15	1

rge.4						
s7.xlarge.2	4	8	2	0.35	25	1
s7.xlarge.4	4	16	2	0.35	25	1
s7.2xlarge.2	8	16	3	0.75	50	2
s7.2xlarge.4	8	32	3	0.75	50	2
s7.4xlarge.2	16	32	6	1.5	100	4
s7.4xlarge.4	16	64	6	1.5	100	4
s7.8xlarge.2	32	64	12	3	200	8
s7.8x	32	128	12	3	200	8

large.4						

s6 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
s6.small.1	1	1	0.8	0.1	10	1
s6.medium.2	1	2	0.8	0.1	10	1
s6.medium.4	1	4	0.8	0.1	10	1
s6.large.2	2	4	1.5	0.2	15	1
s6.large.4	2	8	1.5	0.2	15	1
s6.xlarge.2	4	8	2	0.35	25	1
s6.xlarge.	4	16	2	0.35	25	1

4						
s6.2xlarge.2	8	16	3	0.75	50	2
s6.2xlarge.4	8	32	3	0.75	50	2
s6.4xlarge.2	16	32	6	2	100	4
s6.4xlarge.4	16	64	6	2	100	4

s3 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
s3.small.1	1	1	0.3	0.06	3	1
s3.medium.2	1	2	0.3	0.06	3	1
s3.me	1	4	0.3	0.06	3	1

dium.4						
s3.large.2	2	4	0.5	0.12	6	1
s3.large.4	2	8	0.5	0.12	6	1
s3.xlarge.2	4	8	1	0.25	9	1
s3.xlarge.4	4	16	1	0.25	9	1
s3.2xlarge.2	8	16	1.5	0.5	12	2
s3.2xlarge.4	8	32	1.5	0.5	12	2
s3.4xlarge.2	16	32	2.5	1	18	4
s3.4xlarge	16	64	2.5	1	18	4

.4						

s2 弹性云主机的规格

规格名称	CPU(核)	内存(GB)	最大带宽(Gbps)	基准带宽(Gbps)	最大收发包（万 pps）	网卡多队列
s2.small.1	1	1	0.3	-	-	1
s2.medium.2	1	2	0.3	-	-	1
s2.medium.4	1	4	0.3	-	-	1
s2.large.1	2	2	0.5	-	-	1
s2.large.2	2	4	0.5	-	-	1
s2.large.4	2	8	0.5	-	-	1
s2.xlarge.1	4	4	1	-	-	1
s2.xlarge.2	4	8	1	-	-	1

s2.xlarge.4	4	16	1	-	-	1
s2.2xlarge.1	8	8	1.5	-	-	2
s2.2xlarge.2	8	16	1.5	-	-	2
s2.2xlarge.4	8	32	1.5	-	-	2
s2.4xlarge.1	16	16	2.5	-	-	4
s2.4xlarge.2	16	32	2.5		-	4
s2.4xlarge.4	16	64	2.5	-	-	4
s2.8xlarge.1	32	32	5	-	-	6
s2.8xlarge.2	32	64	5	-	-	6
s2.8xlarge.4	32	128	5	-	-	6

m2 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
m2.medium.8	1	8	-	-	-	-
m2.large.8	2	16	-	-	-	-
m2.xlarge.8	4	32	-	-	-	-
m2.2xlarge.8	8	64	-	-	-	-
m2.4xlarge.8	16	128	-	-	-	-

4.10.3.3 计算型

计算型云主机独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，提供更接近物理服务器的性能。适用于对实例性能有一定要求的计算密集型业务场景。

- 在售规格：c8a、c8e、c8、c7、c6、c3

适用场景

计算型云主机适用于计算密集型业务等场景：

- 大型网站
- 电商营销

计算型弹性云主机特点

规格名称	计算	磁盘类型	网络
c8a	1.CPU/内存配比： 1:2/1:4 2.vCPU 数量范围：2-192 3.处理器：AMD EPYC™ Genoa 处理器 4.基频/睿频： 2.6GHz/3.7GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD 6.XSSD-0、XSSD-1、XSSD-2	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：1800 万 PPS 4.最大内网带宽：100Gbps
c8e	1.CPU/内存配比： 1:2/1:4 2.vCPU 数量范围：2-192 3.处理器：第五代英特尔®至强®可扩展处理器 4.基频/睿频： 2.6GHz/3.1GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD 6.XSSD-0、XSSD-1、XSSD-2	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：1400 万 PPS 4.最大内网带宽：40Gbps
c8	1. CPU/内存配比： 1:2/1:4 2.vCPU 数量范围：2-128 3.处理器：第三代英特尔®至强®可扩展处理器 4.基频/睿频： 2.8GHz/3.5GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD 6.XSSD-0、XSSD-1、XSSD-2	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：3000 万 PPS 4.最大内网带宽：100Gbps
c7	1.CPU/内存配比：	1.普通 IO	1.支持 IPv6

规格名称	计算	磁盘类型	网络
	1:2/1:4 2. vCPU 数量范围：2-96 3. 处理器：第三代英特尔®至强®可扩展处理器 4. 基频/睿频： 2. 8GHz/3. 5GHz	2. 高 IO 3. 通用型 SSD 4. 超高 IO 5. 极速型 SSD	2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：1100 万 PPS 4. 最大内网带宽：40Gbps
c6	1. CPU/内存配比： 1:2/1:4 2. vCPU 数量范围：2-64 3. 处理器：第二代英特尔®至强®可扩展处理器 4. 基频/睿频： 3. 0GHz/4. 0GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO 5. 极速型 SSD	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：1000 万 PPS 4. 最大内网带宽：40Gbps
c3	1. CPU/内存配比： 1:2/1:4 2. vCPU 数量范围：2-32 3. 处理器：英特尔®至强®可扩展处理器 4. 基频/睿频： 2. 1-2. 3GHz/ 3. 2-4. 0GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO	1. 支持 IPv6 2. 实例网络性能与计算规格对应，规格越高网络性能越强 3. 最大网络收发包：260 万 PPS 4. 最大内网带宽：15Gbps

c8a 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8a.large.2	2	4	1.5	10	90	25	2	3	2	1.5
c8a.large.4	2	8	1.5	10	90	25	2	3	2	1.5
c8a.xlarge.2	4	8	2.5	10	100	25	4	4	3	2
c8a.xlarge.4	4	16	2.5	10	100	25	4	4	3	2
c8a.2xlarge.2	8	16	4	15	160	25	8	4	4.5	2.5
c8a.2xlarge.4	8	32	4	15	160	25	8	4	4.5	2.5
c8a.4xlarge.2	16	32	7	15	200	50	16	8	6	3.5

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8a.4xlarge.4	16	64	7	15	200	50	16	8	6	3.5
c8a.8xlarge.2	32	64	10	25	300	100	32	8	8	5
c8a.8xlarge.4	32	128	10	25	300	100	32	8	8	5
c8a.12xlarge.2	48	96	16	25	450	150	48	8	12	8
c8a.12xlarge.4	48	192	16	25	450	150	48	8	12	8
c8a.16xlarge.2	64	128	20	25	600	200	64	8	16	10
c8a.16xlarge.4	64	256	20	25	600	200	64	8	16	10

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8a.24xlarge.2	96	192	32	无	900	300	64	15	24	16
c8a.24xlarge.4	96	384	32	无	900	300	64	15	24	16
c8a.32xlarge.2	128	256	40	无	1200	400	64	15	32	20
c8a.32xlarge.4	128	512	40	无	1200	400	64	15	32	20
c8a.48xlarge.2	192	384	64	无	1800	500	64	15	50	32
c8a.48xlarge.4	192	768	64	无	1800	500	64	15	50	32

c8e 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8e.large.2	2	4	0.8	4	90	25	2	3	2	1.5
c8e.large.4	2	8	0.8	4	90	25	2	3	2	1.5
c8e.xlarge.2	4	8	1.6	8	100	50	4	4	4	2
c8e.xlarge.4	4	16	1.6	8	100	50	4	4	4	2
c8e.2xlarge.2	8	16	3.5	15	160	100	8	4	5	3
c8e.2xlarge.4	8	32	3.5	15	160	100	8	4	5	3
c8e.3xlarge.2	12	24	5	17	240	150	8	8	6	4

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8e.3xlarge.4	12	48	5	17	240	150	8	8	6	4
c8e.4xlarge.2	16	32	7	25	300	200	8	8	7	5
c8e.4xlarge.4	16	64	7	25	300	200	8	8	7	5
c8e.6xlarge.2	24	48	10	25	450	300	12	8	8	6
c8e.6xlarge.4	24	96	10	25	450	300	12	8	8	6
c8e.8xlarge.2	32	64	15	32	600	400	16	8	10	7
c8e.8xlarge.4	32	128	15	32	600	400	16	8	10	7

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8e.12xlarge.2	48	96	20	无	900	600	24	8	12	8
c8e.12xlarge.4	48	192	20	无	900	600	24	8	12	8
c8e.16xlarge.2	64	128	25	无	1200	800	32	8	16	9
c8e.16xlarge.4	64	256	25	无	1200	800	32	8	16	9
c8e.24xlarge.2	96	192	36	无	1200	800	32	8	20	10
c8e.24xlarge.4	96	384	36	无	1200	800	32	8	20	10
c8e.32xlarge.2	128	256	40	无	1400	1000	32	15	25	16

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
c8e.32xlarge.4	128	512	40	无	1400	1000	32	15	25	16
c8e.48xlarge.2	192	384	40	无	1400	1000	32	15	30	20
c8e.48xlarge.4	192	768	40	无	1400	1000	32	15	30	20

c8 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（	云盘基础带宽（Gbit/s）

									万次）	
c8.large.2	2	4	2.5	15	100	25	2	3	2.5	2
c8.large.4	2	8	2.5	15	100	25	2	3	2.5	2
c8.xlarge.2	4	8	4	15	120	50	4	4	5	2.5
c8.xlarge.4	4	16	4	15	120	50	4	4	5	2.5
c8.2xlarge.2	8	16	6	15	160	100	8	4	6	4
c8.2xlarge.4	8	32	6	15	160	100	8	4	6	4
c8.3xlarge.2	12	24	10	15	240	150	12	8	8	5
c8.3xlarge.4	12	48	10	15	240	150	12	8	8	5
c8.4xlarge.2	16	32	12	25	300	200	16	8	10	6
c8.4xl	16	64	12	25	300	20	1	8	10	6

arge.4						0	6			
c8.6xlarge.2	24	48	15	25	450	300	24	8	12	7.5
c8.6xlarge.4	24	96	15	25	450	300	24	8	12	7.5
c8.8xlarge.2	32	64	20	25	600	400	32	8	20	10
c8.8xlarge.4	32	128	20	25	600	400	32	8	20	10
c8.12xlarge.2	48	96	25	无	900	600	48	8	30	12
c8.12xlarge.4	48	192	25	无	900	600	48	8	30	12
c8.16xlarge.2	64	128	32	无	1200	800	64	8	36	20
c8.16xlarge.4	64	256	32	无	1200	800	64	8	36	20
c8.24x	96	192	50	无	180	10	6	1	100	48

large.2					0	00	4	5		
c8.24xlarge.4	96	384	50	无	1800	1000	64	15	100	48
c8.32xlarge.2	128	256	100	无	3000	1000	64	15	100	48
c8.32xlarge.4	128	512	100	无	3000	1000	64	15	100	48

c7 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
c7.large.2	2	4	4	0.8	40	2
c7.large.4	2	8	4	0.8	40	2
c7.xlarge.2	4	8	8	1.6	80	2
c7.xlarge.4	4	16	8	1.6	80	2

c7.2xlarge.2	8	16	15	3	150	4
c7.2xlarge.4	8	32	15	3	150	4
c7.3xlarge.2	12	24	17	5	200	4
c7.3xlarge.4	12	48	17	5	200	4
c7.4xlarge.2	16	32	20	6	280	8
c7.4xlarge.4	16	64	20	6	280	8
c7.6xlarge.2	24	48	25	9	400	8
c7.6xlarge.4	24	96	25	9	400	8
c7.8xlarge.2	32	64	30	12	550	16
c7.8xlarge.4	32	128	30	12	550	16
c7.12xlarge.	48	96	35	18	750	16

2						
c7.12xlarge.4	48	192	35	18	750	16
c7.16xlarge.2	64	128	36	24	1000	28
c7.16xlarge.4	64	256	36	24	1000	28
c7.24xlarge.2	96	192	40	36	1100	32
c7.24xlarge.4	96	384	40	36	1100	32

c6 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
c6.large.2	2	4	4	1.2	40	2
c6.lar	2	8	4	1.2	40	2

ge.4						
c6.xlarge.2	4	8	8	2.4	80	2
c6.xlarge.4	4	16	8	2.4	80	2
c6.2xlarge.2	8	16	15	4.5	150	4
c6.2xlarge.4	8	32	15	4.5	150	4
c6.3xlarge.2	12	24	17	7	200	4
c6.3xlarge.4	12	48	17	7	200	4
c6.4xlarge.2	16	32	20	9	280	8
c6.4xlarge.4	16	64	20	9	280	8
c6.6xlarge.2	24	48	25	14	400	8
c6.6xlarge.4	24	96	25	14	400	8

c6.8xlarge.2	32	64	30	18	550	16
c6.8xlarge.4	32	128	30	18	550	16
c6.16xlarge.2	64	128	40	36	1000	32
c6.16xlarge.4	64	256	40	36	1000	32

c3 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
c3.large.2	2	4	1.5	0.6	30	2
c3.large.4	2	8	1.5	0.6	30	2
c3.xlarge.2	4	8	3	1	50	2
c3.xlarge.	4	16	3	1	50	2

4						
c3.2xlarge.2	8	16	5	2	90	4
c3.2xlarge.4	8	32	5	2	90	4
c3.4xlarge.2	16	32	10	4	130	4
c3.4xlarge.4	16	64	10	4	130	4
c3.8xlarge.2	32	64	15	8	260	8
c3.8xlarge.4	32	128	15	8	260	8

4.10.3.4 内存型

内存型云主机独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，提供更接近物理服务器的性能，同时提供更大规格的 CPU 和内存组合。适用于对实例性能有一定要求的高内存消耗型业务场景。

- 在售规格：m8a、m8e、m8、m7、m6、m3

适用场景

内存型云主机的 CPU 和内存配比可达 1:8，适用于高内存计算应用：

- 大数据分析
- 核心数据库

内存型弹性云主机特点

规格名称	计算	磁盘类型	网络
m8a	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-192 3.处理器：AMD EPYC™ Genoa 处理器 4.基频/睿频： 2.6GHz/3.7GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD 6.XSSD-0、XSSD-1、XSSD-2	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：1800 万 PPS 4.最大内网带宽：100Gbps
m8e	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-192 3.处理器：第五代英特尔®至强®可扩展处理器 4.基频/睿频：	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：

规格名称	计算	磁盘类型	网络
	2.6GHz/3.1GHz	6.XSSD-0、XSSD-1、XSSD-2	1400 万 PPS 4.最大内网带宽：40Gbps
m8	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-128 3.处理器：第三代英特尔®至强®可扩展处理器 4.基频/睿频：2.8GHz/3.5GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD 6.XSSD-0、XSSD-1、XSSD-2	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：3000 万 PPS 4.最大内网带宽：100Gbps
m7	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-96 3.处理器：第三代英特尔®至强®可扩展处理器 4.基频/睿频：2.8GHz/3.5GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：1100 万 PPS 4.最大内网带宽：40Gbps
m6	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-64 3.处理器：第二代英特尔®至强®可扩展处理器 4.基频/睿频：	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型 SSD	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：

规格名称	计算	磁盘类型	网络
	3.0GHz/4.0GHz		1000 万 PPS 4.最大内网带宽：40Gbps
m3	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-32 3.处理器：英特尔®至强®可扩展处理器 4.基频/睿频： 2.1-2.3GHz/ 3.2-4.0GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：260 万 PPS 4.最大内网带宽：15Gbps

m8a 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
m8a.large.8	2	16	1.5	10	90	25	2	3	2.	1.5

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
									5	
m8a.xlarge.8	4	32	2.5	10	100	25	4	4	3	2
m8a.2xlarge.8	8	64	4	15	160	25	8	4	4.5	2.5
m8a.4xlarge.8	16	128	7	15	200	50	16	8	6	3.5
m8a.8xlarge.8	32	256	10	25	300	100	32	8	8	5
m8a.12xlarge.8	48	384	16	25	450	150	48	8	12	8
m8a.16xlarge.8	64	512	20	25	600	200	64	8	16	10
m8a.24x	96	76	32	无	90	30	6	1	2	16

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
large.8		8			0	0	4	5	4	
m8a.32xlarge.8	128	1024	40	无	1200	400	64	15	32	20

m8e 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
m8e.larg	2	16	0.8	4	90	25	2	3	2	1.5

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
e.8										
m8e.xlarge.8	4	32	1.6	8	100	50	4	4	4	2
m8e.2xlarge.8	8	64	3.5	15	160	100	8	4	5	3
m8e.3xlarge.8	12	96	5	17	240	150	8	8	6	4
m8e.4xlarge.8	16	128	7	25	300	200	8	8	7	5
m8e.6xlarge.8	24	192	10	25	450	300	12	8	8	6
m8e.8xlarge.8	32	256	15	32	600	400	16	8	10	7

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
m8e.12xlarge.8	48	384	20	无	900	600	24	8	12	8
m8e.16xlarge.8	64	512	25	无	1400	800	32	8	16	9
m8e.24xlarge.8	96	768	36	无	1200	800	32	8	20	10
m8e.32xlarge.8	128	1024	40	无	1400	1000	32	15	25	16
m8e.48xlarge.8	192	1536	40	无	1400	1000	32	15	30	20

m8 弹性云主机的规格

规格名	CPU(	内	基准	最大	最大	连	网	弹	云	云盘基

称	核）	存（GB）	带宽（Gbps）	带宽（Gbps）	收发包（万pps）	接数（万）	卡多队列	性网卡	盘基础IOPS（万次）	础带宽（Gbit/s）
m8.large.8	2	16	2.5	15	100	25	2	3	2.5	2
m8.xlarge.8	4	32	4	15	120	50	4	4	5	2.5
m8.2xlarge.8	8	64	6	15	160	100	8	4	6	4
m8.3xlarge.8	12	96	10	15	240	150	12	8	8	5
m8.4xlarge.8	16	128	12	25	300	200	16	8	10	6
m8.6xlarge.8	24	192	15	25	450	300	24	8	12	7.5
m8.8xlarge.8	32	256	20	25	600	400	32	8	20	10
m8.12x	48	384	25	无	900	60	4	8	30	12

large.8						0	8			
m8.16xlarge.8	64	512	32	无	1200	800	64	8	36	20
m8.24xlarge.8	96	768	50	无	1800	1000	64	15	100	48
m8.32xlarge.8	128	1024	100	无	3000	1000	64	15	100	48

m7 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
m7.large.8	2	16	4	0.8	40	2
m7.xlarge.8	4	32	8	1.6	80	2
m7.2xlarge.8	8	64	15	3	150	4
m7.3xlarge.8	12	96	17	5	200	4

m7.4xlarge.8	16	128	20	6	280	8
m7.6xlarge.8	24	192	25	9	400	8
m7.8xlarge.8	32	256	30	12	550	16
m7.12xlarge.8	48	384	35	18	750	16
m7.16xlarge.8	64	512	30	24	1000	28
m7.24xlarge.8	96	768	40	36	1100	32

m6 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
m6.large.8	2	16	4	1.2	40	2
m6.xlarge.8	4	32	8	2.4	80	2

m6.2xlarge.8	8	64	15	4.5	150	4
m6.3xlarge.8	12	96	17	7	200	4
m6.4xlarge.8	16	128	20	9	280	8
m6.6xlarge.8	24	192	25	14	400	8
m6.8xlarge.8	32	256	30	18	550	16
m6.16xlarge.8	64	512	40	36	1000	32

m3 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	最大带宽（Gbps）	基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列
m3.large.8	2	16	1.5	0.6	30	2
m3.xl	4	32	3	1.1	50	2

arge.8						
m3.2xlarge.8	8	64	5	2	90	4
m3.4xlarge.8	16	128	10	4.5	130	4
m3.8xlarge.8	32	256	15	9	260	8

4.10.3.5 增强型

网络增强型云主机配套 25GE 智能网卡，通过软硬结合的方式提高虚拟化性能，大幅提升实例的网络带宽能力和网络收发包能力。

- 在售型号：c8ne、m8ne
- 停售型号：c7ne

适用场景

网络增强型云主机适用各种于网络密集型应用与企业级应用：

- NFV/SD-WAN
- 移动互联网
- 视频弹幕
- 电信业务转发
- 大中型数据库系统、缓存、搜索集群
- 大数据分析和机器学习

网络增强型弹性云主机特点

规格名称	计算	磁盘类型	网络
c8ne	1.CPU/内存配比：1:2/1:4 2.vCPU 数量范围：2-192 3.处理器：第五代英特尔®至强®可扩展处理器 4.基频/睿频： 2.6GHz/3.1GHz	1.XSSD-1	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：3600万 PPS 4.最大内网带宽：64Gbps
m8ne	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-192 3.处理器：第五代英特尔®至强®可扩展处理器 4.基频/睿频： 2.6GHz/3.1GHz	1.XSSD-1	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：3600万 PPS 4.最大内网带宽：64Gbps
c7ne	1.CPU/内存配比：1:2/1:4 2.vCPU 数量范围：2-148 3.处理器：第三代英特尔®至强®可扩展处理器 4.基频/睿频： 2.8GHz/3.5GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO 5.极速型	1.支持 IPv6 2.实例网络性能与计算规格对应，规格越高网络性能越强 3.最大网络收发包：3000万 PPS

规格名称	计算	磁盘类型	网络
		SSD	4.最大内网带宽：32Gbps

c8ne 弹性云主机的规格

规格名称	CPU（核）	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
c8ne.large.2	2	4	2	10	200	12	2	3	1	1
c8ne.large.4	2	8	2	10	200	12	2	3	1	1
c8ne.xl	4	8	3	10	25	25	4	4	2	1.5

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
arge.2					0					
c8ne.xlarge.4	4	16	3	10	250	25	4	4	2	1.5
c8ne.2xlarge.2	8	16	5	10	300	50	8	6	3	2
c8ne.2xlarge.4	8	32	5	10	300	50	8	6	3	2
c8ne.4xlarge.2	16	32	10	15	450	80	16	8	5	3
c8ne.4xlarge.4	16	64	10	15	450	80	16	8	5	3
c8ne.8x	32	64	15	20	90	10	1	8	8	4

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
large.2					0	0	6			
c8ne.8xlarge.4	32	128	15	20	900	100	16	8	8	4
c8ne.12xlarge.2	48	96	20	24	1200	150	16	8	12	6
c8ne.12xlarge.4	48	192	20	24	1200	150	16	8	12	6
c8ne.16xlarge.2	64	128	24	32	1600	200	32	8	16	8
c8ne.16	64	25	24	32	16	20	3	8	16	8

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
xlarge.4		6			00	0	2			
c8ne.24xlarge.2	96	192	32	无	2000	300	32	8	24	12
c8ne.24xlarge.4	96	384	32	无	2000	300	32	8	24	12
c8ne.32xlarge.2	128	256	48	无	2400	400	32	12	32	16
c8ne.32xlarge.4	128	512	48	无	2400	400	32	12	32	16

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
c8ne.48xlarge.2	192	384	64	无	3600	500	32	15	48	24
c8ne.48xlarge.4	192	768	64	无	3600	500	32	15	48	24

m8ne 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
m8ne.large.8	2	16	2	10	200	12	2	3	1	1
m8ne.xlarge.8	4	32	3	10	250	25	4	4	2	1.5
m8ne.2xlarge.8	8	64	5	10	300	50	8	6	3	2
m8ne.4xlarge.8	16	128	10	15	450	80	16	8	5	3
m8ne.8xlarge.8	32	256	15	20	900	100	16	8	8	4
m8ne.12xlarge.8	48	384	20	24	1200	150	16	8	12	6

规格名称	CPU(核)	内存（GB）	基准带宽（Gbps）	最大带宽（Gbps）	最大收发包（万pps）	连接数（万）	网卡多队列	弹性网卡	云盘基础IOPS(万)	云盘基础带宽（Gbit/s）
m8ne.16xlarge.8	64	512	24	32	1600	200	32	8	16	8
m8ne.24xlarge.8	96	768	32	无	2000	300	32	8	24	12
m8ne.32xlarge.8	128	1024	48	无	2400	400	32	12	32	16
m8ne.48xlarge.8	192	1536	64	无	3600	500	32	15	48	24

c7ne 云主机规格

实例规格	vCPU（核）	内存（GiB）	最大带宽（Gbps）	基准带宽（Gbps）	网络收发包PPS（万）	多队列	弹性网卡
c7ne.large.2	2	4	10	1.5	180	2	3
c7ne.large.4	2	8	10	1.5	180	2	3
c7ne.xlarge.2	4	8	12	2.5	210	4	4
c7ne.xlarge.4	4	16	12	2.5	210	4	4
c7ne.2xlarge.2	8	16	16	5	280	8	6
c7ne.2xlarge.4	8	32	16	5	280	8	6
c7ne.4xlarge.2	16	32	20	10	400	16	7
c7ne.4xlarge.4	16	64	20	10	400	16	7
c7ne.8xlarge.2	32	64	无	20	750	16	7

c7ne.8xlarge.4	32	128	无	20	750	16	7
c7ne.16xlarge.2	64	128	无	32	1500	30	7
c7ne.16xlarge.4	64	256	无	32	1500	30	7
c7ne.37xlarge.2	148	296	无	32	3000	30	7
c7ne.37xlarge.4	148	592	无	32	3000	30	7

4.10.4 本地盘云主机

使用须知

- 本地盘云主机所在的物理机发生故障时，不支持弹性云主机的迁移恢复：
  - 部分宿主机硬件故障或亚健康等场景，需要用户配合关闭 ECS 完成宿主机硬件维修动作。
- 本地盘云主机不支持规格变更。
- 本地盘云主机不支持本地盘的快照和备份。
- 可使用本地盘和云硬盘两类磁盘存储数据，通过挂载云硬盘，可以提供更大的存储空间。关于本地盘和云硬盘的使用，有如下约束与限制：
  - 系统盘只能部署在云硬盘上，不可以部署在本地盘上。
  - 数据盘可以部署在云硬盘和本地盘上。

- 本地盘云主机的本地磁盘数据有丢失的风险（比如宿主机宕机或本地磁盘损坏时），如果您的应用不能做到数据可靠性的架构，强烈建议您使用云盘搭建您的弹性云主机。
- 删除本地盘云主机后，本地盘中的数据会被自动清除，请提前做好数据备份。删除本地盘数据的时间较长，因此，资源释放的时间较之常规云主机略长。
- 由于本地盘数据的可靠性取决于物理服务器和硬盘的可靠性，存在单点故障风险，建议您在应用层做好数据冗余，以保证数据的可用性，需要长期保存的业务数据建议使用云硬盘存储。
- 对于本地盘云主机，关机后其基础资源（包括 vCPU、内存、镜像）会继续收费。如需停止计费，需删除弹性云主机。
- 您不能单独购买本地盘，本地盘的数量和容量由您选择的弹性云主机规格决定，只能在创建本地盘云主机的同时购买本地盘。
- Ir3 型弹性云主机的本地盘为拆分型本地盘，一块本地盘可能被多个弹性云主机使用。当本地盘损坏时，会影响多个弹性云主机。建议您在创建 Ir3 型弹性云主机时，将弹性云主机加入云主机组，避免出现一块本地盘损坏影响多个弹性云主机的情况。

适用场景

超高 IO 型云主机：

- 超高 I/O 型云主机适用于高性能关系型数据库
- NoSQL 数据库(Cassandra、MongoDB 等)以及 ElasticSearch 搜索等场景

磁盘增强型云主机：

- 互联网行业、金融行业等有大数据计算与存储分析需求的行业客户，进行海量数据存储和计算的业务场景
- ElasticSearch、日志等

超高 IO 型本地盘 Ip3 云主机

Ip3 云主机搭载 NVMe SSD 本地盘，采用第三代英特尔® 至强® 可扩展处理器（Ice Lake），主频 3.0GHz。Ip3 云主机为 CPU 独享型的 x86 架构云主机。满足 I/O 密集型应用对块存储的低时延和高 I/O 性能需求。在提供高达数十万随机 I/O 读写能力的同时，时延水平保持在 μs 级别。本地盘来自单台物理机，数据可靠

性取决于物理机的可靠性，存在单点故障风险，建议用户在应用层做数据冗余，保证数据的可用性，用户可以做软 raid 的形式实现数据冗余。

规格名称	vCPU	内存（GB）	本地盘（GB）	最大带宽/基准带宽(Gbps)	最大收发包能力(万 pps)	网卡多队列数
Ip3.2xlarge.8	8	64	1 * 3200	8/4.5	100	4
Ip3.4xlarge.8	16	128	2 * 3200	15/9	240	4
Ip3.8xlarge.8	32	256	4 * 3200	30/17	500	8
Ip3.10xlarge.8	40	320	5 * 3200	40/20	600	8
Ip3.16xlarge.8	64	512	8 * 3200	45/33	1000	16

超高 IO 型本地盘 Ir3 云主机

Ir3 云主机搭载 NVMe SSD 本地盘，本地盘为拆分型本地盘，一块本地盘可能被多个弹性云主机使用。建议您在创建 Ir3 型弹性云主机时，将弹性云主机加入云主机组，避免出现一块本地盘损坏影响多个弹性云主机的情况。

采用第三代英特尔® 至强® 可扩展处理器（Ice Lake），主频 3.0GHz。Ir3 云主机为 CPU 独享型的 x86 架构云主机。满足 I/O 密集型应用对块存储的低时延

和高 I/O 性能需求。在提供高达数十万随机 I/O 读写能力的同时，时延水平保持在μs 级别。本地盘来自单台物理机，数据可靠性取决于物理机的可靠性，存在单点故障风险，建议用户在应用层做数据冗余，保证数据的可用性，用户可以做软 raid 的形式实现数据冗余。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万PPS）	网卡多队列数	本地盘（GB）	本地盘类型	IOPS	BPS
ir3.large.4	2	8	4/0.7	70	2	1*125	NVMe SSD	30000	280
ir3.xlarge.4	4	16	8/1.5	120	2	1*250	NVMe SSD	60000	560
ir3.2xlarge.4	8	32	15/3	190	4	2*250	NVMe SSD	2*60000	2*560
ir3.4xlarge.4	16	64	30/7	300	8	2*500	NVMe SSD	2*120000	2*1120
ir3.8xlarge.	32	128	40/15	420	16	4*500	NVMe	4*120000	4*11

4							SSD		20
ir3.12xlarge.4	48	192	45/20	600	16	3*1000	NVMe SSD	3*2400000	3*2240

磁盘增强型本地盘 D3 云主机

d3 云主机搭载 SATA HDD 本地盘，CPU 为 intel 4316，为 CPU 独享型的 x86 架构云主机。磁盘增强型实例旨在解决大数据时代下海量业务数据的云上计算和存储难题。

规格名称	vCPU	内存（GB）	本地盘（GB）	最大带宽/基准带宽(Gbps)	最大收发包能力(万 pps)	网卡多队列数
D3.xlarge.4	4	16	1 * 8000	5/2	60	2
D3.2xlarge.4	8	32	2 * 8000	10/5	120	4
D3.4xlarge.4	16	64	4 * 8000	20/14	300	8
D3.6xlarge.4	24	96	6 * 8000	30/21	400	8
D3.8xlarge.4	32	128	8 * 8000	38/32	550	16
D3.12xlarge.4	48	192	10 * 8000	42/37	750	16

D3.14xlarge.4	56	224	12 * 8000	45/42	900	16

本地盘性能

Ip3 型弹性云主机本地盘 IOPS 性能

规格名称	4KB 随机读最大 IOPS
Ip3.2xlarge.8	500000
Ip3.4xlarge.8	1000000
Ip3.8xlarge.8	2000000
Ip3.10xlarge.8	2500000
Ip3.16xlarge.8	4000000

Ip3 型弹性云主机 NVME SSD 本地盘单盘指标

指标	性能
磁盘容量	3.2T
读 IOPS（4KB 随机读）	500000
写 IOPS（4KB 随机写）	200000
读吞吐量	6.2 GiB/s
写吞吐量	4.5 GiB/s

访问时延	微秒级

D3 型弹性云主机 SATA HDD 单盘指标

指标	性能
磁盘容量	8000 GB
最大吞吐量	300 MBps
访问时延	毫秒级

4.10.5GPU 加速/AI 加速云主机

GPU/AI 加速云主机能够提供优秀的浮点计算能力，从容应对高实时、高并发的海量计算场景。特别适合于深度学习、科学计算、CAE、3D 动画渲染、CAD 等应用。

使用 NVIDIA 显卡的 GPU 云主机分为计算加速型（PN8I、PN8S、P8A、PI7、P2V、P2Vs、PI2）和图形加速基础型（G7、G6、G5、G5s），其中计算加速型采用 GPU 直通技术、图形加速基础型采用 GPU 虚拟化技术；国产 AI 加速云主机支持昇腾 AI 加速卡、寒武纪 AI 加速卡（PAK1、PCH1），采用直通技术。建议您先了解产品规格、性能和使用限制等信息，然后按照实际需要进行选择。

NVIDIA GPU 云主机

GPU 计算加速型

GPU 计算加速型 GPU 云主机采用 GPU 硬件直通技术，主要适用于 AI 深度学习训练、推理、科学计算、视频转码、图形处理等场景。

在售：PN8I、PN8S、P8A、PI7、P2V、P2Vs、PI2

计算加速型 GPU 云主机特点

规格名称	显卡型号	显卡数量	单卡 GPU 性能	磁盘类型
PN8I	Nvidia L20	1、2、4、8	119.5 TFLOPS 半精度浮点计算 59.8 TFLOPS 单精度浮点计算 239 TFLOPS FP8 浮点计算 59.8 TFLOPS TF32AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
PN8S	Nvidia L40S	1、2、4、8	366.5 TFLOPS 半精度浮点计算 91.6 TFLOPS 单精度浮点计算 733 TFLOPS FP8 浮点计算 183 TFLOPS TF32AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
P8A	Nvidia Tesla A100（40G PCIE）	1、2、4	312 TFLOPS 半精度浮点计算 19.5 TFLOPS 单精度浮点计算 9.7 TFLOPS 双精度浮点计算 156 TFLOPS TF32AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO

规格名称	显卡型号	显卡数量	单卡 GPU 性能	磁盘类型
PI7	Nvidia Tesla A10	1、2、4	125 TFLOPS 半精度浮点计算 31.2 TFLOPS 单精度浮点计算 62.5 TFLOPS TF32AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
P2V	Nvidia Tesla V100	1、2、4	14 TFLOPS 单精度浮点计算 7 TFLOPS 双精度浮点计算 112 TFLOPS TF32 AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
P2Vs	Nvidia Tesla V100s	1、2、4	16.4 TFLOPS 单精度浮点计算 8.2 TFLOPS 双精度浮点计算 130 TFLOPS TF32 AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
PI2	Nvidia Tesla T4	1、2、4	8.1 TFLOPS 单精度浮点计算 130 TOPS INT8 计算 260 TOPS INT4 计算	普通 IO 高 IO 通用型 SSD 超高 IO

PN8I 型云主机

PN8I 型云主机采用 NVIDIA L20 GPU，采用 GPU 直通技术，使用第五代英特尔 ® 至强® 可扩展处理器，独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，在提供云主机灵活性的同时，提供高性能计算能力和优秀的性价比。单卡能够提供 119.5TFLOPS 的半精度浮点运算能力和 59.8 TFLOPS 的单精度浮点运算能力，显存带宽达到 864GB/s。PN8I 型云主机能够提供超高的计算能力，适用于 AIGC 图形生成、LLM 大模型推理微调、通用 AI 识别等众多场景。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）	连接数（万）
pn8i.4xlarge.8	16	128	1×L20	1×48GB	KVM	12/4	8	175	90
pn8i.8xlarge.8	32	256	2×L20	2×48GB	KVM	20/8	16	350	175
pn8i.16xlarge.8	64	512	4×L20	4×48GB	KVM	32/16	32	700	350
pn8i.32xlarge.8	128	1024	8×L20	8×48GB	KVM	45/32	32	1400	700

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）	连接数（万）
pn8i.43xlarge.8	172	1336	8×L20	8×48GB	KVM	47/45	32	1800	1000

常规软件支持列表

PN8I 型云主机主要用于 AI 深度学习、图形处理场景，例如 LLM 大模型推理微调、通用 AI 识别、AIGC 图形生成等众多场景。应用软件如果使用到 GPU 的 CUDA 并行计算能力，可以使用 PN8I 型云主机。常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等深度学习框架。
- RedShift for Autodesk 3dsMax、V-Ray for 3ds Max 等支持 CUDA 的 GPU 渲染。

使用须知

PN8I 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准。

PN8S 型云主机

该实例目前仅部分地域支持，如有需求，请联系天翼云客服。

PN8S 型云主机采用 NVIDIA L40S GPU，采用 GPU 直通技术，使用第四代英特尔 ® 至强® 可扩展处理器，独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，在提供云主机灵活性的同时，提供高性能计算能力和优秀的性价比。单卡能够提供 366.5TFLOPS 的半精度浮点运算能力和 91.6 TFLOPS 的单精度浮点运算能力，显存带宽达到 864GB/s。PN8I 型云主机能够提供超高的计算能力，适用于 AIGC 图形生成、LLM 大模型推理微调、通用 AI 识别等众多场景。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）	连接数（万）
pn8s.5xlarge.4	20	74	1×L40S	1×48GB	KVM	17.5/5.5	8	125	125
pn8s.11xlarge.4	44	148	2×L40S	2×48GB	KVM	23/11.5	16	250	250
pn8s.22xlarge.4	88	296	4×L40S	4×48GB	KVM	36/23.5	32	500	500

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）	连接数（万）
pn8s.44xlarge.4	176	592	8×L40S	8×48GB	KVM	47/45	32	1300	1000

常规软件支持列表

PN8S 型云主机主要用于 AI 深度学习、图形处理场景，例如 LLM 大模型推理微调、通用 AI 识别、AIGC 图形生成等众多场景。应用软件如果使用到 GPU 的 CUDA 并行计算能力，可以使用 PN8I 型云主机。常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等深度学习框架。
- RedShift for Autodesk 3dsMax、V-Ray for 3ds Max 等支持 CUDA 的 GPU 渲染。

使用须知

PN8S 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准。

P8A 型云主机

P8A 型云主机采用 NVIDIA A100 40GB PCIE GPU，采用 GPU 直通技术，使用第三代英特尔 ® 至强® 可扩展处理器（主频 2.6GHz），独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，在提供云主机灵活性的同时，提供高性能计算能力和优秀的性价比，单卡能够提供最大 312 TFLOPS 的半精度浮点运算能力和 9.746 TFLOPS 的双精度浮点运算能力。P8A 型云主机能够提供超高的通用计算能力，适用于 AI 深度学习、科学计算，在深度学习训练、科学计算、计算流体动力学、计算金融、地震分析、分子建模、基因组学等领域都能表现出巨大的计算优势。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）
p8a.6xlarge.4	24	96	1×A100	1×40GB	KVM	30/11	8	300
p8a.12xlarge.4	48	192	2×A100	2×40GB	KVM	36/23	16	600
p8a.24xlarge.4	96	384	4×A100	4×40GB	KVM	47/45	32	1000

常规软件支持列表

P8A 型云主机主要用于计算加速场景，例如深度学习训练、推理、科学计算、分子建模、地震分析等场景。应用软件如果使用到 GPU 的 CUDA 并行计算能力，可以使用 P8A 型云主机。常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等常用深度学习框架。

使用须知

P8A 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

PI7 型云主机

PI7 型云主机采用专为 AI 推理打造的 NVIDIA A10 Tensor Core GPU，采用 GPU 直通技术，使用第三代英特尔 ® 至强® 可扩展处理器（主频 2.6GHz），独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配， 能够提供超强的实时推理能力，同时也具备图像渲染能力。PI7 型弹性云主机借助 A10，单卡能够提供最大 31.2 TFLOPS 的 FP32 算力。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）
pi7.4xlarge.4	16	64	1×A10	1×24GB	KVM	17/7.5	8	200
pi7.8xlarge.4	32	128	2×A10	2×24GB	KVM	25/15	16	400
pi7.16xlarge	64	256	4×	4×	KV	47/45	3	800

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）
.4			A10	24GB	M		2	

常规支持软件列表

PI7 型主要用于 GPU 推理计算场景，例如图片识别、语音识别、自然语言处理等场景。也可以支持轻量级训练场景和视频编解码场景。

常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等深度学习框架。
- RedShift for Autodesk 3dsMax、V-Ray for 3ds Max 等支持 CUDA 的 GPU 渲染。

使用须知

PI7 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

P2V 型云主机

P2V 型云主机采用 NVIDIA Tesla V100 PCIE GPU，采用 GPU 直通技术，在提供云主机灵活性的同时，提供高性能计算能力和优秀的性价比。P2V 型云主机能够提供超高的通用计算能力，适用于 AI 深度学习、科学计算，在深度学习训练、科

学计算、计算流体动力学、计算金融、地震分析、分子建模、基因组学等领域都能表现出巨大的计算优势。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型
p2v.4xlarge.8	16	128	1*V100	1*32GB	KVM
p2v.8xlarge.8	32	256	2*V100	2*32GB	KVM
p2v.2xlarge.4	8	32	1*V100	1*32GB	KVM
p2v.4xlarge.4	16	64	2*V100	2*32GB	KVM
p2v.8xlarge.4	32	128	4*V100	4*32GB	KVM

常规软件支持列表

P2V 型云主机主要用于计算加速场景，例如深度学习训练、推理、科学计算、分子建模、地震分析等场景。应用软件如果使用到 GPU 的 CUDA 并行计算能力，可以使用 P2V 型云主机。常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等常用深度学习框架。
- RedShift for Autodesk 3dsMax、V-Ray for 3ds Max 等支持 CUDA 的 GPU 渲染。

使用须知

P2V 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

P2Vs 型云主机

P2Vs 型云主机采用 NVIDIA Tesla V100s PCIE GPU，采用 GPU 直通技术，在提供云主机灵活性的同时，提供高性能计算能力和优秀的性价比。P2Vs 型云主机能够提供超高的通用计算能力，适用于 AI 深度学习、科学计算，在深度学习训练、科学计算、计算流体动力学、计算金融、地震分析、分子建模、基因组学等领域都能表现出巨大的计算优势。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型
p2vs.4xlarge.8	16	128	1*V100s	1*32GB	KVM
p2vs.8xlarge.8	32	256	2*V100s	2*32GB	KVM
p2vs.2xlarge.4	8	32	1*V100s	1*32GB	KVM
p2vs.4xlarge.4	16	64	2*V100s	2*32GB	KVM
p2vs.8xlarge.4	32	128	4*V100s	4*32GB	KVM

常规软件支持列表

P2Vs 型云主机主要用于计算加速场景，例如深度学习训练、推理、科学计算、分子建模、地震分析等场景。应用软件如果使用到 GPU 的 CUDA 并行计算能力，可以使用 P2Vs 型云主机。常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等常用深度学习框架。
- RedShift for Autodesk 3dsMax、V-Ray for 3ds Max 等支持 CUDA 的 GPU 渲染。

使用须知

P2Vs 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

PI2 型云主机

PI2 型云主机采用专为 AI 推理打造的 NVIDIA Tesla T4 GPU，采用 GPU 直通技术，能够提供超强的实时推理能力。PI2 型弹性云主机借助 T4 的 INT8 运算器，能够提供最大 130 TOPS 的 INT8 算力。PI2 也可以支持轻量级训练场景。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型
pi2.2xlarge.4	8	32	1×T4	1×16GB	KVM
pi2.4xlarge.4	16	64	2×T4	2×16GB	KVM
pi2.8xlarge.4	32	128	4×T4	4×16GB	KVM

常规支持软件列表

PI2 型云主机主要用于 GPU 推理计算场景，例如图片识别、语音识别、自然语言处理等场景。也可以支持轻量级训练场景。

常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet 等常用深度学习框架。
- RedShift for Autodesk 3dsMax、V-Ray for 3ds Max 等支持 CUDA 的 GPU 渲染。

使用须知

PI2 型云主机当前支持如下类型的操作系统：

- Windows Server
- CentOS
- Ubuntu
- Ctyunos

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

GPU 图形加速基础型

图形加速基础型 GPU 云主机基于 NVIDIA GRID 虚拟化 GPU 技术，公共镜像集成 GRID 驱动，并包含 NVIDIA GRID vWS 的软件 License，能够有效降低小规模需求的使用成本，同时主机共享宿主机的 CPU 资源，适用于图像渲染和小规模 AI 推理等场景。

在售：G7、G6、G5、G5s

图形加速基础型 GPU 云主机特点

规格名称	显卡型号	显卡数量	单卡 GPU 性能	磁盘类型
G7	Nvidia Tesla A10	1/4、1/2、1	31.2 TFLOPS 单精度浮点计算 62.5 TFLOPS TF32AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
G6	Nvidia Tesla T4	1/4、1/2	31.2 TFLOPS 单精度浮点计算 62.5 TFLOPS TF32AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
G5	Nvidia Tesla V100	1/16、1/8、1/4、1/2	14 TFLOPS 单精度浮点计算 7 TFLOPS 双精度浮点计算 112 TFLOPS TF32 AI 加速	普通 IO 高 IO 通用型 SSD 超高 IO
G5s	Nvidia Tesla	1/16、1/8、1/4、	16.4 TFLOPS 单精度浮	普通 IO

规格名称	显卡型号	显卡数量	单卡 GPU 性能	磁盘类型
	V100s	1/2	点计算 8.2 TFLOPS 双精度浮点计算 130 TFLOPS TF32 AI 加速	高 IO 通用型 SSD 超高 IO

G7 型云主机

G7 型云主机基于 NVIDIA GRID 虚拟化 GPU 技术，能够提供全面的专业级的图形加速能力。G7 型云主机使用 NVIDIA A10 Tensor Core GPU 显卡，能够支持 DirectX、OpenGL、Vulkan 接口，提供 6/12/24 GB 三种显存规格，支持 7680*4320 分辨率 ，满足从入门级到专业级的图形处理需求。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型	最大带宽（Gbps）/基准带宽(Gbps)	网卡多队列数	最大收发包能力（万PPS）
g7.2xlarge.4	8	32	A10-6Q	6	KVM	8/2.5	4	110
g7.4xlarge.4	16	64	A10-12Q	12	KVM	15/4.5	8	220
g7.8xlarge.4	32	128	A10-24Q	24	KVM	20/9	16	440

G7 型云主机功能如下：

- 处理器与内存配比为 1:4
- 支持图形加速接口：
  - DirectX 12, Direct2D, DirectX Video Acceleration(DXVA)
  - OpenGL 4.5
  - Vulkan 1.0
- 支持 CUDA 和 OpenCL
- 支持 Quadro vDWS 特性，为专业级图形应用提供加速
- 支持 NVIDIA A10 GPU 卡
- 支持图形加速应用
- 提供 GPU 硬件虚拟化（vGPU）
- 提供和弹性云主机相同的申请流程
- 自动化的调度 G7 型弹性云主机到装有 NVIDIA A10 GPU 卡的可用区
- 可以提供最大显存 24GB，分辨率为 7680*4320 的图形图像处理能力

常规支持软件列表

G7 型云主机主要用于图形加速场景，例如图像渲染、云桌面、3D 可视化。应用软件如果依赖 GPU 的 DirectX、OpenGL 硬件加速能力可以使用 G7 型云主机。常用的图形处理软件支持列表如下：

- AutoCAD
- 3DS MAX
- MAYA
- Agisoft PhotoScan
- ContextCapture

使用须知

G7 型云主机当前支持如下版本的操作系统：

- Windows Server 2019 DataCenter 64bit
- Windows Server 2016 DataCenter 64bit
- Windows Server 2012 DataCenter 64bit

- CentOS 8.1 64bit（目前仅多 AZ 资源池提供）
- CentOS 8.2 64bit（目前仅多 AZ 资源池提供）
- Ubuntu Server 20.04 64bit（目前仅多 AZ 资源池提供）

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

G6 型云主机

G6 型云主机基于 NVIDIA GRID 虚拟化 GPU 技术，使用 NVIDIA Tesla T4 GPU 显卡，能够支持 DirectX、OpenGL、Vulkan 接口，提供 8GiB 显存，理论性能 Pixe Rate：101.8GPixel/s，Texture Rate：254.4GTexel/s，满足专业级的图形处理需求。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型
g6.xlarge.4	4	16	T4-4Q	4	KVM
g6.2xlarge.4	8	32	T4-8Q	8	KVM

常规支持软件列表

G6 型云主机主要用于图形加速场景，例如图像渲染、云桌面、3D 可视化。应用软件如果依赖 GPU 的 DirectX、OpenGL 硬件加速能力可以使用 G6 型云主机。常用的图形处理软件支持列表如下：

- AutoCAD
- 3DS MAX
- MAYA
- Agisoft PhotoScan
- ContextCapture

使用须知

G6 型云主机当前支持如下版本的操作系统：

- Windows Server 2016 Standard 64bit
- Windows Server 2012 Standard 64bit

- CentOS 7.5 64bit
- CentOS 7.6 64bit
- Ubuntu Server 16.04 64bit

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

G5 型云主机

G5 型云主机基于 NVIDIA GRID 虚拟化 GPU 技术，能够提供全面的专业级的图形加速能力。G5 型云主机使用 NVIDIA Tesla V100 PCIE GPU 显卡，能够支持 DirectX、OpenGL、Vulkan 接口，提供 2/4/8/16 GB 四种显存规格，支持最大 4096*2160 分辨率，满足从入门级到专业级的图形处理需求。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型
g5.2xlarge.2.1	8	16	V100-2Q	2	KVM
g5.2xlarge.2	8	32	V100-4Q	4	KVM
g5.2xlarge.8	8	64	V100-16Q	16	KVM
g5.4xlarge.4	16	64	V100-8Q	8	KVM
g5.8xlarge.4	32	128	V100-16Q	16	KVM

G5 型云主机功能如下：

- 处理器与内存配比为 1:4/1:2/1:8
- 支持图形加速接口：
  - DirectX 12, Direct2D, DirectX Video Acceleration(DXVA)
  - OpenGL 4.5
  - Vulkan 1.0
- 支持 CUDA 和 OpenCL
- 支持 Quadro vDWS 特性，为专业级图形应用提供加速
- 支持 NVIDIA V100 GPU 卡

- 支持图形加速应用
- 提供 GPU 硬件虚拟化（vGPU）
- 提供和弹性云主机相同的申请流程
- 自动化的调度 G5 型弹性云主机到装有 NVIDIA V100 GPU 卡的可用区
- 可以提供最大显存 16GB，分辨率为 4096×2160 的图形图像处理能力

常规支持软件列表

G5 型云主机主要用于图形加速场景，例如图像渲染、云桌面、3D 可视化。应用软件如果依赖 GPU 的 DirectX、OpenGL 硬件加速能力可以使用 G5 型云主机。常用的图形处理软件支持列表如下：

- AutoCAD
- 3DS MAX
- MAYA
- Agisoft PhotoScan
- ContextCapture

使用须知

G5 型云主机当前支持如下版本的操作系统：

- Windows Server 2016 Standard 64bit
- Windows Server 2012 Standard 64bit
- CentOS 7.5 64bit
- CentOS 7.6 64bit
- Ubuntu Server 16.04 64bit

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

G5s 型云主机

G5s 型云主机基于 NVIDIA GRID 虚拟化 GPU 技术，能够提供全面的专业级的图形加速能力。G5s 型云主机使用 NVIDIA Tesla V100s PCIE GPU 显卡，能够支持 DirectX、OpenGL、Vulkan 接口，提供 2/4/8/16 GB 四种显存规格，支持最大 4096*2160 分辨率，满足从入门级到专业级的图形处理需求。

规格名称	vCPU	内存（GB）	GPU	显存（GB）	虚拟化类型
g5s.2xlarge.2.1	8	16	V100s-2Q	2	KVM
g5s.2xlarge.2	8	32	V100s-4Q	4	KVM
g5s.2xlarge.8	8	64	V100s-16Q	16	KVM
g5s.4xlarge.4	16	64	V100s-8Q	8	KVM
g5s.8xlarge.4	32	128	V100s-16Q	16	KVM

G5s 型云主机功能如下：

- 处理器与内存配比为 1:4/1:2/1:8
- 支持图形加速接口：
  - DirectX 12, Direct2D, DirectX Video Acceleration(DXVA)
  - OpenGL 4.5
  - Vulkan 1.0
- 支持 CUDA 和 OpenCL
- 支持 Quadro vDWS 特性，为专业级图形应用提供加速
- 支持 NVIDIA V100s GPU 卡
- 支持图形加速应用
- 提供 GPU 硬件虚拟化（vGPU）
- 提供和弹性云主机相同的申请流程。
- 自动化的调度 G5s 型弹性云主机到装有 NVIDIA V100s GPU 卡的可用区
- 可以提供最大显存 16GB，分辨率为 4096×2160 的图形图像处理能力

常规支持软件列表

G5s 型云主机主要用于图形加速场景，例如图像渲染、云桌面、3D 可视化。应用软件如果依赖 GPU 的 DirectX、OpenGL 硬件加速能力可以使用 G5s 型云主机。常用的图形处理软件支持列表如下：

- AutoCAD

- 3DS MAX
- MAYA
- Agisoft PhotoScan
- ContextCapture

使用须知

G5s 型云主机当前支持如下版本的操作系统：

- Windows Server 2016 Standard 64bit
- Windows Server 2012 Standard 64bit
- CentOS 7.5 64bit
- CentOS 7.6 64bit
- Ubuntu Server 16.04 64bit

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

国产 AI 加速云主机

国产 AI 加速云主机支持昇腾 AI 加速卡、寒武纪 AI 加速卡（PAK1、PCH1），采用硬件直通技术。

在售：PAK1、PCH1

国产 AI 加速云主机特点

规格名称	AI 加速卡型号	AI 加速卡数量	单卡性能	磁盘类型
PAK1	Atlas 300I pro	1、2、4	70 TFLOPS 半精度浮点计算 140 TOPS INT8 计算	普通 IO 高 IO 通用型 SSD 超高 IO
PCH1	MLU370-S4	1、2、3、4	72 TFLOPS 半精度浮点计	普通 IO

规格名称	AI 加速卡型号	AI 加速卡数量	单卡性能	磁盘类型
			算 192 TOPS INT8 计算	高 IO 通用型 SSD 超高 I

昇腾计算加速型

PAK1 型云主机

PAK1 型昇腾计算加速型云主机采用专为 AI 推理打造的 Atlas 300I pro 加速卡（Ascend 310 芯片），国产 ARM 架构鲲鹏 CPU，独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，属于计算加速型（直通）规格，云主机的 CPU/内存配比为 1：4，主要适用于搜索推荐、内容审核和 OCR 系统等推理场景。

系列	CPU	内存	AI 加速卡类型	AI 芯片	显存	最大带宽/基准带宽（Gbit/s）	网络收发包（万 PPS）	多队列
pak1.4xlarge.4	18	72	Huawei Atlas 300I pro	Ascend 310	1*24G	18/11.5	240	8
pak1.9xlarge.4	36	144	Huawei Atlas	Ascend 310	2*24G	30/22.5	480	16

系列	CPU	内存	AI 加速卡类型	AI 芯片	显存	最大带宽/基准带宽（Gbit/s）	网络收发包（万PPS）	多队列
			300I pro					
pak1.18xlarge.4	72	288	Huawei Atlas 300I pro	Ascend 310	4*24G	47/45	960	32

支持的镜像

- CTyunOS 2.0.1 64 位 ARM 版昇腾 24.1.0.1 专用
- 银河麒麟高级服务器操作系统 V10 SP3 64 位 ARM 版昇腾 24.1.0.1 专用

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

注意

PAK1 型云主机的公共镜像中均已默认安装了 NPU 驱动及 CANN 包。

寒武纪计算加速型

PCH1 型云主机

PCH1 型寒武纪计算加速型云主机采用专为 AI 推理打造的 MLU370-S4 加速卡，国产 X86 架构海光 CPU，独享宿主机的 CPU 资源，实例间无 CPU 争抢，没有进行资源超配，属于计算加速型（直通）规格，云主机的 CPU/内存配比为 1：4，可广泛支持视觉、语音、自然语言处理等高度多样化的人工智能应用，帮助 AI 推理平台实现超高密度。

系列	CPU	内存	AI 加速卡类型	显存	最大带宽/基准带宽（Gbit/s）	网络收发包（万PPS）	多队列
pch1.4xlarge.4	16	64	Cambricon MLU370 s4	1*24G	18/11.5	200	8
pch1.6xlarge.4	24	96	Cambricon MLU370 s4	1*24G	18/11.5	200	8
pch1.9xlarge.4	36	144	Cambricon MLU370 s4	2*24G	30/22.5	400	16
pch1.12xlarge.4	48	192	Cambricon MLU370 s4	3*24G	30/22.5	400	16
pch1.21xlarge.3	84	252	Cambricon MLU370 s4	4*24G	47/45	800	32

支持的镜像

- CentOS 7.8 64 位 64 位
- CTyunOS 2.0.1 64 位 ARM 版
- KylinOS V10 SP1 64 位

备注：各资源池支持的具体版本可能略有出入，请以控制台实际支持的版本为准

注意

PCH1 型云主机的公共镜像中均已默认安装了 cnmon 驱动及 cntoolkit 工具包。

4.10.6 国产云主机

4.10.6.1 鲲鹏系列

鲲鹏通用型

鲲鹏通用型搭载鲲鹏处理器，云主机实例共享 CPU。提供基础计算能力，可满足基础业务上云需求。

鲲鹏通用型适用于不会经常或始终用尽 vCPU 性能的场景，如小型网站、轻量级研发测试环境、小型数据库等。

规格特点

规格名称	计算	磁盘类型	网络
鲲鹏通用型 ks1	1. CPU/内存配比：1:1/1:2/1:4 2. vCPU 数量范围：1-16 3. 处理器：鲲鹏 920 4. 基频：2.6GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD 4. 超高 IO	1. 实例网络性能与计算规格对应，规格越高网络性能越强 2. 最大网络收发包：100 万 PPS 3. 最大内网带宽：6Gbps

ks1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
ks1.small.1	1	1	0.8/0.1	10	1
ks1.medium.2	1	2	0.8/0.1	10	1

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
ks1.medium.4	1	4	0.8/0.1	10	1
ks1.large.2	2	4	1.5/0.2	15	1
ks1.large.4	2	8	1.5/0.2	15	1
ks1.xlarge.2	4	8	2/0.35	25	1
ks1.xlarge.4	4	16	2/0.35	25	1
ks1.2xlarge.2	8	16	3/0.75	50	2
ks1.2xlarge.4	8	32	3/0.75	50	2
ks1.4xlarge.2	16	32	6/2	100	4
ks1.4xlarge.4	16	64	6/2	100	4

鲲鹏计算型

鲲鹏计算型搭载鲲鹏处理器，云主机实例独享 CPU，提供更稳定的 CPU 性能以及更优秀的网络性能，对业务稳定性提供更好支持。

鲲鹏计算型适用于对计算与网络有更高性能要求的 Web 应用、电商平台、短视频平台、在线游戏、保险金融等各类应用。

规格特点

规格名称	计算	磁盘类型	网络
鲲鹏计算型 kc1	1.CPU/内存配比：1:2/1:4 2.vCPU 数量范围：	1.普通 IO 2.高 IO	1.实例网络性能与计算规格对应，规格越高网络 2.性能越强

规格名称	计算	磁盘类型	网络
	2-32 3. 处理器:鲲鹏 920 4. 基频：2.6GHz	3. 通用型 SSD 4. 超高 IO	3. 最大网络收发包：400 万 PPS 4. 最大内网带宽：30Gbps

kc1 弹性云主机的规格

注意

“央企北京 1”和“佛山 3”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
kc1.large.2	2	4	2/0.5	20	1
kc1.large.4	2	8	3/0.8	30	2
kc1.xlarge.2	4	8	5/1.5	50	2
kc1.xlarge.4	4	16	7/3	80	4
kc1.2xlarge.2	8	16	9/4.5	110	4
kc1.2xlarge.4	8	32	12/6	140	4
kc1.4xlarge.2	16	32	15/8.5	200	8
kc1.4xlarge.4	16	64	18/10	260	8
kc1.8xlarge.2	32	64	25/16	350	16

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
kc1.8xlarge.4	32	128	30/20	400	16

鲲鹏内存型

鲲鹏内存型搭载鲲鹏处理器，云主机实例独享 CPU，提供更稳定的 CPU 性能、更大内存比以及更优秀的网络性能，对内存型应用的运行提供更好支持。

鲲鹏内存型适用于内存要求高，数据量大并且数据访问量大，同时要求快速的数据交换和处理的应用。

规格特点

规格名称	计算	磁盘类型	网络
鲲鹏内存型 km1	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-32 3.处理器：鲲鹏 920 4.基频：2.6GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.超高 IO	1.实例网络性能与计算规格对应，规格越高 2.网络性能越强 3.最大网络收发包：400 万 PPS 4.最大内网带宽：30Gbps

km1 弹性云主机的规格

注意

“央企北京 1”和“佛山 3”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
km1.large.8	2	16	3/0.8	30	1

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
km1.xlarge.8	4	32	7/3	80	2
km1.2xlarge.8	8	64	12/6	140	4
km1.4xlarge.8	16	128	15/8	200	8
km1.8xlarge.8	32	256	30/20	400	16

4.10.6.2 海光系列

海光通用型

海光通用型搭载海光处理器，云主机实例共享 CPU。提供基础计算能力，可满足基础业务上云需求。

海光通用型适用于不会经常或始终用尽 vCPU 性能的场景，如小型网站、轻量级研发测试环境、小型数据库等。

规格特点

规格名称	计算	磁盘类型	网络
海光通用型 hs3	1.CPU/内存配比：1:1/1:2/1:4 2.vCPU 数量范围：1-32 3.处理器：Hygon C86 7493 4.基频：2.9GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.XSSD-0、XSSD-1、XSSD-2	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：180 万 PPS 3.最大内网带宽：12Gbps

规格名称	计算	磁盘类型	网络
海光通用型 hs1	1.CPU/内存配比：1:1/1:2/1:4 2.vCPU 数量范围：1-16 3.处理器：Hygon C86 7285 或 Hygon C86 7380 4.基频：2.0GHz	1.普通 IO 2.高 IO 3.通用型 SSD	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：100 万 PPS 3.最大内网带宽：6Gbps

hs3 弹性云主机的规格

规格名称	CPU(核)	内存（GB）	最大带宽（Gbps）/基准带宽（Gbps）	最大收发包（万 pps）	网卡多队列	云盘基础 IOPS（万次）	云盘基础带宽（Gbit/s）
hs3.small.1	1	1	1.5/0.3	15	1	0.35	0.3
hs3.medium.2	1	2	1.5/0.3	15	1	0.35	0.3
hs3.medium.4	1	4	1.5/0.3	15	1	0.35	0.3
hs3.large.2	2	4	2/0.5	20	1	0.6	0.5

规格名称	CPU(核)	内存（GB）	最大带宽（Gbps）/基准带宽（Gbps）	最大收发包（万pps）	网卡多队列	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
hs3.large.4	2	8	2/0.5	20	1	0.6	0.5
hs3.xlarge.2	4	8	3/1	40	2	1	0.8
hs3.xlarge.4	4	16	3/1	40	2	1	0.8
hs3.2xlarge.2	8	16	4/1.5	60	4	1.5	1.2
hs3.2xlarge.4	8	32	4/1.5	60	4	1.5	1.2
hs3.4xlarge.2	16	32	8/3	100	8	2.2	2
hs3.4xlarge.4	16	64	8/3	100	8	2.2	2
hs3.8xlarge.2	32	64	12/4	180	16	3.2	3
hs3.8xlarge.4	32	128	12/4	180	16	3.2	3

hs1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内　存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
hs1.small.1	1	1	0.8/0.1	10	1
hs1.medium.2	1	2	0.8/0.1	10	1
hs1.medium.4	1	4	0.8/0.1	10	1
hs1.large.2	2	4	1.5/0.2	15	1
hs1.large.4	2	8	1.5/0.2	15	1
hs1.xlarge.2	4	8	2/0.35	25	1
hs1.xlarge.4	4	16	2/0.35	25	1
hs1.2xlarge.2	8	16	3/0.75	50	2
hs1.2xlarge.4	8	32	3/0.75	50	2
hs1.4xlarge.2	16	32	6/2	100	4
hs1.4xlarge.4	16	64	6/2	100	4

海光计算型

海光计算型搭载海光处理器，云主机实例独享 CPU，提供更稳定的 CPU 性能以及更优秀的网络性能，对业务稳定性提供更好支持。

海光计算型适用于对计算与网络有更高性能要求的 Web 应用、电商平台、短视频平台、在线游戏、保险金融等各类应用。

规格特点

规格名称	计算	磁盘类型	网络
海光计算型 hc3	1.CPU/内存配比：1:2/1:4 2.vCPU 数量范围：2-64 3.处理器：Hygon C86 7493 4.基频：2.9GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.XSSD-0、XSSD-1 、XSSD-2	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：700 万 PPS 3.最大内网带宽：35Gbps
海光计算型 hc1	1.CPU/内存配比：1:2/1:4 2.vCPU 数量范围：2-32 3.处理器：Hygon C86 7285 或 Hygon C86 7380 4.基频：2.0GHz	1.普通 IO 2.高 IO 3.通用型 SSD	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：400 万 PPS 3.最大内网带宽：30Gbps

hc3 弹性云主机的规格

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数	云盘基础 IOPS（万次）	云盘基础带宽（Gbit/s）
hc3.large.2	2	4	5/1.5	50	2	1.2	1

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万PPS）	网卡多队列数	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
hc3.large.4	2	8	5/1.5	50	2	1.2	1
hc3.xlarge.2	4	8	8/2.5	90	4	2	1.6
hc3.xlarge.4	4	16	8/2.5	90	4	2	1.6
hc3.2xlarge.2	8	16	15/3	150	8	3	2.4
hc3.2xlarge.4	8	32	15/3	150	8	3	2.4
hc3.4xlarge.2	16	32	20/4	280	16	4.5	4
hc3.4xlarge.4	16	64	20/4	280	16	4.5	4
hc3.8xlarge.2	32	64	30/8	550	32	6.4	6
hc3.8xlarge.4	32	128	30/8	550	32	6.4	6
hc3.16xlarge.2	64	128	35/16	700	32	12.8	12
hc3.16xlarge.4	64	256	35/16	700	32	12.8	12

hc1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内　存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
hc1.large.2	2	4	2/0.5	20	1
hc1.large.4	2	8	3/0.8	30	2
hc1.xlarge.2	4	8	5/1.5	50	2
hc1.xlarge.4	4	16	7/3	80	4
hc1.2xlarge.2	8	16	9/4.5	110	4
hc1.2xlarge.4	8	32	12/6	140	4
hc1.4xlarge.2	16	32	15/8.5	200	8
hc1.4xlarge.4	16	64	18/10	260	8
hc1.8xlarge.2	32	64	25/16	350	16
hc1.8xlarge.4	32	128	30/20	400	16

海光内存型

海光内存型搭载海光处理器，云主机实例独享 CPU，提供更稳定的 CPU 性能、更大内存比以及更优秀的网络性能，对内存型应用的运行提供更好支持。

海光内存型适用于内存要求高，数据量大并且数据访问量大，同时要求快速的数据交换和处理的应用。

规格特点

规格名称	计算	磁盘类型	网络
海光内存型 hm3	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-32 3.处理器：Hygon C86 7493 4.基频：2.9GHz	1.普通 IO 2.高 IO 3.通用型 SSD 4.XSSD-0、XSSD-1 、XSSD-2	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：550 万 PPS 3.最大内网带宽：30Gbps
海光内存型 hm1	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-16 3.处理器：Hygon C86 7285 或 Hygon C86 7380 4.基频：2.0GHz	1.普通 IO 2.高 IO 3.通用型 SSD	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：200 万 PPS 3.最大内网带宽：15Gbps

hm3 弹性云主机的规格

规格名称	**vCPU**	**内存（GB）**	**最大带宽/基准带宽(Gbps)**	**最大收发包能力(万PPS)**	**网卡多队列数**	**云盘基础IOPS（万次）**	**云盘基础带宽（Gbit/s）**
hm3.large.8	2	16	5/1.5	50	2	1.2	1
hm3.xlarge.8	4	32	8/2.5	90	4	2	1.6

规格名称	vCPU	内存（GB）	最大带宽/基准带宽(Gbps)	最大收发包能力(万PPS)	网卡多队列数	云盘基础IOPS（万次）	云盘基础带宽（Gbit/s）
hm3.2xlarge.8	8	64	15/3	150	8	3	2.4
hm3.4xlarge.8	16	128	20/4	280	16	4.5	4
hm3.8xlarge.8	32	256	30/8	550	32	6.4	6

hm1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
hm1.large.8	2	16	3/0.8	30	1
hm1.xlarge.8	4	32	7/3	80	2
hm1.2xlarge.8	8	64	12/6	140	4
hm1.4xlarge.8	16	128	15/8	200	8

4.10.6.3 飞腾系列

飞腾通用型

飞腾通用型搭载飞腾处理器，云主机实例共享 CPU。提供基础计算能力，可满足基础业务上云需求。

飞腾通用型适用于不会经常或始终用尽 vCPU 性能的场景，如小型网站、轻量级研发测试环境、小型数据库等。

规格特点

规格名称	计算	磁盘类型	网络
飞腾通用型 fs1	1. CPU/内存配比：1:1/1:2/1:4 2. vCPU 数量范围：1-16 3. 处理器：飞腾 S2500 4. 基频：2.1GHz	1. 普通 IO 2. 高 IO 3. 通用型 SSD	1. 实例网络性能与计算规格对应，规格越高网络性能越强 2. 最大网络收发包：100 万 PPS 3. 最大内网带宽：6Gbps

fs1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
fs1.small.1	1	1	0.8/0.1	10	1
fs1.medium.2	1	2	0.8/0.1	10	1
fs1.medium.4	1	4	0.8/0.1	10	1
fs1.large.2	2	4	1.5/0.2	15	1

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
fs1.large.4	2	8	1.5/0.2	15	1
fs1.xlarge.2	4	8	2/0.35	25	1
fs1.xlarge.4	4	16	2/0.35	25	1
fs1.2xlarge.2	8	16	3/0.75	50	2
fs1.2xlarge.4	8	32	3/0.75	50	2
fs1.4xlarge.2	16	32	6/2	100	4
fs1.4xlarge.4	16	64	6/2	100	4

飞腾计算型

飞腾计算型搭载飞腾处理器，云主机实例独享 CPU，提供更稳定的计算性能以及更优秀的网络性能，对业务稳定性提供更好支持。

飞腾计算型适用于对计算与网络有更高性能要求的 Web 应用、电商平台、短视频平台、在线游戏、保险金融等各类应用。

规格特点

规格名称	计算	磁盘类型	网络
飞腾计算型 fc1	1.CPU/内存配比：1:2/1:4 2.vCPU 数量范围：2-16 3.处理器：飞腾 S2500	1.普通 IO 2.高 IO 3.通用型 SSD	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：140 万 PPS 3.最大内网带宽：9Gbps

规格名称	计算	磁盘类型	网络
	4.基频：2.1GHz		

fc1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
fc1.large.2	2	4	2/0.5	20	1
fc1.large.4	2	8	3/0.8	30	1
fc1.xlarge.2	4	8	3/0.8	50	1
fc1.xlarge.4	4	16	5/1.5	50	1
fc1.2xlarge.2	8	16	5/1.5	80	2
fc1.2xlarge.4	8	32	7/3	80	2
fc1.4xlarge.2	16	32	7/3	140	4
fc1.4xlarge.4	16	64	9/4.5	140	4

飞腾内存型

飞腾内存型搭载飞腾处理器，云主机实例独享 CPU，提供更稳定的 CPU 性能、更

大内存比以及更优秀的网络性能，对内存型应用的运行提供更好支持。

飞腾内存型适用于内存要求高，数据量大并且数据访问量大，同时要求快速的数据交换和处理的应用。

规格特点

规格名称	计算	磁盘类型	网络
飞腾内存型 fm1	1.CPU/内存配比：1:8 2.vCPU 数量范围：2-16 3.处理器：飞腾 S2500 4.基频：2.1GHz	1.普通 IO 2.高 IO 3.通用型 SSD	1.实例网络性能与计算规格对应，规格越高网络性能越强 2.最大网络收发包：200 万 PPS 3.最大内网带宽：18Gbps

fm1 弹性云主机的规格

注意

“央企北京 1”提供的云主机规格的网络指标（最大带宽、基准带宽、最大收发包能力）与本文内容不一致，请以对应创建页面所展示的数值为准。

规格名称	vCPU	内存（GB）	最大带宽/基准带宽（Gbps）	最大收发包能力（万 PPS）	网卡多队列数
fm1.large.8	2	16	5/1.5	30	1
fm1.xlarge.8	4	32	7/3	80	2
fm1.2xlarge.8	8	64	12/6	140	4
fm1.4xlarge.8	16	128	18/10	200	8

4.10.7 经济型云主机

经济型云主机

经济型云主机共享宿主机的 CPU 资源，提供基本水平的 vCPU 性能、平衡的计算、内存和网络资源，具有较低的价格，有可用性保障，无性能保障。支持小型的业务运行。适用于对价格敏感，性能要求低的场景，如小型网站、开发测试环境、个人云上学习环境等。

适用场景

- 小型网站
- 开发测试环境
- 个人云上学习环境

规格特点

规格名称	计算	磁盘类型
e	1.CPU/内存配比：1:1/1:2/1:4 2.vCPU 数量范围：2-8 3.处理器：英特尔®至强®可扩展处理器	普通 IO 高 IO 通用型 SSD 超高 IO

经济型 e 弹性云主机的规格

实例规格	vCPU	内存（GiB）	网络基准带宽（Gbit/s）/ 最大带宽 网络基准带宽（Gbit/s）	网卡多队列
e.large.1	2	2	0.2/2	1
e.large.2	2	4	0.2/2	1

实例规格	vCPU	内存（GiB）	网络基准带宽（Gbit/s）/ 最大带宽 网络基准带宽（Gbit/s）	网卡多队列
e.large.4	2	8	0.2/2	1
e.xlarge.2	4	8	0.4/3	1
e.xlarge.4	4	16	0.4/4	1
e.2xlarge.2	8	16	0.8/4	1
e.2xlarge.4	8	32	0.8/4	1

4.11 镜像

4.11.1 镜像类型

什么是镜像

镜像是弹性云主机可选择的运行环境模板，一般包括操作系统和预装软件。通过镜像用户可以在弹性云主机上实现应用场景的快速部署。

镜像类型

镜像有多种类型，比如公共镜像为系统默认提供的镜像，私有镜像为用户自己创建的镜像，共享镜像为其他用户共享的私有镜像等。详情可参考：镜像服务-产品简介。

4.11.2Cloud-init

Cloud-init

Cloud-init 或 Cloudbase-init 工具是云平台中 Linux/Windows 镜像里对新开的虚拟机进行初始化的工具，包括初始化主机名、网络配置、初始化密码等功能。Linux 系统公共镜像默认已安装 Cloud-init；Windows 系统公共镜像默认已安装 Cloudbase-init。

关于 Cloud-init 的更多信息，请参见 Cloud-init 官方文档。

Cloud-init 能做什么

Cloud-init 是一款用于初始化云主机的工具，它拥有丰富的模块，能够为云主机提供的能力有：初始化密码、扩容根分区、设置主机名、注入公钥、执行自定义脚本等等，功能十分强大。

安装 Cloud-init

步骤一：检查是否已经安装 Cloud-init 工具

不同的操作系统，命令不同，以 centos 为例，执行以下命令查看系统是否已经安装 Cloud-init：

```
rpm -qa | grep cloud-init
```

步骤二：安装 Cloud-init

无明确版本要求，建议安装系统源上的版本：

采用以下命令安装：

```
yum install cloud-init cloud-utils-growpart -y
```

步骤三：配置 Cloud-init 步骤 3

1. 默认的 cloud.cfg 配置文件/etc/cloud/cloud.cfg 修改如下：

1.1. 确定参数值如下：true/True/1, false/False/0 均可。

```
disable_root: false
ssh_pwauth: true
ssh_deletekeys: false
preserve_hostname: false
```

1.2. 模块包括如下内容（建议参考修改）：

```
cloud_init_modules:
 - migrator
 - seed_random
 - bootcmd
 - write-files
 - growpart
 - resizefs
 - disk_setup
 - mounts
 - set_hostname
 - update_hostname
 - update_etc_hosts
 - ca-certs
 - rsyslog
 - users-groups
 - ssh
```

1.3. 修改 system_info 部分。

distro 内容不做修改。

default_user::默认用户名改为 root，lock_passwd 改为 False。

其余部分涉及系统 yum/apt 源的更改，建议遵循【系统其他配置修改】，保留系统默认，去除 Cloud-init 配置。

system_info 其余内容不做修改。

```
system_info:
   # This will affect which distro class gets used
   distro: openEuler
   # Default user name + that default users groups (if added/used)
   default_user:
     name: root
     lock_passwd: False
```

1.4. 添加 datasource_list 和 datasource 部分，禁用网络托管 。

```
datasource_list: [ ConfigDrive,OpenStack ]
datasource:
  ConfigDrive:
    dsmode: local
  OpenStack:
    metadata_urls: ["http://169.254.169.254"]
    max_wait: 120
    timeout: 10
    retries: 5

network:
  config: disabled
```

4.12 安全

4.12.1 数据保护技术

4.12.1.1Hypervisor 安全

Hypervisor

Hypervisor 是一种运行在物理机上的软件，可以在单个物理机上创建、运行和管理多个虚拟机。Hypervisor 根据需要将底层物理计算资源（如 cpu、内存）进行抽象统一管理，并按需将资源分配给各个虚拟机。Hypervisor 实现了同一物理机上不同虚拟机之间的资源隔离，避免数据窃取或恶意攻击，同时保证虚拟机的资源使用不受周边虚拟机的影响。用户使用虚拟机时，只能访问属于自己的虚拟机的资源（如硬件、软件和数据），不能访问其他虚拟机的资源，保证了虚拟机的数据隔离和安全。

CPU 隔离

CPU 虚拟化是 Hypervisor 中最核心的部分，内存虚拟化和 IO 虚拟化都依赖于 CPU 虚拟化的正确实现。

X86 架构中为了保护指令的运行，提供了指令的 4 个不同特权级别，术语称为 Ring，优先级从高到低依次为：

Ring 0：最高特权级别，被用于运行操作系统内核。

Ring 1：用于操作系统服务。

Ring 2：用于操作系统服务。

Ring 3：用于应用程序。

Hypervisor 运行在最高特权级别，可以控制物理处理器上的所有关键资源；而虚拟机操作系统运行在非最高级特权级别，所以其访问物理资源的敏感指令会陷入到 Hypervisor 中通过软件的方式进行模拟。通过拦截并模拟虚拟机的敏感指令，Hypervisor 实现了虚拟机 vCPU 的隔离，有效防止了虚拟机越权恶意攻击物理机或其他虚拟机。

内存隔离

Hypervisor 通过内存虚拟化技术来实现不同虚拟机之间的内存隔离，使每个虚拟机都拥有独立的内存空间，相互之间不受影响。

相较于传统 OS 两层内存地址映射（“虚拟地址”到“机器地址”），内存虚拟化的核心在于引入一层新的地址空间，叫做 Guest 物理地址空间。Hypervisor 负载管理和分配每个虚拟机的物理内存，虚拟机操作系统看到的是一个虚构的 Guest 物理地址空间，其指令的目标地址也是一个 Guest 物理地址。这样的地址在无虚拟化的情况下，其实就是实际的物理地址，但是在虚拟化场景下，这样的地址是不能直接发送到系统总线上去，需要进行正确的转换后才能交由物理处理器执行。虚拟机内部访问内存时，正确的三层地址转换的逻辑是：虚拟机负责将“Guest 虚拟地址”映射为“Guest 物理地址”，然后 Hypervisor 负责将“Guest 物理地址”映射成“机器地址”。

Hypervisor 的这种地址转换机制，保证了虚拟机无法直接接触实际的机器地址，只能访问 Hypervisor 分配给它的物理内存。

I/O 隔离

物理机外设资源是有限的，为了满足多个虚拟机操作系统的需求，Hypervisor 通过 I/O 虚拟化的方式来复用有限的外设资源。Hypervisor 拦截虚拟机操作系统对设备的访问请求，然后通过软件的方式来模拟真实设备的效果。从处理器的角度看，外设是通过一组 I/O 资源（I/O 端口或 MMIO）来进行访问的，所以设备相关的虚拟化又被称为 I/O 虚拟化。

Hypervisor 采用分离设备模型实现 I/O 的虚拟化。前端负责将虚拟机的 I/O 请求传递到 Hypervisor 中的后端，后端解析 I/O 请求并提交给相应的设备完成 I/O 操作。Hypervisor 保证虚拟机只能访问分配给它的 I/O 资源。

4.12.1.2 用户加密

用户加密，是指使用天翼云产品或其他手段对弹性云主机资源进行加密，从而提升数据的安全性。

云硬盘加密

天翼云云硬盘支持系统盘加密和数据盘加密。

云硬盘加密功能：创建云硬盘时，用户可以选择是否加密此云硬盘，云硬盘创建完成后加密属性无法更改。

云主机系统盘加密：创建云主机时，支持在创建时直接设置系统盘加密。

云主机数据盘加密：创建云主机时，支持在创建时直接设置数据盘加密。

云硬盘加密与快照：加密云硬盘生成的快照及通过这些快照创建的云硬盘将自动继承加密功能属性。

云硬盘加密与备份：加密云硬盘生成的备份及通过这些备份创建的云硬盘将自动继承加密功能属性。

密钥管理

天翼云使用行业标准的 AES-256 算法，利用数据密钥加密您的云硬盘数据，加密云硬盘使用的密钥由天翼云自研密钥管理（KMS，Key Management Service）功能提供，用户可轻松创建并管理密钥，满足数据加解密及数字签名验签等需求，安全便捷。

KMS 通过使用硬件安全模块 HSM（Hardware Security Module）保护密钥的安全，所有的用户密钥都由 HSM 中的根密钥保护，避免密钥泄露。KMS 对密钥的所有操作都会进行访问控制及日志跟踪，并提供所有密钥的使用记录，

满足审计和合规性要求。密钥管理可以轻松满足对小数据和大量数据的加解密。

工作原理

在了解云硬盘加密工作原理之前，首先需要了解两个概念：

- 默认密钥 （Default CMK） ：用户第一次通过对应云服务使用 KMS 加密时，系统自动生成的并托管在用户账号下的服务密钥。
- 用户主密钥 （Customer Master Key，CMK）：用户主密钥包括对称密钥及非对称密钥，主要用于加密保护数据密钥并产生信封，也可直接用于加密少量的数据。用户可以调用 KMS 的 API CreateKey 创建一个用户主密钥。

第一次使用加密云硬盘时，系统会自动创建一个用户主密钥（CMK），该密钥有且仅有一个，且是在 KMS 中的相应地域所创建，并将其存储在受严格的物理和逻辑安全控制保护的密钥管理服务上。

每个地域的加密云硬盘，都需要通过 256 位数据密钥（DK）进行加密，此数据密钥（DK）具备地域唯一性，即每个地域都有且仅有一个。该密钥受 KMS 提供的密钥管理基础设施的保护，能有效防止未经授权的访问。云硬盘的数据密钥（DK）仅在实例所在的宿主机的内存中使用，不会以明文形式存储在任何持久化介质（即使是云硬盘本身）上。

在创建加密云硬盘并将其挂载到实例后，以下数据都将关联此密钥并进行加密：

- 云硬盘中的静态数据
- 云硬盘和实例间传输的数据（实例操作系统内的数据不加密）
- 通过加密云硬盘创建的快照

对弹性伸缩的影响

如果使用加密的弹性云主机创建弹性伸缩配置，通过伸缩配置创建的云主机，加密方式与原云主机保持一致。

4.12.2 服务的访问控制

IAM 身份认证

统一身份认证（Identity and Access Management，简称 IAM）服务，是提供用户进行权限管理的基础服务，可以帮助您安全的控制云服务和资源的访问及操作权限。具体请参考：统一身份认证 IAM 介绍。

企业项目

企业项目管理（Enterprise Project Management Service，简称“EPS”），为客户提供与企业组织架构和业务管理模型匹配的云治理能力。以面向企业资源管理为出发点，帮助企业以部门、项目等组织架构分级管理和项目业务结构来实现企业在云上的人、物、权管理，提供企业人员管理、项目管理、资源管理等能力。

企业项目服务申请开通后免费使用，您只需要为您帐号中的资源进行付费。

企业项目更多相关内容请参见：企业项目管理。

访问控制

虚拟私有云

虚拟私有云（Virtual Private Cloud，VPC）是您在天翼云上申请的隔离的、私密的网络环境。您能够在一个安全可控、隔离的网络环境中实现云资源的高效管理和利用。虚拟私有云具备丰富的产品特性，使得您可以自定义网络地址、路由表、安全组等。同时，虚拟私有云提供丰富的网络连接，可以满足云上虚拟私有云互访、公网访问、通过专线或者 VPN 与线下 IDC 互通等网络场景。

更多内容请参见虚拟私有云-产品介绍。

安全组

安全组是一种网络安全防护机制，用于防止未经授权的访问和保护计算机网络免受恶意攻击。它是一种虚拟防火墙，用于限制入向和出向网络流量。安全组工作在网络层和传输层，它通过检查数据包的源地址、目标地址、协议

类型和端口号等信息来决定是否允许通过。安全组创建后，用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，即受到这些访问规则的保护。

安全组默认规则请参见安全组-安全组概述。

重点操作短信验证

重点操作验证，是天翼云平台为了保障安全，在用户进行重点操作前增加的二次认证，二次认证通过后系统才能执行用户操作。可避免因误操作引起严重后果，提供更高的安全性。

“重点操作”是可能引起服务运行中断、网络中断、数据丢失等情况的操作。

目前二次认证仅支持短信验证码方式认证。若您开启短信验证能力，在控制台进行重点操作时，系统会向您的手机号（绑定天翼云账号的手机号）发送短信验证码，需输入正确的验证码才能执行该重点操作。从而避免您进行误操作，造成云产品数据的丢失；另一方面，确保操作人身份，避免影响业务运行。

更多内容请参见重点操作短信验证。

4.12.3 监控安全风险

您可以使用天翼云云监控产品来监控您的弹性云主机，云监控支持自动实时监控、告警配置和告警通知，让您更好地掌握弹性云主机的运行状态和各项性能指标。

指标监控

对弹性云主机的各项运行指标进行实时监控。还可以查询历史的监控指标情况。

监控告警

通过配置告警规则，在指标发生异常的第一时间对您进行提醒。及时发现问题并处理，可以有效保障部署在弹性云主机上的服务持续可用。

更多内容请参见云监控。

4.12.4 故障恢复

您可以通过天翼云云主机备份产品和云硬盘备份产品，对弹性云主机上的数据进行备份，当云主机或主机上的云硬盘出现故障，或者软件造成的数据丢失损坏以及人为错误导致的数据误删时，您可以借助备份功能自助快速恢复数据。

云主机备份

云主机备份（CT-CSBS，Cloud Server Backup Service）提供对弹性云主机的备份保护服务，支持基于云硬盘快照技术的备份服务，并支持利用备份数据恢复弹性云主机数据。通过云主机备份服务，可以在发生数据丢失、系统故障、人为错误或恶意攻击等情况下，还原云主机数据，确保业务的连续性和数据的安全性。

更多内容请参见云主机备份。

云硬盘备份

云硬盘备份（CT-VBS，Volume Backup Service）是针对云主机的系统盘、数据盘提供的备份服务。用户可对存储重要数据的云硬盘进行备份，并在云硬盘故障、用户误删数据、遭到黑客攻击等情况下，使用备份快速恢复数据，最大限度保证用户数据的安全性。

更多内容请参见云硬盘备份。

云主机备份和云硬盘备份的差异

产品名称	备份对象	产品要点说明	场景说明
云主	云主	整机备份，通过存储快照技术，将云主机包含的多	针对需要对整个云主机进行保护，或者希望通过备份创建云主机达到快速复

机备份	机	个云硬盘的数据备份到对象存储。 周期备份、手动备份。 支持恢复原云主机或创建新的云主机。	制业务运行环境的场景，建议使用云主机备份。
云硬盘备份	云硬盘	通过存储快照技术，将云硬盘数据备份到对象存储。 周期备份、手动备份。 支持恢复原云硬盘，和使用备份数据创建新的云硬盘。	针对系统盘没有自定义数据的场景，可以选择使用云硬盘备份对系统盘和数据盘进行备份，保证数据安全的同时，降低备份成本。

4.13 网络

虚拟私有云

虚拟私有云（Virtual Private Cloud，VPC）是您在天翼云上申请的隔离的、私密的网络环境。您能够在一个安全可控、隔离的网络环境中实现云资源的高效管理和利用。虚拟私有云具备丰富的产品特性，使得您可以自定义网络地址、路由表、安全组等。同时，虚拟私有云提供丰富的网络连接，可以满足云上虚拟私有云互访、公网访问、通过专线或者 VPN 与线下 IDC 互通等网络场景。

更多内容请参见虚拟私有云-产品介绍。

子网

云资源（例如云主机、物理机等）必须部署在子网内。您可以在虚拟私有云内创建一个或多个子网，但是子网的网段必须在虚拟私有云网段范围内。同子网内网络默认互通，同 VPC 下不同子网之间默认互通。子网网段创建后无法修改，请合理规划网络。

更多内容请参见子网管理 - 子网的高级配置。

虚拟 IP

虚拟 IP（Virtual IP Address，简称 VIP）是一个从子网中分配的内网 IP 地址，没有分配给真实弹性云主机网卡。虚拟 IP 地址拥有私有 IP 地址同样的网络接入能力，用户也可以像主私网 IP 地址一样通过虚拟 IP 去访问弹性云主机。

您可以通过将虚拟 IP 与主备弹性云主机绑定，根据是否需要访问公网可以为虚拟 IP 绑定一个弹性 IP，配合高可用软件（例如 Keepalived）使用，实现业务的高可用。

注意

如未结合 Keepalived 使用，请谨慎删除与虚拟 IP 绑定的主服务器或者网卡，有可能会导致备服务器或网卡流量不通等现象。

更多内容请参见虚拟 IP 概述-虚拟 IP 概述。

安全组

安全组是一种网络安全防护机制，用于防止未经授权的访问和保护计算机网络免受恶意攻击。它是一种虚拟防火墙，用于限制入向和出向网络流量。安全组工作在网络层和传输层，它通过检查数据包的源地址、目标地址、协议类型和端口号等信息来决定是否允许通过。安全组创建后，用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，即受到这些访问规则的保护。

更多内容请参见安全组-安全组概述。

弹性 IP

弹性 IP（Elastic IP Address，简称 EIP）是可以独立申请的公网 IP 地址，将弹性 IP 地址和子网中关联的各项云资源绑定和解绑，可以实现 VPC 中的云资源通过固定的公网 IP 地址与互联网互通。弹性 IP 产品既包含公网 IP 地址，也可以提供公网带宽服务，与其他带宽产品组合使用，可以达到访问公网与节省成本的目的。

弹性 IP 需与绑定的云资源在同一个区域进行使用，不允许跨区域（Region）使用弹性 IP。

更多内容请参见弹性 IP-产品介绍 。

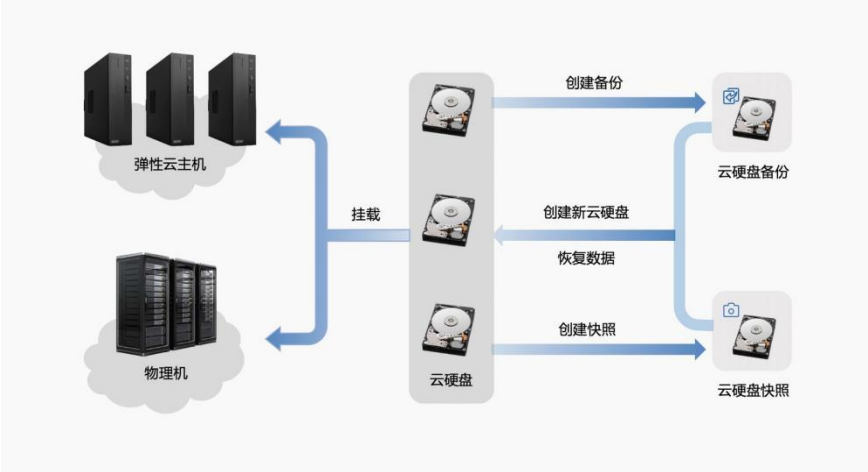
4.14 云硬盘

云硬盘（CT-EVS，Elastic Volume Service）是一种可弹性扩展的块存储设备，可以为弹性云主机和物理机提供高性能、高可靠的块存储服务。天翼云硬盘规格丰富，可满足不同场景的业务需求，适用于分布式文件系统、数据库、开发测试等场景。

用户可以在线操作及管理云硬盘，并可以像使用传统服务器硬盘一样，对挂载到云主机的云硬盘做格式化、创建文件系统等操作。

产品架构

云硬盘可以挂载至弹性云主机和物理机，云硬盘的备份与快照功能可帮助用户恢复数据，创建新的云硬盘，为用户提供了强大的数据保护，提高了数据的可靠性和灵活性。



弹性文件服务、对象存储、云硬盘的区别

天翼云为用户提供了三种数据存储服务，分别是弹性文件服务、对象存储、云硬盘，这三种服务的主要区别如下：

维度	弹性文件服务	对象存储	云硬盘
概念	弹性文件服务提供了一个高度可扩展的文件系统，可在云环境中共享文件数据。具有高可用性、持久性和可靠性。	对象存储具有高度的可扩展性和耐久性，可以存储任意类型的海量数据，并且能够自动处理数据冗余、故障恢复和数据分发。	云硬盘提供了高性能、低延迟、可扩展的块级存储。云硬盘可以被挂载到弹性云主机或物理机上，使其能够持久化地存储数据。
存储方式	弹性文件服务采用文件存储方式。文件存储将数据组织为层次化的目录和文件结构，用户可以通过文件路径和名称来操作文件和目录。	对象存储将数据存储为独立的对象。每个对象由数据本身和与之相关的元数据（例如文件名、文件类型、大小等）组成。	云硬盘采用块存储方式。块存储将数据分为固定大小的块（通常为几 KB 或几 MB），并通过唯一的块地址进行访问。
访问方式	弹性文件服务通过网络共享的方式进行访问。用户可以在需要的弹性云主机实例或容器实例上挂载文件系统，并通过标准的文件系统接口（如 NFS、SMB 等）访问共享的文件系统。	对象存储需要指定桶地址，通过 HTTP 或 HTTPS 等传输协议进行访问。	云硬盘类似于 PC 机的硬盘，无法单独使用，通常通过挂载（Mount）的方式来访问。它可以被挂载到弹性云主机或物理机上，使其在操作系统中可见。
适	如应用程序的配置文件、	如大数据分析，数	如作为弹性云主机或

用场景	日志文件等需要共享的文件数据以及在容器化应用中支持多个容器实例之间的数据共享和同步。	据湖，数据备份和归档等大规模数据存储和分析场景；静态网站托管解决方案存储。	物理机的数据存储介质进行数据存储和持久化；大规模数据处理与分布式计算等高性能计算场景。
容量	弹性文件服务可随业务动态扩展或缩小，容量最高可达 10 PB。	对象存储服务没有容量限制，存储资源可无限扩展。	云硬盘支持按需扩容，最小扩容步长为 1 GB，单个云硬盘可由 10 GB 扩展至 32 TB。
是否支持数据共享	是	是	是
是否支持远程访问	是	是	否

访问方式

天翼云提供如下方式进行云硬盘的配置和管理：

- 控制台：web 化的服务管理平台，即利用管理控制台来配置和管理云硬盘。
- API：天翼云也支持用户通过 API（Application programming interface）方式访问云硬盘，具体操作请参见云硬盘 API 参考。

4.15 地域和可用区

地域

地域（Region）是指物理的数据中心的地理区域。地域从地理位置和网络时延维度划分，同一个 Region 内共享弹性计算、块存储、对象存储、VPC 网络、弹性公网 IP、镜像等公共服务。

天翼云不同地域之间完全隔离，保证不同地域间最大程度的稳定性和容错性。为了降低访问时延、提高下载速度，建议您选择最靠近业务需求的地域。

相关特性

不同地域之间的网络完全隔离，不同地域的云产品默认不能通过内网通信。

如果不同地域之间的云产品之间有通信需求，可以通过公网 IP、VPN 等方式进行通信。

如何选择地域

在天翼云中，资源创建或购买成功后不能更换地域，因此选择地域时，您需要慎重考虑以下几个因素：

1. 地理位置：用户和资源部署地域的距离越近，网络时延越低，访问速度越快。建议您基于业务场景对时延的要求选择地域。

a）中国内地：一般情况下建议选择和您目标用户所在地域最为接近的数据中心，可以进一步提升用户访问速度。如果使用天翼云承载您的全部业务，电信网络可以保证中国内地地域间的快速访问。

b）其他国家及地区：其他国家及地区提供的带宽主要面向非中国内地的用

户。如果您在中国内地，使用这些地域会有较长的访问延迟，不建议您使用。资源价格及资源覆盖：不同区域的资源价格可能有差异，不同地域的产品覆盖可能有差异，请根据您的需求及预算选择合适的地域。

2. 经营性备案：如果您使用弹性云主机实例作为 Web 服务器，您需要完成经营性备案，同时需要在指定的地域购买实例。（各省（或市）通信管理局对经营性备案的审批要求不同，请以当地管理局经营性备案网站公示内容为准。）

可用区

可用区（AZ，Availability Zone）是指在同一地域内，电力和网络互相独立的物理区域。一个 AZ 是一个或多个物理数据中心的集合，具备独立的风火水电，可用区之间距离 100KM 以内，一个 Region 中的多个 AZ 间通过高速光纤相连，以满足用户跨 AZ 构建高可用性系统的需求。

相关特性

- 划分可用区的目标是能够保证可用区间故障相互隔离（大型灾害或者大型电力故障除外），不出现故障扩散，使得用户的业务持续在线服务。通过启动独立可用区内的实例，用户可以保护应用程序不受单一位置故障的影响。
- 处于相同地域不同可用区，但在同一个虚拟私有云下的云产品之间均通过内网互通，可以直接使用内网 IP 访问。

如何选择可用区

在同一地域内，可用区与可用区之间内网互通。各可用区之间可以实现故障隔离，即如果一个可用区出现故障，则不会影响其他可用区的正常运行。是否将实例放在同一可用区内，主要取决于您的应用对容灾能力和网络延时的要求。

- 如果您的应用需要较高的容灾能力，建议您将实例部署在同一地域的不同可用区内。
- 如果您的应用要求实例之间的网络延时较低，建议您将实例创建在同一

可用区内。

地域和可用区的关系

每个地域完全独立，不同地域的可用区完全隔离，但同一个地域内的可用区之间使用低时延链路相连。

地域和可用区之间的关系如图所示：



目前，天翼云已在全球多个地域开放云服务，您可以根据需求选择适合自己的区域和可用区。

相关操作

将实例迁移到其他可用区

一个已经启动的实例是无法更改其可用区的，但是用户可以通过其他方法把实例迁移至其他可用区。迁移过程包括从原始实例创建自定义镜像、使用自定义镜像在新可用区中启动实例以及更新实例的配置。

将镜像复制到其他地域

用户启动实例、查看实例等动作都是区分地域属性的。若用户需要启动实例的镜像在本地域不存在，需要将镜像复制到本地域。

5 购买指南

5.1 计费说明

5.1.1 计费项及其计费方式

一台云主机实例包括计算资源（vCPU 和内存）、镜像、云硬盘等资源，其中涉及计费的资源如下表所示。

资源类型	计费说明	计费方式
计算资源（vCPU 和内存）	以实例规格的形式提供计算资源，包括 vCPU 和内存，收取实例规格费用。 包年包月-计费说明，按量付费-计费说明	包年/包月、按量计费
镜像	镜像当前分为公共镜像、私有镜像、共享镜像与安全产品镜像。 镜像服务-计费说明	根据镜像类型及使用情况决定
云硬盘	针对不同云硬盘类型与容量计费。购买云主机实例时会默认购买 40GiB 的系统盘，用户可以根据需要调整该容量。 云硬盘-计费说明	包年/包月、按量计费
弹性 IP	如有互联网访问需求，您需要购买弹性公网 IP。弹性 IP 可根据带宽、流量进行计费 弹性 IP-计费说明	按带宽包年包月、按带宽按量付费、按实际流量付费。

5.1.2 价格与折扣

包年优惠政策

规格类型	规格族	折扣
通用型	s2	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 4.5 折、5 年 4 折。
	s3、s6、s7、s8e	1 年 6 折、2 年 5.5 折、3 年 3 折、4 年 3 折、5 年 3 折。
	s8r、s8	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 4.5 折、5 年 4 折。
计算型	c3、c6、c7、c8e、c8a	1 年 6 折、2 年 5.5 折、3 年 3 折、4 年 3 折、5 年 3 折。
	c8	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 4.5 折、5 年 4 折。
内存型	m2	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 4.5 折、5 年 4 折。
	m3、m6、m7、m8e、m8a	1 年 6 折、2 年 5.5 折、3 年 3 折、4 年 3 折、5 年 3 折。
	m8	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 4.5 折、5 年 4 折。
增强型	c7ne、c8ne、m8ne	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 4.5 折、5 年 4 折。
本地盘云主机	ip3、ir3、d3	1 年 6 折、2 年 5.5 折、3 年 4 折、4 年 4 折、5 年 4 折。
鲲鹏系列	ks1、kc1、km1	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 5 折、

规格类型	规格族	折扣
云主机		5 年 5 折。
海光系列云主机	hs1、hc1、hm1	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 5 折、5 年 5 折。
	hs3、hc3、hm3	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 5 折、5 年 5 折。
飞腾系列云主机	fs1、fc1、fm1	1 年 8.5 折、2 年 7 折、3 年 5 折、4 年 5 折、5 年 5 折。
GPU 云主机	GPU 云主机价格及优惠政策请参考：GPU 云主机价格总览。	
经济型	e	1 年 3.4 折、2 年 2.5 折、3 年 2.1 折、4 年 2.1 折、5 年 2.1 折。

注意

国际站节点云主机无包年折扣。

通用型云主机

s8e

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	46	0.096
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	17	0.035

s8r

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	46	0.096
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	17	0.035

s8

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	46	0.096
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	17	0.035

s7

国内站节点

于 2023 年 5 月 1 日 0 点完成价格变更，用户可结合实际情况查看相关价格，详细调价方案可咨询客服热线：4008109889。

2023 年 5 月 1 日后采用以下价格：

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	45.69	0.095
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	16.31	0.034

2023 年 5 月 1 日前采用以下价格：

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	49	0.028
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	18	0.088

国际站节点

澳门节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	91.38	0.19
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	32.62	0.07

新加坡节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	80	0.16
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	28	0.06

菲律宾节点

产品规格	vCPU（核）	内存（G）	包月标准价格(元/月）	按需标准价格（元/小时）
s7.small.1	1	1	77	0.134

产品规格	vCPU（核）	内存（G）	包月标准价格(元/月）	按需标准价格（元/小时）
s7.medium.2	1	2	97.26	0.169
s7.medium.4	1	4	137.78	0.239
s7.large.2	2	4	194.52	0.338
s7.large.4	2	8	275.55	0.478
s7.xlarge.2	4	8	389.04	0.675
s7.xlarge.4	4	16	551.1	0.957
s7.2xlarge.2	8	16	778.09	1.351
s7.2xlarge.4	8	32	1102.2	1.914
s7.4xlarge.2	16	32	1556.18	2.702
s7.4xlarge.4	16	64	2204.4	3.827
s7.8xlarge.2	32	64	3112.35	5.403
s7.8xlarge.4	32	128	4408.8	7.654

印尼节点

产品规格	vCPU（核）	内存（G）	包月标准价格(元/月）	按需标准价格（元/小时）
s7.small.1	1	1	92.18	0.16
s7.medium.2	1	2	116.42	0.20
s7.medium.4	1	4	164.93	0.29

产品规格	vCPU（核）	内存（G）	包月标准价格(元/月）	按需标准价格（元/小时）
s7.large.2	2	4	232.86	0.40
s7.large.4	2	8	329.85	0.57
s7.xlarge.2	4	8	465.71	0.81
s7.xlarge.4	4	16	659.71	1.15
s7.2xlarge.2	8	16	931.43	1.62
s7.2xlarge.4	8	32	1319.41	2.29
s7.4xlarge.2	16	32	1862.86	3.23
s7.4xlarge.4	16	64	2638.82	4.58
s7.8xlarge.2	32	64	3725.7	6.47
s7.8xlarge.4	32	128	5277.65	9.16

香港节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	91.38	0.19
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	32.62	0.07

s2/s3/s6

非蒙贵地区

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	45	0.025
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	16	0.08

蒙贵地区

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	35	0.02
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	13	0.065

备注：最低配 1 核 1G 云主机按月订购执行特价 41 元/月，针对一次性付费客户享受一年及以上 8.5 折优惠。

计算型云主机

c8a

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	61	0.127
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

c8e

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	74	0.154
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

c8

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	74	0.154
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

c7

国内站节点

于 2023 年 5 月 1 日 0 点完成价格变更，用户可结合实际情况查看相关价格，详细调价方案可咨询客服热线：4008109889。

2023 年 5 月 1 日后采用以下价格：

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	73	0.152
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

2023 年 5 月 1 日前采用以下价格：

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	96	0.1
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	11	0.07

国际站节点

澳门节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	136.45	0.28
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	26.17	0.05

新加坡节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	124	0.26
产品名称	**包月标准价格（元/核/月）**	**按需标准价格（元/核/小时）**
内存	24	0.05

菲律宾节点

产品规格	vCPU（核）	内存（G）	标准包月价格（元/月）	标准按需价格（元/小时）
c7.large.2	2	4	278.3	0.483

产品规格	vCPU（核）	内存（G）	标准包月价格（元/月）	标准按需价格（元/小时）
c7.large.4	2	8	358.3	0.622
c7.xlarge.2	4	8	556.59	0.966
c7.xlarge.4	4	16	716.6	1.244
c7.2xlarge.2	8	16	1113.17	1.933
c7.2xlarge.4	8	32	1433.2	2.488
c7.3xlarge.2	12	24	1669.77	2.899
c7.3xlarge.4	12	48	2149.79	3.732
c7.4xlarge.2	16	32	2226.36	3.865
c7.4xlarge.4	16	64	2866.39	4.976
c7.6xlarge.2	24	48	3339.53	5.798
c7.6xlarge.4	24	96	4299.58	7.465
c7.8xlarge.2	32	64	4452.71	7.73
c7.8xlarge.4	32	128	5732.78	9.953
c7.12xlarge.2	48	96	5759.42	9.999
c7.12xlarge.4	48	192	7356.1	12.771
c7.16xlarge.2	64	128	8905.42	15.461
c7.16xlarge.4	64	256	11465.55	19.905
c7.24xlarge.2	96	192	11518.85	19.998

产品规格	vCPU（核）	内存（G）	标准包月价格（元/月）	标准按需价格（元/小时）
c7.24xlarge.4	96	384	14712.19	25.542

印尼节点

产品规格	vCPU（核）	内存（G）	标准包月价格（元/月）	标准按需价格（元/小时）
c7.large.2	2	4	309.21	0.54
c7.large.4	2	8	398.11	0.69
c7.xlarge.2	4	8	618.44	1.07
c7.xlarge.4	4	16	796.22	1.38
c7.2xlarge.2	8	16	1236.86	2.15
c7.2xlarge.4	8	32	1592.44	2.76
c7.3xlarge.2	12	24	1861.5	3.23
c7.3xlarge.4	12	48	2377.56	4.13
c7.4xlarge.2	16	32	2473.73	4.29
c7.4xlarge.4	16	64	3184.88	5.53
c7.6xlarge.2	24	48	3710.59	6.44
c7.6xlarge.4	24	96	4755.11	8.26
c7.8xlarge.2	32	64	4947.45	8.59
c7.8xlarge.4	32	128	6369.75	11.06
c7.12xlarge.2	48	96	7446	12.93

产品规格	vCPU（核）	内存（G）	标准包月价格（元/月）	标准按需价格（元/小时）
c7.12xlarge.4	48	192	9510.24	16.51
c7.16xlarge.2	64	128	9894.92	17.18
c7.16xlarge.4	64	256	12680.32	22.01
c7.24xlarge.2	96	192	14891.99	25.85
c7.24xlarge.4	96	384	19020.47	33.02

香港节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	136.45	0.28
产品名称	**包月标准价格（元/核/月）**	**按需标准价格（元/核/小时）**
内存	26.17	0.05

c3/c6

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	85	0.18
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	10	0.02

内存型云主机

m8a

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	42	0.088
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

m8e

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	60	0.125
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

m8

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	60	0.125
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

m7

国内站节点

于 2023 年 5 月 1 日 0 点完成价格变更，用户可结合实际情况查看相关价格，详细调价方案可咨询客服热线：4008109889。

2023 年 5 月 1 日后采用以下价格：

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	58.43	0.124
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.029

2023 年 5 月 1 日前采用以下价格：

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	92	0.077
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	11	0.054

国际站节点

澳门节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	101.62	0.22
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	24.35	0.05

新加坡节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	93	0.2

产品名称	包月标准价格（元/G/月）	按需标准价格（元/G/小时）
内存	22	0.05

菲律宾节点

产品规格	vCPU（核）	内存（G）	标准包月价格（元/月）	标准按需价格（元/小时）
m7.large.8	2	16	441.03	0.766
m7.xlarge.8	4	32	882.05	1.531
m7.2xlarge.8	8	64	1764.11	3.063
m7.3xlarge.8	12	96	2646.16	4.594
m7.4xlarge.8	16	128	3528.23	6.125
m7.6xlarge.8	24	192	5292.33	9.188
m7.8xlarge.8	32	256	7056.44	12.251
m7.12xlarge.8	48	384	10602.11	18.406
m7.16xlarge.8	64	512	14112.88	24.502
m7.24xlarge.8	96	768	21204.22	36.813

印尼节点

产品规格	vCPU	内存	标准价格（元/月）	标准价格（元/小时）
m7.large.8	2	16	490.03	0.85
m7.xlarge.8	4	32	980.06	1.70
m7.2xlarge.8	8	64	1960.12	3.40

产品规格	vCPU	内存	标准价格（元/月）	标准价格（元/小时）
m7.3xlarge.8	12	96	2940.18	5.10
m7.4xlarge.8	16	128	3920.25	6.81
m7.6xlarge.8	24	192	5880.37	10.21
m7.8xlarge.8	32	256	7840.49	13.61
m7.12xlarge.8	48	384	11760.74	20.42
m7.16xlarge.8	64	512	15680.98	27.22
m7.24xlarge.8	96	768	23521.48	40.84

香港节点

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	101.62	0.22
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	24.35	0.05

m2

非蒙贵地区

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	45	0.025
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	16	0.08

蒙贵地区

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	35	0.02
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	13	0.065

m3/m6

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	85	0.18
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	10	0.02

超高 IO 型云主机

ip3

规格名称	vcpu	内存（GB）	本地盘（GB）	本地盘类型	包月标准价格（元/月）	按需标准价格（元/小时）
Ip3.2xlarge.8	8	64	1*3200	NVMe SSD	1927	4
Ip3.4xlarge.8	16	128	2*3200	NVMe SSD	3854	8
Ip3.8xlarge.8	32	256	4*3200	NVMe SSD	7708	16
Ip3.10xlarge.8	40	320	5*3200	NVMe	9635	20

规格名称	vcpu	内存（GB）	本地盘（GB）	本地盘类型	包月标准价格（元/月）	按需标准价格（元/小时）
				SSD		
Ip3.16xlarge.8	64	512	8*3200	NVMe SSD	15416	32

ir3

规格名称	vcpu	内存（GB）	本地盘（GB）	本地盘类型	包月标准价格（元/月）	按需标准价格（元/小时）
ir3.large.4	2	8	1*125	NVMe SSD	307	0.64
ir3.xlarge.4	4	16	1*250	NVMe SSD	614	1.28
ir3.2xlarge.4	8	32	2*250	NVMe SSD	1228	2.56
ir3.4xlarge.4	16	64	2*500	NVMe SSD	2456	5.11
ir3.8xlarge.4	32	128	4*500	NVMe SSD	4912	10.23
ir3.12xlarge.4	48	192	3*1000	NVMe SSD	7368	15.33

磁盘增强型云主机

d3

规格	vcpu	内存（GB）	本地盘（GB）	本地盘类型	包月标准价格（元/月）	按需标准价格（元/小时）
d3.xlarge.4	4	16	1 * 8000	SATA HDD	786	1.83
d3.2xlarge.4	8	32	2* 8000	SATA HDD	1572	3.66
d3.4xlarge.4	16	64	4 * 8000	SATA HDD	3144	7.32
d3.6xlarge.4	24	96	6 * 8000	SATA HDD	4716	10.98
d3.8xlarge.4	32	128	8 * 8000	SATA HDD	6288	14.64
d3.12xlarge.4	48	192	10 * 8000	SATA HDD	9432	21.96
d3.14xlarge.4	56	224	12 * 8000	SATA HDD	11004	25.63

增强型云主机

c8ne

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	81	0.169

产品名称	包月标准价格（元/G/月）	按需标准价格（元/G/小时）

产品名称	包月标准价格（元/G/月）	按需标准价格（元/G/小时）
内存	16	0.033

m8ne

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	66	0.138
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	16	0.033

c7ne

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	75	0.156
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	15	0.031

鲲鹏系列云主机

鲲鹏通用型 ks1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	52	0.1084
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	15	0.0313

鲲鹏计算型 kc1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	104	0.2167
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	15	0.0313

鲲鹏内存型 km1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	104	0.2167
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	15	0.0313

海光系列云主机

海光通用型 hs3

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	45	0.0938
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	16	0.0333

海光通用型 hs1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	45	0.0938

产品名称	包月标准价格（元/G/月）	按需标准价格（元/G/小时）
内存	16	0.0333

海光计算型 hc3

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	89	0.1854
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.0292

海光计算型 hc1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	89	0.1854
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.0292

海光内存型 hm3

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	89	0.1854
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.0292

海光内存型 hm1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	89	0.1854
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	14	0.0292

飞腾系列云主机

飞腾通用型 fs1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	45	0.0938
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	16	0.0333

飞腾计算型 fc1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	85	0.1771
产品名称	**包月标准价格（元/G/月）**	**按需标准价格（元/G/小时）**
内存	10	0.0208

飞腾内存型 fm1

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
vCPU	76	0.1583

产品名称	包月标准价格（元/G/月）	按需标准价格（元/G/小时）
内存	12	0.0250

经济型云主机

经济型 e

产品名称	包月标准价格（元/核/月）	按需标准价格（元/核/小时）
VCPU	24	0.047619

产品名称	包月标准价格（元/G/月）	按需标准价格（元/G/小时）
内存	12	0.023810

5.2 计费模式

5.2.1 按量计费模式

适用场景

按需付费是一种灵活的计费模式，适用于需要灵活调整资源、业务不稳定或资金有限的场景。在选择计费模式时，应结合业务需求和实际情况来做出合适的选择。

收费方式

一种后付费模式，即先使用再付费。

计费说明

- 开通按需计费资源时，您的天翼云账户余额+通用代金券须大于 100 元，并且大于按需计费资源（当前已开通+即将要开通）1 小时费用的总和。
- 按需计费根据资源的结算周期进行结算。一般按需计费的结算周期有小时/日/月等，在达到结算周期时，系统会生成账单，进行扣费。
- 如果账户余额不足，资源将进入保留期，需要在保留期完成缴费，超过保留期，所使用资源将被关停并收回资源。

收费项

- 云主机收费项：CPU、内存、GPU（针对 GPU 型云主机）、本地盘（针对本地盘型云主机）
- 云硬盘收费项：存储容量、存储类型
- 公网带宽收费项：带宽大小

关机计费规则

目前采用的关机规则：

在云主机开通期间，如果用户主动执行了关机操作（含普通关机和节省关机），在云主机关机状态下不会收取 CPU、内存、GPU、本地盘的费用，其他关联资源继续计费。

计划修改为新规则，新规则生效时间以产品公告为准：

- 针对非本地盘型的云主机，两种关机模式的收费情况如下：

关机模式	是否继续收费
普通关机	各项资源均保留且正常计费
节省关机	计算资源（CPU、内存、GPU）被回收且不再收费。其余资源如系统盘、数据盘、带宽、流量包仍正常计费。

- 针对本地盘型云主机，关机期间依旧保留其基础资源（CPU、内存、本地盘），且在关机期间继续对云主机实例采用按量计费：

关机模式	是否继续收费
普通关机	各项资源均保留且正常计费

注意

若选择节省关机模式，当您对此类型的弹性云主机重新开机时，可能会因资源不足导致云主机实例无法正常启动。介于此类情景，请您根据自身业务使用情况，及时调整业务策略。

删除规则

- 删除云主机时，云主机的 CPU、内存、GPU（针对 GPU 型云主机）、本地盘（针对本地盘型云主机）的费用停止计费，其他未删除的关联资源继续计费。
- 云主机删除后，数据不会保留，会立即释放资源。

账户欠费

如用户账户出现欠费，账户一旦充值，系统将会自动优先扣除欠费金额。

欠费影响

当您的账号因自动扣费导致欠费后，按需资源状态变为“已冻结”，系统会关机中断服务，您将无法对处于冻结的按需计费资源执行任何操作，请及时续费使用。系统会为你的资源保留一段时间的保留期（不超过 15 天），保留期到期后，若您仍未续费和支付账户欠款，那么云主机（计算资源，包括 vCPU 和内存）、云硬盘和弹性公网 IP 都将被释放，数据无法恢复。

续费后，若实例停机前为运行中，系统会尝试自动重启。如果偶尔出现自动重开机失败的情况，您也可以手动启动实例。若实例停机前为关机，系统将不会自动重启，保持关机。

提醒/通知规则

- 提醒及通知方式：邮件、短信、站内信。
- 充值成功通知：当用户充值成功后，会发送 1 次充值成功通知。
- 余额不足通知：当用户账户余额不足 100 元，或不足以支付当前所有按量资源 1 天费用时，会发送 1 次余额不足提醒。
- 账户欠费通知：当用户欠费时，会向用户发送 1 次欠费提醒。
- 资源销毁通知：当用户的云主机销毁后，会向用户发送 1 次销毁通知。

5.2.2 包年包月计费模式

适用场景

包年包月是一种计费模式，适用于多种场景，尤其是需要稳定资源并长期使用的情况：

1. 长期稳定使用

常驻服务： 适用于需要长期稳定运行的服务，如网站托管、数据库服务器、企业应用等。

2. 成本可控和节约

成本控制： 对于需要长期使用的资源，包年包月可以降低长期成本，相比按需付费更为经济。

预算规划： 有助于企业预算规划，避免长期高额的变动费用，提供更稳定的费用支出。

3. 稳定资源需求

资源保障： 对于有稳定资源需求的业务，如特定配置的服务器、存储或计算资源，包年包月提供稳定的资源保障。

约束与限制

包年包月的计费模式虽然有其优势，但也存在一些约束和限制，这些限制可能因服务提供商和具体服务套餐而异：

1. 长期绑定

合约期限： 通常需要按照一定期限购买，可能是一年或更长时间的合约，无法随时更改或取消。

费用一次性支付： 一次性支付包年包月费用，无法根据实际使用情况灵活调整费用。

2. 非弹性和限制

固定资源： 购买后资源固定，无法根据实际需求灵活调整，可能导致资源过剩或不足。

限制升级： 一些服务可能在包年包月期间限制资源的升级，只能在合同到期后进行。

3. 限制服务范围

部分服务限制： 某些服务或特性可能无法包含在包年包月套餐内，需要额外付费或选择其他计费模式。

4. 资费约束

费用固定性： 购买后的费用是固定的，无法根据实际使用情况进行变动或调整。

收费方式

一种预付费模式，即先付费再使用。

按月计费，以自然月为计费单位。

收费项

- 云主机收费项：CPU、内存
- 云硬盘收费项：存储容量、存储类型

- 公网带宽收费项：带宽大小

升级规则

云主机升级时间不满整月的，升级后配置按当月实际发生天数计费。

续订规则

参照弹性云主机-购买指南-续费说明-规则说明。

退订规则

参照弹性云主机-购买指南-退费说明-规则说明。

提醒/通知规则

- 提醒及通知方式：邮件、短信、站内信。
- 充值成功通知：当用户充值成功后，会发送 1 次充值成功通知。
- 资源到期通知：云主机到期前 7 天、3 天以及到期当天，会分别发送到期提醒。
- 资源释放通知：云主机到期后 3 天、7 天，会分别发送释放提醒。
- 资源销毁通知：当用户的云主机销毁后，会向用户发送 1 次销毁通知。

5.3 变更配置

5.3.1 变更概述

云主机变配

每个云主机规格已预定义 vCPU 和内存配置，修改云主机规格时，您需要选择目标实例规格，不能单独修改 vCPU 或内存，具体可修改的实例规格以变配页面显示为准。

如需了解云主机变配规则，请参考：支持变配的实例规格

如需了解云主机变配操作，请参考：变更规格

注意
修改云主机规格将产生变更订单，请于 5 天内完成支付，否则变配操作失败，云主机的数据将被清空。

公网带宽变配

如需修改公网带宽配置，请参考：修改带宽

云硬盘变配

如需修改云硬盘配置，请参考：云硬盘变配

5.3.2 支持变配的实例规格

- X86 与 ARM 云主机不可互相变配。
- X86 下海光云主机不可与其他云主机互相变配。
- ARM 计算架构中，基于不同厂商芯片的云主机不可互相变配。例如，鲲鹏云主机不可与飞腾云主机互相变配。
- GPU 云主机和非 GPU 云主机不可互相变配。
- 只允许往同代或更高代云主机变配，反之不可以。例如，s3 可变配为 s6、c6 或 m6，反之不可以。
- 同代系同类型之间可以升配亦可以降配。

5.3.3 云主机冷变配规则说明

冷变配，即关机变配，是指云主机在关机状态下进行变更规格的操作。冷变配规则如下：

- 支持冷升配，可以变更为更高代系或同代系更大规格。例如，s3.large.2 可变配为 s6.large.2。
- 支持冷降配，可以变更为同代系较小规格。例如：s6.4xlarge.2 可变配为 s6.xlarge.2。
- 支持跨 S/C/M 类型冷变配。例如：c3.large.2 可变配为 m3.large.8。
- X86 与 ARM 云主机不可互相冷变配。
- X86 下海光云主机不可与其他云主机互相冷变配。
- 通用云主机、本地盘云主机、GPU 云主机、网络增强型云主机之间不能相互冷变配。
- ARM 计算架构中，基于不同厂商芯片的云主机不可互相冷变配。例如，鲲鹏云主机不可与飞腾云主机互相变配。

注意

- 在变配操作中，当宿主机余量不足时，可能会出现云主机漂移至其他宿主机的情况。
- 部分资源池由于版本问题可能不支持跨代系或跨 S/C/M 类型变配，如遇相关问题可联系天翼云客服。

4.3.4 云主机热变配规则说明

热变配，即开机变配，是指云主机在运行中状态下进行变更规格的操作。热变配规则如下：

- 仅支持热升配，不支持热降配。
- 仅支持 vCPU 在 32C 及以上的规格进行热升配。
- 仅支持 8 代机/7 代机/6 代机/3 代机进行同代热升配，即变更为同代系更大规格。例如，支持 s3.8xlarge.2 热变配为 s3.8xlarge.4，不支持 s3.8xlarge.2 热变配为 s6.8xlarge.4。

- 仅支持 S 类型之间、C/M 类型之间进行同类型热升配，不支持跨类型热升配。例如，支持 c3.8xlarge.2 热变配为 c3.8xlarge.4，支持 c3.8xlarge.2 热变配为 m3.8xlarge.8，不支持 s6.8xlarge.2 热变配为 c6.8xlarge.4。

  支持热变配的镜像列表：

系统	版本
CentOS	CentOS 7.2 64 位
	CentOS 7.3 64 位
	CentOS 7.4 64 位
	CentOS 7.5 64 位
	CentOS 7.6 64 位
	CentOS 7.7 64 位
	CentOS 7.8 64 位
	CentOS 8.0 64 位

系统	版本
	CentOS 8.1 64 位
	CentOS 8.2 64 位
CTyunOS	CTyunOS 2.0.1-21.06.4 64 位
	CTyunOS 2.0.1-21.06.4 64 位 ARM 版
	CTyunOS 3-23.01 64 位

- 存量云主机，即在热变配功能上线前已开通的云主机，不支持热变配。
- 已加入云主机组的云主机不支持热变配。

注意

- 热变配过程约在 15~60 秒，并且可能对业务有一定影响，例如可能造成读业务性能下降约 10%，持续时间不超过 5 秒。规格跨度越大变配耗时越久、业务影响越大，请您根据自身业务情况谨慎操作。
- 请勿连续进行热变配操作，建议两次热变配操作至少间隔五分钟。
- 部分资源池由于版本问题可能不支持热变配，如遇相关问题可联系天翼云客服。

5.4 续费说明

5.4.1 规则说明

- 只有通过实名认证的客户，才可以执行续订操作。
- 按需资源、包年/包月转按需（已完成转按需或正在进行转按需）的资源不可续订。
- 未完成订单中的资源不允许续订，如开通中的资源、规格变更中的资源、退订中的资源。
- 已退订或释放的资源不可续费。
- 若资源到期后续费，续费周期自资源续订解冻开始，计算新的服务有效期，按照新的服务有效期计算费用。例如，客户资源 2020 年 9 月 30 号到期，10 月 11 号续订 1 个月，那么资源新的服务开始时间为 10 月 11 号，到期时间为 11 月 10 号。相关费用自 10 月 11 号开始计算。
- 成套订购的套餐类组合产品，续订时套餐内产品整体续订；非成套订购、具有挂载关系的资源，续订时需要对单资源分别续订。
- 续费后，若实例停机前为运行中，系统会尝试自动重启。如果偶尔出现自动重开机失败的情况，您也可以手动启动实例。若实例停机前为关机，系统将不会自动重启，保持关机。

5.4.2 手动续订

续订规则

用户随时可以手动续订包月包年且未退订、未释放的资源，延长相关资源的使用时间。

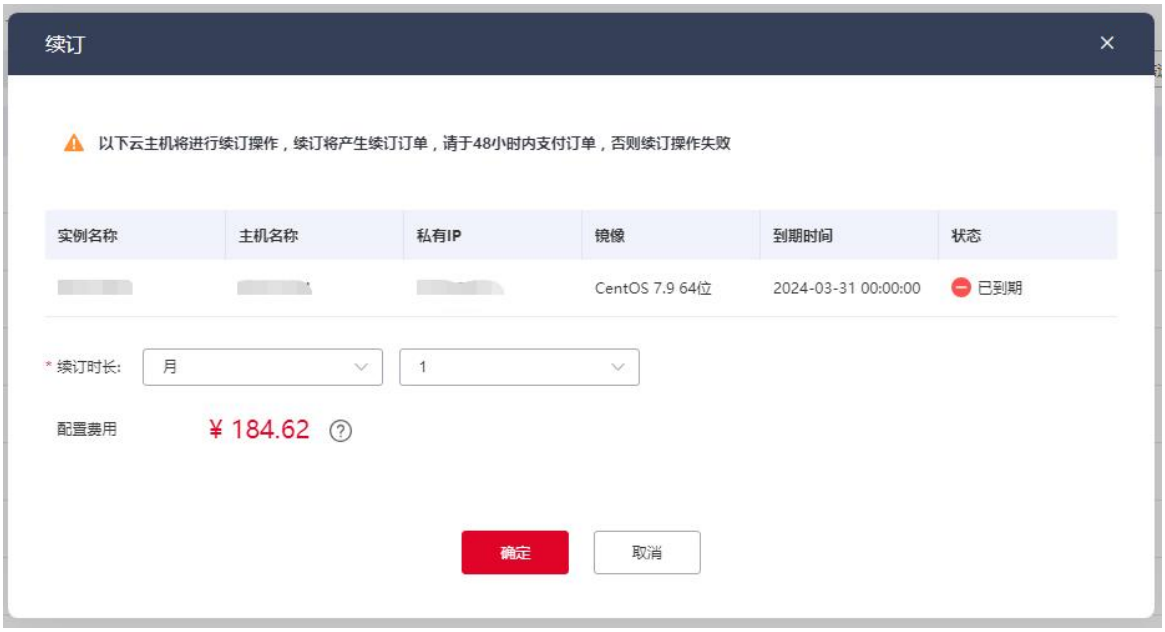
手动续订操作说明

手动续订方式 1

控制台资源管理界面找到所需续订的资源，点击“操作”下面的更多，点击“续费”，在弹窗中调整所需续订周期后，提交续订订单。

当续订周期达到 1 年或以上时，续订单将可享受包年折扣，续订金额显示折后价。

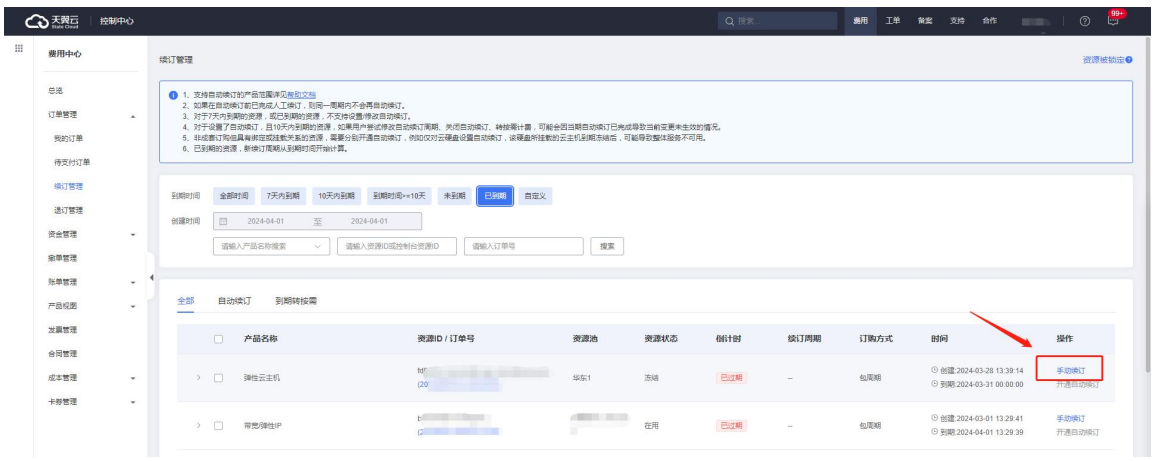


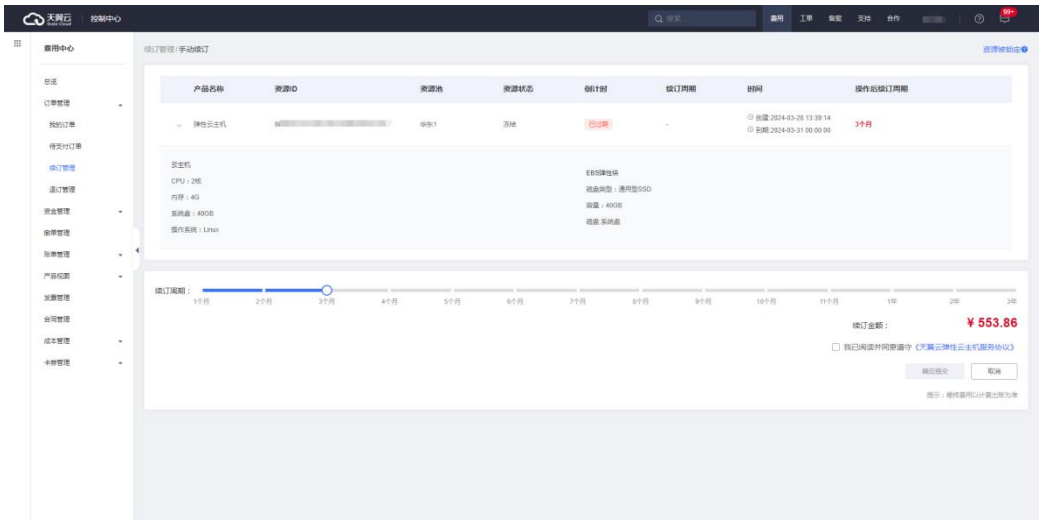


手动续订方式 2

控制台费用中心续订管理页面找到所需续订的资源，点击“手动续订”，根据使用需要调整续订周期后，提交续订订单。

当续订周期达到 1 年或以上时，续订单将可享受包年折扣，实际价格以下单价格为准。





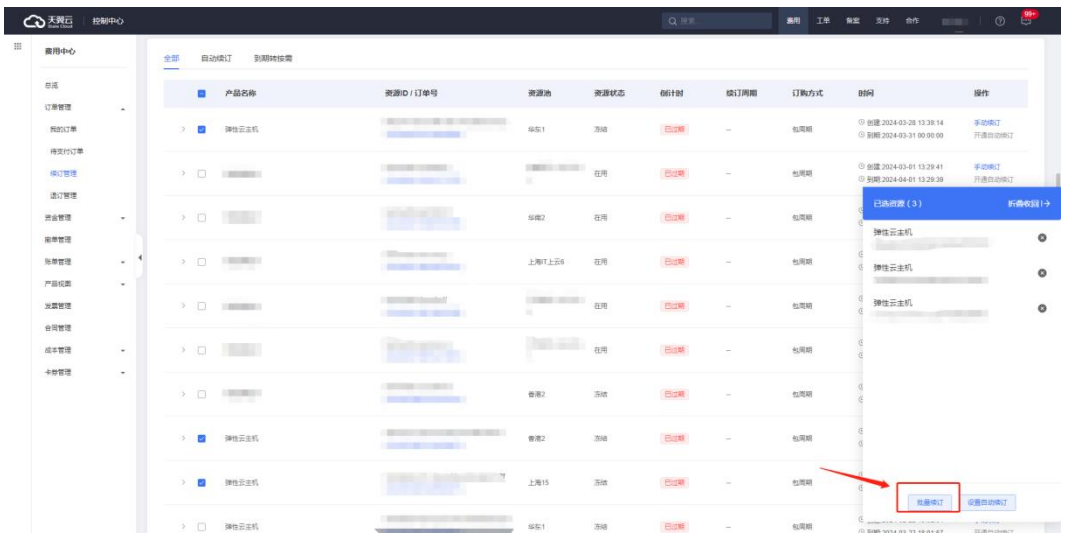
手动批量续订

用户可以一次性手动续订多个包月包年且未退订、未释放资源，最多一次性续订 50 个资源。

批量续订方式：进入控制台产品中心续订管理页面，通过筛选资源到期时间、筛选产品类型等，找到并勾选所需续订的多个资源，点击“批量续订”。

下单前如需确认资源详情，可点击右侧的下拉键查看资源 ID、配置等。

批量续订的多个资源续订周期相同，如需对不同资源续订不同周期，需要分别下不同的续订订单。



5.4.3 自动续订

为避免由于未及时对资源采取续订操作，资源被到期冻结或超期释放，客户购买包月包年产品后，可设置开通自动续订。开通自动续订后，系统将在资源到期前自动续订，无需客户再手动操作。

适用范围

- 自动续订仅针对采用包月、包年计费模式的资源。
- 已到期资源不支持设置/修改自动续订。
- 自动续订仅适于　付费用户　，不适用于非付费用户。
- 目前支持设置自动续订的产品有：弹性云主机、GPU 云主机、物理机、云桌面、云硬盘、对象存储经典版、云主机备份、专属云计算独享型、专属云存储独享型、弹性 IP、共享带宽、天翼云 SD-WAN、云间高速、关系数据库 MySQL 版、关系数据库 Postgre SQL 版、分布式关系数据库、分布式缓存服务 Redis 版、分布式缓存服务 Memcache、文档数据库服务、云 HBASE 数据库、分布式消息服务 RocketMQ、分布式消息服务 RabbitMQ、分布式消息服务 Kafka、Web 应用防火墙、服务器安全卫士、域名无忧、DDos 高防 IP、云解析、登录保护、网站安全监测、内容安全。
- 单次最多支持 20 个资源实例批量续订。

开通、变更、关闭自动续订

- 用户在续订管理页可开通自动续订功能，变更自动续约周期，或关闭自动续订。
- 不关闭自动续订的情况下，只要预付费账户余额充足，或为后付费客户，系统将持续按设定的周期自动续订下去。
- 预付费用户可在官网自主控制自动续订功能的开通、变更、关闭。后付费用户需要客户经理协助开启自动续订权限后才可以自主管理。

自动续订周期

- 包月产品默认自动续订周期为 1 个月，包年产品默认自动续订周期为 1 年，用户可按需调整自动续订周期。

自动续订价格

- 自动续订下单扣费时按当时的标准价自动续订，续订 1 年或以上可享受包年折扣。
- 0 元、秒杀等特价促销活动产品订购后，自动续订下单扣费时将恢复标准价。
- 预付费用户暂不支持代金券支付，仅支持余额支付，用户需确保账户余额充足。

自动续订扣费规则

- 支付方式及支付时间：将在资源到期前 10 天或前 7 天进行自动续订下单及扣费。
- 自动续订订单出账后不可取消。客户如有问题，可发起退订，自动续订订单的退订与退订规则保持一致，退订的同时，该资源的自动续订自动关闭。

自动续订和手动续订的关系

- 在 7 天或更短时间内到期的资源，或已到期资源，需手动续订，无法设置自动续订。
- 开通自动续订功能后，也可以进行手动续订。在自动续订扣费日前进行手动续订，系统将按照手动续订后的到期日期，重新计算下一次自动续订的下单时间。

自动续订操作流程

1、开通自动续订

步骤 1 进入“续订管理”页面。

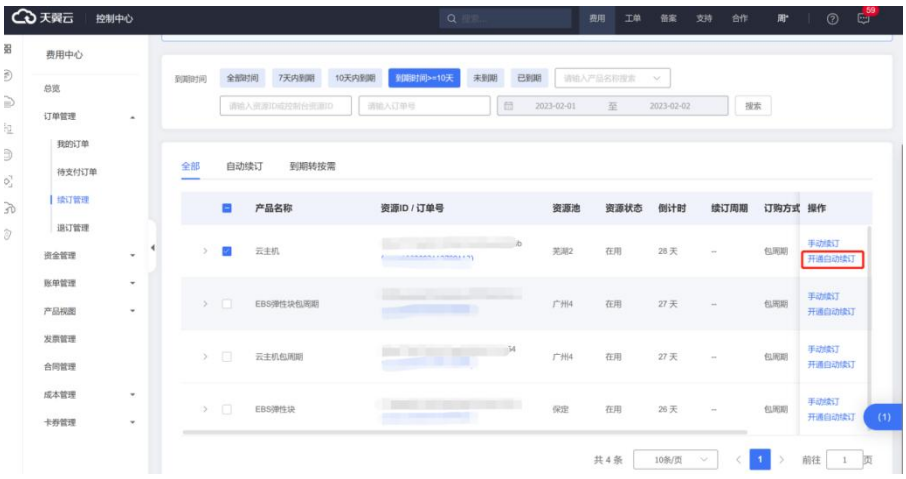
步骤 2 设置查询条件。

可综合利用到期时间、产品类型、是否开通自动续订查询资源。

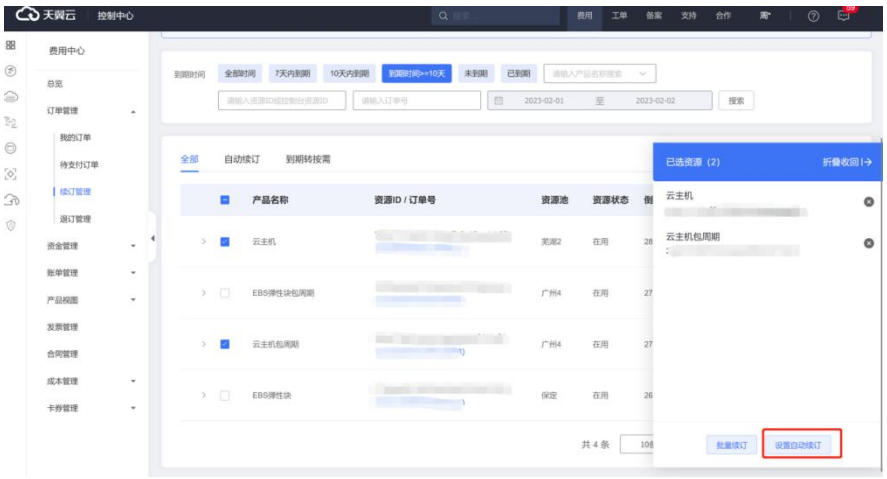
由于自动续订两次下单时间为到期前 10 天和前 7 天，建议您选择“到期时间≥10 天”，选择未开通自动续订功能的云主机、云硬盘、带宽、RDS 资源开通自动续订。

步骤 3 单个自动续订与批量自动续订可使用不同的操作方式：

- 单个自动续订：在资源页面找到待续订的资源，单击操作列的“开通自动续订”。



- 批量自动续订：在资源页面勾选需要续订的资源，单击资源列表下方的“设置自动续订”。



步骤 4 设置“自动续订周期”，仔细阅读《天翼云自动续订服务协议》，如果同意全部约定，则勾选“我已阅读并同意遵守《天翼云自动续订服务协议》的约定”，单击“确定提交”。

说明

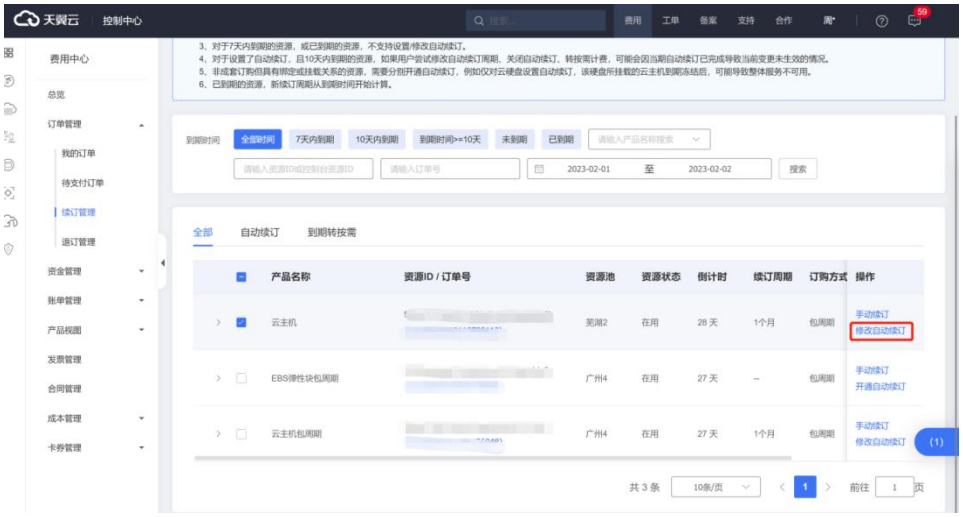
- 开通自动续订功能并不会直接下单扣费，系统将在资源到期前第 10 天、第 7 天下单扣费。
- 自动续订页面的续订金额目的在于提醒用户，预付费用户的账户余额需至少大于等于此处续订金额（如案例中为 145 元）才可自动续订成功，后付费用户则需了解账单（如案例中为 3 月出账）将增加一笔资源续订费用（如案例中为 145 元）。
- 批量开通自动续订，操作后续订周期相同，用户可随时对其中任意资源的自动续订周期进行调整。

2、修改自动续订周期

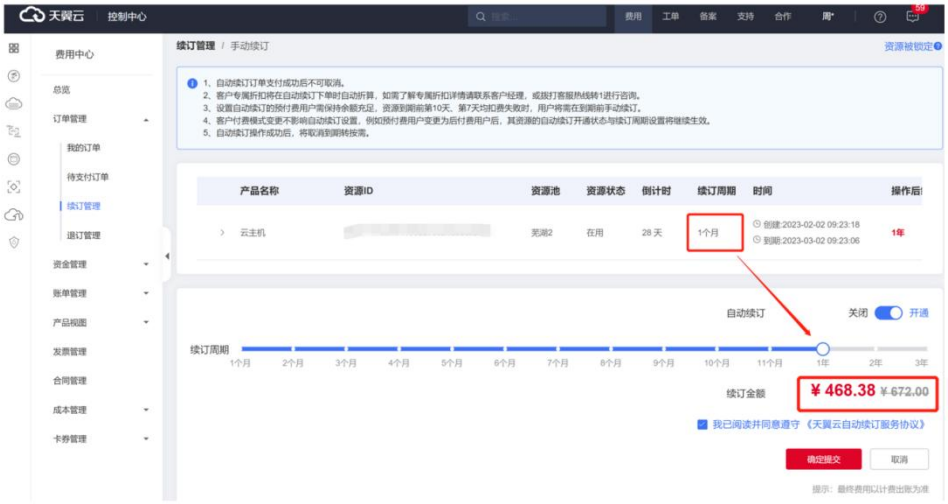
（1）修改单个资源的自动续订周期

步骤 1 进入“续订管理”页面。

步骤 2 在资源页面找到待修改自动续订的资源，单击操作列的“修改自动续订”。



步骤 3 拖动“续订周期”可修改自动续订周期，当自动续订周期达 1 年或以上时，将可享受包年折扣。

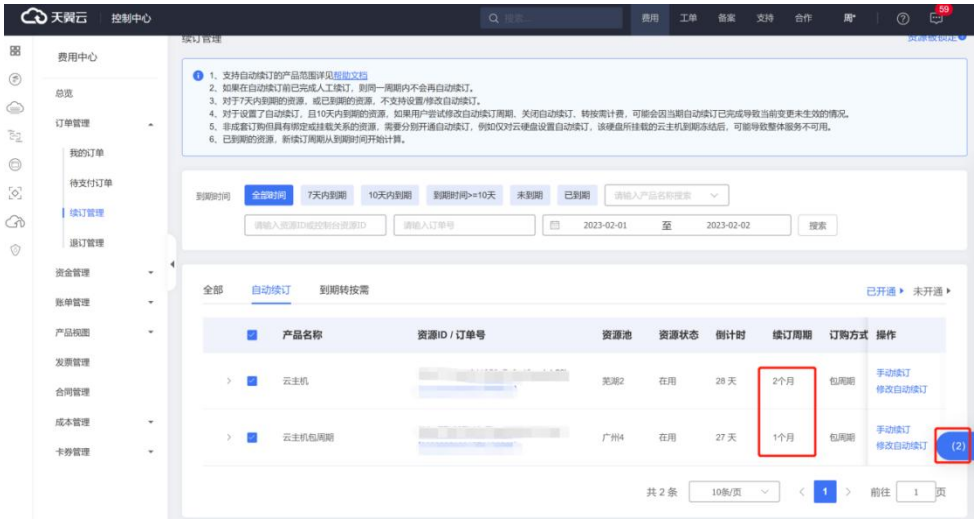


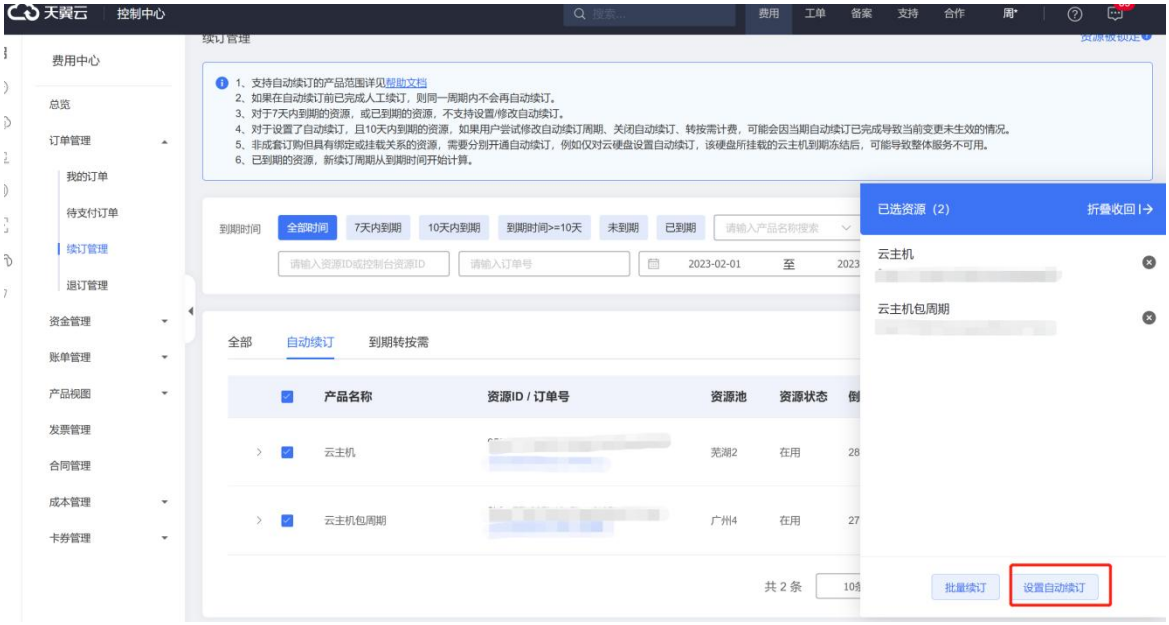
步骤 4 勾选“我已阅读并同意遵守《天翼云自动续订服务协议》的约定”，点击“确定提交”。

（2）修改多个资源的自动续订周期

步骤 1 进入“续订管理”页面。

步骤 2 在“自动续订”列表中查看已开通自动续订功能的资源，选择需要批量修改的资源，点击列表右侧的蓝框数字，确认批量资源信息，确认后点击下方的“设置自动续订”进行批量修改。





步骤 3 拖动“续订周期”修改这些资源的自动续订周期，如果这些资源的原自动续订周期不同，批量修改后使自动续订周期统一。



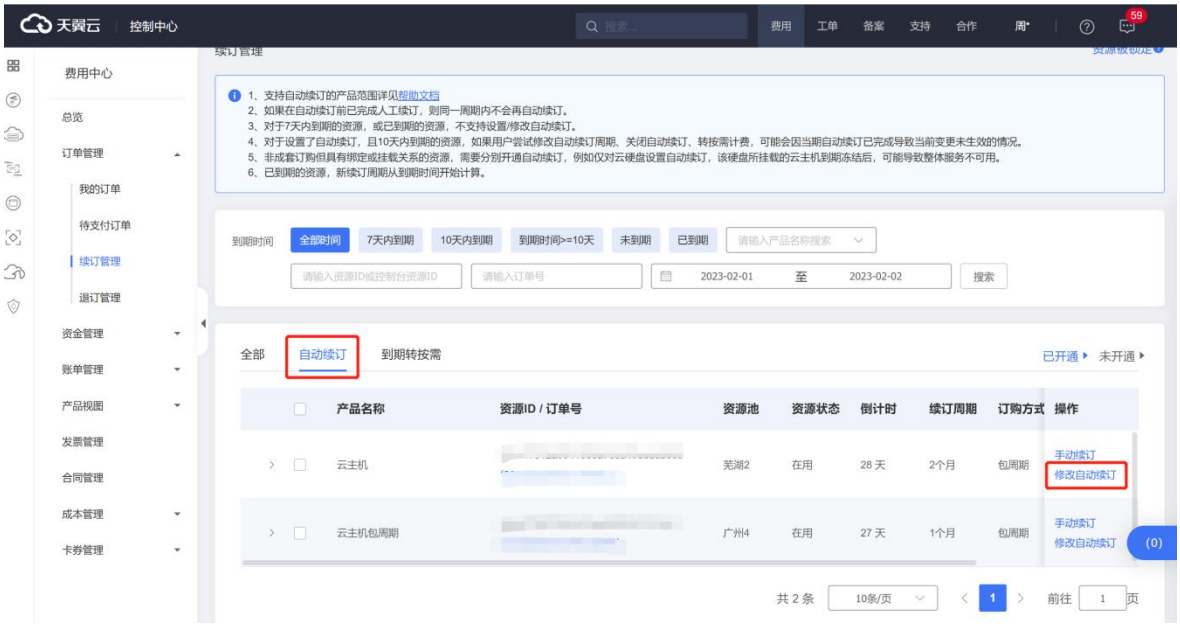
步骤 4 勾选“我已阅读并同意遵守《天翼云自动续订服务协议》的约定”，点击“确定提交”。

3、关闭自动续订

（1）关闭单个资源的自动续订功能

步骤 1 进入“续订管理”页面。

步骤 2 在“自动续订“页面页面找到待修改自动续订的资源，单击操作列的“修改自动续订”。



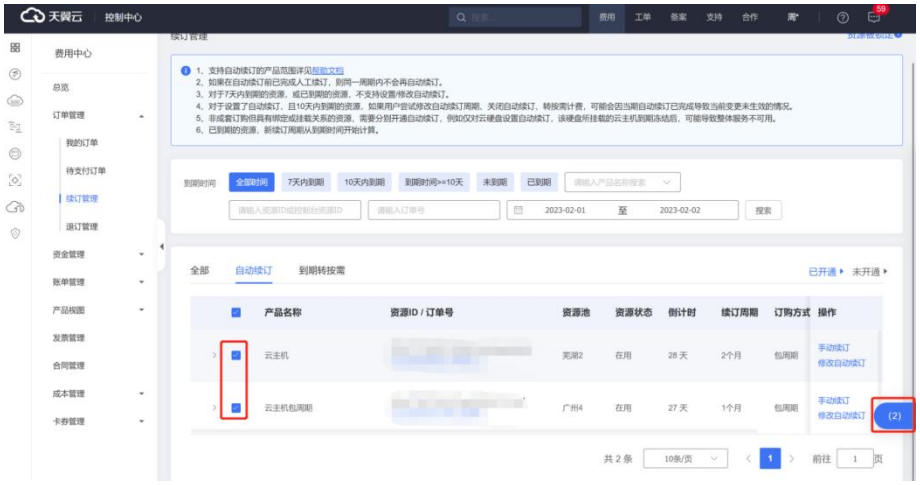
步骤 3 点击“自动续订”后方的关闭/开通按钮，单击“确定提交”。

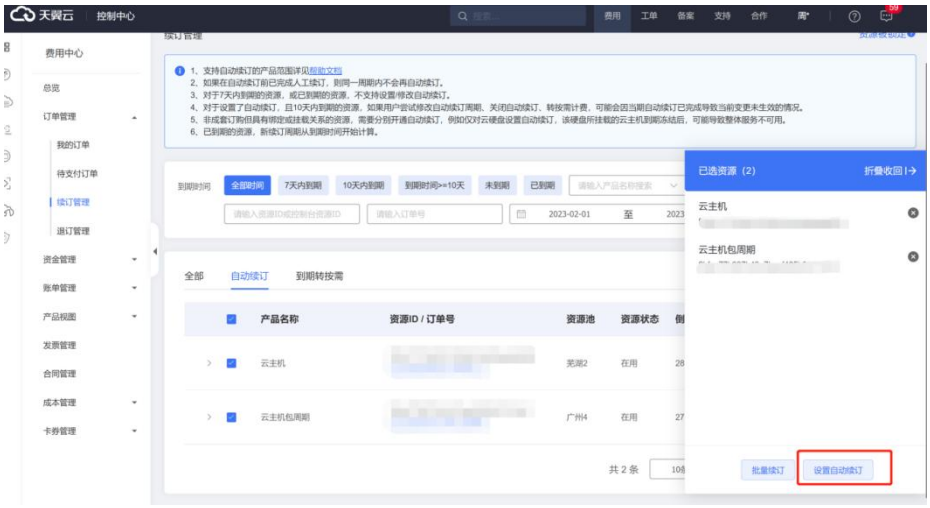


（2）关闭多个资源的自动续订功能

步骤 1 进入“续订管理”页面。

步骤 2 在“自动续订”列表中查看已开通自动续订功能的资源，选择需要批量修改的资源，点击列表右侧的蓝框数字，确认批量资源信息，确认后点击下方的“设置自动续订”进行批量修改。





步骤 3 点击“自动续订”后方的关闭/开通按钮，单击“确定提交”。



5.5 退费说明

5.5.1 规则说明

退订规则说明

客户（天翼云用户）可根据需要，在符合天翼云退订规则的前提下，灵活退订资源。目前退订包含七天无理由全额退订，非七天无理由退订，不可退订，以及其他退订。

- 七天无理由全额退订，用户使用新购资源（含该笔新购资源订单的续订资源）七天内（含第七天），可随时通过系统完成全额退订。
  - 发起退订操作的账号（“退订账号”）七天无理由全额退订次数不超过 3 次（每账号享有 3 次七天无理由全额退订次数）；
  - 同一用户累计使用的七天无理由全额退订次数不超过 15 次。其中，同一用户是指：根据不同天翼云账号在注册、登录、使用中的关联信息，关联信息相同天翼云判断其实际为同一用户。关联信息举例：同一名称、同一邮箱、同一负责人证件、同一手机号、同一设备、同一 IP 地址等（包括已注销的账号）。 客户同意天翼云使用上述信息核查同一用户情况；
  - 七天无理由退订标准（含退订次数）将根据业务发展适时调整，具体结果以退订时官网展示和系统显示为准；
  - 促销活动中说明“不支持七天无理由退订”的云服务，无法申请七天无理由退订；
  - 当同时退订多个资源时，且超过 3 次免费退订次数，系统按照单资源实付金额从高到低依次退订，免费次数优先覆盖高价值资源。

尽管有上述规则，客户不得利用退订规则频繁订购并退订服务，恶意占用天翼云及其他用户资源。如天翼云有合理理由怀疑客户存在频繁退订恶意占用天翼云及其他用户资源的，则天翼云有权取消该客户七天无理由全额退订的权利，并根据《中国电信天翼云用户协议》及网站相关规则和相关服务协议约定，采取相应措施直至终止服务，并追究客户的违约责任。

- 非七天无理由退订，指使用时长已超过七天以上的用户新购资源（含续订资源），属于非七天无理由退订。非七天无理由退订，不限制退订次数，但退订需要收取相应的使用费用和退订手续费，且不退还代金券及优惠券，但符合下文“其他退订”情形的除外。
- 不可退订，指天翼云部分资源因技术属性、营销政策、活动优惠、特定使用群体、商务协议等各类原因，此类资源暂不支持退订。不支持退订的场景范围如下：
  - 按需套餐包产品；

- 含硬件类产品和人工服务类产品；
- 未完工的订单；
- 云服务参加特殊赠送活动，明确不可退订的（如促销套餐、促销优惠券、促销优惠券礼包、优惠券计划、抽奖活动、推荐送活动等）具体以特殊赠送活动规则为准；
- 明确不允许退订的产品或服务。

- 其他退订。指因天翼云原因导致的，如创建资源失败或系统 BUG，产品迭代等情况导致的用户退订。其他退订不限制退订次数，无需用户承担退订费用。

注意事项：

- 退订前的资源确认：执行退订操作前，请确保退订的资源数据已完成备份或迁移，退订成功后的资源将被完全删除，且不可恢复，请谨慎操作。退订可能会影响其他存在的关联业务，请用户做好评估；
- 退订未生效的续订资源：用户在使用的资源成功完成续订，在续订订单未到生效日期前，用户退订该资源，在使用的资源和未生效的续订资源将被全部退订；
- 退订具有绑定关系的资源群：用户该资源与其他资源具有绑定关系，若退订该资源，与该资源具有绑定关系的其他资源仍将继续产生费用；退订将解除原资源绑定关系进而可能影响业务使用；
- 参与活动购买的云产品，如若本退订规则与活动规则冲突，以活动规则为准；活动中说明“不支持退订”的云服务资源不支持退订。

表 1 规则说明

退订分类	退订次数	收取手续费	收取已消费金额	返还代金券	返还优惠券
七天无理由全额退订	同一用户（含已注销账号）累计 15 次。每个账号享有 3 次。	否	否	是	否

退订分类	退订次数	收取手续费	收取已消费金额	返还代金券	返还优惠券
非七天无理由退订	不限次数	是	是	否	否
其他退订	不限次数	否	否	是	否

说明

1. 代金券指以券面金额代替现金用来支付订单费用的一类使用券。
2. 优惠券指折扣券和满减券，即购买某种云产品可以打折或者满足支付一定现金金额后可以抵扣部分订单费用的一类使用券。
3. 若用户使用代金券或优惠券订购资源并仅退订其中部分资源，退订发起页面提示退订完成后不返还该订单已使用的代金券和优惠券，并由用户确认是否依然执行退订。
4. 券类如果存在有效使用期，系统将不返还超出有效期的券类。

七天无理由全额退订规则说明

- 用户使用新购资源（含该笔新购资源订单的续订资源）七天内（含第七天），可随时通过系统完成全额退订（全额退还实付金额，已使用的代金券全额退还）。同一用户（含已注销账号）累计可享有 15 次七天无理由全额退订。每个账号享有 3 次七天无理由全额退订。
- 促销活动中说明“不支持七天无理由退订”的云服务，无法申请七天无理由退订。
- 七天无理由退订次数计算规则
  - 支持一个订单下的退订一个或多个资源，每退订一个资源则计算一次退订次数；
  - 支持退订整个订单，退订次数以资源数量为准，每退订一个资源计算一次退订次数；
  - 支持多订单退订资源，每退订一个资源计算一次退订次数；
  - 对于明确需要用户订购多个不同资源才能提供某类服务或功能，即具备绑定关系的资源群退订，退订其中任一资源，每退订一次，则记为 1 次退订次数；

- 对于成套订购的资源，成套退订时视为 1 次退订次数。

- 示例：
  - 用户 A 一个订单购买 12 台云主机，在开通的 7 天内全部退订，此时七天无理由退订计为 3 次，其余 9 个资源按照非七天无理由退订规则执行；
  - 用户 B 一个订单购买 12 台云主机，在开通的 7 天内全部退订，退订时又勾选了其中一台云主机挂载的额外购买的云硬盘，此时 7 无理由退订计为 3 次，其余 10 个资源按照非七天无理由退订规则执行；
  - 用户 C 订购一套由云主机、磁盘、带宽组成成套资源，退订云主机时，磁盘、带宽同时退订，此时 7 无理由退订计为 1 次。

非七天无理由全额退订规则说明

资源退订时按照原始订单实付价格折算退订金额。例如：原订单支付时 1C1G 月价 100 元/月，用户申请退订的时候 1C1G 已经降价为 50 元/月，则系统仍然按照 100 元实付金额参与退订计算。

预付费业务退款金额=订单实付金额-已消费金额-退订手续费，不退还代金券、优惠券。

- 订单实付金额即用户现金支付金额。
- 已消费金额=订单实付金额*实际使用天数/该订单总天数，实际使用天数=成功退订生效日期-订单开始日期，退订当天算一天使用费。
- 退订手续费与订购年限及实际消费时长有关，具体收取标准如表 2 所示。

表 2 退订手续费收费标准

产品类型	消费时长≤1 年	1 年＜消费时长 ≤2 年	2 年＜消费时长≤3 年	消费时长>3 年
3 年（不含）以上预付费产品	已消费金额 ×20%	已消费金额 ×15%	已消费金额 ×10%	已消费金额 ×10%

产品类型	消费时长≤1年	1年＜消费时长 ≤2年	2年＜消费时长≤3年	消费时长>3年
3年预付费产品	已消费金额 ×15%	已消费金额 ×10%	已消费金额 ×5%	-
2年预付费产品	已消费金额 ×10%	已消费金额 ×5%	-	-
1年预付费产品	已消费金额 ×5%	-	-	-
包月预付费产品	已消费金额 ×5%	-	-	-

注意

扣费比例天翼云将根据业务发展、风险管控、客户服务等情况动态调整，具体结果以官网展示和系统显示为准。

- 若用户订单整个生命周期中出现升级或续订后，再操作退订，系统将记录该资源订购+升级或续订总费用，再分别减去各阶段规格的实际资源使用费，从而计算出应退费用。

其他退订

如因天翼云原因导致用户退订的情况，退订不占用用户每账户的退订次数，不产生费用。

退款金额流向

退款方式分原路返回和退回余额两种情况。

- 订单充值支付（不含对公充值）在 3 个月内的订单，退订后退款金额会原路返回，按照原支付路径实时退还至原付款账户。
- 订单充值支付（不含对公充值）超过 3 个月的订单，退订后退款金额会退回账户余额，客户可根据需要进行提现 。提现操作详见文档余额提现。

5.5.2 退费流程

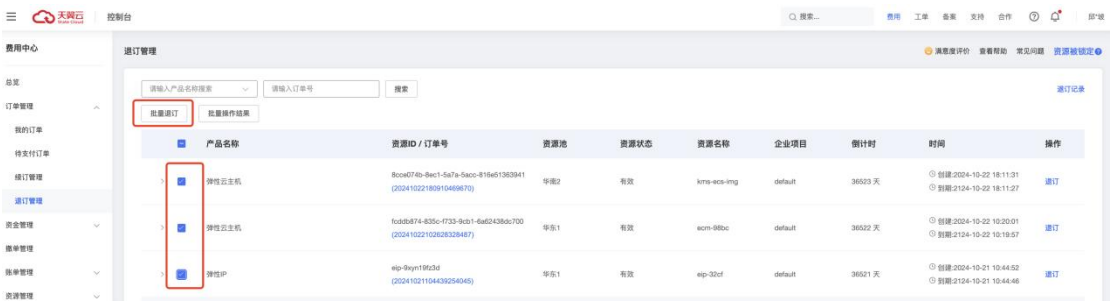
操作流程

通过管理中心退订

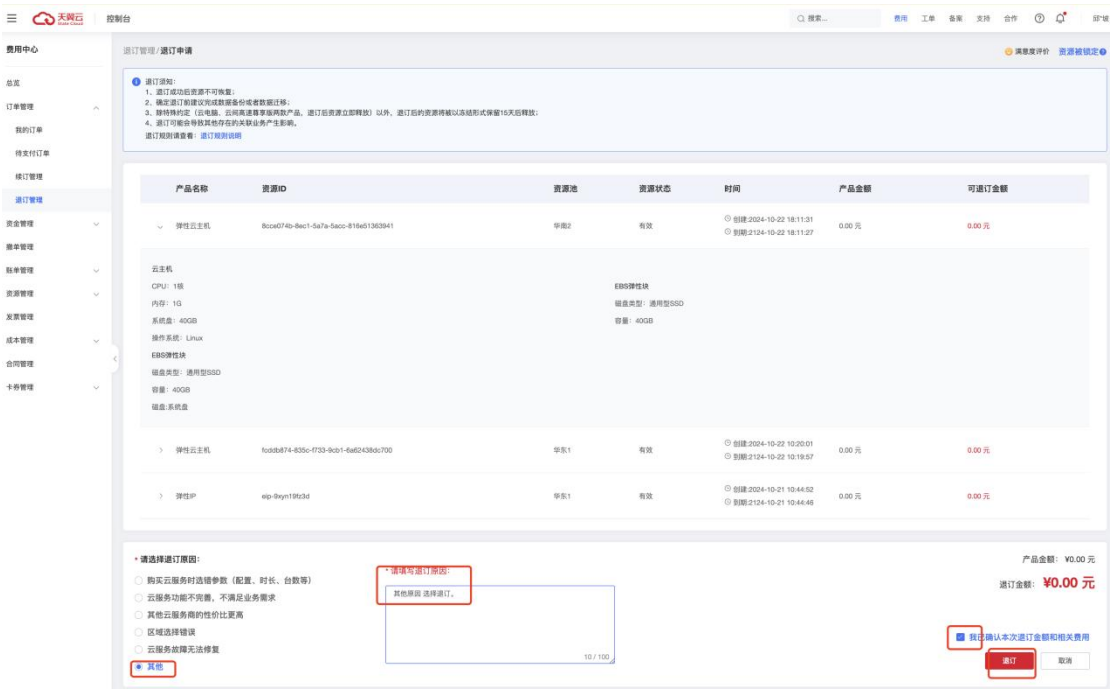
1. 单个退订。登录天翼云官网，进入管理中心-退订管理-查看需退订产品名称 ，点击退订。



2. 批量退订。管理中心-退订管理-勾选多个产品名称，点击批量退订。



3. 退订前仔细阅读退订须知，然后确认退订资源信息，包括产品名称、资源 ID、规格、时间及可退订金额。



4. 信息确认无误后，勾选并填写退款原因-勾选协议-点击退订并再次确认-确认后即刻完成退订。退订后资金退回账户余额，可以通过“余额提现”进行提现，提现详细操作请参考余额提现。

通过产品控制台退订

您可通过产品控制台退订，此处以云主机退订为例进行说明。

1. 登录天翼云，进入 控制中心-云主机控制台。
2. 选择需要退订的资源，点击右侧“更多-退订“，进入退订详情页面。



3. 确认退订信息，信息确认无误后勾选协议，点击退订并再次确认，确认后即刻完成退订。

4.2024 年 4 月 26 日前退订的云主机不在控制台展示会占用相关配额。自 2024 年 4 月 26 日后，包年包月计费模式的云主机实例在退订后可在控制台查看，此时相关计算资源（vCPU、内存、GPU）进入资源保留状态，保留 15 天后自动释放。“包周期退订中”、“包周期已退订”状态的云主机将会正常占用配额但不会继续收取费用，您可以点击“更多--立即释放”提前彻底释放此计算资源，释放完成后将不会占用配额。若您需要对当前状态云主机进行恢复，请提交工单联系相关客服。



5.6 包周期按量互转

概述

目前天翼云支持包周期、按量付费两种计费方式，您可根据您的实际情况选择不同的计费方式。天翼云支持计费方式相互转换，即包周期计费可以转换成按量付费，按量付费可以转换为包周期计费。

按量付费转换为包周期计费

按量转包周期指的是按量计费模式转为包周期计费模式，用户进行按量转包周期操作后，立即生效，原来按量计费资源即转为包周期计费。

1、 在云主机控制台列表页面，点击按量计费实例的“更多”，在下拉菜单中选择“转包周期”。

2、在弹出的转包周期对话框中，选择“确定”按钮。

3、进入支付页面，点击“去支付”完成支付操作。按量转换为包周期计费完成。

包周期转换为按量计费

包周期转按量指的是包周期计费模式转换为按量计费模式，用户进行转按量操作后，在资源到期时生效，计费模式转换为按量计费。

1、在云主机控制台列表页面，点击包周期实例的“更多”，在下拉菜单中选择“到期转按需”。

2、在弹出的对话框中，选择“确定”按钮。进入支付页面，点击“去支付”完成支付操作。包周期转换为按量计费完成。

注意

若距离实例到期时间不足 1 天，无法进行包周期转按量计费操作。

说明

资源到期前可取消“转按需”设置，操作入口在费用中心-订单管理-续订管理下的“到期转按需”页面进行操作。

5.7 成本优化

相比自建数据中心，使用云资源时无须投入硬件、物理环境人力等成本，单位资源成本相对线性，所有资源按需取用，交付便利。除此之外，云资源支持多种付费模式，方便进一步优化成本。

针对使用云资源如何优化成本，我们给出以下建议：

追踪成本

- 从费用账单了解消费情况，追踪成本并确定如何优化；
- 使用标签等功能分类资源，以便统计相应成本。

优化资源

- 监控资源使用情况

- 监控资源利用率，评估当前配置是否过高。例如 CPU、内存、云盘、带宽等资源的利用率。
- 监控闲置的资源，避免浪费。例如升配但未挂载的云盘、未关联的弹性 IP 等。
- 监控资源使用周期。如果长期使用按量付费实例、云盘等资源，考虑以更实惠的方式购买，例如包年包月、资源包等。
- 监控资源生命周期，了解包年包月资源的到期日，及时续费。例如包年包月实例、存储容量单位包等。

- 选择合适的规格

根据业务场景选择最佳性价比的实例规格，并调整合适的数量，在满足业务需求的同时追求高资源利用率，可参考选型最佳实践。

- 组合多种付费模式

不同类型的业务对资源使用周期有不同要求。为每一类业务确定合适的付费模式，灵活组合达到最优效果。如业务负载稳定使用包年包月方式，业务负载动态变化使用按量付费方式。

节约意识

按需取用是云计算最大的一个特点，您可以将成本优化融入到日常工作中，如：定期盘点资源使用情况，明确闲置资源的通知和处置流程；定期和成本相关方（例如财务、研发等团队）评审预算执行情况，改进优化策略；按时续费，提前申请包周期预算，避免到期释放后重新购买部署增加额外成本。

5.8 计费 FAQ

弹性云主机支持哪些计费方式？

提供包年包月、按量购买两种计费方式，您可根据实际情况进行选择。

同一台云主机是否支持两种计费方式？

不可以。同一台云主机只能选择 1 种计费方式，无法同时选择。

按量订购有什么限制？

按量付费方式的订购，要求您的账户有可用余额，余额需要满足以下要求：

- 不少于 100 元。
- 如果您有其他按量付费资源，余额数量应至少满足下一小时的扣费金额。此限制只在部分资源池生效，具体以资源池可见为准（下单时系统会进行校验，不满足的情况下会给予对应提醒）。

场景 1：用户首次注册，需要通过按量付费方式购买资源，因为用户没有已开通的按量付费资源，用户只需要保证余额不少于 100 元即可。

场景 2：用户已经开通了大量按量付费资源，每小时扣费金额 300 元。如果用户还需要继续下单购买按量付费资源，需要余额不少于 300 元。

场景 3：用户已经开通了少量按量付费资源，每小时扣费金额 50 元。如果用户还需要继续下单购买按量付费资源，需要余额不少于 100 元。

注意，余额剩余情况不影响包周期资源的使用。

按量订购的资源欠费，是否会发送提醒？

欠费后将冻结您账号下所有的按量资源，并会发送短信及邮件提醒您。

不同资源池的云产品内网是否可以互通？

默认不互通，但可通过云间高速实现两个资源池的 VPC 的打通。

已购买的弹性云主机是否可以更换资源池？

很抱歉，已有的弹性云主机不支持更换资源池。

操作系统、镜像是否需要单独付费？

目前操作系统、公共镜像及私有镜像无需另外付费。

按量购买时，购买页面上并没有显示价格，如何获知价格？

您可以点击“价格计算器”查看。

已购买的弹性云主是否支持升级？

目前弹性云主机所有资源池支持 CPU、内存、云硬盘、带宽升级。

CPU 和内存升级必须以套餐的形式升级，必须符合天翼云套餐搭配。

云主机到期后忘记续费了，会出现什么后果？

如果您忘记续费，到期当天您的云主机将无法操作。建议您尽快进行手动续费，

否则到期 15 天后数据就会清除。

退订云主机前有哪些限制条件？

如果您需要退订云主机，需要符合天翼云 7 天无理由退款条件。

- 订购时间超过 7 天，以及订购 7 天内但执行过升级、续订操作的主机均不能执行退订操作。
- 由于退订操作会导致资源回收和清理，云主机上的数据将无法恢复，因此，在退订前请您做好数据备份工作。
- 按量订购的云主机不支持退订操作。

包年包月订购的弹性云主机如何退订？

登录天翼云控制台，勾选需要退订的云主机，选择”更多-退订“，在页面中点击确认，将完成退订订单提交，弹出提示信息，并可在用户中心的订单管理页面看到该退订订单，此时订单状态为“待审核”。

待审核通过后，天翼云将回收资源，并返还现金支付的金额。此时，订单状态将变为“工单完成”，在云主机列表里也无法看见该主机。

退订弹性云主机后，退款如何返还？

系统提示退订申请已提交，3 个工作日内进行审核，审核结果将以短信告知，退订成功过后，现金支付部分将实时返还至您的现金帐户中，代金券支付部分不予退还。

申请开具发票需要注意哪些事项？

- 增值税发票仅面向中国大陆客户，港澳台及海外企业客户无法开具此类发票。
- 个人认证账号可以申请增值税普通发票；

  企业认证账号（类型是政府、军队、事业单位、社团团体、其他）可以申请增值税普通发票；

  企业认证账号（类型是企业）可以申请增值税普通发票以及增值税专用发票；
- 实名认证为企业的客户，创建企业抬头的发票模板时，发票抬头必须和实名认证名称保持一致，否则只能通过重新实名认证进行修改；创建个人抬头的发票模板时，发票抬头可以手动设置，支持修改。
- 实名认证为个人的客户，创建发票模板时，发票抬头可以手动设置，支持修改。

- 代金券消费不纳入可开票金额。
- 月产生的按需产品账单在下个月 3 日后才可以索取发票。
- 申请纸质发票， 如金额不足 100 元，快递费用需要自行承担 。申请成功后，纸质发票预计 15 个工作日内寄出。
- 账号每个月可进行 5 次发票申请操作（含退票），超过则无法操作。

成功购买弹性云主机后多久可以申请开具发票？

每月 3 日产生上月账单，账单生成后才可以索取发票索。

具体信息及操作请查看：发票申请。

已到期资源如果续订，续订周期如何计算？

若资源到期后续费，续费周期自资源续订解冻开始，计算新的服务有效期，按照新的服务有效期计算费用。例如，客户资源 2020 年 9 月 30 号到期，10 月 11 号续订 1 个月，那么资源新的服务开始时间为 10 月 11 号，到期时间为 11 月 10 号。相关费用自 10 月 11 号开始计算。

按量购买云主机时账户余额不足是否有提醒？

用户按量购买的云主机账户余额不足时会有短信提醒。另外，当用户余额小于 100 元，也会有短信提醒。

按量购买的弹性云主机关机后还会计费吗？

按量付费的弹性云主机关机后，基础资源（包括 vCPU、内存）不计费，但系统盘仍会收取容量对应的费用。 如有其他绑定的产品，如云硬盘、弹性 IP、带宽等，按各自产品计费方法收费。

我账户里面有钱，但是我无法创建按量付费云主机？

请您检查您的账户金额是否小于 100 元，如果小于 100 元，则无法创建新的云主机。

我一个账户最多可以创建多少按量付费的云主机？

目前账户默认配额是：50 台云主机、200 核 CPU 和 409600GB 内存。

在不超出配额的情况下，您可根据需求创建云主机，如果云主机创建需求超出默认配额，可提交工单申请增加配额。

是否支持自动删除云主机？

不支持，在您不需要的时候请您自行删除。建议整点前几分钟删除，避免时间损

失。

6 快速入门

6.1 注册账号

1. 打开天翼云门户网站，点击“免费注册”；



2. 在注册页面，在注册页面，选择短信注册或账号注册，并填写相应的信息。点击获取验证码，如 1 分钟内手机未收到验证码，请再次点击“获取验证码”按钮。





3. 阅读并勾选协议，点击“注册” 按钮。

4. 注册成功后，可到邮箱激活您的账号，即可体验天翼云。

5. 如需实名认证，请参考会员服务-实名认证。

6.2 创建弹性云主机

购买流程

步骤 1：基本配置

步骤 2：网络配置

步骤 3：高级配置

步骤 4：确认配置

步骤 1：基本配置

准备工作

1. 创建账号，以及完善账号信息。本教程创建的是按量付费实例。开通按量付费 ECS 资源时，您的天翼云账户余额（即现金余额）和代金券的总值不得小于 100.00 元人民币。充值方式请参见费用中心-在线充值。
2. 可选：在创建弹性云主机时，如果该地域没有 VPC，天翼云提供一个默认的 VPC，如果您不想使用默认的，可以在目标地域创建 VPC。具体操作，请参见虚拟私有云-创建 VPC、子网搭建私有网络。
3. 可选：天翼云提供一个默认的安全组，如果您不想使用默认的，可以在目标地域创建一个安全组。具体操作，请参见虚拟私有云-安全组-创建安全组。

进入创建云主机页面

1. 登录天翼云，点击页面右上方“控制中心”，进入控制台。
2. 单击管理控制台左上角的 ，选择区域，此处我们选择华东 1。
3. 点击计算-弹性云主机进入创建云主机页面。
4. 单击右上角创建云主机进入云主机购买页面。

基础配置

1. 选择计费模式：

- 包年包月：一种预付费模式，即先付费再使用。一般适用于固定业务应用，例如网站服务。需要先支付包年包月资源账单，才能开始使用包年包月资源。

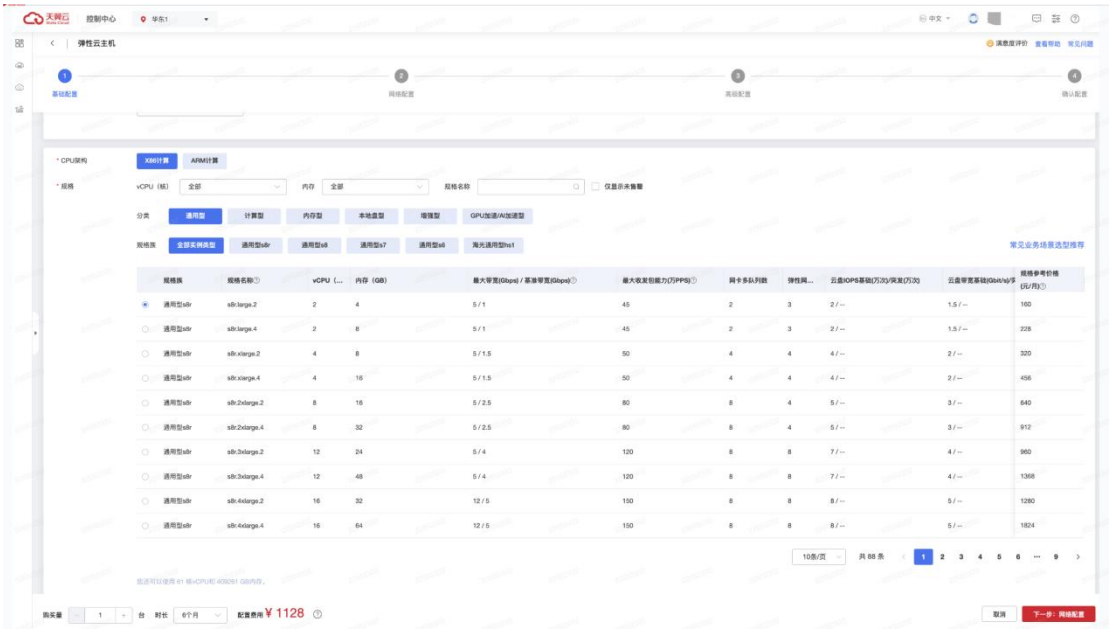
- 按量付费： 一种后付费模式，即先使用再付费。一般适用于有业务变化的应用，例如临时扩展、临时测试。可以先开通并使用按量付费资源，系统在每个结算周期生成账单并从账户中扣除相应费用。
- 此处我们选择按量付费。

2. 选择地域，默认为控制台左上角所设置的区域，可根据实际需求进行选择。
3. 选择可用区，请根据实际需求进行选择。如果您需要购买多台云主机，建议选择不同可用区，实现容灾效果。
4. 设置企业项目，该参数针对企业用户使用。如需使用该功能，请联系客服申请开通。企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理，默认项目为 default。
5. 设置虚拟私有云，您可以选择使用已有的虚拟私有云网络，或者创建新的虚拟私有云。
6. 设置“实例名称”和“主机名称”。其命名规则为：

参数	命名规则
实例名称	长度为 2-63 字符。
主机名称	Windows 系统，长度为 2～15 个字符，允许使用大小写字母、数字或连字符（-）。不能以连字符（-）开头或结尾，不能连续使用连字符（-），也不能仅使用数字。其他操作系统（Linux 等），长度为 2～64 个字符，允许使用点号（.）分隔字符成多段，每段允许使用大小写字母、数字或连字符（-），但不能连续使用点号（.）或连字符（-）。不能以点号（.）或连字符（-）开头或结尾。修改主机名称需要重启生效。

7. （可选）设置云主机描述：您可以为云主机添加描述，方便备注和区分云主机。在云主机创建后，您可以通过“编辑云主机属性”对云主机描述进行修改。

8. 选择 CPU 架构，天翼云提供 X86 和 ARM 两种 CPU 架构，可以根据您的业务需求进行选择。
9. 选择规格，天翼云提供了多种类型的弹性云主机供您选择，针对不同的应用场景，可以选择不同内存和 CPU 核数的弹性云主机。如果您对自身业务需要购买的云主机规格不明确，可以点击“常见业务场景选型推荐”，在帮助文档中查看不同场景所应该购买的主机规格。



10. 选择镜像

- 公共镜像：常见的标准操作系统镜像，所有用户可见，包括操作系统以及预装的公共应用。请根据您的实际情况自助配置应用环境或相关软件。
- 私有镜像：用户基于弹性云主机创建的个人镜像，仅用户自己可见。包含操作系统、预装的公共应用以及用户的私有应用。选择私有镜像创建弹性云主机，可以节省您重复配置弹性云主机的时间。
- 共享镜像：您将接受其他用户共享的私有镜像，作为自己的镜像进行使用。
- 安全产品镜像：仅用于加载部分安全产品服务。
- 应用镜像：与基础的公共镜像相比，应用镜像预装了一些常见应用，可以实现快速部署特殊应用的目的。

说明

选择公有镜像时，用户可选择是否开启安全防护功能，关于安全防护功能的具体介绍可参考服务器安全卫士介绍。当前所有资源池云主机创建页支持开启免费版本安全防护，部分资源池支持开启企业版本安全防护，具体请以资源池实际情况为准。

弹性云主机创建成功后，将自动安装服务器安全卫士的 Agent 并开启防护，这个过程需要一定的时间，请耐心等待。

11. 设置“存储”。您可以为弹性云主机添加多块数据盘，系统盘大小目前默认为 40GB。系统盘和数据盘可提供“普通 IO”、“高 IO”、“通用型 SSD”、“超高 IO”及“极速型 SSD”等存储类型（不同资源池提供不同的存储类型）。
12. 单击“下一步：网络配置”。

步骤 2：网络配置

1. 设置网络，包括“安全组”、“网卡”、“弹性 IP”等信息。第一次使用网络服务时，系统将自动为您创建一个默认虚拟私有云，包括安全组、网卡。

参数	说明
虚拟私有云	弹性云主机网络使用虚拟私有云（VPC）提供的网络，包括子网、安全组等。您可以选择使用已有的虚拟私有云网络，或者创建新的虚拟私有云。更多关于虚拟私有云的信息，请参见《天翼云虚拟私有云用户使用指南》。
安全组	安全组用来实现安全组内和安全组间云主机的访问控制，加强云主机的安全保护。用户可以在安全组中定义各种访问规则，当云主机加入该安全组后，即受到这些访问规则的保护。创建弹性云主机时，可支持选择多个安全组（建议不超过 5 个）。此时，云主机的访问规则遵循几个安全组规则的并集。注意：1.如需通过远程桌面连接到 Windows 云主机，请在安全组中增加如下规则，方向：入方向，协议：TCP，端口范围：3389。2.如需通过远程桌面连接到 Linux 弹性云主机，请在安全组中增加如下规则，方向：入方向，协议：TCP，端口范围：22。3.如需 Ping 云主机地址，请在安全组中增加如下规则，方向：入方向，协议：ICMP，地址

参数	说明
	和掩码：0.0.0.0/0
网卡	包括主网卡和扩展网卡。订购主机时可设置主网卡和扩展网卡。网卡使用限制可参考弹性云主机-产品使用限制。
弹性IP	弹性 IP 是指将公网 IP 地址和路由网络中关联的云主机绑定，以实现虚拟私有云内的云主机通过固定的公网 IP 地址对外提供访问服务。可以根据实际情况选择以下三种方式：1.不使用：云主机不能与互联网互通，仅可作为私有网络中部署业务或者集群所需云主机进行使用。2.自动分配：自动为每台云主机分配独享带宽的弹性 IP，带宽值可以由您设定。3.使用已有：为云主机分配已有弹性 IP。使用已有弹性 IP 时，不能批量创建云主机。

说明

创建主机时，最多支持挂载 5 块弹性网卡及 1 个弹性 IP，若用户需要挂载更多弹性网卡及 IP 请在完成主机创建后，前往主机详情页进行挂载。

2. 单击“下一步：高级配置”。

步骤 3：高级配置

1. 选择登录方式

- 密钥对：指使用密钥对作为云主机的鉴权方式。您可以选择使用已有的密钥，或者单击“查看密钥对”创建新的密钥。 注：如果选择使用已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录云主机。
- 密码：指使用设置初始密码方式作为云主机的鉴权方式，此时，您可以通过用户名密码方式登录云主机，Linux 操作系统使用 root 或 ecs-user 用户和初始密码，Windows 操作系统使用 Administrator 用户和初始密码。密码复杂度需满足：

参数	规格
密码	8～30 个字符，必须同时包含三项（大写字母、小写字母、数字、 ()`~!@#$%^&*_-+=|{}[]:;'<>,.?/ 中的特殊符号），不能以斜线号（/）开头。

说明

当创建 Windows 云主机时，仅支持选择密码登录方式。

您可以在“创建密码”选项处选择“稍后创建”按钮，在云主机创建成功后，您可通过“重置密码”操作设置密码。重置密码操作详见在控制台重置密码。

2. 【可选】设置云主机组，云主机组内的云主机在创建时，将尽量分散地创建在不同的宿主机上，提高业务的可靠性；或尽量集中地创建在同一宿主机上，提升业务的性能。创建云主机组，详情请参见云主机实例-云主机组管理。
3. 【可选】用户数据，主要用于创建云主机时向云主机注入实例自定义数据。配置后，云主机首次启动时会自行注入数据信息。

- 以文本形式：在下方文本框内输入用户数据内容。
- 以文件形式：主要用于创建云主机时注入的脚本文件或其他文件。
- 例如：您可以通过注入一段脚本，激活待创建云主机的 root 用户权限，注入成功后，您可以使用 root 用户登录弹性云主机。

4. 单击“下一步，确认配置”。

步骤 4：确认配置

1. 在“确认配置”页面，查看云主机配置详情。
2. 设置购买量。

- 购买时长：“包年/包月”方式需要设置购买时长，最短为 1 个月，最长为 5 年。
- 自动续订：“包年/包月”方式可选是否开启自动续订。按月购买的自动续订周期为 1 个月，按年购买的自动续订周期为 1 年。

- 购买数量：设置购买弹性云主机的数量。为了保证所有资源的合理分配，如果您需要的弹性云主机数量超过当前您可以购买的最大数值，您要提交工单申请扩大配额。申请通过后，您可以购买到满足您需要的弹性云主机数量。

3. 协议：阅读并勾选同意协议。
4. 如果您确认配置无误，单击“立即购买”。
5. 点击“立即支付”进行付款，付款成功即可创建弹性云主机。弹性云主机创建成功后，您可以在弹性云主机信息页面看到您新创建的弹性云主机。

6.3 登录 Linux 弹性云主机

只有运行中的弹性云主机才允许用户登录，登录 Linux 弹性云主机的方式有：SSH 方式登录、VNC 方式登录。

- SSH 方式登录：仅适用于 Linux 弹性云主机。包括密码方式和密钥方式两种，您可以使用远程登录工具（例如 PuTTY），登录弹性云主机。此时，需要该弹性云主机绑定弹性 IP。
- VNC 方式登录：未绑定弹性 IP 的弹性云主机可通过控制中心提供的远程登录方式直接登录。

说明

首次登录使用用户名可以设置为 root 或 ecs-user。root 和 ecs-user 说明如下：

- root： root 用户为 Linux 系统的超级用户，在系统安装时默认创建，拥有操作系统的所有权限，不可被删除。
- ecs-user(推荐)： ecs-user 用户为 Linux 系统的一个普通用户， 没有系统级的权限，只能执行被授权的操作，若涉及系统敏感的操作需要通过 sudo 命令临时获得 root 权限执行操作。比如：使用 systemctl 命令重启服务，使用 mount 命令挂载磁盘设备等，需要在命令前加入 sudo 进行提权。
- 使用 root 用户登录可能会导致云主机安全风险，基于云主机安全考虑，我们建议您创建云主机时选择“立即创建”，并设置 ecs-user 作为初始登录用户；若您在创建云主机时选择稍后创建，则默认会使用 root 用户登录云主机。

密钥对方式登录

前提条件

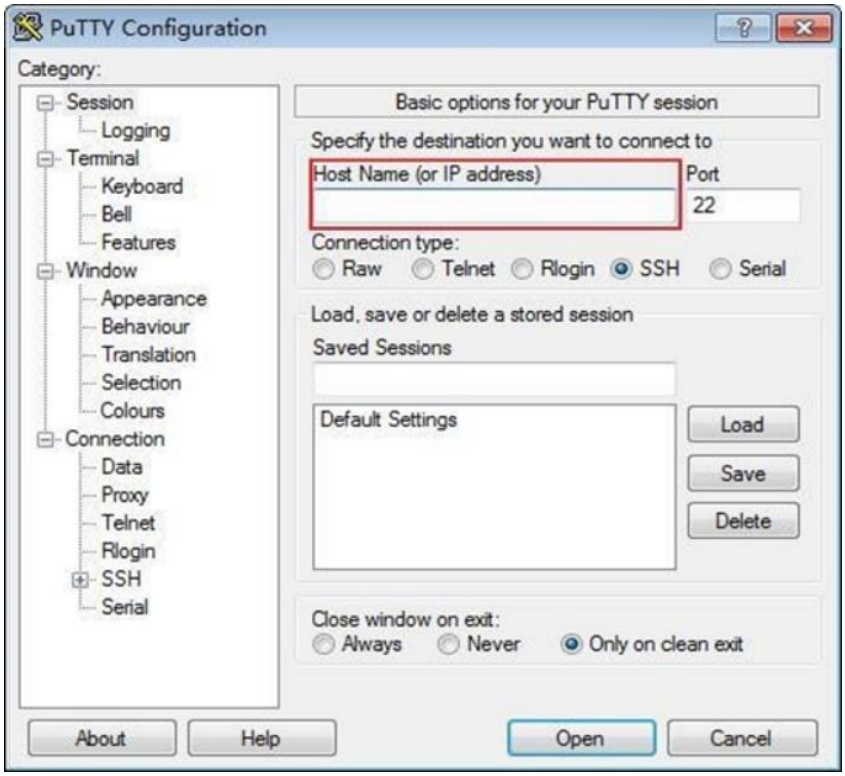
1.已获取该弹性云主机的密钥文件，获取方式参见通过管理控制中心创建密钥对；
2.弹性云主机已经绑定弹性 IP；
3.已配置安全组入方向的访问规则。例如，默认的 22 端口没有被防火墙屏蔽；
4.使用的登录工具（如 PuTTY）与待登录的弹性云主机之间网络连通。

客户端使用 Windows 操作系统

如果您本地使用 Windows 操作系统登录 Linux 弹性云主机，可以按照下面方式登录弹性云主机：

我们以 PuTTY 为例介绍如何登录弹性云主机。使用 PuTTY 登录弹性云主机前，需要先将私钥文件转化为.ppk 格式。

1.在以下路径中下载 PuTTY 和 PuTTYgen。PuTTYgen 是密钥生成器，用于创建密钥对，生成一对公钥和私钥供 PuTTY 使用；PuTTY 下载
2.运行 PuTTYgen；
3.在“Actions”区域，单击“Load”，并导入创建弹性云主机时保存的私钥文件； 导入时注意确保导入的格式要求为“All files (.)”；
4.单击“Save private key”；
5.保存转化后的私钥到本地。例如：ssh-example.ppk；
6.双击“PUTTY.EXE”，打开“PuTTY Configuration”；
7.选择“Connection > data”，在 Auto-login username 处输入镜像的用户名；
8.选择“Connection > SSH > Auth”，在最下面一个配置项“Private key file for authentication”中，单击“Browse”，选择步骤 5 转化完成的密钥；
9.单击“Session”，在“Host Name (or IP address)”下的输入框中输入弹性云主机的弹性 IP 地址；



10. 单击“Open”。登录弹性云主机。

客户端使用 Linux 操作系统

如果您本地使用 Linux 操作系统登录 Linux 弹性云主机，可以按照下面方式登录。下面步骤以私钥文件是 ssh-exampl.pem 为例进行介绍。

1.在您的 linux 计算机的命令行中执行如下命令，变更权限。下列命令的 path 为密钥文件的存放路径；

```
chmod 400 /path/ssh-exampl
```

2.执行如下命令，登录云主机；

```
ssh -i /path/ssh-exampl 默认用户名@云主机
```

3.假设 Linux 云主机的默认用户名是 linux，则命令如下：

```
ssh -i /path/ssh-exampl linux@弹性 IP 地址
```

path 为密钥文件的存放路径；

弹性 IP 地址为云主机绑定的弹性 IP 地址。

密码方式登录

客户端使用 Windows 操作系统

如果客户端使用的计算机系统为 Windows 操作系统，可以按照下面方式登录云主机。下面步骤以 PuTTY 为例：

1.运行 PuTTY；
2.单击“Session”，在“Host Name (or IP address)”下的输入框中输入云主机的弹性 IP；
3.单击“Window”，在“Translation”下的“Received data assumed to be in which character set:”选择“UTF-8”；
4.单击“Open”；
5.输入用户名和创建云主机时设置的密码登录云主机。

客户端使用 Linux 操作系统

如果客户端使用的计算机系统为 Linux 操作系统，您可以在计算机的命令行中运行如下命令登录云主机器。

1.输入 SSH 命令：ssh 用户名@弹性 IP；
2.输入用户名和创建云主机时设置的密码登录云主机。

6.4 登录 Windows 弹性云主机

只有运行中的云主机才允许用户登录，登录云主机的方式有：VNC 方式登录、MSTSC 方式登录。

- VNC 方式登录：未绑定弹性 IP 的云主机可通过控制中心提供的远程登录方式直接登录。
- MSTSC 方式登录：仅适用于 Windows 云主机。您可以通过在本机运行 MSTSC 方式登录云主机。此时，需要该云主机绑定弹性 IP。

说明

如果使用控制中心提供的“远程登录”功能无法满足您的访问需求，请自行在云主机上安装符合要求的远程访问工具（如 Tight VNC）。 Tight VNC 下载地址：Tight VNC 下载

首次登录使用用户名 administrator。

VNC 方式登录

约束与限制

- 当前提供的远程登录功能是通过系统配置的自定义端口进行访问的，所以在使用远程登录功能时，请确保需要使用的端口未被防火墙屏蔽。例如：远程登录的链接为“xxx:8002”，则需要确保端口 8002 没有被防火墙屏蔽。
- 如果客户端操作系统使用了本地代理，且用户无法配置该本地代理的防火墙端口，请关闭代理模式后再使用远程登录功能。

操作步骤

1. 进入天翼云官网，点击页面右上方“控制中心”按钮，登录控制中心；
2. 选择“计算”>“弹性云主机”；
3. VNC 方式登录云主机时，需已知其密码，然后再采用 VNC 方式登录：
4. 在云主机列表中的右上角，输入云主机名、IP 地址或 ID 进行搜索；
5. 在搜索到的云主机的“操作”列下，单击“远程登录“；
6. （可选）如果界面提示“Press CTRL+ALT+DELETE to log on”，请单击远程登录操作面板右上方的“Send CtrlAltDel”按钮进行登录；



7. 根据界面提示，输入云主机的密码，回车完成登录验证。

MSTSC 密码方式登录

前提条件

- Windows 云主机已经绑定弹性公网 IP；
- 已配置安全组 3389 端口入方向的访问规则；
- 使用的登录工具与待登录的 Windows 云主机之间网络连通。例如，默认的 3389 端口没有被防火墙屏蔽。

操作步骤

打开 Windows 云主机远程登录 RDPWindows（Remote Desktop Protocol）协议。首次登录云主机时，请先使用 VNC 方式登录云主机，打开 RDP，然后再使用 MSTSC 方式连接。

1.检查 Windows 云主机的 RDP 是否开启；

1）VNC 方式登录云主机；

2）单击“开始”菜单，选择“控制面板 > 系统和安全 > 系统 > 远程设置”，系统进入“系统属性”页面；



3）选择“远程”页签，在“远程桌面”栏，选择“允许远程连接到此计算机”；

4）单击“确定”；

2.在用户本地计算机，使用远程登录工具 MSTSC 登录 Windows 云主机；

1）单击“开始”菜单；

2）在“搜索程序和文件”中，输入“mstsc”；

3）根据提示登录云主机。

3.（可选）通过远程桌面连接（Remote Desktop Protocol，RDP）方式登录云主机后，如果需要使用 RDP 提供的“剪切板”功能，将本地的大文件（文件大小超过 2GB）复制粘贴至远端的 Windows 云主机中，由于 Windows 系统的限制，会导致操作失败。具体解决办法请查看微软官方提供的帮助手册：Copying files exceeding 2 GB fails - Windows Server，或咨询微软官方。

5.5 初始化数据盘

具体操作流程请参见云硬盘-快速入门-初始化数据盘。

6.5 入门实践

背景介绍

Discuz!是一款通用的社区论坛软件系统，它采用 PHP 和 MySQL 组合的基础架构，为您提供高效的论坛解决方案。

前提条件

已购买弹性云主机实例，实例满足以下条件：

1.实例已分配公网 IP 地址或绑定弹性公网 IP（EIP）。
2.操作系统必须为 CentOS 7.x。
3.实例安全组的入方向规则已放行 22、80、443 端口。

操作步骤

安装 Apache

Apache 是世界使用排名第一的 Web 服务器软件。它可以运行在几乎所有广泛使用的计算机平台上，由于其跨平台和安全性被广泛使用，是最流行的 Web 服务器端软件之一。它快速、可靠并且可通过简单的 API 扩充，将 Perl/Python 等解释器编译到服务器中。

1.安装软件 httpd

httpd 是 Apache 超文本传输协议(HTTP)服务器的主程序。执行以下命令：

```
yum install httpd -y
```



2.启动主程序 httpd

启动主程序 httpd，用以接受请求。执行以下命令：

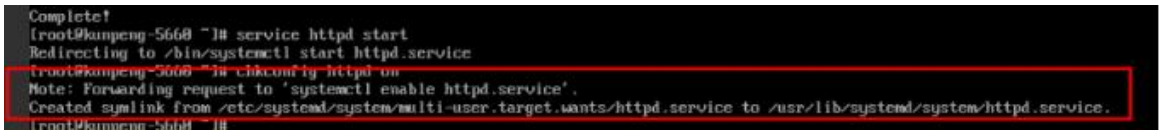
```
service httpd start
```



3.设置 httpd 开机自启动

httpd 开机自动启动可以保证虚拟机重启之后，服务可以同时启动，执行以下命令：

```
chkconfig httpd on
```



4. 安装 php 编译器

安装 php 编译器，为后续的编写工作做准备。执行以下命令：

```
yum install php  -y
```



5. 安装 php-mysql 插件

安装 php 编译器与 mysql 链接的插件，用户进行数据库连接。执行以下命令：

```
yum install php-mysql -y
```



Apache 软件安装完成。

安装 MariaDB

Mariadb 是一个数据库，主要用于后续存储论坛数据。

1. 安装 mariadb

执行以下命令

```
yum install mariadb-server mariadb  -y
```



2. 启动 mariadb

执行以下命令：

```
systemctl start mariadb
```

3. 设置 mariadb 开机自启动

mariadb 开机自动启动可以保证虚拟机重启之后，服务可以同时启动；执行以下命令：

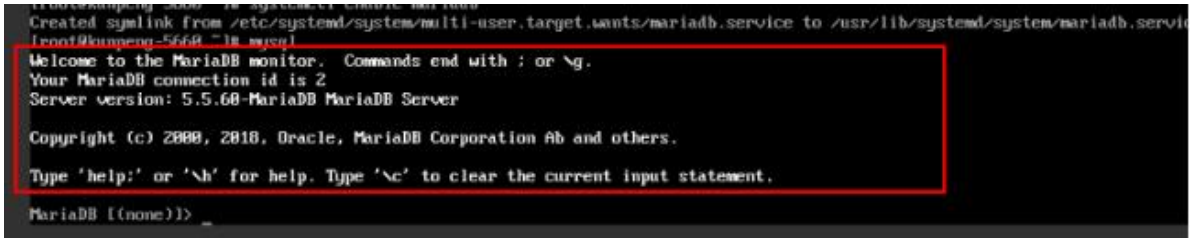
```
systemctl enable mariadb
```



4. 进入 mysql 数据库

执行以下命令：

```
mysql
```



5. 创建数据库的用户名和密码

创建访问数据库的用户名和密码，并进行授权；如我们需要创建如下信息：

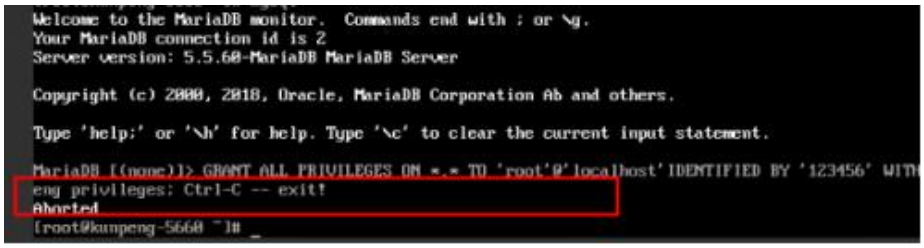
- 用户名：root
- 密码：123456

- 主机：localhost（本机）
- 数据库名字：tianyi

需要执行以下命令：

```
GRANT ALL PRIVILEGES ON *.* TO 'root'@'localhost'IDENTIFIED BY '123456' WITH GRANT OPTION;CREATE DATABASE tianyi;flush privileges;
```

代码配置完成后，输入 ctrl+c 退出数据库编辑。

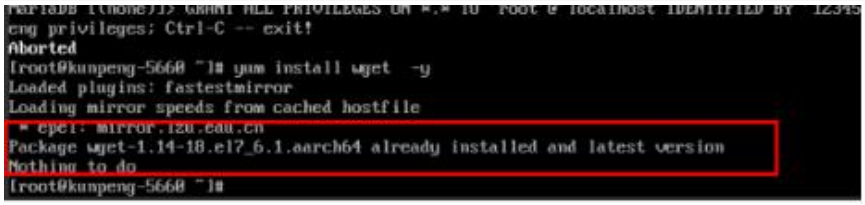


Discuz!安装

1.安装下载工具

下载 discuz!之前需要先下载 wget 工具，执行以下命令：

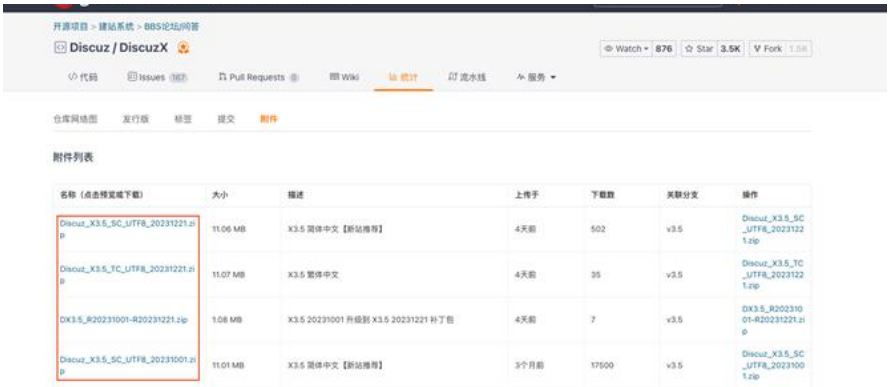
```
yum install wget -y
```



2.下载 Discuz!源码包

请先登录 gitee 官网获取 Discuz 下载地址：

执行以下命令：

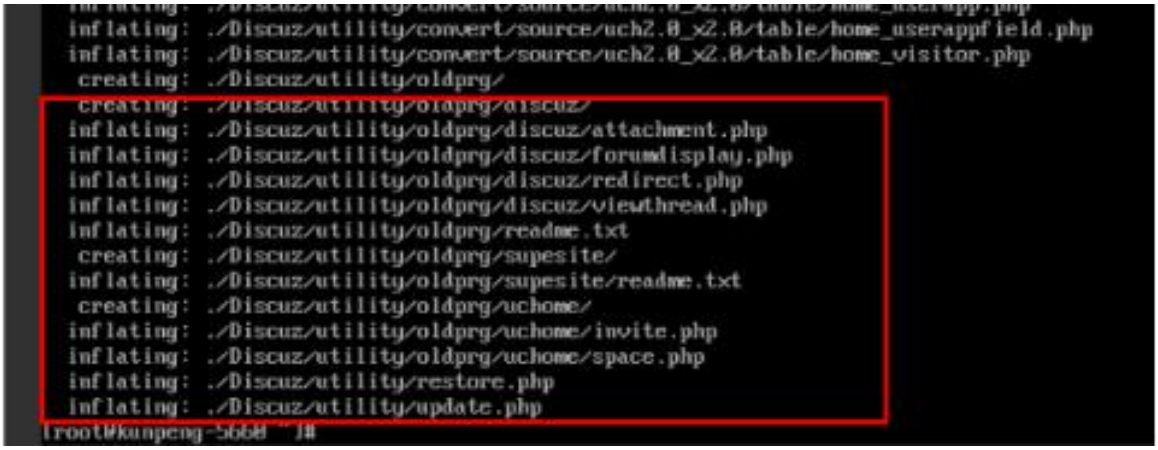


```
wget https://gitee.com/Discuz/DiscuzX/attach_files/XXX/download
```

3.解压 Discuz!源码包

执行以下命令：

```
unzip -d ./Discuz ./Discuz_X3.5_SC_UTF8_20250205.zip
```



4.Discuz!源码移动

将 discuz!移动到 httpd 启动的目录，使用户可以通过 http 访问到 discuz!网站。

执行以下命令：

```
mv ./Discuz/upload/* /var/www/html/
```



5.修改 Discuz!源码权限

修改源码读写权限，使网站能够被所有用户正常读写访问。执行以下命令：

```
chmod -R 777 /var/www/html/*
```



6.重启 httpd

重启 httpd，使所有设置好的环境变量生效。执行以下命令：

```
service httpd restart
```



7.浏览器访问 Discuz

输入 http://+云主机的 EIP 地址，访问 discuz!论坛网站。

Demo 体验

下面我们进行 Discuz!安装，请大家按照步骤进行操作。

1.Discuz!安装，点击“我同意”。



2.设置运行环境，点击“下一步”。

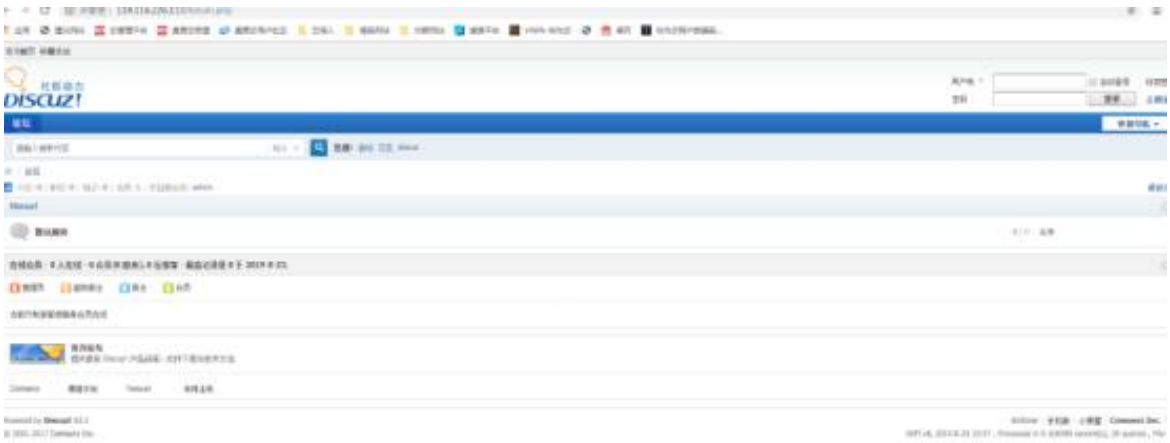


3.安装数据库，填入数据库信息、管理员信息。



4.完成安装，登录 Discuz!论坛网站。





当最终出现 Discuz!论坛网站页面时，证明操作成功。在天翼云弹性云主机实例上搭建 Discuz!论坛网站就完成了。

6.6 修订记录

修订记录（快速入门）

发布日期	修订记录
2023-9-1	第一次正式发布。 本次变更说明：全新改版。

7 用户指南

7.1 云主机实例

7.1.1 实例生命周期

实例状态分为稳定状态和中间状态，中间状态是实例在到达稳定状态前暂时处于的状态，如果实例长时间处于中间状态，说明可能出现了异常。

在实例生命周期中可能的状态如下表所示：

状态	状态类型	说明
创建中	中间状态	创建实例之后在实例开通完成之前的中间状态，创建完成后该实例会变为“运行中”。
运行中	稳定状态	实例正常运行状态，用户可正常运行相关业务。
停止中	中间状态	对处于“运行中”状态的实例执行关机操作后的中间状态，关机完成后该实例变为“关机”。

状态	状态类型	说明
节省关机中	中间状态	对处于“运行中”状态的实例执行节省关机操作后的中间状态，关机完成后该实例变为“节省关机”。
关机	稳定状态	操作实例关机，经过“停止中”到达的稳定状态。在“关机”状态下的实例，不能对外提供服务。
节省关机	稳定状态	操作实例节省关机，经过“节省关机中”到达的稳定状态。在“节省关机”状态下的实例，不能对外提供服务。
启动中	中间状态	对处于“关机”状态的实例执行开机操作后的中间状态，启动完成后该实例变为“运行中”。
重启中	中间状态	对处于“运行中”状态的实例执行重启操作后的中间状态，重启完成后该实例变为“运行中”。
变配中	中间状态	对实例执行变配操作后的中间状态，变配成功后会置为执行变配前原本的状态。
重建中	中间状态	对处于“关机”状态的实例执行“一键重装”后的中间状态，该状态为实例执行重装操作系统的过程。重装完成后该实例会变为“关机”。
删除中	中间状态	实例处于正在被删除的状态，用户删除成功将会收到退订短信。

状态	状态类型	说明
已删除	稳定状态	已删除的实例将从控制台列表显示中移除。此状态不在控制台展示。
已冻结	稳定状态	订购按需计费模式的实例，账号欠费后，云主机进入该状态。完成账号续费后，该状态解除。
已到期	稳定状态	订购包周期计费模式的实例，到期后进入该状态。续订实例后，该状态解除。
错误	稳定状态	在上述任何操作中导致云主机出现错误，则会置为 error。此时可以通过两种方式解决问题： 方式一，请参考帮助文档中如何解决云平台创建云主机状态错误和云主机错误状态及解决方案进行自助修复； 方式二，请提交工单，由技术人员协助解决问题。
包周期退订中	中间状态	控制台对包年包月计费模式的云主机实例执行退订操作后的中间状态，退订完成后该实例变状态为“包周期已退订”。
包周期已退订	稳定状态	订购包周期计费模式的云主机实例，手动退订后，云主机进入该状态。

7.1.2 查看实例

7.1.2.1 查看实例状态

操作场景

在您提交创建弹性云主机后，可以查看该弹性云主机的状态。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在弹性云主机列表页面，找到您需要查看的弹性云主机，对应“状态”栏展示其当前状态，通常提交创建后会显示为“创建中”的中间状态，创建完成后会变为“运行中”的稳定状态。
5. 更多状态和说明可查看。

7.1.2.2 查看实例详情

操作场景

在您创建弹性云主机后，通过控制台进入弹性云主机列表页与详情页，可以查看弹性云主机的列表与详细信息。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。

3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”，进入弹性云主机控制台，此处展示弹性云主机列表页，您可以查看弹性云主机名称、规格、镜像等基本信息。
4. 在弹性云主机列表页，点击目标弹性云主机名称，即可跳转至该云主机的详情页，可以进一步查看弹性网卡、云硬盘、安全组、弹性 IP、监控等详细信息。

7.1.2.3 导出实例信息

操作场景

您可以将当前帐号下的弹性云主机信息，如 ID、名称、规格、镜像、云盘、IP、安全组、创建时间、付费方式等信息，以.xls 文件的形式导出至本地。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在弹性云主机列表页，单击右上角的导出按钮，即可将当前账号下当前区域的所有弹性云主机信息自动导出，并下载至本地。
5. 在本地计算机的下载路径，可以获取到导出的弹性云主机信息文件。

7.1.2.4 搜索实例

操作场景

购买弹性云主机后，您可以通过控制台提供的搜索功能快速检索当前区域的云主机资源。

当前支持使用云主机名称、云主机 ID、IP 地址属性进行搜索。

操作步骤

1. 登录控制台，进入弹性云主机列表页。
2. 在搜索框直接输入云主机名称、ID 或 IP 地址，点击搜索按钮或回车，将为您识别并匹配目标云主机。

7.1.3 登录实例

7.1.3.1 登录 Windows 弹性云主机实例

7.1.3.1.1Windows 弹性云主机登录方式概述

约束与限制

- Windows 云主机处于“运行中”状态。
- 已获得 Windows 云主机的用户名和密码。如果忘记密码，须重置云主机密码，请参考在控制台重置弹性云主机密码。
- Windows 云主机所在安全组的入方向已开放 3389 端口。
- Windows 防火墙未开启或开启后放通 3389 端口。
- 登录工具与 Windows 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Windows 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。
- Windows 云主机已开启远程桌面协议 RDP。
- GPU 实例中，部分实例不支持云平台提供的远程登录功能，需要自行安装 VNC Server 进行登录。推荐使用 MSTSC 方式登录弹性云主机。
- 使用 MSTSC 方式访问 GPU 型云主机时，使用 WDDM 驱动程序模型的 GPU 将被替换为一个非加速的远程桌面显示驱动程序，造成 GPU 加速能力无法实现。因此，如果需要使用 GPU 加速能力，您必须使用不同的远程访问工具，如 VNC 工具。如果使用管理控制中心操作界面的“远程登录”功能无法满足您的访问需求，请自行在弹性云主机上安装符合要求的远程访问工具（如 Tight VNC）。

登录方式概述

请根据需要选择登录方式，登录 Windows 云主机。

云主机操作系统	本地主机操作系统	连接方法	条件
Windows	Windows	使用控制中心远程登录方式： 可参见：Windows 弹性云主机控制中心远程登录（VNC 方式）。	不依赖弹性 IP
	移动设备	安装远程连接工具，例如 Microsoft Remote Desktop 在移动设备上登录。 可参见：在移动设备上登录 Windows 云主机	云主机绑定弹性 IP（如果通过内网环境登录云主机，可以不绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。）
	Windows	使用 mstsc 方式登录云主机。在本地主机单击“开始”菜单，输入 mstsc 命令，打开远程桌面对话框。 可参见：远程桌面连接（MSTSC 方式）	
	Linux	安装远程连接工具，例如 rdesktop，执行连接命令。 可参见：在 Linux 主机上登录 Windows 云主机	
	Mac OS 系统	安装远程连接工具，例如 Microsoft Remote Desktop for Mac 在 Mac OS 系统上登录。 可参见：Mac OS 系统登录 Windows 云主机	

开启远程桌面协议 RDP

首次登录弹性云主机时，请先使用 VNC 方式登录弹性云主机，打开 RDP（Remote

Desktop Protocol），然后再使用 MSTSC 方式连接。

说明

使用公共镜像创建的 Windows 云主机，默认已打开 RDP。

1. 使用 VNC 方式登录 Windows 云主机。请参考 Windows 弹性云主机控制中心远程登录（VNC 方式）。
2. 在 Windows 操作系统中，点击"开始"菜单。
3. 在开始菜单中，选择"控制面板"，然后选择"系统和安全"，接着选择"系统"。
4. 在系统页面中，点击左侧的"远程设置"链接。
5. 在远程设置页面的"远程桌面"栏，勾选"允许远程连接到此计算机"选项。
6. 点击"确定"按钮，以保存并应用所做的更改。

7.1.3.1.2Windows 弹性云主机控制中心远程登录（VNC 方式）

操作场景

本节为您介绍如何通过控制中心操作界面的远程登录功能（即 VNC 方式）登录到 Windows 弹性云主机上。

约束与限制

- 当前提供的远程登录功能是通过系统配置的自定义端口进行访问的，所以在使用远程登录功能时，请确保需要使用的端口未被防火墙屏蔽。例如：远程登录的链接为“xxx:8002”，则需要确保端口 8002 没有被防火墙屏蔽。
- 如果客户端操作系统使用了本地代理，且用户无法配置该本地代理的防火墙端口，请关闭代理模式后再使用远程登录功能。
- GPU 实例中，部分实例不支持控制中心操作界面的远程登录功能，需要自行安装 VNC Server 进行登录。推荐使用 MSTSC 方式登录弹性云主机。

登录 Windows 弹性云主机

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。

3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 获取 Windows 弹性云主机密码。VNC 方式登录弹性云主机时，需已知其密码方可登录。

- 当您的 Windows 弹性云主机是采用密码方式鉴权时，请直接使用创建云主机时设置的密码进行登录。
- 当您的 Windows 弹性云主机是采用密钥方式鉴权时，可以使用重置密码功能为云主机设置新密码。

5. 选择要登录的 Windows 弹性云主机，单击“操作”列下的“远程登录”，弹出带有操作系统界面的远程登录窗口。
6. 如果界面提示“按 Ctrl+Alt+DEL 解锁”，请单击远程登录操作面板右上方的“Send CtrlAltDel”进入到用户登录界面。
7. 输入弹性云主机密码完成登录。

7.1.3.1.3 远程桌面连接（MSTSC 方式）

操作场景

在本节中，我们将介绍如何通过 Windows 操作系统中的远程桌面连接（MSTSC 方式）登录到 Windows 弹性云主机，您可以方便地访问和管理弹性云主机的操作系统界面，以进行远程操作和配置。

前提条件

- Windows 云主机处于“运行中”状态。
- 已获得 Windows 云主机的用户名和密码。如果忘记密码，须重置云主机密码，请参考在控制台重置弹性云主机密码。
- Windows 云主机所在安全组的入方向已开放 3389 端口。
- Windows 防火墙未开启或开启后放通 3389 端口。
- 登录工具与 Windows 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Windows 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。

- Windows 云主机已开启远程桌面协议 RDP。

操作步骤

1. 在本地 Windows 主机的开始菜单中，点击"开始"按钮。
2. 在“搜索程序和文件”中，输入“mstsc”并点击打开远程桌面连接工具。
3. 在“远程桌面连接”的对话框中，点击“选项”。
4. 在选项中,输入待登录的云主机的弹性 IP 和用户名(默认为 Administrator)。
5. （可选）如果需要在远程会话中使用本地主机的资源，点击“本地资源”选项卡，并根据需求进行配置。可以勾选“剪贴板”来实现从本地主机复制到云主机的操作。
6. （可选）如果需要调整远程桌面窗口的大小，可以选择“显示”选项卡并调整窗口大小。
7. 点击“确定”，根据提示输入密码，登录云主机。首次登录云主机时，需要更改密码以提高安全性。
8. （可选）注意，使用远程桌面连接（RDP）方式登录云主机后，如果需要通过 RDP 将本地的大文件（文件大小超过 2GB）复制粘贴到远程的 Windows 云主机中，由于 Windows 系统的限制，可能导致操作失败。

7.1.3.1.4 在 Linux 主机上登录 Windows 云主机

操作场景

本节将介绍如何在 Linux 主机上使用远程桌面协议（RDP）登录到 Windows 云主机。通过这种方式，您可以使用 Linux 主机来远程管理和操作 Windows 云主机的桌面环境。

前提条件

- Windows 云主机处于“运行中”状态。
- 已获得 Windows 云主机的用户名和密码。如果忘记密码，须重置云主机密码，请参考在控制台重置弹性云主机密码。

- Windows 云主机所在安全组的入方向已开放 3389 端口。
- Windows 防火墙未开启或开启后放通 3389 端口。
- 登录工具与 Windows 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Windows 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。
- Windows 云主机已开启远程桌面协议 RDP。

操作步骤

如果本地主机为 Linux 操作系统，您可以使用远程连接工具（例如 rdesktop）连接 Windows 云主机。

1. 打开终端，确保您的 Linux 主机已经安装了 rdesktop 工具。您可以执行以下命令来检查是否安装了 rdesktop：

```
rdesktop
```

如果提示"command not found"，表示 rdesktop 未安装。您可以参考 rdesktop 工具官方网站获取 rdesktop 安装包并进行安装。

2. 输入以下命令以登录到 Windows 云主机。

```
rdesktop -u 用户名 -p 密码 -g 分辨率 弹性 IP 地址
```

将命令中的以下信息替换为实际的值：

- 用户名：您用于登录 Windows 云主机的用户名。
- 密码：您用于登录 Windows 云主机的密码。
- 分辨率：您希望设置的远程桌面分辨率，以宽度和高度的形式表示。
- 弹性 IP 地址：Windows 云主机的弹性 IP 地址。

例如：

```
rdesktop -u administrator -p password -g 1024*720 121.xx.xx.xxx
```

表 1 远程登录命令参数

参数	说明
-u	用户名，Windows 实例默认用户名是 Administrator。
-p	登录 Windows 实例的密码。
-f	默认全屏，需要用 Ctrl+Alt+Enter 组合键进行全屏模式切换。
-g	分辨率，中间用星号连接。可省略，省略后默认为全屏显示。例如：1024*720
弹性 IP 地址	需要远程连接的服务器 IP 地址。需要替换为您的 Windows 实例的弹性 IP 地址。

7.1.3.1.5 在移动设备上登录 Windows 云主机

操作场景

本节以“Windows Server 2012 R2 数据中心版 64 位”操作系统为例，我们将介绍如何通过 Microsoft Remote Desktop 客户端，从移动设备上登录到 Windows 云主机。

前提条件

- Windows 云主机处于“运行中”状态。
- 已获得 Windows 云主机的用户名和密码。如果忘记密码，须重置云主机密码，请参考在控制台重置弹性云主机密码。
- Windows 云主机所在安全组的入方向已开放 3389 端口。
- Windows 防火墙未开启或开启后放通 3389 端口。
- 登录工具与 Windows 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Windows 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。
- 移动设备已安装 Microsoft Remote Desktop 客户端。

操作步骤

1. 启动 Microsoft Remote Desktop (RD Client)应用程序。
2. 在“Remote Desktop”页面的右上角，点击添加图标，并选择“Desktop”。
3. 在“Add desktop”页面中，设置登录信息，并点击“SAVE”。

- PC name：输入需要登录的 Windows 实例的弹性 IP 地址。
- 按以下步骤设置“User name”：

✧ 点击“User name”，然后从下拉列表中选择“Add user account”，弹出“Add user account”对话框。

✧ 输入 Windows 实例的帐号“administrator”，并输入实例的登录密码，然后点击“SAVE”。

4. 在“Remote Desktop”页面中，点击要登录的 Windows 实例的图标。
5. 确认信息后，点击“CONNECT”。

7.1.3.1.6Mac OS 系统登录 Windows 云主机

操作场景

本节将以使用"Microsoft Remote Desktop for Mac"工具远程连接到运行"Windows Server 2012 R2 数据中心版 64 位"操作系统的云主机为例，介绍如何在 Mac OS 系统上登录 Windows 云主机的操作步骤。

前提条件

- Windows 云主机处于“运行中”状态。
- 已获得 Windows 云主机的用户名和密码。如果忘记密码，须重置云主机密码，请参考在控制台重置弹性云主机密码。
- Windows 云主机所在安全组的入方向已开放 3389 端口。
- Windows 防火墙未开启或开启后放通 3389 端口。
- 登录工具与 Windows 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Windows 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。
- Windows 云主机已开启远程桌面协议 RDP。

- 已在 Mac OS 系统上安装了 Microsoft Remote Desktop for Mac 或其他适用于 Mac OS 系统的远程连接工具。您可以下载 Microsoft Remote Desktop for Mac。

  注意

  请注意，微软官方已停止提供 Remote Desktop 客户端的下载链接，您可以尝试下载 Microsoft Remote Desktop Beta 版本进行使用。

  操作步骤

1. 启动 Microsoft Remote Desktop 应用。
2. 点击"Add Desktop"（添加桌面）。
3. 在"Add PC"（添加计算机）页面，进行登录信息的设置。

- PC name（计算机名称）：输入要登录的 Windows 实例的弹性 IP 地址。
- User account（用户账户）：从下拉列表中选择"Add user account"（添加用户账户），然后弹出"Add user account"对话框。
- 在"Add user account"对话框中，输入 Windows 实例的用户名（例如"administrator"）和登录密码，然后点击"Add"（添加）。

4. 在"Remote Desktop"页面，双击要登录的 Windows 实例的图标。
5. 确认信息后，点击"Continue"（继续）。

7.1.3.2 登录 Linux 弹性云主机实例

7.1.3.2.1Linux 弹性云主机登录方式概述

约束与限制

- Linux 云主机状态处于“运行中”。
- Linux 操作系统一般情况首次登录使用用户名可以设置为 root 或 ecs-user，出于安全考虑，推荐您使用 ecs-user，CoreOS 的默认用户名为“core”。
- Linux 云主机所在安全组的入方向已开放 22 端口。
- Linux 云主机防火墙未开启或开启后放通 22 端口。

- 您所使用的登录工具与 Linux 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Linux 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。

  说明

  出于安全考虑，通常会修改 Linux 远程登录端口。如果您需要修改默认登录端口，请参考 Linux 服务器 SSH 登录的安全加固。

  登录方式概述

  请根据需要选择登录方式，登录云主机。

云主机操作系统	本地主机操作系统	连接方法	条件
linux	Windows	使用控制中心远程登录方式： 可参见：Linux 弹性云主机远程登录（VNC 方式）。	不依赖弹性 IP
	Windows	安使用 PuTTY、Xshell 等远程登录工具： 密码方式鉴权可参见：SSH 密码方式登录（本地使用 Windows 操作系统）。 密钥方式鉴权可参见：SSH 密钥方式登录（本地使用 Windows 操作系统）。	云主机绑定弹性 IP （如果通过内网环境登录云主机，可以不绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。）

云主机操作系统	本地主机操作系统	连接方法	条件
	Linux	使用命令连接： 密码方式鉴权可参见：SSH 密码方式登录（本地使用 Linux 操作系统）。 密钥方式鉴权可参见：SSH 密钥方式登录（本地使用 Linux 操作系统）。	
	移动设备	使用 Termius、JuiceSSH 等 SSH 客户端工具登录云主机。 可参见：在移动设备上登录 Linux 云主机。	
	Mac OS 系统	使用系统自带的终端（Terminal）： 可参见：Mac OS 系统登录 Linux 弹性云主机。	

7.1.3.2.2Linux 弹性云主机远程登录（VNC 方式）

操作场景

本节将介绍如何通过云主机控制台操作界面的远程登录功能（VNC 方式）登录到 Linux 云主机。

约束与限制

- Linux 云主机状态处于“运行中”。
- Linux 操作系统一般情况首次登录使用用户名可以设置为 root 或 ecs-user，出于安全考虑，推荐您使用 ecs-user，CoreOS 的默认用户名为“core”。
- 云主机控制台操作界面的远程登录方式，需要使用用户名和密码。如果创建 Linux 云主机时登录方式设置的为密钥，请通过重置密码功能，设置 Linux 云主机的登录密码后执行本节操作。重置密码请参考：在控制台重置弹性云主机密码。
- 远程登录功能使用系统配置的自定义端口进行访问。确保所需使用的端口未被防火墙屏蔽。例如，如果远程登录的链接是“xxx:8002”，请确保端口 8002 没有被防火墙屏蔽。
- 如果客户端操作系统使用了本地代理，且用户无法配置代理的防火墙端口，请在使用远程登录功能之前关闭代理模式。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的 ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择要登录的 Linux 弹性云主机，单击“操作”列下的“远程登录”，弹出带有操作系统界面的远程登录窗口。
5. 输入用户名，回车。
6. 输入密码，回车。

说明

Linux 的命令行界面，输入密码时无任何字符显示。输入完成后按下回车键，如果用户名和密码准确，将会自动登录成功。如果用户名和密码任何一个信息错误，将会提示重新输入。

7.1.3.2.3SSH 密钥方式登录

操作场景

本节操作介绍在 Windows 和 Linux 环境中使用 SSH 密钥对方式远程登录 Linux 云主机的操作步骤。

使用 SSH 密钥对登录，无需输入密码。该方式可以提供更高的安全性和便利性。

前提条件

- Linux 云主机状态处于“运行中”。
- Linux 操作系统一般情况首次登录使用用户名可以设置为 root 或 ecs-user，出于安全考虑，推荐您使用 ecs-user，CoreOS 的默认用户名为“core”。
- 您已经获得用于创建 Linux 云主机时所使用的密钥对文件，该文件为私钥。
- Linux 云主机所在安全组的入方向已开放 22 端口。
- Linux 云主机防火墙未开启或开启后放通 22 端口。
- 您所使用的 SSH 登录工具（例如 PuTTY）已经正确配置，并且与 Linux 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Linux 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。

  说明
  出于安全考虑，通常会修改 Linux 远程登录端口。如果您需要修改默认登录端口，请参考 Linux 服务器 SSH 登录的安全加固。

  本地使用 Windows 操作系统

  在本地使用 Windows 操作系统登录 Linux 云主机，您可以按照以下方式使用 PuTTY 进行登录。

  方式一：使用 PuTTY 登录

  我们以 PuTTY 为例介绍如何登录 Linux 云主机。在使用 PuTTY 登录 Linux 云主机之前，需要将私钥文件转换为.ppk 格式。

1. 检查私钥文件是否已经是.ppk 格式。

- 如果是.ppk 格式，请执行步骤 7。
- 如果不是.ppk 格式，请执行步骤 2。

2. 在以下链接中下载 PuTTY 和 PuTTYgen：

   https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html

   说明

   PuTTYgen 是密钥生成器，用于创建密钥对并生成一对公钥和私钥供 PuTTY 使用。

3. 运行 PuTTYgen。
4. 在“Actions”区域，单击“Load”， 导入您在创建 Linux 云主机时保存的私钥文件。请注意，在导入时确保选择了"All files (.)"格式。
5. 单击“Save private key”。
6. 将转换后的私钥保存到本地，例如：kp-101.ppk。
7. 双击“PUTTY.EXE”，打开“PuTTY Configuration”。
8. 在"Session"中，输入 Linux 云主机的弹性 IP 地址到"Host Name (or IP address)"输入框中。
9. 在"Connection”>“Data"中，输入镜像的用户名到"Auto-login username"处。
10. 在"Connection > SSH > Auth"中的"Private key file for authentication"配置项下，单击"Browse"，选择转换后的密钥文件。
11. 单击“Open”，登录 Linux 云主机。

方式二：使用 Xshell 登录

如果您希望使用 Xshell 工具登录 Linux 云主机，请按照以下步骤进行操作：

1. 打开 Xshell 工具。
2. 使用弹性 IP 执行以下命令，通过 SSH 远程连接 Linux 云主机：

   ```
   ssh 用户名@弹性 IP
   ```

3. （可选）如果系统弹出提示窗口显示"SSH 安全警告"，请点击"接受并保存"。

4. 选择"Public Key"，并单击"用户密钥(K)"栏的"浏览"。
5. 在“用户密钥”窗口中，单击“导入”。
6. 选择本地保存的密钥文件，并打击“打开”。
7. 单击“确定”，登录 Linux 云主机。

本地使用 Linux 操作系统

如果您本地使用 Linux 操作系统登录 Linux 云主机，可以按照以下方式进行登录。下面的步骤以私钥文件名为"ssh-example.pem"为例进行说明：

1. 打开终端，进入保存私钥文件的目录
2. 使用以下命令设置私钥文件的访问权限：

```
chmod 400 /path/ssh-example.pem
```

说明

上述命令中的"/path"指的是私钥文件的存放路径。

3. 执行以下命令登录 Linux 云主机。

```
ssh -i /path/ssh-example.pem 默认用户名@弹性 IP
```

假设 Linux 云主机的默认用户名是"root",弹性 IP 为 9.9.9.9,则命令如下：

```
ssh -i /path/ssh-example.pem root@9.9.9.9
```

说明

- "/path"是私钥文件的存放路径。
- 弹性 IP 地址为 Linux 云主机绑定的弹性 IP 地址。

7.1.3.2.4SSH 密码方式登录

操作场景

本节操作介绍在 Windows 和 Linux 环境中使用 SSH 密码方式远程登录 Linux 云主机的操作步骤。

前提条件

- Linux 云主机状态处于“运行中”。
- Linux 操作系统一般情况首次登录使用用户名可以设置为 root 或 ecs-user，出于安全考虑，推荐您使用 ecs-user，CoreOS 的默认用户名为“core”。
- Linux 云主机所在安全组的入方向已开放 22 端口。
- Linux 云主机防火墙未开启或开启后放通 22 端口。
- 如果创建 Linux 云主机时登录方式设置的为密钥，请通过重置密码功能，设置 Linux 云主机的登录密码后执行本节操作。重置密码请参考：在控制台重置弹性云主机密码。
- 您所使用的 SSH 登录工具（例如 PuTTY）已经正确配置，并且与 Linux 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Linux 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。

本地使用 Windows 操作系统

如果本地主机为 Windows 操作系统，可以按照以下方式登录云主机。下面的步骤以 PuTTY 为例：

1. 在以下路径中下载 PuTTY 和 PuTTYgen：
   https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html
2. 运行 PuTTY。
3. 在 PuTTY 的配置界面中，填写以下信息：

- Host Name (or IP address)：输入 Linux 云主机的弹性 IP。
- Port：输入 22。
- Connection Type：选择 SSH。

- Saved Sessions：为任务设置一个名称，以便下次使用 PuTTY 时可以直接打开该任务。

4. 单击"Window"，在"Translation"下的"Received data assumed to be in which character set:"选择"UTF-8"。
5. 单击“Open”。如果首次登录服务器，PuTTY 会显示安全警告对话框，询问是否接受服务器的安全证书。单击"是"将证书保存到本地注册表中。
6. 建立到 Linux 云主机的 SSH 连接后，根据提示输入用户名和密码登录云主机。

本地使用 Linux 操作系统

如果本地主机为 Linux 操作系统，您可以在计算机的命令行中运行如下命令登录 Linux 云主机。

```
ssh 弹性云主机绑定的弹性 IP
```

假设 Linux 云主机绑定的弹性 IP 为 9.9.9.9，则命令为：

```
ssh 9.9.9.9
```

7.1.3.2.5 在移动设备上登录 Linux 云主机

操作场景

本节操作介绍如何在移动设备上连接 Linux 实例，轻松地连接和管理您的 Linux 云主机。

- 以 iTerminal-SSH Telnet 为例介绍如何在 iOS 设备上连接 Linux 实例，详细操作请参考 iOS 设备上登录 Linux 云主机。
- 以 JuiceSSH 为例介绍如何在 Android 设备上连接 Linux 实例，详细操作请参考 Android 设备上登录 Linux 云主机。

前提条件

- Linux 云主机状态处于“运行中”。

- Linux 操作系统一般情况首次登录使用用户名可以设置为 root 或 ecs-user，出于安全考虑，推荐您使用 ecs-user，CoreOS 的默认用户名为“core”。
- Linux 云主机所在安全组的入方向已开放 22 端口。
- Linux 云主机防火墙未开启或开启后放通 22 端口。
- 确保使用的登录终端与 Linux 云主机具有网络连通性。如果使用处于公网中的终端登录，需要您的 Linux 云主机已绑定弹性 IP。如果通过内网使用此方式登录，则不需要绑定弹性 IP，例如 VPN、云专线等内网网络连通场景。

iOS 设备上登录 Linux 云主机

如果您使用 iOS 设备,请确保已经安装了 SSH 客户端工具,我们以 Termius 为例。本示例中使用 CentOS 7.6 操作系统，使用用户名和密码进行认证。

1. 在 iOS 设备上启动 Termius 应用，并单击"New Host"。
2. 在 SSH 页面上，输入连接信息后，单击 Save。需要提供以下连接信息：

- Alias（别名）：为连接设置一个名称，例如"ecs01"。
- Hostname（主机名）：输入要连接的 Linux 实例的公网 IP 地址。
- Use SSH（使用 SSH）：确保 SSH 登录配置已打开。
- Host（主机）：输入要连接的 Linux 实例的固定公网 IP 或弹性 IP。
- Port（端口）：输入端口号 22。
- 用户名：输入用户名"root"。
- 密码：输入实例的登录密码。

3. 点击右上角的"Save"按钮，保存登录信息。在 Hosts 页面上，点击连接的名称以远程连接服务器。
4. 当出现欢迎登录提示时，您已成功连接到 Linux 云主机。

Android 设备上登录 Linux 云主机

如果您使用 Android 设备，请确保已经安装了 JuiceSSH。本示例中使用 CentOS 7.6 操作系统，通过用户名和密码进行认证。

1. 启动 JuiceSSH 应用，点击"连接"。
2. 在"连接"页面，点击右下角的添加按钮。
3. 在"新建连接"页面上，填写基本设置和高级设置的信息，并保存。需要提供以下信息：

- 昵称：为登录会话指定一个名称，例如"linux_test"。
- 类型：使用默认的"SSH"。
- 地址：输入要登录的 Linux 实例的弹性 IP 地址。
- 认证：按照以下步骤设置认证：

✧ 点击"认证"，然后点击下拉列表中的"新建"。

✧ 在"新建认证"页面上，填写以下信息，并点击右上角的确认按钮。

- 昵称：可选项，您可以根据需要为身份设置一个名称，例如"linux_test"。
- 用户名：输入用户名“root”。
- 密码：单击“设置（可选）”后，输入实例的登录密码，单击“确定”。
- 端口：输入端口号“22”。

4. 在"连接"页面，点击创建好的连接。
5. 在确认提示信息后，点击"接受"。
6. （可选）第一次连接时，JuiceSSH 可能会提示您如何设置字体大小和弹出键盘等选项。请确认相关信息后，点击"好的，我知道了！"。

7.1.3.2.6Mac OS 系统登录 Linux 云主机

操作场景

本节操作以使用 Mac OS 系统自带的“终端（Terminal）”远程连接 Linux 操作系统云主机。

前提条件

- 云主机状态为“运行中”。
- Linux 操作系统一般情况首次登录使用用户名可以设置为 root 或 ecs-user，出于安全考虑，推荐您使用 ecs-user，CoreOS 的默认用户名为“core”。

- 弹性云主机已经绑定弹性 IP。
- 所在安全组入方向已开放 3389 端口。

操作步骤

您可以通过 Mac OS 系统自带的终端（Terminal）登录 Linux 云主机：

- SSH 密码方式

1. 打开系统自带的终端（Terminal），执行以下命令

```
ssh 用户名@弹性 IP
```

说明
如果是公共镜像（包括 CoreOS），用户名为“root”。

2. 输入登录密码并按下回车键，以完成登录。

- SSH 密钥方式

1. 打开系统自带的终端（Terminal）应用程序。
2. 在终端中执行以下命令，变更密钥文件的权限。下面的步骤以私钥文件"kp-101.pem"为例进行介绍。

```
chmod 400 /path/kp-101.pem
```

说明
上述命令中的"path"为密钥文件的存放路径。

3. 执行以下命令登录云主机。

```
ssh -i /path/kp-101.pem 用户名@弹性 IP
```

说明
- 如果是“CoreOS”的公共镜像，用户名为“core”。
- 如果是“非 CoreOS”的公共镜像，用户名为“root”。

7.1.3.3 管理远程登录用户

Linux 云主机初始登录用户

Linux 云主机初始登录用户指您在创建 Linux 云主机时所设置的登录用户，对于

Linux 云主机，首次登录使用用户名可以设置为 root 或 ecs-user。root 和 ecs-user 说明如下：

- root： root 用户为 Linux 系统的超级用户，在系统安装时默认创建，拥有操作系统的所有权限，不可被删除。
- ecs-user(推荐)： ecs-user 用户为 Linux 系统的一个普通用户， 没有系统级的权限，只能执行被授权的操作，若涉及系统敏感的操作需要通过 sudo 命令临时获得 root 权限执行操作。比如：使用 systemctl 命令重启服务，使用 mount 命令挂载磁盘设备等，需要在命令前加入 sudo 进行提权。

说明

- 使用 root 用户登录可能会导致云主机安全风险，基于云主机安全考虑，我们建议您创建云主机时选择“立即创建”，并设置 ecs-user 作为初始登录用户；若您在创建云主机时选择稍后创建，则默认会使用 root 用户登录云主机。
- 在系统安装时，root 用户会自动创建，如果您在创建云主机时选择使用 ecs-user 用户，那么会在您云主机内部操作系统中创建 ecs-user 用户，即系统内会同时存在 root 和 ecs-user 用户，但是 root 不用于登录云主机。
- 若 Linux 云主机系统中存在 ecs-user 用户,连接云主机默认用该用户。

Windows 云主机初始登录用户

Windows 云主机初始登录用户指您在创建 Windows 云主机时所设置的登录用户，对于 Windows 云主机,首次登录使用用户名为 administrator。该用户为 Windows 系统的超级管理员，拥有整个 Windows 系统的所有操作权限。

7.1.4 管理实例属性

7.1.4.1 修改实例名称

操作场景

弹性云主机创建成功后，您可以根据需求，修改云主机的名称。

修改弹性云主机实例名称

方法一：

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 将鼠标移动至目标云主机的“实例名称”列。
5. 单击“编辑” ，在“名称”输入框中修改实例名称。命名规则为：

参数	命名规则	备注
实例名称	长度为 2-63 字符。	中英文、数字、字符均可。可与其他云主机的实例名称重复。

6. 单击“确定”，完成弹性云主机名称修改。

方法二：

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 将鼠标移动至目标云主机的“实例/主机名称”列，点击进入详情页。
5. 找到实例名称项目，点击“编辑”按钮 ，在“实例名称”输入框中修改实例名称。

方法三：

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 将鼠标移动至目标云主机的“操作”列，点击“编辑云主机属性”。
5. 找到实例名称项目，在输入框中修改云主机实例名称。
6. 点击“确认”，修改完毕。

7.1.4.2 修改主机名称（hostname）

操作场景

您可以根据需求修改实例的主机名称（hostname）。

操作前提

云主机需处于运行中状态。

以下镜像暂不支持通过控制台手动修改 hostname：

- CentOS 6 系列
- Windows Server 2008 系列

修改弹性云主机的主机名

方法一：

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 将鼠标移动至目标云主机的“操作”列。
5. 单击“更多” ，编辑“编辑云主机属性”，在输入框中修改主机名称。

参数	命名规则
主机名称	Windows 系统，长度为 2～15 个字符，允许使用大小写字母、数字或连字符（-）。不能以连字符（-）开头或结尾，不能连续使用连字符（-），也不能仅使用数字。其他操作系统（Linux 等），长度为 2～64 个字符，允许使用点号（.）分隔字符成多段，每段允许使用大小写字母、数字或连字符（-），但不能连续使用点号（.）或连字符（-）。不能以点号（.）

参数	命名规则
	或连字符（-）开头或结尾。修改主机名称需要重启生效。

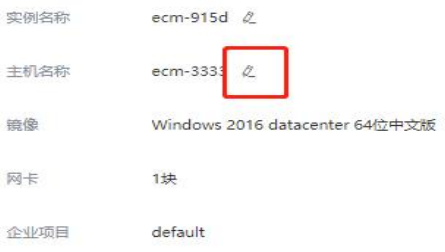
6. 单击“确定”，完成弹性云主机名称修改。

方法二：

1. 登录控制中心。
2. 单击控制中心顶部的 ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 单击在云主机列表页需要修改主机名的云主机蓝色实例名称，进入云主机详情页。



5. 单击主机名称的编辑标识，在输入框中输入要修改的主机名称。



6. 单击“确定”后，点击右上角的“重启”使编辑的主机名（hostname）生效。



注意

重启 ECS 实例必须通过控制台界面或调用 RebootInstance API 来完成，在操作系统内重启不能使修改后的新主机名生效。

7.1.4.3 批量设置有序的实例名称或主机名称

操作场景

购买多台云主机时，有以下两种方式设置有序的云服务器名称。

- 自动排序：购买多台云服务器时自动按序增加 4 位数字后缀，例如：我的云主机-0001。
- 自定义排序：按照 name_prefix{R，begin_number}name_suffix 格式为多台云服务器设置有序的名称。例如：my_name{R:99}ecs ，若创建 2 台实例最终生成的名称分别为：my_name0099ecs、my_name0100ecs，begin_number 取值区间为 [1,9799]。

  本节操作分别介绍两种命名方式的使用方法。

  自动排序

  名称可自定义，但需符合命名规则。

  实例名称：

  - 支持长度为 2-63 字符。
  - 支持{R:数字}形式，且只支持使用 1 次，此处数字需为小于等于 9799 的正整数，不支持冒号“:”以及大括号“{}”两类字符单独存在或其它组合方式。

  主机名称：

  Windows 系统

  - 2～15 个字符，大小写字母、数字或连字符（-）组成。
  - 禁止以连字符（-）开头或结尾，禁止连续使用连字符（-）。
  - 禁止使用纯数字。

• 支持{R:数字}形式，且只支持使用 1 次，此处数字需为小于等于 9799 的正整数，不支持冒号“:”以及大括号“{}”两类字符单独存在或其它组合方式。

其他操作系统(如 linux)

- 2～64 个字符，大小写字母、数字或连字符（-）组成。
- 允许使用点号（.）分段，每段包含小写字母、数字或连字符。
- 禁止连续使用连续使用点号（.）或连字符（-）。
- 禁止以点号（.）或连字符（-）开头或结尾。
- 禁止使用纯数字。

• 支持{R:数字}形式，且只支持使用 1 次，此处数字需为小于等于 9799 的正整数，不支持冒号“:”以及大括号“{}”两类字符单独存在或其它组合方式。

自动排序说明

一次创建多台弹性云服务器时，系统自动增加 4 位数字后缀。例如:输入 ecs，云主机名称为 ecs-0001，ecs-0002，……。再次创建多台云服务器时，命名从本次最小值连续增加，例如：输入 ecs，已有云服务器 ecs-0010，新创云主机名称为 ecs-0001、ecs-0002、……。

允许重名：允许创建的云服务器名称相同。

- 示例 1:创建时输入云主机名称 ecs-f526。云主机名称依次为 ecs-f526-0001、ecs-f526-0002、ecs-f526-0003...
- 示例 2：创建时输入云主机名称 ecs-f526，若已有 ecs-f526-0010，仍从 ecs-f526-0001 开始命名。云主机名称依次为 ecs-f526-0001、ecs-f526-0002、ecs-f526-0003...

自定义排序

按照 name_prefix{R，begin_number}name_suffix 格式为多台云服务器设置有序的名称。例如：my_name{R:99}ecs ，若创建 2 台实例最终生成的名称分别为：my_name0099ecs、my_name0100ecs，begin_number 取值区间为[1,9799]。

字段名称	是否固定	说明	示例
name_prefix	否	云主机名称前缀	ecs
R	是	用于激活自定义排序	R
begin_number	否	云主机名称的开始数字，只能为 1 到 9799 的整数。	0063
name_suffix	否	云主机名称后缀。	f405

自定义排序示例

示例：创建时输入云主机名称使用自定义排序，输入 name_prefix {R，9} name_suffix。云服务器名称依次为 name_prefix0009name_suffix、name_prefix0010name_suffix、name_prefix0011name_suffix...

7.1.4.4 修改实例描述

操作场景

创建云主机后，如果云主机描述需要更新或编辑，可以为云主机修改描述。

操作步骤

1. 登录控制中心。

2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在弹性云主机列表中，找到您想要变更规格的云主机，单击云主机操作栏的"编辑云主机属性"按钮。
5. 在“编辑云主机属性”弹窗中的“云主机描述”栏对云主机描述进行编辑。
6. 确认无误后，点击“确认”按钮。
7. 返回云主机列表，修改云主机描述操作完成。

7.1.4.5 更改实例时区

操作场景

弹性云主机默认时区设置取决于您制作镜像时选择的时区。如果您需要修改弹性云主机的时区，按照本节内容进行操作，将系统时间与本地时间同步或更改为其他所需的时区。

更改 Linux 弹性云主机的时区

更改 Linux 弹性云主机的时区可以通过修改系统配置文件来实现。以下是在 CentOS 中更改时区的详细操作方法：

1. 登录到您的 CentOS 弹性云主机：使用您的凭据登录到弹性云主机。
2. 打开终端或使用远程登录工具：获取访问弹性云主机的命令行界面。
3. 查看当前时区：您可以使用以下命令来查看当前的时区设置：

```
timedatectl
```

您可以使用以下命令列出系统中可用的时区：

```
timedatectl list-timezones
```

4. 使用以下命令将新时区应用到系统中，将“Your_Time_Zone”替换为您要设置的时区：

```
sudo timedatectl set-timezone Your_Time_Zone
```

例如“Asia/Shanghai”：



5. 保存更改并退出编辑器：在 Vi 编辑器中，按下 Esc 键，然后输入":wq"保存更改并退出编辑器。

```
:wq
```

6. 使用以下命令再次运行 timedatectl 来验证时区是否已更改：

```
timedatectl
```

系统将显示当前日期和时间，确保显示的时间与您所设置的时区一致，如下图示。



7. 至此，您已完成 CentOS 中更改时区。

更改 Windows 弹性云主机的时区

1. 使用您的凭据登录到您的 Windows 弹性云主机，本次以 Windows Server 2016 为例。
2. 点击 Windows 开始菜单，选择“控制面板”。
3. 在控制面板中，选择“日期和时间”或“时钟和区域”（具体名称可能会略有不同）。
4. 在日期和时间设置页面，点击“更改时区”按钮。
5. 在时区设置对话框中，选择适合您所在地区的时区。您可以通过从下拉列表中选择时区，或者在搜索框中键入相关的地区名称进行搜索。

6. 在选择了适当的时区后，点击“确定”按钮以保存更改。
7. 打开系统的日期和时间显示，或者运行命令提示符（CMD）并输入"date"命令，确保显示的时间与您所设置的时区一致。
8. 至此，您已在 Windows 弹性云主机上成功更改时区。

7.1.4.6 开启和关闭实例保护功能

如果您购买的主机计费类型为按需计费。那么您可以为您的主机开启删除保护功能，尤其是对于承载了关键业务应用或数据的主机，通过使用该功能，可以避免由于操作不当，导致主机被误删。该功能将从天翼云控制台及 OpenAPI 两个维度进行限制，以下将详细讲述如何操作。

前提条件

购买计费类型为按需付费的云主机

注意事项

以下情况，删除保护功能将不生效

- 用户账号欠费后，已过保留期，资源将进行销毁；
- 已加入弹性伸缩内的主机不受该功能影响；不影响弹性伸缩逻辑；
- 已开启实例删除保护功能的主机按需转包周期后，功能失效。

开启/关闭删除保护功能操作步骤

1. 点击云主机控制台，进入云主机列表页面。
2. 点击"操作 > 更多"，选择"开启删除保护功能"或"关闭删除保护功能"。

   注：实例默认为未开启删除保护功能。
3. 点击"确认"，则开启/关闭删除实例保护功能。

   开启实例删除保护功能后，实例"删除"功能将无法操作。

查询删除保护功能状态操作步骤

1. 点击云主机控制台，进入云主机列表页面。
2. 点击目标云主机名称，进入云主机详情页。
3. 查看详情页字段"实例删除保护"，显示当前实例开启关闭状态。

7.1.4.7 配置主机名与 IP 地址的映射

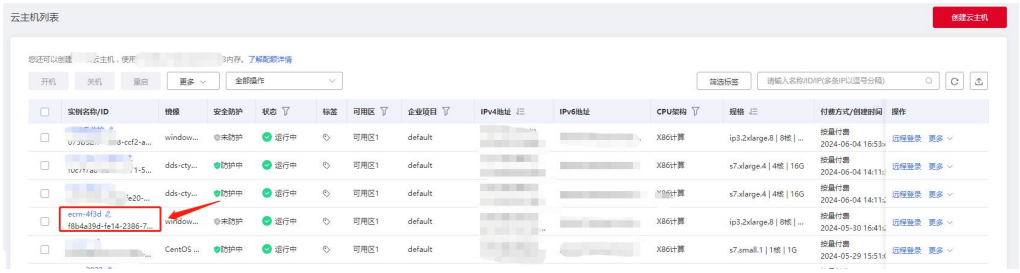
约束限制

仅适用于同一 VPC 内 Linux 弹性云主机 。

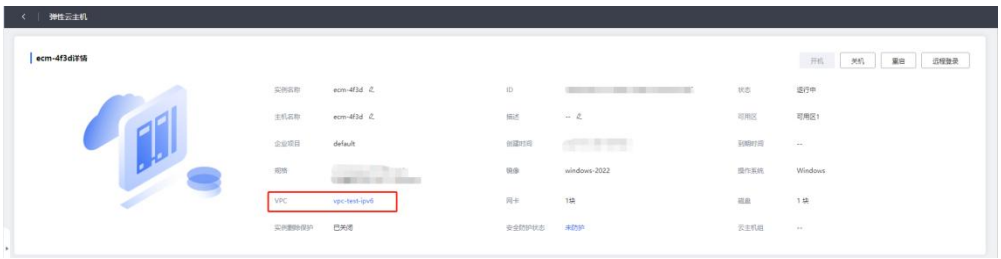
可以通过如下方式查看弹性云主机的 vpc：

a. 登录云主机控制台。

b. 点击实例。



c. 查看 vpc。



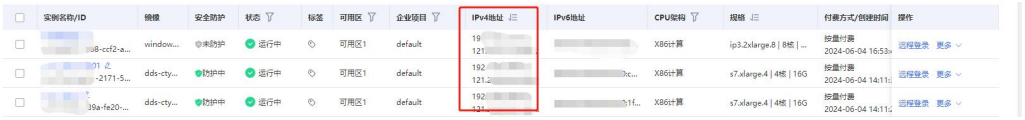
操作步骤

假设同一 VPC 内共有 2 台弹性云主机：ecm-53a1 和 ecm-9d59。通过如下操作，2 台弹性云主机可以通过主机名互相通信。

1. 获取 2 台弹性云主机的 IP 地址。

a. 登录云主机控制台。

b. 在云主机列表中查看云主机的 IPv4 地址。



【示例】假设查询的 IP 地址如下：

ecm-53a1：10.0.0.4

ecm-9d59：10.0.0.5

2. 分别获取 2 台弹性云主机的主机名。

a. 登录弹性云主机。

b. 执行以下命令，查询弹性云主机的主机名。

```
sudo hostname
```

【示例】假设查询的主机名如下：

ecm-53a1：ecm-53a1

ecm-9d59：ecm-9d59

3. 建立主机名与 IP 地址之间的映射关系，并添加同一 VPC 内其他弹性云主机的信息。

a. 登录弹性云主机 ecm-53a1。

b. 执行以下命令，切换至 root 权限。

```
sudo su -
```

c. 执行以下命令，编辑 hosts 文件。

```
vi /etc/hosts
```

d. 按“ i ”，进入编辑模式。

e. 按照如下格式添加语句，建立映射关系。

IP 地址 主机名

【示例】需添加的语句为：

10.0.0.4 ecm-53a1

10.0.0.5 ecm-9d59

f. 按“Esc”退出编辑模式，并执行以下命令，保存并退出。

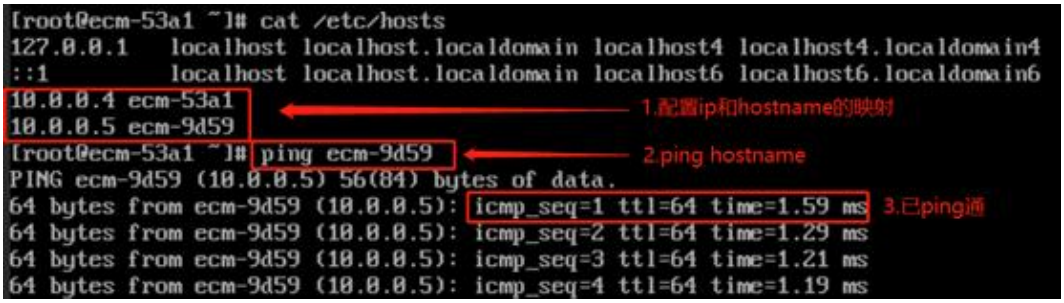
```
:wq
```

g. 登录 ecm-9d59，重复执行 3.b~3.f。

4. 测试验证能否通过主机名正常通信。

分别登录配置好的的弹性云主机，执行以下命令，看 ping 包是否能正常送达。如有下图所示的结果，则表示已经可以通过主机名正常通信。

```
ping hostname
```



7.1.4.8 实例自定义数据注入

操作场景

当您需要通过脚本简化弹性云主机配置、通过脚本初始化系统、在创建弹性云主机时上传已有脚本，或者完成其他需要使用脚本的功能时，可以考虑使用实例自定义数据注入功能来配置弹性云主机。

使用限制

Windows

- 用于创建弹性云主机的镜像必须安装了 Cloud-init 组件。
- 实例自定义数据大小限制：小于等于 16KB。

- 如果通过文本方式上传实例自定义数据，数据只能包含 ASCII 码字符；如果通过文件方式上传实例自定义数据，只能上传 text/*格式的文件，且文件大小小于等于 16KB。
- 必须使用公共镜像、继承于公共镜像的私有镜像，或自行安装了 Cloud-init 组件的私有镜像。
- 实例自定义脚本必须满足相应弹性云主机自定义脚本类型的格式要求。

Linux

- 用于创建弹性云主机的镜像必须安装了 Cloudbase-init 组件。
- 实例自定义数据大小限制：小于等于 16KB。
- 如果通过文本方式上传实例自定义数据，数据只能包含 ASCII 码字符；如果通过文件方式上传实例自定义数据，可以包含任意字符，同时，要求文件大小小于等于 16KB。
- 必须是公共镜像，或继承于公共镜像的私有镜像，或自行安装了 Cloudbase-init 组件的私有镜像。

实例自定义数据使用方法

1. 根据实际需要写 User-Data 脚本，脚本格式必须符合 Cloud-init/Cloudbase-init 官网相关规范。
2. 在创建弹性云主机时，在“高级配置 > 用户数据 > 立即配置 > 以文本形式”的文本框中粘贴实例自定义数据脚本内容，或选择以文件形式上传。

   说明

   实例自定义数据注入支持两种方式：

   - 文本：将实例自定义数据脚本内容粘贴到文本框中。
   - 文件：将实例自定义数据脚本内容保存至文本地文件后再上传。
3. 系统创建弹性云主机后，弹性云主机将自动执行 Cloud-init 或 Cloudbase-init 组件，读取实例自定义数据脚本。

关于 Windows 弹性云主机的实例自定义数据脚本

Windows 弹性云主机的实例自定义数据脚本使用开源的 cloudbase-init 架构实现（详情请参考 Cloud-init 开源说明），它利用弹性云主机的元数据

为数据来源，实现对弹性云主机的自动化配置和初始化脚本执行。下面是关于 Windows 弹性云主机实例自定义数据脚本的详细说明和使用示例：

- 脚本类型：包括“Batch 批处理程序”和“PowerShell 脚本”。

-	Batch 批处理程序	PowerShell 脚本
格式	以“rem cmd”开头且作为首行，前面不能有空格。	在采用 Base64 编码前，以“#ps1”开头且作为首行，前面不能有空格。
限制	在采用 Base64 编码前，脚本内容（包括首行）不能超过 16KB。	在采用 Base64 编码前，脚本内容（包括首行）不能超过 16KB。

- 查看注入 Windows 弹性云主机的自定义数据：

1. 登录弹性云主机。
2. 执行以下命令，使用 Administrator 用户权限查看实例自定义数据：

```
http://169.254.169.254/openstack/latest/user_data
```

- 脚本使用示例：

  该示例介绍如何以不同格式输入 Windows 弹性云主机的脚本，并查看脚本的运行结果。

示例：PowerShell 脚本

1. 创建弹性云主机时，选择"文本" 形式，并输入以下实例自定义数据脚本：

```
#ps1echo "Hello, Powershell Test" > C:\aaaa.txt
```

2. 创建成功后，连接并启动弹性云主机，在 C:\盘中查看脚本的运行结果：将创建一个名为 "aaaa" 的文本文档。
3. 要查看注入到 Windows 弹性云主机的实例自定义数据，可以在浏览器中访问：

```
http://169.254.169.254/openstack/latest/user_data
```

关于 Linux 云主机的实例自定义数据脚本

Linux 弹性云主机的实例自定义数据脚本采用开源的 Cloud-init 架构实现（详情请参考 Cloud-init 开源说明），使用云主机的元数据作为数据

来源，用于自动化配置。下面是关于 Linux 弹性云主机实例自定义数据脚本的详细说明和使用示例：

- 脚本的执行时间：在弹性云主机状态变为"running" 之后，但在执行 /etc/init 文件之前。
- 脚本类型：实例自定义数据脚本支持两种类型，即"User-Data Script" 和 "Cloud-Config Data"。

-	User-Data Script	Cloud-Config Data
描述	使用脚本作为自定义配置的手段，可以是 shell 脚本或 Python 脚本。	使用 Cloud-init 预定义的方式配置云主机的某些服务，例如 yum 源、SSH 密钥等。
格式	首行必须以"#!" 开头（例如 "#!/bin/bash" 或 "#!/usr/bin/env python"），前面不能有空格首次启动时，以 rc.local-like 级别执行，rc.local-like 意味着在启动序列中较后面的位置。	首行必须以 "#cloud-config" 开头，前面不能有空格
限制	在采用 Base64 编码前，脚本内容（包括首行）不能超过 16KB。	在采用 Base64 编码前，脚本内容（包括首行）不能超过 16KB。
频率	仅在首次启动弹性云主机时执行一次。	配置的服务不同，执行频率也会不同。

- 查看注入的实例自定义数据:

1. 登录弹性云主机。
2. 以 root 用户权限执行以下命令来查看实例自定义数据：

```
curl http://169.254.169.254/openstack/latest/user_data
```

- 脚本使用示例：

  该示例介绍如何以不同格式输入 Linux 弹性云主机的脚本，并查看脚本的运行结果。

  示例一：脚本类型为 User-Data Script

1. 创建弹性云主机时，选择"文本" 形式，并输入以下实例自定义数据脚本：

```
#!/bin/bashecho "Hello, the time is now $(date -R)" | tee
/root/output.txt
```

2. 创建成功后，连接并启动弹性云主机，执行命令 cat /root/output.txt 查看脚本的运行结果：

```
[root@XXXXXXXX ~]# cat /root/output.txt

Hello, the time is now Mon, 03 Jul 2023 14:49:08+0800
```

  示例二：脚本类型为 Cloud-Config Data

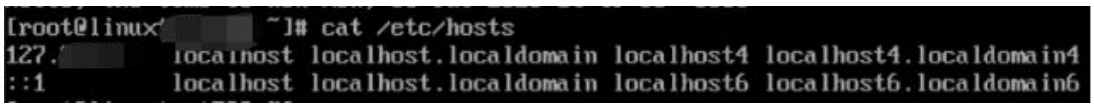
1. 创建弹性云主机时，选择"文本" 形式，并输入以下实例自定义数据脚本：

```
#cloud-config

bootcmd:

- echo 192.168.1.130 us.archive.ubuntu.com >> /etc/hosts
```

2. 创建成功后，连接并启动弹性云主机，执行命令 cat /etc/hosts 查看脚本的运行结果。



  案例 1

  以下是通过实例自定义数据注入来重置 Linux 弹性云主机密码的示例。

  假设您要将 root 用户的密码重置为”******“，新密码必须符合密码复杂度要求，如下表 1 密码设置规则所示：

表 1 密码设置规则

参数	规则	样例
密码	• 密码长度范围为 8 到 26 位。 • 密码至少包含以下 4 种字符中的 3 种： 大写字母 小写字母 数字 Windows 操作系统云主机特殊字符：包括“$”、“!”、“@”、“%”、“-”、“ ”、“=”、“+”、“[”、“]”、“:”、“.”、“/”、“,”和“?” Linux 操作系统云主机特殊字符：包括“!”、“@”、“%”、“-”、“”、“=”、“+”、“[”、“]”、“:”、“.”、“/”、“^”、“,”、“{”、“}”和“?” • 密码不能包含用户名或用户名的逆序。 • Windows 操作系统的云主机，不能包含用户名中超过两个连续字符的部分。	UTX8@81JCA.1d8sd

请按照以下示例注入实例自定义数据（请严格遵守下面示例中的缩进标准）：

```
#cloud-configchpasswd:

 list: |

   root:******  expire: False
```

创建弹性云主机成功后，您可以使用重置的密码登录弹性云主机。为了确保密码安全，建议您在第一次登录弹性云主机后修改 root 用户的密码。

案例 2

以下是通过实例自定义数据注入来为Windows弹性云主机新增用户并设置密码的示例。

假设您要创建的用户名为abc，密码为"******"的普通用户，并将其添加到administrators用户组。

说明
新密码必须符合密码复杂度要求，如表1密码设置规则。

请按照以下示例注入实例自定义数据：

```
rem cmdnet user abc ****** /addnet localgroup administrators abc /add
```

创建弹性云主机成功后，您可以使用新创建的用户名和密码登录弹性云主机，请确保密码符合密码复杂度要求。

案例3

以下是通过实例自定义数据注入来为Linux弹性云主机更新系统软件包并启动httpd服务的示例。

请按照以下示例注入实例自定义数据：

```
#!/bin/bash

yum update -y

service httpd start

chkconfig httpd on
```

注入成功后，您的弹性云主机将会更新系统软件包，并启动httpd服务。您可以通过访问弹性云主机的公网IP地址来使用httpd服务。请确保您的操作系统支持yum包管理器，并且已经安装了httpd服务。

案例4

以下是通过实例自定义数据注入来激活Linux弹性云主机的root用户远程登录权限的示例。

实例自定义数据注入示例：

```
#cloud-config

disable_root: false

runcmd:

- sed -i 's/^PermitRootLogin.*$/PermitRootLogin without-password/' /etc/ssh/sshd_config

- sed -i '/^KexAlgorithms.*$/d' /etc/ssh/sshd_config

- service sshd restart
```

注入成功后，您将能够使用 SSH 密钥方式以 root 帐户登录弹性云主机。该示例中的自定义数据使用 cloud-config 格式，并通过 runcmd 字段执行了一系列命令，包括修改/etc/ssh/sshd_config 文件中的配置，重启 sshd 服务使更改生效。

说明
激活 root 用户远程登录权限可能存在安全风险，建议在完成必要的任务后再禁用 root 用户的远程登录权限，以提高系统的安全性。

7.1.4.9 实例自定义元数据

云主机的元数据（metadata）是描述云主机信息的数据，包含了天翼云系统中的默认元数据信息与用户自定义的元数据信息。在云主机内部，通过内网可以方便地查看该部分信息，并基于元数据信息配置或管理云主机。

注意
本文仅适用于多可用区资源池，非多可用区资源池暂不可参考本文。

云主机默认元数据

云主机创建过后，天翼云云主机会含有默认的元数据，以下为多可用区类型资源池下默认的元数据：

元数据说明	元数据项

元数据说明	元数据项
拥有者天翼云账号 ID	/meta-data/owner-account-id
云主机 ID	/meta-data/instance-id
云主机名称	/meta-data/hostname
云主机规格名称	/meta-data/flavor_name
云主机的资源池 ID	/meta-data/region-id
云主机的可用区 ID	/meta-data/zone-id
云主机的自定义数据	/meta-data/user_data
创建云主机时所使用的镜像 ID	/meta-data/image-id
云主机所属 VPC 的 ID	/meta-data/vpc-id
云主机主网卡的内网 IPv4 地址	/meta-data/private-ipv4
网卡的 mac 地址列表	/meta-data/network/interfaces/macs/

云主机自定义元数据

除了上述云主机元数据项，用户也可以增加相应的自定义元数据项。在云主机 openapi 元数据管理中，api 对云主机操作的就是该部分元数据内容。

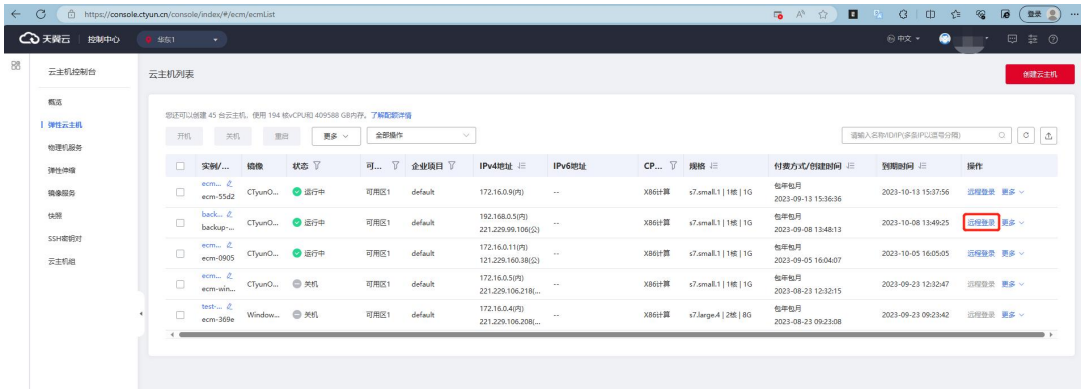
元数据说明	元数据项
一次获取所有自定义元数据(返回格式为	/spec-meta-data

元数据说明	元数据项
json）	
单一自定义元数据获取	/meta-data/instance/specs/[key]

元数据的使用

查看默认元数据

通过 console 云主机列表界面选择对应云主机进行远程登录。



请求查询元数据的内网地址为　169.254.169.254。

我们可以根据上述默认元数据的表格，对所需元数据进行查看，对于 Windows 操作系统，使用用网页 http 访问；对于 Linux 操作系统，我们可以使用 curl 命令进行查看，命令组成为：curl 169.254.169.254{元数据项}。

示例 shell 脚本：

```
#!/usr/bin/bash

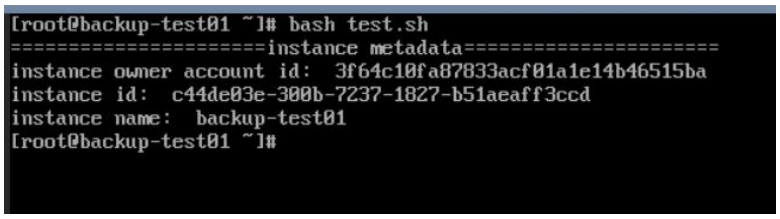
echo "======================instance metadata======================"

echo "instance owner account id: " $(curl -s
169.254.169.254/meta-data/owner-account-id)

echo "instance id: " $(curl -s 169.254.169.254/meta-data/instance-id)
```

```
echo "instance id: " $(curl -s 169.254.169.254/meta-data/instance-id)
```

效果：



云主机元数据 openapi 的使用

目前，天翼云提供的元数据 openapi 提供对云主机自定义元数据进行操作与查看的能力，主要分为创建、更新、查询和删除。通过 python 示例文档　与 openapi 对应的文档，可以写出对应的 openapi 请求。

1、创建自定义元数据

首先，我们通过 openapi 创建元数据 openapi 文档来创建自定义元数据（{"test1“: "123", "test2": "abc"}）。

```
# 按照创建云数据的 openapi 文档填写相关参数
create_metadata_params = {
    "regionID": huadong1_region_id,
    "azName": huadong1_az1,
    "instanceID": huadong1_instance,
    "metadata": {
        "test1": "123",
        "test2": "abc"
    }
}
# 这里省略 post 方法的定义，该方法的实现请参考天翼云官网文档中的
url = "https://ctecs-global.ctapi.ctyun.cn/v4/ecs/metadata/create"
```

```
res_metadata_create = post(url, params=create_metadata_params)
```

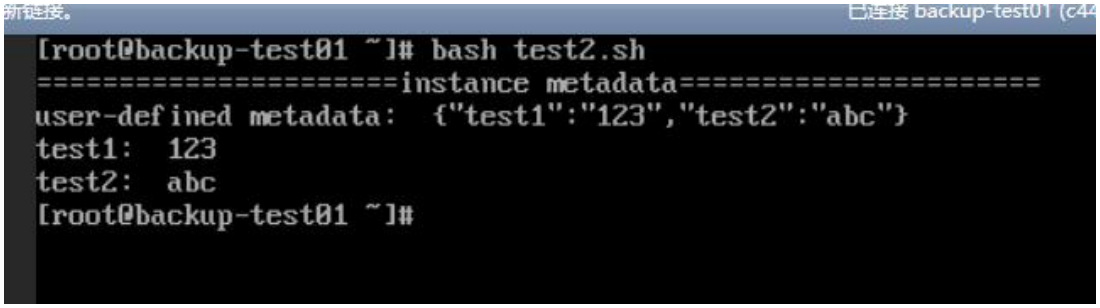
调用结果：

```
{
    "returnObj": {
        "instanceID": "c44de03e-300b-7237-1827-b51aeaff3ccd",
        "metadata": {
            "test1": "123",
            "test2": "abc"
        }
    },
    "details": "",
    "message": "SUCCESS",
    "description": "成功",
    "statusCode": 800
}
```

通过之前的方式，使用自定义元数据项来查询自定义的元数据信息（test2.sh脚本如下）：

```
#!/usr/bin/bash
echo "======================instance metadata======================"
echo "user-defined metadata: " $(curl -s 169.254.169.254/spec-meta-data)
echo "test1: " $(curl -s
169.254.169.254/meta-data/instance/specs/test1)
echo "test2: " $(curl -s
169.254.169.254/meta-data/instance/specs/test2)
```

结果：



2、查询自定义元数据

同时，我们也可以调用 openapi 接口来查询自定义元数据。

```
res_metadata_details =
get("https://ctecs-global.ctapi.ctyun.cn/v4/ecs/metadata/details",
params=dict(regionID=huadong1_region_id, azName=huadong1_az1,
instanceID=huadong1_instance))
```

调用结果：

```
{
    "returnObj": {
        "metadata": {
            "test1": "123",
            "test2": "abc"
        }
    },
    "message": "SUCCESS",
    "description": "成功",
    "statusCode": 800
}
```

3、更新自定义元数据

调用更新元数据接口：

```
# 按照创建云数据的 openapi 文档填写相关参数

update_metadata_params = {

    "regionID": huadong1_region_id,

    "azName": huadong1_az1,

    "instanceID": huadong1_instance,

    "metadata": {

        "test1": "789"

    }

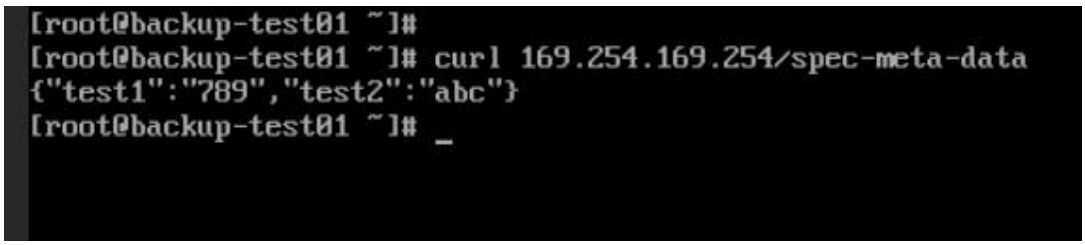
}

# 这里省略 post 方法的定义，该方法的实现请参考天翼云官网文档中的

url = "https://ctecs-global.ctapi.ctyun.cn/v4/ecs/metadata/update"

res_metadata_update = post(url, params=update_metadata_params)
```

在云主机内部进行 curl 查看情况：



*覆盖方式更新（isForce=true）

```
# 按照创建云数据的 openapi 文档填写相关参数

update_metadata_params = {

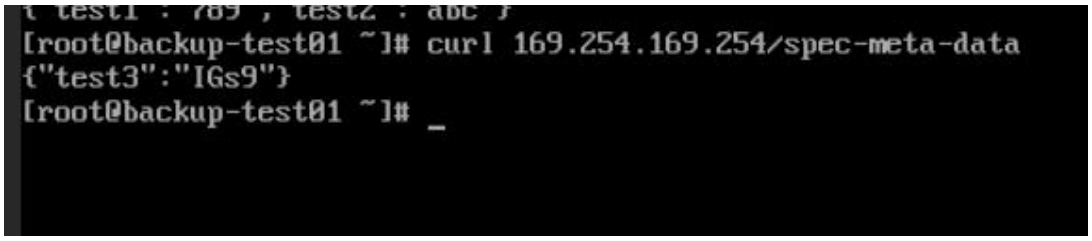
    "regionID": huadong1_region_id,

    "azName": huadong1_az1,

    "instanceID": huadong1_instance,
```

```
    "isForce": True,
    "metadata": {
        "test3": "IGs9"
    }
}
# 这里省略 post 方法的定义，该方法的实现请参考天翼云官网文档中的
url = "https://ctecs-global.ctapi.ctyun.cn/v4/ecs/metadata/update"
res_metadata_update = post(url, params=update_metadata_params)
```

在云主机内部进行 curl 查看情况：

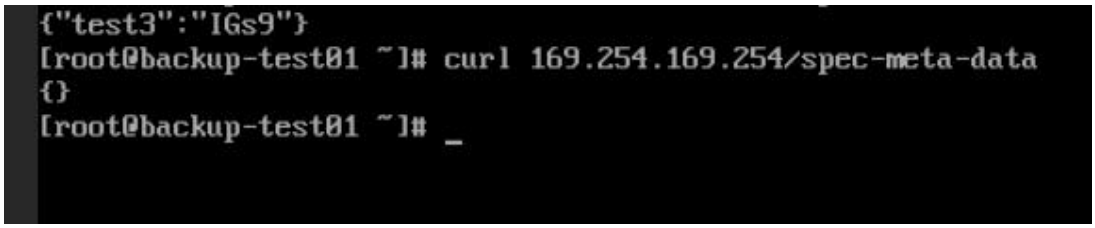


4、删除自定义元数据

调用删除接口；

```
res_metadata_delete =
post("https://ctecs-global.ctapi.ctyun.cn/v4/ecs/metadata/delete",
params=dict(regionID=huadong1_region_id, azName=huadong1_az1,
instanceID=huadong1_instance))
```

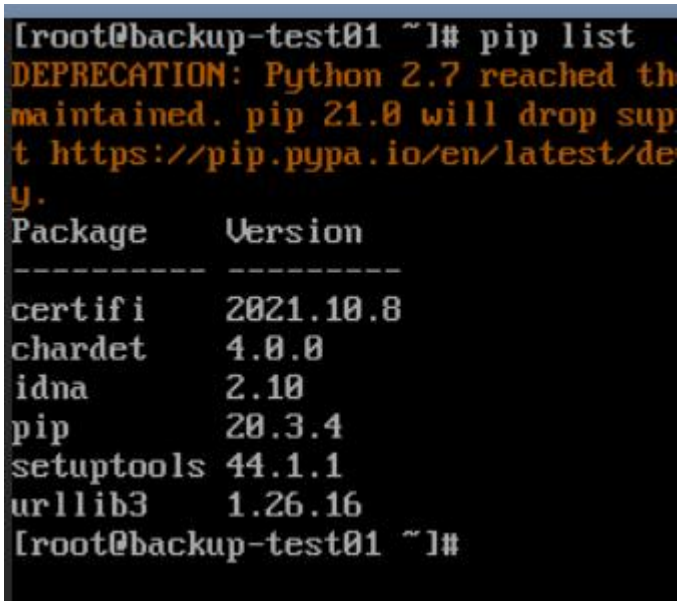
在云主机内部进行 curl 查看情况：



示例：通过元数据配置云主机的 python 环境

场景：在使用 python 过程中，我们经常需要安装很多依赖包，并且每个包之间互相有版本限制（比如在深度学习框架中，Keras 与 TensorFlow 之间需要版本对应）。此时我们可以通过创建元数据，来记录该云主机所需要的 python 包环境的版本信息。

首先，console 界面远程登录云主机，进入远程界面后，查看当前 python 包信息：



我们需要将关键的包的版本信息写入。

这里我们以部分包为例，执行 openapi 创建云主机的自定义元数据：

```
# 按照创建云数据的 openapi 文档填写相关参数

create_metadata_params = {

    "regionID": huadong1_region_id,

    "azName": huadong1_az1,

    "instanceID": huadong1_instance,

    "metadta": {

        "requeirements": {

            "requests": "2.27.1",
```

```
            "numpy": "1.16.6",
            "matplotlib": "2.2.5"
        }
    }
}
# 这里省略 post 方法的定义，该方法的实现请参考天翼云官网文档中的[python 示例文档](https://www.ctyun.cn/document/10026730/10044097)
url = "https://ctecs-global.ctapi.ctyun.cn/v4/ecs/metadata/create"
res_metadata_create = post(url, params=create_metadata_params)
```

执行结果：

```
{
    "returnObj": {
        "instanceID": "c44de03e-300b-7237-1827-b51aeaff3ccd",
        "metadata": {
            "requeirements": {
                "requests": "2.27.1",
                "numpy": "1.16.6",
                "matplotlib": "2.2.5"
            }
        }
    },
    "details": "",
    "message": "SUCCESS",
    "description": "成功",
```

```
    "statusCode": 800
}
```

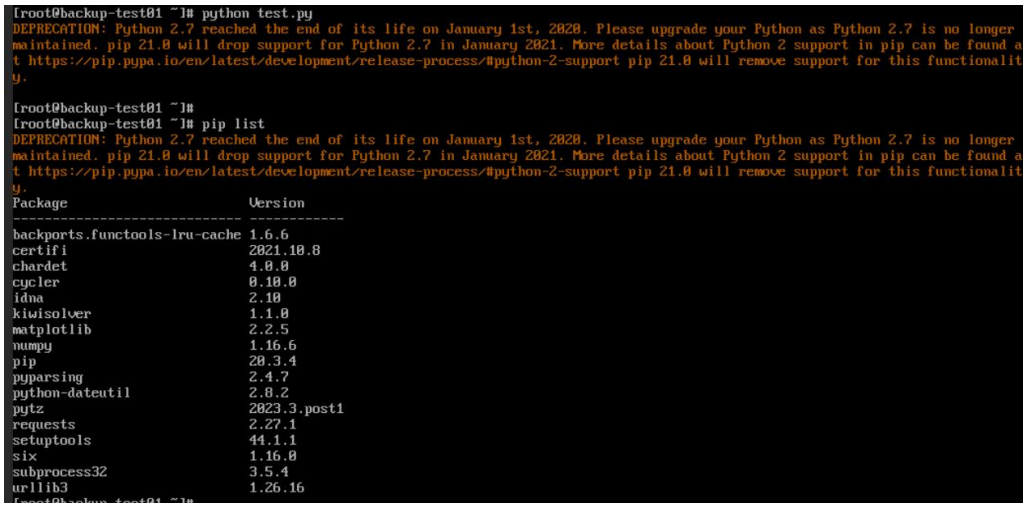
在创建过元数据后，我们可以在云主机内使用相关脚本，对 python 进行相应的配置。

python 示例脚本（test.py）：

```
import jsonimport subprocess
HOST = "169.254.169.254"
URL = HOST + "/spec-meta-data"
FILE_REQUEIRE = "requirements.txt"
def send_command(command):
    command_ls = command.split()
    process = subprocess.Popen(command_ls, stdout=subprocess.PIPE)
    output, err = process.communicate()
    return output.decode("utf-8")
if __name__ == "__main__":
    res = send_command("curl -s %s" % URL)
    metadata = json.loads(res)
    requeire_info = metadata.get("requeirements")
    with open(FILE_REQUEIRE, "w") as f:
        for k, v in requeire_info.items():
            write_info = "%s==%s" % (str(k), str(v))
            f.writelines(write_line)
    tsinghua_image = "https://pypi.tuna.tsinghua.edu.cn/simple" # 清华镜像站
```

```
    send_command("pip install -r %s -i %s" % (FILE_REQUEIRE,
tsinghua_image))
```

执行结果：



可以看到，云主机的 python 环境中已经下载了对应的 python 包。在云主机出现问题时，后续也可以根据对应的 python 包信息进行相应的再次部署。

7.1.5 管理实例状态

7.1.5.1 开机实例

操作场景

开机操作可以将弹性云主机从关机状态切换为运行状态，以便正常运行应用程序和服务。

对单台云主机进行开机的操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在云主机列表中，使用搜索功能输入云主机的名称、ID 或 IP 地址以定位目标云主机。

5. 选择目标云主机，点击云主机列表左上角的开机操作按钮。
6. 在弹出的提示信息中，确认操作是否正确。

注意

请注意云主机状态的说明。如果云主机在中间状态停留超过 30 分钟（如启动中、停止中、重启中），表示可能出现异常情况，请及时提交工单以寻求进一步处理。

对多台云主机进行开机的操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在云主机列表中，选择需要开机的多台云主机。
5. 点击云主机列表左上角的开机操作按钮。
6. 在弹出的提示信息中，确认操作是否正确。

注意

请注意云主机状态的说明：如果云主机在中间状态停留超过 30 分钟（如启动中、停止中、重启中），表示可能出现异常情况，请及时提交工单以寻求进一步处理。

7.1.5.2 关机实例

操作场景

关机操作可以将弹性云主机从运行状态切换为关机状态，以进行维护等操作。

使用须知

- 按需付费的弹性云主机支持节省关机和普通关机两种计费模式（节省关机目前仅部分资源池支持）。

- 节省关机模式下，计算资源（CPU、内存、GPU)不再收费，其余收费的资源正常计费，如系统盘、数据盘、带宽等，且再次开机后不会导致内网IP 或弹性公网 IP 地址变更。
- 由于计算资源被回收，节省关机模式下，再次开机时可能因为资源不足导致启动失败，您可以稍后尝试再次开机或者尝试变配为其他规格，如一直没有可用资源，云主机有一直不能开机的风险。
- 如下场景不支持节省关机

✧ 所开通的云主机为包周期计费模式

✧ 所开通的云主机为本地盘云主机

✧ 已经加入主机组的云主机

✧ 已经加入云主机快照、云主机备份策略的云主机

注意

按量付费云主机的计费规则详见：按量计费模式-弹性云主机-购买指南-计费模式 - 天翼云（ctyun.cn）

操作步骤

1. 登录控制中心。

2. 单击控制中心顶部的  ，选择“地域”。

3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。

4. 根据实际需求对云主机进行关机。

(1) 对单台云主机进行关机操作：在云主机列表中，选择需要开机的云主机，点击云主机列表左上角的开机操作按钮。

(2) 对多台云主机进行关机操作：在云主机列表中，选择需要开机的多台云主机，点击云主机列表左上角的开机操作按钮。

5. 在关机弹窗中，设置关机方式和模式，并确认操作是否正确，无误后点击确定。

设置	说明

设置	说明
关机方式	关机：正常关机流程。但为避免关机失败，当您选择关机时，普通关机超时会自动执行强制关机操作。 强制关机：等同于断电处理，可能丢失云主机操作系统中未写入磁盘的数据。
关机模式	普通关机模式：普通关机后保留计算、存储、网络资源。 节省关机模式：节省关机后，计算资源（CPU、内存、GPU)均被释放且不计费，其余网络、存储资源继续保留并收费。

注意

请注意云主机状态的说明：如果云主机在中间状态停留超过 30 分钟（如启动中、停止中、重启中），表示可能出现异常情况，请及时提交工单以寻求进一步处理。

7.1.5.3 重启实例

操作场景

重启操作是维护云主机的一种常用方式，如系统更新、重启保存相关配置等。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 根据实际需求对云主机进行重启。

   （1）对单台云主机进行重启操作：在云主机列表中，选择需要重启的云主机，点击云主机列表左上角的重启操作按钮。

（2）对多台云主机进行重启操作：在云主机列表中，选择需要重启的多台云主机，点击云主机列表左上角的重启操作按钮。

5. 在重启弹窗中，设置重启方式，并确认操作是否正确，无误后点击确定。

设置	说明
重启方式	重启：正常重启流程。 强制重启：等同于断电重启，可能丢失云主机操作系统中未写入磁盘的数据。

注意

请注意云主机状态的说明：如果云主机在中间状态停留超过 30 分钟（如启动中、停止中、重启中），表示可能出现异常情况，请及时提交工单以寻求进一步处理。

7.1.5.4 重装操作系统

操作场景

当弹性云主机的操作系统无法正常启动或需要进行优化以达到最佳状态时，用户可以通过重装操作系统的功能进行操作。

重装操作系统的功能允许用户更换当前使用的操作系统，以满足业务需求。这包括升级到更高版本的操作系统，以满足特定软件的要求。在天翼云中，您可以选择不同类型的镜像，包括公共镜像和私有镜像，并且可以在 Windows 到 Windows、Linux 到 Linux，以及 Windows 到 Linux 之间进行切换。这样，您可以灵活选择适合实际业务需求的操作系统。

重装须知

- 重装操作系统后，您的弹性云主机的 IP 地址和 MAC 地址将保持不变。

- 重装操作系统将清除系统盘上的所有数据，包括系统分区和其他所有分区。因此，请务必提前备份好重要的数据。
- 重装操作系统不会影响数据盘上的数据，您的数据盘中的信息将保持不变。
- 在重装操作系统后的几分钟内，系统将注入密码或密钥信息。在此期间，请不要执行任何其他操作，以避免导致密码或密钥信息注入失败，从而无法登录到您的云主机。请耐心等待注入过程完成。

约束与限制

- 如果您使用了私有镜像创建的弹性云主机，请确保原有镜像仍然存在，以便成功完成重装操作。
- 如果原来的云主机是按需计费模式，请确保您的账户余额足够支付相关费用。
- 如果原来的云主机是包周期计费模式，请确保该云主机仍在有效期内。
- 在一键重装操作系统之前，请确保先将云主机关机，或根据页面提示勾选"立即关机"选项。

重装单台操作系统

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在待重装操作系统的弹性云主机下方的"操作"列，点击"更多 > 一键重装"。
5. 根据界面提示根据业务需求选择“镜像类型”，“重装选择”进行操作系统及版本选择，并完成密码及用户配置。
6. 单击“确定”。
7. 一旦提交了重装系统的申请，弹性云主机的状态将变为"重建中"。当该状态消失后，表示重装操作已完成。

后续处理

如果重装操作系统失败，页面会显示重装操作系统失败的提示。您可以使用云服务平台提供的重试功能，重新执行第 3 至第 7 步，以尝试重装弹性云主机的操作系统。

如果重试仍然不成功，建议直接联系客服。客服人员将在后台进行人工恢复操作，解决问题。

重装多台操作系统

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在“云主机列表”左侧勾选要重装的多台云主机。
5. 云主机列表页面正上方，点击"更多"按钮，点击" 一键重装"
6. 根据界面提示根据业务需求选择“镜像类型”，“重装选择”进行操作系统及版本选择，并完成密码及用户配置。
7. 单击“确定”。
8. 一旦提交了重装系统的申请，弹性云主机的状态将变为"重建中"。当该状态消失后，表示重装操作已完成。

7.1.5.5 变更规格

操作场景

创建云主机后，如果当前配置无法满足您的业务需求，可以为云主机变更规格，即修改云主机的 vCPU 和内存配置。

变更须知

- 账号具有适当的访问权限以执行规格变更操作。

- 不同云主机规格的资源配额有差异，进行降配操作时，为了保持与目标规格的资源配额要求数一致，可能会要求您解绑一些资源，比如弹性 IP。
- 确保云主机为关机状态。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在弹性云主机列表中，找到您想要变更规格的云主机，单击云主机操作栏的"变配"按钮。
5. 在弹出的变更规格页面中，选择想要变更的目标规格，会根据您所选规格及云主机的状态进行判断和提示，比如只可冷变配至该规格，会提示进行关机操作。
6. 确认选择无误后，单击"确定"按钮。
7. 进入订单支付界面，确定配置、费用等项无误后，单击“立即支付”。
8. 完成支付过程后，等待规格变更操作完成。

变更规格可能出现的磁盘挂载问题

检查磁盘挂载状态

变更规格时，可能会发生磁盘挂载失败的情况，因此，变更规格后，需检查磁盘挂载状态是否正常。如果正常，则变更成功。

- Windows 弹性云主机详细操作请参考 Windows 云主机变更规格后磁盘脱机怎么办？
- Linux 弹性云主机详细操作请参考 Linux 云主机变更规格后磁盘脱机怎么办？

后续处理

如果变更规格失败，请在支持中心进行工单处理。天翼云会用最快时间帮您解决变更规格失败的问题。

7.1.5.6 创建相同配置实例

操作场景

对于已购买成功的弹性云主机，如需再次购买相同配置的，建议您基于已创建的云主机，使用“创建相同配置”功能，快速购买同一配置的弹性云主机，即主机规格、磁盘个数和大小、镜像等配置都相同的云主机。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择目标弹性云主机，并单击“操作”列下的“更多 > 创建相同配置”。
5. 在打开的主机订购页面中，会自动按照已选的云主机的参数配置好相同的参数（除公网弹性 IP 外），您可以直接订购，也可以修改参数。

   说明

   弹性 IP 信息需要单独配置。

7.1.5.7 克隆实例

操作场景

- 当您需要多个相似的云主机或环境时，可以克隆现有的主机来快速部署所需的配置、应用程序和设置，节省了配置时间和劳动成本。
- 如果主机出现故障或问题，克隆主机可以快速替代原始主机，减少服务中断时间。

- 在需要快速增加服务器以应对流量增加或负载均衡的情况下，通过克隆现有主机，可以更迅速地扩展服务器规模，以满足需求。

前提条件

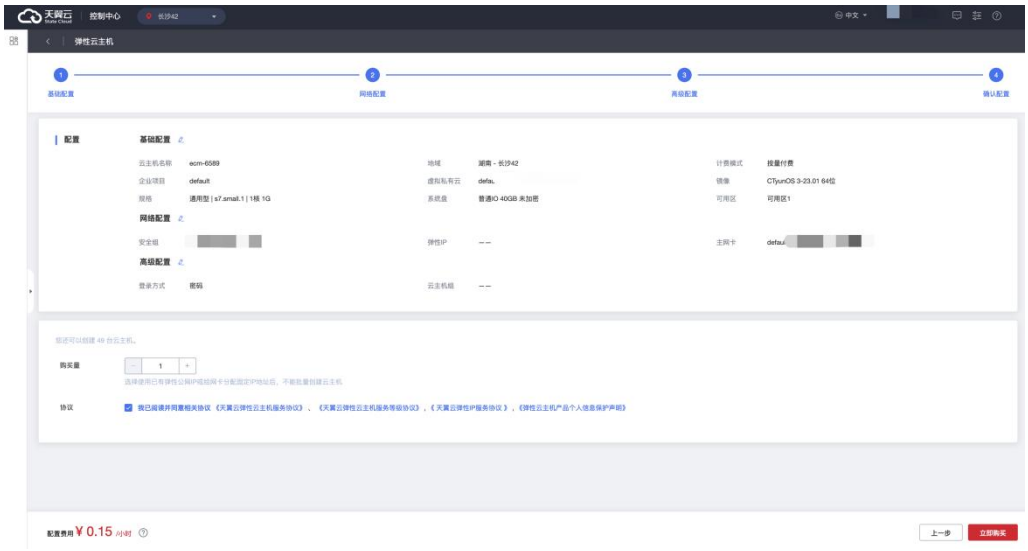
- 选择支持克隆云主机操作的区域。
- 待克隆的云主机处于“运行中”或“关机”状态，处于中间态的云主机无法进行克隆。
- 本地盘云主机不支持克隆。
- 其余类型云主机状态为创建中、停止中、启动中、重启中、规格变更中、重建中、快照创建中、备份创建中、按量删除中、已冻结、已到期、包周期退订中、包周期已退订、节省关机中、节省关机、错误 不支持克隆。

操作步骤

1. 登录控制中心；
2. 单击控制中心顶部的 ，选择“地域”；
3. 选择“计算 >云主机”；
4. 在待克隆的云主机的“操作”列下，单击“更多 >克隆云主机”；



5. 在克隆云主机界面，填写克隆云主机的相关信息后，单击“立即购买”，支付成功后完成克隆。



7.1.6 管理云主机组

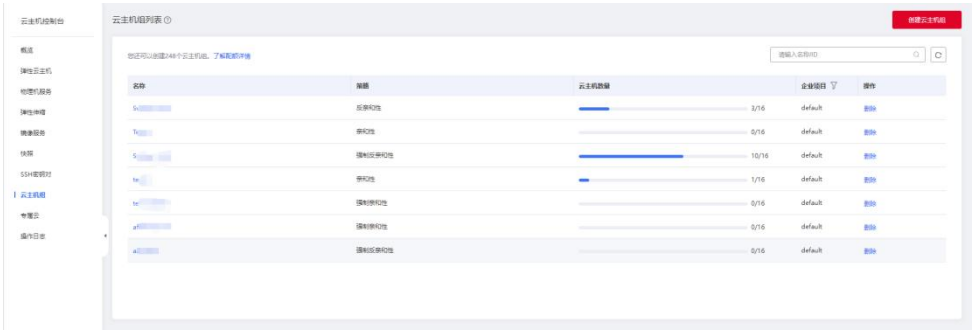
通过云主机组功能，云主机在创建时，将尽量分散地创建在不同的宿主机上，提高业务的可靠性，或尽量集中地创建在同一宿主机上，提升业务的性能。

新建主机组

如果没有任何主机组，需新建一个主机组，配置主机组策略。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“云主机组”。



5. 点击“创建云主机组”。



6. 确定主机组名称：只能由中文、英文字母、数字、下划线、中划线、点组成；长度：2-64 位。
7. 选择主机组策略：

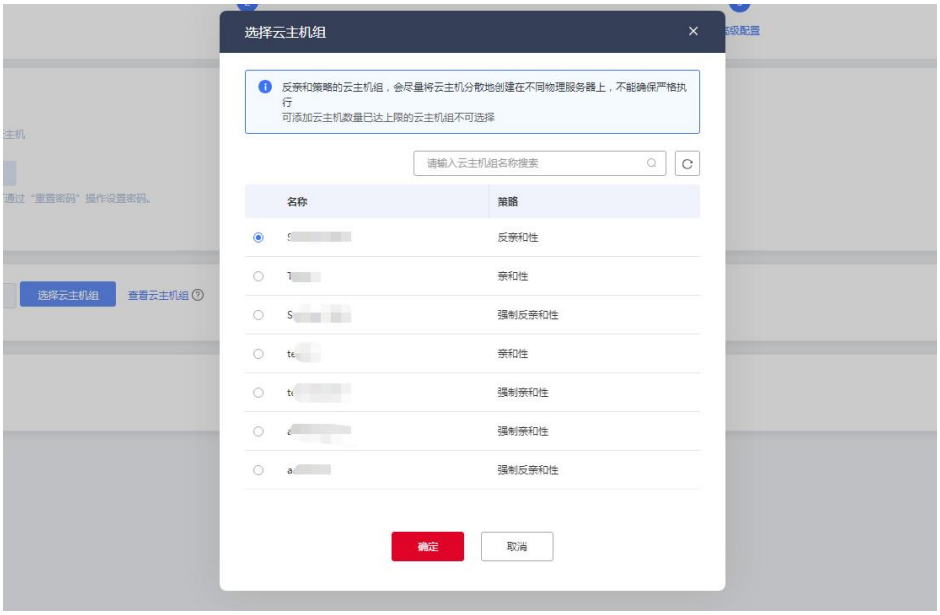
- 强制亲和策略的云主机组中的云主机，严格地执行将云主机创建在同一台宿主机上；
- 亲和策略则尽量将云主机创建在同一台宿主机上，不能确保严格执行；
- 强制反亲和策略的云主机组中的云主机，严格地执行将云主机分散地创建在不同的宿主机上；
- 反亲和策略则尽量将云主机分散地创建在不同宿主机上，不能确保严格执行。

8. 点击确定。

新建云主机添加至主机组

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 点击“创建云主机”。
5. 在“高级配置”页“云主机组”模块，点击“选择云主机组”。



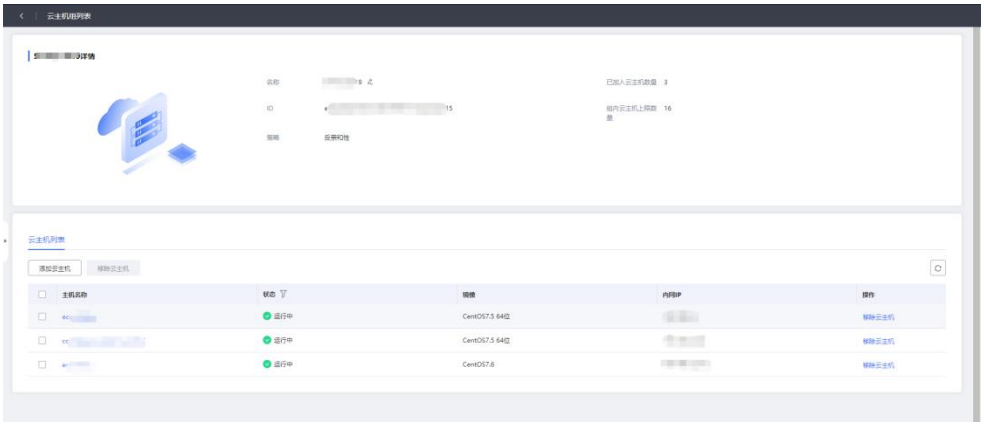
6. 选择云主机组（一个云主机组默认可添加 16 台云主机，可点击“查看云主机组”查看云主机组中的云主机数量）；
7. 点击确定。

已有云主机添加至主机组

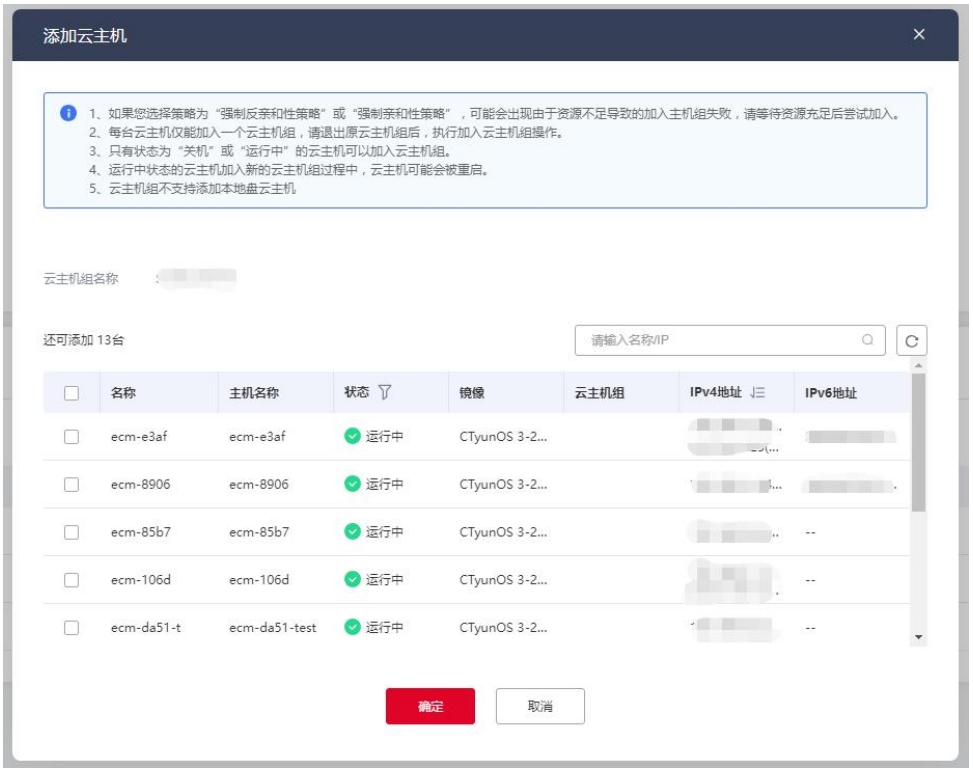
操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。

4. 选择左侧的“云主机组”。
5. 点击目标云主机组名称，进入云主机组的详情页。



6. 选择“添加云主机”，选择要加入该云主机组的云主机。



注意

运行中状态的云主机加入新的云主机组过程中，云主机可能会被重启。

7. 点击“确定”。

删除云主机组

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“云主机组”。
5. 找到想要删除的云主机组，点击右侧操作栏中的“删除”按钮。
6. 在删除确认弹窗中点击“确认”按钮。
7. 返回云主机组列表，刷新页面，云主机组删除操作完成。

从云主机组批量迁出云主机

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“云主机组”。
5. 点击目标云主机组名称，进入云主机组的详情页。
6. 在云主机列表中勾选需要迁移的云主机，或通过全选复选框勾选全部云主机，点击“移除云主机”按钮。
7. 在确认弹窗中确认云主机信息，点击“确定”，完成批量移除操作。

7.1.7 安装驱动和工具包（可选）

7.1.7.1NVIDIA 驱动安装指引

驱动类型选型概述

天翼云 GPU 云主机支持安装以下两种 NVIDIA 驱动：

- GPU 驱动：用于驱动物理 GPU，也称 Tesla 驱动。

- GRID 驱动：用于获得图形加速能力。

实例类型	场景	驱动类型	驱动安装方式
GPU 计算加速型	通用计算	Tesla 驱动	创建 GPU 实例时自动安装 GPU 驱动（仅支持 Linux 操作系统） NVIDIA 官网下载并安装 GPU 驱动
	图形渲染	GRID 驱动	在购买时选择已预装 GRID 驱动的计费镜像 向 NVIDIA 或其代理商购买对应的 License 并自行安装 GRID 驱动
GPU 图形加速基础型	通用计算	GRID 驱动	公共镜像中已预装 GRID 驱动，无需单独付费
	图形渲染	GRID 驱动	公共镜像中已预装 GRID 驱动，无需单独付费

注意

目前仅部分资源池的 GPU 云主机支持自动安装 GPU 驱动/提供预装 GRID 驱动的计费镜像，其他资源池请您手动安装驱动。如您需手动安装 GPU 驱动，请参见 安装 Tesla 驱动 和 安装 GRID 驱动。

各实例规格支持自动安装/预装的驱动版本

实例类型	支持驱动类型	自动安装支持的驱动版本	预装驱动的镜像	手动安装驱动版本支持
GPU 计算加速型 P8A/PI7/P2V/P2VS/PI2	Tesla 驱动	Driver 470.82.01CUDA 11.4.3 CUDNN 8.8.1.3	-	无特殊要求，NVIDIA 官方支持版本即可。
	GRID 驱动	-	Windows2019-DataCenter-GRID13.2 注：预装驱动版本 GRID13.2；部分资源池暂不支持。	-GRID14（- Driver 514.08）NVIDIA Virtual GPU (vGPU) Software Documentation
GPU 图形加速基础型 G5/G6	GRID 驱动	-	Windows Server 2016 Standard 64bit Windows Server 2012 Standard 64bit CentOS 7.5 64bit CentOS 7.6 64bit Ubuntu Server 16.04 64bit 注：预装驱动版本 GRID9.1	GRID9.0-9.4（Driver 430.30-432.44）
GPU 图形加速基础型 G	GRID 驱动	-	非多 AZ 资源池 Windows Server 2019 DataCenter 64bit Windows Server 2016 DataCenter 64bit	GRID13.0-13.8(Driver 470.63.01-474..44)

7/G5S			Windows Server 2012 DataCenter 64bit 多 AZ 资源池 Windows Server 2019 DataCenter 64bit Windows Server 2016 DataCenter 64bit Windows Server 2012 DataCenter 64bit CentOS 8.1 64bit CentOS 8.2 64bit Ubuntu Server 20.04 64bit 注：预装驱动版本 GRID13.2	

7.1.7.2 创建配备 GPU 驱动的 GPU 云主机（Linux）

准备工作

创建账号，以及完善账号信息。本教程创建的是按量付费实例。开通按量付费 GPU 云主机资源时，您的天翼云账户余额（即现金余额）不得小于 100.00 元人民币。充值方式请参见费用中心 >账户充值。

可选：在创建弹性云主机时，如果您的账号在本地域没有创建 VPC，天翼云会提供一个默认的 VPC，如果您不想使用默认 vpc，可以在本地域创建 vpc。具体操作，请参见虚拟私有云 >创建 VPC、子网搭建私有网络。

操作步骤

步骤 1：进入创建云主机页面

点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。

单击“服务列表>弹性云主机”。

单击“创建云主机”，系统进入创建页。

步骤 2：基础配置

选择“计费模式”。

包年包月：一种预付费模式，即先付费再使用。一般适用于固定业务应用，例如网站服务。需要先支付包年包月资源账单，才能开始使用包年包月资源。

按量付费：一种后付费模式，即先使用再付费。一般适用于有业务变化的应用，例如临时扩展、临时测试。可以先开通并使用按量付费资源，系统在每个结算周期生成账单并从账户中扣除相应费用。

选择“地域和可用区”。

设置“实例名称”，长度为 2~63 个字符。

设置“主机名称”，长度为 2～15 个字符，允许使用大小写字母、数字或连字符（-）。不能以连字符（-）开头或结尾，不能连续使用连字符（-），也不能仅使用数字。

选择分类为“GPU 加速/AI 加速型架构”。

根据自身需求选择规格族及具体规格。

预装驱动，如下两种方式可以预装驱动：

直接选择预装驱动的公共镜像，如 CTyunOS 2.0.1 64 位 预装 NVIDIA Tesla 550.90.07 驱动(40GB)、Ubuntu Server 22.04 64 位预装 NVIDIA Tesla 550.90.07 驱动(40GB)等名称中有“预装 NVIDIA Tesla 550.90.07 驱动”字样的镜像。

选择目标的 Linux 的公共镜像，勾选“安装 GPU 驱动”，选择对应的 CUDA、Driver、CUDNN 版本。

注意

目前仅部分资源池支持勾选自动安装 GPU 驱动，若目标资源池不支持该功能，且提供的预装驱动镜像无法满足您的要求，请您参见安装 Tesla 驱动，手动安装 GPU 驱动。

设置“存储”。 磁盘包括系统盘和数据盘。您可以为云主机添加多块数据盘，系统盘大小目前默认为 40GB。

步骤 3：网络配置

设置网络，包括“虚拟私有云”、“安全组”、“网卡”等信息。

参数	说明
虚拟私有云	云主机网络使用虚拟私有云（VPC）提供的网络，包括子网、安全组等。您可以选择使用已有的虚拟私有云网络，或者单击“前往控制台创建”来创建新的虚拟私有云。
安全组	安全组用来实现安全组内和安全组间云主机的访问控制，加强云主机的安全保护。用户可以在安全组中定义各种访问规则，当云主机加入该安全组后，即受到这些访问规则的保护。创建云主机时，可支持选择多个安全组。此时，云主机的访问规则遵循几个安全组规则的并集。 注意：如需通过远程桌面连接到 Windows 云主机，请在安全组中添加如下规则 方向：入方向 协议：TCP 端口范围：3389 如需通过 ssh 连接到 Linux 云主机，请在安全组中添加如下规则 方向：入方向 协议：TCP 端口范围：22 如需 Ping 云主机地址，请在安全组中添加如下规则 方向：入方向 协议：ICMP 类型：Any
网卡	添加一张主网卡，可使用已有内网 IP 地址，或自动分配。

参数	说明
弹性IP	将云主机与弹性 IP 绑定，使云主机通过固定的公网 IP 地址与互联网互通。 您可以根据实际情况选择以下三种方式： 不使用：云主机不能直接与互联网互通，仅可作为私有网络中部署业务或者集群所需云主机进行使用。 自动分配：自动为每台云主机分配独享带宽的弹性 IP，带宽值可以由您设定。 使用已有：为云主机分配已有弹性 IP。使用已有弹性 IP 时，不能批量创建云主机。

步骤 4：高级配置

1. 设置“登录方式”，并创建对应的密码或密钥对。

- 密钥对：指使用密钥对作为云主机的鉴权方式。您可以选择使用已有的密钥，或者单击“查看密钥对”创建新的密钥。注：如果选择使用已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录云主机。
- 密码：指使用设置初始密码方式作为云主机的鉴权方式。此时，您可以通过用户名密码方式登录云主机，Linux 操作系统时为 root 用户的初始密码，Windows 操作系统时为 Administrator 用户的初始密码。密码复杂度需满足：

参数	**规则**
密码	8～30 个字符，必须同时包含三项（大写字母、小写字母、数字、()`~!@#$%^&*_-+={}[]:;'<>,.?/中的特殊符号），且不能以斜线号（/）开头

2. 选择“云主机组”。

3. 选择“用户数据”配置方式（目前仅部分资源池支持）。
4. 单击“下一步，确认配置”。

步骤 5：确认配置

1. 在“确认配置”页面，查看云主机配置详情。
2. 企业项目：企业项目是对多个资源实例进行归类管理的单位，不同云服务区域的资源和项目可以归到一个企业项目中。企业可以根据不同的部门或项目组，将相关的资源放置在相同的企业项目内进行管理，支持资源在企业项目之间迁移。
3. 设置购买量。

- 购买时长：“包年/包月”方式需要设置购买时长，最短为 1 个月，最长为 3 年。
- 自动续订：“包年/包月”方式可选是否开启自动续订。按月购买的自动续订周期为 1 个月，按年购买的自动续订周期为 1 年。
- 购买数量：设置购买弹性云主机的数量。为了保证所有资源的合理分配，如果您需要的弹性云主机数量超过当前您可以购买的最大数值，您要提交工单申请扩大配额。申请通过后，您可以购买到满足您需要的弹性云主机数量。

4. 协议：阅读并勾选同意协议。
5. 如果您确认配置无误，单击“立即购买”。
6. 点击“立即支付”进行付款，付款成功即可创建弹性云主机。弹性云主机创建成功后，您可以在弹性云主机信息页面看到您新创建的弹性云主机。

7.2 云主机启动模版

7.2.1 创建云主机启动模版

云主机启动模板是天翼云提供的创建云主机的配置信息模板，包括镜像、规格、系统盘及数据盘类型和容量、网络配置、安全组、标签、登录信息等。当前已在部分资源池支持。

模板配置成功后，即可多次快速创建相同配置云主机，无需重新指定模板中已配置的参数，缩短部署时间。

若模版中配置不能满足您本次云主机创建时的需要，您可以在创建云主机环节在创建云主机页面修改相关参数。需注意的是，修改仅影响此次创建云主机操作，不影响模板内容。具体使用可查看使用云主机启动模版创建云主机。

前提条件及限制

不可跨地域使用云主机启动模板，例如不能使用华东 1 地域的模板创建华北 2 地域的云主机；

单地域最多可创建 20 个启动模板，如需调整请提交工单；

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 >弹性云主机 > 云主机启动模版”，进入云主机启动模版控制列表页。
4. 点击页面右上角“创建启动模版”。
5. 在启动模版中配置云主机基础配置、网络配置、高级配置信息，配置项填写可参考创建弹性云主机中步骤 1、步骤 2 及步骤 3。
6. 确认配置中，优先需要确认本次填写的所选配置，其次需要填写模版名称与模版描述。
7. 确认无误后，点击右下角“创建启动模版”。

7.2.2 查看云主机启动模版

操作场景

在您创建云主机启动模版后，通过控制台进入云主机启动模版列表页与详情页，可以查看云主机启动模版的列表与详细信息。

操作步骤

1. 登录控制中心。

2. 单击控制中心顶部的，选择“地域”。

3. 单击左侧导航栏“产品服务列表”，选择“计算 >弹性云主机 > 云主机启动模版”，进入云主机启动模版控制列表页。您可以查看启动模版名称、管理的镜像、规格等基本信息。

4. 在云主机启动模版列表页，点击目标模版名称，即可跳转至该启动模版的详情页，可以进一步模版中配置的等详细信息。

5. 部分情况下，实例模版中的配置可能被删除或已不可用，此时在实例模版详情页中将进行相关提醒。



提示语	出现的原因
资源已被删除	当前模版中选中的资源已被删除

提示语	出现的原因
资源不可用	当前资源已经停止服务
当前资源无权限访问	IAM 侧用户的权限被调整，当前用户无权限使用此资源

7.2.3 修改云主机启动模版

操作场景

您创建的云主机启动模版中部分资源已不可再使用，您可以修改云主机启动模版。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 >弹性云主机 > 云主机启动模版”，进入云主机启动模版控制列表页。

4. 找到需要修改的云主机启动模版，点击操作中“修改模版”



5. 在修改页面中更新相关配置信息，确认信息后，点击页面“保存本次修改”。

7.2.4 使用云主机启动模版创建云主机

操作场景

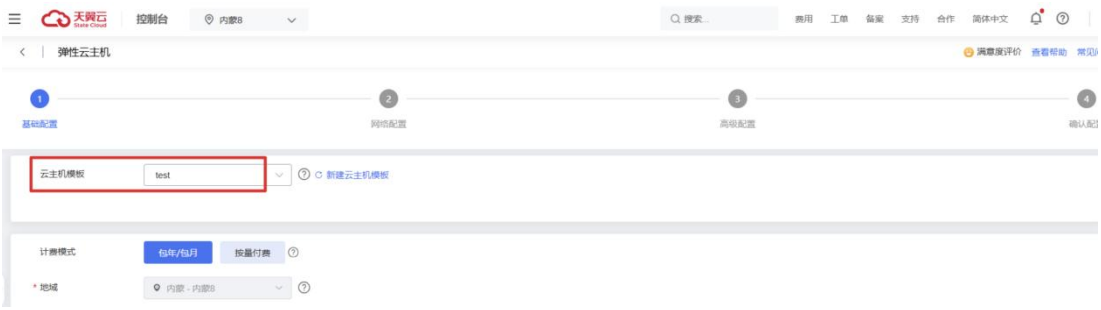
您创建的云主机启动模版可用来快速创建同样配置的云主机。使用云主机启动模板创建云主机时无需重新指定模板已包含的参数，缩短业务部署时间。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 >弹性云主机 > 云主机启动模版”，进入云主机启动模版控制列表页。
4. 找到想要使用的云主机启动模版，点击操作中“创建云主机”。



5. 在创建云主机页面将默认选中该条云主机启动模版。创建云主机需要的参数按照模版中内容已默认填写。



6.（可选）若启动模版中配置不能满足您此次云主机创建操作，您也可以进行修改。需注意的是，修改配置仅影响此次创建操作，不影响启动模板内容。

7.2.5 删除云主机启动模版

操作场景

当您创建的云主机启动模版已不再需要时，您可删除云主机模版。删除云主机模版后不会影响当前已创建的云主机实例或弹性伸缩。

操作步骤

1. 登录控制中心。

2. 单击控制中心顶部的 ，选择“地域”。

3. 单击左侧导航栏“产品服务列表”，选择“计算 >弹性云主机 > 云主机启动模版”，进入云主机启动模版控制列表页。

4. 找到想要使用的云主机启动模版，点击操作中“删除”。



7.3 云主机镜像

7.3.1 镜像概述

云主机镜像是预先配置好的操作系统和软件环境的模板，用于创建云主机实例。它包含操作系统、预装软件、系统配置和安全更新等信息，为用户提供快速部署和可靠的基础环境。通过使用云主机镜像，用户可以省去操作系统和软件的配置工作，快速启动适合自己需求的云主机实例。

本文关联产品为镜像服务，更多详细内容，请参考镜像服务产品介绍。相关约束限制，请参考镜像服务约束与限制。

公共镜像

公共镜像是云厂商向所有用户提供的镜像服务，涵盖了广泛的主流操作系统。这些镜像经过正版授权，您可以基于这些公共镜像进行个性化的配置和部署，以满足其特定的应用环境需求。

公共镜像的广泛适用性和高度稳定性使用户能够快速启动和部署云主机实例，并在其基础上构建自己的应用环境。您无需担心操作系统授权和稳定性问题，可以专注于应用程序的开发和部署工作。

公共镜像计费说明

部分公共镜像收费，例如 gpu 云主机的 Windows2019-DataCenter-GRID13.2 镜像，其余公共镜像免费。

天翼云提供的公共镜像规格

镜像规格请参考：镜像服务-产品规格

私有镜像

私有镜像是用户根据自身需求进行定制的镜像，仅对该用户可见。用户可以通过两种方式创建私有镜像：一种是基于现有的弹性云主机进行创建，另一种是将线下环境中的镜像导入到弹性云主机所在的地域。私有镜像具有高度的灵活性，可以满足用户个性化的需求。

私有镜像计费说明

- 通过云主机创建系统盘镜像、数据盘镜像：免费。
- 系统盘镜像、数据盘镜像存储在 ZOS 桶中，但并非用户的私有桶，对用户不可见，镜像的管理维护必须通过镜像服务来实现，目前这部分存储免费供用户使用。
- 通过镜像文件创建系统盘镜像、数据盘镜像或 ISO 镜像：收取镜像文件在 ZOS 桶的存储费用。
- 创建整机镜像：涉及的费用为备份存储库的费用。通过云主机创建整机镜像，需要使用备份存储库，用户需要对备份存储库付费。

共享镜像

共享镜像是用户之间可以共享和重复使用的预先配置的操作系统和软件环境模板，它避免了重复的操作和配置工作。当一个用户创建满足特定需求的镜像时，其他用户可以通过共享该镜像快速获取相同配置的环境，提高部署效率和保持一致性。通过共享镜像，用户能够更高效地利用云计算资源，加快应用程序的部署和开发速度。

共享镜像计费说明

来源于他人共享的系统盘镜像、数据盘镜像：免费。

安全产品镜像

安全产品镜像仅用于加载 CSSP、DAS、LAS、OSM 等部分安全产品服务。

安全产品镜像使用说明

- 来源于天翼云平台共享给客户。
- 仅用于加载部分安全产品服务

应用镜像

与基础的公共镜像相比，应用镜像预装了一些常见应用，可以实现快速部署特殊应用的目的。

7.3.2 通过镜像创建云主机

操作场景

您可以通过使用公共镜像或私有镜像来完成云主机创建，无论您选择使用哪种镜像类型，都能为您提供快速、可靠和灵活的云主机部署。

- 公共镜像为您提供了广泛的操作系统和预配置环境选项，使您能够迅速启动云主机并开始应用部署。
- 私有镜像则允许您根据自身需求定制和管理云主机的镜像，确保您的应用环境与业务需求完美契合。

使用须知

使用公共镜像和私有镜像创建云主机的区别是：

- 公共镜像：创建的云主机包含所需操作系统和预装的公共应用，需要您自行安装应用软件。
- 私有镜像：创建的云主机包含操作系统、预装的公共应用以及用户的私有应用。

操作步骤

1.登录控制中心。

2.单击“左侧导航栏>服务列表”，选择"计算>镜像服务"。

3.单击"公共镜像"或"私有镜像"，进入相应的镜像列表。

4.在公共镜像所在行右侧的"操作"列下，单击"申请云主机"，或在私有镜像操作下点击"申请云主机"。

5.填写云主机配置信息，包括云主机名称、网络、数量等。

6.单击"立即购买"。

7.单击"提交申请"。

8.返回云主机列表页，点击云主机列表右上角的刷新按钮，查看云主机的创建情况。

7.3.3 通过云主机创建 Windows 私有镜像

操作场景

本节将为您详细介绍如何通过云主机创建 Windows 私有镜像的步骤和方法。通过本节您将了解到如何在弹性云主机上创建您自己的 Windows 私有镜像，以满足您特定的应用需求和个性化配置。

前提条件

- 已有安装 Windows 操作系统的云主机。
- 确保创建私有镜像所使用的云主机网卡设置为 DHCP 的方式动态获取网络地址，系统默认满足；
- 当前“关机”状态的云主机才可以用来创建私有镜像，如果选择的云主机状态是“运行中”，请先关机。

操作步骤

1.登录控制中心，切换到相应节点；

2.单击“左侧导航栏>服务列表”，选择“计算>镜像服务”。

3.在“私有镜像”页面，单击“创建私有镜像”。

4.从弹出的云主机列表中选择相应的 Windows 云主机。

5.输入“镜像名称”，单击“下一步”按钮。

6.检查确认创建私有镜像参数，确认无误后勾选“我已阅读”，单击“确认下单”。

7.镜像创建时间与镜像文件本身大小有关可能稍慢，当镜像的状态为“正常”时，表示创建完成。

7.3.4 通过云主机创建 Linux 私有镜像

操作场景

您可以通过云主机创建自定义的 Linux 镜像，通过创建 Linux 私有镜像，您可以获得更高度定制化和灵活性，确保应用环境的一致性和可重复性。

本节为您介绍如何通过云主机创建 Linux 私有镜像。

前提条件

- 已有安装 Linux 操作系统的云主机。
- 需要确保创建私有镜像所使用的云主机网卡设置为 DHCP 的方式来动态获取网络地址，系统默认满足此配置。
- 清理 udev 配置规则。
- 配置私有镜像创建的 Linux 云主机密码初始化执行命令。
- 卸载创建私有镜像所使用的云主机中挂载的所有数据盘。
- 当前“关机”状态的云主机才可以用来创建私有镜像，如果选择的云主机状态是“运行中”，请先关机。

操作步骤

1.登录控制中心。

2.单击“左侧导航栏>服务列表”，选择“镜像服务”。

3.在“私有镜像”页面，单击“创建私有镜像”。

4.从弹出的云主机列表中选择相应的 Linux 云主机。

5.输入“镜像名称”，单击“下一步”按钮。

6.检查确认创建私有镜像参数，确认无误后勾选“我已阅读”，单击“确认下单”。

7.镜像创建时间与镜像文件本身大小有关可能稍慢，当镜像的状态为“正常”时，表示创建完成。

7.3.5 镜像更多操作

- 通过云主机创建系统盘镜像
- 通过镜像文件创建系统盘镜像
- 通过云主机创建数据盘镜像
- 通过镜像文件创建数据盘镜像
- 通过云主机快照创建系统盘镜像

7.4 云硬盘

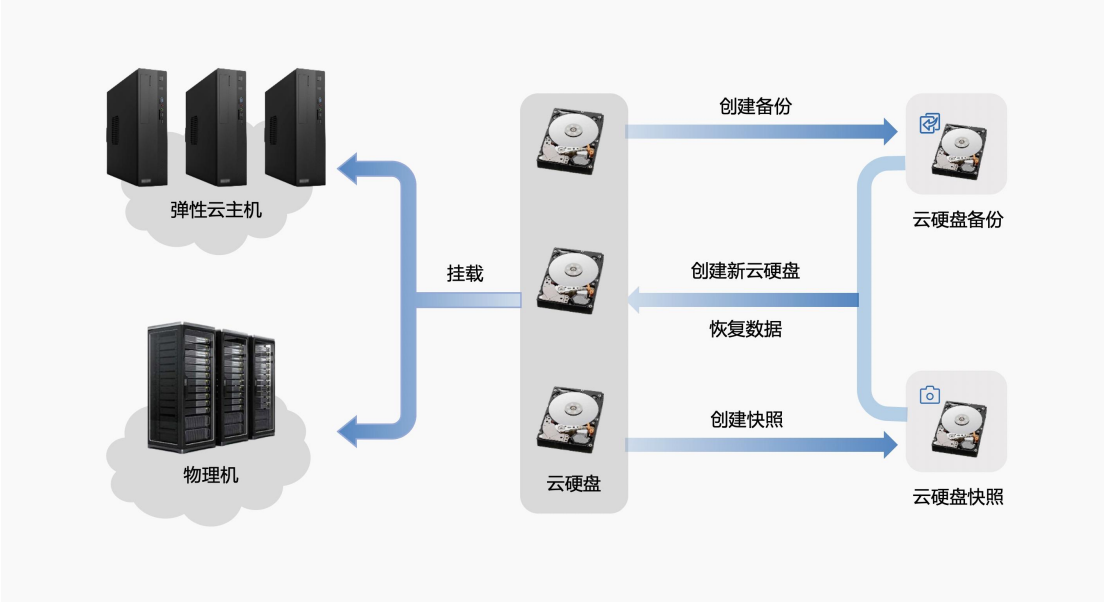
7.4.1 云硬盘概述

云硬盘（CT-EVS，Elastic Volume Service）是一种可弹性扩展的块存储设备，可以为弹性云主机和物理机提供高性能、高可靠的块存储服务。天翼云硬盘规格丰富，可满足不同场景的业务需求，适用于分布式文件系统、数据库、开发测试等场景。

用户可以在线操作及管理云硬盘，并可以像使用传统服务器硬盘一样，对挂载到云主机的云硬盘做格式化、创建文件系统等操作。

产品架构

云硬盘可以挂载至弹性云主机和物理机，云硬盘的备份与快照功能可帮助用户恢复数据，创建新的云硬盘，为用户提供了强大的数据保护，提高了数据的可靠性和灵活性。



弹性文件服务、对象存储、云硬盘的区别

天翼云为用户提供了三种数据存储服务，分别是弹性文件服务、对象存储、云硬盘，这三种服务的主要区别如下：

维度	弹性文件服务	对象存储	云硬盘
概念	弹性文件服务提供了一个高度可扩展的文件系统，可在云环境中共享文件数据。具有高可用性、持久性和可靠性。	对象存储具有高度的可扩展性和耐久性，可以存储任意类型的海量数据，并且能够自动处理数据冗余、故障恢复和数据分发。	云硬盘提供了高性能、低延迟、可扩展的块级存储。云硬盘可以被挂载到弹性云主机或物理机上，使其能够持久化地存储数据。
存储方	弹性文件服务采用文件存储方式。文件存储将数据组织为层次化的目录和文件	对象存储将数据存储为独立的对象。每个对象由数	云硬盘采用块存储方式。块存储将数据分为固定大小的块（通常为

维度	弹性文件服务	对象存储	云硬盘
式	结构，用户可以通过文件路径和名称来操作文件和目录。	据本身和与之相关的元数据（例如文件名、文件类型、大小等）组成。	几 KB 或几 MB），并通过唯一的块地址进行访问。
访问方式	弹性文件服务通过网络共享的方式进行访问。用户可以在需要的弹性云主机实例或容器实例上挂载文件系统，并通过标准的文件系统接口（如 NFS、SMB 等）访问共享的文件系统。	对象存储需要指定桶地址，通过 HTTP 或 HTTPS 等传输协议进行访问。	云硬盘类似于 PC 机的硬盘，无法单独使用，通常通过挂载（Mount）的方式来访问。它可以被挂载到弹性云主机或物理机上，使其在操作系统中可见。
适用场景	如应用程序的配置文件、日志文件等需要共享的文件数据以及在容器化应用中支持多个容器实例之间的数据共享和同步。	如大数据分析，数据湖，数据备份和归档等大规模数据存储和分析场景；静态网站托管解决方案存储。	如作为弹性云主机或物理机的数据存储介质进行数据存储和持久化；大规模数据处理与分布式计算等高性能计算场景。
容量	弹性文件服务可按需扩展，单文件系统容量默认最大为 32TB。如需更大容量的文件系统，可提工单申请。	对象存储服务没有容量限制，存储资源可无限扩展。	云硬盘支持按需扩容，最小扩容步长为 1 GB，单个云硬盘可由 10 GB 扩展至 32 TB。
是否	是	是	是

维度	弹性文件服务	对象存储	云硬盘
支持数据共享			
是否支持远程访问	是	是	否

访问方式

天翼云提供如下方式进行云硬盘的配置和管理：

- 控制台：Web 化的服务管理平台，即利用管理控制台来配置和管理云硬盘。
- API：天翼云也支持用户通过 API（Application Programming Interface）方式访问云硬盘，具体操作请参见云硬盘 API 参考。

7.4.2 创建云硬盘

操作场景

系统盘在创建云主机或物理机时自动添加，无需单独创建。数据盘可以在创建云主机或物理机时创建，由系统自动挂载给云主机或物理机，也可以在创建了云主机或物理机之后，单独创建云硬盘并挂载给云主机或物理机。

约束与限制

- 系统盘只能在购买云主机或物理机时自动创建并挂载，云硬盘创建页面只能创建数据盘。
- 随云主机或物理机创建的云硬盘，计费模式与云主机或物理机保持一致。
- 共享盘或磁盘模式为 SCSI 或磁盘类型为极速型 SSD 的云硬盘不支持加密模式。
- 随云主机或物理机一起创建的云硬盘会自动挂载至云主机或物理机上，不可卸载，无法再挂载至其他云主机或物理机中使用。
- 随云主机一起创建的云硬盘，磁盘类型可以在 VBD 或 SCSI 中进行选择，随着物理机一起创建的云硬盘磁盘类型默认为 VBD 类型。
- 云硬盘只能被挂载在同一个可用区的云主机或物理机上，可用区在创建完成后不支持修改。
- 创建云硬盘时若选择“立即挂载”，则无法批量创建，一次只能创建一个云硬盘。

操作步骤

1. 登录控制中心。
2. 单击控制中心左上角的  ，选择地域，此处我们选择华东 1。
3. 单击“存储>云硬盘”，进入云硬盘主页面。
4. 单击“创建云硬盘”，进入云硬盘创建页面。
5. 根据界面提示，配置云硬盘的基本信息。各配置项说明如下：

参数	是否必选	说明
地域	是	请选择云硬盘的地域，尽量选择距离您更近的区域以提高访问速度。
可用区	是	● 在拥有多个可用区的资源池中，选择其中一

		个可用区创建云硬盘。 ● 云硬盘只能挂载至同一个可用区的云主机上。云硬盘创建完成后不支持修改可用区。
付费方式	是	云硬盘支持的计费类型有包年/包月和按需计费两种。 ● 包年包月是一种先付费后使用的计费方式，按订单的购买周期结算。 ● 按需计费是一种先使用后付费的计费方式，按照云硬盘的实际使用时长计费，每小时结算一次。
是否挂载	是	选择是否在云硬盘创建完成后自动挂载到弹性云主机。 ● 暂不挂载（默认）：云硬盘创建完成后不自动挂载至云主机。 ● 立即挂载：云硬盘创建完成后自动挂载至用户选择的云主机。 注意 挂载到云主机的云硬盘，您还需要登录云主机进行数据盘初始化才可以正常使用。
数据源	否	云硬盘支持从快照、备份或镜像创建云硬盘。 从快照创建： ● 从快照创建云硬盘可选择目标地域目标可用区下的快照。 ● 从快照创建云硬盘所创建的云硬盘磁盘模式、磁盘类型、加密属性、所在地域可用区

		与源快照保持一致，不支持修改。 ● 从快照创建云硬盘创建磁盘容量不小于快照容量。 从备份创建： ● 从备份创建云硬盘可选择目标地域下的备份副本。 ● 从备份创建云硬盘所创建的云硬盘磁盘模式、加密属性和备份源云硬盘保持一致。 ● 从备份创建云硬盘所创建的云硬盘支持切换磁盘类型、支持跨可用区创建。 ● 从备份创建云硬盘创建磁盘容量不小于备份容量。 从镜像创建： ● 从镜像创建云硬盘可选择目标地域下的镜像。 ● 从镜像创建云硬盘仅支持 VBD 模式、且不支持磁盘加密。 ● 从镜像创建云硬盘所创建的云硬盘支持切换磁盘类型、支持跨可用区创建。 ● 从镜像创建云硬盘创建磁盘容量不小于镜像容量。 ● 仅支持“正常”状态的私有镜像或“已接受”状态的共享镜像。 注意 无论是从快照、备份或镜像创建，都不支持一次性创建多个云硬盘。
磁盘规格	是	● 云硬盘的类型：普通 IO，高 IO，通用型 SSD，

		超高 IO，极速型 SSD，XSSD-0，XSSD-1，XSSD-2。具体规格可参见产品规格。 ● 云硬盘的容量范围可参考云硬盘使用限制。 说明 ● 从备份创建云硬盘时，容量大小不小于备份源盘大小，默认容量为备份源盘大小。 ● 从快照创建云硬盘时，容量大小不小于快照源盘大小，默认容量为快照源盘大小。 ● 未选择数据源时，XSSD-2 默认容量大小为 512GB，其他类型云硬盘的默认容量大小为 40GB。
当前配置	-	用于显示当前所选磁盘类型和容量计算出来的基础 IOPS 上限及 IOPS 突发上限。具体计算方式可参见磁盘类型及性能介绍。
云硬盘备份	否	创建云硬盘时支持同时订购云硬盘备份服务。可以选择暂不配置、使用已有、现在购买。 ● 使用已有：若您想要配置备份且已有可用的存储库，可以选择“使用已有”。 ● 现在购买：若您想要配置备份但没有可用的存储库，可以选择“现在购买”。 配置存储库和备份策略并成功购买云硬盘后，系统会按此备份策略定期对云硬盘进行备份。云硬盘备份服务的计费说明请参考云硬盘备份服务计费说明。
共享盘	否	● 勾选“共享盘”，则创建的是共享云硬盘，共享云硬盘最多可同时挂载至 16 台云主机或物理机。

		● 不勾选“共享盘”，则默认为非共享云硬盘，只能挂载至 1 台云主机或物理机。
是否编辑标签	否	标签用于标识资源，可通过标签管理功能对云硬盘进行分类和筛选。 勾选后需要输入标签键和标签值，具体操作可参考标签功能概述。
磁盘模式	否	● VBD（默认）：VBD 类型的云硬盘仅支持简单的读写命令。 ● SCSI：支持 SCSI 指令透传，允许云主机操作系统直接访问存储介质。 ● FCSAN：FCSAN 云硬盘仅适用于物理机。
磁盘加密	否	● 支持创建加密云硬盘，磁盘加密能够最大限度的为您的数据提供安全防护。 ● 加密磁盘生成的快照/备份及通过这些快照/备份创建的磁盘将自动继承加密属性。
释放设置	否	支持设置释放策略。 可设置云硬盘释放时是否同步释放该盘的全部快照。
磁盘名称	是	自定义所创建的磁盘名称。 不能使用中文，且长度为 2-63 字符。
企业项目	是	支持为云硬盘选择企业项目。
数量	是	创建云硬盘的数量，默认为“1”，表示只创建一个云硬盘。

		注意 目前一次最多可批量创建 100 个云硬盘。
创建时长	是	如果计费类型选择“包年/包月”，则需要选择购买时长，可选取的时间范围为 1 个月到 5 年。
自动续订	否	启用自动续订，包年/包月订购的云硬盘到期后会自动续订。

6. 确定云硬盘的配置信息后，点击“下一步”。
7. 在“资源详情”页面，您可以再次核对云硬盘信息。确认无误后，阅读并勾选服务协议，单击“确认下单”，开始创建云硬盘。如果还需要修改，点击“上一页”，修改参数。
8. 在云硬盘主页面，查看云硬盘状态。待云硬盘状态变为“未挂载”时，表示创建成功。

7.4.3 挂载数据盘

操作场景

创建云硬盘后，需要将云硬盘挂载给云主机或物理机，供云主机或物理机作为数据盘使用。

挂载云硬盘通常分为两种，一种是挂载非共享云硬盘，另一种为挂载共享云硬盘。

挂载非共享云硬盘

约束与限制

- 非共享云硬盘只可以挂载至 1 台云主机或物理机。
- 挂载非共享云硬盘时，挂载的实例需要与此云硬盘位于同一区域下的同一可用区。
- 极速型 SSD 云硬盘仅支持挂载至 vCPU 数量至少为 16 且为 6 代以上的通用计算增强型和内存优化型云主机，并且一台云主机只允许挂载最多 3 块极速型 SSD 云硬盘。

操作步骤

用户可以从云主机与云硬盘两个页面入口进行挂载，具体操作步骤如下：

从云主机页面进行挂载

1. 登录控制中心。
2. 单击控制中心左上角的，选择地域，此处我们选择华东 1。
3. 单击选择“计算>弹性云主机”，进入云主机列表页面。
4. 单击待挂载云硬盘的弹性云主机名称，进入弹性云主机详情页面。
5. 单击“云硬盘”页签，在云硬盘页面中，点击“挂载磁盘”按键，弹出磁盘挂载窗口。



6. 在跳转窗口中选择要挂载的云硬盘，点击“确定”，即可为弹性云主机挂载云硬盘，且在云主机的云硬盘详情页面可看到挂载的云硬盘详情信息。

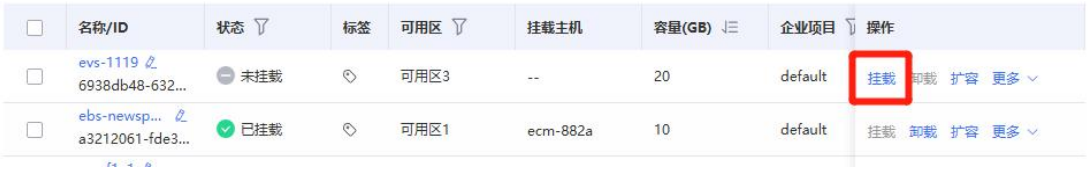


从云硬盘页面进行挂载

1. 登录控制中心。
2. 单击控制中心左上角的，选择地域，此处我们选择华东 1。

3. 单击“存储 > 云硬盘”，进入云硬盘主页面。

4. 在待挂载的云硬盘所在行，单击“挂载”。



5. 在弹出的“挂载磁盘”对话框中选择磁盘待挂载的云主机或物理机。



6. 单击“确定”。此时在弹出的窗口中，可点击“初始化数据盘”，也可继续点击“确定”。

7. 返回磁盘列表页面，当磁盘状态为“已挂载”时，表示挂载成功。

挂载共享云硬盘

约束与限制

- 共享云硬盘只能用作数据盘，不能用作系统盘。
- 共享盘最多可同时挂载至 16 台云主机或物理机。
- 当挂载共享云硬盘时，弹性云主机或物理机必须与共享云硬盘位于同一区域下的同一可用区。
- 磁盘模式为 SCSI 的共享云硬盘只能挂载至位于同一个强制反亲和性云主机组内的云主机，否则 SCSI 锁无法正常使用，可能会造成数据丢失风险。
- 共享云硬盘挂载的多台云主机只能为 Windows 或 Linux 操作系统中的一种。

操作步骤

用户可以从云硬盘页面入口对共享云硬盘进行挂载，具体操作步骤如下：

1. 登录控制中心。

2. 单击控制中心左上角的，选择地域，此处我们选择华东 1。

3. 单击“存储 > 云硬盘”，进入云硬盘列表页。

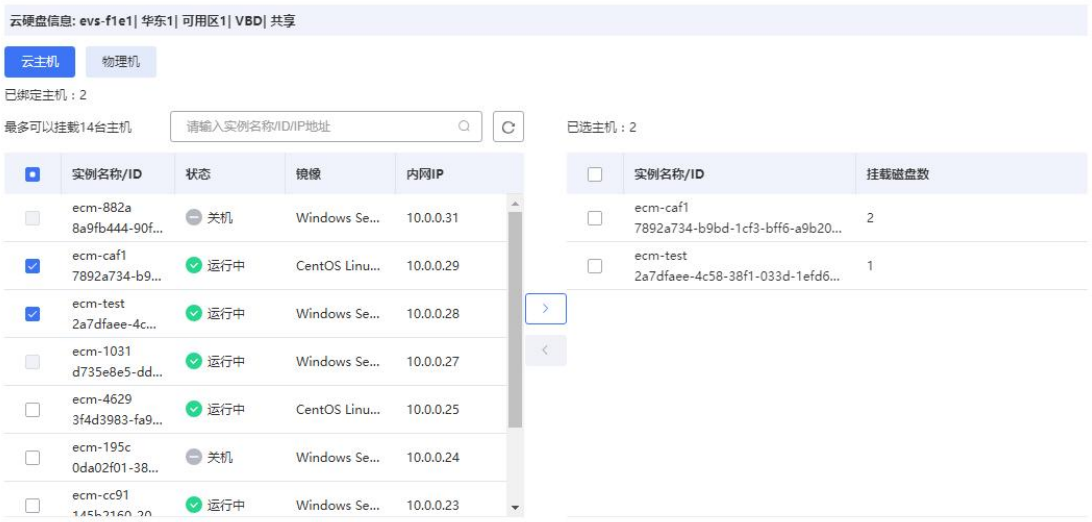
4. 在待挂载的共享云硬盘所在行，单击“挂载”。

注意

共享云硬盘与非共享云硬盘的区别为：当非共享云硬盘已经挂载至一台云主机后，其“挂载”操纵按钮是置灰状态，无法继续挂载。共享云硬盘则可以继续进行挂载。



5. 当单击“挂载”之后，会弹出“挂载磁盘”窗口，在此窗口中，会为您显示当前云硬盘已挂载主机台数和最多还可挂载的主机台数，用户可在此窗口中勾选需要挂载的云主机或物理机，并点击中间位置的按钮，最后点击“确定”按钮，即可将此云主机或物理机作为挂载点之一进行挂载。选中的云主机或物理机也可以点击按钮，从已选主机列表中移除。



6. 单击“确定”。此时在弹出的窗口中，可点击“初始化数据盘”，也可继续点击“确定”。

注意

已经初始化过或者有数据的数据盘不需要初始化。

至此，已将一块共享云硬盘挂载至一台弹性云主机或物理机上。

7.4.4 卸载数据盘

操作场景

只有数据盘支持卸载操作，系统盘不支持卸载。

当卸载数据盘时，支持离线卸载或在线卸载，即可在挂载该数据盘的云主机处于“运行中”或“关机”或“节省关机”状态时进行卸载。

当前有两种卸载云硬盘场景，一种为弹性云主机卸载，一种为物理机卸载。

- 弹性云主机：用户可为弹性云主机卸载磁盘，本文将详细介绍弹性云主机卸载磁盘的操作步骤。
- 物理机：支持将云硬盘挂载至物理机用作数据盘，数据盘可在物理机处于“关机”或“运行中”状态时进行卸载。具体操作步骤请参见物理机卸载数据盘。

注意

随云主机一起订购的数据盘不支持在云硬盘列表中卸载，需要前往云主机详情页面进行卸载。

前提条件

- 卸载已挂载到 Windows 云主机的云硬盘前，请确保没有程序正在对该云硬盘进行读写操作。否则，有可能造成数据丢失或卸载失败。
- 卸载已挂载到 Linux 云主机的云硬盘前，请先登录云主机，执行 umount 命令，并确保没有程序正在对该云硬盘进行读写操作。否则，有可能造成数据丢失或卸载失败。

操作步骤

用户可通过两种方式卸载云硬盘，分别为：

- 通过云硬盘管理控制台卸载
- 通过弹性云主机管理控制台卸载

通过云硬盘管理控制台卸载

1. 登录控制中心。
2. 单击控制中心左上角的，选择地域，此处我们选择华东 1。
3. 单击“存储>云硬盘”，进入云硬盘主页面。
4. 在云硬盘主页面，单击待卸载云硬盘所在行“操作 > 卸载”，弹出“卸载云硬盘”对话框。



5. 在“卸载云硬盘”对话框中，确认云硬盘需要卸载的云主机并勾选，然后单击“确定”。
6. 回到云硬盘列表，此时云硬盘状态为“卸载中”，表示云硬盘处于正在从云主机卸载的过程中。当云硬盘状态为“未挂载”时，表示卸载成功。

通过弹性云主机管理控制台卸载

1. 登录控制中心。
2. 单击控制中心左上角的，选择地域，此处我们选择华东 1。
3. 单击选择“计算>弹性云主机”，进入云主机列表页面。
4. 点击待卸载云硬盘的弹性云主机名称，进入弹性云主机详情页面。
5. 点击进入详情页面下方的“云硬盘”页签，可查看当前云主机下所有云硬盘的具体信息。单击待卸载云硬盘所在行“操作 > 卸载”，弹出“卸载云硬盘”对话框，请阅读完注意事项后单击“确定”。



6. 若卸载的云硬盘是“随云主机一起订购的数据盘”，则点击“确定”后，会先进行订单拆分。待订单拆分完毕后，再次点击“操作 > 卸载”，才能成功完成卸载云硬盘操作。

7. 若卸载的云硬盘是“单独订购后挂载至该云主机的数据盘”，则点击“确定”后即可完成卸载云硬盘操作。

说明

如果共享云硬盘同时挂载至多个云主机，只从其中的一个云主机卸载，卸载成功后，云硬盘状态依然为“已挂载”。只有当共享云硬盘已经从所有的云主机上卸载成功时，状态会变为“未挂载”。

7.4.5 扩容云硬盘

操作场景

当云硬盘空间不足时，用户可以扩大云硬盘的容量，也就是云硬盘扩容。云硬盘扩容可以有如下两种处理方式：

- 申请一块新的云硬盘，并挂载给云主机。
- 扩容原有云硬盘空间。系统盘和数据盘均支持扩容。

本文主要对方式二进行介绍。在该方式中，您可以对状态为“已挂载”或者“未挂载”的云硬盘进行扩容。状态为“已挂载”的云硬盘，即当前需要扩容的云硬盘已经挂载给云主机。状态为“未挂载”的云硬盘，即当前需要扩容的云硬盘未挂载至任何云主机。系统盘支持的最大容量为 2TB，数据盘支持的最大容量为 32TB（X 系列数据盘的最大容量为 64TB），最小扩容容量为 1GB，扩容步长为 1GB。

约束与限制

- 若扩容的云硬盘处于“已挂载”状态：
  - ✧ 云硬盘为非共享盘，则其挂载的云主机状态必须为“运行中”或“关机”才支持扩容。
  - ✧ 云硬盘为共享盘，则其挂载的所有云主机状态必须都为“运行中”或“关机”才支持扩容。

注意

按需云主机支持“节省关机模式”，此模式下关闭云主机，云主机也处于“关机”状态，同样支持扩容云硬盘。

- 使用公共镜像创建的系统盘及使用公共镜像制作的私有镜像创建的系统盘支持在“已挂载”状态下扩容。
- 云硬盘扩容功能支持扩大云硬盘容量，不支持缩小云硬盘容量。
- 当磁盘已经投入使用后，请在扩容前务必检查磁盘的分区形式，具体说明如下：

分区形式	注意事项
MBR 分区	容量最大支持 2TB（2048GB），超过 2TB 的部分无法使用。
GPT 分区	容量最大支持 18EB（19327352832 GB）。云硬盘服务支持的最大数据盘容量为 32TB（32768 GB），即您最大可将数据盘扩容至 32TB。
MBR 分区需扩容至 2TB 以上	必须将磁盘分区形式由 MBR 切换成 GPT，期间会中断业务，并且更换磁盘分区形式时会清除磁盘的原有数据，请在扩容前先对数据进行备份。

操作步骤

一、在控制中心扩容云硬盘容量

1. 登录控制中心。
2. 单击控制中心左上角的，选择地域，此处我们选择华东 1。
3. 单击“存储>云硬盘”，进入云硬盘主页面。
4. 在云硬盘列表中，选择指定的云硬盘所在行“操作”列下的“扩容”，进入扩容磁盘界面。



5. 根据界面提示，设置“新增容量”参数，设置完成后，单击“确定”。



6. 支付成功后，在云硬盘主页面，查看云硬盘扩容结果。当云硬盘状态由“扩容中”变为“未挂载”或“已挂载”时，此时可以看到容量已增加，表示扩容成功。



扩容成功后，需要对扩容部分的云硬盘进行后续处理。

二、登陆云主机扩展分区和文件系统

不同操作系统的云主机处理方式不同，具体请参见后续章节：

- 扩展磁盘分区和文件系统（Windows 2008）
- 扩展磁盘分区和文件系统（Windows 2016）
- Linux 扩展磁盘分区和文件系统

7.4.6 加密云硬盘

7.4.6.1 支持云硬盘加密功能

什么是云硬盘加密？

当您的业务因为等保合规或安全要求等原因，需要对存储在云硬盘上的数据进行加密保护时，您可以在创建云硬盘时勾选加密选项，即可对新创建的云硬盘进行加密。

密钥管理

天翼云使用行业标准的 AES-256 算法，利用数据密钥加密您的云硬盘数据，加密云硬盘使用的密钥由天翼云自研密钥管理（KMS，Key Management Service）功能提供，用户可轻松创建并管理密钥，满足数据加解密及数字签名验签等需求，安全便捷。

KMS 通过使用硬件安全模块 HSM（Hardware Security Module）保护密钥的安全，所有的用户密钥都由 HSM 中的根密钥保护，避免密钥泄露。密钥管理可以轻松满足对小数据和大量数据的加解密。

工作原理

服务端加密支持选择默认密钥及用户自行创建的用户主密钥，具体可选择的密钥类型如下。

密钥创建者	密钥类型	密钥算法	服务版本
云产品	默认密钥	AES_256（默认）	按需版&包周期版
用户自行创建	用户主密钥-软件	AES_256	按需版
	用户主密钥-硬件	AES_256 SM4	按需版

在了解云硬盘加密工作原理之前，首先需要了解两个概念：

- **默认密钥**
  - 系统为云产品自动创建的用于服务端加密的默认密钥，默认密钥与云产品对应，每个天翼云账号下的每个云产品，在每个资源支持创建 1 个默认密钥。
  - 默认密钥的别名定义为 alias_<云产品代码>，例如 alias_ecs。
  - 默认密钥的密钥材料由 KMS 生成，不支持导入外部密钥材料，同时不支持自动轮转、启用/禁用、删除等操作。
- **用户主密钥**

- 用户主密钥云产品加密时，可选用户在 KMS 服务中自建的用户主密钥，密钥类型为对称密钥，算法支持 AES_256、SM4，保护级别可选软件保护、硬件保护。
- 用户主密钥按照 KMS 按需及包周期版的服务标准资费进行计费，请您确保账户余额充足、到期前及时续费，避免 KMS 服务冻结，冻结后云产品无法进行正常的加解密操作，云产品可能会出现异常。
- 用户主密钥支持计划删除，操作计划删除前请确保该密钥非云产品加密使用的密钥，避免误删除导致云产品无法正常加解密而出现异常。为避免误删，您可以为密钥开启删除保护功能。

注意

当前云硬盘加密仅支持选择按需版本中的自建用户主密钥，当前包周期版本中的用户主密钥暂不支持做云硬盘加密使用。

若您为 2024 年 9 月 10 日之后购买了 KMS 包周期服务，您可选择使用默认密钥进行云产品加密。

第一次使用加密云硬盘时，系统会自动创建一个用户主密钥（CMK），该密钥有且仅有一个，且是在 KMS 中的相应地域所创建，并将其存储在受严格的物理和逻辑安全控制保护的密钥管理服务上。查看如何通过 KMS 实现服务端加密。

每个地域的加密云硬盘，都需要通过 256 位数据密钥（DEK）进行加密，此数据密钥（DEK）具备地域唯一性，即每个地域都有且仅有一个。该密钥受 KMS 提供的密钥管理基础设施的保护，能有效防止未经授权的访问。云硬盘的数据密钥（DEK）仅在实例所在的宿主机的内存中使用，不会以明文形式存储在任何持久化介质（即使是云硬盘本身）上。

在创建加密云硬盘并将其挂载到实例后，以下数据都将关联此密钥并进行加密：

- 云硬盘中的静态数据
- 云硬盘和实例间传输的数据（实例操作系统内的数据不加密）
- 通过加密云硬盘创建的快照

云硬盘加密功能与快照、备份、镜像之间的关系：

云硬盘加密功能与云主机系统盘、云主机数据盘、快照、备份、镜像的关系介绍如下：

云硬盘加密功能

- 创建云硬盘时，用户可以选择是否加密此云硬盘，云硬盘创建完成后加密属性无法更改。

云主机系统盘加密

- 创建云主机时，支持在创建时直接设置系统盘加密。

云主机数据盘加密

- 创建云主机时，支持在创建时直接设置数据盘加密。

云硬盘加密与快照

- 加密云硬盘生成的快照及通过这些快照创建的云硬盘将自动继承加密功能属性。

云硬盘加密与备份

- 加密云硬盘生成的备份及通过这些备份创建的云硬盘将自动继承加密功能属性。

7.4.6.2 管理加密云硬盘

工作原理

服务端加密支持选择默认密钥及用户自行创建的用户主密钥，具体可选择的密钥类型如下。

密钥创建者	密钥类型	密钥算法	服务版本
云产品	默认密钥	AES_256（默认）	按需版&包周期版
用户自行创建	用户主密钥-软件	AES_256	按需版

密钥创建者	密钥类型	密钥算法	服务版本
	用户主密钥-硬件	AES_256 SM4	按需版

在了解云硬盘加密工作原理之前，首先需要了解两个概念：

- **默认密钥**
  - 系统为云产品自动创建的用于服务端加密的默认密钥，默认密钥与云产品对应，每个天翼云账号下的每个云产品，在每个资源池支持创建 1 个默认密钥。
  - 默认密钥的别名定义为 alias_<云产品代码>，例如 alias_ecs。
  - 默认密钥的密钥材料由 KMS 生成，不支持导入外部密钥材料，同时不支持自动轮转、启用/禁用、删除等操作。
- **用户主密钥**
  - 云产品加密时，可选用户在 KMS 服务中自建的用户主密钥，密钥类型为对称密钥，算法支持 AES_256、SM4，保护级别可选软件保护、硬件保护。
  - 用户主密钥按照 KMS 按需及包周期版的服务标准资费进行计费，请您确保账户余额充足、到期前及时续费，避免 KMS 服务冻结，冻结后云产品无法进行正常的加解密操作，云产品可能会出现异常。
  - 用户主密钥支持计划删除，操作计划删除前请确保该密钥非云产品加密使用的密钥，避免误删除导致云产品无法正常加解密而出现异常。为避免误删，您可以为密钥开启删除保护功能。

注意

当前云硬盘加密仅支持选择按需版本中的自建用户主密钥，当前包周期版本中的用户主密钥暂不支持做云硬盘加密使用。

若您为 2024 年 9 月 10 日之后购买了 KMS 包周期服务，您可选择使用默认密钥进行云产品加密。

第一次使用加密云硬盘时，系统会自动创建一个用户主密钥（CMK），该密钥有且仅有一个，且是在 KMS 中的相应地域所创建，并将其存储在受严格的物理和逻辑安全控制保护的密钥管理服务上。查看如何通过 KMS 实现服务端加密。

每个地域的加密云硬盘，都需要通过 256 位数据密钥（DEK）进行加密，此数据密钥（DEK）具备地域唯一性，即每个地域都有且仅有一个。该密钥受 KMS 提供的密钥管理基础设施的保护，能有效防止未经授权的访问。云硬盘的数据密钥（DEK）仅在实例所在的宿主机的内存中使用，不会以明文形式存储在任何持久化介质（即使是云硬盘本身）上。

在创建加密云硬盘并将其挂载到实例后，以下数据都将关联此密钥并进行加密：

- 云硬盘中的静态数据
- 云硬盘和实例间传输的数据（实例操作系统内的数据不加密）
- 通过加密云硬盘创建的快照

创建加密云硬盘

加密数据盘的灵活度较高，用户可以选择跟随弹性云主机一起购买，也可以在云硬盘管理控制台上单独购买，为其加密可以通过以下方法：

- 跟随弹性云主机购买云硬盘时，用户可在云硬盘高级属性中选择是否为其加密，当勾选加密属性后，此云硬盘即可成功加密。系统盘只能跟随弹性云主机一起订购，因此只能在订购云主机时选择加密。
- 在云硬盘管理控制台单独购买一个空的数据盘，且不选择其数据来源时，也可以在高级属性设置中选择其为加密数据盘，且购买成功后，用户无法更改其加密属性。
- 在云硬盘管理控制台勾选数据来源购买一个数据盘，用户可选择的数据来源有备份与快照，此云硬盘的加密属性和备份、快照的源云硬盘加密属性保持一致。即快照与备份的源云硬盘是加密云硬盘，则此云硬盘也具有加密属性。

创建云硬盘的具体操作步骤可参见创建云硬盘。

卸载加密云硬盘

若加密云硬盘使用的是用户主密钥，在卸载之前请确认云硬盘的用户主密钥是否可用。

- 若此加密云硬盘的用户主密钥是可用的，卸载云硬盘时，数据不会丢失，也可以正常重新挂载。
- 若此加密云硬盘的用户主密钥不可用，即使当前该云硬盘还可以正常读写，但是不能保证此云硬盘一直可以正常使用，且有可能造成重新挂载的失败，因此用户需要随时确保用户主密钥的状态，再进行卸载。

卸载加密云硬盘的具体操作请参见卸载云硬盘。

数据盘加密场景

数据盘可以跟随弹性云主机一起购买，也可以单独购买。数据盘是否加密主要涉及如下场景：

购买方式	数据源	说明
随弹性云主机一起购买数据盘	不选择数据源	随弹性云主机一起购买的空白数据盘，可以选择加密或不加密。 创建完成后不可更改加密属性。
单独购买数据盘	不选择数据源	创建的空白数据盘，可以选择加密或不加密。 创建完成后不可更改加密属性。
	从备份创建（备份源云硬盘加密）	通过加密云硬盘创建的备份属性为加密。 使用加密备份作为数据源创建的云硬盘继承备份的加密属性和加密密钥。

	从备份创建 （备份源云硬盘未加密）	通过未加密云硬盘创建的备份为未加密备份。 使用未加密备份作为数据源创建的云硬盘未加密。
	从快照创建（快照源云硬盘加密）	通过加密云硬盘创建的快照为加密快照。 使用加密快照作为数据源创建的云硬盘继承快照的加密属性和加密密钥。
	从快照创建 （快照源云硬盘未加密）	通过未加密云硬盘创建的快照为未加密快照。 使用未加密快照作为数据源创建的云硬盘未加密。
	从镜像创建 （数据盘镜像不支持加密）	仅未加密云硬盘可以创建数据盘镜像。 使用未加密镜像作为数据源创建的云硬盘未加密。

加密云硬盘相关限制

使用加密云硬盘时还需要注意以下限制：

限制项	**限制说明**
支持资源池	南宁 23/华东 1/南昌 5/华南 2/西安 7/太原 4/华北 2/郑州 5/西南 2-贵州/杭州 7/庆阳 2/呼和浩特 3/长沙 42/芜湖 4。

支持加密的云硬盘类型	普通 IO、高 IO、通用型 SSD、超高 IO。
其他限制项	云硬盘的加密属性在云硬盘创建完成后不支持修改。
	磁盘模式为 SCSI 或 FCSAN 的云硬盘不支持加密。
	共享盘不支持加密。

7.4.7 设置云硬盘释放策略

随实例释放按需计费云硬盘

针对按需计费云硬盘，您可以手动释放云硬盘，也可以设置随实例释放云硬盘（如果云硬盘开启了随实例释放，则在释放云主机实例时云硬盘会自动一起释放）。

应用场景

以下几种按需计费云硬盘，可以对其设置“随实例释放云硬盘”，即可实现在释放云主机时同时释放该云硬盘。您也可以根据业务需要随时关闭该选项。

- 单独创建的云硬盘。
- 随云主机一起订购，但已进行产品拆分的数据盘。

约束与限制

- 仅按需计费云硬盘支持该功能。
- 仅非共享盘支持该功能。
- 按需计费云硬盘，删除后放入回收站，保留此设置信息；从回收站恢复云硬盘，该设置仍然生效。
- 云硬盘状态为“未挂载”或“已挂载”状态下，支持设置“云硬盘随实例释放”。

操作步骤

1. 登录控制中心。
2. 单击控制中心左上角的  ，选择地域，此处我们选择华东 1。
3. 单击“存储>云硬盘”，进入云硬盘列表页。

4. 在待设置释放策略的云硬盘所在行，单击“操作 >更多> 随实例释放设置”，进入“随实例释放设置”窗口。
5. 在“随实例释放设置”窗口，您可以为该盘设置或修改释放策略。
   - 勾选“云硬盘随实例释放”，云硬盘和实例将保持一致的删除动作；
   - 不勾选“云硬盘随实例释放”，云硬盘不随实例删除动作而删除。

   对于已设置过释放策略的云硬盘，可通过该操作修改设置。
6. 确定云硬盘释放设置后，您可以点击“确定”，设置释放策略的命令将被下发。若您不需设置释放策略，可点击“取消”退出当前操作。
7. 释放策略设置成功后，您可以在云硬盘详情页面查看云硬盘随实例释放的配置详情是否和预期一致。

随云硬盘释放快照

云硬盘的快照默认不随云硬盘自动释放，如果您不再需要某些云硬盘快照，可以将其手动释放。您还可以通过设置云硬盘释放策略，选择在释放云硬盘的同时释放该盘的全部快照。

应用场景

释放方式

不同订购方式的云硬盘的快照支持的释放方式如下：

资源类型	**释放方式**
随弹性云主机一起购买数据盘的全部快照	只支持手动释放。
单独购买数据盘的全部快照	支持手动释放。 支持自动释放，默认关闭全部快照随盘释放。

释放影响

对于单独购买的数据盘的全部快照，开启与关闭全部快照随云硬盘释放有以下影响：

- 开启全部快照随云硬盘释放后，包周期或按需云硬盘被释放时，该云硬盘的所有快照（包括自动快照、手动快照）将会同步自动释放。
- 关闭全部快照随云硬盘释放后，包周期或按需云硬盘被释放时，该云硬盘的所有快照（包括自动快照、手动快照）将不会同步自动释放。

注意

开启全部快照随云硬盘释放后，云硬盘的全部快照将随云硬盘删除/退订动作直接被删除，不会进入云硬盘回收站保留。

约束与限制

只有单独购买的数据盘支持设置快照释放策略。

操作步骤

1. 登录控制中心。
2. 单击控制中心左上角的  ，选择地域，此处我们选择华东 1。
3. 单击“存储>云硬盘”，进入云硬盘列表页。
4. 在待设置释放策略的云硬盘所在行，单击“操作 >更多> 云硬盘释放设置”，进入“释放设置”页面。
5. 在“释放设置”页面，您可以为该盘设置或修改释放策略。勾选“全部快照随云硬盘释放”，快照和云硬盘将保持一致的退订/删除动作；不勾选“全部快照随云硬盘释放”，快照不随云硬盘退订/删除动作而删除。对于已设置过释放策略的云硬盘，可通过该操作修改设置。
6. 确定云硬盘释放设置后，您可以点击“确定”，设置释放策略的命令将被下发。若您不需设置释放策略，可点击“取消”退出当前操作。
7. 释放策略设置成功后，您可以在云硬盘详情页面查看全部快照是否随云硬盘释放的配置详情是否和预期一致。

7.5 网卡

7.5.1 弹性网卡概述

弹性网卡是虚拟私有网络 VPC 中的虚拟网络接口,用于连接 ECS 与私有网络。

每个弹性网卡可以包含以下主要属性：

- 专有网络 VPC、子网信息
- VPC 子网的 IPv4 地址范围内的一个主要私有 IPv4 地址
- VPC 子网的 IPv4 地址范围内的一个或者多个辅助私有 IPv4 地址
- 和一个公网 IPv4 地址绑定，将网卡的私网 IP NAT 成为公网 IP
- 一个 IPv6 地址
- 一个 MAC 地址
- 一个或多个安全组

网卡类型

网卡类型包括主网卡和辅助网卡。随着云主机、物理机服务器等实例一起创建的网卡称为主网卡。VPC 网络中的每个实例都有默认的主网卡，您无法独立创建和从实例卸载主网卡。您可以创建并额外绑定到实例上的网卡，称为辅助网卡，辅助网卡可以灵活的和实例解绑、绑定。

私有 IPv4 地址

每个弹性网卡都有一个子网地址范围内的私有 IPv4 地址，称为主私网地址。您可以在创建辅助弹性网卡时指定主私网地址，如果不指定，我们将为您随机分配。此外，您还可以为弹性网卡分配一个或者多个辅助私有 IPv4 地址，辅助私有 IP 地址在取消分配后可以回收，然后再分配给其他弹性网卡。

私有 IPv6 地址

如果虚拟私有网络 VPC 和子网开通了 IPv6 网段，您可以为弹性网卡分配一个主私网 IPv6 地址。主私网 IPv6 地址在取消分配后可以回收。VPC 内的 IPv6 地址支持 VPC 内网通信和公网通信，网卡绑定 IPv6 带宽或者加入共享带宽后，可以使用 IPv6 来访问公网或被公网访问。

弹性公网 IP 地址

弹性网卡绑定弹性公网 IP，用于公网通信。使用弹性 IP 时会将网卡的私网 IP NAT 成为公网 IP，以此来访问公网或者被公网访问。

多队列

网卡多队列是指实例规格支持的最大网卡队列数。单台 ECS 实例 vCPU 处理网络中断存在性能瓶颈时，您可以将实例中的网络中断分散给不同的 vCPU 处理，从而提升性能。

修改内网 IP/切换 VPC

云主机、物理机的主网卡提供以下能力：修改为同子网下其他 IP、切换同 VPC 下的其他子网、切换至其他 VPC。

您可以通过修改云主机、物理机服务器的主网卡 IP 或者切换 VPC、子网来实现迁移功能，在网络规划存在冲突或者需要隔离某个服务器时，可以使用此功能。

注意
弹性网卡仅适用于可用区资源池，不同资源池列表见产品简介-资源池区别页面，实际情况以控制台展现为准。

7.5.2 创建弹性网卡

操作场景

创建弹性网卡可用于实现灵活、高可用的网络方案配置。

前提条件

- 注册天翼云账号，并完成实名认证。具体操作，请参见天翼云账号注册流程。
- 您已经创建虚拟私有云 VPC、子网、弹性云主机。

操作步骤

1. 登录控制中心。

2. 在控制中心页面左上角点击  ，选择区域，本文我们选择华东-华东 1。

3. 依次选择“网络”，单击“虚拟私有云”，进入网络控制台页面。

4. 单击“弹性网卡”，进入弹性网卡列表页面。单击“创建弹性网卡”，进入创建弹性网卡界面。

5. 在创建弹性网卡详情页填写基本信息。弹性网卡具有子网属性，创建后只能挂载到对应子网的云主机、物理机实例上。可以通过分配辅助私网 IP 来实现单网卡多 IP。

参数	描述
名称	弹性网卡名称，由数字、字母、中文、-、_组成，不能以数字、_和-开头。
VPC	选择弹性网卡归属的 VPC，必填项。
子网	选择弹性网卡归属的子网，必填项。
主私网 IPv4 地址	自动分配
辅助私网 IP	可以选择不设置、自动分配，或者指定地址。
安全组	选择弹性网卡所属安全组。

6.填写基本信息后，点击“确定”按钮，弹性网卡创建成功。

7.5.3 管理弹性网卡

添加网卡

当您的弹性云主机需要多个网卡时，可为弹性云主机添加网卡。

前提条件

- 用户在开关机状态下均可添加/删除网卡。
- 主网卡不支持删除，扩展网卡支持删除。

操作步骤

1. 进入弹性云主机详情页面，点击“添加网卡”。
2. 在弹窗中配置安全组、扩展网卡信息。
3. 点击确定，完成添加网卡。

删除网卡

弹性云主机可以绑定多张弹性网卡，其中包括一个主网卡，且主网卡不可删除。当您的弹性云主机需要删除主网卡以外的其他网卡时，可以参考下面步骤。

1.进入弹性云主机详情页面，选择要删除的网卡，点击“删除网卡”。

说明

对于该弹性云主机的主网卡（默认为网卡列表中显示的第一个网卡），用户不能执行删除操作。

2.在弹出的对话框中单击“确定”，删除网卡；

修改内网 IP 地址

前提条件

- 云主机处于关机状态；
- 只有主网卡支持修改内网 IP，必须先删除辅助网卡；
- 如果网卡绑定了虚拟 IP 或者 DNAT 规则，需要先解绑；
- 如果网卡上有 IPv6 地址，无法修改（包括 IPv4 和 IPv6 的）内网 IP 地址；
- 如需修改弹性负载均衡后端服务器的内网 IP 地址，请先移出后端服务器组后再修改内网 IP；
- 如果云主机作为静态路由的下一跳，必须先删除静态路由再修改内网 IP。

操作步骤

1.登录控制中心；

2.在系统首页，单击"控制中心 >弹性云主机"，进入云主机列表页；

3.选择对应的云主机，在云主机详情页选择对应的主网卡（辅助网卡需要删除），点击更改内网 IP 或者修改 VPC。

7.5.4 修改弹性网卡基本信息、分配 IPv6、辅助私网 IP

操作场景

您可以根据业务需求修改弹性网卡的名称、安全组，管理辅助私网 IP。

前提条件

- 注册天翼云账号，并完成实名认证。具体操作，请参见天翼云账号注册流程。
- 您已经创建虚拟私有云 VPC、子网、弹性云主机。

操作步骤

1.登录控制中心。

2. 在控制中心页面左上角点击 ，选择区域，本文我们选择华东-华东 1。

3. 依次选择“网络”，单击“虚拟私有云”；进入网络控制台页面。

4. 单击“弹性网卡”，进入弹性网卡列表页面。

5. 在弹性网卡列表页找到对应的弹性网卡，点击“修改”，填写名称、安全组、描述，点击“确定”即可完成修改。

6. 在目标弹性网卡的“操作”列，点击“管理辅助私网 IP”。



7. 在弹窗页面可以取消分配主私网 IPv6 地址，取消分配后 IPv6 地址回收。再次分配可以重新获得 IPv6 地址。您也可以分配新的辅助私网 IP（IPv4）或者取消分配。

注意

分配辅助私网 IP 后，您需要登录服务器实例，在实例内部配置已分配的辅助私网 IP 地址。

8. 当您不需要此弹性网卡时，点击弹性网卡列表“删除”，即可删除该弹性网卡。

7.5.5 如何修改内网 IP、切换 VPC

使用场景

当您遇到 VPC 中的两个实例 IP 地址冲突，或者在进行网络重构或迁移时，需要调整 VPC 网络架构、子网划分等，这可能会导致需要修改云主机内网 IP 或更换 VPC。

前提条件

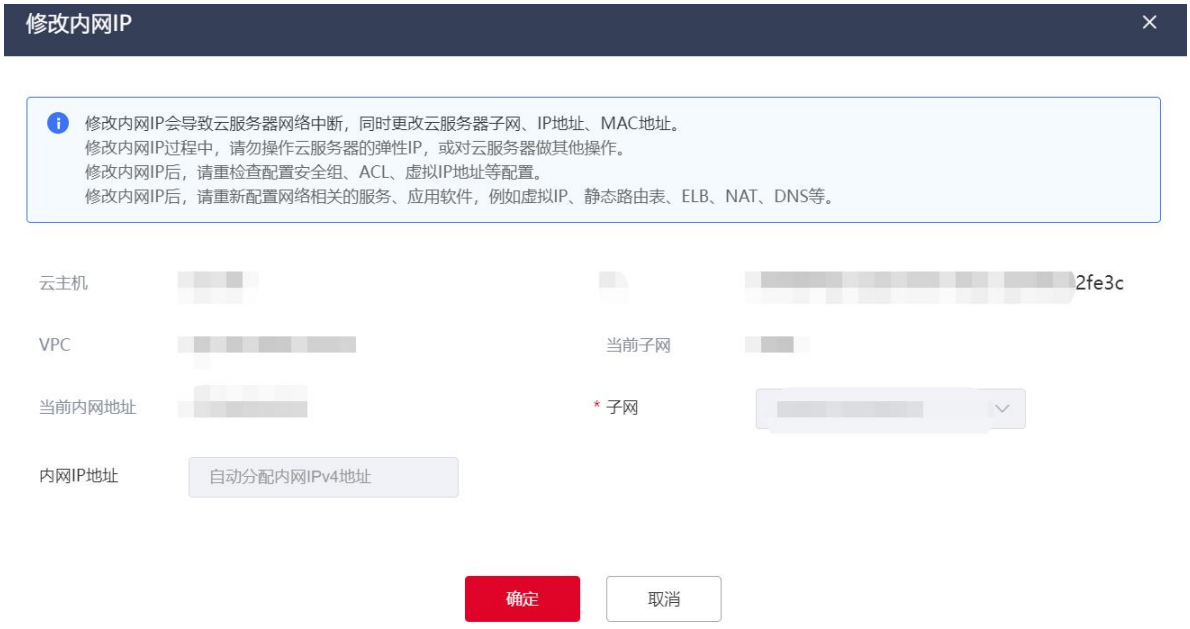
- 弹性云主机处于关机状态。
- 只有主网卡支持修改内网 IP，必须先删除辅助网卡。
- 如果网卡绑定了虚拟 IP 或者 DNAT 规则，需要先解绑。
- 如果网卡上有 IPv6 地址，无法修改（包括 IPv4 和 IPv6 的）内网 IP 地址，请先删除 IPv6 地址。
- 如需修改弹性负载均衡后端服务器的内网 IP 地址，请先移出后端服务器组后再修改内网 IP。
- 如果弹性云主机作为静态路由的下一跳，必须先删除静态路由再修改内网 IP。

操作步骤

1. 登录控制中心。
2. 在控制中心页面左上角点击  ，选择区域，本文我们选择华东-华东 1。
3. 依次选择“计算”，单击“弹性云主机”，进入云主机控制台页面。
4. 在云主机列表，选择需要修改内网 IP 或者更换 VPC 的弹性云主机，在云主机详情页选择“弹性网卡”页签，点击“修改内网 IP”或者“更改 VPC”。



5. 在修改内网 IP 弹框页面或者更换 VPC 弹框页面，输入相关参数，点击“确定”按钮。





注意

- 修改内网 IP 会导致云主机网络中断，同时更改云主机子网、IP 地址、MAC 地址。
- 修改内网 IP 过程中，请勿操作云主机的弹性 IP，或对云主机做其他操作。
- 修改内网 IP 后，请重检查配置安全组、ACL、虚拟 IP 地址等配置。
- 修改内网 IP 后，请重新配置网络相关的服务、应用软件，例如虚拟 IP、静态路由表、ELB、NAT、DNS 等。

7.5.6 开启网卡多队列功能

操作场景

随着网络 IO 的带宽不断提升，单核 CPU 处理网络中断存在瓶颈，不能完全满足网卡的需求，通过开启网卡多队列功能，您可以将弹性云主机中的网卡中断分散给不同的 CPU 处理，以满足网卡的需求，从而提升网络 PPS 和带宽性能。

假设以下场景所述的弹性云主机满足规格和虚拟化类型要求：

- 使用网卡多队列支持列表中的公共镜像创建的弹性云主机，默认已开启网卡多队列，无需执行本节操作。

- 对于私有镜像场景，如果您的外部镜像文件的操作系统在网卡多队列支持列表范围内，需要按照如下流程开启网卡多队列：

a. 将外部镜像文件导入镜像服务控制台。

b. 使用私有镜像创建弹性云主机。

c. 执行网卡多队列的配置脚本。

说明

云主机开启网卡多队列功能后，如果后续有新增或删除网卡，切换 VPC 等操作，需要重新对云主机设置网卡多队列。

网卡多队列支持列表

网卡多队列的支持情况和实例规格、虚拟化类型、镜像的操作系统有关，只有同时满足这些要求，弹性云主机才能开启网卡多队列功能。

- 支持网卡多队列的规格请参考弹性云主机-产品规格，网卡多队列数大于 1 的表示支持多队列。
- 所有公共镜像均支持网卡多队列。

将外部镜像文件导入镜像服务控制台

将外部镜像文件导入镜像服务控制台，详细操作请参考镜像服务-通过镜像文件创建系统盘镜像。

使用私有镜像创建弹性云主机

使用注册好的私有镜像创建弹性云主机。在配置参数时，需要注意以下两点：

- 区域：必须选择私有镜像所在的区域。
- 镜像：选择“私有镜像”，并在下拉列表中选择需要的镜像。

执行网卡多队列的配置脚本

Windows 操作系统使用的半虚拟化驱动 pvdriver 有一套动态调整网卡队列数的策略，会根据 vCPU 数计算队列数并进行动态调整，因此无需对 windows 网卡多队列数进行设置。

Linux 操作系统提供了自动开启网卡多队列功能的配置脚本，配置成功后，弹性云主机支持网卡多队列功能。

下载配置脚本“multi-queue-ctyun”。

下载地址。

执行以下命令，添加执行权限。

```
chmod +x multi-queue-ctyun
```

执行以下命令，将脚本“multi-queue-ctyun”放到目录/etc/init.d 下。

```
mv multi-queue-ctyun /etc/init.d
```

执行以下命令，运行脚本“multi-queue-ctyun”。

```
/etc/init.d/multi-queue-ctyun start
```

运行脚本后，立即生效。但关机后，网卡多队列功能将自动失效。

为了使网卡多队列功能开机自动生效，各个 OS 需要增加开机启动配置：

- CentOS/Redhat/Fedora/EulerOS/Suse/OpenSuse 使用如下命令，增加开机启动项，使网卡多队列配置开机：

```
chkconfig multi-queue-ctyun on
```

- Ubuntu 使用如下命令，增加开机启动项：

```
update-rc.d multi-queue-ctyun defaults 90 10
```

- Debian 使用如下命令，增加开机启动项：

```
systemctl enable multi-queue-ctyun
```

查看网卡队列数

1.开启网卡多队列功能

2.登录弹性云主机。

3.执行以下命令，查看网卡支持和已开启的队列数。

ethtool -l 网卡

示例：

[root@localhost ~]# ethtool -l eth0 #查询网卡 eth0 的队列数

```
Channel parameters for eth0:
Pre-set maximums:
RX:             0
TX:             0
Other:          0
Combined: 4  #表示此网卡最多支持设置开启 4 个队列
Current hardware settings:
RX:             0
TX:             0
Other:          0
Combined: 4 #表示当前开启的是 4 个队列
```

7.5.7 虚拟 IP 概述

虚拟 IP（Virtual IP Address，简称 VIP）是一个从子网中分配的内网 IP 地址，没有分配给真实弹性云服务器网卡。虚拟 IP 地址拥有私有 IP 地址同样的网络接入能力，用户也可以像主私网 IP 地址一样通过虚拟 IP 去访问弹性云服务器。

您可以通过将虚拟 IP 与主备弹性云服务器绑定，根据是否需要访问公网可以为虚拟 IP 绑定一个弹性 IP，配合高可用软件（例如 Keepalived）使用，实现业务的高可用。

详细信息可参考：虚拟私有云-虚拟 IP。

7.5.8 IPv6 带宽

什么是 IPv6 带宽

IPv6 带宽是可以独立购买，将 IPv6 带宽和 VPC 网络中资源实例上使用的 IPv6 地址绑定，可以实现 VPC 中的云资源通过绑定的 IPv6 地址与互联网互通。IPv6 带宽可绑定的云资源有：弹性云主机、物理机、弹性负载均衡等云资源。

关于 IPv6 带宽的具体介绍及使用限制请参考：IPv6 带宽概述。

购买 IPv6 带宽

当前 IPv6 带宽仅支持通过网络控制台购买，具体操作步骤请参考：购买 IPv6 带宽。

云主机绑定 IPv6 带宽实例

IPv6 开启后，云主机如需访问公网，则需要绑定 IPv6 带宽。当前云主机绑定 IPv6 仅支持通过网络控制台操作。

前提条件

如需绑定 IPv6 带宽，云主机的网卡需绑定已开启 IPv6 地址的 VPC 及子网。

检查云主机网卡绑定的 VPC 是否已开启 IPv6

说明

本操作只涉及可用区资源池（云主机创建时有“可用区”选项），若您的资源在区域资源池（云主机创建时无“可用区”选项），请跳过这一步骤。

1. 登录云主机控制台，点击云主机 ID，进入云主机详情页。

2. 点击云主机详情信息中的 VPC 字段，跳转至网络控制台-虚拟私有云详情页。

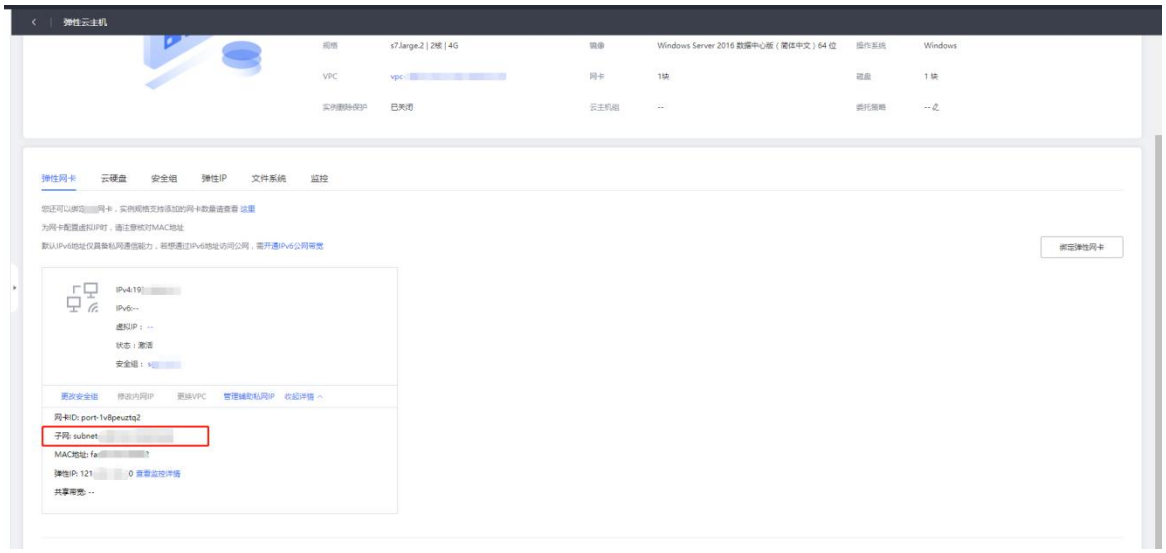
3. 虚拟私有云详情信息中，若“开启 IPv6”开关已打开、“VPC 网段（IPv6）”字段展示 IPv6 地址信息，则该 VPC 已开启 IPv6 地址。



检查云主机网卡绑定的子网是否已开启 IPv6

1. 登录云主机控制台，点击云主机 ID，进入云主机详情页。

2. 在云主机详情页-弹性网卡页签中，点击“展开详情”，查看云主机的子网名称。



3. 前往网络控制台-子网页，在子网列表中通过子网名称检索到上述云主机的子网，点击子网名称，进入子网详情页。

4 若详情信息中“子网网段（IPv6）”字段展示了 IPv6 网段信息，则说明子网已开启 IPv6。

5. 若当前子网未开启 IPv6，如需开启可点击子网列表-操作列-“修改”按钮，在修改子网弹窗中勾选“子网 IPv6 网段”复选框，点击“确定”，即可开启子网 IPv6 网段。



操作步骤

当前云主机绑定 IPv6 仅支持通过网络控制台操作，具体操作步骤请参考：绑定实例。

云主机解绑 IPv6 带宽实例

当前云主机解绑 IPv6 仅支持通过网络控制台操作。

具体操作请参考：解绑实例。

7.6 弹性 IP

7.6.1 弹性 IP 概述

弹性 IP（Elastic IP Address，简称 EIP）是可以独立申请的公网 IP 地址，将弹性 IP 地址和子网中关联的各项云资源绑定和解绑，可以实现 VPC 中的云资源通过固定的公网 IP 地址与互联网互通。弹性 IP 产品既包含公网 IP 地址，也可以提供公网带宽服务，与其他带宽产品组合使用，可以达到访问公网与节省成本的目的。

弹性 IP 需与绑定的云资源在同一个区域（Region）进行使用，不允许跨区域（Region）使用弹性 IP。

详细信息请参考：弹性 IP。

7.6.2 绑定弹性 IP

操作场景

云主机需要使用公网，访问公网资源，或对外发布服务。

前提条件

- 已购买弹性 IP 与弹性云主机
- 弹性云主机有未绑定弹性 IP 的网卡。
- 弹性 IP 处于未绑定的状态。

操作须知

绑定弹性 IP 不需要对虚拟机进行关机，绑定成功后实时生效。

操作步骤

1. 登录控制中心。

2. 在控制中心左上角选择地域，此处选择华东-华东 1。

3. 选择“计算>弹性云主机”，进入云主机控制台。

4. 在云主机列表中，点击待绑定弹性 IP 的云主机名称，进入云主机详情页面。5. 选择弹性 IP 页签，单击“绑定弹性 IP”，在弹出页面选择网卡，选择待绑定的弹性 IP，单击“确定”，完成弹性 IP 的绑定。



6. 绑定成功后，可在弹性 IP 页签看到已绑定的弹性 IP。



说明

后绑定的弹性 IP 和云资源如果计费方式、到期时间不一致，可能导致其中一方到期而影响整体业务。包年包月订购的资源您需要格外关注资源到期时间，并在到期前进行资源续费。按量计费的资源需要关注账户余额，在余额用尽前及时充值。

常见问题

为什么 EIP 无法绑定到云主机实例上？

EIP 无法绑定到云主机实例的常见原因如下：

- EIP 和云主机实例不在同一个地域。
- 云主机实例已经分配了固定公网 IP 或绑定了其他 EIP。
- 云主机实例状态异常，只有运行中或已关机状态的云主机实例才能绑定 EIP。

7.6.3 解绑弹性 IP

操作场景

如果您的云主机不再需要使用公网，可参考本文章解绑弹性 IP。

前提条件

只有已绑定云主机的弹性 IP 地址才可以进行解绑操作。

操作须知

弹性 IP 解绑后不会暂停收费。

操作步骤

1. 登录控制中心。
2. 在控制中心左上角选择地域，此处选择华东-华东 1。
3. 选择“计算>弹性云主机”，进入云主机控制台。
4. 在云主机列表中，点击待解绑弹性 IP 的云主机名称，进入云主机详情页面。
5. 选择弹性 IP 页签，在待解绑弹性 IP 的操作栏单击“解绑”，在弹出页面单击确定，完成弹性 IP 的解绑。



7.6.4 更换弹性 IP

前提条件

弹性云主机已绑定弹性 IP；

已购买新的弹性 IP，购买流程参见申请弹性 IP。

操作步骤

当您的弹性云主机在使用过程中需要换绑弹性公网 IP 时，需要先解绑当前的弹性 IP，再绑定新的弹性 IP。

解绑当前弹性 IP

1. 进入云主机控制台，在产品服务列表点击“弹性云主机”按钮。
2. 在弹性云主机列表找到要解绑弹性 IP 的云主机，点击主机名称进入到该云主机详情页。
3. 找到弹性 IP，点击操作列的“解绑”按钮，在弹出的提示框中选择“确定”即可完成解绑。

绑定新的弹性 IP

1. 进入控制台，在服务列表中点击“弹性 IP”按钮。
2. 在弹性 IP 列表找到待绑定的新弹性 IP 地址，单击操作列中的“绑定”按钮。
3. 在弹出窗口中，如果需要绑定云主机，选择“一对一映射”。
4. 在弹窗中的资源列表中，选择需要绑定的主机资源，选择完毕后点击【确定】完成绑定操作。
5. 绑定后，在该弹性云主机详情中可以看到已绑定新的弹性 IP。

说明

后绑定的弹性 IP 和云资源如果计费方式、到期时间不一致，可能导致其中一方到期而影响整体业务。包年包月订购的资源您需要格外关注资源到期时间，并在到期前进行资源续费。按量计费的资源需要关注账户余额，在余额用尽前及时充值。

7.6.5 修改弹性 IP 带宽

操作场景

如果您在使用弹性 IP 地址的过程中，使用场景发生变化，对带宽需求发生改变，您无须重新购买弹性 IP，只要对正在使用中的弹性 IP 进行变配即可满足要求。

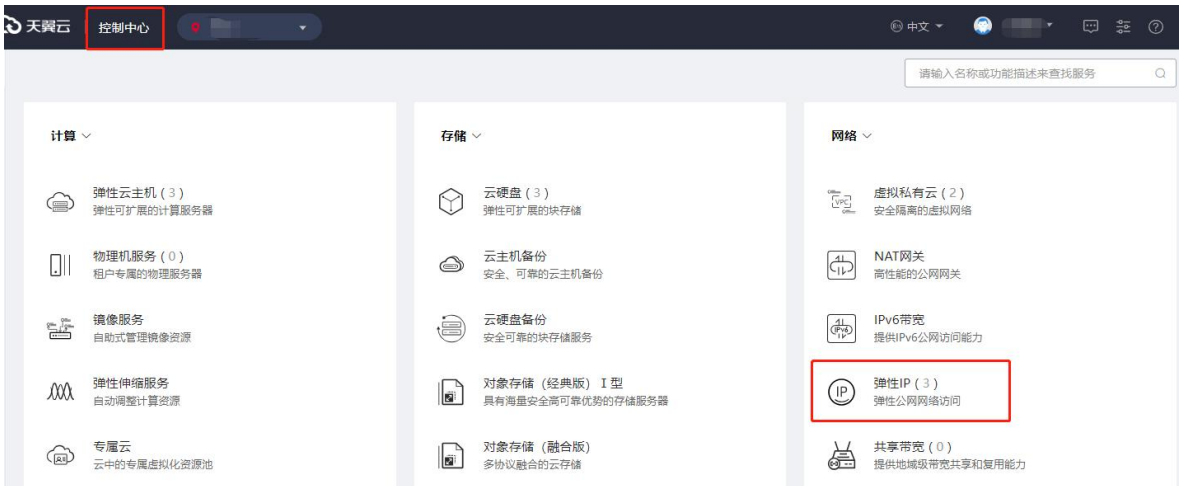
前提条件

要修改带宽的弹性 IP 不能是已过期状态，只有未过期的弹性 IP 才可以修改带宽，已过期的弹性 IP 要想修改带宽必须先完成续费。

弹性 IP 处于“未绑定”状态，如果已绑定需要优先解绑。

操作步骤

1. 进入控制中心，在服务列表中点击“网络>弹性 IP”。



2. 在弹性公网 IP 列表找到需要变配的弹性 IP，鼠标移入操作列中的“更多>变配”。



3. 在弹出的界面中按照您的需要修改带宽，下方显示变配差价。修改完成后点击“确定”，根据提示完成订单即可。



7.6.6 使用 SNAT 访问公网

背景介绍

当您有大量需要访问公网的弹性云主机，您可以通过 NAT 网关的 SNAT 功能统一公网出口，同时避免云主机 IP 直接暴露在公网上。

准备工作

环境准备

- 已创建 VPC，具体操作参见虚拟私有云－创建 VPC 、子网搭建私有网络。
- 该 VPC 下已创建 1 台不带弹性 IP 的弹性云主机。具体操作参见弹性云主机－创建弹性云主机。

操作步骤

步骤一：购买弹性 IP

步骤二：购买 NAT 网关

步骤三：配置路由（仅部分资源池需要）

步骤四：配置 SNAT 规则

步骤五：测试连通性

步骤一：购买弹性 IP

步骤说明

弹性 IP 主要用来绑定 NAT 网关的 SNAT 规则，以实现访问公网的功能。

操作步骤

1. 登录天翼云控制台，选择“网络>弹性 IP”。
2. 进入弹性 IP 控制台，点击右上角“申请弹性 IP”。
3. 按需求选择带宽规格，单击“下一步”。
4. 确定规格，选择我已阅读并同意相关协议，单击“确认下单”，完成弹性 IP 创建。

步骤二：购买 NAT 网关

步骤说明

购买公网 NAT 网关必须指定公网 NAT 网关所在 VPC、子网。此处指定弹性云主机所在的 VPC 和子网。

操作步骤

1. 登录天翼云控制台，选择“网络>NAT 网关”，
2. 进入 NAT 网关控制台，点击右上角“创建 NAT 网关”
3. 选择付费方式，填写名称，选择可用区，VPC 选择环境准备中的创建的 VPC，选择规格，点击“下一步”。
4. 确认规格，选择我已阅读并同意相关协议，单击“确认下单”，完成 NAT 网关的创建。

步骤三：配置路由（仅部分资源池需要）

步骤说明

部分资源池在购买 NAT 网关后，自动加载路由到 VPC 中，不需要此步操作。

部分资源池需要在默认路由中添加路由指向 NAT 网关。

关于资源池与操作的对应关系，请参考虚拟私有云相关产品资源池区别。

操作步骤

1. 添加默认路由指向 NAT 网关。单击“虚拟私有云”，进入虚拟私有云控制台，选择环境中创建的 VPC，点击名称打开 VPC 详情页，在子网页签点击子网名称，进入子网详情页。
2. 在子网详情页单击“路由管理”页签，点击已绑定的“默认路由表”进入路由规则页面。
3. 在路由规则页面单击“创建”，在弹出页面，选择

- IP 类型：IPv4
- 目的地址：0.0.0.0/0

- 下一跳类型：NAT 网关
- NAT 网关：选择刚创建的 NAT 网关

4. 点击确定，完成默认路由指向 NAT 网关。

注意

在同一个子网内云主机同时绑定公网 IP 和 SNAT 规则时，由于指向 NAT 网关的路由优先级高于指向 IPv4 网关的系统路由，默认会通过 SNAT 访问公网，云主机绑定的 EIP 无法访问公网。因此，不建议同一个子网内的云主机同时绑定公网 IP 或者 NAT 网关。

具体内容参见弹性 IP-如何通过弹性公网 IP 或 NAT 网关访问公网。

步骤四 ：配置 SNAT 规则

步骤说明

公网 NAT 网关创建成功后，您需要创建 SNAT 规则。通过创建 SNAT 规则，您可以将该子网下的云主机通过共享弹性公网 IP 访问互联网。

操作步骤

1. 单击上一步创建的 NAT 网关，进入 NAT 网关详情页，下拉至 SNAT 页签，点击添加 SNAT 规则。
2. 在弹出页面选择配置：
- 源网段类型：选择 VPC 内子网
- 子网：选择云主机所在子网
- 弹性 IP：选择步骤一创建的弹性 IP

3. 单击“确定”，完成 SNAT 规则的添加。
4. 等待 SNAT 规则状态变为运行中，即完成 SNAT 规则配置。

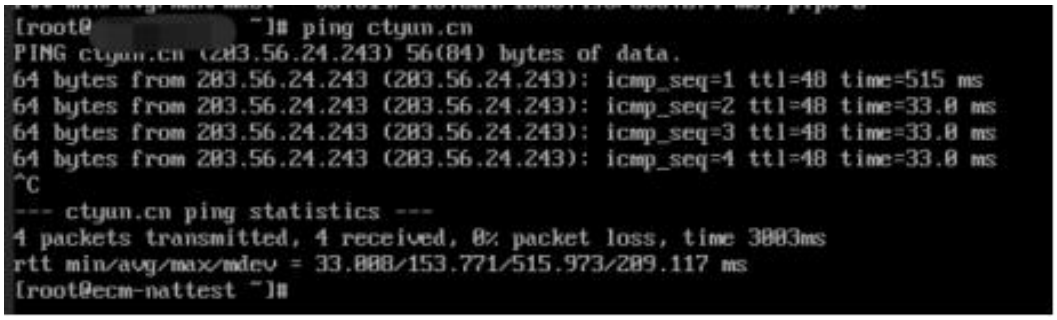
步骤五 ：测试连通性

步骤说明

我们通过验证未绑定EIP的弹性云主机是否可以通过NAT网关访问公网来测试网络连通性。

操作步骤

1.登录天翼云控制台，选择“计算>弹性云主机”。
2.进入弹性云主机控制台，选择准备工作中创建的不带弹性IP的弹性云主机，点击右侧“远程登录”登录弹性云主机。
3.登录完成，通过ping外网地址，测试云主机是否能够通过NAT网关访问公网。



说明弹性云主机可以连接公网。

7.7 密码和密钥对

7.7.1 密码

7.7.1.1 密码使用场景介绍

密码

密码指使用设置初始密码方式作为云主机的鉴权方式，此时，您可以通过用户名密码方式登录云主机，Linux操作系统时用root用户和初始密码，Windows操作系统时用Administrator用户和初始密码。

使用场景

密码在弹性云主机系统中用于保护客户重要私人信息，为确保账户安全，建议妥善保管密码。在使用弹性云主机服务过程中，您会在以下场景遇到密码相关设置：

- 创建弹性云主机时设置密码。

  如果用户购买弹性云主机进行高级配置时选择“密码”作为登陆方式，您可以单击“立即创建”创建密码。具体步骤可参见创建弹性云主机。
- 重置弹性云主机的登录密码。

  如果用户忘记密码或密码已过期，可通过“重置密码”进行密码重置。具体步骤可参见在控制台重置弹性云主机密码。

7.7.1.2 在控制台重置密码

操作场景

如果在创建弹性云主机时未设置密码，或者密码丢失或过期，您可以通过以下步骤在控制中心重置密码。

前提条件

- 弹性云主机处于开机状态。
- 弹性云主机在密码丢失或过期前已安装 qemu-guest-agent。
- 弹性云主机的 qemu-guest-agent 正常运行，否则无法使用重置密码功能。

  Linux 云主机执行 systemctl status qemu-guest-agent 查看运行状态，

  Windows 云主机在 cmd 执行 services.msc 进入服务窗口,查看运行状态。

说明

公共镜像创建的弹性云主机默认安装 qemu-guest-agent。没有安装 qemu-guest-agent 或未正常运行的云主机不支持重置密码，您可以通过重装云主机为其设置新密码。

- 弹性云主机使用的 VPC 网络 DHCP 不能被禁用。
- 弹性云主机的网络正常通行。

单台云主机重置密码操作步骤

1.登录控制中心，单击顶部的  ，选择“地域”。
2.单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
3.将鼠标移动至目标云主机的“操作”列，选择“更多”。
4.点击“重置密码”，在“重置密码”弹框中输入要重置的云主机新密码。
5.点击“确定”，完成弹性云主机密码重置。

多台云主机重置密码操作步骤

1.登录控制中心，单击顶部的  ，选择“地域”。
2.单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
3.选中多台需要重置密码的云主机，点击云主机列表左上角的“更多-重置密码”按钮。
4.在弹框中输入要重置的云主机新密码。
5.点击“确定”，完成多台弹性云主机密码重置。

7.7.2 密钥对

7.7.2.1 密钥对使用场景介绍

密钥对

密钥对，也称为 SSH 密钥对，是一种用于远程登录云主机的身份验证方式，与传统的用户名和密码登录方式有所区别。

密钥对由公钥和私钥组成，公钥保存在系统中，私钥由用户保存在本地。如果用户将公钥配置在 Linux 云主机中，便可以使用私钥登录 Linux 云主机，无需输入密码。通过使用密钥对登录 Linux 云主机，可以防止密码被拦截或破解导致帐户密码泄露，从而提高 Linux 云主机的安全性。

您可以通过创建、导入、绑定、查看、重置、替换、解绑、删除密钥对等操作管理密钥对。

使用场景

建议用户在购买弹性云主机时选择使用密钥对进行身份认证。

通过将公钥与云主机关联，实现 SSH 认证，确保只有持有对应私钥的用户能够访问主机。这提高了远程访问的安全性，防止未经授权的访问。

如果用户购买的是 Linux 操作系统的弹性云主机，可以直接使用密钥对进行远程登录到云主机，无需输入密码。

- 创建弹性云主机时，选择 Linux 镜像，选择密钥对登录方式，详细操作，请参见新手入门：购买并登录 Linux 弹性云主机的“高级配置“
- 创建弹性云主机完成后，通过“为云主机绑定/更换密钥对”的方式为云主机绑定密钥对。

创建密钥对操作指引

您可以使用已有密钥对或新建一个密钥对，用于远程登录身份验证。

- 新建密钥对

如果没有可用的密钥对，需要新建一个密钥对，在登录弹性云主机时提供私钥进行鉴权实现远程登录身份验证。可用以下方法创建密钥对：
在控制台创建密钥对：公钥自动保存在系统中，私钥由用户保存在本地。

注意
为保证云主机安全，私钥只能下载一次，私钥丢失后将无法登录云主机，请您妥善保管。

新建密钥对后可以通过查看 SSH 密钥对列表页，点击列表右上角的刷新按钮，可以查看 SSH 密钥对的创建情况。

- 使用已有密钥对

对如果您希望使用本地已有的密钥对，可以在管理控制台导入密钥对公钥，由系统维护您的公钥文件。具体操作请参见导入密钥对。

约束与限制

- 密钥对仅支持远程登录 Linux 云主机。
- 密钥对只能在本区域的云主机使用。
- 通过管理控制台创建的基于 SSH2 协议的密钥对，仅支持“RSA-2048”加解密算法。
- 通过外部导入的密钥对支持的加解密算法为：
  - RSA-1024
  - RSA-2048
  - RSA-4096
  - ECDSA-256
  - ECDSA-384
  - ECDSA-512
  - ED25519-256
- 为保证云主机安全，私钥只能下载一次，私钥丢失后将无法登录云主机，请您妥善保管。

7.7.2.2 在控制台创建密钥对

操作场景

您可以通过管理控制中心来创建密钥对。创建完成后，公钥将自动保存在系统中，而私钥则由用户保存在本地。

操作步骤

1. 登录控制中心。
2. 单击“左侧导航栏>服务列表”，选择"计算>弹性云主机"。
3. 在左侧导航栏，单击"SSH 密钥对"，进入密钥对列表。

4.单机右上角的"创建密钥对"按钮，输入“名称”并选择“企业项目”后点击“确定”按钮开始创建。

5.创建密钥对成功后，弹窗会自动下载“XXXX.pem”密钥到本地。

注意

为保证云主机安全，私钥只能下载一次，私钥丢失后将无法登录云主机，请您妥善保管。

6.返回 SSH 密钥对列表页，点击列表右上角的刷新按钮，查看 SSH 密钥对的创建情况。

7.7.2.3 导入密钥对

操作场景

如果您已经在本地生成了密钥对（例如使用 PuTTYgen 工具），您可以将公钥导入到控制中心，让系统维护您的公钥文件。这样，您就可以轻松地在创建云主机时选择导入的密钥对进行身份认证。

操作步骤

1.登录控制中心，单击顶部的  ，选择“地域”。

2.单击左侧导航栏“产品服务列表”，选择"计算>弹性云主机"。

3.在左侧导航栏，单击"SSH 密钥对"，进入密钥对列表。

4.单机右上角的"导入密钥对"按钮，在“文件”选框上传公钥文件。

说明

最多支持 1024 字符长度（包括 1024 字符）的公钥导入。

仅支持 RSA 类型的密钥。

1.复制内容到“公钥内容”文本框，手动输入公钥名称。

2.下拉选择“企业项目”。

3.点击“确定”按钮，完成 SSH 密钥对导入。

4.返回 SSH 密钥对列表页，查看导入的 SSH 密钥对。

7.7.2.4 为云主机绑定/更换密钥对

操作场景

绑定密钥对适用于更新已创建云主机的密钥对。

如果您创建的一台 Linux 云主机未绑定 SSH 密钥对，本功能可以为该云主机绑定密钥对；如果您创建的一台 Linux 云主机已绑定 SSH 密钥对，本功能可以为该云主机更换密钥对。

使用限制

本功能仅适用于 Linux 类型操作系统，不适用于 Windows 类型操作系统。

本功能仅对状态处于“运行中”的云主机生效。

如果您使用的私有镜像未安装 qga(通常由用户自己上传的镜像不具备 qga)，则此功能无法生效。

操作步骤

绑定密钥对有两个操作入口。

入口一 ：通过云主机控制台执行绑定密钥对，本入口仅对单台云主机操作。

1.登录控制中心，进入弹性云主机控制台；

2.选择一台处于“运行中”的 Linux 云主机，点击操作列的【更多】；

3.点击【绑定/更换密钥对】；

4.选择一个密钥对，点击【确定】，完成更新。

入口二 ：通过 SSH 密钥对控制台执行绑定密钥对，本入口可以对多台云主机批量操作。

1.登录控制中心，进入弹性云主机控制台，左侧导航栏点击【SSH 密钥对】，进入 SSH 密钥对列表页；
2.选择一个 SSH 密钥对，点击操作列的【绑定密钥对】，弹出绑定密钥对操作界面；
3.选择需要绑定密钥对的云主机，可多选操作；
4.点击【确定】，完成更新。

注意
批量绑定密钥对时，请不要选择非“运行中”状态的云主机、操作系统为 Windows 的云主机，否则会绑定失败。

7.7.2.5 为云主机解绑密钥对

操作场景

解绑密钥对适用于将已绑定云主机的密钥对与该云主机解绑。

操作解绑后，将无法使用原密钥对登录云主机。如果您需要再次登录云主机，需要绑定新的密钥对，或者使用密码登录。

使用限制

本功能仅适用于 Linux 类型操作系统，不适用于 Windows 类型操作系统。

本功能仅对状态处于“运行中”的云主机生效。

如果您使用的私有镜像未安装 qga（通常由用户自己上传的镜像不具备 qga），则此功能无法生效。

操作步骤

解绑密钥对有两个操作入口。

入口一 ：通过云主机控制台执行解绑密钥对，本入口仅对单台云主机操作。

1.登录控制中心，进入弹性云主机控制台；
2.选择一台处于“运行中”的 Linux 云主机，点击操作列的【更多】；
3.点击【解绑密钥对】；
4.浏览提示后，点击【确定】，完成解绑。

入口二 ：通过 SSH 密钥对控制台执行解绑密钥对，本入口可以对多台云主机批量操作。

1.登录控制中心，进入弹性云主机控制台，左侧导航栏点击【SSH 密钥对】，进入 SSH 密钥对列表页；
2.选择一个 SSH 密钥对，点击操作列的【解绑密钥对】，弹出解绑密钥对操作界面；
3.本界面展示已绑定该密钥对的云主机，选择需要解绑密钥对的云主机，可多选操作；
4.点击【确定】，完成解绑。

注意
批量解绑密钥对时，请不要选择非“运行中”状态云主机，否则会解绑失败。

7.8 云主机快照

7.8.1 云主机快照管理

注意
因业务调整，不再推荐使用云主机快照。详情请查看公告，点击公告链接。
云主机快照属于公测产品，公司将不对任何服务可用性、可靠性做出承诺，公司有权根据运营情况对功能做出调整。您选择使用云主机快照功能代表您同意了公测协议。

云主机快照
概述

天翼云云主机快照是一种云主机数据备份方式，云主机快照服务可以备份或者恢复云主机所有云硬盘的数据，常用于重要变更前的数据临时备份、数据恢复等。

云主机快照不支持跨可用区，跨地域使用。相关需求请使用云主机备份，镜像服务等功能。

云主机快照主要适用于系统配置变更时的临时备份场景。鉴于其特性，并不推荐将其用于日常数据备份。若需进行日常数据备份操作，建议优先选择云主机备份或云硬盘备份方案，以保障数据的安全性与完整性。

云主机快照能够记录云主机在某个时刻的数据，通过回滚将数据恢复至快照时间点。还可以通过快照快速创建出多个具有相同数据的云主机用于业务部署。

云主机快照依赖原云主机，退订删除云主机需要先删除快照，所以快照数据不能单独保留。

一致性快照

云主机快照支持非一致性快照和文件系统一致性快照，在创建快照和配置快照策略时，可以根据用户需求选择文件系统一致性快照选项，选中时创建文件系统一致性快照。云主机快照不支持应用一致性快照，应用一致性需要根据用户自身应用特点从应用架构和实现去保障，比如停用应用，应用一致性同步，分布式，主备，异地多活等手段。

文件系统一致性快照会先同步未落盘数据，然后冻结文件系统，快照完成后解冻文件系统。 非一致性不会执行同步未落盘数据和冻结文件操作，不能保证所有云硬盘的数据时序一致性。

在创建文件系统一致性快照时，可能会对业务系统产生影响。因此，建议用户选择在业务低峰期或主机空闲时段进行操作。若在高 IO 等场景下创建一致性快照，可能导致文件系统长时间冻结，进而影响客户业务。这种情况下，推荐采用非一致性快照方式。

注意

文件系统一致性快照表示创建时会尝试同步未落盘数据，冻结文件系统，同步或冻结失败时会生成非一致性快照。创建时会对业务系统造成影响，如需避免影响建议使用非一致性方式或系统空闲时使用。

非一致性快照表示创建时不会同步未落盘数据，冻结文件系统。

关机情况下系统已满足同步数据和冻结文件，会生成一致性快照。Windows 系统不支持同步数据和冻结文件，开机情况下生成的都是非一致性快照。

适用场景

● 数据保护和恢复：快照可以用作数据的临时备份和恢复的手段。当数据发生意外删除、损坏或错误修改时，可以使用快照来恢复到之前的状态。

● 数据恢复测试：可以使用快照来进行数据恢复测试。在进行关键操作之前，先创建一个快照，然后在测试过程中可以随时回滚到快照状态，以确保操作的安全性和可靠性。

● 批量部署多个业务：通过同一个快照可以快速创建出多个具有相同数据的云主机，从而可以同时为多种业务提供数据资源。例如数据挖掘、报表查询和开发测试等业务。

使用限制

限制项	限制说明
单个云主机快照总数	一个云主机最多支持创建 7 个快照。
跨地域，跨可用区	云主机快照不支持跨可用区，跨地域使用。相关需求请使用云主机备份，镜像服务。
重装系统	重装系统前需要先删除快照。
保留	云主机快照依赖原云硬盘，云主机删除前，需要删除云主机快照。
创建限制	建议在业务低峰或主机空闲时段创建云主机快照，以免影响业务。

限制项	限制说明
一致性	在创建文件系统一致性快照时，可能会对业务系统产生影响。因此，建议用户选择在业务低峰期或主机空闲时段进行操作。若在高 IO 等场景下创建一致性快照，可能导致文件系统长时间冻结，进而影响客户业务。这种情况下，推荐采用非一致性快照方式。
挂载卸载盘	增加新盘后，云主机快照仍然可以用于恢复，不影响新增的盘。卸载盘后，主机快照详情的盘快照会删除，被卸载的盘不能恢复，再次挂载也不能恢复。

创建快照

前提条件

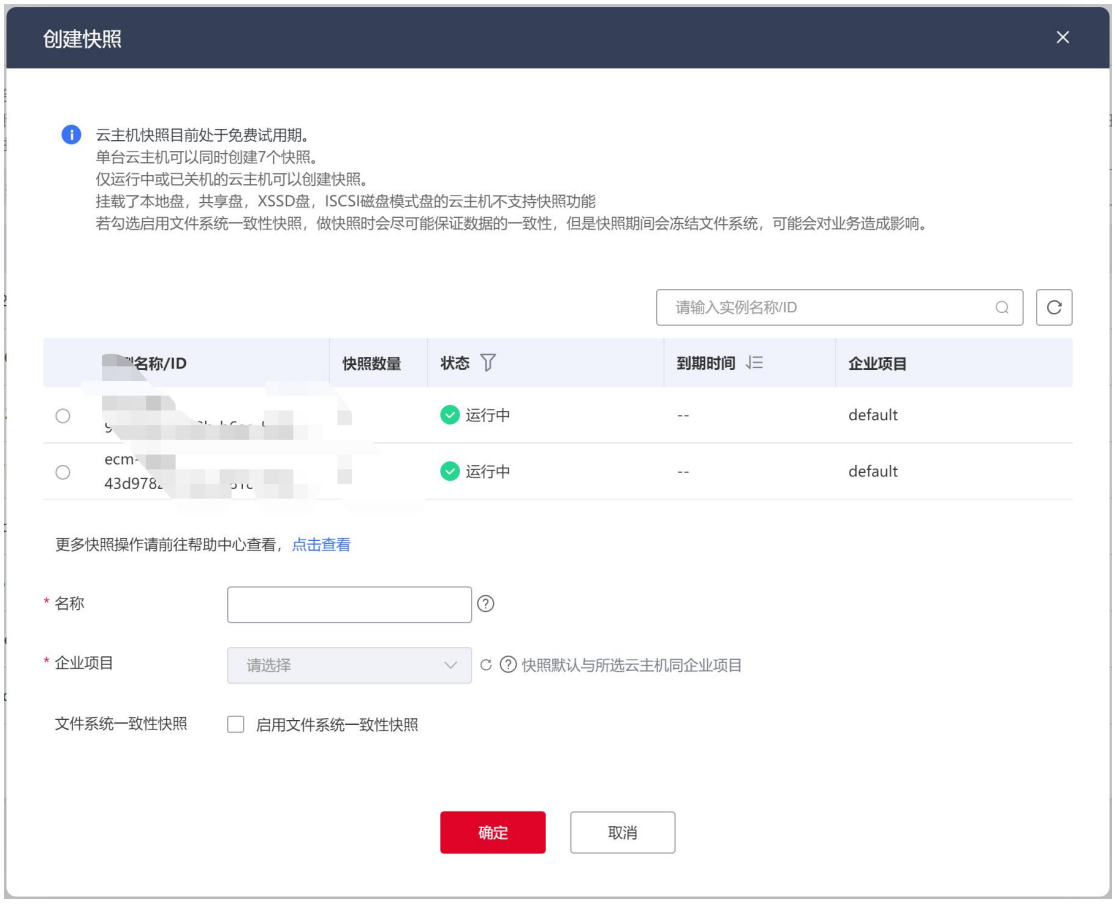
- 单台云主机最大支持创建 7 个快照。
- 仅运行中或已关机的云主机可以创建快照。
- 当您需要创建云主机快照时，您可以通过手动创建和策略定时自动创建。
- 挂载了本地盘，共享盘，XSSD 盘，ISCSI 磁盘模式盘的云主机不支持快照功能，同时绑定了快照策略的云主机不支持挂载本地盘，共享盘，XSSD 盘，ISCSI 磁盘模式盘，如果有需求请解绑快照，再挂载。

方式一：在快照列表页创建快照

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在快照列表页点击“创建快照”，打开创建快照弹窗。
6. 在弹窗中选择云主机，输入快照名称，点击确认。

说明

快照名称长度为 2-63 字符，文件系统一致性快照选中时支持文件系统一致性快照。



方式二：在主机列表页创建快照

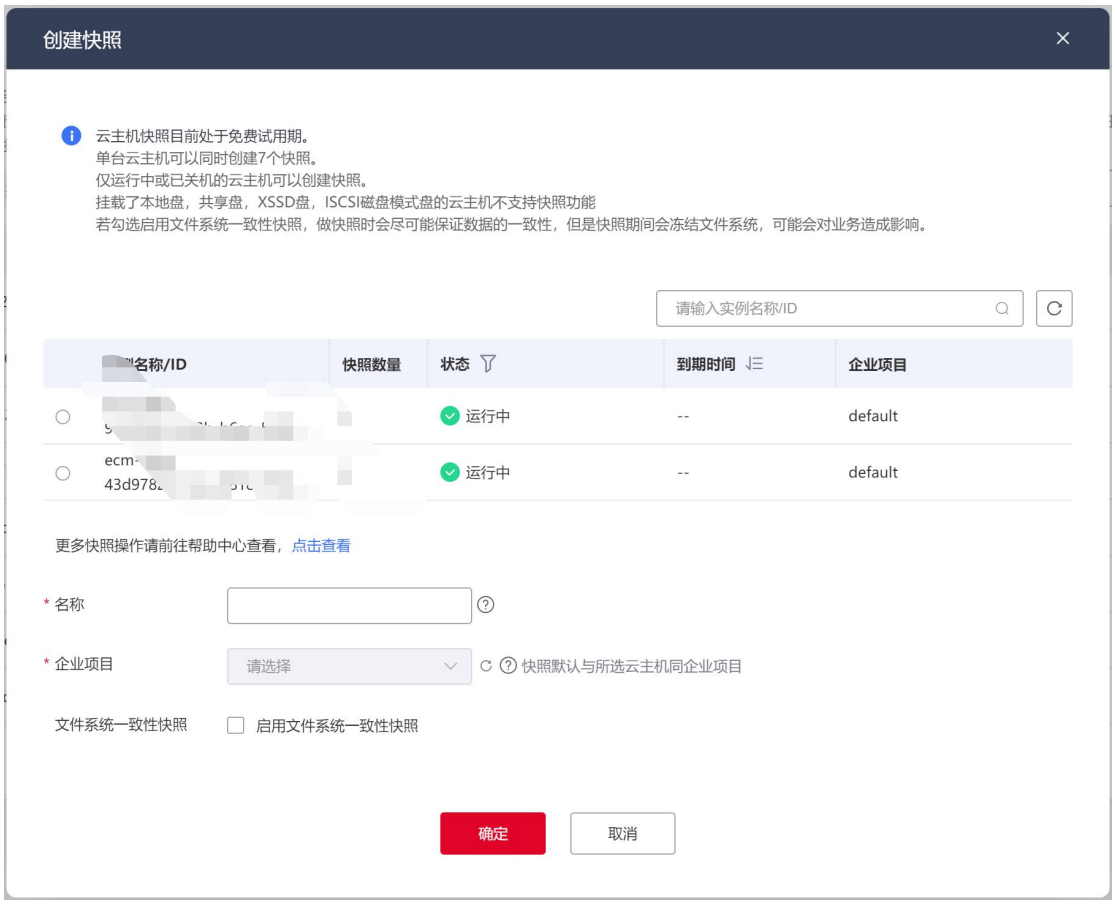
1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在云主机列表页，在云主机列表中，选择需要创建快照的云主机，在操作栏中，点击“创建快照”。



5. 在弹窗中输入快照名称，点击确认。

说明

快照名称长度为 2-63 字符，文件系统一致性快照选中时支持文件系统一致性快照。



恢复数据

恢复数据指通过云主机快照将云主机恢复至快照创建时的配置和状态。

前提条件

- 快照状态必须为“正常”。
- 要恢复的云主机必须为关机状态。
- 当您需要将云主机恢复至快照创建时的配置和状态时，您可以通过两种方式进行恢复。

方式一：在快照列表页恢复数据

1. 登录控制中心。

2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在云主机快照列表中选择要进行操作的快照，点击右侧“操作”中的“恢复数据”（云主机需处于关机状态）。

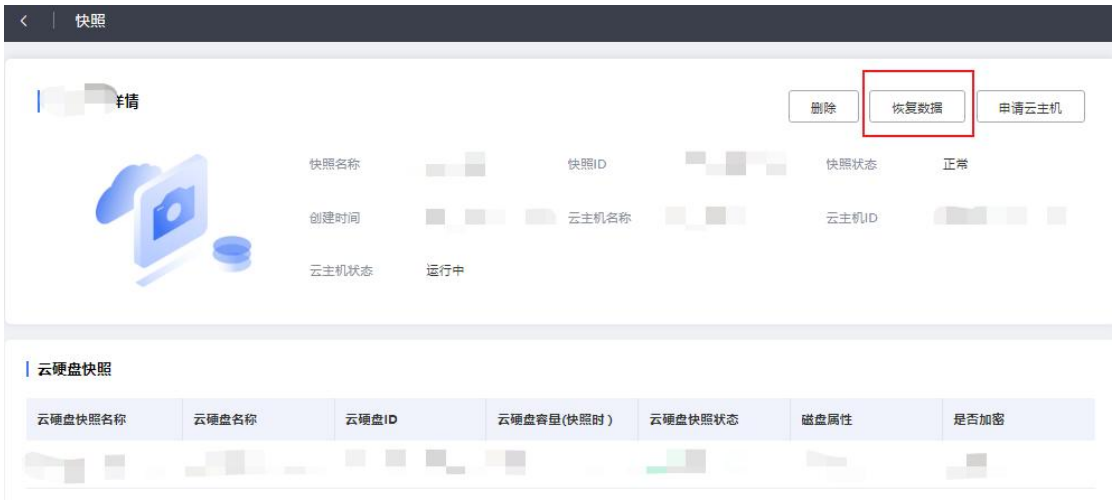


6. 在弹出的恢复数据的弹窗中核对相关信息，点击确定。



方式二：在快照详情页恢复数据

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在快照列表中点击某快照的名称。
6. 在快照详情页点击“恢复数据”。



7. 在弹出的恢复数据的弹窗中核对相关信息，点击确定。



说明

云主机制作快照后，又挂载了新数据盘：可恢复数据，但新创建的数据盘不会被恢复。

云主机制作快照后，又卸载了数据盘：主机快照中仅保留剩余云硬盘的快照信息，仍可恢复数据。

云主机制作快照后，又升级了规格：可恢复数据，恢复数据不改变云主机规格。

云主机制作快照后，对系统盘或数据盘进行了扩容：可恢复数据，恢复至制作快照时的状态。不影响磁盘的总容量。

利用快照申请云主机

申请云主机指使用快照中的配置和数据状态来创建新的云主机。创建的云主机其

规格、配置、数据等均与快照一致。新创建的云主机与普通云主机一致，可进行变配、扩容、开关机等所有普通云主机可进行的操作。

前提条件

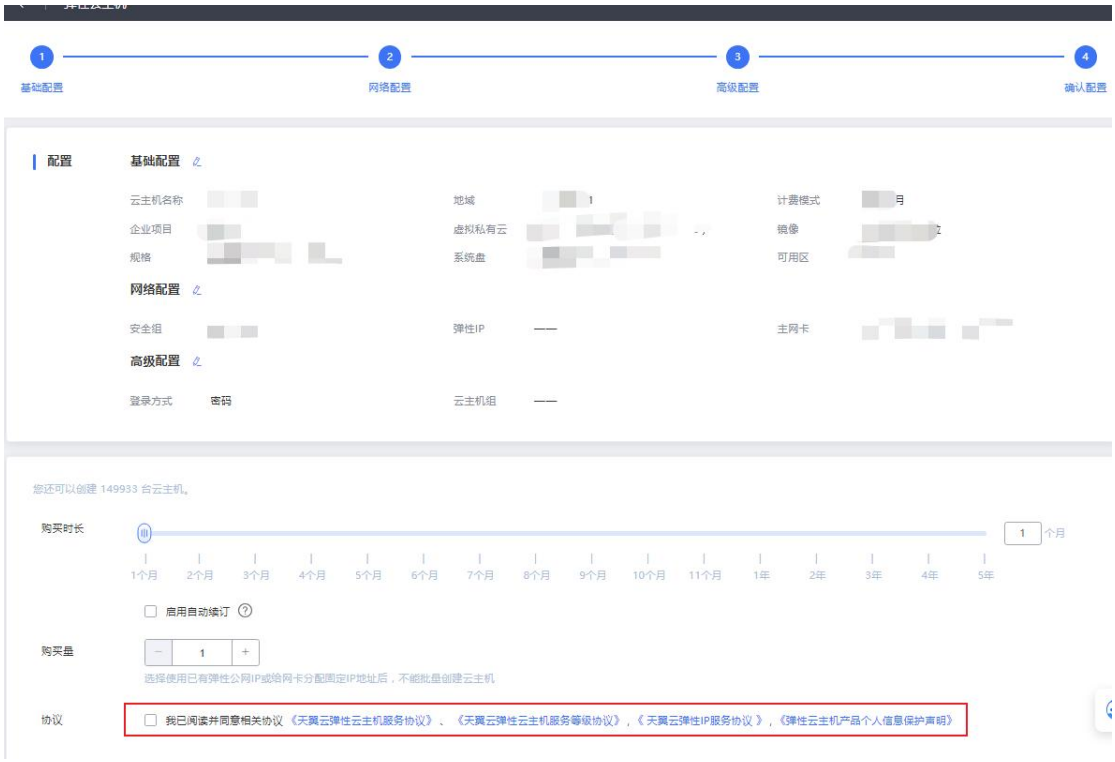
快照状态必须为“正常”。

方式一：在快照列表页申请云主机

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“快照”。
5. 在云主机快照列表中选择要进行操作的快照，点击右侧“操作”中的“申请云主机”。



6. 在跳转出的云主机订购页选择云主机配置（云主机规格、镜像、磁盘大小和类型不可修改）。
7. 单击“立即购买”。
8. 在确认规格页面，您可以查看云主机详情并提交申请。如果您确认配置信息无误，勾选协议，单击“提交订单”。



9. 点击“立即支付”进行付款，付款成功即可创建弹性云主机。弹性云主机创建成功后，您可以在弹性云主机信息页面看到您新创建的弹性云主机。

方式二：在快照详情页申请云主机

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“快照”。
5. 在快照列表中点击某快照的名称。
6. 在快照详情页点击“申请云主机”。



7. 后续确认页面和支付流程等均与申请云主机方式一一致。

删除快照

删除快照指将快照删除，用户列表不可见、不可再使用，请确认后再删除。

前提条件

- 快照状态不能为“创建中”，“恢复数据中”等中间状态。
- 您可以通过两种方式来删除云主机快照。

方式一：在快照列表页进行删除

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“快照”。
5. 在云主机快照列表中选择要进行操作的快照，点击右侧“操作”中的“删除”。



6. 在删除弹窗中核对将要删除的快照信息，单击确定。



方式二：在快照详情页进行删除

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。

3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在快照列表中点击某快照的名称。
6. 在云主机快照详情页单击“删除”。



7. 在删除弹窗中核对将要删除的快照信息，单击确定。

7.8.2 云主机快照策略

注意

因业务调整，不再推荐使用云主机快照。详情请查看公告，点击公告链接。

概述

云主机快照策略通过创建策略并绑定云主机后，实现定时创建云主机快照。云主机快照同时支持按数量和时间保留策略。

策略绑定和解绑云主机不会影响已经创建的快照。

使用限制

- 一个云主机快照策略支持绑定多个云主机，一个云主机只能绑定一个策略。
- 单台云主机最大支持创建 7 个快照。手动创建的快照个数和策略自动创建的快照个数总和达到单台云主机快照个数限制时，快照策略将删除策略最早生成的自动快照。快照策略不会删除手动快照，单台云主机手动快照数量达到限额将会造成快照策略任务失败。您可以在快照策略任务列表查看详细信息。
- 请避免同一时刻或策略的同一时段内对云主机和云硬盘进行快照、备份、制作镜像，否则会造成失败。云主机只能绑定一个策略。

创建快照策略

用户可以创建快照策略，支持设置不同的策略内容。

前提条件

一个账号在一个资源池仅可以同时创建 20 个快照策略。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在快照策略列表页点击“创建快照策略”，打开创建快照策略弹窗。
6. 在弹窗中输入快照名称。
7. 选择是否启用，开启后策略才会执行定时创建快照。
8. 选择快照时间。
9. 设定快照周期。
10. 选择是否启用文件系统一致性快照。
11. 设定保留规则，点击“确定”即可完成快照策略的创建。

说明

快照名称不能使用中文，且长度为 2-32 字符。

快照时间均为整点。

快照周期可按周或按天，按天最多支持每 10 天。

保留规则支持按时间或按数量。



启用/停用快照策略

用户可以手动对快照策略进行启用与停用。启用的快照才可用，停用的快照不可用。

前提条件

只有“停用”状态的策略支持启用，只有“启用”状态的策略支持停用。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。

5. 在快照策略列表，选择“停用”的策略，在操作栏点击“启用”按钮。
6. 在快照策略列表，选择“启用”的策略，在操作栏点击“停用”按钮。



立即执行策略

快照策略除了根据策略设定的时间执行之外，用户也可以手动触发策略的执行。

前提条件

- 快照策略必须为“启用”状态。
- 快照策略必须绑定云主机。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“快照”。
5. 在快照策略列表，选择“启用”状态的策略，在操作栏点击“立即执行”按钮。



修改快照策略

快照策略创建完成后，支持修改其策略内容。

操作步骤

1. 登录控制中心。

2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧的“快照”。
5. 在云主机快照策略列表中选择要进行操作的快照，点击右侧“操作”中的“修改”。



6. 在修改弹窗中填入要修改的信息，单击确定。



删除快照策略

快照策略创建完成后，用户可以删除不再需要的策略。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在云主机快照策略列表中选择要进行操作的快照，点击右侧“操作”中的“删除”。



绑定/解绑云主机

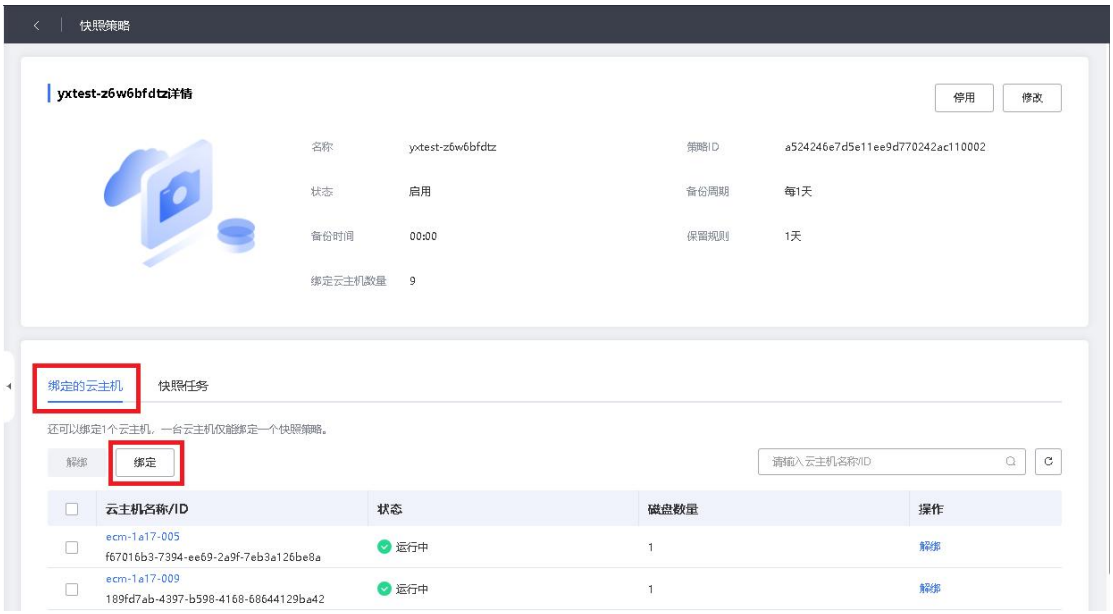
快照策略创建后，需要绑定云主机后策略才能正常执行。云主机绑定后可以进行解绑。

前提条件

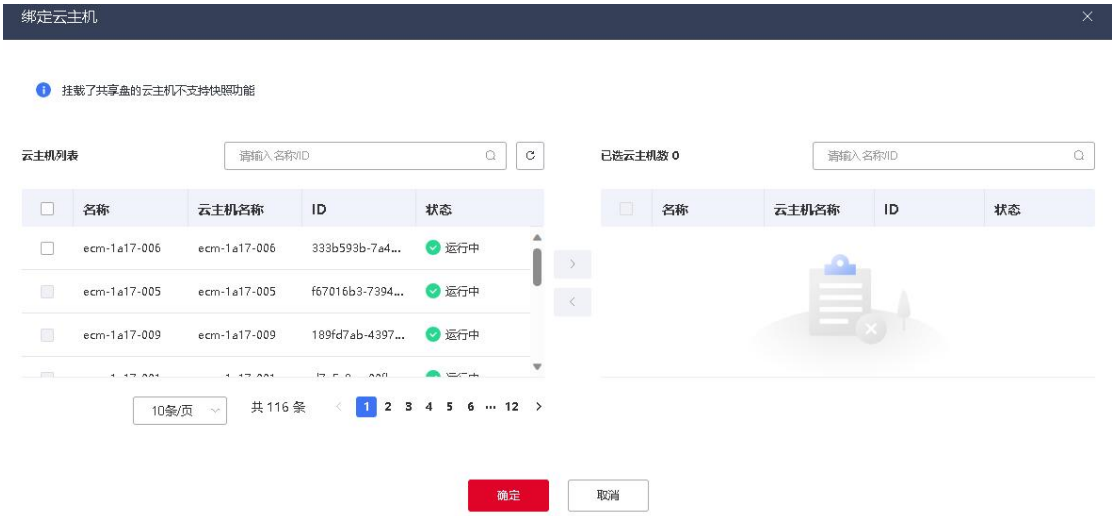
- 运行中的云主机才能绑定到策略。
- 一个快照策略最多只能绑定 10 台云主机。
- 一台云主机仅能绑定一个快照策略。
- 挂载了共享盘的云主机不支持绑定。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在云主机快照策略列表中选择要进行操作的快照，点击策略名称进入详情页。
6. 在绑定的云主机列表，点击按钮“绑定”，进入绑定云主机页面。



7. 在绑定云主机页面，选择需要绑定的云主机，点击“确定”即可绑定云主机。



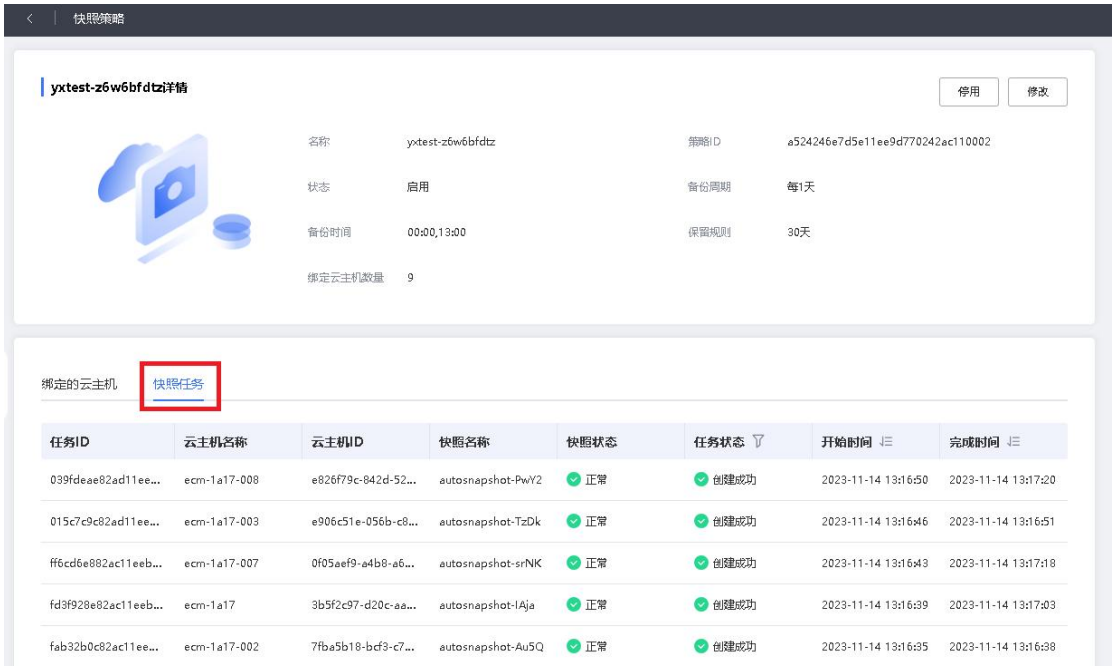
8. 在绑定的云主机列表，选择云主机，点击按钮“解绑”，即可解绑云主机。

查看快照任务

快照策略执行后，用户可以在快照任务列表进行查看任务的执行情况。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 选择左侧导航栏的“快照”。
5. 在云主机快照策略列表中选择要进行操作的快照，点击策略名称进入详情页。
6. 点击“快照任务”，即可查看快照策略的执行情况。



7.9 云主机备份

7.9.1 备份弹性云主机

操作场景

备份服务提供申请即用的服务，使您的数据更加安全可靠。例如，当云主机或磁盘出现故障或者人为错误导致数据误删时，可以自助快速恢复数据。本节操作介绍备份云主机和云硬盘的操作步骤。

使用须知

目前弹性云主机备份可以通过“云主机备份”和“云硬盘备份”功能实现：

- 云主机备份（推荐）：如果是对弹性云主机中的所有云硬盘（系统盘和数据盘）进行备份，推荐使用云主机备份功能，同时对所有云硬盘进行备份，避免因备份创建时间差带来的数据不一致问题。
- 云硬盘备份：如果对指定的单个或多个云硬盘（系统盘或数据盘）进行备份，推荐使用云硬盘备份功能，在保证数据安全的同时降低备份成本。

云主机备份操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“存储 >云主机备份”。

4. 单击右上角“创建存储库”，绑定云主机及备份策略。
   - 如果云主机已绑定存储库，则根据界面提示，配置备份信息。
   - 如果云主机未绑定存储库，请先购买存储库，配置备份信息。
5. 单击“下一步”。系统会自动为云主机创建备份。

在“备份副本”页签，产生的备份的“备份状态”为“可用”时，表示备份任务执行成功。

为了保证数据完整性，建议备份完成后再执行重启。

详细操作请参见云主机备份介绍

云硬盘备份操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的  ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“存储 >云硬盘备份”。
4. 单击右上角“创建存储库”，绑定云硬盘及备份策略。
   - 如果云主机已绑定存储库，则根据界面提示，配置备份信息。
   - 如果云主机未绑定存储库，请先购买存储库，配置备份信息。
5. 单击“下一步”。系统会自动为云硬盘创建备份。

在“备份副本”页签，产生的备份的“备份状态”为“可用”时，表示备份任务执行成功。

如果在备份过程中对磁盘的数据进行删除等操作，被删除的文件可能不会被备份成功。为了保证数据完整性，建议备份完成后再对数据进行操作。

详细操作请参见云硬盘备份介绍

7.10 文件系统

7.10.1 一键挂载文件系统

新购云主机一键挂载 NAS 文件系统

前提条件

已创建文件系统。具体操作请参考：弹性文件-快速入门-创建文件系统

使用限制

- 镜像：仅支持使用公共镜像的云主机挂载 NAS 文件系统，暂不支持私有镜像。
- 挂载目录：仅支持挂载 NAS 文件系统的根目录，不支持挂载 NAS 文件系统的子目录。
- VPC：NAS 文件系统与云主机必须属于同一个 VPC；同时挂载多个 NAS 文件系统时，每个文件系统的挂载地址必须属于同一个 VPC。
- 子网：如果 NAS 文件系统选择了 ipv6 的挂载地址，云主机需要选择双栈类型（开启 ipv6）的子网。
- 挂载路径：同时挂载多个 NAS 文件系统时，每个文件系统的挂载路径不可重复，不可嵌套。
- 挂载 NAS 文件系统的云主机不支持更换 VPC。
- 挂载 NAS 文件系统的云主机不支持一键重装。

操作步骤

1. 登录控制中心。
2. 单击控制中心顶部的 ，选择“地域”。
3. 单击左侧导航栏“产品服务列表”，选择“计算 > 弹性云主机”。
4. 在弹性云主机列表页，单击“创建云主机”。
5. 在创建云主机的第 1 步【基础配置】，可选择文件系统进行挂载，重要配置说明如下：

- 文件系统：Linux 镜像仅支持 NFS 协议文件系统，Windows 镜像仅支持 CIFS 协议的文件系统。
- 挂载地址：用于 NAS 和云主机间的网络连通，需要和云主机属于同一 VPC。
- 挂载路径：从云主机上访问 NAS 的本地路径。
    - 如果镜像已选择 Linux
        - 单目录的长度不超过 255 个字符，总长度不能超过 4095 个字符；
        - 必须以"/"开头，由数字，字母，点，下划线，减号组成，通过"/"分割；
        - 不能是系统路径：/，/bin，/usr，/boot，/dev，/etc，/lib，/lib64，/proc，/run，/sys，/var，/tmp，/sbin；

- 每条配置填写的挂载路径不能重复；
- 推荐在/mnt 下新建本地路径作为挂载路径，如：/mnt/docs；

○ 如果镜像已选择 Windows

- 只能输入【E~Z】内的单个字母作为盘符；
- 推荐选择 Z 作为挂载盘符，如果 Z 被占用，请重新输入其他字母作为盘符；

- 协议类型：NFS 文件系统可以选择 NFSv3 协议或 NFSv4 协议，CIFS 文件系统选择 CIFS 协议即可。
- 挂载参数：根据所选协议类型自动填写，不可修改。
- 开机自动挂载：勾选后，每次云主机启动时都会自动挂载文件系统。

6. 成功创建云主机后，NAS 会自动完成挂载，假如 NAS 挂载失败，不影响云主机创建结果。

已购云主机一键挂载 NAS 文件系统

前提条件

已创建文件系统。具体操作请参考：弹性文件-快速入门-创建文件系统

使用

挂载文件系统，具体操作请参考：挂载文件系统

7.11 监控

7.11.1 监控弹性云主机

为保证弹性云主机的可靠性、可用性和可观测性，对云主机进行监控已经成为一种必要且重要的手段。天翼云控制平台提供的云主机监控功能，可方便用户更快、更直观的了解弹性云主机的运行情况、资源使用情况及其他性能指标，同时可根据实时监控情况，执行告警通知等操作，帮助客户更好的管理云主机

云主机监控分为基础监控、操作系统监控和进程监控。

1.基础监控：无需安装监控 Agent 就可获得的监控指标。指标采集粒度为 1 分钟，监控 CPU 使用率、内存使用率、网络 IO、磁盘 IO 等指标。

2.操作系统监控：通过在云主机中安装 Agent 插件，提供更细颗粒度的监控。除 CPU 与内存使用率等指标外，还可以支持 CPU 负载、磁盘和网卡等相关指标的监控服务。

3.进程监控：通过在云主机中安装 Agent 插件，对主机内活跃的进程进行监控，默认监控进程数量、进程的 CPU 使用率，以及打开的文件数等具体信息。

说明

GPU 实例分为图形加速基础型（G 系列）和计算加速型（P 系列）两类 GPU 云主机实例，若要对 P 系列云主机进行 GPU 监控，需要安装 Agent 的同时，对应的安装 GPU 的驱动。

7.11.2 监控 Agent 概览

监控 Agent 简介

通过在主机上安装云监控 Agent，为您提供主机的系统监控、进程监控等服务。

系统要求

目前只支持 Linux 操作系统和 Windows 操作系统。支持的系统如下：

操作系统	版本
Linux	天翼云提供的公共 Linux 镜像均支持
Windows	天翼云提供的公共 Windows 镜像均支持

在部分区域，对于使用标准公共版本镜像的服务器，可以在云主机详情页-监控单击“一键安装监控 Agent”来实现一键安装新的监控。



安装位置

监控 Agent 在云主机上的安装位置如下：

- Linux：/usr/bin/telegraf
- Windows：C:\Program Files\ctyuntelegraf\telegraf

端口说明

云主机监控 Agent 通过 TCP 协议访问远程服务器的 10063 和 10064 端口，用于心跳检测与监控数据上报，以及用于查看配置和更改日志级别等。

性能说明

资源	性能说明
CPU	当监控插件稳定运行时，一般 CPU 的消耗小于 1%
内存	当监控插件稳定运行时，一般内存的占用小于 64M
插件安装包	64M

支持安装监控 Agent 的地区及 Agent 安装包下载路径

地区名称	Agent 安装包下载路径
华北 2	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
西南 2	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
青岛 20	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
上海 15	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1

地区名称	Agent 安装包下载路径
南昌 5	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
华南 2	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
广州 6	Linux：http://100.127.108.130/cttg/linux/update_agent.sh Windows： http://100.127.108.130/cttg/windows/update_agent.ps1
西南 1	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
芜湖 4	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
贵阳 42	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
沈阳 8	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
杭州 7	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
北京 9	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
西安 7	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1

地区名称	Agent 安装包下载路径
西安 5	Linux：http://100.127.204.130/cttg/linux/update_agent.sh Windows： http://100.127.204.130/cttg/windows/update_agent.ps1
太原 4	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
郑州 5	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
常州 69	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
上海 32	Linux：http://100.126.15.1/cttg/linux/update_agent.sh Windows：http://100.126.15.1/cttg/windows/update_agent.ps1
武汉 41	Linux：http://100.126.26.1/cttg/linux/update_agent.sh Windows：http://100.126.26.1/cttg/windows/update_agent.ps1
华东 1	Linux：http://100.123.136.65/cttg/linux/update_agent.sh Windows： http://100.123.136.65/cttg/windows/update_agent.ps1
南宁 23	Linux：http://100.123.89.1/cttg/linux/update_agent.sh Windows：http://100.123.89.1/cttg/windows/update_agent.ps1
上海 36	Linux：http://100.123.113.1/cttg/linux/update_agent.sh Windows:http://100.123.113.1/cttg/windows/update_agent.ps1
上海 7	Linux：http://100.127.20.130/cttg/linux/update_agent.sh

地区名称	Agent 安装包下载路径
	Windows： http://100.127.20.130/cttg/windows/update_agent.ps1
内蒙 6	Linux：http://100.127.8.130/cttg/linux/update_agent.sh Windows：http://100.127.8.130/cttg/windows/update_agent.ps1
北京 5	Linux：http://100.127.244.130/cttg/linux/update_agent.sh Windows： http://100.127.244.130/cttg/windows/update_agent.ps1
九江	Linux：http://100.127.124.130/cttg/linux/update_agent.sh Windows： http://100.127.124.130/cttg/windows/update_agent.ps1
拉萨 3	Linux：http://100.127.36.130/cttg/linux/update_agent.sh Windows： http://100.127.36.130/cttg/windows/update_agent.ps1
杭州 2	Linux：http://100.127.16.130/cttg/linux/update_agent.sh Windows： http://100.127.16.130/cttg/windows/update_agent.ps1
长沙 42	Linux：http://100.86.2.1/cttg/linux/update_agent.sh Windows：http://100.86.2.1/cttg/windows/update_agent.ps1
中卫 5	Linux：http://100.123.60.9/cttg/linux/update_agent.sh Windows：http://100.123.60.9/cttg/windows/update_agent.ps1
石家庄	Linux：http://100.123.26.171/cttg/linux/update_agent.sh

地区名称	Agent 安装包下载路径
20	Windows： http://100.123.26.171/cttg/windows/update_agent.ps1
南京 5	Linux：http://100.123.0.1/cttg/linux/update_agent.sh Windows：http://100.123.0.1/cttg/windows/update_agent.ps1
晋中	Linux：http://100.127.252.50/cttg/linux/update_agent.sh Windows： http://100.127.252.50/cttg/windows/update_agent.ps1
贵州 3	Linux：http://100.127.12.130/cttg/linux/update_agent.sh Windows： http://100.127.12.130/cttg/windows/update_agent.ps1
辽阳 1	Linux：http://100.123.84.171/cttg/linux/update_agent.sh Windows： http://100.123.84.171/cttg/windows/update_agent.ps1
福州 25	Linux：http://100.123.20.200/cttg/linux/update_agent.sh Windows： http://100.123.20.200/cttg/windows/update_agent.ps1
合肥 2	Linux：http://100.127.224.130/cttg/linux/update_agent.sh Windows： http://100.127.224.130/cttg/windows/update_agent.ps1

注意

除以上地区外，若需要下载监控 Agent 安装包，请参考云监控-Agent 管理。

监控 Agent 下载

版本	监控 Agent 安装包
v1.3.1	telegraf-monitor.tar.gz

7.11.3 监控 Agent 版本说明

版本-1.3.1

- 发布时间：2025-06
- 更新说明：修复已知问题。

版本-1.2.0

- 发布时间：2024-11
- 更新说明：新增云硬盘监控指标，包含：swap 内存利用率、平均 I/O 服务时长、磁盘 I/O 使用率、平均队列长度，详细监控指标请查看查看云硬盘监控数据。
- 版本-1.0.0 时间：2024-07
- 更新说明：支持操作系统监控，包含：CPU 相关监控指标、内存相关监控指标、CPU 负载相关监控指标、磁盘相关监控指标、磁盘 I/O 相关监控指标、文件系统相关监控项、网络相关监控项、GPU 相关监控指标，详细监控指标请查看云主机支持的操作系统监控。

- 发布

7.11.4 安装监控 Agent

为保证您创建出来的云主机监控数据、功能正常，天翼云推荐您通过控制台上的相关功能自动为云主机安装云监控插件，同时提供手动安装云监控插件功能，可根据您的需求场景选择不同的方式。

约束与限制：

确保需要安装监控 Agent 所在得资源池支持监控 Agent 能力，可参考监控概览-支持安装监控 Agent 的地区及 Agent 安装包下载路径查询。

确保安装时，云主机状态为运行中。

确保云主机可以正常访问 curl 169.254.169.254:10063 及 169.254.169.254:10064。

安装成功后的监控进程不会被其他软件关闭。

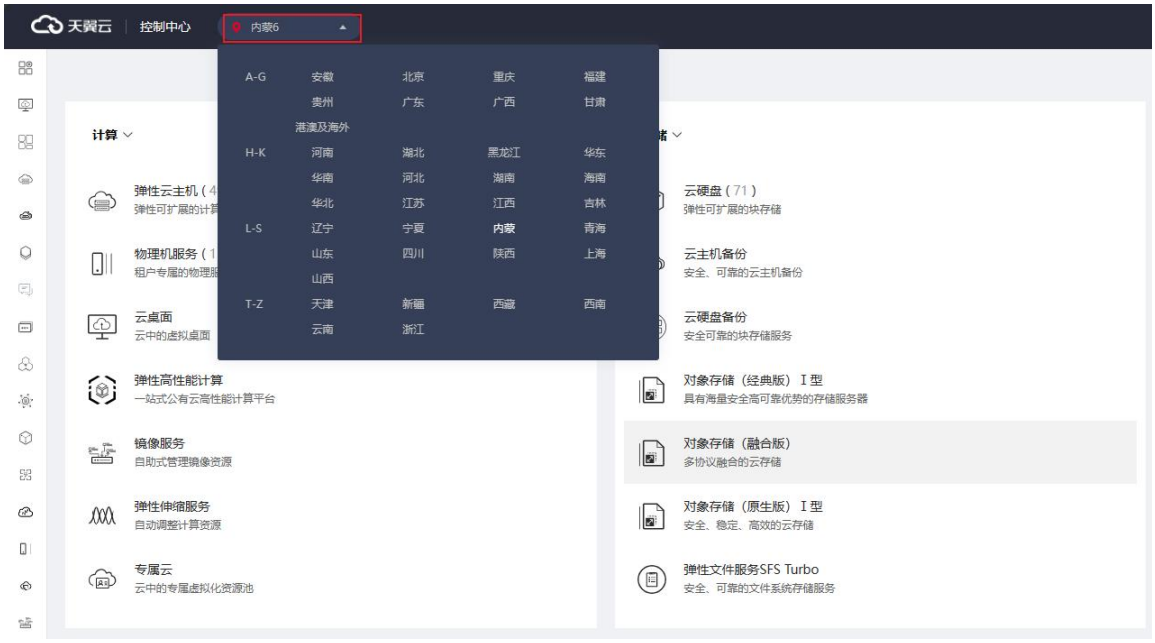
说明

若云主机无法正常访问，请检查防火墙、安全组等服务是否屏蔽监控服务器地址。

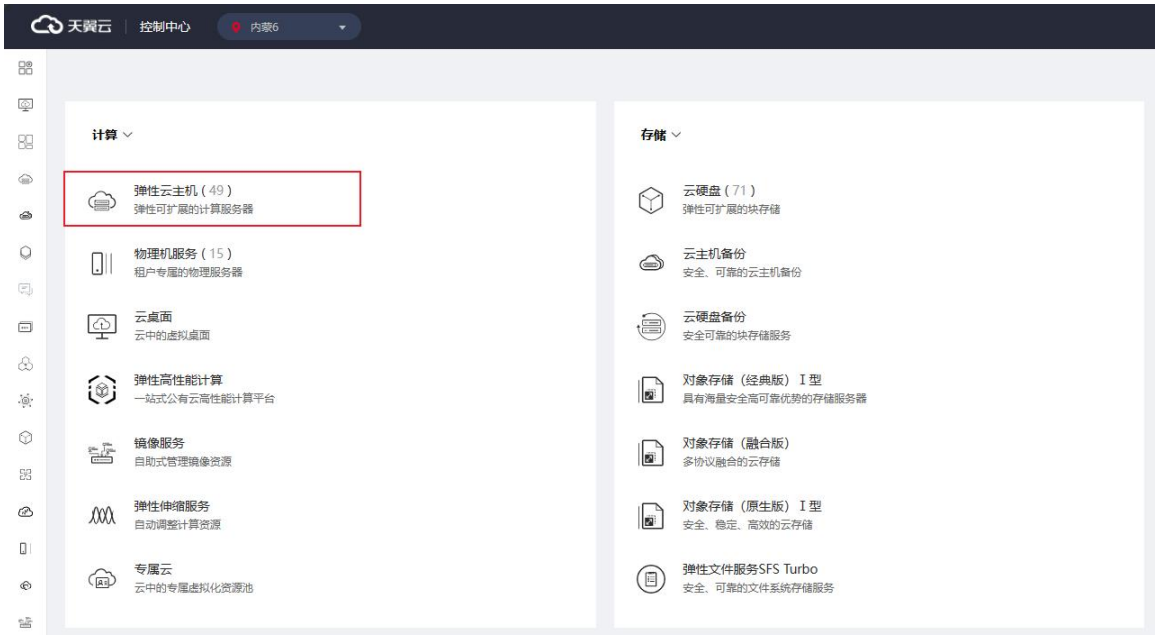
创建云主机时自动安装监控插件

操作步骤

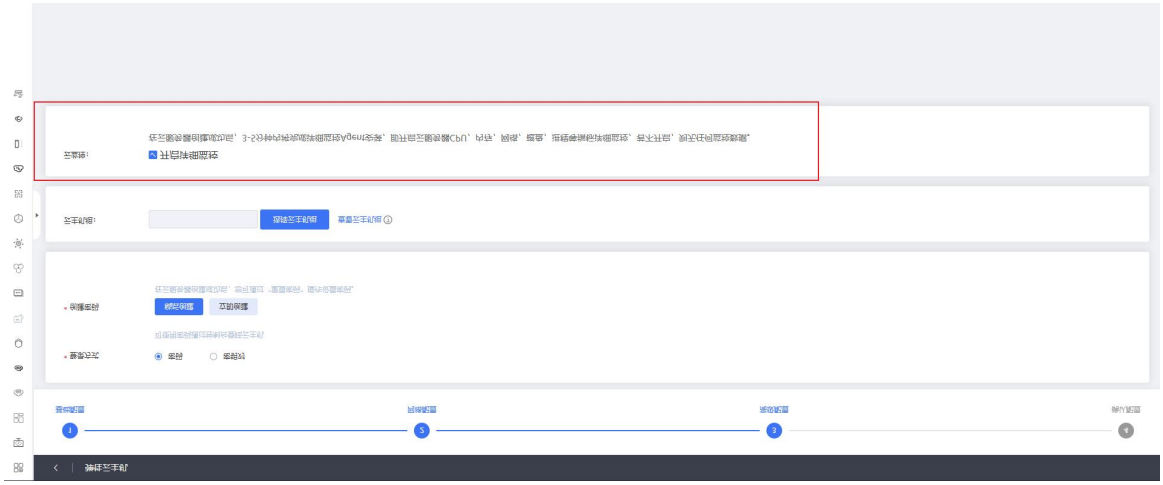
1. 登录控制中心。
2. 单击左上角选择即将创建云主机所在的地域。



3. 单击“计算>弹性云主机”，进入云主机控制台。



4. 单击“创建云主机”，在高级配置页签，选中“开启详细监控”选项。按此配置创建出的云主机，将被自动安装监控插件。



注意

若通过快照、备份、私有镜像、克隆创建云主机时，则：

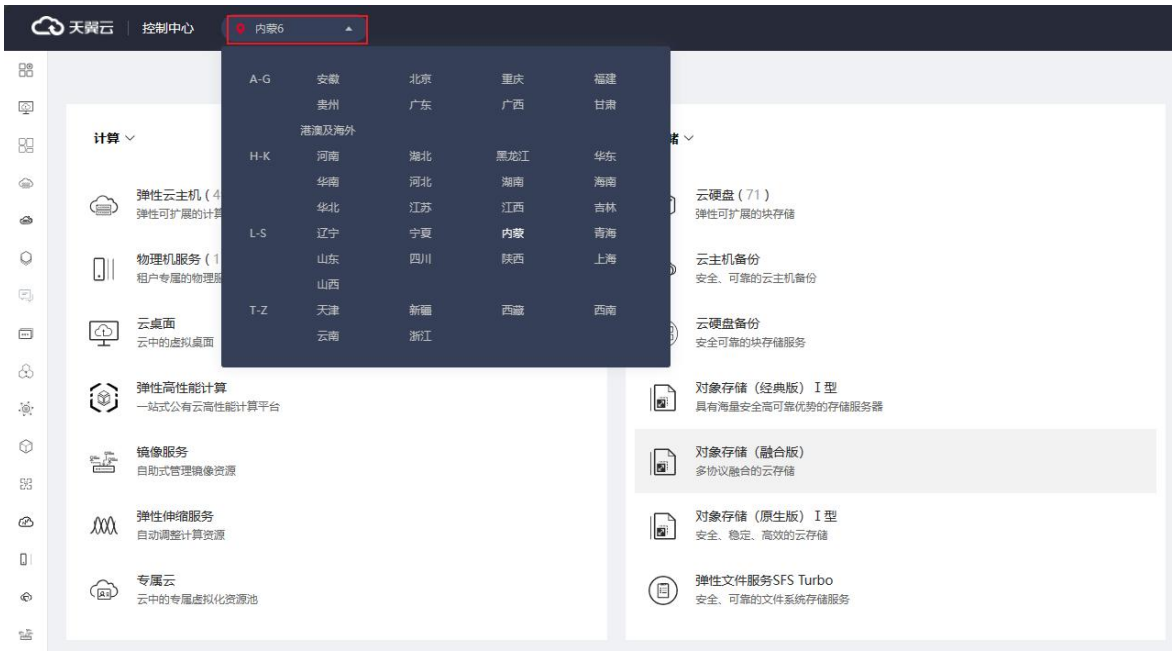
选择开启详细监控：新创建的云主机会安装最新版本的云主机。

选择不开启详细监控：新创建的云主机与原快照、备份、私有镜像、被克隆云主机监控 Agent 保持一致。即，原场景已安装 Agent，新创建云主机保持安装，安装版本与原场景一致；原场景未安装 Agent，则新创建云主机也不安装。

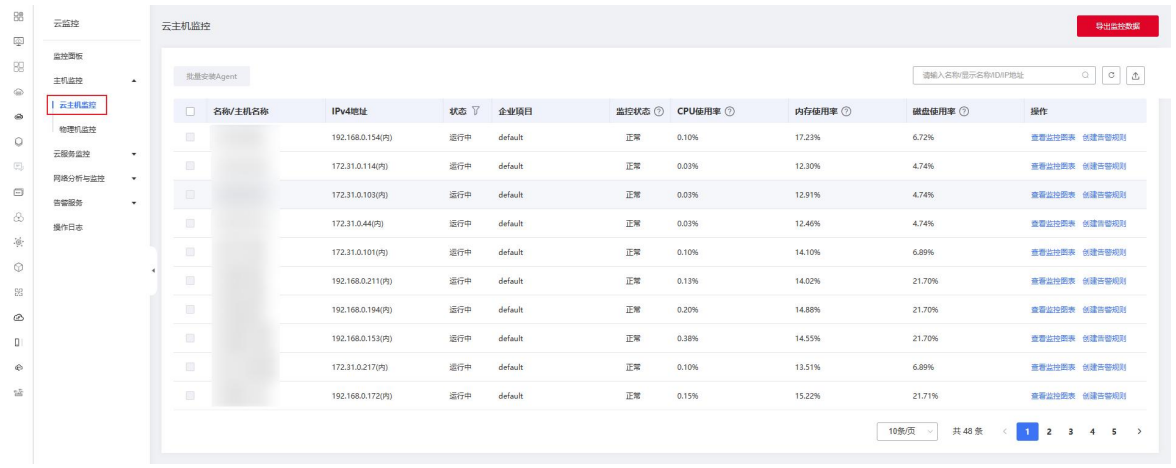
为已创建的云主机安装、升级监控插件

操作步骤

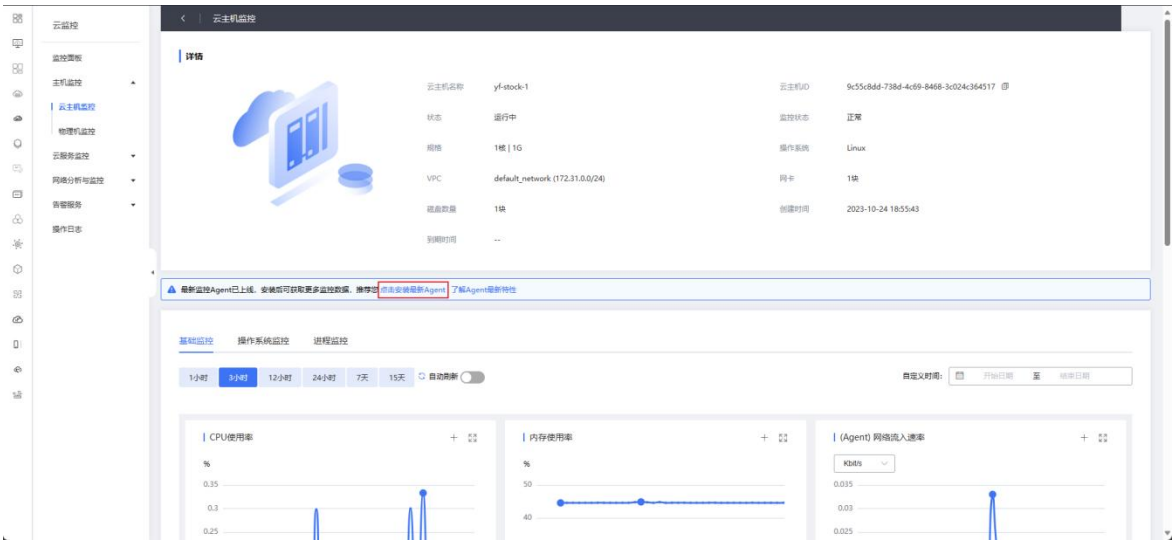
1. 单击左上角选择需要安装或升级的云主机所在的地域。



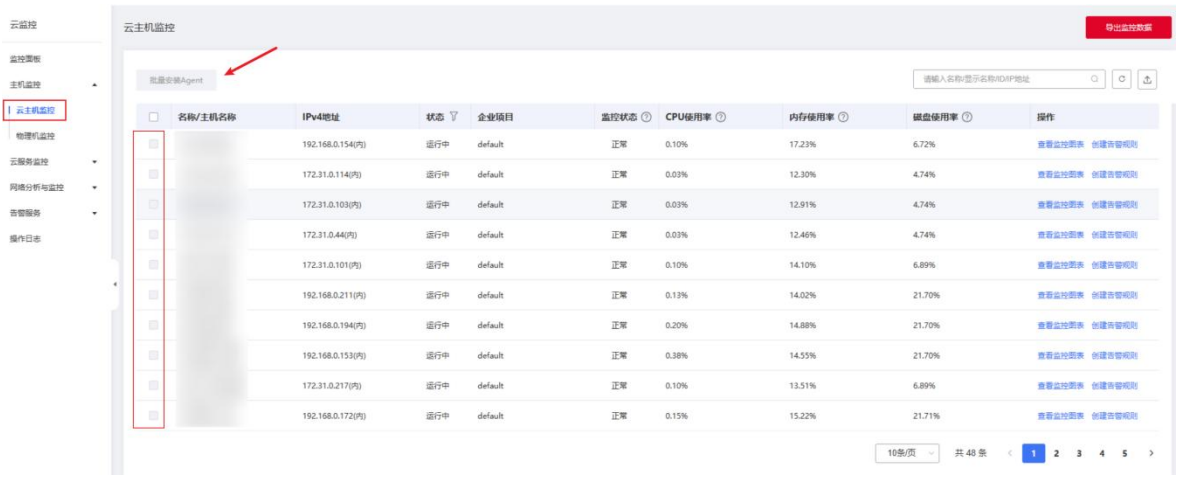
2. 单击“管理与部署>云监控”，进入“主机监控>云主机监控”菜单。



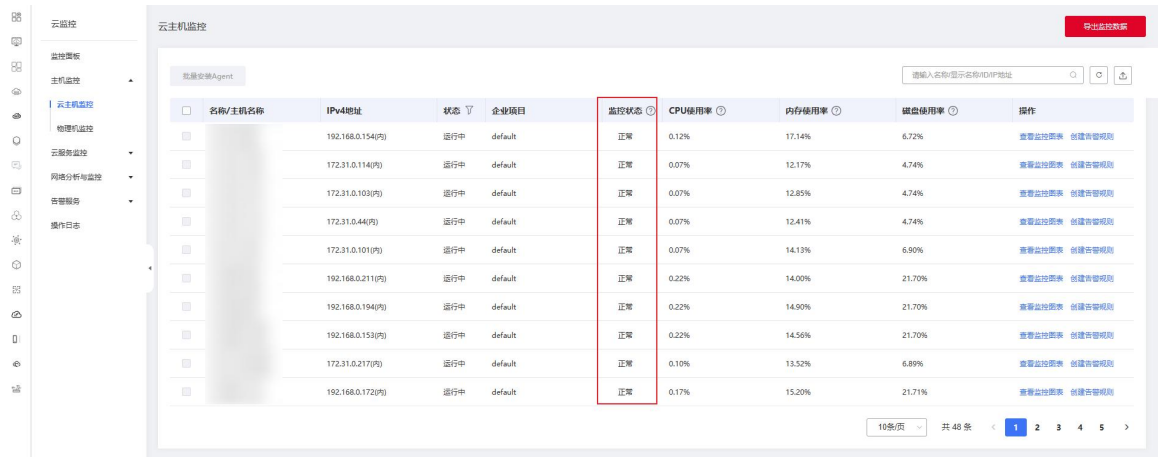
3. 选中需要安装或升级的云主机，点击“安装/更新 Agent”。



也可选中多台需要升级的云主机，进行批量升级云主机监控 Agent 操作，单次最大允许批量操作 50 台云主机。



4. 当云主机监控列表监控状态为“正常”时，即代表监控插件安装成功。



手动为云主机安装、升级监控插件

前提条件：

请确保云主机可正常访问 objectStorageIP 以获取监控 Agent 安装包。

Linux 系统操作步骤：

1. 以 root 用户登录待安装插件的云主机。

2. 执行以下监控插件安装命令：

```
curl -o update_agent.sh
http://{obejctStorageIP}:{port}/cttg/linux/update_agent.sh

chmod +x update_agent.sh

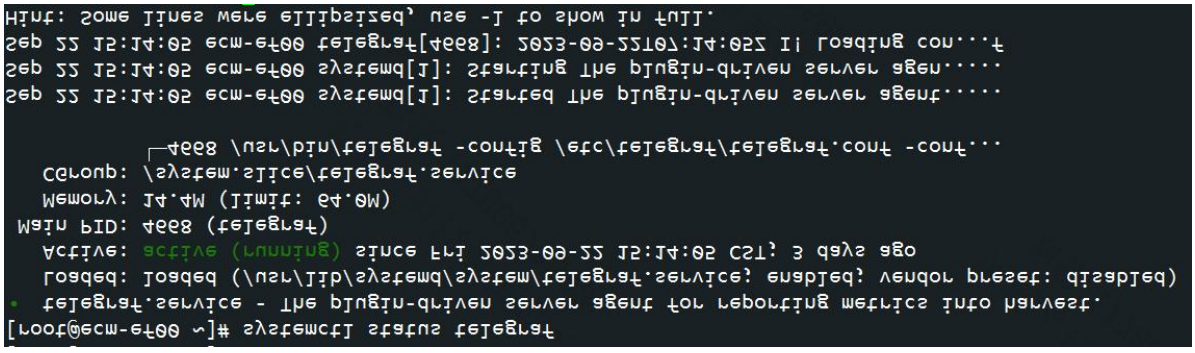
./update_agent.sh http://{obejctStorageIP}:{port} 1
```

注意

每个资源池 objectStorageIP 存在差异，可参考监控概览-支持安装监控 Agent 的地区及 Agent 安装包地址查询。

3. 执行以下命令查询监控安装状态：

```
systemctl status telegraf
```



Windows 系统操作步骤：

1.以 Administrator 用户登录待安装云监控插件的主机。

2.执行以下命令，安装云监控插件。

- 若 Windows 系统支持-curl 命令，在 Powershell 中分别执行以下命令：

```
curl -Uri
http://{objectStorageIP}:{port}/cttg/windows/update_agent.ps1
-OutFile update_agent.ps1

./update_agent.ps1 http://{objectStorageIP}:{port} 2
```

- 若 Windows 系统不支持-curl 命令，在 Powershell 中执行以下命令：

```
& "C:\Program Files\Cloudbase
Solutions\Cloudbase-Init\Python\python.exe"-c "import urllib;from
urllib import
request;url='http://{objectStorageIP}:{port}/cttg/windows/curl.exe';u
rllib.request.urlretrieve(url,'C:/Windows/System32/curl.exe')"



& "C:\Windows\System32\curl.exe"
http://{objectStorageIP}:{port}/cttg/windows/curl_update_agent.ps1 -o
curl_update_agent.ps1

./curl_update_agent.ps1 http://{objectStorageIP}:{port} 2
```

注意

每个资源池 objectStorageIP 存在差异，可参考监控概览-支持安装监控 Agent 的地区及 Agent 安装包地址查询。

4.在 Powershell 中执行以下命令查询监控安装状态：

```
Get-Service ctyuntelegraf
```



7.11.5 云主机支持的操作系统监控指标（安装监控 Agent）

前提条件

请确保您的云主机上已成功安装监控 Agent。具体操作，请参见安装监控 Agent。

监控指标说明

安装监控 Agent 后，您可以查看不同分类的操作系统监控指标，采集频率均为 1 分钟/次。

操作系统监控指标已支持：CPU 相关监控指标、内存相关监控指标、CPU 负载相关监控指标、磁盘相关监控指标、磁盘 I/O 相关监控指标、文件系统相关监控项、网络相关监控项、GPU 相关监控指标。

CPU 相关监控指标

监控指标	监控指标名称	监控指标含义	单位	维度	监控周期

监控指标	监控指标名称	监控指标含义	单位	维度	监控周期
CPU 使用率	cpu_util	云主机 CPU 使用率。	%	云主机	1 分钟
用户空间 CPU 使用率	cpu_user_time	云主机用户空间占用 CPU 使用率。	%	云主机	1 分钟
内核空间 CPU 使用率	cpu_system_time	云主机内核空间占用 CPU 使用率。	%	云主机	1 分钟
其他 CPU 使用率	other_cpu_util	云主机其他占用 CPU 使用率。	%	云主机	1 分钟
CPU 空闲时间占比	cpu_idle_time	云主机 CPU 空闲时间占比。	%	云主机	1 分钟
Nice 进程 CPU 使用率	cpu_nice_time	云主机 Nice 进程 CPU 使用率。	%	云主机	1 分钟

监控指标	监控指标名称	监控指标含义	单位	维度	监控周期
iowait 状态占比	cpu_iowait_time	云主机 iowait 状态占用 CPU 的比率。	%	云主机	1 分钟
CPU 中断时间占比	cpu_interrupt_time	云主机 CPU 处理中断用时占用 CPU 时间的比率	%	云主机	1 分钟
CPU 软中断时间占比	cpu_softirq_time	云主机 CPU 处理软中断时间占用 CPU 时间的比率。	%	云主机	1 分钟

内存相关监控指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
内存使用率	mem_util	云主机内存使用率。	%	云主机	1 分钟
free_memory	可用内存	云主机可用内存。	byte	云主机	1 分钟

监控指标	监控指标名称	监控指标含义	单位(原始数据)	维度	监控周期
used_memory	已用内存量	云主机已使用内存量。	byte	云主机	1 分钟
buffer_memory	Buffers 占用量	云主机 Buffers 内存量。	byte	云主机	1 分钟
cache_memory	Cached 占用量	云主机 Cache 内存量。	byte	云主机	1 分钟

CPU 负载相关指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
1 分钟平均负载	mem_util	云主机过去 1 分钟的系统平均负载。	无	云主机	1 分钟
5 分钟平均负载	可用内存	云主机过去 5 分钟的系统平均负载。	无	云主机	1 分钟
15 分钟平均负载	已用内存量	云主机过去 15 分钟的系统平均负载。	无	云主机	1 分钟

磁盘相关指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
磁盘分配率	disk_util_inband	磁盘使用率。	%	云主机-挂载点	1 分钟
磁盘已用存储量	used_disk_space_on	磁盘的已用存储空间。	byte	云主机-挂载点	1 分钟
磁盘剩余存储量	free_disk_space_on	磁盘的剩余存储空间。	byte	云主机-挂载点	1 分钟
磁盘存储总量	total_disk_space_on	磁盘存储总量。	byte	云主机-挂载点	1 分钟

磁盘 I/O 相关指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
磁盘读速率	disk_read_bytes_rate	从磁盘每秒读取的数据量。	KB/s	云主机-磁盘	1 分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
磁盘读请求速率	disk_read_requests_rate	从磁盘每秒读请求次数。	次/s	云主机-磁盘	1 分钟
磁盘写速率	disk_write_bytes_rate	向磁盘每秒写入的数据量。	KB/s	云主机-磁盘	1 分钟
磁盘写请求速率	disk_write_requests_rate	向磁盘每秒写请求次数。	次/s	云主机-磁盘	1 分钟

文件系统操作指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
inode 空间大小	total_inode_on	inode 空间。	files	云主机-挂载点	1 分钟
inode 已使用空	used_inode_on	已使用的 inode 空	files	云主机-挂	1 分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
间		间。		载点	
inode 已使用占比	pused_inode_on_percentage	已使用的 inode 占比。	%	云主机-挂载点	1 分钟

网络相关监控指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
网络流入流量的速率	network_incoming_bytes_rate_inband	云主机内网每秒接收的数据量。	bit/s	云主机	1 分钟
网络流出流量的速率	network_outing_bytes_rat	云主机内网每秒发送的数据量。	bit/s	云主机	1 分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
	e_inband				
网卡包接收的速率	network_incoming_packets_rate_inband	云主机内网每秒接收的数据包数量。	个/s	云主机	1分钟
网卡包发送的速率	network_outing_packets_rate_inband	云主机内网每秒发送的数据包数量。	个/s	云主机	1分钟
接收误包率	network_incoming_errs_rate_inband	云主机内网每秒接收的错误数据包数量占所接收的数据包的比率。	%	云主机	1分钟
发送误包率	network_outing_e	云主机内网每秒发送的错误数据包数量占所接收的数据包的比率。	%	云主	1分

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
	rrs_rate_inband			机	钟
接收丢包率	network_incoming_drop_rate_inband	云主机内网每秒接收并已丢弃的数据包数量占所接收的数据包的比率。	%	云主机	1分钟
发送丢包率	network_outing_drop_rate_inband	云主机内网每秒发送并已丢弃的数据包数量占所接收的数据包的比率。	%	云主机	1分钟
TCP 连接数	net_tcp_total	云主机内所有状态的 TCP 连接总数。	个	云主机	1分钟
ESTABLISHED TCP 连接数	net_tcp_established	云主机内 ESTABLISHED 状态的 TCP 连接数量。	个	云主机	1分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
SYS_SENT TCP 连接数	net_tcp_syn_sent	云主机内 SYS_SENT 状态的 TCP 连接数量。	个	云主机	1 分钟
SYS_RECV TCP 连接数	net_tcp_syn_recv	云主机内 SYS_RECV 状态的 TCP 连接数量。	个	云主机	1 分钟
FIN_WAIT 1TCP 连接数	net_tcp_fin_wait1	云主机内 FIN_WAIT1 状态的 TCP 连接数量。	个	云主机	1 分钟
FIN_WAIT 2TCP 连接数	net_tcp_fin_wait2	云主机内 FIN_WAIT2 状态的 TCP 连接数量。	个	云主机	1 分钟
TIME_WAIT TCP 连接数	net_tcp_time_wait	云主机内 TIME_WAIT 状态的 TCP 连接数量	个	云主机	1 分钟
CLOSE TCP 连接数	net_tcp_close	云主机内 CLOSE 状态的 TCP 连接数量。	个	云主机	1 分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
CLOSE_WAITTCP 连接数	net_tcp_close_wait	云主机内 CLOSE_WAIT 状态的 TCP 连接数量。	个	云主机	1 分钟
LAST_ACK TCP 连接数	net_tcp_last_ack	云主机内 LAST_ACK 状态的 TCP 连接数量。	个	云主机	1 分钟
LISTEN TCP 连接数	net_tcp_listen	云主机内 LISTEN 状态的 TCP 连接数量。	个	云主机	1 分钟
CLOSING TCP 连接数	net_tcp_closing	云主机内 CLOSING 状态的 TCP 连接数量。	个	云主机	1 分钟

进程相关监控指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
进程 CPU 使用率	process_cpu_used	云主机内进程消耗的 CPU 百分比。	%	云主机	1 分钟
进程内存使用率	process_memory_used	云主机内进程消耗的内存百分比。	%	云主机	1 分钟
打开文件数	open_files_num	云主机内进程打开的文件数。	个	云主机	1 分钟
运行中进程数	proc_num	云主机内运行的进程数量。	个	云主机	1 分钟
僵死进程数	zomb_num	云主机内僵死的进程数量。	个	云主机	1 分钟
睡眠进程数	sleep_num	云主机内睡眠的进程数量。	个	云主机	1 分钟
指定进程数量	process_number	云主机内指定进程数量。	个	云主机-进程	1 分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
指定进程CPU使用率	process_cpu_util	指定进程消耗的CPU百分比。	%	云主机-进程	1分钟
指定进程内存使用率	process_mem_util	指定进程消耗的内存百分比。	%	云主机-进程	1分钟
指定进程打开文件数	process_open_files	指定进程打开的文件数。	个	云主机-进程	1分钟

注意

以上进程相关指标Windows操作系统尚未支持。

GPU相关监控指标

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
显存使用率	memusage_on_gpu	GPU云主机显存使用率。	%	云主机-GPU	1分钟

监控指标	监控指标名称	监控指标含义	单位（原始数据）	维度	监控周期
GPU 使用率	gpuutil_on_gpu	GPU 使用率。	%	云主机-GPU	1 分钟
显卡功率	npu_chip_info_power	GPU 功率。	W	云主机-GPU	1 分钟
显卡温度	npu_chip_info_temperature	GPU 卡温度。	℃	云主机-GPU	1 分钟
显存使用量	npu_chip_info_used_memory	GPU 显存使用量。	byte	云主机-GPU	1 分钟

7.11.6 设置告警规则

操作场景

支持根据需要灵活创建告警规则，既可以使用系统默认告警模板，也可以对具体监控指标进行自定义告警规则的设置。

当资源的监控指标达到告警条件，云监控将向您发送告警消息，报告异常监控数据，帮助您及时掌握异常状态并处理，保证业务顺畅进行。

前提条件

注册天翼云账号，并完成实名认证。具体操作，请参见天翼云账号注册流程。

操作步骤

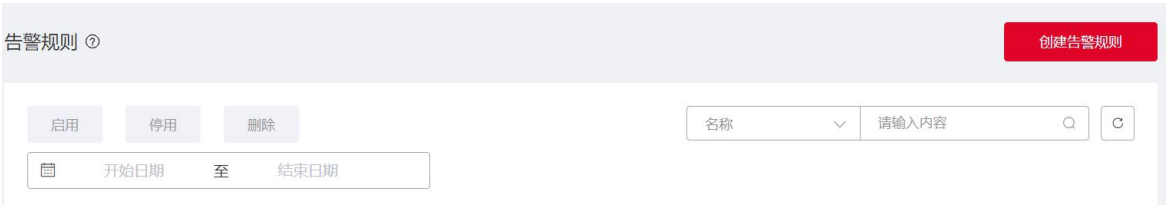
1. 登录控制中心。

2. 在控制中心页面左上角点击 ，选择区域，本文我们选择华东 1。

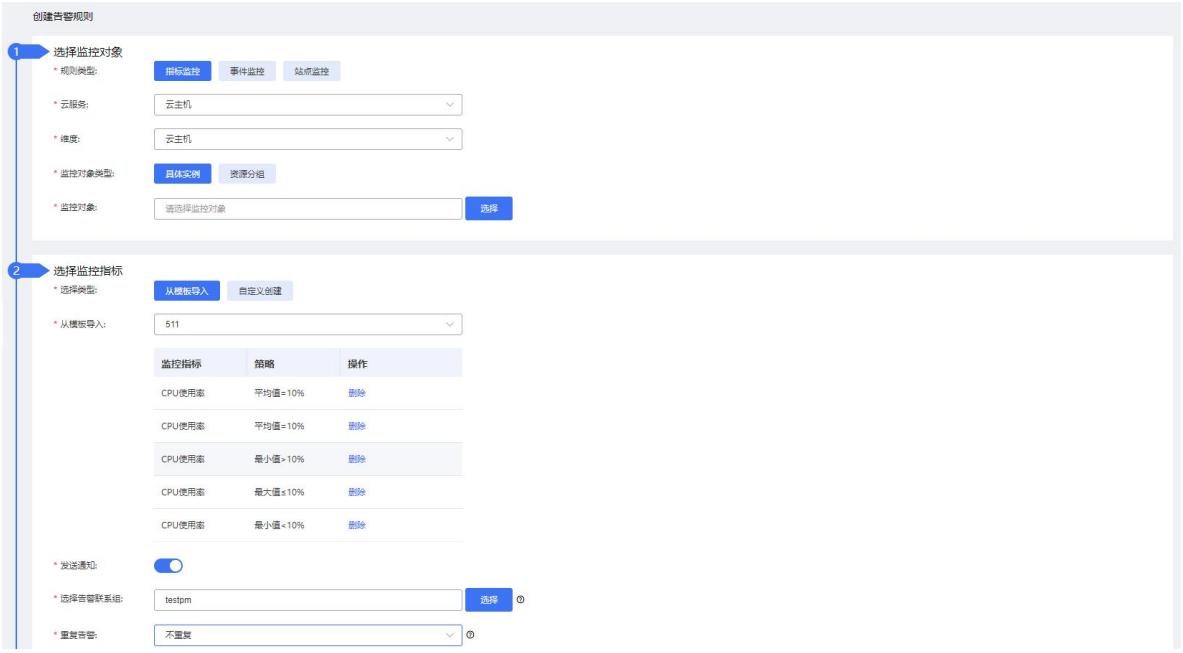
3. 依次选择“管理与部署”，单击“云监控”，进入监控概览页面。

4. 单击“告警服务”下拉菜单，单击“告警规则”，进入告警规则详情页面。

5. 在“告警规则”界面，单击“创建告警规则”按钮，进入告警规则创建页面。



6. 在“创建告警规则”页面，根据界面提示配置参数。



配置参数如下：

参数	参数说明

参数	参数说明
规则类型	选择规则的类型，主要包括指标监控、事件监控、站点监控三种。
云服务	配置告警规则监控的云服务资源类型。
维度	用于指定告警规则对应指标的维度名称。
监控对象	用来配置该告警规则针对的具体资源，可以是一个或多个。
选择类型	可以选择从模板导入或自定义创建。
模板	选择需要导入的模板。
发送通知	配置是否发送邮件通知用户，可以选择“是”（推荐选择）或者“否”。
选择告警联系组	配置发生告警通知的用户组。
重复告警	告警发生后，如果监控项未恢复正常，之后间隔多久再次发送告警。

参数	参数说明
触发场景	触发告警邮件的场景，可在告警及恢复时发送提醒信息。
通知周期	配置告警通知的周期时间。
通知时段	配置告警通知的时间段。
通知方式	配置告警通知的通知方式，支持邮箱及短信。
名称	该告警规则的自定义名称。
企业项目	选择告警规则适用的企业项目。
描述	添加对该告警规则描述（此参数非必填项）。

7.11.7 查看监控指标

操作场景

云监控会自动获取您当前账号下云产品的资源并关联相应监测指标，帮助您实时了解云产品的性能指标，及时掌握各项资源的运行状态。

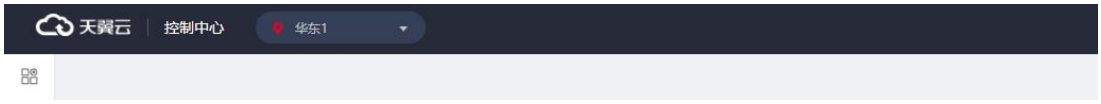
前提条件

- 注册天翼云账号，并完成实名认证。具体操作，请参见天翼云账号注册流程。
- 您已经完成云产品的创建。

操作步骤

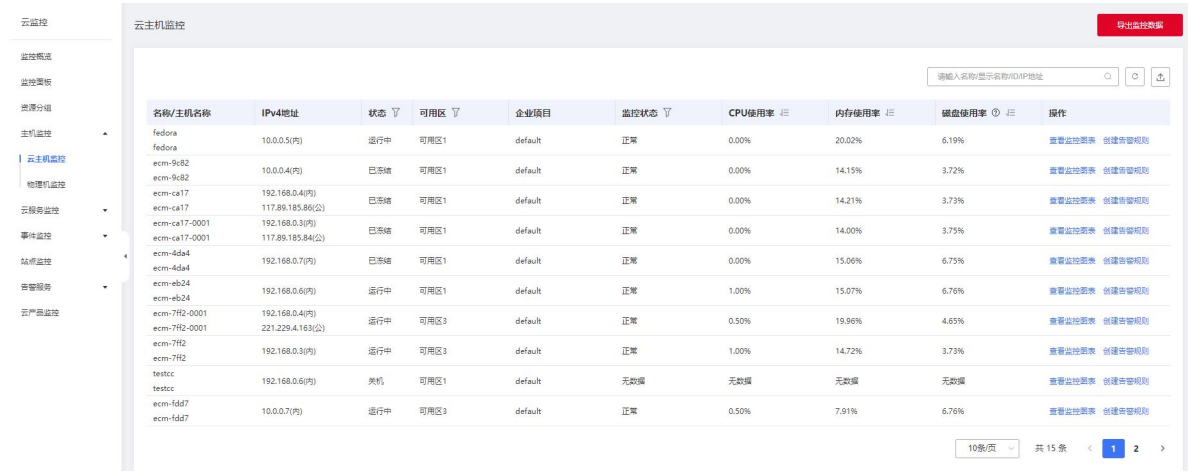
1. 登录控制中心。

2. 在控制中心页面左上角点击 ，选择区域，本文我们选择华东 1。

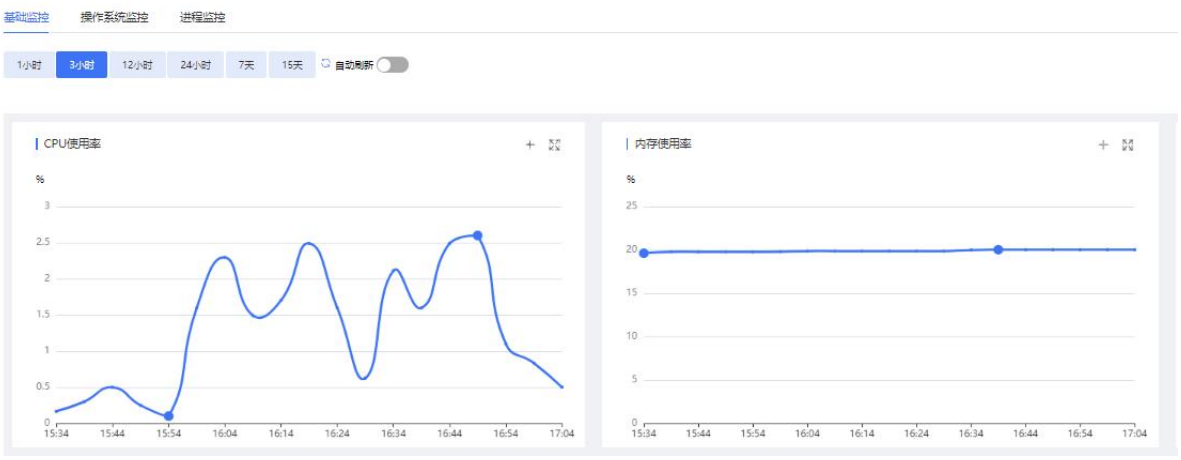


3. 依次选择“管理与部署”，单击“云监控”，进入监控概览页面。

4. 单击“主机监控”“云主机监控”选项，进入对应云产品的监控页面。



5. 单击待查看的云主机资源所在行的“查看监控图表”，可查看该云产品中指定资源的监控图表。



6. 云主机支持近两天实时监控数据的导出。在云产品监控页面的右上角，单击“导出监控数据”按钮，添加需要导出的监控项，单击“确定”按钮，可导出数据。



7.11.8 弹性云主机支持的基础监控指标

指标名称	中文介绍	单位
cpu_user_time	用户空间 CPU 使用率	%
cpu_system_time	内核空间 CPU 使用率	%
other_cpu_util	其他 CPU 使用率	%

指标名称	中文介绍	单位
cpu_util	CPU 使用率	%
cpu_idle_time	CPU 空闲时间占比	%
cpu_nice_time	Nice 进程 CPU 使用率	%
cpu_iowait_time	iowait 状态占比	%
cpu_interrupt_time	CPU 中断时间占比	%
cpu_softirq_time	CPU 软中断时间占比	%
processor_load_1_min_average_per_core	cpu 的 1 分钟平均负载	%
processor_load_5_min_average_per_core	cpu 的 5 分钟平均负载	%
processor_load_15_min_average_per_core	cpu 的 15 分钟平均负载	%

指标名称	中文介绍	单位
mem_util	内存使用率	%
pused_memory	内存使用率(细粒度)	%
free_memory	可用内存	GB
used_memory	已用内存量	GB
buffer_memory	Buffers 占用量	GB
cache_memory	Cached 占用量	GB
disk_read_bytes_rate	磁盘读速率	byte/s
disk_read_requests_rate	磁盘读操作速率	次/秒
disk_write_bytes_rate	磁盘写速率	byte/s
disk_write_requests_rate	磁盘写操作速率	次/秒

指标名称	中文介绍	单位
network_incoming_bytes_rate_inband	网络流入速率	kb/s
network_outing_bytes_rate_inband	网络流出速率	kb/s
network_incoming_packets_rate_inband	网卡包接收的正确率	%
network_outing_packets_rate_inband	网卡包发送的正确率	%
network_incoming_errs_rate_inband	接收误包率	%
network_outing_errs_rate_inband	发送误包率	%
network_incoming_drop_rate_inband	接收丢包率	%
network_outing_drop_rate_inband	发送丢包率	%
process_cpu_used	进程 CPU 使用率	%
open_files_num	打开文件数	个

指标名称	中文介绍	单位
proc_num	运行中进程数	个
zomb_num	僵死进程数	个
sleep_num	睡眠进程数	个
disk_util_inband	磁盘分配率	%
perf_on_gpu	gpu 性能状态	Dec-00
memusage_on_gpu	GPU 显存用量	%
gpuutil_on_gpu	GPU 使用率	%
process_memory_used	进程内存使用率	%

7.12 安全

7.12.1 提升云主机安全的方法

账号安全

天翼云账号注册时要求您设置复杂度极高的密码。您应该妥善保管天翼云的账号

密码并定期对密码进行更换。

同时天翼云还支持账号二次认证，您可以在个人中心-安全设置中开启安全验证。开启后每次登录均需要短信验证码进行登录。

主机密码安全

主机密码是弹性云主机访问时最常用的安全验证。您在创建时输入的主机密码应该确保复杂度满足要求，并且妥善保管密码。

主机密码的复杂度应满足如下要求：

- 不少于 8 个字符。
- 必须同时包含大写字母、小写字母、数字、符号中的三项。
- 符号支持：

  [()`~!@#$%^& _-+=|{}[]:;'<>,.?/](mailto:()%60~!@#$%^&*

  _-+=|{}[]:;'<>,.?/)。
- 密码不能以斜线号（/）开头。

密钥对

密钥对，也称为 SSH 密钥对，是一种用于远程登录云主机的身份验证方式，与传统的用户名和密码登录方式有所区别。通过使用密钥对，您可以提升云主机的安全性，并避免因密码被拦截或破解导致的密码泄露问题。

密钥对由公钥和私钥组成，如果用户将公钥配置在 Linux 云主机中，便可以使用私钥登录 Linux 云主机，无需输入密码。通过使用密钥对登录 Linux 云主机，可以防止密码被拦截或破解导致帐户密码泄露，从而提高 Linux 云主机的安全性。

更多内容请参见密码和密钥对。

监控云主机

您可以使用天翼云云监控产品来监控您的弹性云主机，云监控支持自动实时监控、告警配置和告警通知，让您更好地掌握弹性云主机的运行状态和各项性能指标。

指标监控：对弹性云主机的各项运行指标进行实时监控。还可以查询历史的监控指标情况。

监控告警：通过配置告警规则，在指标发生异常的第一时间对您进行提醒。及时发现问题并处理，可以有效保障部署在的弹性云主机上的服务持续可用。

更多内容请参见云监控。

备份云主机

您可以通过天翼云云主机备份产品和云硬盘备份产品，对弹性云主机上的数据进行备份，当云主机或主机上的云硬盘出现故障，或者软件造成的数据丢失损坏以及人为错误导致的数据误删时，您可以借助备份功能自助快速恢复数据。

云主机备份

云主机备份（CT-CSBS，Cloud Server Backup Service）提供对弹性云主机的备份保护服务，支持基于云硬盘快照技术的备份服务，并支持利用备份数据恢复弹性云主机数据。通过云主机备份服务，可以在发生数据丢失、系统故障、人为错误或恶意攻击等情况下，还原云主机数据，确保业务的连续性和数据的安全性。

更多内容请参见云主机备份。

云硬盘备份

云硬盘备份（CT-VBS，Volume Backup Service）是针对云主机的系统盘、数据盘提供的备份服务。用户可对存储重要数据的云硬盘进行备份，并在云硬盘故障、用户误删数据、遭到黑客攻击等情况下，使用备份快速恢复数据，最大限度保证用户数据的安全性。

更多内容请参见云硬盘备份。

云硬盘加密概述

使用 KMS 加密数据密钥管理服务（Key Management Service，KMS）是一站式密钥管理和数据加密服务平台，提供安全合规、可靠易用的资源托管及密码运算服务。同时与天翼云云硬盘、对象存储、弹性文件、关系型数据库 MYSQL 等云产品无缝集成，实现云上原生数据的加密保护。

更多内容请参见云硬盘加密概述。

云硬盘快照策略

使用云硬盘快照策略对云硬盘进行周期性备份天翼云云硬盘快照是一种数据备份方式，云硬盘快照服务可以备份或者恢复整个云硬盘的数据，常用于数据备份、制作镜像、应用容灾等。

更多内容请参见云硬盘快照策略。

云主机访问控制

当您的云主机需要访问公网或被公网访问时您可以使用安全组和网络 ACL 对主

机进行访问控制设置，屏蔽不使用的端口，尽可能减少暴露在公网的端口数量。可以有效提高云主机的安全性。

安全组

安全组是一种网络安全防护机制，用于防止未经授权的访问和保护计算机网络免受恶意攻击。它是一种虚拟防火墙，用于限制入向和出向网络流量。安全组工作在网络层和传输层，它通过检查数据包的源地址、目标地址、协议类型和端口号等信息来决定是否允许通过。安全组创建后，用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，即受到这些访问规则的保护。

更多内容请参见安全组。

网络 ACL

网络 ACL 是一个子网级别的流量防护策略，您可以自定义设置网络 ACL 规则，并将网络 ACL 与子网绑定，实现对子网中云主机实例流量的访问控制。通过出方向/入方向规则控制出入子网的流量数据。

更多内容请参见网络 ACL。

常见的高危端口

TCP：42、135、137、138、139、444、445、593、1025、1068、1434、3127、3128、3129、3130、4444、4789、5554、5800、5900、9996

UDP：135-139、1026-1028、1068、1433、1434、4789、5554、9996

定期升级操作系统

您应该定期对云主机内的操作系统进行升级。了解操作系统发布的漏洞公告，及时安装漏洞补丁，以防您的主机收到侵害。天翼云并不承诺对客户操作系统内进行升级和维护。

服务器安全卫士

服务器安全卫士（原生版）（CT-CSS，Cloud Security System）是一款全方位保障云上服务器安全的产品，能全面识别并管理服务器中的信息资产、实时监测服务器风险并阻止非法入侵行为，当发现服务器出现安全问题时，第一时间向您发出告警通知。主要包括资产清点、漏洞扫描、入侵检测、基线检查、弱口令检测、病毒查杀、文件防勒索等功能，帮助您构建服务器安全防护体系。

根据支持功能不同，服务器安全卫士（原生版）分为基础版、企业版、旗舰版和

增值服务。不同规格功能差异请参见：服务器安全卫士产品规格。

Linux 服务器 SSH 登录的安全加固

SSH 是远程登录 Linux 服务器的主要方式，但它也是黑客进行暴力破解和恶意攻击的主要入口之一。为了加强 Linux 云主机的 SSH 登录安全，进行安全加固。

更多内容请参见 Linux 服务器 SSH 登录的安全加固。

基线检测

对服务器操作系统、数据库、关键应用软件配置等进行检测，帮助用户提前识别出可能导致安全漏洞的不安全配置项，例如不必要的服务开启、过时的软件版本、未启用的安全特性等。通过基线检测，可以及时发现并修复这些潜在的安全风险。

更多内容请参见基线检测

漏洞扫描

服务器安全卫士（原生版）支持扫描 Linux 软件漏洞、Windows 系统漏洞、Web-CMS 漏洞、应用漏洞，并提供漏洞的修复建议和一键修复功能，帮助您及时了解云主机操作系统中存在的风险，及时修复主机漏洞。

更多内容请参见扫描漏洞

弱口令检测

通过与弱口令库对比，检测系统账号和应用账号口令是否属于常用的弱口令，提示用户修改不安全的口令。

更多内容请参见弱口令检测

病毒查杀

服务器安全卫士（原生版）的病毒查杀功能，支持对挖矿木马、蠕虫、勒索病毒等进行有效的检测，提供灵活的检测方式，支持一键检测和定时检测方式，通过简单操作即可完成对病毒的处理，支持对病毒文件进行隔离、删除和信任。

更多内容请参见病毒查杀

文件防勒索

围绕事前、事中、事后三个阶段进行防护，可以有效地应对勒索病毒的威胁。

更多内容请参见文件防勒索

网页防篡改

网页防篡改可实时监控网站目录并通过备份恢复被篡改的文件或目录，保障重要

系统的网站信息不被恶意算改，防止出现挂马、黑链、非法植入恐怖威胁、色情等内容。此功能为付费的增值服务，需要单独购买。

更多内容请参见网页防篡改

7.12.2 安全组

7.12.2.1 安全组概述

什么是安全组

安全组是一种网络安全防护机制，用于防止未经授权的访问和保护计算机网络免受恶意攻击。它是一种虚拟防火墙，用于限制入向和出向网络流量通行。安全组工作在网络层和传输层，它通过检查数据包的源地址、目标地址、协议类型和端口号等信息来决定是否允许通过。安全组创建后，用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，即受到这些访问规则的保护。

不同资源池的安全组存在部分差异，如表 1 所示：

表 1 可用区资源池和地域资源池的差异

对比项	多可用区资源池	地域资源池
作用机制	“白名单”机制，即不匹配规则时，默认拒绝所有访问。	“白名单”机制，即不匹配规则时，默认拒绝所有访问。
创建安全组	需指定和 VPC 的关联关系	无需指定和 VPC 的关联关系
默认安全组	一个 VPC 一个默认安全组	一个资源池一个默认安全组
自定义模板安全组的默认规则	自定义模板类型的安全组不存在默认规则	每个自定义模板类型的安全组存在两条默认规则
默认安全组规	存在默认安全组规则，具体规	存在默认安全组规则，具体规

对比项	多可用区资源池	地域资源池
则	则以表 3 为准	则以表 2 为准
安全组是否有状态	出/入方向均有状态	仅出向有状态
安全组模板	自定义、通用 Web 服务器、开放全部端口	自定义、通用 Web 服务器、开放全部端口

默认安全组：

- 对于地域资源池，系统会为每个用户默认创建一个安全组，多个 VPC 可以共用同一个安全组。默认安全组的默认规则可参考表 2“地域资源池默认安全组规则”。
- 对于可用区资源池，系统会为每个 VPC 默认创建一个安全组。一般来讲不同 VPC 之间是不同的业务，VPC 之间是相互隔离的，相同业务一般部署在一个 VPC。大多数情况下，不同的 VPC 由于业务的差异，所使用的安全组规则应该是不一样的。每个 VPC 自动建立一个默认安全组，可以满足用户不同业务需要不同安全组的场景。默认安全组的默认规则可参考表 3“可用区资源池默认安全组规则”。

安全组状态：

- 对于地域资源池来说，安全组出向是有状态的。如果您从实例发送一个出站请求，且该安全组的出站规则是放通的话，那么无论其入站规则如何，都将允许该出站请求的响应流量流入。
- 对于可用区资源池来说，安全组出/入方向均有状态的。如果您从实例发送一个出站请求，且该安全组的出站规则是放通的话，那么无论其入站规则如何，都将允许该出站请求的响应流量流入。同理，如果该安全组的入站规则是放通的，那无论出站规则如何，都将允许入站请求的响应流量可以出站。

自定义安全组：

- 对于地域资源池来说，用户创建安全组时，模板类型为自定义的安全组会存

在两条默认规则，即出向默认放通所有 IP 地址（0.0.0.0/0、::/0）流量的数据报文通过。

- 对于可用区资源池来说，用户创建安全组时，模板类型为自定义的安全组不存在默认规则，如未添加规则，则默认安全组出、入方向将均拒绝所有访问。

不同资源池列表见产品简介-资源池区别页面，实际情况以控制台展现为准。

安全组规则

为了更好地管理安全组的入出方向，您可以设置安全组规则，去控制云服务器的出入向流量。通过配置适当的规则，控制和保护加入安全组的弹性云服务器的访问。

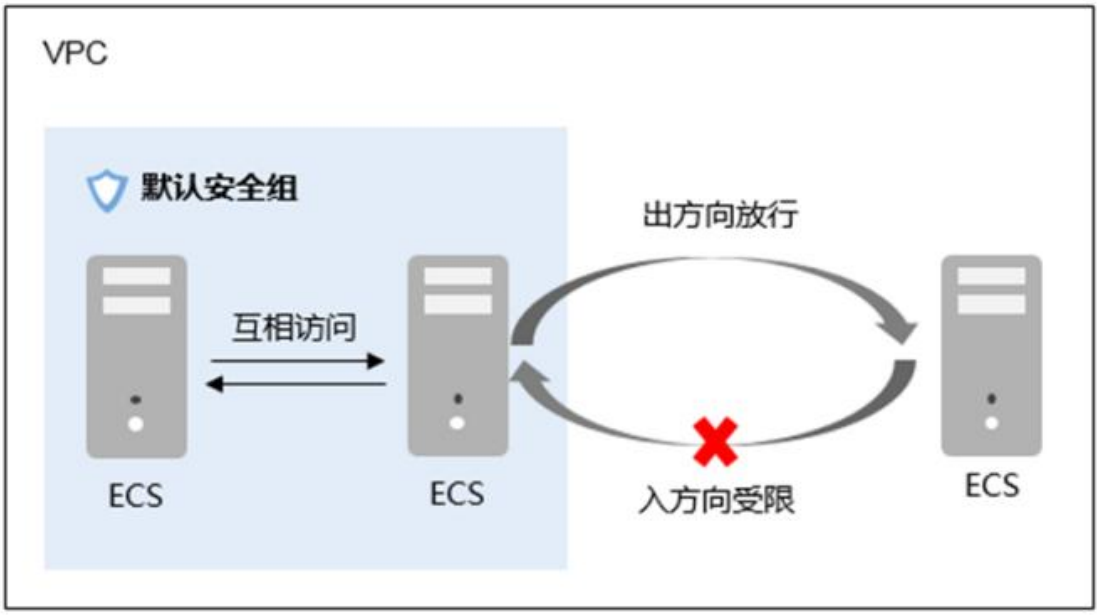
- 安全组规则可分为入向规则和出向规则。入向规则用于控制流入服务器实例的流量，出向规则用于控制从服务器实例流出的流量。默认安全组会自带一些默认规则，您也可以自定义添加安全组规则。
- 安全组规则主要遵循白名单机制，具体说明如下：

  入方向规则：入方向指外部访问安全组内的云服务器的指定端口。当外部请求匹配上安全组中入方向规则的源地址，并且授权策略为“允许”时，允许该请求进入，其他请求一律拦截。通常情况下，您一般不用在入方向配置授权策略为“拒绝”的规则，因为不匹配“允许”规则的请求均会被拦截。

  出方向规则：出方向指安全组内的云服务器访问外部的指定端口。在出方向中放通全部协议和端口，配置全零 IP 地址，并且策略为“允许”时，允许所有的内部请求出去。0.0.0.0/0 表示所有 IPv4 地址，::/0 表示所有 IPv6 地址。

地域资源池：

对于地域资源池来说，系统会为每个用户默认创建一个安全组，默认安全组包含一系列默认规则，主要是在出方向上的数据报文全部放行，入方向访问受限，安全组内的云服务器无需添加规则即可互相访问。



默认安全组规则如下表 2：

表 2 地域资源池默认安全组规则

方向	授权策略	类型	协议	端口范围	目的地址/源地址	说明
出方向	允许	IPv4	Any	Any	0.0.0.0/0	允许所有 IPv4 类型的出站流量的数据报文通过。
出方向	允许	IPv6	Any	Any	::/0	允许所有 IPv6 类型的出站流量的数据报文通过。
入方向	允许	IPv4	Any	Any	默认安全组 ID（例如：sg-xxxxx）	仅允许安全组内的云服务器彼此通信，丢弃其他入站流量的全部数据报文。

方向	授权策略	类型	协议	端口范围	目的地址/源地址	说明
入方向	允许	IPv6	Any	Any	默认安全组 ID（例如：sg-xxxxx）	仅允许安全组内的云服务器彼此通信，丢弃其他入站流量的全部数据报文。
入方向	允许	IPv4	TCP	22	0.0.0.0/0	允许所有 IP 地址通过 SSH 远程连接到 Linux 云服务器。
入方向	允许	IPv4	TCP	3389	0.0.0.0/0	允许所有 IP 地址通过 RDP 远程连接到 Windows 云服务器。
入方向	允许	IPv4	ICMP	Any	0.0.0.0/0	使用 ping 程序测试云服务器之间的通讯状况。

可用区资源池：

对于可用区资源池来说，系统会为每个 VPC 默认创建一个安全组，默认安全组包含一系列默认规则，主要是在出方向上的数据报文全部放行，入方向访问受限。默认安全组规则如下表 3：

表 3 可用区资源池默认安全组规则

方向	授权策略	类型	优先级	协议	端口范围	目的地址/源地址	说明
出方向	允许	IPv4	100	Any	Any	0.0.0.0/0	允许所有 IPv4 类型的出站流量的数据报文通过。
出方向	允许	IPv6	100	Any	Any	::/0	允许所有 IPv6 类型的出站流量的数据报文通过。
入方向	允许	IPv4	99	ICMP	Any	0.0.0.0/0	使用 ping 程序测试云服务器之间的 IPv4 地址通讯状况
入方向	允许	IPv6	99	ICMP	Any	::/0	使用 ping 程序测试云服务器之间的 IPv6 地址通讯状况
入方向	允许	IPv4	99	TCP	22	0.0.0.0/0	允许所有 IPv4 地址通过 SSH 远程连接到 Linux 云服务器。
入方向	允许	IPv6	99	TCP	22	::/0	允许所有 IPv6 地址通过 SSH 远程连接到 Linux 云服务器。
入方	允许	IPv4	99	TCP	3389	0.0.0.0/0	允许所有 IPv4 地址通过 RDP 远程连接到 Windows 云

方向	授权策略	类型	优先级	协议	端口范围	目的地址/源地址	说明
向							服务器。
入方向	允许	IPv6	99	TCP	3389	::/0	允许所有 IPv6 地址通过 RDP 远程连接到 Windows 云服务器。
入方向	拒绝	IPv4	100	Any	Any	0.0.0.0/0	禁止所有 IPv4 类型的入站流量的数据报文通过。
入方向	拒绝	IPv6	100	Any	Any	::/0	禁止所有 IPv6 类型的入站流量的数据报文通过。

实践建议

以下是一些关于安全组的建议和实践经验，帮助确保您的云环境的安全性，您可根据具体需求和环境来制定适合的安全组策略：

- 原则上，安全组规则取最小权限原则，通过设置所需的端口和协议，限制对必要 IP 地址的访问。只允许最少必要的流量进出您的资源实例。
- 定期更新安全组规则，以适应您的业务需求的变化。不再需要的规则应被删除，根据业务变化添加新的安全组规则。
- 根据资源的安全需求，将资源实例划分为不同的安全组。通过多层级的安全组可以实现细粒度的安全控制，确保安全性与灵活性之间的平衡。
- 通过有规范的命名等方式有意识地规划和管理安全组规则，易于查找。
- 安全组应与其他安全措施如网络 ACL、防火墙等结合使用，以提供更全面的安全保护。

- 建议您不要直接修改线上环境使用的安全组，修改后的安全组会自动应用在安全组内的所有云服务器上，因此您可以先克隆一个安全组，在测试环境中进行调试，确保修改后云服务器之间的通讯正常，再将修改后的安全组同步到线上环境。

7.12.2.2 安全组配置示例

不同安全组内的弹性云服务器内网互通

场景举例：

在同一个 VPC 内，用户需要将某个安全组内一台弹性云主机上的资源拷贝到另一个安全组内的弹性云主机上时， 用户可以将两台弹性云主机设置为内网互通后再拷贝资源。

安全组配置方法：

由于同一个 VPC 内，在同一个安全组内的弹性云服务器默认互通，无需配置。但是，在不同安全组内的弹性云服务器默认无法通信，此时需要添加安全组规则，使得不同安全组内的弹性云主机内网互通。

在两台弹性云主机所在安全组中分别添加一条入方向安全组规则，放通来自另一个安全组内的实例的访问，实现内网互通，安全组规则如下所示。

协议	方向	端口范围/ICMP 协议类型	源地址
设置内网互通时使用的协议类型（支持 TCP/UDP/ICMP/All）	入方向	设置端口范围或者 ICMP 协议类型	IPv4 地址、IPv4 CIDR 或者另一个安全组的 ID

仅允许特定 IP 地址远程连接弹性云主机

场景举例：

为了防止弹性云主机被网络攻击，用户可以修改远程登录端口号，并设置安全组规则只允许特定的 IP 地址远程登录到弹性云主机。

安全组配置方法：

以仅允许特定 IP 地址（例如，192.168.20.2）通过 SSH 协议访问 Linux 操作系

统的弹性云主机的 22 端口为例，安全组规则如下所示。

协议	方向	端口范围	源地址
SSH（22）	入方向	22	IPv4 地址、IPv4 CIDR 或者另一个安全组的 ID。例如：192.168.20.2

SSH 远程连接 Linux 弹性云主机

场景举例：

创建好 Linux 弹性云主机后，为了通过 SSH 远程连接到弹性云主机，您可以添加安全组规则。

安全组配置方法：

协议	方向	端口范围	源地址
SSH（22）	入方向	22	0.0.0.0/0

RDP 远程连接 Windows 弹性云主机

场景举例：

创建好 Windows 弹性云主机后，为了通过 RDP 远程连接弹性云主机，您可以添加安全组规则。

安全组配置方法：

协议	方向	端口范围	源地址
RDP（3389）	入方向	3389	0.0.0.0/0

公网 ping 弹性云主机

场景举例：

创建好弹性云主机后，为了使用 ping 程序测试弹性云主机之间的通讯状况，您需要添加安全组规则。

安全组配置方法：

协议	方向	端口范围	源地址
ICMP	入方向	All	0.0.0.0/0

弹性云主机做 Web 服务器

场景举例：

如果您在弹性云主机上部署了网站，即弹性云主机作 Web 服务器用，希望用户能通过 HTTP 或 HTTPS 服务访问到您的网站，您需要在弹性云主机所在安全组中添加以下安全组规则。

安全组配置方法：

协议	方向	端口范围	源地址
TCP	入方向	80（HTTP）	0.0.0.0/0
TCP	入方向	443（HTTPS）	0.0.0.0/0

弹性云主机做 DNS 服务器

场景举例：

如果您将弹性云主机设置为 DNS 服务器，则必须确保 TCP 和 UDP 数据可通过 53 端口访问您的 DNS 服务器。您需要在弹性云服务器所在安全组中添加以下安全组规则。

安全组配置方法：

协议	方向	端口范围	源地址
TCP	入方向	53	0.0.0.0/0
UDP	入方向	53	0.0.0.0/0

使用 FTP 上传或下载文件

场景举例：

如果您需要使用 FTP 软件向弹性云主机上传或下载文件，您需要添加安全组规则。

安全组配置方法：

您需要在弹性云主机上先安装 FTP 服务器程序，再查看 20、21 端口是否正常工作。

协议	方向	端口范围	源地址
FTP	入方向	20-21	0.0.0.0/0

7.12.2.3 配置安全组规则

7.12.2.3.1 修改安全组规则

操作场景

安全组创建成功后，当您的云主机需要与外部网络通讯时，您可根据业务需求自定义添加新的出方向、入方向安全组规则，这可以帮助保护云主机免受未经授权的访问。

入方向：指从外部访问安全组规则下的弹性云主机。

出方向：指安全组规则下的弹性云主机访问安全组外的实例。

安全组规则有数量限制，您应尽量保持规则的精简。

前提条件

- 您已经完成云主机的创建，已关联安全组。

操作步骤

1. 登录控制中心。
2. 在控制中心页面左上角选择地域，此处选择华东-华东 1。
3. 选择“计算>弹性云主机”，进入云主机控制台。
4. 在云主机列表中，点击待添加安全组规则的云主机名称，进入云主机详情页面。
5. 选择安全组页签，在安全组列表中，点击安全组名称左侧的展开符，可将安全组详情信息展开
6. 修改安全组规则需要单击安全组名称跳转进入安全组详情。



7. 在安全组详情页面，配置安全组规则后，返回云主机控制台刷新。

- 安全组规则具体配置请参考“安全组”。

注意

安全组规则的修改将影响所有当前使用该安全组的云主机。

操作指引

- 关于使用安全组时的一些常见类问题，可参考常见问题“安全类”。
- 关于配置安全组规则放通弹性云主机对应的常见端口，可参考文档“云主机的常用端口”。

7.12.2.4 更改安全组

操作场景

安全组是一种重要的网络安全防护策略，用于管理和控制云主机的网络访问。当您创建好云主机并关联安全组后，可以根据业务需求随时更改安全组。

前提条件

- 已完成弹性云主机的创建，已关联安全组。
- 已创建另一个安全组，作为更换后的安全组。

操作步骤

1. 登录控制中心。
2. 在控制中心页面左上角选择地域，此处选择华东-华东 1。
3. 选择“计算>弹性云主机”，进入云主机控制台。
4. 在云主机列表中，点击待更换安全组规则的云主机名称，进入云主机详情页面。
5. 在云主机详情页面，选择安全组页签，点击“更改安全组”按钮，弹出更改安全组弹窗。
6. 在更改安全组弹窗中，可下拉选择其他安全组进行关联。关联完成后单击“确定”，完成安全组的更改。

7. 更改完成后，可以在安全组页签下看到更改后的安全组。

7.12.3 数据保护技术

7.12.3.1 用户加密

天翼云支持以下方式，助您对弹性云主机资源进行加密，从而提升数据的安全性。

云硬盘加密

天翼云云硬盘支持系统盘加密和数据盘加密。

云硬盘加密功能：创建云硬盘时，用户可以选择是否加密此云硬盘，云硬盘创建完成后加密属性无法更改。

云主机系统盘加密：创建云主机时，支持在创建时直接设置系统盘加密。

云主机数据盘加密：创建云主机时，支持在创建时直接设置数据盘加密。

云硬盘加密与快照：加密云硬盘生成的快照及通过这些快照创建的云硬盘将自动继承加密功能属性。

云硬盘加密与备份：加密云硬盘生成的备份及通过这些备份创建的云硬盘将自动继承加密功能属性。

密钥管理

天翼云使用行业标准的 AES-256 算法，利用数据密钥加密您的云硬盘数据，加密云硬盘使用的密钥由天翼云自研密钥管理（KMS，Key Management Service）功能提供，用户可轻松创建并管理密钥，满足数据加解密及数字签名验签等需求，安全便捷。

KMS 通过使用硬件安全模块 HSM（Hardware Security Module）保护密钥的安全，所有的用户密钥都由 HSM 中的根密钥保护，避免密钥泄露。KMS 对密钥的所有操作都会进行访问控制及日志跟踪，并提供所有密钥的使用记录，满足审计和合规性要求。密钥管理可以轻松满足对小数据和大量数据的加解密。

工作原理

在了解云硬盘加密工作原理之前，首先需要了解两个概念：

默认密钥（Default CMK）：用户第一次通过对应云服务使用 KMS 加密时，系统自动生成的并托管在用户账号下的服务密钥。

用户主密钥（Customer Master Key，CMK）：用户主密钥包括对称密钥及非对称密钥,主要用于加密保护数据密钥并产生信封,也可直接用于加密少量的数据。用户可以调用 KMS 的 API CreateKey 创建一个用户主密钥。

第一次使用加密云硬盘时，系统会自动创建一个用户主密钥（CMK），该密钥有且仅有一个，且是在 KMS 中的相应地域所创建，并将其存储在受严格的物理和逻辑安全控制保护的密钥管理服务上。

每个地域的加密云硬盘，都需要通过 256 位数据密钥（DK）进行加密，此数据密钥（DK）具备地域唯一性，即每个地域都有且仅有一个。该密钥受 KMS 提供的密钥管理基础设施的保护，能有效防止未经授权的访问。云硬盘的数据密钥（DK）仅在实例所在的宿主机的内存中使用，不会以明文形式存储在任何持久化介质（即使是云硬盘本身）上。

在创建加密云硬盘并将其挂载到实例后，以下数据都将关联此密钥并进行加密：

云硬盘中的静态数据

云硬盘和实例间传输的数据（实例操作系统内的数据不加密）

通过加密云硬盘创建的快照

对弹性伸缩的影响

如果使用加密的弹性云主机创建弹性伸缩配置，通过伸缩配置创建的云主机，加密方式与原云主机保持一致。

7.12.4 重点操作短信二次验证

7.12.4.1 验证规则

短信验证规则

1. 若您已开启短信验证，在您进行重点操作时，在点击最后一步"确认"时弹出二次认证窗口，进行短信验证，如下图所示。



2. 点击“发送验证码”，系统向绑定的天翼云手机号发送二次认证码，每间隔 1 分钟才能获取验证码。在您进行短信验证时，验证码单次有效时间 10 分钟，请在有效时间内完成验证，如下图所示。

注意

请勿关闭弹窗并及时验证，关闭该弹窗验证码即失效。



3. 点击“确认”按钮后，即可进行重点操作。短信验证功能一旦开启，所有重点操作统一开启，短信验证的有效期为 15 分钟，15 分钟内做其他操作不需要重复验证。

7.12.4.2 开启重点操作短信验证

操作说明

开启重点操作短信验证保护后，您在进行对云主机进行关机、重启、重装等重点操作的场景时，需输入验证码，以防止错误操作造成的损失和风险。

需要注意的是：短信验证功能一旦开启，所有重点操作统一开启，短信验证的有效期为 15 分钟，15 分钟内做其他操作不需要重复验证。

操作流程

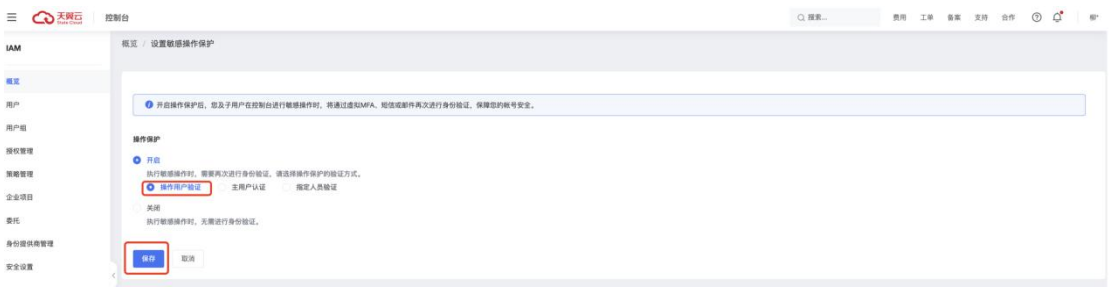
1. 登录天翼云账号，点击右上角头像下方的【个人中心】进入基本信息，下拉点击“安全设置”再点击“登录保护”如下图所示。



2. 点击【立即修改】跳转至概览/设置敏感操作保护，选择“开启”，如下图所示。



3. 选择操作用户验证，如下图所示。



4. 点击“保存”，即可开启重点操作保护。

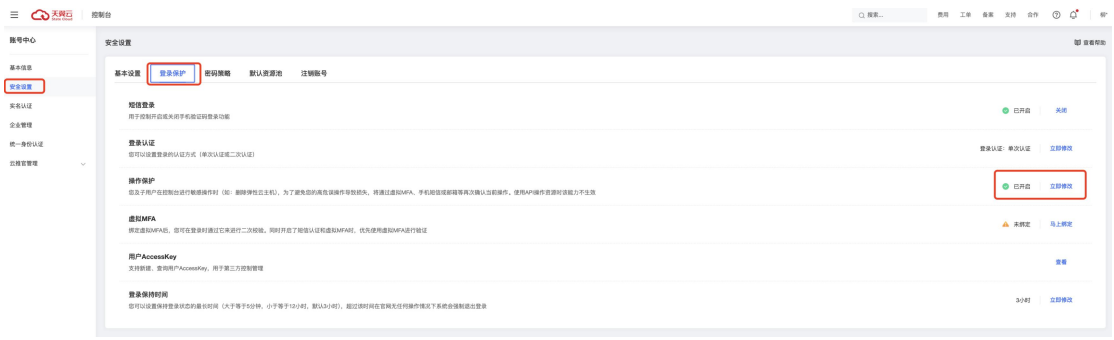
7.12.4.3 关闭重点操作短信验证

操作说明

关闭后，您在进行重点操作时，无需输入验证码进行验证。

操作流程

1. 登录天翼云账号，点击右上角头像下方的【个人中心】进入基本信息，下拉点击“安全设置”再点击“登录保护”如下图所示。



2. 点击【立即修改】跳转至概览/设置敏感操作保护，如下图所示。



3. 选择关闭，如下图所示。



4. 点击【保存】，即可关闭重点操作保护。

7.12.5 云服务器 CTIAM

7.12.5.1 统一身份认证 IAM 介绍

IAM 为您提供的主要功能包括：精细的权限管理、安全访问、通过用户组批量管理用户权限、委托其他帐号管理资源等。

身份管理

访问控制 IAM 中的身份包括 IAM 用户、IAM 用户组。

IAM 用户有确定的登录密码和访问密钥，IAM 用户组则用于分类职责相同的 IAM 用户，IAM 用户和 IAM 用户组均可以被赋予一组权限策略。在需要协同使用

资源的场景中，避免直接共享天翼云账号的密码等信息，缩小不同 IAM 子用户的信息可见范围，可为 IAM 子用户和 IAM 用户组按需授权，即使不慎泄露机密信息，也不会危及天翼云账号下的所有资源。

权限管理

统一身份认证 IAM 通过权限策略描述授权的具体内容，权限策略包括固定的基本元素“Action”“Effect”等，更多信息，请参见“2.1.2 创建自定义策略“。为 IAM 用户、IAM 用户组在全局授权或企业项目授权中添加一组权限策略后，即可让其有权限访问指定资源。

权限策略分为系统策略和自定义策略:

系统策略

预置的系统策略，您只能使用不能修改。云服务器 ECS 相关的系统策略包含如下：

Ecs Admin：弹性云主机服务的管理者权限，包含弹性云主机所有控制权限（不含订单类权限）；

Ecs Viewer:弹性云主机服务的观察者权限，包含弹性云主机服务的列表页与详情页面权限；

自定义策略

您按需自行创建和维护的权限策略，关于自定义策略的操作和示例，请参见“创建自定义策略”。

弹性云主机接口对应权限表

如下是弹性云主机服务相关权限三元组及生效范围：

控制台接口	权限三元组	配置支持

		IAM（资源池/全局）	企业项目（资源组）
创建云主机	ecs:cloudServers:create	√	√
云主机列表获取	ecs:cloudServers:list	√	√
云主机详情获取	ecs:cloudServers:get	√	√
修改云主机名称	ecs:cloudServers:put	√	√
开机	ecs:cloudServers:start	√	√
关机	ecs:cloudServers:stop	√	√
重启	ecs:cloudServers:reboot	√	√
续订	ecs:cloudServers:create	√	√

退订(包周期)	ecs:cloudServers:delete	√	√
删除(按需付费)	ecs:cloudServers:delete	√	√
重置密码	ecs:cloudServers:resetServerPwd	√	√
一键重装	ecs:cloudServers:rebuild	√	√
全部重启	ecs:cloudServers:reboot	√	√
远程登录	ecs:cloudServers:rlogin	√	√
创建快照	ecs:snapshot:create	√	√
制作镜像	ims:serverImages:create	√	√
变配	ecs:cloudServers:update	√	√
到期转按需	ecs:switchRequired:create	√	√

转包周期		ecs:switchPeriod:create	√	√
云主机详情页	云主机详情获取	ecs:cloudServers:get	√	√
	弹性网卡列表获取	vpc:cloudServerNics:list	√	
	弹性网卡详情获取	vpc:cloudServerNics:get	√	
	绑定弹性网卡	ecs:cloudServerNics:binding	√	
	解绑弹性网卡	ecs:cloudServerNics:unbinding	√	
	修改内网 IP	ecs:publicIps:change	√	
		vpc:subnets:list	√	
	更换 VPC（绑定）	ecs:vpcs:change	√	
		vpc:subnets:list	√	

		vpc:securityGroups:list	√	
		vpc:vpcs:list	√	
	云硬盘列表获取	evs:volumes:list	√	√
	云硬盘详情获取	evs:volumes:get	√	√
	卸载	evs:volumes:detach	√	√
	挂载磁盘	evs:volumes:attach	√	√
	安全组列表获取	vpc:securityGroups:list	√	√
	安全组详情获取	vpc:securityGroups:get	√	√
	弹性 IP 列表获取	vpc:publicIps:list	√	√
	弹性 IP 详情获取	vpc:publicIps:get	√	√

	绑定弹性 IP	vpc:publicIps:update	√	√
	解绑弹性 IP	vpc:publicIps:detach	√	√
云主机组	创建云主机组	ecs:ServersGroups:create	√	
	云主机组列表获取	ecs:ServersGroups:list	√	
	云主机组详情获取	ecs:ServersGroups:get	√	
	添加云主机	ecs:ServersGroups:attach	√	
	移除云主机	ecs:ServersGroups:detach	√	
	云主机列表获取	ecs:cloudServers:list	√	
	删除云主机组	ecs:ServersGroups:delete	√	
快照	创建快照	ecs:snapshot:create	√	√

	云主机列表获取	ecs:cloudServers:list	√	√
	快照列表获取	ecs:snapshot:list	√	√
	快照详情获取	ecs:snapshot:get	√	√
	删除快照	ecs:snapshot:delete	√	√
	恢复数据	ecs:snapshot:rollback	√	√
	申请云主机	ecs:cloudServers:create	√	√
	删除快照	ecs:snapshot:delete	√	√
	创建快照策略	ecs:snapshotpolicy:create	√	
	快照策略列表	ecs:snapshotpolicy:list	√	
	快照策略详情	ecs:snapshotpolicy:get	√	

	修改快照策略	ecs:snapshotpolicy:update	√	
	停用快照策略	ecs:snapshotpolicy:stop	√	
	删除快照策略	ecs:snapshotpolicy:delete	√	
	启用策略/立即执行快照策略	ecs:snapshotpolicy:start	√	
SSH 密钥对	创建密钥对	ecs:serverKeypairs:create	√	√
	导入密钥对	ecs:serverKeypairs:create	√	√
	密钥对列表获取	ecs:serverKeypairs:list	√	√
	删除	ecs:serverKeypairs:delete	√	√

天翼云支持对用户组/子用户，进行资源池或全局维度的权限授权；同时也支持在企业项目中，对用户组进行资源组维度的权限授权。部分没有企业项目属性的接口或资源，授权只能以资源池或全局维度进行。以资源池或全局维度进行的授权判断，其优先级高于企业项目中的资源组维度授权。

7.12.5.2 通过 IAM 用户控制资源访问

在协同使用资源的场景下，根据实际的职责权限情况，您可以创建多个 IAM 用户并为其授予不同的权限，实现不同 IAM 子用户可以分权管理不同的资源，从而提高管理效率，降低信息泄露风险。本文介绍如何创建 IAM 子用户并授予特定权限策略，从而控制对云服务器 ECS 资源的访问。

操作步骤

创建 IAM 子用户

具体操作，请参见统一身份认证 IAM。

创建自定义策略

天翼云提供了访问云服务器 ECS 资源的系统策略，更多信息，请参见“1.2 权限管理”。如果系统策略不能满足需求，您还可以创建自定义策略，具体操作，请参见统一身份认证 IAM。

策略分为用户可以自行定义的自定义策略，以及预定义在平台录入的系统策略两类。

细粒度授权策略结构包括策略版本号（Version）及策略授权语句（Statement）列表。

策略版本号：Version 标识策略结构的版本号，目前为 1.1。

策略授权语句：Statement，包括了基本元素：作用（Effect）和权限集（Action）。

作用（Effect）包含两种：允许（Allow）和拒绝（Deny）。

授权项（Action）是对资源的具体操作权限，支持单个或多个操作权限。

1.脚本配置策略示例一：为 IAM 子用户配置云主机查看者权限。

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ecs:cloudServers:list",
                "ecs:cloudServers:get",
                "ecs:ServersGroups:list",
                "ecs:ServersGroups:get",
                "ecs:snapshot:list",
                "ecs:snapshot:get",
                "ecs:snapshotpolicy:list",
                "ecs:snapshotpolicy:get",
                "ecs:serverKeypairs:list",
                "ecs:generalView:list"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

2.脚本配置策略示例二：为 IAM 子用户配置云主机所有操作权限，以及 vpc 的所有操作权限（*代表取所有值）。

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ecs:*:*",
                "vpc:*:*"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

授权自定义策略

授予 IAM 用户访问所创建的自定义策略范围中的资源，具体操作，请参见统一身份认证 IAM。

授权系统策略

您也可以直接使用天翼云预制的产品系统策略对 IAM 子用户进行授权。

7.12.6 云审计

操作场景

本服务现已对接天翼云云审计服务，云审计服务提供对各种云资源操作的记录和查询功能，用于支撑合规审计、安全分析、操作追踪和问题定位等场景，同时提供事件跟踪功能，将操作日志转储至对象存储实现永久保存。

云审计可提供的功能服务具体如下：

- 记录审计日志：支持用户通过管理控制台或 API 接口发起的操作，以及各服务内部自触发的操作。
- 审计日志查询：支持在管理控制台对 7 天内操作记录按照事件类型、事件来源、资源类型、筛选类型、操作用户和事件级别等多个维度进行组合查询。

使用限制

- 云审计服务本身免费，包括时间记录以及 7 天内时间的存储和检索。
- 用户通过云审计能查询到多久前的操作事件：7 天。
- 用户操作后多久可以通过云审计查询到数据：5 分钟。
- 其它限制请参考使用限制-云审计。

操作步骤

1. 开通云审计服务。

参见开通云审计服务-云审计。

2. 查看云审计事件。

参见查看审计事件-云审计。

3. 在事件列表中，选择事件来源为“计算”，资源类型选择“云主机”，上方时间选择需要筛选的时间段。点击查询即可。

4. 在审计事件右侧点击详情，可以看到更详细的事件信息。

更多云审计相关使用说明和常见问题请参考用户指南、常见问题。

关键操作列表（多可用区资源池）

操作事件	字段
云主机创建	create_server
云主机开机	start_server
云主机关机	stop_server
云主机重启	restart_server

操作事件	字段
云主机重装	rebuild_server
云主机删除	delete_server
云主机变配	resize_server
云主机重置密码	change_server_password
云主机续订	renew_server
云主机退订	refund_server
创建密钥对	create_keypair
删除密钥对	delete_keypair

关键操作列表（单可用区资源池）

操作事件	字段
云主机创建	create_server
云主机退订	delete_server
云主机变配	resize_server
云主机开机	start_server

操作事件	字段
云主机关机	stop_server
云主机重启	reboot_server
一键重装	rebuild_server
远程登录	get_server_vnc
获取密码	getVMPassword
重置密码	change_server_password
导入密钥对	import_keypair
创建密钥对	create_keypair
删除秘钥对	delete_keypair

7.13 迁移

7.13.1 迁移云主机实例

产品定义

天翼云云迁移服务（CT-CMS，Cloud Migration Service）是一款由天翼云自主研发的专业迁移服务产品。基于 P2V/V2V 技术，CT-CMS 可以帮助用户将单台或多台迁移源从私有云、公有云或其他特殊环境，迁移至天翼云弹性云主机

（CT-ECS）。旨在提供有序、安全、便捷、轻松的数字资产、服务、IT 资源及应用程序迁移解决方案，同时保证云上业务的可用性、安全性和连续性。

迁移流程



操作场景

ECS 相关场景

天翼云 ECS 实例间迁移：

当需要把不同天翼云账号下的主机资源整合到一个天翼云账号下，或者将同账户不同区域的 CT-ECS 资源整合到同一个区域下时，可以采用主机迁移服务实现 CT-ECS 实例间的快速迁移。

其他云平台云主机迁移至天翼云：

CMS 支持将您在不同环境如各云服务平台等云主机、IDC 机房虚拟机、本地虚拟机迁移至天翼云，通过 CMS 可以实现相应业务需求。

应用场景

- 业务快速拓展：业务系统需要具备快速拓展的能力，但传统的云下设备拓展能力有限，因此存在上云的需求。在迁移过程中，需要保证业务正常运行，并在上云后获得优于云下体验，同时还要支持快速的业务拓展。
- 本地 IDC 陆续退网、业务持续上云：拥有自己的本地 IDC 机房，但随着时间的推移，机房内的服务器集群已经运行了 8-10 年，逐渐到达设备的运行年限。在这种情况下，大部分企业会选择将达到运行年限的设备进行退网处理。同时，随着设备退网，原有的业务系统、应用集群、数据库、对象存储等也面临着上云迁移的需求。
- 高密特殊上云迁移：政府单位以及部分特殊企业，具有较高的安全需求；需

要采取一系列措施来确保数据和信息的安全性。在这种场景下，通常会对内外网进行隔离，限制上云迁移软件的使用。

- 各环境迁移至天翼云：业务在不同环境中运行，包括在 IDC 机房、本地虚拟机、天翼云不同区域以及其他云服务平台上。现在需要将这些不同环境中的业务迁移至天翼云平台。

服务场景详细描述可参考：云迁移-产品应用场景。

约束与限制

项目	约束和限制
迁移源和迁移任务数量限制	每位用户可激活的迁移源数量上限为 1000 台；每位用户可创建迁移任务数量为 1000 个；每位用户可并发执行的迁移任务数量为 50 台。
迁移源绑定限制	每个迁移源同一时刻仅能绑定一个目标端，对应生成唯一一个"未完成"状态的迁移任务；已完成或异常的迁移任务不能作为迁移源绑定目标端。
迁移源软硬件、授权限制	未授权的应用软件、盗版软件、集群数据库、U-Key 等可能影响产品的使用，应避免使用。
迁移	目标机内存必须大于 4GB；目标机配置建议与迁移源相同，如果迁移

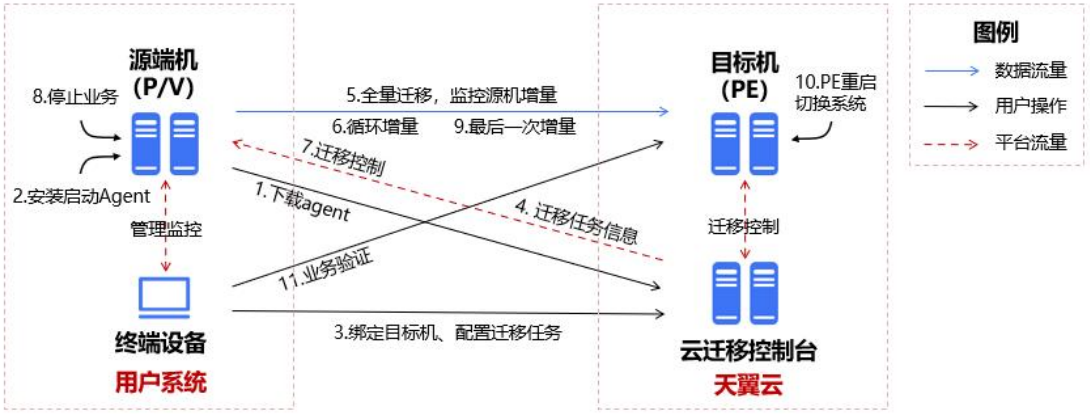
源目标端配置限制	源内存小于 4GB，迁移完成后可以根据需求进行缩扩容等修改。
迁移源目标端限制	不支持裸金属服务器。
迁移源网络限制	仅支持独享 IP。
迁移盘符限制	引导位置必须在第一个盘的第一个分区。
迁移源存储服务限制	不支持共享存储。
迁移源磁盘数、容量限制	单盘容量无限制，但磁盘个数应确定在 24 个以下。如果超过该限制，需要对源机磁盘进行减少或通过工单进行技术咨询。

数据库与集群服务限制	Oracle RAC 需要使用共享存储、部分使用 ASM 磁盘，CMS 无法提供 ASM 磁盘、Oracle RAC 迁移。大型库不建议使用 CMS 进行迁移，建议使用数据库专业迁移工具进行迁移。如果使用 CMS 进行迁移，在使用过程中应当在“停止增量”步骤前停止相关应用进程。
网络环境因素限制	网络环境问题可能影响迁移速率，应先测试迁移源出口带宽、目标端入流量带宽以及 CMS 控制台带宽限制配置；注意是否存在特殊的网络、专线以及是否具备公网环境。
复杂业务拉起、支撑限制	迁移复杂业务时，需要记录业务关联信息，并根据需求按业务系统进行迁移、切换、测试；增量迁移时，需要停止相关业务进程，以保持一致性。业务上云后，也需相关熟悉业务人员进行测试验证，以此减少业务停机时间。
特定需求限制	如果存在涉密文件、需要保持 IP 不变等特定需求，需要通过工单进行技术咨询，以满足相应需求。
迁移源资源与性能限制	CMS 代理端部署于源机，需占用源机的 CPU 与内存，当源机 CPU 使用率或内存使用率过高时，CMS 代理端会影响业务，甚至导致迁移源宕机。正常 4C4G 的机器能正常满足迁移所需资源。CMS 需在源机 agent 启动后，持续对源机进行监控，并给出一定周期内 CPU、内存使用情况、硬盘数据增长量等数据。用户根据给出数据对系统进行评估。若无法准确评估需找相关专业人员确认。
目标端设	目标机规格不支持 GPU 类相关规格，包括 GPU 图形加速基础型与 GPU 计算加速型。

备规格的规格限制	
操作系统	支持迁移的 Windows 操作系统参见 Windows 兼容性列表。支持迁移的 Linux 操作系统参见 Linux 兼容性列表。不支持迁移多操作系统。

操作步骤

云迁移服务的操作步骤如下：



1. **在迁移源服务器上安装并运行 Agent：** 在需要迁移的源服务器上安装和运行迁移 Agent。该 Agent 用于采集源服务器的配置信息、数据和日志等，以便后续的迁移操作。

2. **安装启动 Agent ：** 启动 Agent 后会将其采集到的必要信息回调给云迁移服务（CMS）的控制台。这些信息包括源服务器的配置、数据和日志等，用于后续的迁移任务配置和分析。

3. **绑定目标机、配置迁移任务：** 在 CMS 控制台上，您可以绑定目标机并进行迁移任务配置，管理员可以根据采集到的信息配置迁移任务。这些任务包括迁移范围、迁移策略、迁移时间窗口等。管理员可以根据业务需求和实际情况来制定合理的迁移计划。

4. **迁移任务信息：** 一旦迁移任务配置完成，迁移 Agent 即可获取并执行来自云迁移服务的迁移指令，开始迁移任务。

5. **全量迁移、监控源机增量：** 开始迁移任务后，CMS 开始全量迁移。如果任务配置勾选增量，CMS 会在全量开始时监控全量过程产生的修改、增量数据。全量迁移完成后，会弹窗提醒全量迁移完成，可进行后续操作。

6. **循环增量：** 若您迁移时启用了增量，迁移程序会在全量迁移完成后，自动进入循环增量模式。每 5 分钟向目标机同步上一个时间段内的增量文件。

7. **迁移控制：** CMS 可在迁移过程中您可在“操作”选项卡，对不同任务进度进行操作。包括手动增量、停止增量、开始核查、取消核查、开始修复、取消修复、引导修复等操作。

8. **停止业务：** 迁移任务配置中若启用了增量同步，在全量迁移完成后或增量阶段割接前，您需要停止迁移源机应用，再进行后续迁移操作。

若未配置增量同步，则在“开始迁移”前您需要停止迁移源机应用以保证数据一致性。

9. **最后一次增量：** 在“操作”选项卡单击“停止增量”，进行最后一次增量同步，同步完毕后，进入后续的迁移步骤。

10. **PE 重启切换系统：** 在源机完成全量迁移后或停止增量同步后，可以下发引导修复任务，目标机进入引导修复状态，引导修复完毕后，目标机自动重启。引导修复操作会根据任务配置中的网卡配置信息，自动配置目标机 IP 地址。自动进行驱动安装和天翼云 agent 安装。

11. **业务验证：** 迁移任务完成后，您可登录目标机查看迁移结果，并验证迁移后目标机应用及服务可用性。

具体操作指导可参考云迁移-用户指南。

访问方式

使用 CMS 迁移服务，您需在迁移源服务器和 CMS 控制台进行操作。

控制中心方式

如果用户已注册公有云，可在登录后开始使用云迁移服务。

1. 进入天翼云首页，在上方“产品”栏展开左侧服务导航栏并选择“安全及管理 > 管理工具 > 云迁移”。
2. 单击“立即开通”，进入云迁移服务界面。
3. 您可以在上方 切换所在区域，并使用云迁移服务。

7.13.2 迁移本地盘云主机数据

操作场景

本地盘云主机，同时具有本地盘和云盘两种数据存储的方式，当用户出现以下应用场景时，需要考虑自行迁移本地盘云主机数据：

- 需要切换数据到另一台主机。
- 需要扩大本地盘容量，但由于当前不支持本地盘的变配操作，故需要用户自行迁移数据至更大规格的本地盘中。

操作步骤

1. 创建云硬盘。

用户需要前往云硬盘产品控制台界面，根据迁移的数据量大小，购买合适大小的云硬盘。

2. 挂载云硬盘。

将新创建购买的云硬盘，挂载到当前需要进行数据迁移的本地盘云主机上，可在云硬盘控制台列表通过“挂载”操作执行。

3. 自行备份本地盘数据。

用户需要登陆到本地盘云主机，自行操作，将本地盘数据迁移至新挂载的云硬盘上。

至此用户已完成数据的迁移，用户继续将包含迁移数据的云硬盘挂载到其它主机上，以完成数据在不同主机间的迁移，步骤如下：

4. 创建云主机。

用户需要前往云主机产品控制台界面，根据迁移的数据量大小，购买合适大小的本地盘云主机。

5. 挂载云硬盘。

将云硬盘，挂载到新创建的云主机上，可通过控制台的"挂载"操作进行。

6. 迁移云盘数据。

将已挂载的云硬盘数据迁移到当前新创建的本地盘云主机中，进行存储。

7.14 运维

7.14.1 云助手

7.14.1.1 云助手产品介绍

产品概述

天翼云云助手（CT-CA， Cloud Assistant）是专门为云服务器打造的原生自动化运维工具，免密码、免登录、无需使用跳板机，即可批量执行命令（Shell、PowerShell、Bat、Python 等），完成运行自动化运维脚本、轮询进程、安装卸载软件、启动或停止服务、安装补丁或安装安全更新等任务。

产品优势

批量执行命令

通过云助手，您能够以安全可靠的方式远程管理大规模的实例，无需登录服务器、使用堡垒主机或 SSH。

提供集成工具

云助手可作为运维部署工具，免费支持 API，可供集成。

查询历史记录

通过云助手执行的操作系统脚本命令，您可以查询到其历史记录信息。

实现自动化

您可以预先创建常用的操作命令，并通过云助手高效完成您的日常任务。例如：注册表编辑、用户管理、软件和补丁安装等。

应用场景

当您涉及以下业务场景时，可通过云助手快速实现需求：

- 上传并运行自动化运维脚本。
- 执行常见操作任务。
- 运行已保存的脚本。
- 批量安装软件或应用。
- 安装补丁或安全更新。
- 上传文件或证书到指定目录。

使用限制

云助手具备以下限制项：

- 2024 年 6 月 30 日之后使用公共镜像创建的实例，默认已预装云助手客户端。若您的实例于 2024 年 6 月 30 日前购买，请参考 安装云助手客户端 进行安装。
- 单次执行命令实例数量上限为 100。
- 云助手仅保存最近三个月的命令执行历史记录，且记录数量最大保存 10 万条。
- 创建的 Bat、PowerShell、Python 或者 Shell 脚本和自定义参数的使用场景与文件大小说明如下：
  1. 上传文件：文件大小不能超过 24 KB。
  2. 创建命令：命令大小不能超过 24 KB。

云主机只能在以下操作系统中运行云助手命令：

- Linux
- Windows

物理机只能在以下操作系统中运行云助手命令：

- Linux

资源占用

主机资源	Linux 操作系统	Windows 操作系统

主机资源	Linux 操作系统	Windows 操作系统
CPU	平均 CPU 使用率不到 1%	平均 CPU 使用率不到 1%
物理内存	小于 50 MB	小于 50 MB
磁盘 I/O	平稳运行时几乎没有 I/O，仅在下载升级安装包和保存命令脚本等场景下产生磁盘 I/O	平稳运行时几乎没有 I/O，仅在下载升级安装包和保存命令脚本等场景下产生磁盘 I/O
网络 I/O	平稳运行时仅有心跳上报等数据产生的少量 I/O	平稳运行时仅有心跳上报等数据产生的少量 I/O

名词解释

云助手的常用名词及其具体描述如下表所示。

常见名词	说明
云助手	云助手官方名称，可以帮您在弹性云主机或物理机实例上自动以及批量执行日常维护任务。
云助手 Agent	安装在弹性云主机或物理机实例中的轻量级插件，所有在实例中完成的命令都会通过云助手 Agent 执行。
命令执行路径	云助手会将用户的命令内容以文件形式先保存到实例上，然后再执行文件，具体默认保存路径如下： Windows 默认保存路径：C:/Windows/System32/ Linux 默认保存路径：/tmp/
命令	需要在实例中执行的具体命令操作，如一份 Shell 脚本或者 PowerShell 脚本。
单次执行	在一台或者多台实例中执行某个命令，即为一次执行（Invocation）。

命令执行状态

单条命令执行状态

API 状态	状态显示	描述
Pending	下发中	系统正在校验或发送命令。
Running	执行中	命令正在实例上运行。
Success	执行成功	执行成功，退出码为 0。
Failed	执行失败	执行完成，退出码非 0。

Openapi

欢迎使用云助手 openapi，统一的参数风格和公共错误码，给您带来简单快捷的使用体验。

Openapi 概览

接口名称	接口功能	频率限制(次/秒)
run-command	执行命令	单用户 20 次/秒
describe-invocation-results	查询命令执行结果	单用户 20 次/秒
create-command	创建命令	单用户 20 次/秒
invoke-command	触发命令	单用户 20 次/秒
get-commands	查询已创建的或公共的命令	单用户 20 次/秒

接口名称	接口功能	频率限制(次/秒)
get-command	查询命令详情	单用户 20 次/秒
modify-command	修改命令	单用户 20 次/秒
delete-command	删除命令	单用户 20 次/秒
get-ca-agent	查询实例中是否安装了云助手 agent	单用户 60 次/秒
send-file	上转文件到实例	单用户 20 次/秒
describe-send-file-results	查询文件上传结果	单用户 20 次/秒

常见问题

参考　云助手常见问题　。

7.14.1.2 安装云助手客户端

背景信息

- 使用公共镜像创建的云主机实例，默认预装了云助手 Agent。
- 2024 年 6 月 30 日以后使用公共镜像创建的物理机服务器已默认预装云助手 Agent。如果您的实例是 2024 年 6 月 30 日之前购买的或使用自行上传的自定义镜像创建的物理机实例，若需要使用云助手相关功能，请参考本文自行安装云助手 Agent。

操作场景

本文为您介绍如何在弹性云主机、物理机服务器实例上安装云助手客户端。

如何查看云助手客户端状态？

- Linux 云主机：

查看是否安装了 QGA：执行 ps -ef | grep qemu-ga 查看进程是否存在，或者 systemctl status qemu-guest-agent 服务为 active 。

- Windows 云主机：

查看是否安装了 QGA：打开任务管理器，选择服务，查看 QEMU Guest Agent 服务是否正在运行。

- Linux 物理机：

查看是否安装了 ca-agent：执行 ps -ef | grep ca-agent 查看进程是否存在，或者 systemctl status ca-agent 服务为 active 。

在弹性云主机实例上安装云助手 Agent

Linux 云主机安装 QGA，详情请参见： 安装 QEMU-Guest-Agent 。

Windows 云主机安装 QGA，详情请参见： 安装 QEMU-Guest-Agent 。

在物理机器实例上安装云助手 Agent

目前仅支持 Linux 系统的物理机服务器安装云助手 Agent，并提供相应的安装脚本供用户安装使用。

Linux 系统的物理机实例安装云助手 Agent

1. 以管理员身份登录 Linux 物理机服务器。
2. 下载云助手 Agent 安装包：ca-agent-install.tar.gz 。

```
curl http://169.254.169.254:20300/api/ebm/caAgent/latest/download -o ca-agent-install.tar.gz
```

3. 解压缩安装包，该压缩包内包含安装脚本 install_linux.sh、最新版云助手 ca-agent 以及云助手服务相关配置文件 。

```
tar -zxvf ca-agent-install.tar.gzcd ca-agent-install;bash -x install_linux.sh
```

4. 执行命令，查看 ca-agent 服务状态为 active 。

```
systemctl status ca-agent.service
```

7.14.1.3 购买指南

云助手功能免费，但是执行命令依赖于一台在线的服务器，需要您账号下有可用的弹性云主机或物理机实例。如您账号下没有可用实例，需要您新购一个实例。实例费用请参见 弹性云主机计费项及其计费方式 及 物理机计费说明 。

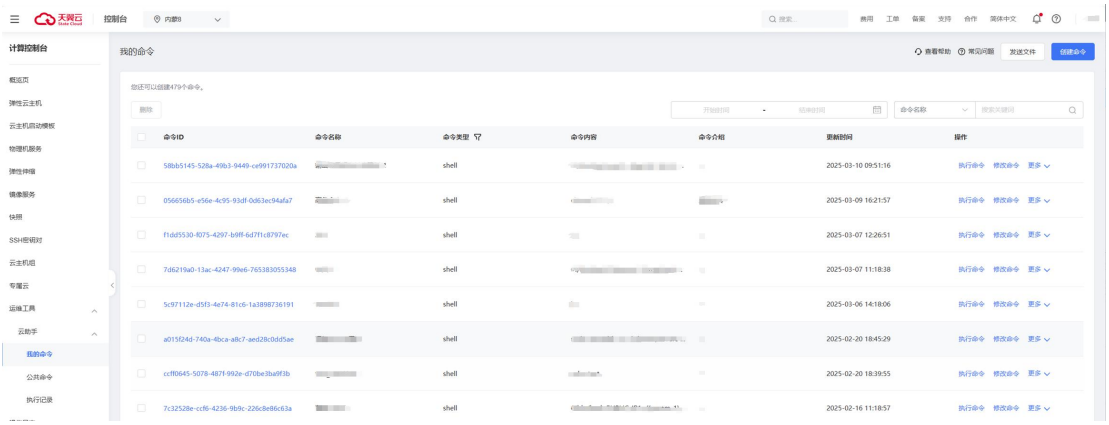
7.14.1.4 创建命令

使用限制

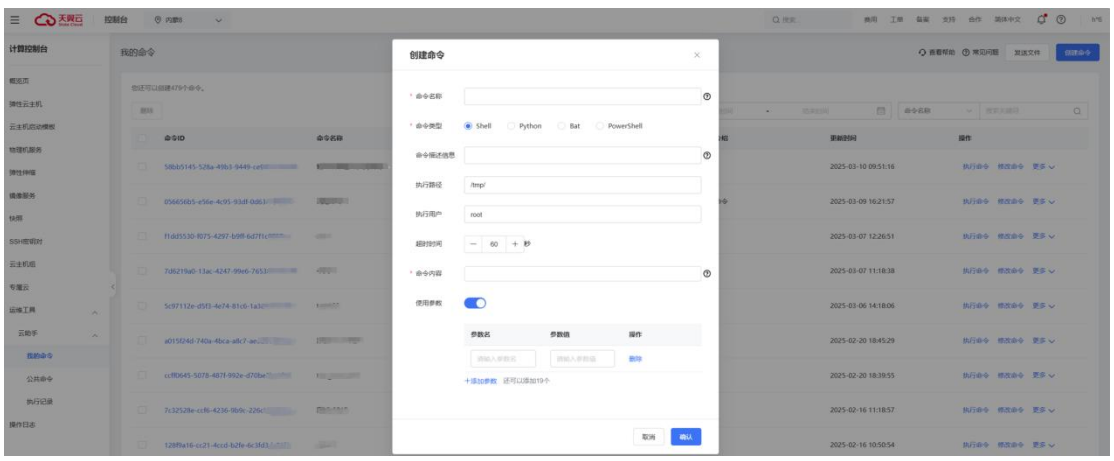
- 在同一地域下，您可以保存 500 条云助手命令。
- 创建的脚本大小不能超过 24KB。

操作步骤

1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维工具。
2. 展开运维工具下云助手标签页，选择我的命令标签页，如下图所示：



3. 点击创建命令按钮，如下图所示：



4. 填写命令名称、命令内容、选择命令类型、填写命令参数等。

5. 填写完成后点击保存按钮保存命令。

参数说明

参数	说明	是否必填
命令名称	设置命令名称，长度不超过 128 个字符。	是
命令类型	按需选择命令类型。	是
命令描述信息	按需填写命令描述信息，长度不超过 512 个字符。	否
执行路径	自定义命令的执行路径。 Linux 实例默认路径为 root 用户的 /tmp/ 目录。 Windows 实例默认路径为 System 权限的 C:/Windows/System32/ 目录。	是
执行用户	自定义执行命令的用户。 Linux 实例默认为 root 用户。 Windows 实例默认为 System 用户。	是
超时时间	设置命令在实例中的超时时间，当执行命令的任务超时后，云助手将强制终止任务进程。单位为秒，默认为 60 秒，取值范围为[10, 86400]。	是
命令内容	编辑或者粘贴您的命令。	是
使用参数	是否在命令中使用参数，自定义参数不能超过 20 个。	否

7.14.1.5 免登录执行命令

操作场景

本文介绍如何在不登录云主机或物理机实例时，执行云助手命令。

前提条件

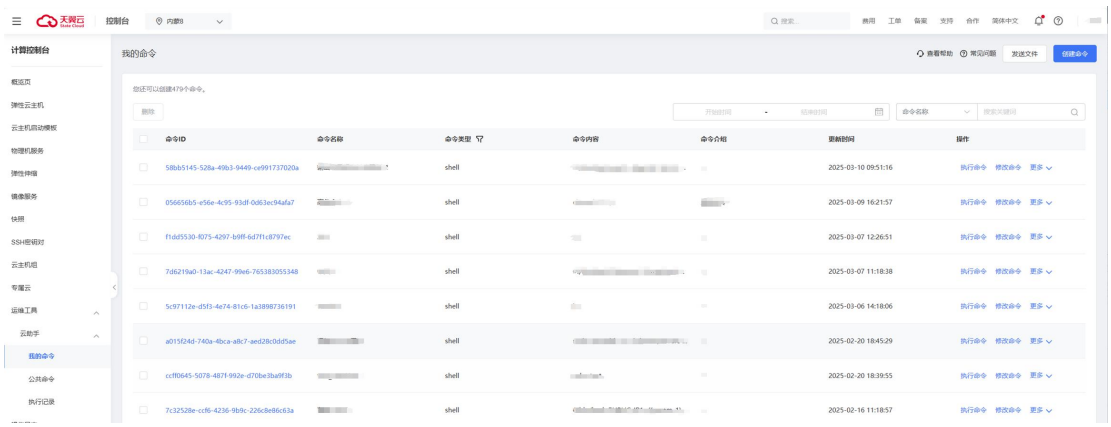
- 实例状态需处于运行中 。
- 目标实例已安装云助手客户端，且云助手客户端状态处于运行中。

背景信息

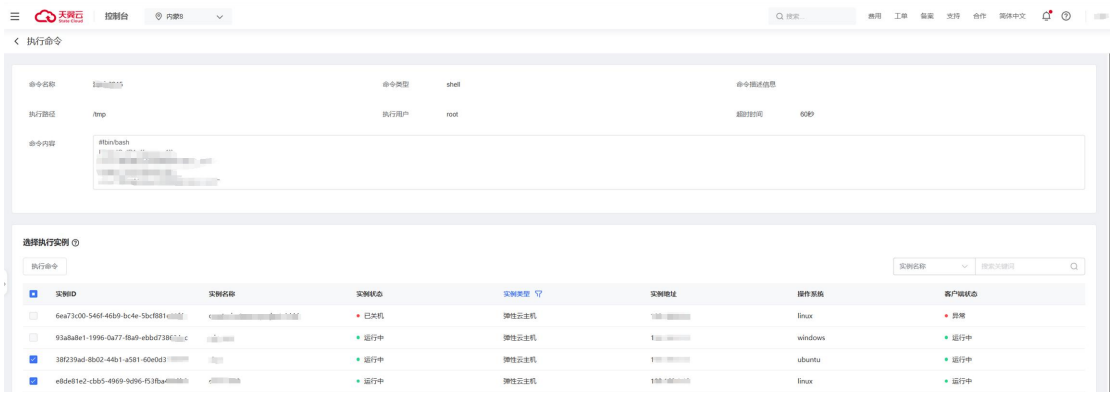
调用 API 执行命令时最多可以选择 100 台实例。

操作步骤

1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维工具。
2. 展开运维工具下云助手标签页，选择我的命令标签页，如下图所示：



3. 选择需要执行的指令，点击执行命令按钮。
4. 选择需要执行命令的实例，点击执行命令即可实现免登录执行命令，如下图所示：



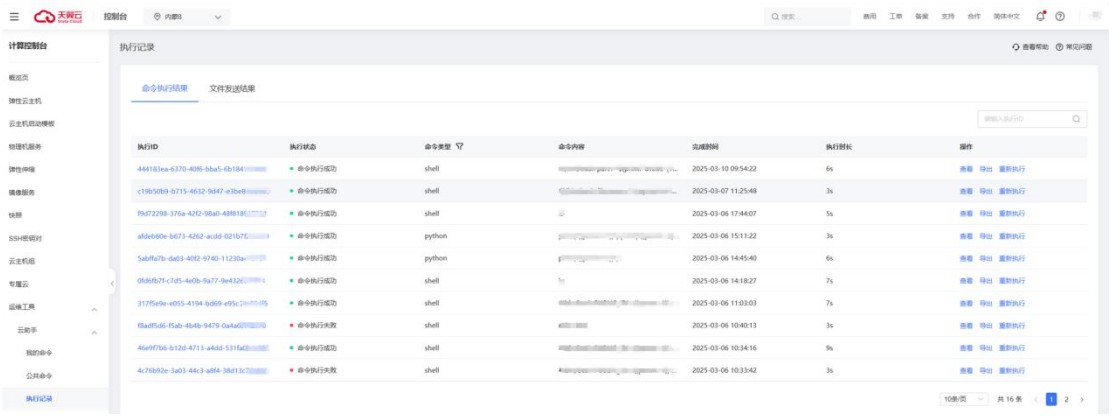
7.14.1.6 查看命令执行结果

背景信息

当出现实例缺乏相关依赖、网络异常、命令语义错误、脚本调试失败或者实例状态异常等情况时，命令会呈现不同的执行状态与执行结果。您可以通过控制台或 API 查看执行结果中的错误信息、诊断并修复问题。

操作步骤

1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维工具。
2. 展开运维工具下云助手标签页，选择执行记录标签页，如下图所示：



3. 选择需要查看的命令执行结果，点击右侧查看按钮，可查看执行列表、执行信息，如下图所示：



4. 点击实例名称左侧的箭头可以查看命令执行的输出信息或命令执行失败的报错信息。

7.14.1.7 查看和执行公共命令

背景信息

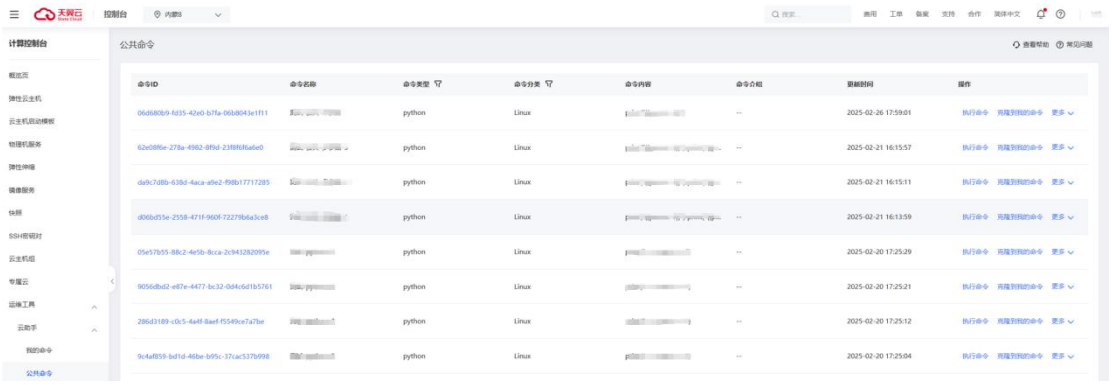
公共命令是天翼云创建的云助手命令，对所有天翼云用户可见。通常包含一些比较复杂的服务器配置、健康或安全检测、文件处理、系统补丁安装、更改系统配置的脚本。

相比较普通命令，公共命令的内容、发布以及升级会由天翼云统一负责维护。公共命令发布后，您可以直接查看命令的详细内容，并可以在弹性云主机或物理机实例上执行命令及查看执行进度和结果。使用公共命令，可以快速地完成某些复杂配置，很大程度提升您的操作和运维效率。

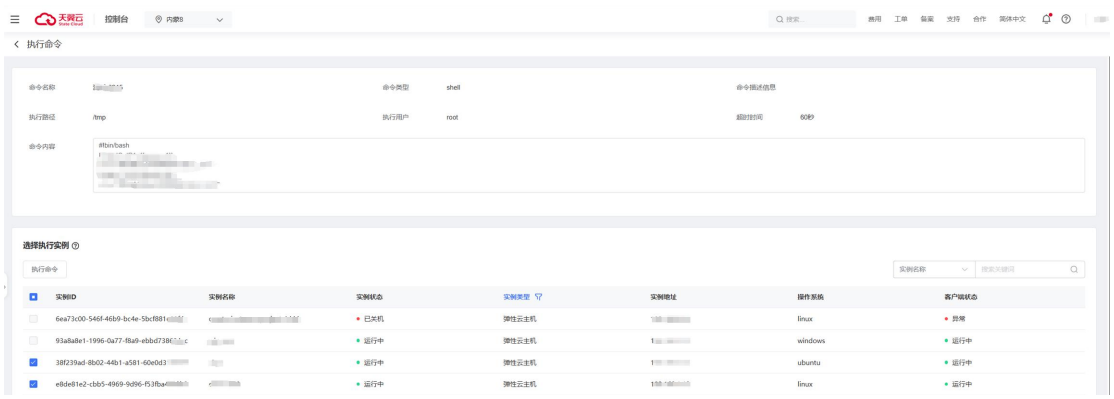
备注：天翼云创建的公共命令按需陆续更新。

操作步骤

1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维工具。
2. 展开运维工具下云助手标签页，选择公共命令标签页，如下图所示：



3. 可以根据命令名称了解命令的主要功能，选择想要查看的公共命令，点击右侧更多->查看详情 可以查看命令的详细信息。
4. 选择需要执行的公共命令，点击执行命令。
5. 选择需要执行命令的实例，点击执行命令即可实现免登录执行命令，如下图所示：



7.14.1.8 上传文件

前提条件

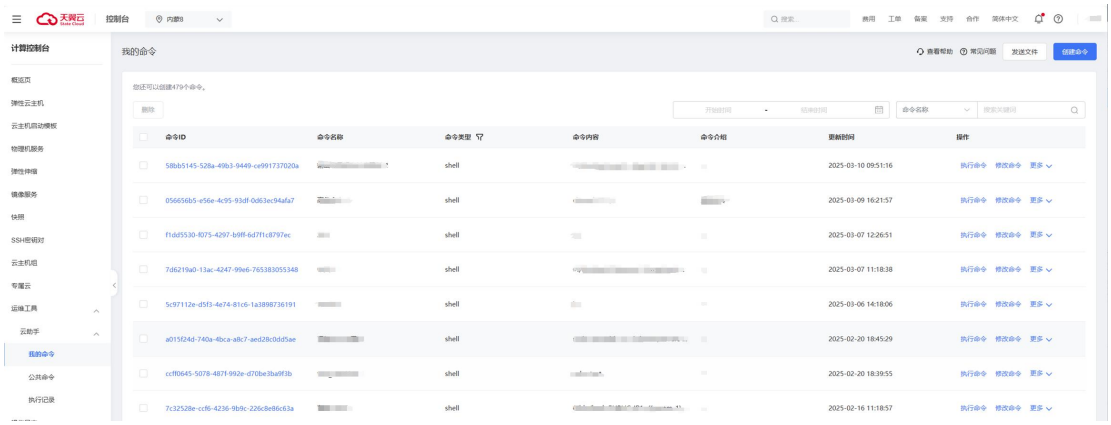
- 实例状态需处于运行中 。
- 目标实例已安装云助手客户端，且云助手客户端状态处于运行中。
- 待上传的文件不能超过 24 KB 且 base64 编码长度不可超过 32 KB。
- 目前仅支持： .py .pl .go .sh .crt .json .sql .der .pem .txt 格式的文件上传。

背景信息

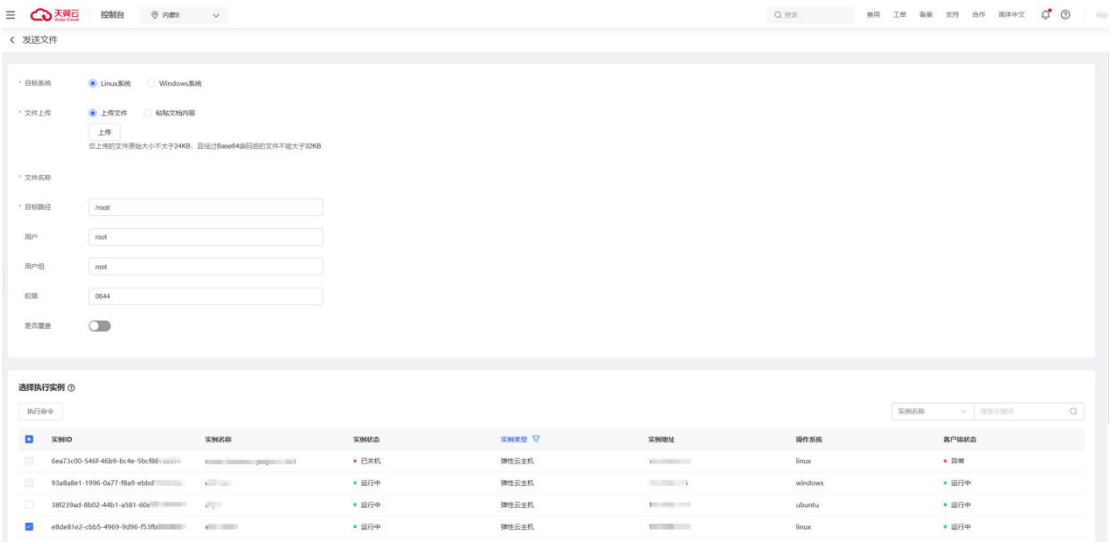
调用 API 执行命令时最多可以选择 100 台实例。

操作步骤

1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维工具。
2. 展开运维工具下云助手标签页，选择我的命令标签页，如下图所示：



3. 点击发送文件按钮，如下图所示：



4. 选择目标系统后，点击上传按钮选择要上传的文件。
5. 填写目标路径、用户、用户组、权限等。
6. 填写完成后选择要上传的目标实例，点击执行命令。

参数说明

参数	说明	是否必填
文件名称	设置文件名称，长度不超过 128 个字符。 如果参数是否覆盖未开启，文件名称不能和实例目标路径下已有文件名称相同。	是
目标路径	文件在实例上的保存路径，必须是已存在的绝对路径。默认情况下，目标路径如下： Linux：/root 。 Windows： C:/Users/Administrator/Documents 。	是
用	设置文件所属用户。	否

参数	说明	是否必填
户	说明： 仅 Linux 实例有此参数。	
用户组	设置文件所属用户组。 说明： 仅 Linux 实例有此参数。	否
权限	设置文件权限，请用 chmod 命令。 默认值为 0644，该值表示：文件所有者有读取、写入的权限；同组用户和公共用户具有读取的权限。 说明： 仅 Linux 实例有此参数。	否
是否覆盖	设置是否覆盖同路径下的同名文件。 默认不覆盖。	否

7.14.1.9 查看文件上传结果

背景信息

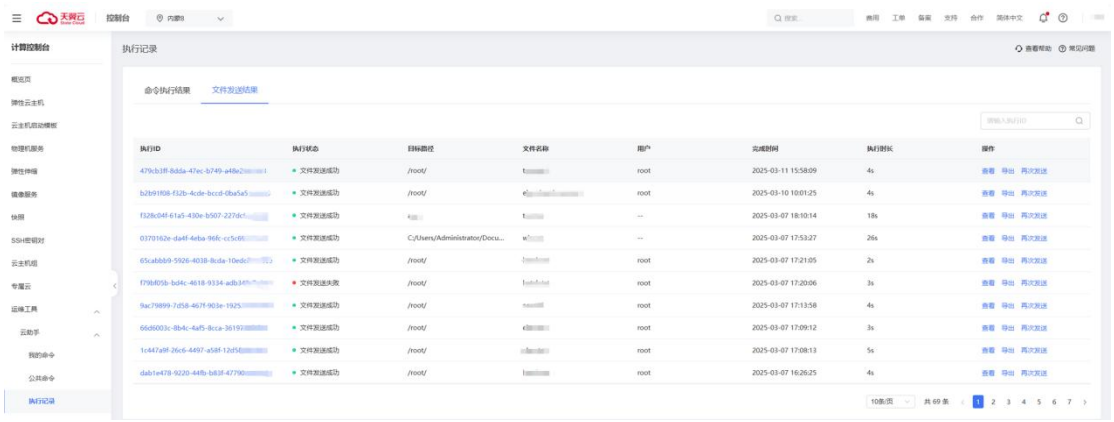
当出现实例网络异常或者实例状态异常等情况时，文件上传会呈现不同的执行状态与执行结果。您可以通过控制台或 API 查看执行结果中的错误信息、诊断并修复问题。

操作步骤

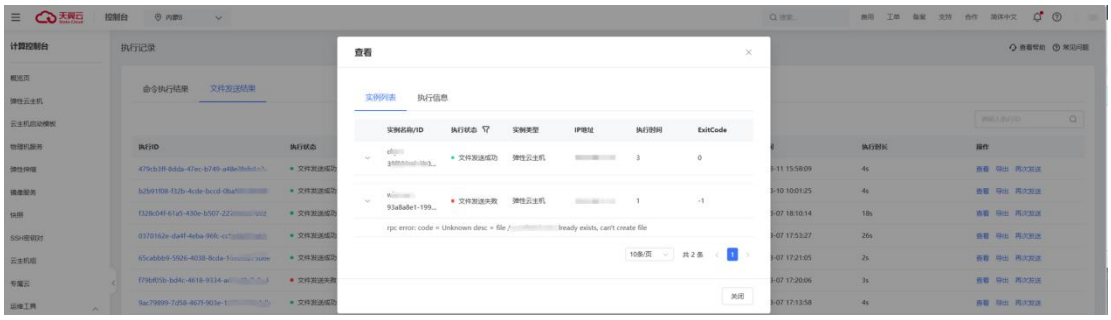
1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维

工具。

2. 展开运维工具下云助手标签页，选择执行记录标签页，如下图所示：



3. 选择需要查看的文件上传结果，点击右侧查看按钮，可查看执行列表、执行信息，如下图所示：



4. 点击实例名称左侧的箭头可以查看文件上传的输出信息或上传失败的报错信息。

7.15 资源与标签

7.15.1 标签管理

7.15.1.1 标签概述

应用场景

当项目中云资源较多，或当前资源信息不足以有效地为资源分组归类时，可以通过为资源添加不同维度（例如所属业务、部署环境、操作系统、开发团队等）的标签，实现资源分类。

基于标签筛选功能，您还可以快速查找一组资源，进行批量管理。

例如：

在团队管理中，为资源添加团队标签（如 department：R&D），标识所属开发部门，方便企业快速定位云资源所属部门。

在项目管理中，批量为业务 1 中所有 ECS 绑定业务标签 business：service1，通过 ESC 标签筛选，快速完成对业务 1 中云主机运行情况的日常检查。

使用说明

- 每个资源最多可绑定 10 个标签。
- 每个资源下的标签键是唯一的，不可绑定相同标签键。
- 每个资源下的标签键对应的标签值唯一，如修改已有标签键对应的标签值，新标签值将会覆盖旧标签值。例如，将云主机实例的 business：service1 的标签值 service1，修改为 service2，则新标签 business：service2 将覆盖旧标签 business：service1。
- 标签键值命名规则：

字段	规则
标签键	● 不能为空 ● 首字符不能为空格 ● 长度不超过 32 个字符
标签值	● 不能为空 ● 首字符不能为空格 ● 长度不超过 32 个字符

7.15.1.2 标签设计原则及示例

标签设计原则概述

当您账户中的云资源不断增多，管理云资源的难度也会随之加大，使用标签功能对账户内云资源的快速分组与管理，可以帮助您从不同维度对具有相同特征的云资源进行筛选，让资源管理变得更加轻松。在创建标签时，遵循以下原则会让您的管理更加高效：

全面原则 ：所有资源都得需要绑标签。从整体出发，考虑标签的组织结构，在规划资源的同时进行标签的键值规划，确保所有资源都可基于标签被分组。注意：标签键值建议采用标准格式。

有效管理幅度原则 ：每个标签键所对应的标签值数量应该是可控的，管理幅度不应过大，避免创建多余的标签值。

精简高效原则 ：精简标签键的设计，去除没有实际管理意义的标签键，防止因为标签键设计混乱导致的管理问题。

互斥原则 ：确保标签键值的含义的唯一性，避免设计含义相同或近似的标签键值。

标签键设计示例

业务维度	类型	子类型
组织架构	department group team organization	组织名称
角色层级	role user	相关角色名称

业务维度	类型	子类型
项目/任务	project task	项目/任务名称
环境	evironment	环境名称
使用者	owner	使用者名称

7.15.1.3 绑定标签

前提条件

已登录弹性云主机控制台。

创建弹性云主机完成，弹性云主机运行状况良好。

操作步骤

单个资源绑定标签

1. 登录管理控制台。
2. 在控制台顶部菜单左侧，选择区域。
3. 选择 “计算 > 弹性云主机”。
4. 点击列表操作栏中的 ″更多″-“编辑标签”，弹出“编辑标签”弹窗。
5. 在“编辑标签”弹窗中，按照使用限制与配置参数说明要求填写/选择标签键值，新增标签键值对。新增完成后，弹性云主机中计算资源将绑定该新增标签。您还可以修改标签键值对，执行编辑标签操作。
6. 点击确定，绑定标签完成。

批量绑定标签

1. 登录管理控制台。

2. 在控制台顶部菜单左侧，选择区域。
3. 选择 “计算 > 弹性云主机”。
4. 勾选需要批量绑定标签的弹性云主机，点击左上角的"更多"下拉菜单-“批量绑定标签”，弹出“批量绑定标签”弹窗。
5. 在“批量绑定标签”弹窗中，展示了所选弹性云主机，您可以按照使用限制与配置参数说明要求填写/选择标签键值，新增标签键值对。新增完成后，所有被勾选的弹性云主机中计算资源都将绑定该新增标签。
6. 您还可以在“批量绑定标签”弹窗中，修改标签键值对，执行编辑标签操作。
7. 点击确定，绑定标签完成。

使用限制与配置参数说明

- 每个资源最多可绑定 10 个标签。
- 每个资源下的标签键是唯一的，不可绑定相同标签键。
- 每个资源下的标签键对应的标签值唯一，如修改已有标签键对应的标签值，新标签值将会覆盖旧标签值，且该过程不可逆。例如，将云主机实例的 business: service1 的标签值 service1，修改为 service2，则新标签 business: service2 将覆盖旧标签 business: service1。
- 配置参数说明如下：

字段	字段说明	校验规则
标签键	输入新标签键或选择原有标签键。 选择框支持模糊搜索，下拉框支持展示所有标签。输入标签键则代表创建标签，选择原有标签键则代表修改现有标签。	不能为空。 首字符不能为空格。 长度不超过 32 个字符。

字段	字段说明	校验规则
标签值	输入新标签值或不修改原有标签值。 选择框支持模糊搜索，下拉框支持展示所有标签。 输入标签值则代表修改标签。	不能为空。 首字符不能为空格。 长度不超过 32 个字符。

执行结果

标签绑定成功后，可以在弹性云主机列表中的 “标签” 栏，鼠标点击查询标签绑定情况。

7.15.1.4 使用标签筛选资源

前提条件

已登录云主机控制台。

操作步骤

1. 登录管理控制台。
2. 在控制台顶部菜单左侧，选择区域。
3. 选择 “计算 > 弹性云主机”。
4. 在弹性云主机列表右上方，点击 “标签筛选” 按钮，弹出 “标签筛选” 弹窗。
5. 在“标签筛选”弹窗中，按照使用限制与配置参数说明要求，输入目标标签键值，新增筛选标签。
6. 点击确定，执行筛选标签操作，列表将展示标签筛选结果。

使用限制与配置参数说明

- 最多可筛选 10 个标签，筛选结果将展示所有包含筛选标签项的资源。
- 筛选的标签键唯一，不可筛选相同标签键。
- 标签筛选和其他筛选搜索操作不可并存。若列表存在筛选、搜索、排序条件时添加标签筛选条件，则清空当前筛选、搜索、排序条件；若列表存在筛选标签时添加筛选、搜索、排序条件，则清空当前筛选标签条件。

执行结果

弹性云主机按照标签筛选成功后，云主机列表将展示所有包含筛选标签项的资源。

7.15.1.5 解绑标签

前提条件

- 已登录到控制台
- 资源已绑定标签

操作步骤

解绑单个资源的标签

1. 登录管理控制台。
2. 选择目标产品与地域。
3. 在控制台列表中选择目标资源，在操作栏"更多"下拉菜单中找到“编辑标签”按钮，点击后弹出编辑标签弹窗。
4. 在弹窗中找到想要解绑的标签项，点击右侧的删除按钮。
5. 点击弹窗下方确定按钮，完成解绑操作。

批量解绑标签

1. 登录管理控制台。
2. 选择目标产品与地域。
3. 在控制台列表中，点击左侧复选框勾选目标资源，在列表上方"更多"下拉菜单中找到“批量解绑标签”按钮，点击后弹出弹窗。
4. 在弹窗中找到想要为已选资源解绑的标签项，勾选左侧复选框。

5. 点击弹窗下方确定按钮，完成批量解绑操作。

执行结果

执行成功后，会将全部已选择资源在弹窗中勾选的标签项解绑，并弹出提示语“操作成功！”，表示解绑标签成功。

7.15.2 云主机服务委托

前提条件

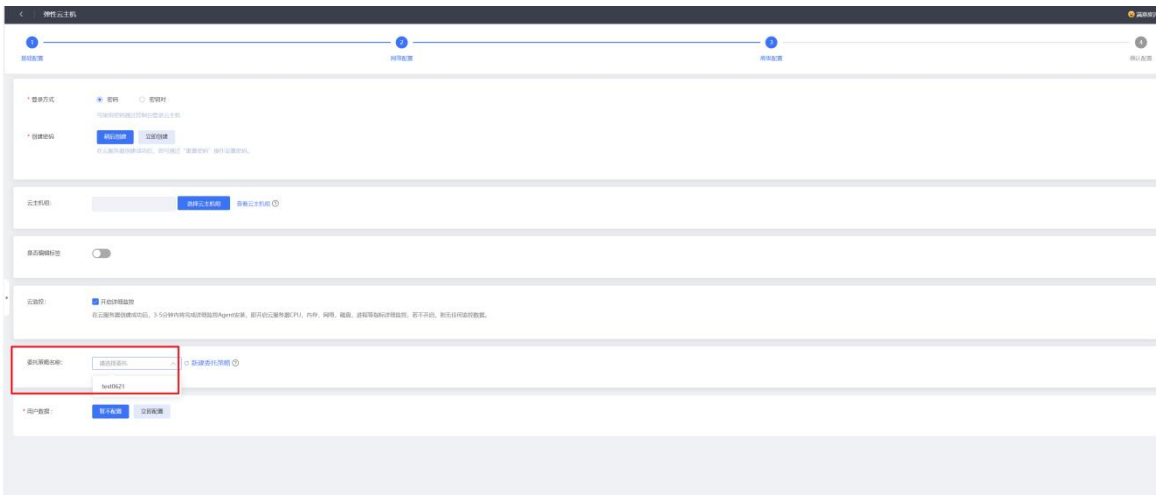
已登录弹性云主机控制台。

已有可使用的委托策略。

操作步骤

新建云主机时绑定委托策略

1. 登录管理控制台。
2. 在控制台顶部菜单左侧，选择区域。
3. 点击计算-弹性云主机进入创建云主机页面。
4. 单击右上角创建云主机进入云主机购买页面。
5. 在步骤 3“高级配置“页面”，选择“委托策略名称”，此条策略后续将生效于该台云主机。
6. 云主机创建其他信息填写完毕，点击“确认配置”以及“立刻购买”。

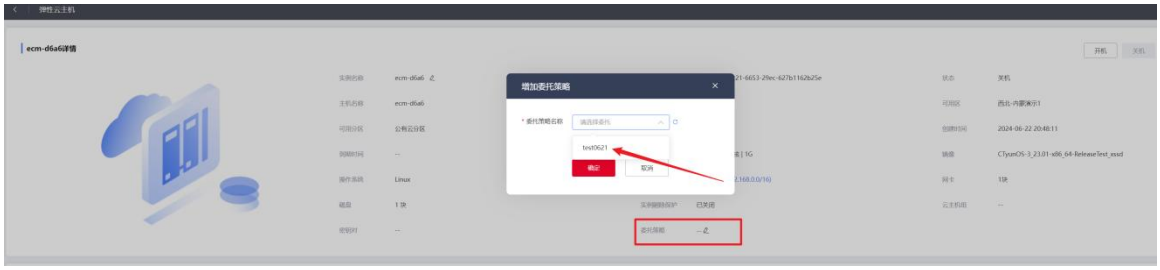


注意

委托策略的创建需要管理员登录 IAM 控制台进行创建。具体操作文档请参考统一身份认证--创建委托。

云主机创建后绑定委托策略

1. 点击指定弹性云主机的名称，系统跳转至该弹性云主机详情页面。
2. 详情页面中点击委托策略后方的编辑标识，为当前云主机绑定新委托策略。



注意

若您的云主机上已绑定委托策略，绑定新的委托策略后会原策略会失效，新策略将生效于当前云主机。

7.15.3 配额调整

配额基本介绍

配额是对有限资源的一种管理和分配，天翼云对您可以使用的资源数量，进行了配额的控制。

如何查看弹性云主机配额

进入弹性云主机页面，页面顶部展示您的基本配额信息，如需了解更多配额，可点击“了解配额详情”。



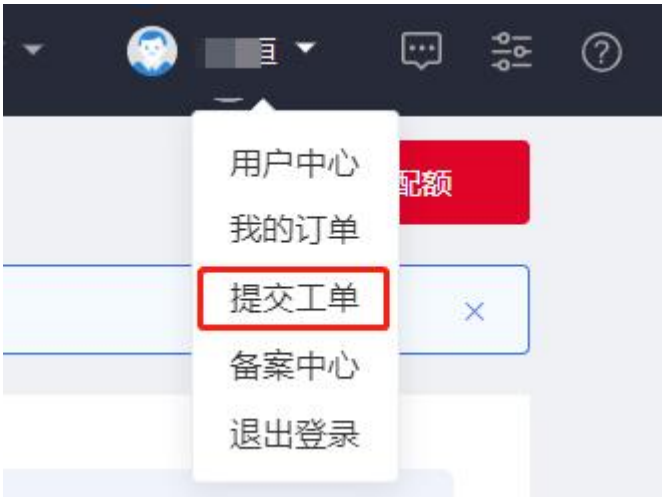
配额详情页面您可以阅读到主机和其他产品的配额信息，如图：

服务	资源	已用配额	总配额
弹性云主机	云主机(台)	0	50
	vCPU(核)	0	200
	内存(GB)	0	409600
	快照	0	30
	SSH密钥对	0	100
	云主机组可添加数（台）	--	16

如何提升配额

配额仅仅是为了防止资源滥用，如果有需要，可以提工单申请扩大配额。

点击用户头像-提交工单，进入工单管理页面：



点击“配额类”板块中的“提问”：



点击“配额申请”按钮，进入创建工单页面：



输入您的工单标题“xx 配额申请”，简要描述您的申请理由、申请资源池、联系方式等信息并提交：



提交后客服将很快与您取得联系，确认信息并帮您申请配额。

配额基本介绍

配额是对有限资源的一种管理和分配，天翼云对您可以使用的资源数量，进行了配额的控制。

如何查看弹性云主机配额

进入弹性云主机页面，页面顶部展示您的基本配额信息，如需了解更多配额，可点击“了解配额详情”。



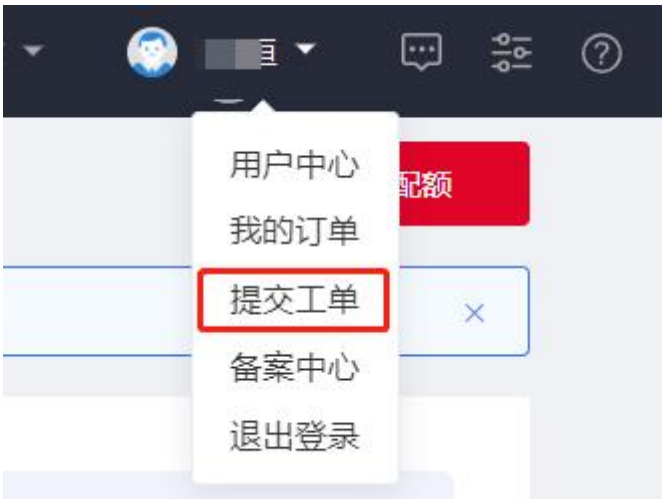
配额详情页面您可以阅读到主机和其他产品的配额信息，如图：

服务	资源	已用配额	总配额
弹性云主机	云主机(台)	0	50
	vCPU(核)	0	200
	内存(GB)	0	409600
	快照	0	30
	SSH密钥对	0	100
	云主机组可添加数（台）	--	16

如何提升配额

配额仅仅是为了防止资源滥用，如果有需要，可以提工单申请扩大配额。

点击用户头像-提交工单，进入工单管理页面：



点击“配额类”板块中的“提问”：



点击“配额申请”按钮，进入创建工单页面：



输入您的工单标题“xx 配额申请”，简要描述您的申请理由、申请资源池、联系方式等信息并提交：



提交后客服将很快与您取得联系，确认信息并帮您申请配额。

8 常见问题

8.1 高频常见问题

远程登录

- 无法远程登录 windows 云主机怎么办？
- 无法远程登录 Linux 云主机怎么办？
- 远程登录时需要输入的帐号和密码是多少？
- Windows 云服务器如何配置多用户登录？（Windows 2012）

云主机故障/卡顿

- Windows 云主机卡顿怎么办？

- Linux 云主机卡顿怎么办？
- 如何排查带宽超过限制？

网站或应用无法访问

- 弹性云主机启动缓慢怎么办？
- 云主机端口不通怎样排查？
- 弹性云主机访问中国大陆外网站时加载缓慢怎么办？

无法访问外网

- Windows 云主机无法访问外网怎么办？
- Linux 云主机无法访问外网怎么办？
- 未绑定弹性公网 IP 的弹性云主机能否访问外网？
- 弹性云主机访问中国大陆外网站时加载缓慢怎么办？
- 云主机网络延迟和丢包，如何定位？

密码与密钥对

- 远程登录时需要输入的帐号和密码是多少？
- 怎样在控制台重置弹性云主机密码？

Ping 不通

- 为什么弹性云主机可以远程连接，但是无法 ping 通？
- ping 不通或丢包时如何进行链路测试？

8.2 地域与可用区类

8.2.1 如何选择地域？

选择地域时，建议考虑以下几个因素：

1. 地理位置：用户和资源部署地域的距离越近，网络时延越低，访问速度越快。
   - 中国内地：一般情况下建议选择和您目标用户所在地域最为接近的数据中心，可以进一步提升用户访问速度。如果您使用天翼云承载全部业务，电信网络可以保证中国内地地域间的快速访问。
   - 其他国家及地区：其他国家及地区提供的带宽主要面向非中国内地的用户。如果您在中国内地，使用这些地域会有较长的访问延迟，不建议使用。
2. 资源价格及资源覆盖：不同地域的资源价格、产品覆盖可能有差异，请根据您的需求及预算选择合适的地域。
3. 产品之间的关系：如果购买了多款天翼云产品搭配使用，需要注意不同地域的云主机、对象存储、负载均衡等服务内网不互通。
4. 经营性备案：如果使用弹性云主机实例作为 Web 主机，您需要完成经营性备案，同时需要在指定的地域购买实例。（各省或市通信管理局对经营性备案的审批要求不同，请以当地管理局经营性备案网站公示内容为准。）

8.2.2 地域之间是否存在差异？

云计算产品地域之间确实存在差异，主要体现在产品覆盖度、产品价格、网络延迟性等方面。

1. 产品覆盖度：由于市场需求与资源差异等因素，不同地域能提供的实例规格不全相同。
2. 产品价格：受成本、供需、政策等因素影响，同规格实例在不同地域可能存在价格差异。
3. 网络延迟性：实例所处地域与用户服务所处地域之间的距离差异，会导致数据在传输过程中有不同程度的网络延迟，数据从发送方到接收方所需时间也会不同。一般来说，实例距用户越近，网络延迟越小，数据传输速度和用户体验越好。

综上几点差异，请根据您的服务需求及预算选择合适的地域。

8.2.3 已购买的弹性云主机，是否支持更换地域？

已购买的弹性云主机不支持更换地域。

如果您需要在另一个地域使用弹性云主机，有以下两种方案：

1）选择目标地域重新购买实例，建议选择与您的业务需求最接近的区域，以减少网络延迟和数据传输时间。

2）采用主机迁移服务实现实例间的快速迁移，具体操作可以参考 天翼云 ECS 实例间迁移最佳实践 。

8.2.4 什么是可用区，怎样选择可用区？

可用区（AZ，Availability Zone）是指在同一地域（Region）内电力和网络互相独立的物理地域。一个 AZ 是一个或多个物理数据中心的集合，具备独立的风火水电，一个 Region 中的多个 AZ 间通过高速光纤相连，以满足用户跨 AZ 构建高可用性系统的需求。

选择可用区可参考以下标准：

在同一地域内，可用区与可用区之间内网互通。各可用区之间可实现故障隔离，如果一个可用区出现故障，不会影响其他可用区的正常运行。

是否将实例放在同一可用区，主要取决于应用对容灾能力和网络延时的要求。

- 如果您的应用需要较高的容灾能力，建议将实例部署在同一地域的不同可用区。
- 如果您的应用要求实例之间较低的网络延时，建议将实例创建在同一可用区。

8.2.5 可用区之间的数据传输是否需要收费？

同一地域的可用区内网互通，内网可实现数据传输，不需要额外购买网络产品，无额外花费。

不同地域可用区之间传输数据涉及购买弹性 IP 等产品，需要为额外购买的产品付费。

8.2.6 是否可以将应用的不同组件分散部署到不同的地域？

建议将应用的不同组件部署到同一个地域，各个组件间可以通过内网网络进行通信，节省用户带宽费用。

也可以将应用的不同组件部署到同一个地域的不同可用区，这样既能保证分布式的应用架构，也能保证组件间网络的通信量。

8.2.7 是否提供支持应用灾备的相关服务？

目前天翼云弹性云主机暂未为用户提供标准的应用灾备方案，如果您有具体需求，可以通过客服电话等方式联系产品售前经理或解决方案经理，我们会根据您的业务为您定制合适的解决方案。

天翼云多活容灾服务（MDR）即将上线，届时您可通过 MDR 平台部署容灾业务，敬请期待。

8.2.8 是否可以在不同的区域之间实施应用灾备？

天翼云多活容灾服务（MDR）即将上线，通过 MDR 平台，您可以将应用的主备节点部署到不同的地域，平时使用主节点提供服务，当主节点应用出现故障，备节点应用可以继续为您的客户提供服务。

此外，您也可以在部署时使用云主机组功能，通过反亲和性/强制反亲和性将云主机尽量/强制地创建在不同的宿主机上，提高业务的可靠性。

8.2.9 如何获取弹性云主机的物理机房位置？

您可以在天翼云官网通过“提交工单”的方式获取弹性云主机的物理机房位置。

具体操作如下：

1. 登录管理控制台，进入“新建工单”。
2. “产品类”选择“弹性云主机”。
3. 填写工单需求，如：获取 XX 云主机的物理机房位置，同时填写联系方式等，点击提交工单。
4. 等待工单工程师回复，获取具体信息。

8.3 购买与退订冻结相关问题

8.3.1 弹性云主机提供哪些计费方式，区别是什么？

提供包年包月、按量计费两种计费方式，详细介绍、适用场景、使用限制等可查看 弹性云主机-购买指南-计费模式 。

8.3.2 同一台云主机是否同时支持两种计费方式？

不可以。同一台云主机在同一时间只能选择一种计费方式。

8.3.3 是否支持更改已有云主机的计费方式？

支持。按量计费云主机可以转为包年包月计费，包年包月计费云主机到期后可以转为按量计费。

- 按量计费转换为包年包月计费：如果您需要长期使用当前弹性云主机，可以将按量计费的云主机转为包年包月计费，经济实惠节省开支。
- 包年包月计费转换为按量计费：如果您需要更灵活的计费方式，可以将包年包月计费的云主机转为按量计费。**注意：包年包月计费转为按量计费，需在包年包月计费资源到期后，按量计费方式才会生效。**

8.3.4 账户里面有余额，为什么无法创建按量付费云主机？

请检查您的账户金额是否小于 100 元，余额不足 100 元时，天翼云不允许订购按量资源，请您及时充值。

当您的账户余额不足 100 元时，会收到账户余额低于 100 元的提醒。**注意：此种情况不影响包周期资源的正常使用。**

8.3.5 弹性云主机关机后还会继续计费吗？

弹性云主机关机后是否计费根据计费方式不同，情况有所不同：

- 包年包月计费方式：包年/包月资源一次性付费，到期自动停止使用。
- 按量计费方式：按量计费的弹性云主机关机后相关资源仍继续计费。

8.3.6 一个账户最多可以创建多少按量付费的云主机？

账户默认配额可查看 弹性云主机-产品使用限制 。

在不超出配额的情况下，您可根据需求创建云主机，如果云主机创建需求超出默认配额，您可提交工单申请增加配额。

8.3.7 如何选择适合业务的云主机实例？

以搭建网站为例，您可以从以下几方面确认实际的业务需求，并参见 弹性云主机-产品规格 选择合适的云主机实例。

- 网站的类型。
- 网站的日均 PV。
- 网站的首页大小。
- 网站的数据容量。

在购买云主机之后，您也可以根据实际需求随时调整实例配置和数量。

8.3.8 购买实例时，资源已经售罄怎么办？

如果在创建实例时遇到资源售罄的情况，建议您采取以下措施：

- 更换地域
- 更换可用区
- 更换资源配置

如果调整需求后仍然没有资源，建议您等待一段时间再购买。实例资源是动态的，如果资源不足，天翼云会尽快补充资源，但是需要一定时间。

8.3.9 什么是 ARM CPU 架构与 X86 CPU 架构？

弹性云主机实例主要包含两种架构，X86 CPU 架构和 ARM CPU 架构。

- x86 CPU 架构

采用复杂指令集 CISC（Complex Instruction Set Computer），CISC 是一种计算机体系结构，其中每个指令可以执行一些较低阶的硬件操作，指令数目多而且复杂，每条指令的长度并不相同。由于指令执行较为复杂所以每条指令花费的时间较长。

- ARM CPU 架构

采用精简指令集 RISC（Reduced Instruction Set Computer），RISC 是一种微处理器，旨在执行较少类型计算机指令，以便能够以更高的速度执行操作，使计算机的结构更加简单、合理地提高运行速度。

ARM CPU 架构包含鲲鹏云主机和飞腾云主机。海光云主机属于 X86 CPU 架构。目前可以在央企北京 1 资源池购买鲲鹏、飞腾、海光云主机，可以在佛山 3 购买鲲鹏云主机。

8.3.10 控制台无法看到云主机，如何处理？

如果您在控制台上找不到已购买的实例，可能存在以下几种原因：

- 资源不在当前地域

如果您未切换到实例所在地域，会导致找不到实例。可以通过以下方法进行确认：在弹性云主机实例页上方切换地域。

- 资源到期或欠费被释放

对于欠费或到期的资源，在一定时间后会被彻底释放，因此在控制台找不到，被释放的资源是无法找回的。

- 资源不足导致退款

刚购买的实例在控制台没有找到，可能是因为资源不足而直接退款了。您可以在订单管理中确认退款信息。在您确认退款后，您可以更换地域或可用区尝试再次购买。

8.3.11 如何处理支付订单后云主机开通失败？

如果遇到支付订单后云主机开通失败的场景时，您可以致电客服电话400-810-9889，说明您的订单信息和问题情况，客服人员会帮助您检查订单状态，并尝试解决问题。

如果云主机开通失败是因为技术问题，客服人员会帮助您联系技术支持团队，以获取更详细的帮助与问题排查。

如果云主机开通失败是因为订单支付问题，客服人员可以帮助您解决支付问题，例如重新验证您的支付信息或与您协商退款。

8.3.12 申请开通弹性云主机需要多久？

弹性云主机的开通一般在分钟级别。

具体的开通时间受多方面因素影响，比如规格大小、关联资源（云硬盘、弹性 IP 等）、开通数量。如果开通较大的规格和较多的关联资源，开通时间可能会稍微长一些，如果当前可用区负载已经很高，开通时间可能也会受到影响。

如果一直无法开通成功或在遇到其它任何问题，您可以提交工单或致电客服电话 400-810-9889 寻求技术支持。客服人员会帮助您解决问题，确保您的弹性云主机能够顺利开通并投入使用。

8.3.13 什么情况下弹性云主机会被冻结，冻结后怎么办？

如果没有及时续费或充值，在欠费后将冻结您账号下所有的按量资源，同时会发送短信及邮件提醒您。

当云主机资源被冻结后，您可通过续费或充值来解冻资源，恢复云主机正常使用。

- 资源冻结时：资源将被限制访问和使用，会导致您的业务中断。
- 资源解冻时：资源将被解除限制，但是需要您自行检查并恢复业务。
- 资源释放时：资源将被释放，存储在资源中的数据将被删除，数据无法找回。

建议您尽快续费或充值来解冻资源，否则在冻结 15 天后，存储在云主机中的数据将被删除、云主机资源将被释放。

8.3.14 如何退订弹性云主机？

仅包年包月的弹性云主机支持退订，首先需将云主机关机，关机完成后单击操作栏“更多-退订”按钮，在弹出页面点击“确认”，完成退订。

注意

- 退订操作会回收资源，在退订前请您做好数据备份工作。
- 按量订购的云主机不支持退订操作。

8.3.15 已到期资源如果续订，续订周期如何计算？

若资源进入冻结状态后续费，所有资源续费周期的开始时间为解冻时间。例如，用户 A 于 2020 年 4 月 1 日订购包月云主机 1 个月，资源于 2020 年 4 月 30 日到期，用户 A 于 2020 年 5 月 10 日续订 1 个月，新的资源到期时间为 6 月 10 日。

为了您的业务不受影响，请在到期之前进行续订。

8.3.16 是否支持账户余额变动提醒？

支持。当您的账户余额发生变化时，系统会自动发送邮件和短信通知，以提醒您及时了解账户余额情况。

除了账户余额变动提醒外，天翼云还提供消费账单查看功能。您可以在费用中心控制台查看自己的消费账单，包括汇总账单、流水账单和账单详情。这些账单提供了详细的费用信息，帮助您更好地了解自己的云服务消费情况和成本情况。您还可以根据需要导出账单数据，以便进行更详细的分析和核对。具体操作可查看 费用中心-账单管理 。

8.3.17 弹性云主机怎样停止计费？

按量计费弹性云主机根据使用时间计费，与之绑定的云硬盘（包括系统盘、数据盘）、弹性公网 IP、带宽等资源按各自产品的计费方法进行计费。如果用户不再使用该产品，需彻底停止计费，请直接删除相应产品。

包年包月弹性云主机用户在购买时一次性付费，到期自动停止计费。若在到期日期之前想进行退订，当前天翼云目前支持 7 天无理由全额退订和非七天无理由退订以及其他退订，具体细则可参考 费用中心-退订管理 。

8.4 创建与删除

8.4.1 创建类

支付成功后为什么不能马上看到创建中的弹性云主机？

支付成功后可能需要少许时间生成订单，待订单完成后即可看到创建中的弹性云主机。若您等待一段时间后仍未在云主机列表中看到云主机，可先在“天翼云官网-我的-费用中心-订单管理-我的订单”中查看是否已生成订单。

注意
如果创建后长时间看不到云主机，您可以提交工单或致电客服电话 400-810-9889 寻求技术支持。客服人员会帮助您解决问题，确保您的弹性云主机能够顺利开通并投入使用。

如何处理支付订单后云主机开通失败？

如果您支付订单后云主机开通失败，系统会自动进行撤单，该云主机的订单状态会变为将“撤单中”，撤单完成后订单状态变为“撤单完成”，撤单完成订单费用退还至您的账户。您可在“天翼云官网-我的-费用中心-订单管理-我的订单”中查看订单详情及进度。

重启/关机弹性云主机时，长时间处于“正在重启”/“正在关机”状态，怎么办？

如果对云主机实例执行重启/关机操作后，云主机长时间处于“正在重启”/“正在关机”状态，其可能出现的情况包括：

- 云主机操作系统内存在无法自动关闭的进程，导致了操作过程无法继续进行。
- 出现系统层面错误，导致后台操作未能正常执行。

对于这些问题，我们建议您采取以下步骤：

1. 尝试手动重启或关机：您可以尝试通过远程连接到云主机，并手动执行重启或关机命令。这可能会帮助您解决问题，或者至少提供更多关于问题的信息。
2. 查看日志：您可以查看云主机的系统日志，以了解是否有任何与重启或关机操作相关的错误信息。这些信息可以帮助您确定问题的原因。
3. 提交工单：如果以上步骤无法解决问题，您可以提交工单或致电客服电话 400-810-9889 寻求技术支持。他们将能够进行更深入的调查，并提供适当的解决方案。

购买云主机时配置“反亲和性”云主机组，提示失败怎么办？

加入强制反亲和组的云主机，必须分散地创建在不同的宿主机上，当没有足够的宿主机资源，就会出现配置了反亲和组的云主机创建失败的情况。您可以考虑切换资源充足的区域购买，或等待当前区域资源充足后再重新购买。

8.4.2 删除类

已删除的弹性云主机可以再开通吗？

已经删除的云主机不能再次开通。所以请在执行删除操作前，确保将云主机上的数据已完成备份或者迁移，删除云主机后，数据无法找回，请谨慎操作。

为了确保您的数据安全，请在删除云主机之前按照确认以下操作：

1. 确认数据备份：您需要确保您的数据已经完整地备份到了另一个安全的位置，例如另一个云主机或已形成镜像或快照。
2. 确认数据迁移：如果您的数据需要迁移到另一个云主机或其他云产品上，您需要确保数据迁移已经完成并且数据已经成功同步。
3. 执行删除操作：一旦您确认数据已经备份或迁移并且您已经确认可以进行删除，您可以按照云主机控制台的指导执行删除操作。
4. 云主机备份与迁移可参考：备份云主机、迁移云主机。

如何删除弹性云主机？

登录天翼云弹性云主机控制台，选中需要删除的弹性云主机。

只有按量计费类型弹性云主机支持删除操作，弹性云主机只有在关机状态才可以删除，首先点击弹性云主机列表左上角的关机按钮。



等待关机完成，点击云主机实例最右侧的更多-删除





在弹出页面点击确定即可删除弹性云主机。

云主机如果存在快照，云主机可以删除吗？

云主机快照的恢复数据与申请云主机的功能依赖云主机的数据，如果云主机已经创建快照且快照没有删除，则不允许删除此云主机，控制台删除按钮为置灰状态。特别的，如果云主机绑定了快照策略但是无快照的情况下，云主机是可以删除的。

“删除”按钮是什么意思？

计费模式为按量计费的云主机可见“删除”按钮，“删除”云主机，即删除云主机的资源与数据，且无法找回，同时停止计费。

删除云主机的操作步骤如下。

1. 登录云主机控制台。
2. 选择按量计费的云主机，操作列点击【更多】。
3. 点击【删除】，并确认删除。

说明

包年包月计费模式的云主机，请于操作列表的【更多】中执行【退订】操作。

如何重启弹性云主机

前提：仅状态处于“运行中”的云主机可以执行重启操作。

1. 登录控制中心，选择云主机所在区域。

2. 点击【计算】-【弹性云主机】进入云主机控制台。

3. 选择需要重启的云主机实例，点击操作列的【更多】。

4. 点击【重启】，并确认重启。

8.5 登录与连接

8.5.1VNC 登录类

8.5.1.1VNC 方式登录后，播放音频文件没有声音怎么办？

使用 VNC 方式登录 Windows 云主机时，播放音频文件没有声音，是因为 VNC 方式不具备音频调用能力。

我们可以通过 MSTSC 方式登录弹性云主机，完成音频播放。

1.本地计算机输入“mstsc”，并单击“确定”。系统打开“远程桌面连接”窗口。



2.单击左下角的“选项”，并选择“本地资源”页签。



3.在“远程音频”栏，单击“设置”，配置远程音频设置。



4.在“远程音频播放”栏，选择“在此计算机上播放”。

8.5.1.2 如何修改 Windows 弹性云主机的分辨率

不同版本的 Windows 系统方法有所不同，这里以 Windows Server 2012 数据中心版为例。

VNC 登录方式

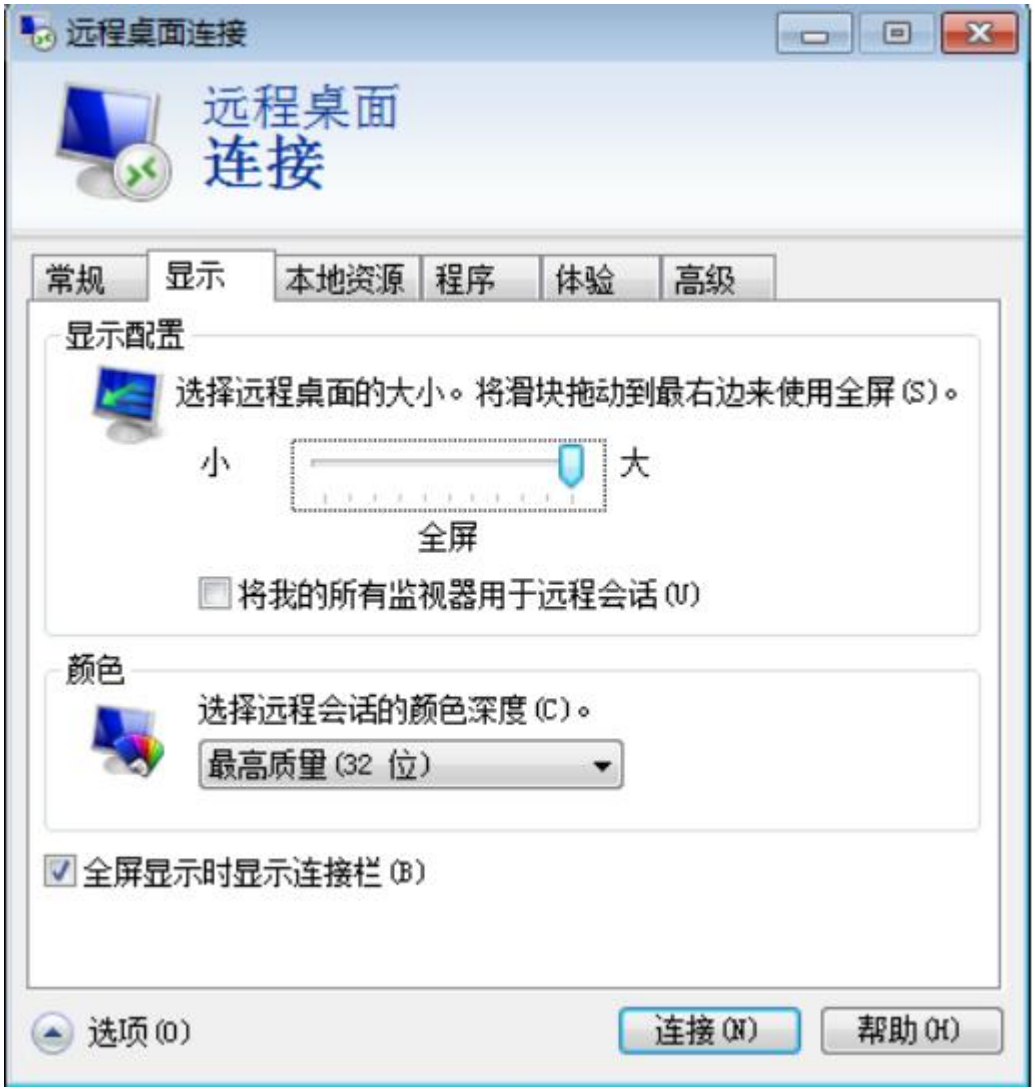
VNC 登录弹性云主机，右键单击桌面，选择屏幕分辨率，在弹出页面，选择合适的分辨率，点击确定、应用即可。



- MSTSC 方式

1. 本地极端及输入“mstsc”，并单击“确定”。系统打开“远程桌面连接”窗口。

2. 选择“显示”页签，在“显示配置”中设置分辨率大小。



3. 分辨率设置完成后，使用 MSTSC 方式连接弹性云主机。

8.5.1.3 怎样修改远程登录的端口？

修改 Windows 系统实例默认远程端口，以 Windows Server 2012 数据中心版为例。

1. 登录控制中心。

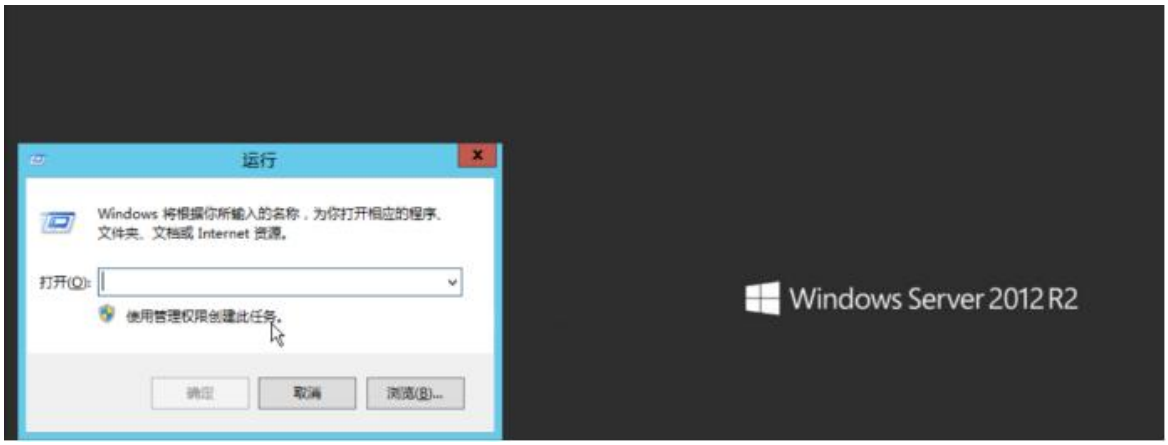
2. 选择地域 华东-华东 1。

3. 单击“计算>弹性云主机”，进入弹性云主机页面。

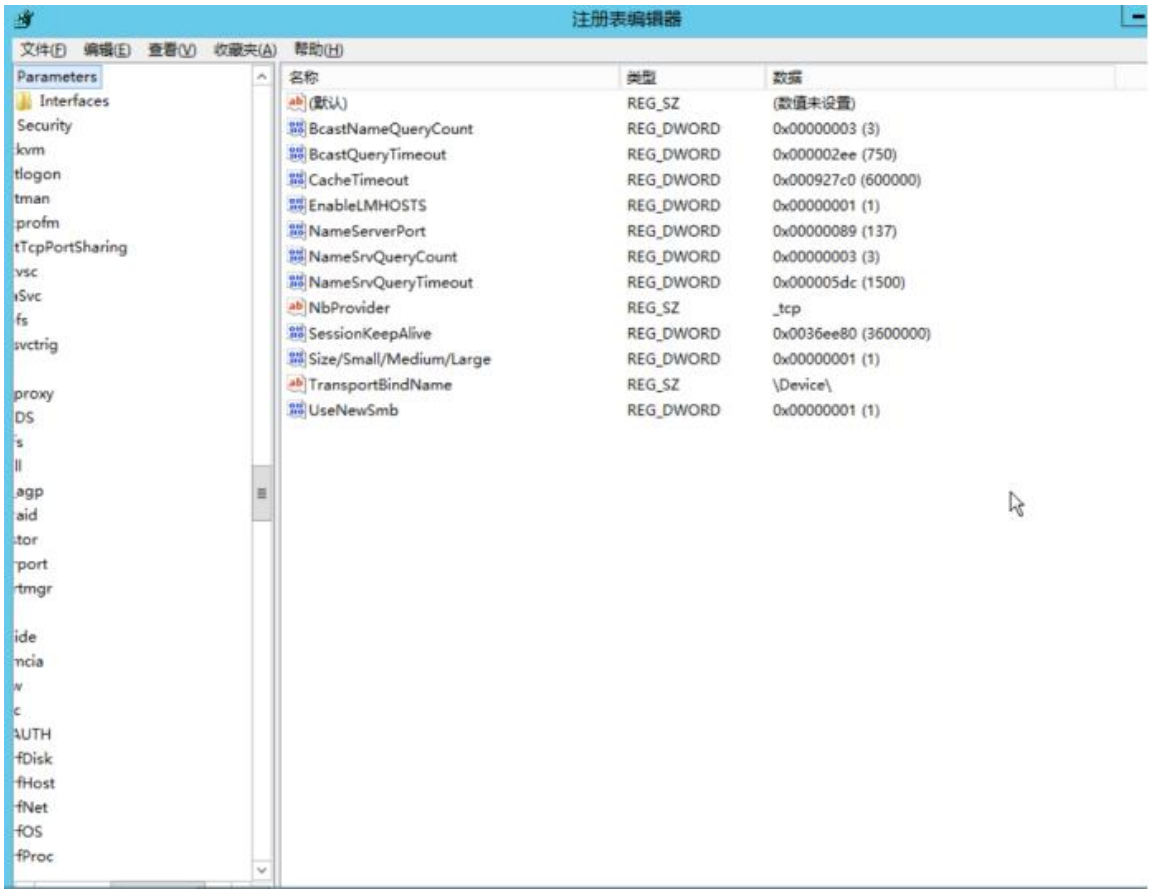
4. 参考 Windows 弹性云主机管理控制台远程登陆（VNC 方式），远程登录待修改端口的 Windows 弹性云主机实例。

5. 修改注册表子项 PortNumber 的值。

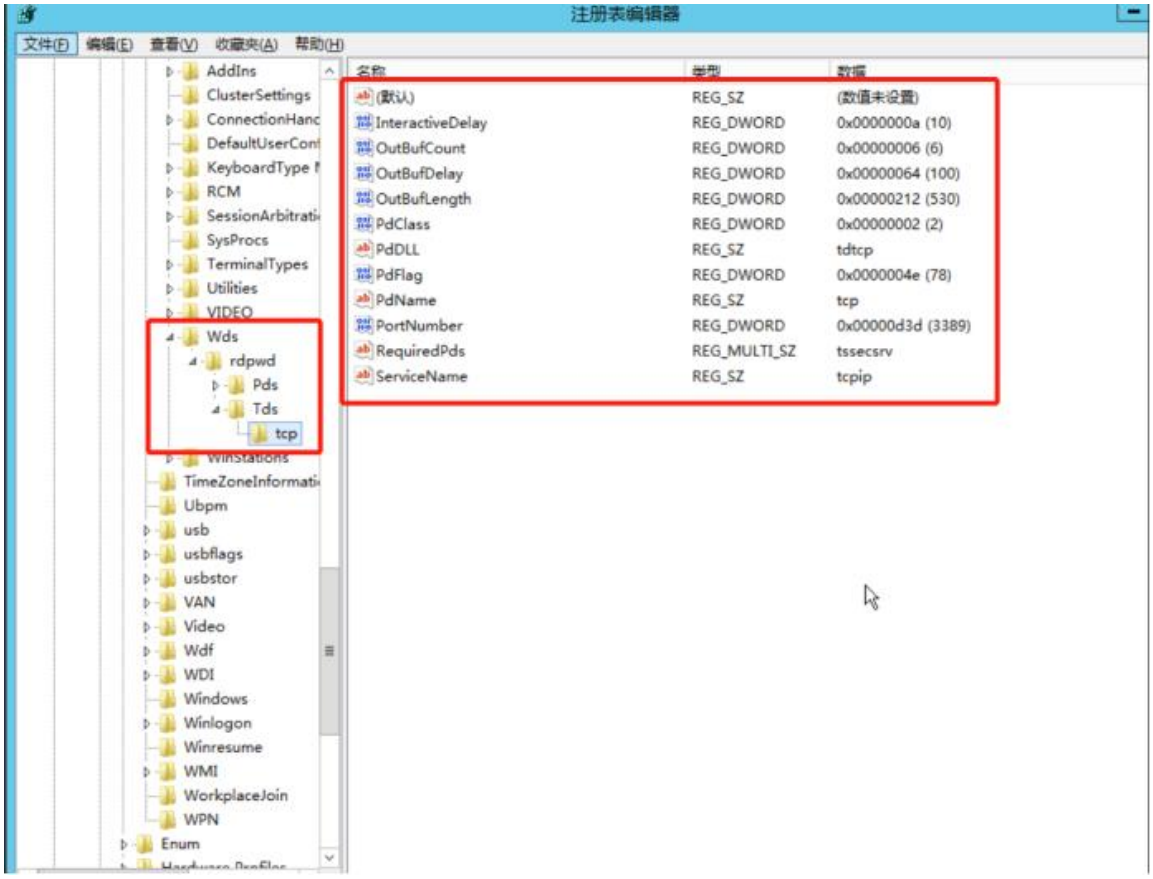
1）右键单击 Windows 徽标键，选择运行，启动运行窗口。



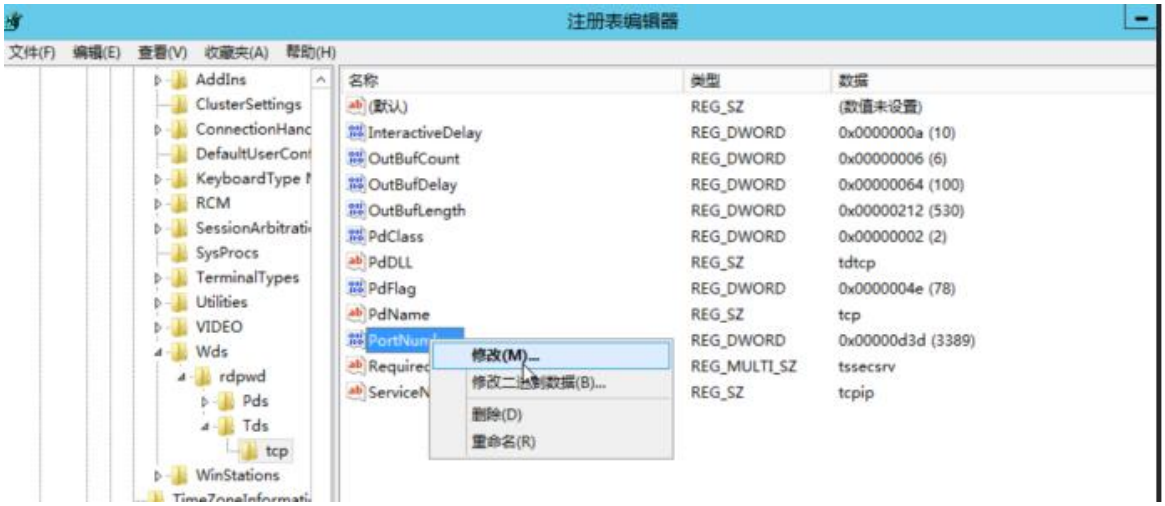
2）在运行窗口的文本框内输入 regedit.exe 后按回车键，打开注册表编辑器。



3）在左侧导航树中逐层选择 HKEY_LOCAL_MACHINE > System > CurrentControlSet > Control > Terminal Server > Wds > rdpwd > Tds > tcp，如下图所示：



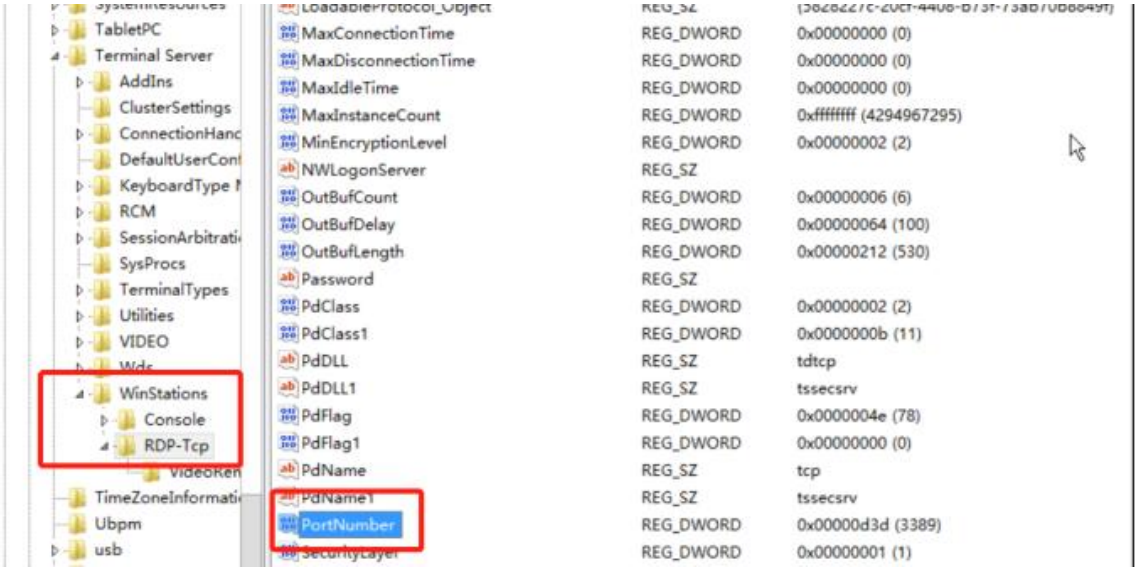
4）在列表中找到注册表子项 PortNumber 并右键单击，选择修改，进入修改窗口。



5）在弹出的修改窗口中，在数值数据的文本框中输入新的远程端口号，以 3398 为例。在基数框中勾选十进制，然后单击确定。



6）在左侧导航树中逐层选择 HKEY_LOCAL_MACHINE > System > CurrentControlSet > Control > Terminal Server > WinStations > RDP-Tcp。



7）在右侧列表中找到注册表子项 PortNumber 并右键单击，选择修改（可以使用键盘方向键向下寻找）。

8）在弹出的对话框中，在数值数据的文本框中输入新的远程端口号，在本示例中即 3398。在基数框中勾选十进制，然后单击确定。



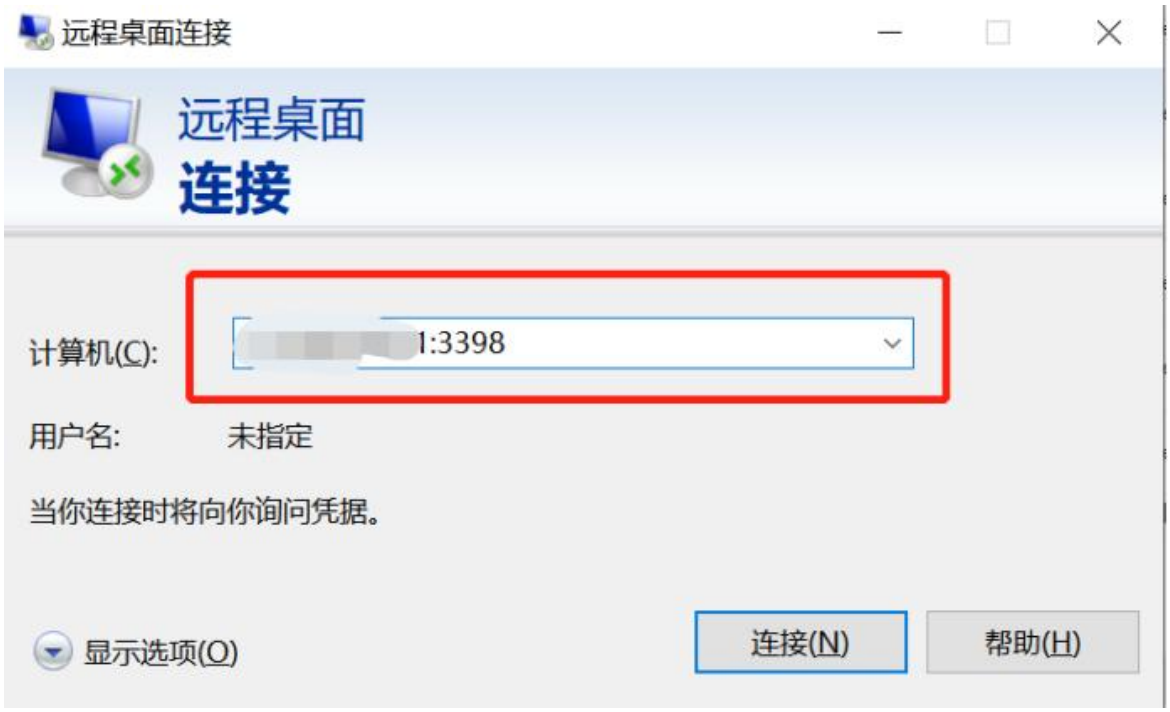
6. 在弹性云主机管理控制台中将此台云主机进行重启，具体如下图：



7. 为该实例添加安全组规则，允许新配置的 3398 远程端口进行连接。

具体操作，请参见配置安全组规则。

8. 最后使用远程桌面连接功能远程访问实例，在远程地址后面添加新端口号 3398 即可连接实例。



修改 Linux 系统实例默认远程端口，以 CentOS 8.0 64 位为例。

1. 登录控制中心。

2. 选择地域 华东-华东 1。

3. 单击“计算>弹性云主机”，进入弹性云主机页面。

4. 单击待修改的弹性云主机行的“操作>远程登录”按钮，远程连接弹性云主机实例。



5. 输入用户名 root，密码为购买弹性云主机时用户自定义的密码，登录成功之后如图：

```
CentOS Linux 8 (Core)
Kernel 4.18.0-80.el8.x86_64 on an x86_64

ecm-ctrlx login: root
Password:
Last login: Thu Jun 29 14:55:51 on tty1
[root@ecm-ctrlx ~]# _
```

6. 因 sshd_config 是 Linux 中重要的配置文件，为了避免误操作所带来的故障，在运行前首先对 sshd_config 进行备份，请运行以下命令：

7. 进入到/etc/ssh 路径下输入 ll 命令查看，具体信息如下图：

```
cp /etc/ssh/sshd_config /etc/ssh/sshd_config_bak
```

```
[root@ecm-ctrlx ~]# cp /etc/ssh/sshd_config /etc/ssh/sshd_config_bak
[root@ecm-ctrlx ~]# cd /etc/ssh
[root@ecm-ctrlx ssh]# ll
total 604
-rw-r--r--. 1 root root 563386 May 11  2019 moduli
-rw-r--r--. 1 root root   1727 May 11  2019 ssh_config
drwxr-xr-x. 2 root root     28 Mar  8  2021 ssh_config.d
-rw-------  1 root root   4443 Mar 10  2022 sshd_config
-rw-------  1 root root   4443 Jun 29 16:25 sshd_config_bak
-rw-------  1 root root   1381 Sep  1  2022 ssh_host_dsa_key
-rw-r--r--  1 root root    601 Sep  1  2022 ssh_host_dsa_key.pub
-rw-------  1 root root    505 Sep  1  2022 ssh_host_ecdsa_key
-rw-r--r--  1 root root    173 Sep  1  2022 ssh_host_ecdsa_key.pub
-rw-------  1 root root    399 Sep  1  2022 ssh_host_ed25519_key
-rw-r--r--  1 root root     93 Sep  1  2022 ssh_host_ed25519_key.pub
-rw-------  1 root root   1811 Sep  1  2022 ssh_host_rsa_key
-rw-r--r--  1 root root    393 Sep  1  2022 ssh_host_rsa_key.pub
[root@ecm-ctrlx ssh]#
```

8. 修改 sshd 服务的端口号，因为已经在/etc/ssh 路径下，因此直接运行 vim sshd_config 编辑 sshd_config 配置文件，在键盘上按 i 键，进入编辑状态，添加新的远程服务端口，本节以 2222 端口为例。在 Port 22 下输入 Port 2222。

```
# If you want to change the port on a SELinux system, you have to tell
# SELinux about this change.
# semanage port -a -t ssh_port_t -p tcp #PORTNUMBER
#
#Port 22
#AddressFamily any
#ListenAddress 0.0.0.0
#ListenAddress ::
```

```
# If you want to change the port on a SELinux system, you have to tell
# SELinux about this change.
# semanage port -a -t ssh_port_t -p tcp #PORTNUMBER
#
#Port 22
port 2222
#AddressFamily any
#ListenAddress 0.0.0.0
#ListenAddress ::
```

9. 在键盘上按 Esc 键，输入:wq 后保存并退出编辑模式。

10. 运行以下命令重启 sshd 服务。

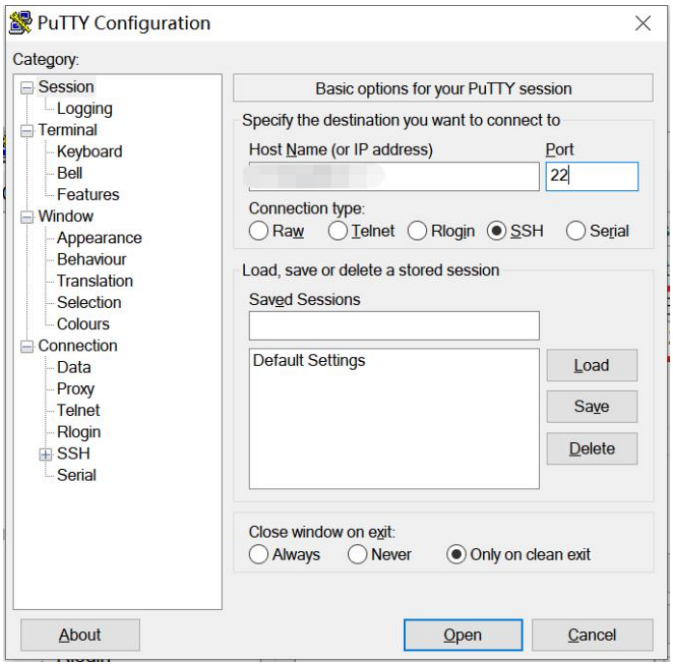
```
systemctl restart sshd
```

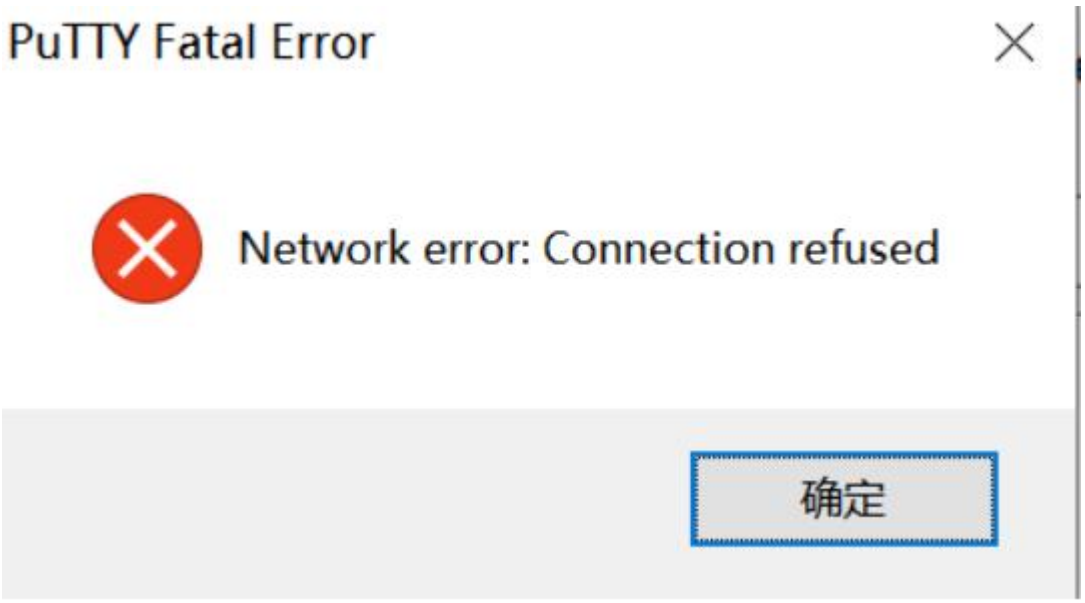
```
[root@ecm-ctrlx ssh]# systemctl restart sshd
[root@ecm-ctrlx ssh]#
```

至此，此弹性云主机的远程登录默认端口已经从 22 改为了 2222。

功能验证：

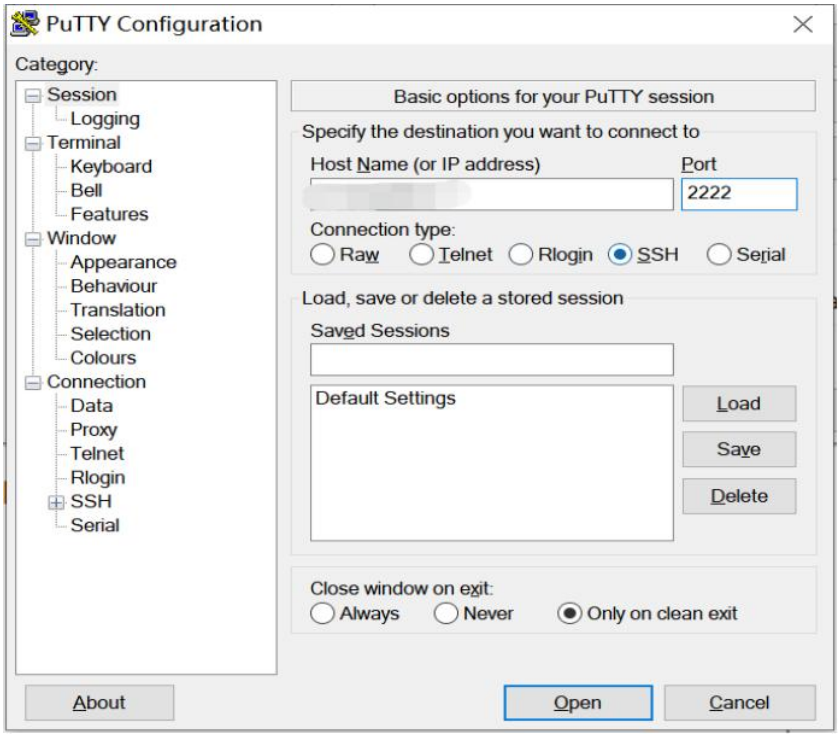
1）打开远程连接工具 putty，输入此弹性云主机实例的弹性公网 IP，先使用默认 22 端口登录，可以看到连接被拒绝，网络已中断，说明目前 22 端口已经无法进行远程登录。

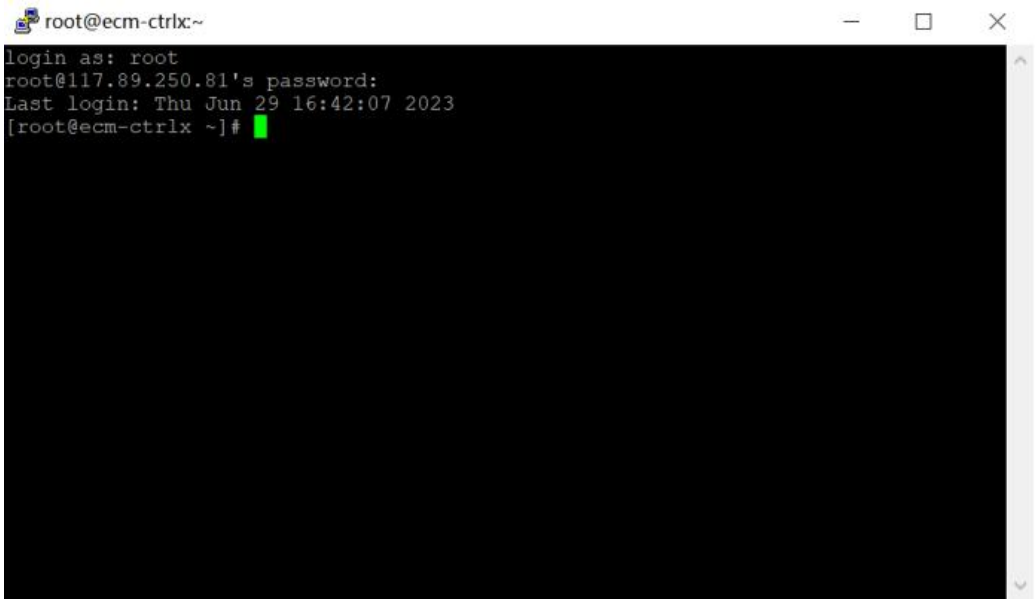




2)配置实例的安全组，放行 TCP 协议 2222 端口，具体操作，请参见配置安全组规则

3)配置完成之后，使用 SSH 工具连接新端口 2222，在 port 下输 2222，可看到能够成功连接。





8.5.1.4 修改 Linux 弹性云主机的默认 SSH 端口后，使用新端口无法登录？

新端口无法登录的原因可能有：安全组没有放通新端口，防火墙没有放通新端口，sshd 文件未添加新端口等。

- 检查安全组规则

  以新的 SSH 端口号为 2222 为例，安全组规则在确保“入方向”需要放通 2222 端口。如果没有此规则，须在安全组中添加 2222 入方向端口

- 检查防火墙规则

  检查 iptables 防火墙是否放通新的 SSH 端口，以 2222 端口为例。

  1）登录 Linux 弹性云主机。以 CentOS 7.6 操作系统为例，执行以下命令编辑 iptables 文件。

```
vi /etc/sysconfig/iptables
```

  2）添加 2222 端口规则

```
-A INPUT -m state -state NEW -m tcp -p tcp -dport 2222 -j ACCEPT
```

  3）修改完成后重启 iptables 服务。

```
systemctl restart iptables
```

- 检查 SSH 配置文件

1）登录弹性云主机检查 ssh 配置文件。执行如下命令，查看是否配置了 Port 2222。

```
vi /etc/ssh/sshd_config
```

2）如未配置，请将“#Port 22”替换为“Port 2222”。

3）执行如下命令，重启 ssh 服务。

```
service sshd restart
```

8.5.1.5 使用 VNC 远程登录，想要复制本地字符发送到虚机中，发现一次性无法达到描述的 3000 个字符

出现该现象有可能是以下两个原因导致：

1.该主机购买的较早，用户需要手动同步升级该功能。具体升级方法，需要用户该关闭主机再重新启动后方可完成更新。关机前请用户注意保存数据，防止数据丢失。

2.以下主机规格不支持 3000 个字符，最多可支持 160 字符，实际新情况请以控制台为准：

- 云主机：s2
- GPU 云主机：P2v、G5、PI2、G6、P2vs、G5s。

7.5.1.6 VNC 方式登录弹性云主机后，较长时间不操作，界面无响应怎么办

通过 VNC 方式登录到弹性云主机，但是在一段时间内没有进行任何操作，VNC 登录界面可能会出现无响应的情况，键盘和鼠标也无法正常操作。这种情况下，你可以尝试重新启动云主机来解决这个问题。

要重新启动云主机，你可以按照以下步骤进行操作：

1.打开终端或命令行界面。

]2.点击右上方 “Send CtrlAltDel”按钮。

3.等待一段时间，以确保主机已重启完毕。

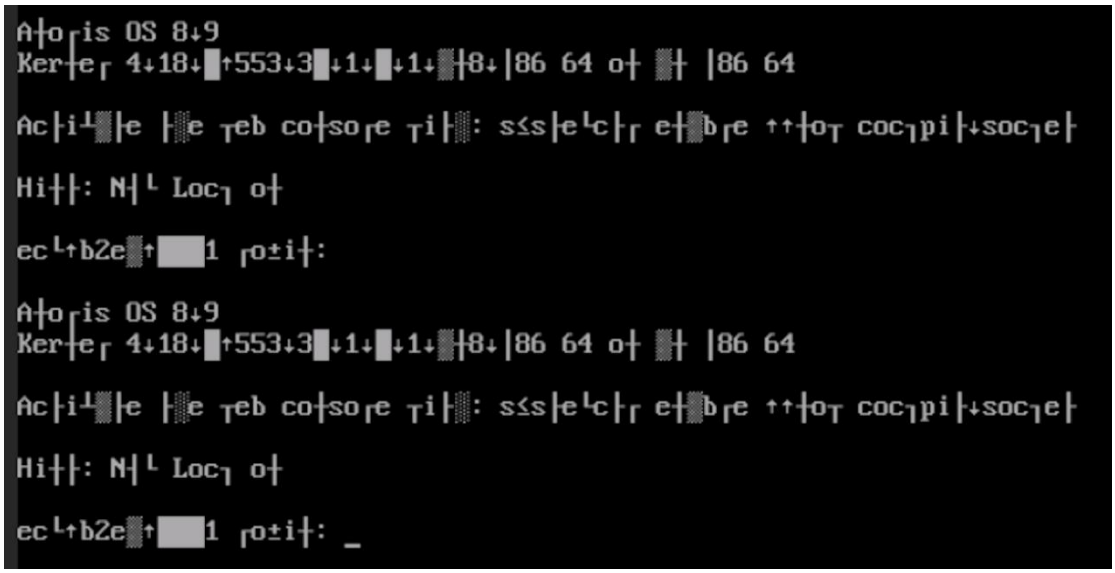
请注意，重新启动会中断正在进行的任务，因此请确保在重新启动之前保存好所有未保存的工作。如果重新启动云主机后仍然遇到问题，您可以提交工单或致电客服电话 400-810-9889 寻求技术支持。

8.5.2 控制台登录类

8.5.2.1VNC 方式登录弹性云主机时，登录界面显示乱码怎么办？

问题现象

当用户通过VNC方式登录Linux系统云主机时，界面可能出现乱码（如下图所示），您可以通过执行以下操作处理该故障。



可能原因

查看的文件过大或物理按键被意外持续按住。

处理方法

方法一： 检查是否有物理按键下按情况，包括 Ctrl、Alt、Shift 按键，用户可尝试按下该按键进行调整。

方法二： 初始化终端界面，刷新当前终端屏幕，在命令行，执行以下命令。

```
reset
```

若以上方法均无法解决问题，请联系工单进行处理。

8.5.2.2 通过控制台登录弹性云主机时提示 1006 或 1000 怎么办？

问题现象

当用户通过 VNC 远程登录方式登录云主机，VNC 登录界面出现 disconnected（code：1006）或者 disconnected（code：1000）提示字样。

可能原因

出现以上的现象，有可能是由于以下原因造成：

- 连接期间长时间未操作，链接已断开。
- 弹性云主机非正常可用状态。
- 存在其他子用户登录 VNC。

排查步骤

当出现该现状时，用户可首先通过重新连接 VNC 远程登录界面进行排查，操作步骤如下：

1.退出当前 VNC 登录界面。

2.选中云主机，选择远程登录，重新登录主机。若登录失败，则请用户检查当前云主机运行状态，是否处于“运行中”，若此时云主机状态异常，请联系人工客服解决问题。

3.确认是否有其他子用户正在使用该台云主机的 VNC 界面。

8.5.2.3 VNC 方式登录弹性云主机时，系统黑屏输入无反应？

VNC 方式登录弹性云服务器时，系统黑屏输入无反应？

若出现此种现象，那有可能是存在其它用户登录了远程 VNC。当前 VNC 远程登录功能，只支持一个用户登录，不支持多用户同时登录，若存在多个用户登录的情况，已登录用户的 VNC 链接将会自动断开，故此时将会出现黑屏。

8.5.2.4 VNC 方式登录弹性云主机后，查看数据失败，VNC 无法正常使用？

用户通过 VNC 方式登录弹性云主机后，查看数据出现乱码、错误码等情况，导致用户无法正常使用 VNC。

以上现象出现的原因有如下几点，用户可依次对照排查：

- 查看大文件长时间未读取完成；
- 游戏像素过高、播放视频清晰度过高；
- 浏览器缺陷，占用内存大。

如存在以上场景，则用户重新登录云主机或者更换浏览器即可。

8.5.3 登录前准备类

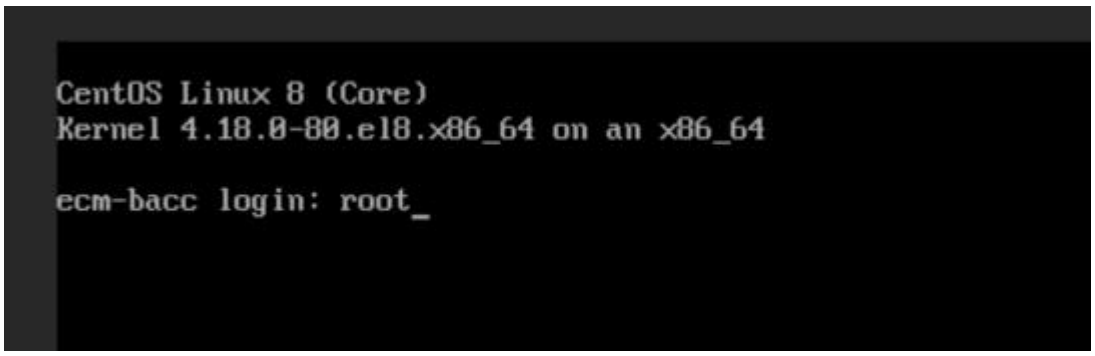
8.5.3.1 使用创建时的用户名和密码无法 SSH 方式登录云主机

操作步骤

1. 进入天翼云弹性云主机控制台，选中要操作的云主机。



2. 通过 VNC 的登录方式链接到云主机控制台。



3. 进入命令行操作界面，输入 root 用户名及设置的登录密码。

4. 输入以下命令。

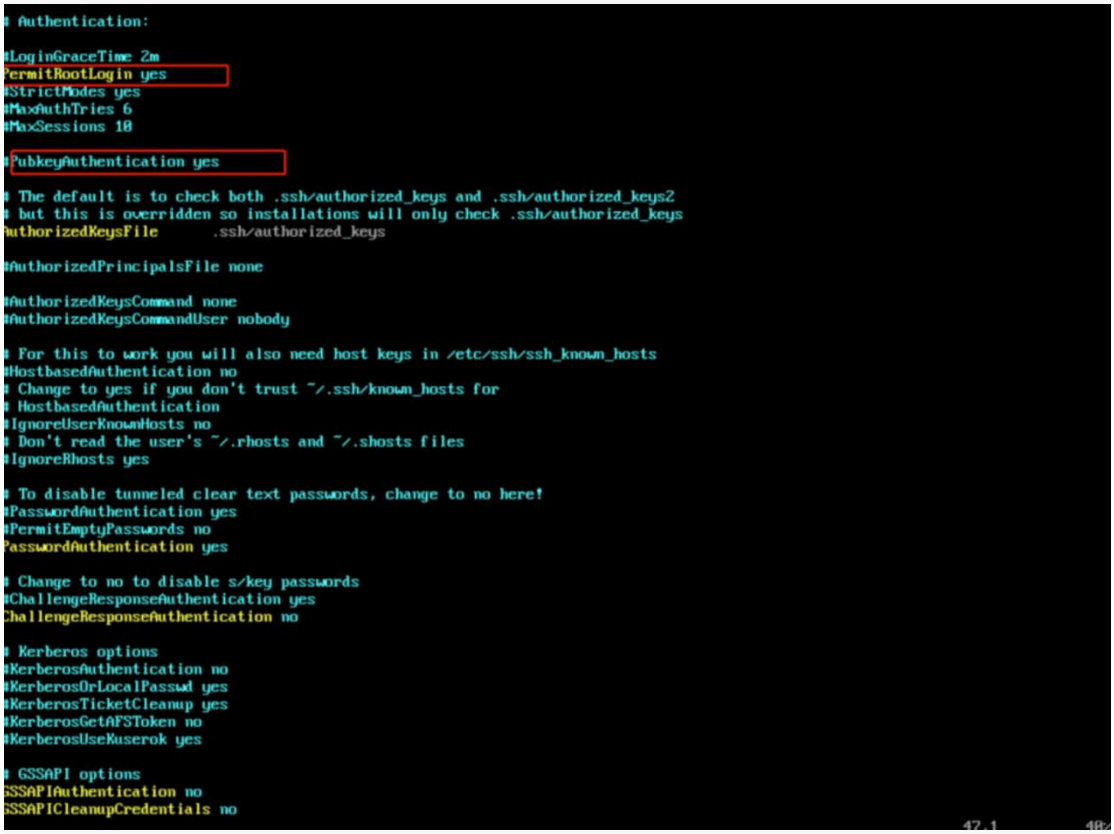
```
vim  /etc/ssh/sshd_config
```

5. 修改以下配置项。

```
PermitRootLogin yes

PasswordAuthentication yes

AllowUsers root（无则添加）
```



6. 完成修改够，退出保存。

```
:Wq!
```

7.重启 sshd 服务，使修改项生效。

```
service sshd restart
```

至此，已完成 SSH 服务配置的修改，用户可再次尝试远程登录操作。

8.5.3.2 远程登录时需要输入的帐号和密码是什么

登录云服务器的用户名和密码：

Windows 操作系统用户名：Administrator

Linux 操作系统用户名：root

如忘记登录密码或创建云主机时未设置密码，可通过 “重置密码” 功能设置新密码。

8.5.3.3 弹性云主机登录前准备工作有哪些

- Windows 操作系统

1.获取登录密码

使用创建云主机时设置的密码登录弹性云主机，如您忘记密码，可以使用“重置密码”重新设置密码。

2.MSTSC 方式登录时，需要确保云主机绑定弹性 IP 及开启远程桌面协议。如果没有，可以绑定弹性 IP，以及使用控制台远程登录工具登录云主机开启远程桌面协议。

- Linux 操作系统

1.获取登录密码或密钥对

使用创建云主机时设置的密码或绑定的密钥对，如您忘记密码，可以使用“重置密码”重新设置密码。

3. SSH 方式登录时，需要确保云主机绑定弹性 IP。

8.5.3.4 远程登录忘记密码怎么办

如忘记登录密码，可通过 “ 重置密码 ” 功能设置新密码。

您可以通过弹性云主机控制台重置密码，点击更多-重置密码更改密码，使用新密码再进行登录。具体操作步骤如下：

1.打开弹性云主机控制台页面。
2.选择您需要重置密码的弹性云主机，点击“更多”按钮。
3.在弹出的下拉菜单中，选择“重置密码”选项。
4.在弹出的重置密码窗口中，您需要输入一个新的密码，并确认一遍输入正确的密码。请注意密码的规则为 8～30 个字符，必须同时包含三项（大写字母、小写字母、数字、 ()`~!@#$%^&*_-+=|{}[]:;'<>,.?/ 中的特殊符号），且不能以斜线号（/）开头，不能包含连续字符。
5.点击“确定”按钮，系统会提示您重置密码成功。
6.使用新密码登录弹性云主机控制台，进行后续操作。

更多可以参考在控制台重置密码。

8.5.4 远程连接类

8.5.4.1MSTSC 方式登录 Windows 2012 的弹性云主机，登录失败怎么办？

问题描述

当用户本地使用 Windows 10 操作系统时，不能以初始密码直接远程连接操作系统为 Windows 2012 的弹性云主机。

可能原因

Windows10 操作系统自身限制不能以初始密码远程连接到 Windows 2012 的弹性云主机。

处理方法

方式一：修改弹性云主机的初始密码。

1. 首次登录，以 VNC 方式登录操作系统为 Windows 2012 的弹性云主机。
2. 登录成功后，按照如下方式修改弹性云主机密码。
（1）按 Windows 键并键入管理工具。
（2）双击计算机管理。
（3）展开本地用户和组。
（4）单击用户。
（5）右键单击管理员。单击设置密码 >继续。
（6）在两个字段中输入您的新密码，然后点击 OK。
3. 使用修改后的密码，以 MSTSC 方式远程登录。

方式二：修改.rdp 格式的远程桌面文件

1. 在远程桌面连接弹窗中输入弹性公网 IP、用户名“administrator”以及创建弹性云主机时设置的登录密码，点击确定。
2. 收到报错提示后，单击“远程桌面连接”页面左下角的“显示选项”，单击“连接设置”栏的“另存为”，保存“.rdp”格式的远程桌面文件。
3. 打开保存的远程桌面文件，在文件的最后一行，添加如下语句并保存：enablecredsspsupport:i 0。
4. 双击更新后的“.rdp”文件，打开远程桌面连接。
5. 单击“连接”，重新连接操作系统为 Windows 2012 的弹性云主机。

8.5.4.2 Windows 2012 系统卸载某些软件后无法进入系统桌面怎么办？

问题描述

当主机使用 Windows2012 浏览器时，部分场景下需要将 Windows2012 自带的.net framework 4.5 版本卸载，卸载之后可能会遇到黑屏、无法进入系统桌面的问题。

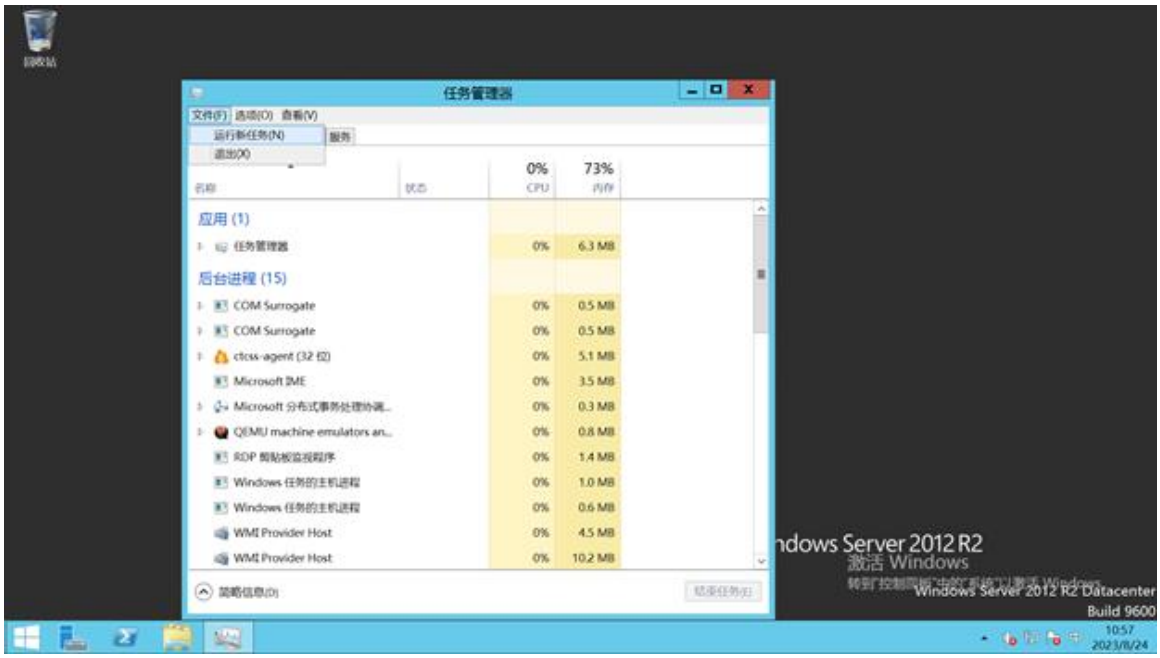
可能原因

如遇这种情况，可能是因为系统由完整模式 Full 变为了核心模式 Core。

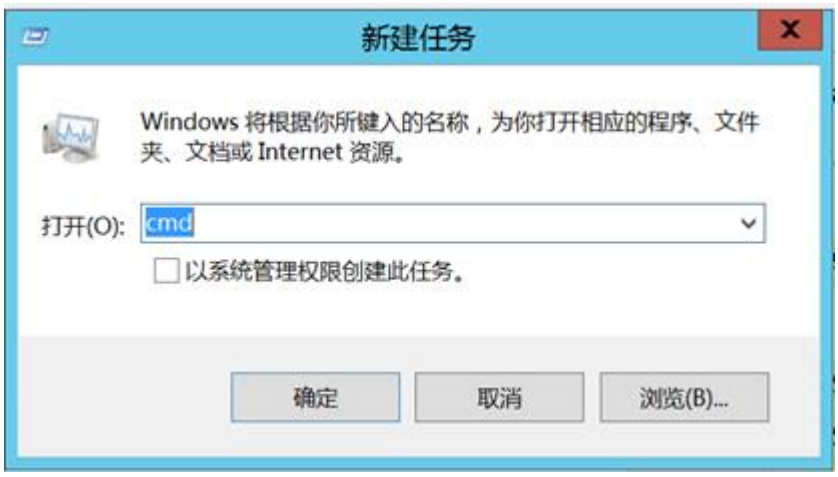
处理方法

1.登录弹性云主机，点击 Win 键，在开始菜单中选择任务管理器。

2.选择“文件 > 运行新任务”。打开“新建任务”窗口。



3.输入“cmd”，然后按回车键。



4.在弹出的命令行窗口执行以下命令，将系统由核心模式切换到完整模式。

```
Dism /online /enable-feature /all /featurename:Server-Gui-Mgmt
/featurename:Server-Gui-Shell /featurename:ServerCore-FullServer
```

3. 根据提示重启云主机，再次登录系统后就可以正常显示桌面。

8.5.4.3 远程登录弹性云主机时，对浏览器版本的要求？

当您使用控制台 VNC 登录云主机时，使用的浏览器版本应满足如下要求，否则会出现样式、功能不正常等问题。

浏览器版本
Chrome≥87
FireFox≥78
Safari≥14

浏览器版本
Edge≥88

8.5.4.4 使用私有镜像创建 Windows 弹性云主机时设置的密码不生效怎么办？

问题描述

使用私有镜像创建 Windows 弹性云主机时设置的密码不生效，登录时提示密码错误。

可能原因

Cloudbase-Init 是 Windows Server 操作系统初始化的工具。它会对云主机进行一些列初始化操作，比如配置网络，注入 kms server 的密钥，扩展磁盘，以及运行用户的自定义脚本等等。天翼云公共镜像默认已安装 Cloudbase-Init，不需要执行安装及配置操作。如您遇到描述问题可能是以下原因。

原因一：Cloudbase-Init 正在按流程执行相关初始化操作，需要一定时间。

原因二：您使用的私有镜像未安装或未正确配置 Cloudbase-Init。

处理方法

针对原因一：如果您确认正确安装并配置了 Cloudbase-Init，您可以在看到用户登录界面后稍微等待一段时间，受多方面因素影响，Cloudbase-Init 按流程执行相关初始化操作，到完成初始化密码需要的时间可能比期望的长。

针对原因二：此时您可以直接通过控制台进行重置密码操作，操作成功后使用新的密码即可进行登录。

8.5.4.5 弹性云主机实例出现了异地登录怎么办

如果出现异地登录，解决办法如下：

1.确认异地登录的时间点，是否是自己或者其他管理员登录。

2.如果不是合法管理员登录，可执行以下操作：

- 立即重置密码。
- 排查是否被病毒攻击。
- 使用安全组功能设置只允许特定的 IP 登录。

8.5.4.6 无法远程登录 Windows 云主机怎么办

无法远程登录 Windows 云主机可以使用以下方式进行问题排查。

1. 检查弹性 IP 是否能 ping 通。

   如果不能 ping 通，请参见文档“ping 不通或丢包时如何进行链路测试？”进行检查。

   如果能 ping 通，请执行 2。

2. 检查弹性云主机的 3389 端口是否能够访问。

   如果不能访问，请检查安全组是否开放 3389 端口。

   如果能访问，请执行 3。

3. 检查弹性云主机内的 RDP 服务是否存在异常。

   登录同一内网段的其他 Windows 主机，在该主机上远程登录问题主机。

   如果不能登录说明 RDP 服务可能存在异常，通过 VNC 登录到问题主机，查看 RDP 服务是否存在异常，重启 RDP 服务后，重新尝试连接。

请按照以下方法重启 RDP 服务。

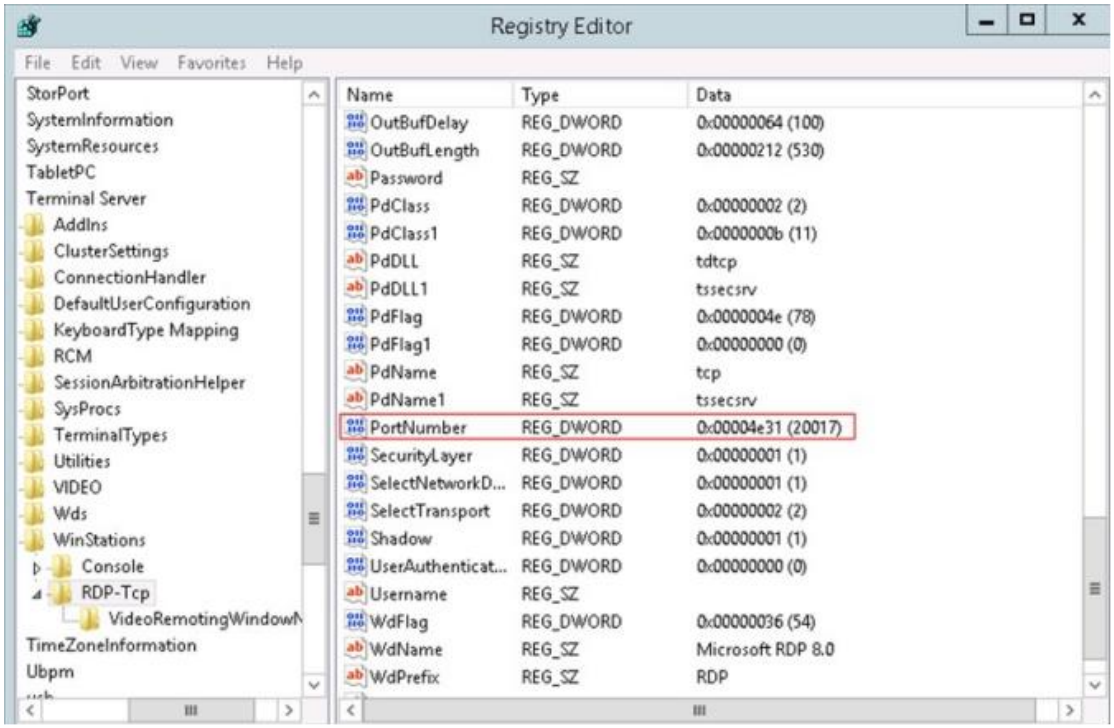
- 单击“开始”，选择“运行”，输入“services.msc”。
- 在服务列表中，选择“Remote Desktop Services”服务，重新启动该服务。



4. 查看弹性云主机是否修改远程连接端口。

查看

“HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\TerminalServer\WinStations\RDP-Tcp\PortNumber”下，配置的端口是否是默认的 3389，若不是，远程连接的时候将端口改为配置中的端口。



5. 查看弹性云主机是否限制连接数量。

检查 OS 内部远程桌面配置：

- 单击“开始”，选择“运行”，输入“gpedit.msc ”，打开组策略。
- 选择“计算机配置 > 管理模板 > 远程桌面服务 > 远程桌面会话主机 > 连

接 ，“打开“限制连接的数量”，查看是否启用限制。



说明

如果此处有配置最大连接数，可能导致 Windows 远程登录无法连接，如果同时有多个用户正在登录该主机，且数量已经超过最大连接数，此时最大连接数外的用户接收不到该主机的仲裁，导致会话会一直处于正在配置中的状态。 请选择禁用该配置或者设置更大的最大连接数。

6. 其它解决方案

通过上述排查后，仍然不能连接 Windows 实例，请您保存自助诊断结果，通过提交工单联系天翼云的技术支持寻求帮助。

8.5.4.7 无法远程登录 Linux 云主机怎么办

无法远程登录 Linux 云主机可以使用以下方式进行问题排查：

1. 检查资源状态是否正常

请在控制台检查云服务器资源状态是否为“运行中”。如果云服务器状态为“关机”，请先开机再登录云服务器。

2.检查登录凭证

请检查您在创建 ECS 时设置的登录凭证。

密码：请确认使用的登录密码是否准确，如果忘记密码可以通过重置密码功能重新设置云服务器登录密码。重置密码后请确保已重启云服务器使新密码生效。

创建后设置：如果创建时未设置登录凭证，云服务器创建成功后请单击“操作”列下的“重置密码”，根据界面提示，为弹性云服务器设置密码，请确保已重启云服务器使新密码生效。

3.检查云主机负载是否过高

云主机的带宽和 CPU 利用率过高可能会导致无法登录。您可以通过云监控服务创建告警任务，当 CPU 或带宽利用率高时，系统会自动发送告警给您。

如果是 CPU 占用过高导致的无法登录请参考以下操作降低 CPU 使用率：

- 关掉一些暂时不使用的进程。
- 重启云主机。
- 变更云主机规格以升级 vCPU、内存。
- 检查是否是带宽超限导致的无法登录，可尝试进行扩大带宽操作，扩大带宽的操作请参考弹性 IP-修改带宽。

完成上述操作后，再次重试远程连接云主机。

4.检查安全组配置是否正确

在云主机的详情页面选择“安全组”页签，查看安全组入方向规则中已添加 22 端口。

如果没有，添加入方向的 22 端口，具体操作参见虚拟私有云-添加安全组规则。

完成上述操作后，再次重试远程连接云主机。

5.其他解决方案

通过上述排查后，仍然不能连接 Linux 实例，请您保存自助诊断结果，通过提交工单联系天翼云的技术支持寻求帮助。

8.6 使用故障类

8.6.1 弹性云主机使用时有什么限制

使用弹性云主机有以下限制：

- 弹性云主机不支持安装虚拟化软件和二次虚拟化。
- 不支持声卡应用。
- 不支持直接加载外接硬件设备（如硬件加密狗、U 盘、外接硬盘、银行 UKey 等）。
- 请勿卸载云主机硬件的驱动程序。
- 请勿修改网卡的 MAC 地址。

8.6.2 宿主机故障时，弹性云主机是否能自动恢复

宿主机故障时，弹性云主机可以自动恢复。

虽然天翼云提供了多种机制来保障系统的可靠性、容错能力和高可用性，但宿主机仍存在故障的可能性，如电源等部件因不可抗力损坏。

弹性云主机是运行在宿主机上的，当宿主机无法正常运行时，天翼云会自动将弹性云主机迁移至正常的宿主机，以疏散或者热迁移的方式迅速恢复，最坏情况为弹性云主机重启，将宿主机宕机的影响降到最小。

8.6.3 如何处理管理控制台异常提示信息

问题描述

您可以通过本节内容解决如下问题：

- 用户在管理控制台执行弹性云主机相关操作后出现异常，针对管理控制台提示的异常信息，应该如何处理？
- 用户根据弹性云主机《API 参考》调用相关的 API 接口时，如果返回错误码，

应该如何处理？

背景信息

用户通过管理控制台执行弹性云主机的相关操作后，弹性云主机列表页面将显示相应操作的申请状态。通过申请状态中显示的信息，用户可以获悉当前操作请求的执行状态。

- 如果操作请求正常执行完毕，则任务提示栏中的记录将自动消失。
- 如果操作请求在执行过程中出现异常，则任务提示栏将返回错误说明。

处理方法

- 针对管理控制台提示的异常信息，您可以通过表 1 中的处理建议进行下一步操作，处理相应的异常。

表 1 常见错误信息处理建议

管理控制台提示信息	按量付费处理建议	包年/包月处理建议
当前账号余额不足 100 元或不支持按量付费功能，请确认后，重新购买。	在用户中心充值，确保现金金额不小于 100 元。	不涉及。
您的配额不足，无法创建或购买对应服务，请您申请更多配额。	按照如下操作申请配额：提交工单->新建工单->配额类->配额申请。	按照如下操作申请配额：提交工单->新建工单->配额类->配额申请。
云主机中间状态下不可执行该操作。	等待云主机上一个操作完成后重试。	等待云主机上一个操作完成后重试。
辅助私网 IP 地址必须在主私网 IP 地址段内。	选取主私网 IP 地址段内的 IP 作为辅助私网	选取主私网 IP 地址段内的 IP 作为辅助私网

	IP 地址。	IP 地址。
内部错误，请联系客户经理协助处理。	联系客户经理或者提交工单。	联系客户经理或者提交工单。

- 如果在调用弹性云主机相关 API 接口时出错，可以根据调用的 API 接口文档里的错误码描述表查看错误的详细信息，并对参数进行相应的调整。

8.6.4 如何修改弹性云主机（Windows）的 SID

操作场景

Windows 操作系统是通过安全标识符（SID）对计算机和用户的识别。由于基于同一镜像生产的云服务器实例 SID 相同，会引起无法入域的问题。如果您需要搭建 Windows 域环境，则需要通过修改 SID 以达到入域的目的。本文档以 Windows Server 2012 操作系统云服务器为例，介绍如何使用系统自带 sysprep 以及 sidchg 工具修改 SID。

注意
本说明仅适用于 Windows Server 2008 R2 、Windows Server 2012 以及 Windows Server 2016 及更高版本。
如果有批量修改 SID 的需求，可通过制作自定义镜像（选择 “执行 sysprep 制作镜像”）解决。
修改云主机 SID 可能导致数据丢失或系统损坏，建议您提前做好数据备份。

操作步骤

使用 sysprep 修改 SID

注意

1. 使用 sysprep 修改 SID 后，系统参数很多都被重新设置，包括 IP 配置信息等，您必须手动重新设置。

2. 使用 sysprep 修改 SID 后，C:\Users\Administrator 将会被重置，系统盘部分数据将被清理，请注意做好数据备份。

1. 使用远程登录方式登录云主机。

2. 在操作系统界面，点击运行，输入 cmd，按 Enter，打开管理员命令行工具。

在管理员命令行工具中，执行以下命令，保存当前网络配置。

```
ipconfig /all
```

3. 以管理员身份使用 PowerShell 执行以下命令。

```
$unattendContent = @"<unattend
xmlns="urn:schemas-microsoft-com:unattend">
    <settings pass="oobeSystem">
        <component name="Microsoft-Windows-Shell-Setup"
processorArchitecture="amd64" publicKeyToken="31bf3856ad364e35"
language="neutral" versionScope="nonSxS">
            <OOBE>
                <HideEULAPage>true</HideEULAPage>
                <NetworkLocation>Work</NetworkLocation>
                <ProtectYourPC>3</ProtectYourPC>
                <SkipUserOOBE>true</SkipUserOOBE>
                <SkipMachineOOBE>true</SkipMachineOOBE>
            </OOBE>
        </component>
    </settings></unattend>
```

```
"@

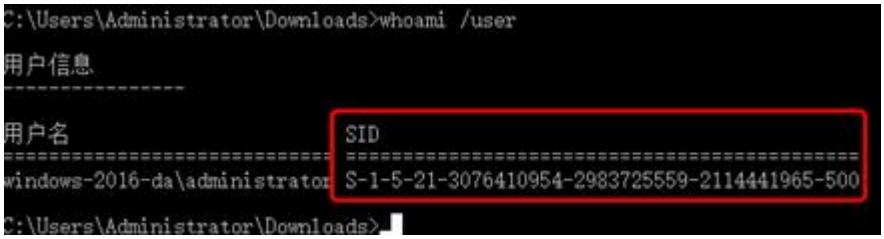
$unattendContent | Out-File -FilePath
C:\Windows\System32\Sysprep\unattend.xml -Encoding UTF8 -Force

C:\Windows\System32\Sysprep\sysprep.exe /generalize /oobe /shutdown
/unattend:C:\Windows\System32\Sysprep\unattend.xml
```

4.执行完成后，系统会自动关机。此时，您可在云平台制作私有镜像。

5.启动云主机进入操作系统后执行下方命令，验证 SID 是否已修改，返回类似如下信息，则表示 SID 已完成修改。

```
whoami /user
```



6.根据步骤 2 保存的网络配置信息，重新设置网卡相关信息(如 IP 地址、网关地址、DNS 等)。

使用 sidchg 修改 SID

1.登录云主机。

2.通过 IE 浏览器访问和下载 sidchg 工具：sidchg 工具下载地址。

说明

Server 版本请下载 64bit 版本。

3.通过执行如下命令修改云主机的 SID。

```
sidchg64-3.0j.exe /R
```



说明

/R 表示修改后自动重启，/S 表示修改后关闭，使用详情请参照 SIDCHG 官方说明。

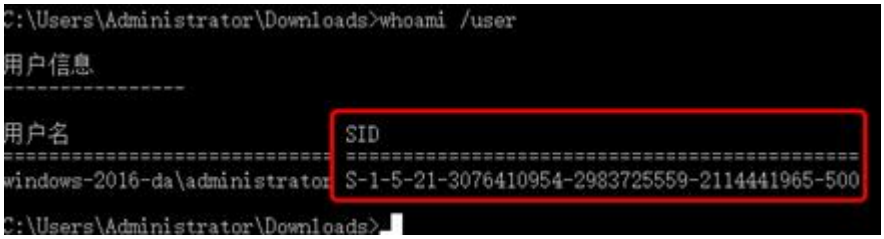
4. 根据界面提示，输入 license key 或者 trial key，按 Enter。

5. 根据界面提示，输入"Y"，按 Enter。

6. 在修改 SID 的提示框中，单击确定，进行 SID 重置，重置过程中，系统将会被重启如下图所示：



7. 待完成重启后，右键单击>运行，输入 cmd，按 Enter，打开管理员命令行工具，执行命令 whoami /user，验证 SID 是否已修改。



8.6.5Windows 弹性云主机没有声音

购买 Windows 弹性云主机后，通过 MSTSC 远程连接，发现没有声音。通过 MSTSC 远程连接的 Windows 弹性云主机如何播放音频？

约束限制

本节内容适用于 Windows Server 2008、Windows Server 2012、Windows Server 2016、Windows Server 2019 系统的弹性云主机。

故障原因

Windows 弹性云主机默认禁用音频设备，导致无法通过远程桌面的方式使用音频设备。如需播放音频、使用多媒体音频功能，可参见本节内容进行设置。

解决步骤

第 1 步：启动 Windows Audio 服务

1. 打开“运行”窗口。

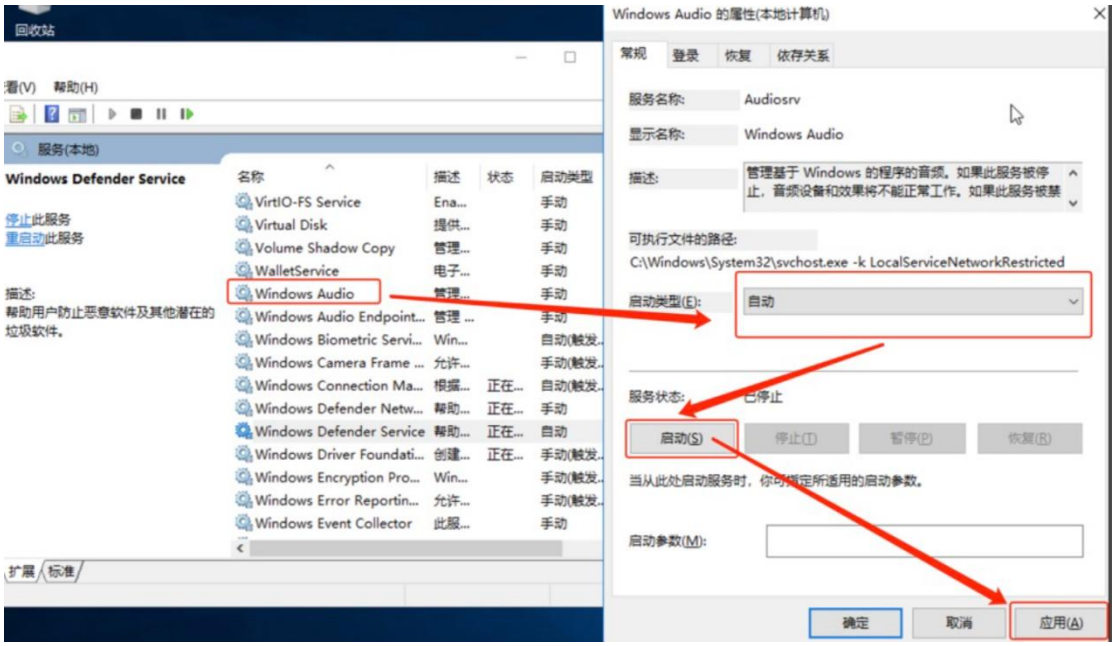
2. 输入“services.msc”，打开“服务”。

3. 找到“Windows Audio”服务，并按如下方式设置。

启动类型：自动

服务状态：启动

以 Windows 2016 操作系统为例，如下图所示。



第 2 步：开启音频和视频播放功能

操作系统不同， “音频和视频播放”功能开启方法不同。

Windows 2008 系统

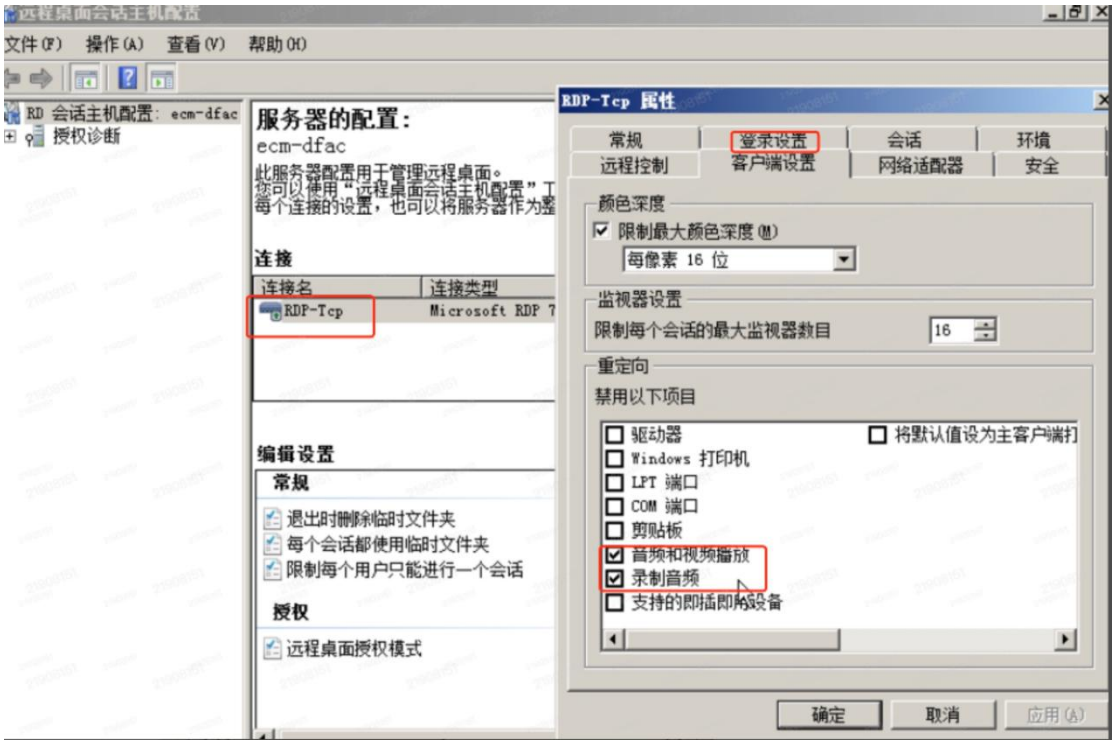
1. 启用 RDP-TCP 的“音频和视频播放”以及“录制音频”。

a. 打开“远程桌面会话主机配置”控制台。

- 打开“开始”菜单，选择“控制面板”。
- 单击右上角的“查看方式”下拉菜单，选择“类别”。
- 选择“系统和安全 > 管理工具 > 远程桌面服务 >远程桌面会话主机配置”。

b. 取消勾选“音频和视频播放”、“录制音频”。

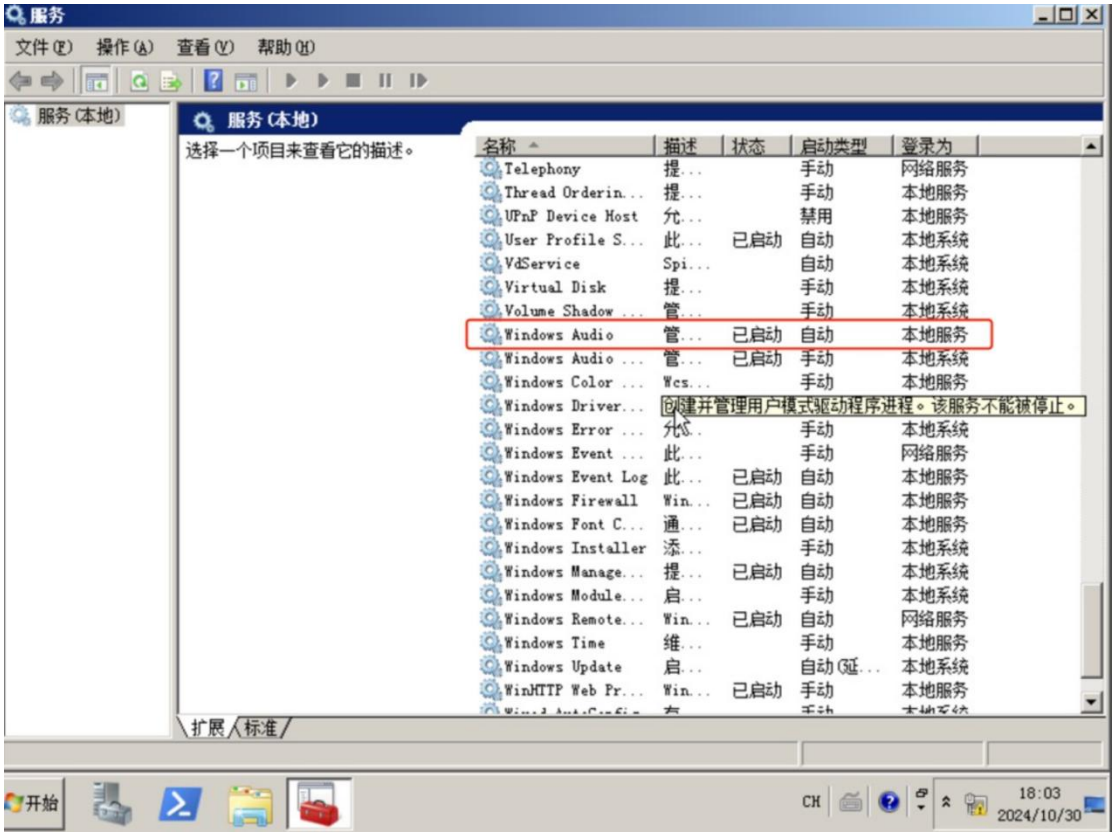
在“连接”列表里面双击“RDP-Tcp”，选择“客户端设置”，取消勾选“音频和视频播放”和“录制音频”，如下图所示。



c. 单击“确定”，激活音频设备。

重启弹性云主机并登录。

2. 重启弹性云主机后，声卡的标识依旧是显示音频服务未运行。这是因为服务未开启，操作开启音频服务后如下图所示。

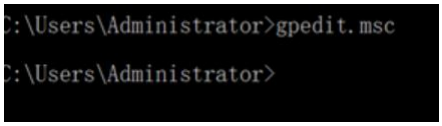


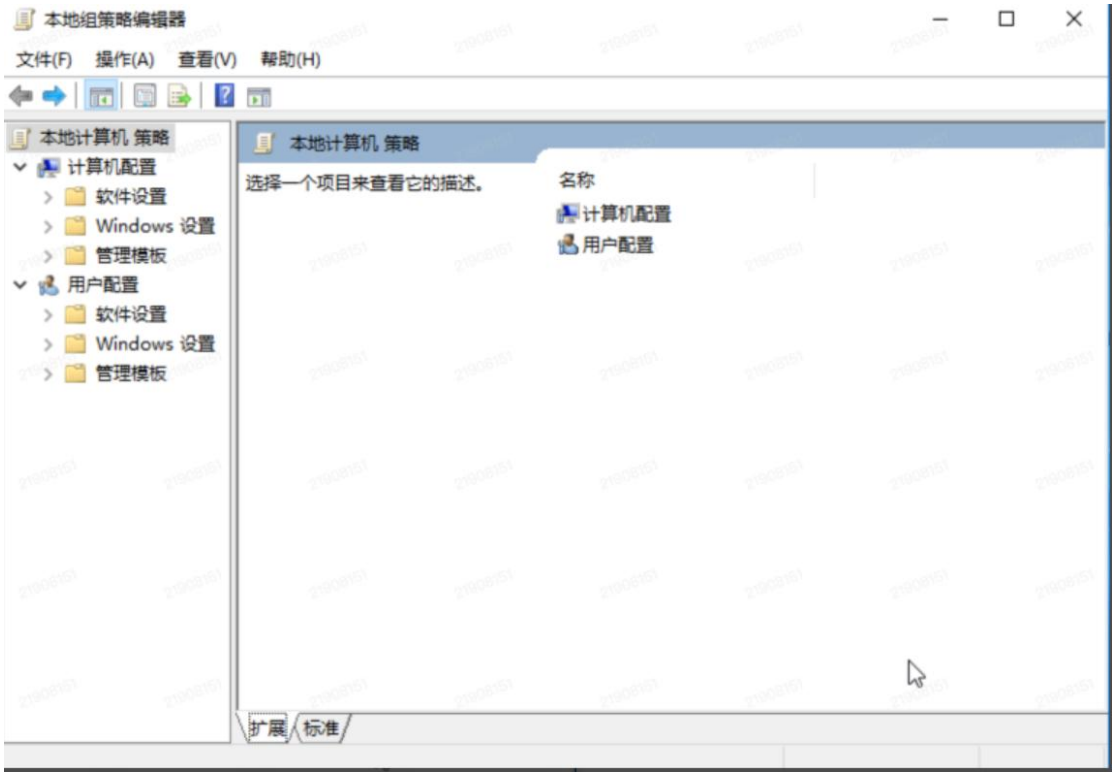
3.打开网页播放音乐，即可验证播放音频成功。

以 Windows 2016 为例，如下所示：

a. 打开“运行”窗口。

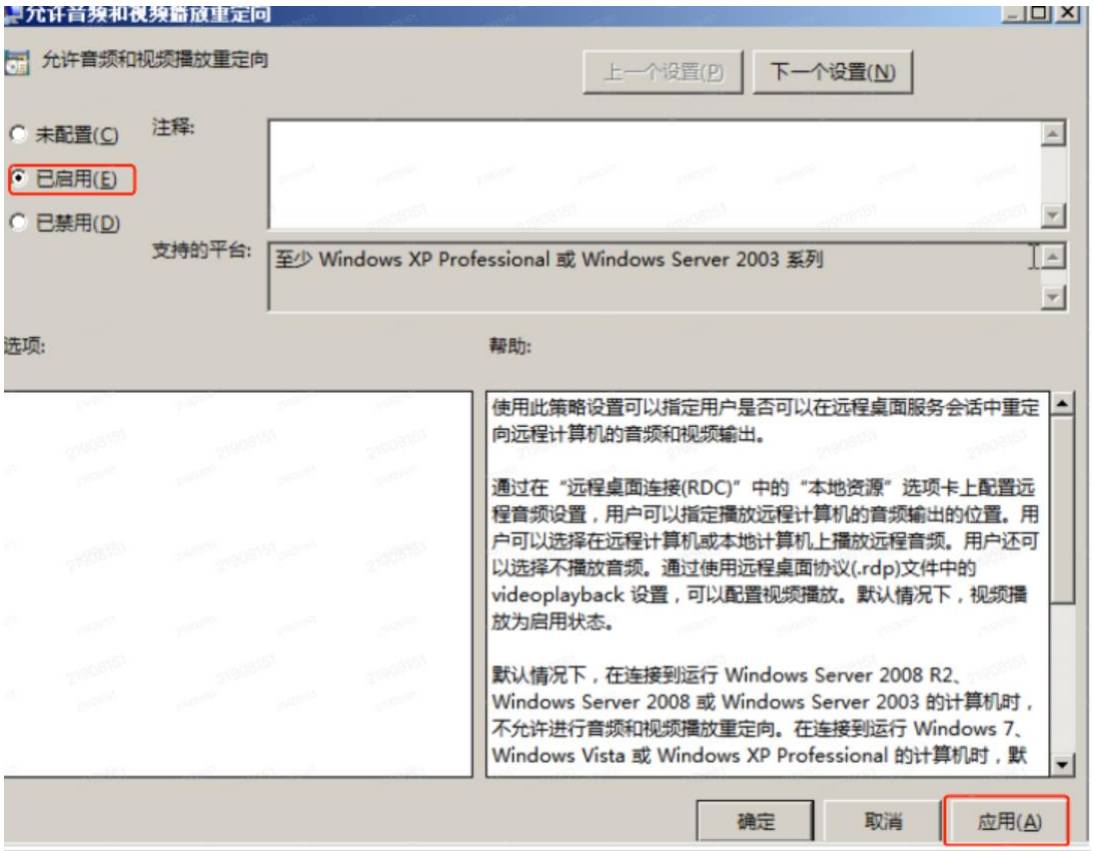
b. 输入 gpedit.msc，打开“组策略”。



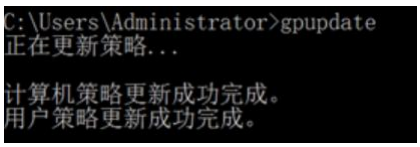


c. 依次单击“计算机配置 > 管理模版 >windows 组件 >远程桌面服务 > 远程桌面会话主机 > 设备和资源 重定向”，打开“允许音频和视频播放重定向”。

d. 选择“已启用”，按“应用”确定。



e. 打开“运行”窗口，执行以下命令，刷新组策略。



第 3 步：远程音频设置

以上配置调整完成后，打开本地远程桌面软件（mstsc），单击“选项 > 本地资源”，在远程音频处单击“设置”，弹出的选项卡中的远程音频播放处，选择“在此计算机中播放”，单击“确定”。





至此完成远程音频设置，请通过 MSTSC 登录云服务器，检查音频是否可以正常播放。

8.6.6 弹性云主机配置的邮件无法正常发送

问题描述

弹性云主机配置的邮件无法正常发送。

问题原因以及处理方法

使用邮件客户端收发邮件

问题原因：

在使用邮件客户端收发邮件时，邮件收发使用不同的协议

- 发件协议：

使用 SMTPS 协议，其端口号为 465。或使用 SMTP 协议，其端口号为 25。

Tips ：优先推荐使用 465 端口，465 端口是为 SMTP SSL（SMTP-over-SSL）协议服务开放的，这是 SMTP 协议基于 SSL 安全协议之上的一种变种协议，它继承了 SSL 安全协议的非对称加密的高度安全可靠性，可防止邮件泄露。

- 收件协议：

使用 POP3 协议，其端口号为 110。

处理方法：

添加“出方向”（协议为“TCP”，端口为“25”或“465”）和“入方向”（协议为“TCP”，端口为“110”）的规则。

操作步骤：

1. 在云主机详情页面找到对应的安全组，进入安全组列表页面，如图 1 所示。

图 1 安全组列表页



2. 在安全组页面，单机对应的安全组，点击添加规则，进行规则添加页面，如图 2 所示。

图 2 添加规则入口



3. 添加“出方向”（协议为“TCP”，端口为“25”或“465”）的规则，如图 3 所示。

图 3 “出方向”规则添加



4. 添加“入方向”（协议为“TCP”， 端口为“110”）的规则，如图 4 所示。

图 4 “入方向”规则添加



使用浏览器发送邮件

问题原因：

在使用浏览器登录邮箱时直接使用 HTTP 协议，其端口号默认为 80，邮件主机之间还是使用邮件发送协议：SMTP 协议。

处理方法：

需要开放 TCP 出方向端口为 80。

操作步骤：

1. 在云主机详情页面找到对应的安全组，进入安全组列表页面，如图 5 所示。

图 5 安全组列表页面



2. 在安全组页面，单机对应的安全组，点击添加规则，进行规则添加页面，如图 6 所示。

图 6 添加规则入口



3. 规则选择“出方向”、“TCP”协议、端口 80，点击确定完成规则添加，如图 7 所示。

图 7 添加规则页面



8.6.7 使用 free -m 命令查询弹性云主机内存大小与实际配置不符

问题描述

弹性云主机创建成功后，使用 free -m 命令查询内存大小，查询结果与实际配置不符，较之创建时的配置要小一些。

问题原因

首先，系统启动时会初始化相关设备，该过程会占用内存，内核启动时，也会占用一部分的内存。其中，kdump 占用的内存是可以自行设置的，如无特殊要求，请勿自行修改 kdump 占用的内存大小。

其次，free -m 命令查询的是弹性云主机的可用内存，dmidecode -t memory 命令查询的是实际硬件内存大小。

因此，使用 free -m 命令查询到的内存大小比实际的要小一些，属于正常情况，非问题。

8.6.8inux 虚机内关键文件路径

1. /etc

存放系统的所有配置文件。

- 关键文件 ：

o /etc/fstab：定义文件系统的挂载信息，误改可能导致系统无法挂载分区。

o /etc/passwd：存储用户账户信息，错误修改可能导致用户无法登录。

o /etc/shadow：存储用户密码的加密信息，错误修改可能破坏用户认证。

o /etc/hostname：定义主机名，修改后可能影响网络配置。

o /etc/resolv.conf：DNS 配置文件，误改可能导致域名解析失败。

2. /boot

存放系统启动相关文件，包括内核、初始化 RAM 磁盘等。

- 关键文件 ：

o /boot/vmlinuz-*：Linux 内核。

o /boot/initrd-* 或**/boot/initramfs-***：初始化 RAM 磁盘文件。

o 修改风险 ：误删或修改可能导致系统无法启动。

---

3. /bin 和 /sbin

存放基本系统命令和管理命令。

- 关键文件 ：

o /bin/bash：Shell 程序。

o /bin/ls、/bin/cp 等：基本命令。

o /sbin/init 或**/sbin/systemd**：初始化系统的程序。

o 修改风险 ：破坏基本命令可能导致系统管理和用户交互失效。

---

4. /usr

存放用户空间的应用程序和库文件。

- 关键目录 ：

o /usr/bin：用户级命令。

o /usr/sbin：高级系统命令。

o /usr/lib 和**/usr/lib64**：存储 32 位和 64 位库文件。

o 修改风险 ：影响系统和应用程序的运行。

---

5. /var

存储动态数据，包括日志、缓存和临时文件。

- 关键文件 ：

o /var/log/：系统日志文件，误删可能导致日志分析丢失。

o /var/lib/dpkg/ 或**/var/lib/rpm/**：包管理器的数据库，误改可能导致包管理器失效。

o 修改风险 ：可能导致无法记录日志或无法管理软件包。

---

6. /dev

存放设备文件，代表内核发现的硬件设备。

- 关键文件 ：

o /dev/sda、/dev/nvme0n1 等：磁盘设备文件。

o /dev/null、/dev/zero 等：特殊设备文件。

o 修改风险 ：误改可能导致硬件无法正常工作或系统无法识别设备。

---

7. /proc 和 /sys

虚拟文件系统，用于访问内核和系统信息。

- 关键文件 ：

o /proc/cpuinfo：CPU 信息。

o /proc/meminfo：内存信息。

o 修改风险 ：通常只读，擅自修改某些可写条目可能导致系统不稳定。

---

8. /root

超级用户的主目录。

- 关键文件 ：

o 配置文件如**.bashrc**、.profile。

o 修改风险 ：错误修改可能影响 root 用户的环境配置。

---

9. /lib 和 /lib64

存放系统启动和运行所需的基础库。

- 关键文件 ：

o /lib/systemd/systemd：系统初始化程序。

o 修改风险 ：修改可能导致系统无法正常启动。

---

10. /grub2 或 /boot/grub

用于存储 GRUB 引导加载程序配置。

- 关键文件 ：

o grub.cfg：GRUB 配置文件。

o 修改风险 ：配置错误可能导致系统无法引导。

建议和注意事项

备份重要文件 ：

o 在修改关键文件前，确保有完整的备份以便于出问题时快速恢复。

使用包管理器 ：

o 对于软件、库文件的安装和卸载，应尽量使用包管理器（如 yum、dnf、apt）。

权限限制 ：

o 确保非必要时不以 root 权限操作，避免意外修改关键系统文件。

谨慎配置内核和引导 ：

o 修改**/boot**、grub 等启动相关文件时，需要非常小心，并确保在修改后重新生成或验证配置。

擅自更改这些关键目录和文件可能导致严重的系统问题，如无法启动、用户无法登录、服务中断等，因此必须谨慎操作。

8.6.9 鲲鹏、海光、飞腾系列云主机挂载云硬盘和弹性网卡异常

问题描述：

鲲鹏、海光、飞腾系列云主机在新建、重启、启动等场景中，在云主机启动后立即操作挂载云硬盘和绑定弹性网卡可能会存在异常情况。

问题现象：

在云主机里新挂载的云硬盘或绑定的弹性网卡无法显示，并且无法使用。不影响云主机已有的云硬盘和弹性网卡使用。

原因分析：

操作系统启动在 5-6s 之间，因为当前 PCIE 设备初始化过程中，此时内核不能正确处理热插拔事件，如果立即热插 PCIE 设备，不能被发现加载。

解决步骤：

重启云主机即可恢复。

使用建议：

云硬盘和弹性网卡随着云主机一起创建，或者选择在云主机关机后挂载，或者等待云主机运行一段时间后挂载。

8.7 规格变更类

8.7.1 弹性云主机怎样变更规格

当您购买的弹性云主机规格无法满足业务需要时，您可以随时变更您的弹性云主机规格，升级 vCPU、内存。

在弹性云主机列表页，单击“操作”列下的“更多 > 变配”，即可变更弹性云主机的规格。

规格变更包括升配和降配：

“按量计费”模式的弹性云主机：升配和降配均立即生效，按照变更后规格的费用按需计费。

“包年/包月”计费模式的弹性云主机：

- 升配：新配置价格高于老配置价格，客户需要支付新老配置的差价。
- 降配：新配置价格低于老配置价格，天翼云会将新老配置的差价退给客户。

8.7.2 国产化云主机变配有什么限制

当前云主机产品中所提供的国产化云主机，涵盖 ARM 与 x86 两种架构，如鲲鹏、飞腾为 ARM 架构，海光为 x86 架构。

在云主机变配功能中具有以下限制：

1、X86 架构与 ARM 架构无法互相变配，即鲲鹏的 kc1.xlarge.2 无法通过变配功能变更为 hc1.2xlarge.4，也不可变更为通用云主机的规格（如 s、c、m 等系列），反之亦然。

2、不同厂商芯片的云主机不可相互变配，包含下述两种情况

（1）x86 架构中，海光系列的云主机规格无法与通用云主机（如 s、c、m 等系列）相互变配

（2）ARM 架构中，鲲鹏与飞腾系列的云主机无法相互变配。

8.7.3 主机升配后 CPU 的 ID 是否会变化

云主机执行升配操作，云主机的资源 ID 与所关联云硬盘的资源 ID 不会变化。

CPU 是云主机内部资源，在云平台无独立 id 标识。

如果您需要查看云主机 CPU 信息，可以参考下述操作：

Linux 操作系统，可以使用 cat /proc/cpuinfo 命令查看 CPU 信息。

Windows 操作系统，可以通过【控制面板】-【系统】路径查看 CPU 信息。

8.7.4Windows 弹性云主机变更规格后磁盘脱机怎么办？

故障描述

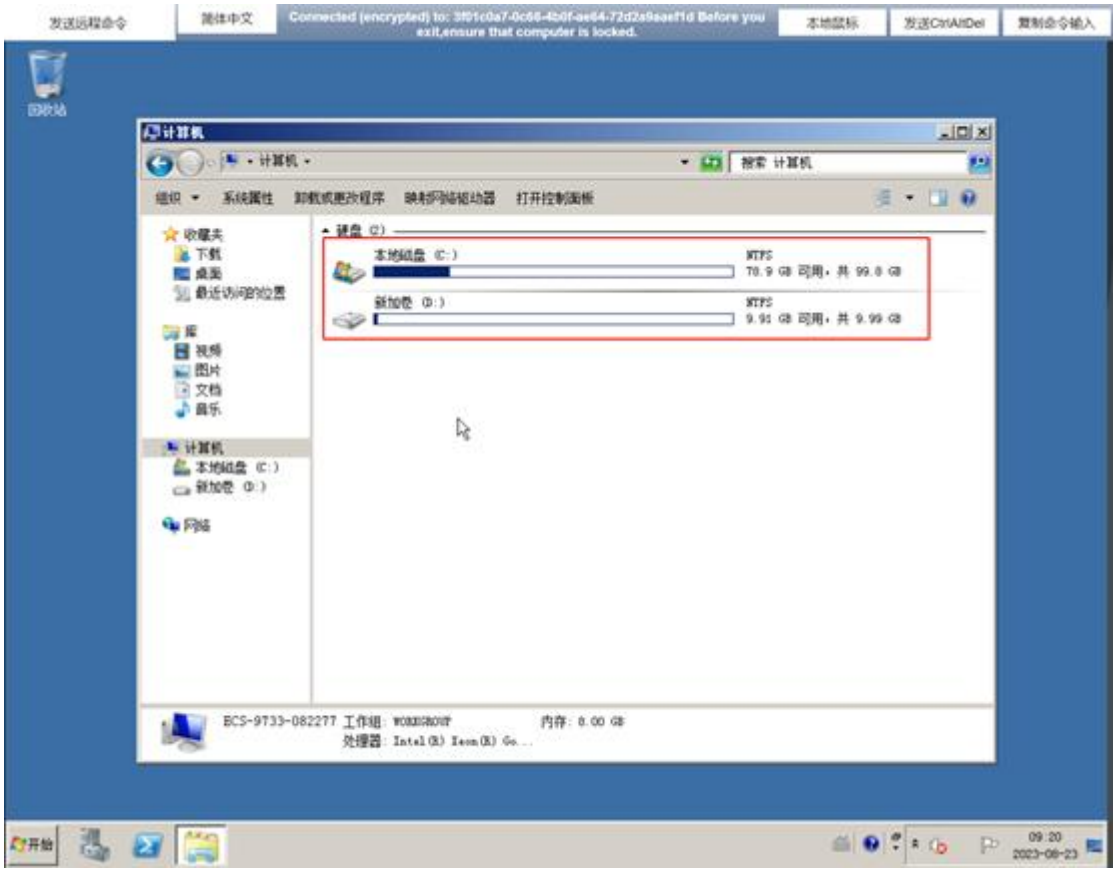
Windows 弹性云主机进行规格变更后，可能会发生磁盘挂载失败的情况从而导致磁盘处于脱机的状态。

操作步骤

1.规格变更前后在“计算机”中查询磁盘个数。

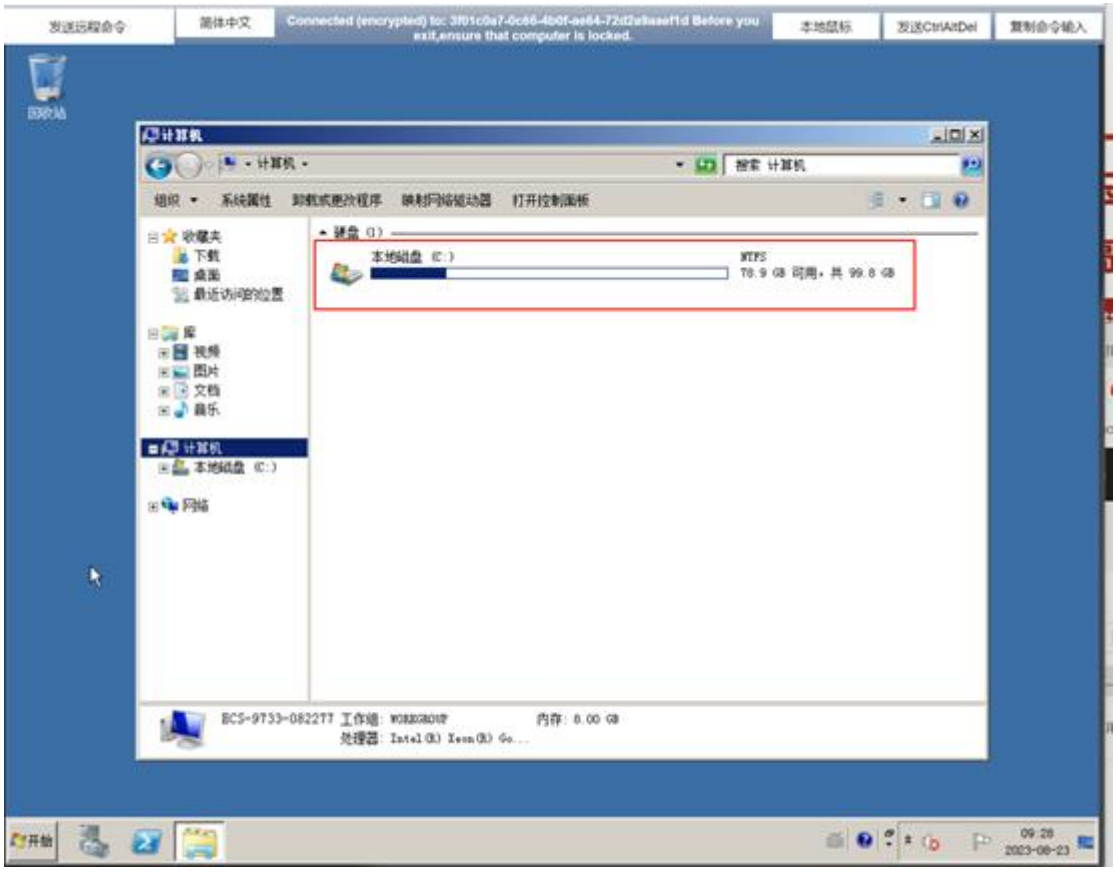
如图 1 所示，共有 2 块磁盘。

图 1 规格变更前磁盘挂载情况



如图 2 所示，只查询到 1 块系统磁盘。

图 2 规格变更后磁盘挂载情况



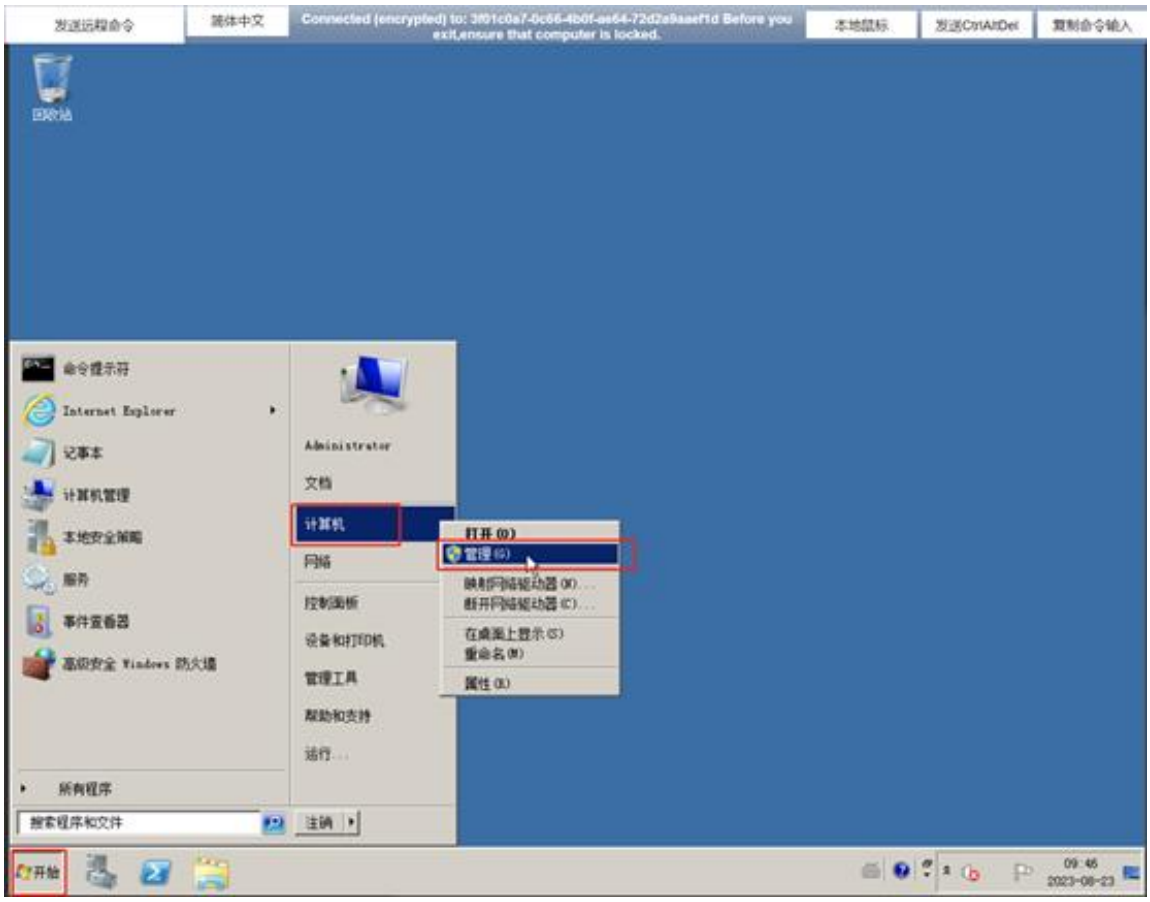
2. 对比查询到的磁盘个数。

- 如果一致，则表示云主机规格变更成功，不用执行以下操作。
- 如果不一致，表示有磁盘处于”脱机”状态。需要对所有“脱机“的磁盘分别执行下述 3 步骤。

3. 将磁盘设置为“联机“状态。

（1）选择“开始“菜单，右键单击”计算机“，并选择”管理“。如图 3 所示。

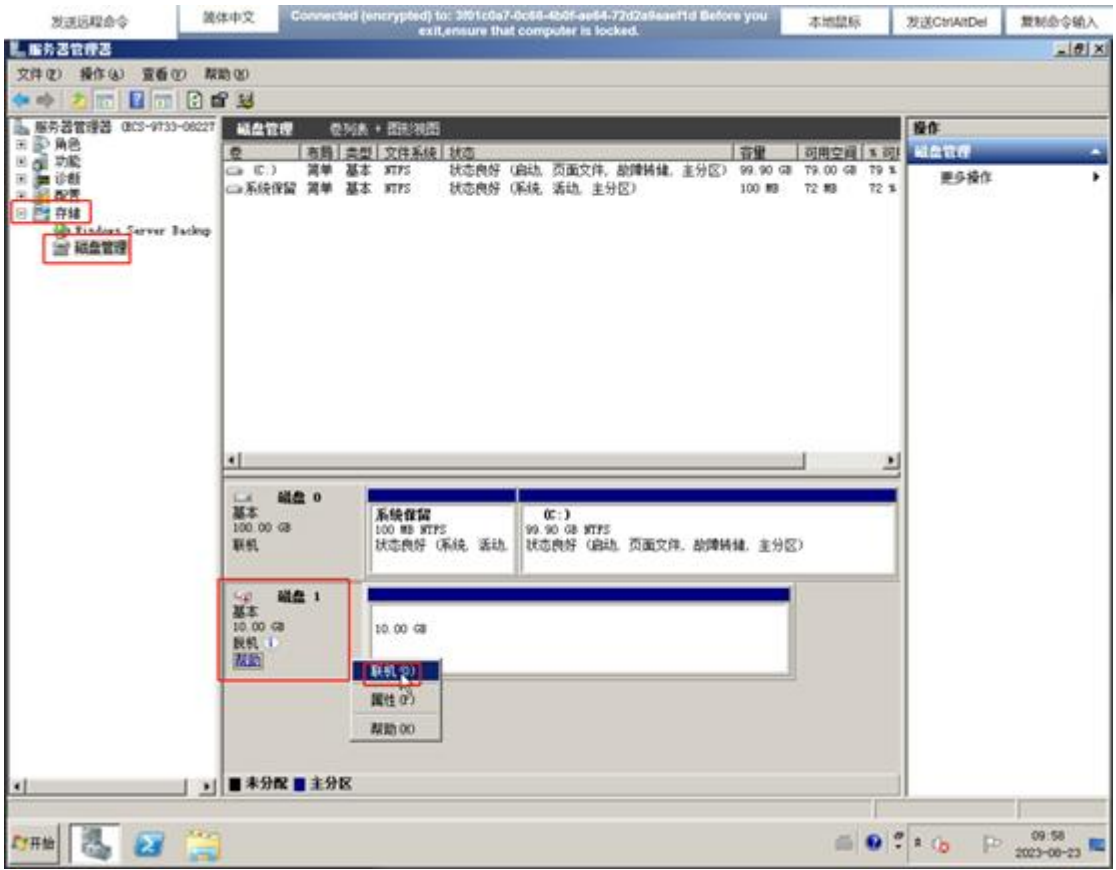
图 3 打开“管理“



（2）在左侧导航栏，选择“存储“，并选择”磁盘管理“。

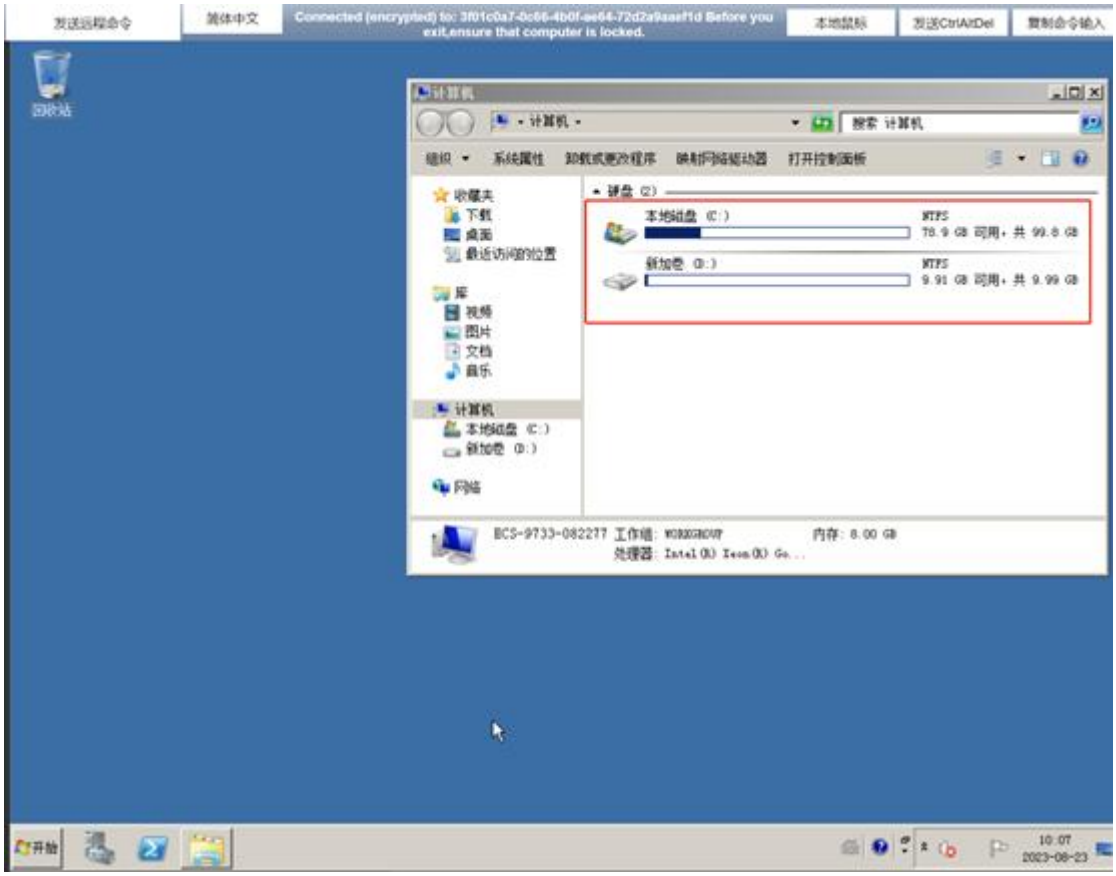
（3）右键点击处于“脱机“状态的磁盘，并选择”联机“。如图 4 所示。

图 4 执行“联机操作“



4.再次在“计算机”检查磁盘挂载情况。

图 5 检查磁盘挂载情况



如图 5 所示，磁盘数量为 2 块，和规格变更前磁盘数量一致。如果不一致您可以再次对处于“脱机”状态的磁盘进行“联机”操作。如果“联机”失败，您可以联系客服获取技术支持。

8.7.5Linux 弹性云主机变更规格后磁盘脱机怎么办？

故障描述

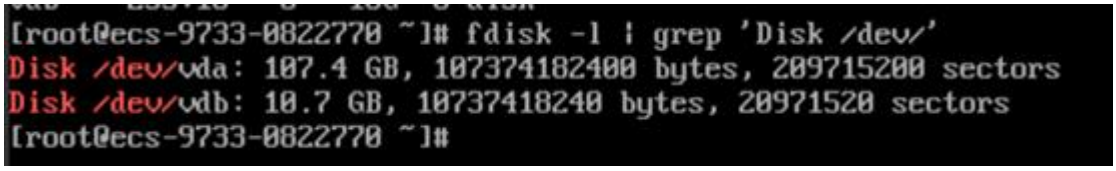
Linux 弹性云主机进行变配后，可能会发生磁盘挂载失败的情况从而导致磁盘处于脱机的状态。

操作步骤

1.请确保使用 root 用户登录弹性云主机。

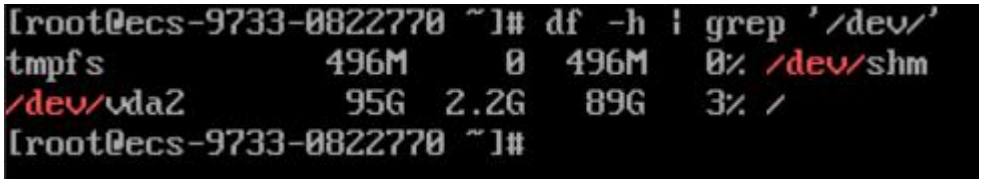
2.首先通过执行 fdisk -l | grep 'Disk /dev/' 命令查询云主机的磁盘信息。之后执行 df -h | grep '/dev/' 命令，查看磁盘挂载情况。

图 1 查询云主机磁盘信息



如图 1 所示，云主机共有 2 块磁盘：/dev/vda、/dev/vdb。

图 2 查询磁盘挂载情况



如图 2 所示，云主机置挂载了 1 块磁盘/dev/vda。

3.检查上述步骤中查询到的磁盘个数和已挂载的磁盘个数是否一致。

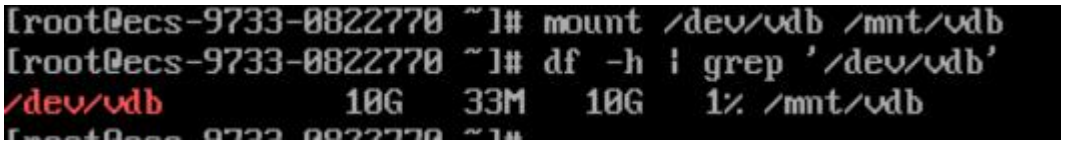
- 如果一致，则表示云主机规格变更成功，不用执行以下操作。
- 如果不一致，则需要对挂载失败的所有磁盘分别执行步骤 4。

4.使用 mount 命令挂载磁盘，并用 df 命令检查磁盘挂载情况。

挂载示例：mount /dev/vdb /mnt/vdb

其中/dev/vdb 是待挂载的磁盘；/mnt/vdb 是待挂载磁盘的文件夹路径， 此文件夹必须是空文件夹 ，否则挂载磁盘会失败。

图 3 挂载磁盘并检查



如图 3 所示，/dev/vdb 磁盘已经挂载成功。如果挂载失败您可以联系客服获取技术支持。

5.再次通过执行 fdisk -l | grep 'Disk/dev/' 和 df -h | grep '/dev/' 命令，对比查询到的磁盘个数和已挂载的磁盘个数是否一致。

图 4 检查磁盘个数是否一致

```
[root@ecs-9733-0822770 ~]# fdisk -l | grep 'Disk /dev/'
Disk /dev/vda: 107.4 GB, 107374182400 bytes, 209715200 sectors
Disk /dev/vdb: 10.7 GB, 10737418240 bytes, 20971520 sectors
[root@ecs-9733-0822770 ~]# df -h | grep '/dev/'
tmpfs           496M     0  496M   0% /dev/shm
/dev/vda2        95G  2.2G   89G   3% /
/dev/vdb         10G   33M   10G   1% /mnt/vdb
```

如图 4 所示，检查到磁盘个数一致。如果不一致您可以执行步骤 4 或者联系客服获取技术支持。

8.7.6 弹性云主机变配操作无法成功提交

问题描述

对已关机的弹性云主机进行变配操作时，无法正常提交。

处理方法

弹性云主机进行变配操作，请确保在弹出的变配页面中选择与当前弹性云主机不同的云主机规格。如果您弹性云主机当前规格是 s7.small.1，再次选择同样规格，提示如图所示：



选择您的目标弹性云主机规格，还需要符合如下限制条件约束：

1.x86 计算架构和 ARM 计算架构的云主机不可互相变配 。

2.x86 计算架构下海光云主机不可与其他云主机变配 。

3.ARM 计算架构中，基于不同厂商芯片的云主机不可互相变配 。

4.只能向同代或更高代云主机变配。

当变配操作符合规则成功提交后，如果您是包年包月的弹性云主机，请您确保升配订单被成功支付，再次返回云主机控制台即可查看到弹性云主机状态变为“规格变更中”，稍后刷新列表，可以看到弹性云主机规格变更完成。

8.8 操作系统管理

8.8.1 图形化系统安装类

8.8.1.1 云主机是否有图形界面

Windows 操作系统是桌面管理，有图形化界面。

Linux 操作系统是命令行，用户如果需要可以自己设置图形管理。

弹性云主机安装图形化界面前，请确保云主机内存不小于 2GB，否则可能出现图像化界面安装失败，或安装后无法启动的问题。

8.8.1.2CentOS 7 系列弹性云主机如何安装图形化界面

约束与限制

弹性云主机安装图形化界面前，请确保云主机内存不小于 2GB，否则可能出现图像化界面安装失败，或安装后无法启动的问题。

操作步骤

本文操作以 CentOS7.9 64 位弹性云主机为例。

1.登录弹性云主机。

2.执行以下命令，安装图形桌面组件。

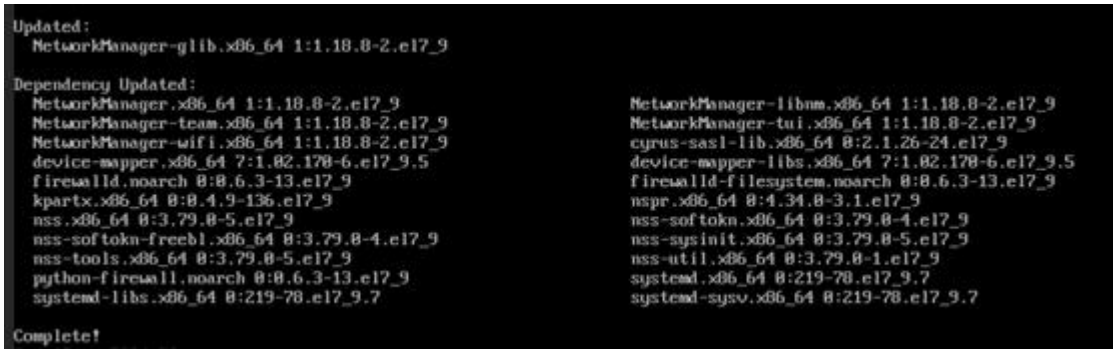
```
# yum groupinstall "Server with GUI"
```

说明

说明：安装前请确保云服务器可以连通互联网，可以通过申请并绑定弹性 IP 连通互联网。



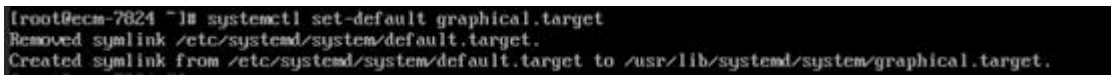
安装完成示例：



3.安装完成后，执行以下命令设置默认启动级别为 graphical.target。

```
#systemctl set-default graphical.target
```

设置命令示例：



4.执行以下命令启动 graphical.target。

```
systemctl start graphical.target
```

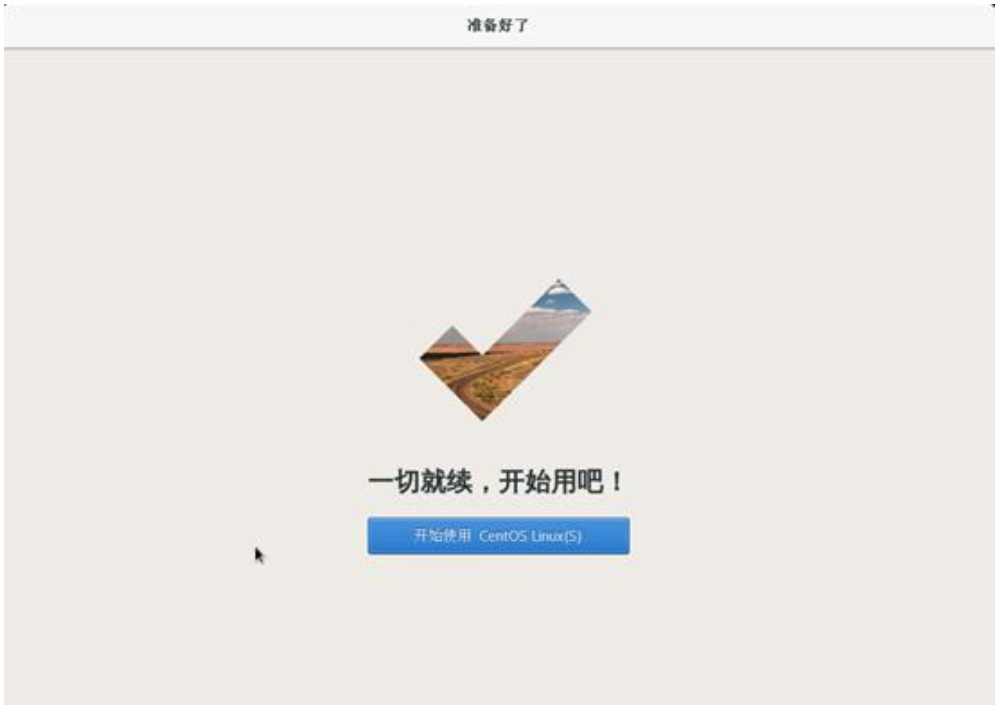
5.重启服务器。

通过控制台提供的远程登录方式连接云主机，并按照桌面启动的提示设置语言、时区、用户名及密码等。

连接成功后，设置示例：



配置成功示例：



8.8.1.3CentOS 8 系列弹性云主机如何安装图形化界面

约束与限制

弹性云主机安装图形化界面前，请确保弹性云主机内存不小于 2GB，否则可能出现图像化界面安装失败，或安装后无法启动的问题。

操作步骤

本文操作 CentOS8.4 64 位弹性云主机为例。

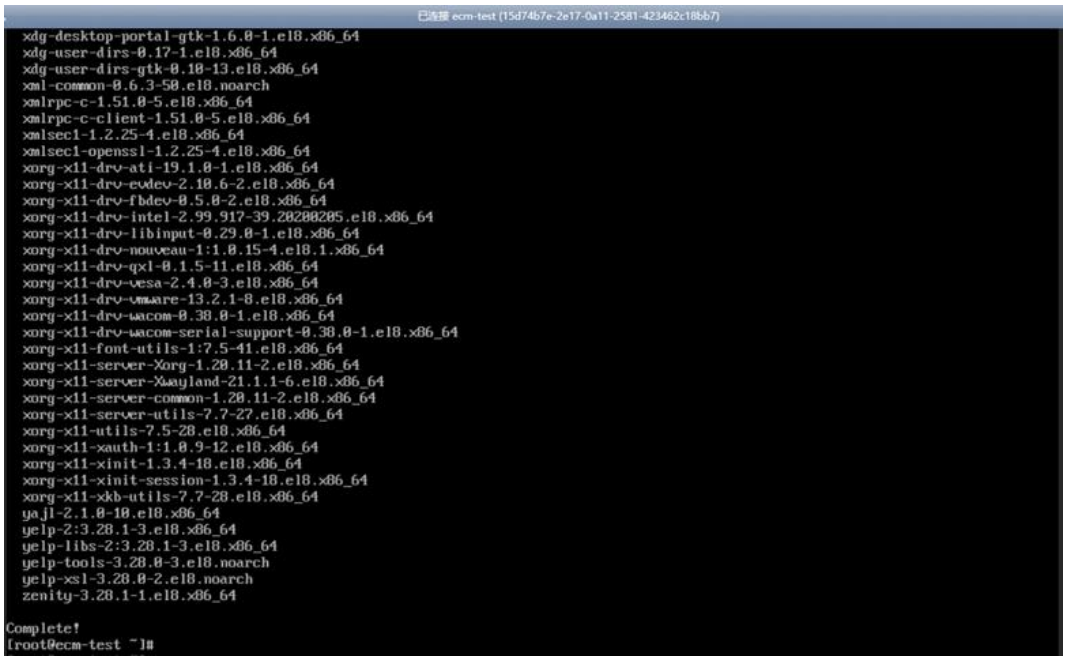
1. 登录弹性云主机。
2. 执行以下命令，安装图形桌面组件。

```
# yum groupinstall "Server with GUI"
```

说明：
安装前请确弹性云主机可以连通互联网，可以通过申请并绑定弹性 IP 连通互联网。



安装完成示例：



3. 安装完成后，执行以下命令设置默认启动级别为 graphical.target。

```
# systemctl set-default graphical.target
```

设置命令示例：

```
[root@ecm-test ~]# systemctl set-default graphical.target
Removed /etc/systemd/system/default.target.
Created symlink /etc/systemd/system/default.target → /usr/lib/systemd/system/graphical.target.
[root@ecm-test ~]#
```

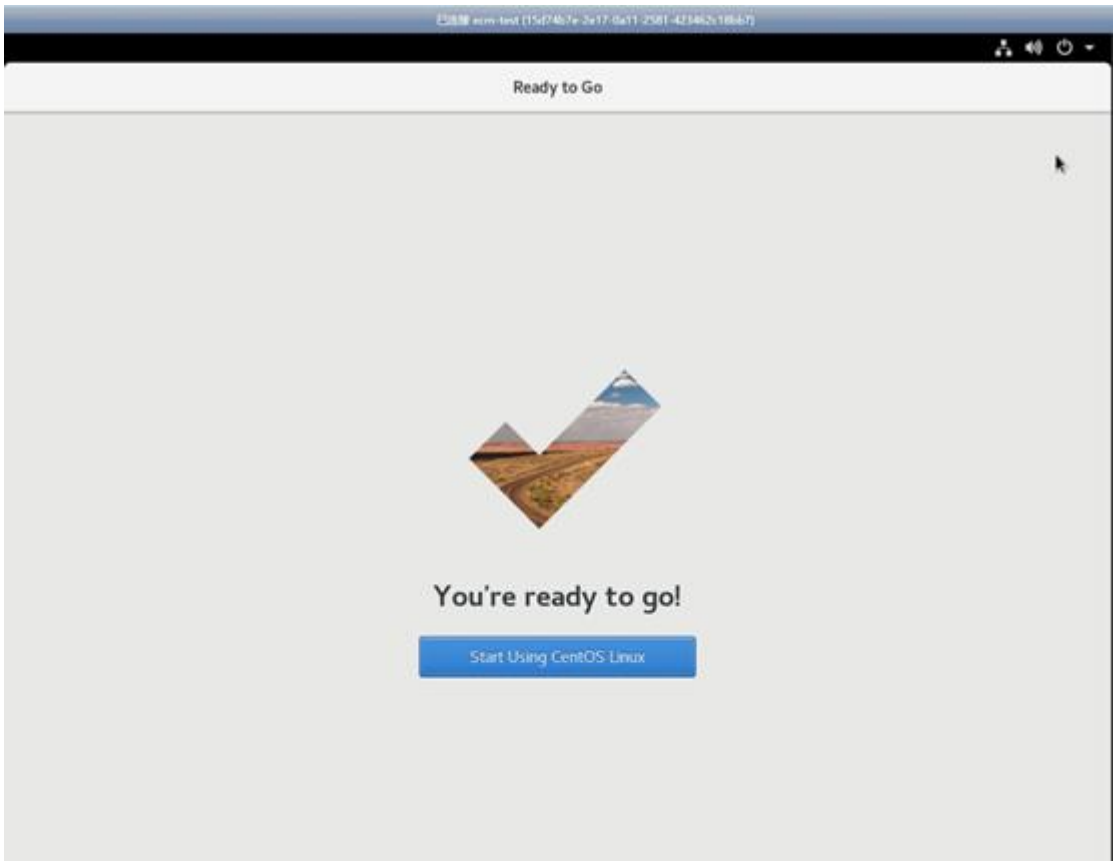
4.执行以下命令启动 graphical.target。

```
# systemctl start graphical.target
```

5.重启弹性云主机。

通过控制台提供的远程登录方式连接主机，并按照桌面启动的提示设置语言、时区、用户名及密码等。

配置成功示例：



8.8.1.4Ubuntu 系列弹性云主机如何安装图形化界面

约束与限制

弹性云主机安装图形化界面前，请确保云主机内存不小于 2GB，否则可能出现图像化界面安装失败，或安装后无法启动的问题。

操作步骤

本文操作以 Ubuntu 22.04 64 位弹性云主机为例。

1. 登录弹性云主机。

2. 执行以下命令，安装 Ubuntu 图形桌面组件。

```
# apt-get update
```

- 对于 Ubuntu16.04 版本，请执行以下命令：

```
# apt-get install -y scite xorg xubuntu-desktop
```

- 对于 Ubuntu18.04、20.04 和 22.04 版本，请执行以下命令：

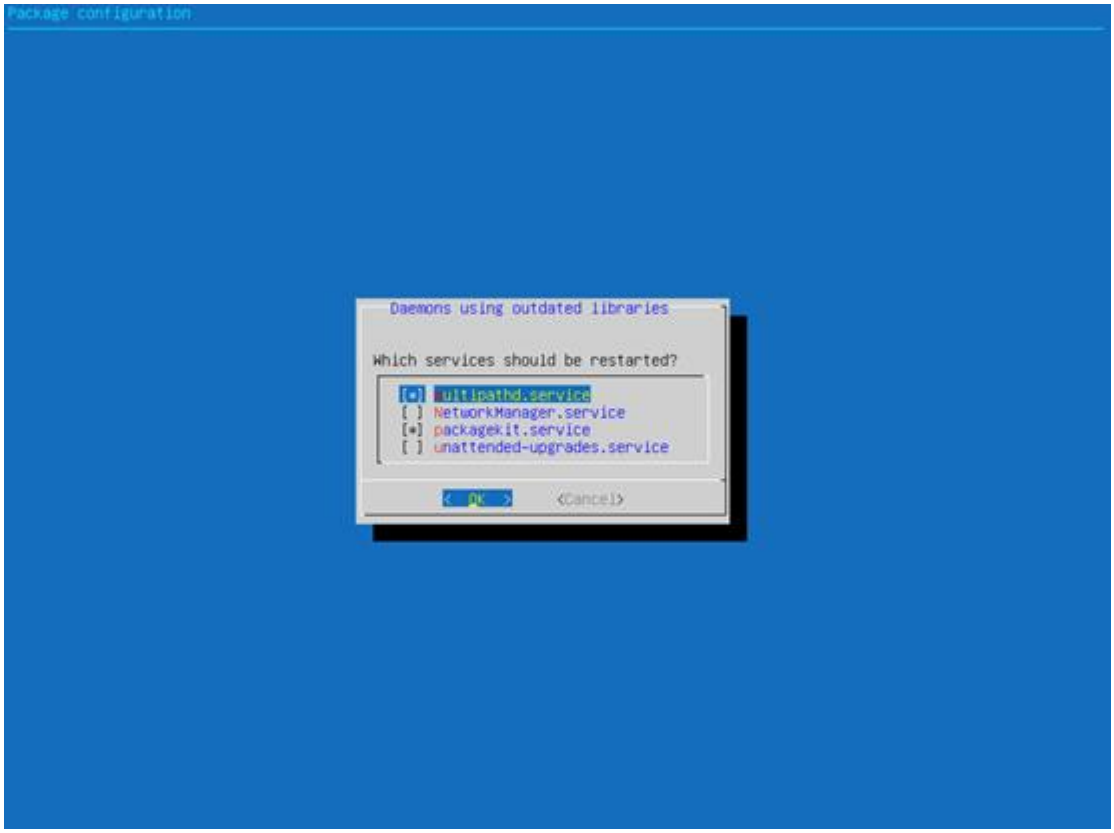
```
# apt-get install -y ubuntu-desktop
```

说明

说明：安装前请确保云主机可以连通互联网，可以通过申请并绑定弹性 IP 连通互联网。



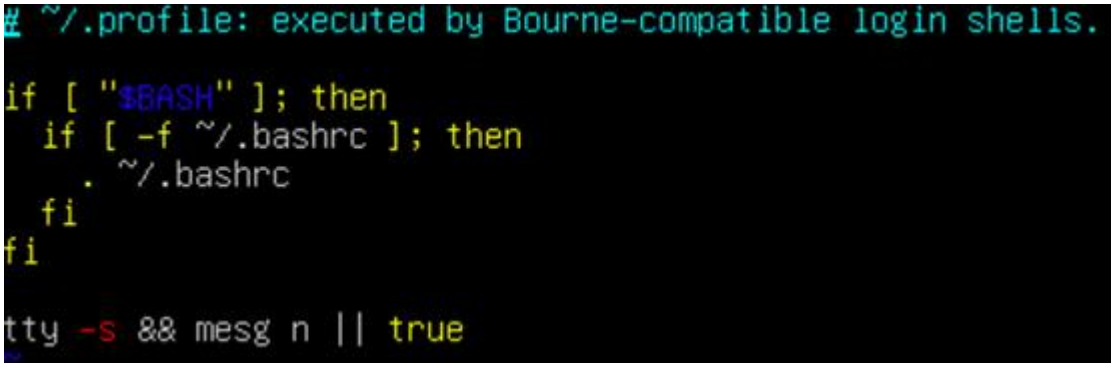
安装完成示例：



3.安装完成后，编辑/root/.profile 文件。

```
vim /root/.profile
```

修改末尾行 mesg n || true 为 tty -s && mesg n || true，修改后该文件内容如下：



4.单击“Esc”退出编辑模式。

5.执行以下命令，保存并退出文件。

```
# :wq
```

6.(Ubuntu 20.04 操作系统必选)添加子账号。

安装图形化桌面后，Ubuntu 20.04 操作系统禁止 root 登录，因此需要在安装完成后添加子账号用于登录图像化桌面。

以添加 user01 用户为例，执行以下命令。

```
# adduser user01
```

根据提示设置新用户的密码。

```
Adding user `user01' ...
Adding new group `user01' (1001) ...
Adding new user `user01' (1001) with group `user01' ...
Creating home directory `/home/user01' ...
Copying files from `/etc/skel' ...
New password:
BAD PASSWORD: The password fails the dictionary check - it is based on a dictionary word
Retype new password:
passwd: password updated successfully
```

设置 user01 的其他信息，该信息均为可选，如果不设置可以按回车键跳过，最后系统将提示您确认输入的信息是否正确无误。

正确输入"Y"。

```
Changing the user information for user01
Enter the new value, or press ENTER for the default
        Full Name []:
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
```

7. 执行 reboot 命令，重启服务器。

8. 通过控制台提供的 VNC 登录方式连接主机，使用 ctyun 帐号或 6 中添加的子账号登录图形化界面。



8.8.1.5Debian 系列弹性云主机安装图形化界面流程

操作场景

为给用户提供纯净的云主机系统，目前使用的 Debian 系列弹性云主机默认没有安装图形化界面，若有相关需要，详见本篇内容进行安装。

约束与限制

本文所用系统为 Debian 9.0.0 64 位。

安装图形化界面前请确保云主机内存不小于 2GB，以免出现安装失败等问题。

操作步骤

1.远程登陆云主机后，执行以下命令对软件库进行更新和升级。

```
apt update

apt upgrade
```

```
root@debian-test1:~# apt update
Ign:1 http://archive.debian.org/debian stretch InRelease
Hit:2 http://archive.debian.org/debian stretch-backports InRelease
Hit:3 http://archive.debian.org/debian-security stretch/updates InRelease
Hit:4 http://archive.debian.org/debian stretch Release
Reading package lists... Done
Building dependency tree
Reading state information... Done
All packages are up to date.
root@debian-test1:~# apt upgrade
Reading package lists... Done
Building dependency tree
Reading state information... Done
Calculating upgrade... Done
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
root@debian-test1:~# _
```

2. 若没有安装 tasksel 执行以下命令进行安装。

```
apt install tasksel
```

3. 执行以下命令使用 tasksel 安装 gnome 图形化界面并等待其安装完成。

```
tasksel install desktop gnome-desktop
```

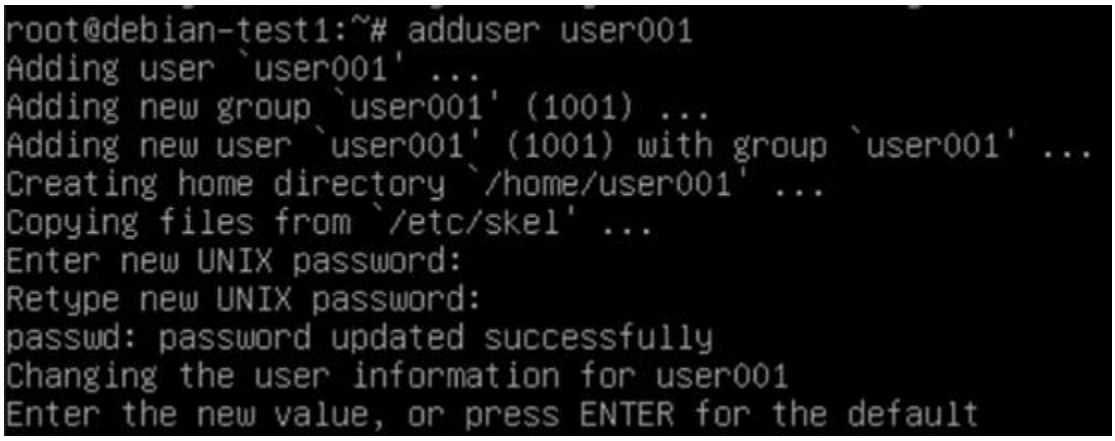


4. 执行以下命令设置图形化界面为默认启动。

```
systemctl set-default graphical.target
```

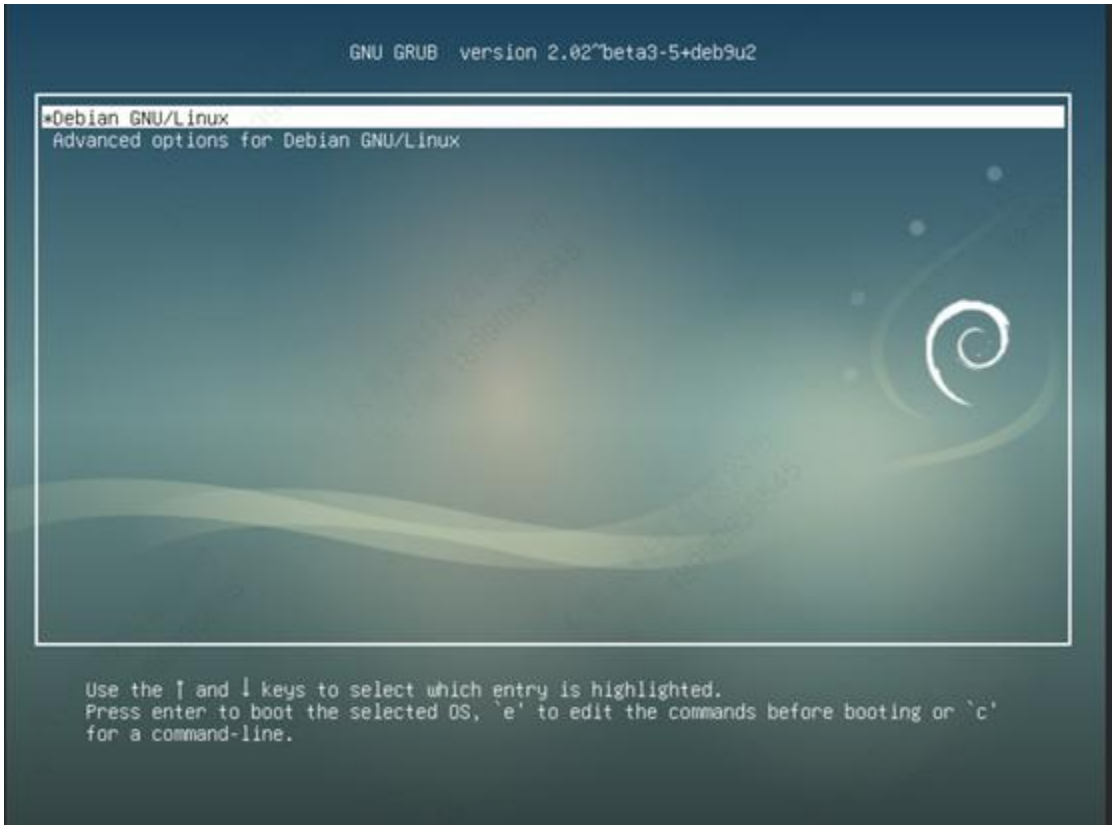
5. 图形化桌面安装后系统禁止 root 用户登陆，需要新添加子账号用于登陆图像化桌面，此处以 user001 为例。

```
adduser user001
```



除密码外后续信息均可回车跳过，最后输入“Y”确定。

6. 执行 reboot 命令重启云主机并使用步骤 5 中创建的子账号登陆即可完成。



8.8.2 操作系统故障类

8.8.2.1 针对 Intel 处理器芯片存在的 Meltdown 和 Spectre 安全漏洞，应该如何规避？

问题描述

2018 年 1 月 3 日，Intel 处理器芯片被曝出存在严重的 Meltdown 和 Spectre 安全漏洞，漏洞详情如下：

漏洞名称：Intel 处理器存在严重芯片级漏洞

漏洞编号：CVE-2017-5753、CVE-2017-5715、CVE-2017-5754

漏洞级别：严重

漏洞描述：攻击者可以利用 Meltdown （CVE-2017-5754） 和 Spectre （CVE-2017-5715/CVE-2017-5753）

绕过内存安全隔离机制，越权访问操作系统和其他程序的核心数据，造成敏感信息泄露。

问题影响

该漏洞不会引起不同弹性云主机之间的攻击，但可能会引起如下问题：

- 弹性云主机内多个应用之间，可能存在攻击。
- 对于同一弹性云主机，多个帐号之间可能存在攻击。

使用公共镜像的弹性云主机，云平台会对公共镜像依次修复，不会对您的业务带来影响。

使用私有镜像的弹性云主机，请根据漏洞影响评估是否更新补丁，以规避风险，更新补丁的具体操作请参见本节内容。

前提条件

为避免发生意外，修复漏洞前，建议进行充分测试，并完成弹性云主机的数据备份操作，避免发生意外。

Windows 弹性云主机处理方法

1.登录弹性云主机。

2.更新补丁。

方式一：使用 Windows 自动更新功能安装补丁。

a. 打开 Windows Update，并单击“检查更新”。

b. 根据需要下载安装相关安全补丁。

方式二：手动下载补丁并安装。

下载官方发布的补丁进行安装，如 KB4078130。

3.重启弹性云主机，使补丁生效。

4.验证是否升级成功。

a. 检查系统运行情况是否正常。

b. 检查已安装的补丁清单是否满足微软官方的要求。

Linux 弹性云主机处理方法

1.登录弹性云主机。

2.更新补丁，升级 kernel 内核。

说明

升级 kernel 内核后，请务必执行 reboot 命令重启弹性云主机。

3.验证是否升级成功。

检测 Linux 操作系统安全漏洞是否已修补完成

1. 在 github 网站搜索 spectre-meltdown-checker，获取 spectre-meltdown-checker.sh 检测脚本。

2.将步骤 1 获取的脚本上传至云主机。

3.在云主机执行以下命令，并根据脚本提示判断 Meltdown 或 Spectre 漏洞是否已经修复。

```
chmod +x spectre-meltdown-checker.sh

sudo bash spectre-meltdown-checker.sh
```

回显信息如图 1 所示。

图 1 执行脚本后的回显信息



OK 为已修复漏洞,KO 为未修复,如图 1 所示代表 CVE-2017-5753、CVE-2017-5715、CVE-2017-5754 漏洞均已修复。

打开或关闭 Linux 操作系统的安全漏洞补丁开关

CPU 的预测执行是一种性能优化技术，因此修复 Meltdown 或 Spectre 漏洞后可能导致在特定工作负载下的性能下降。

如果您希望可以禁用部分或全部漏洞安全保护策略以避免漏洞修复带来的性能下降，那么可以参考以下操作启用或者禁用安全保护策略。

您可以根据如下具体情况配置系统来达到理想的安全策略：

- Meltdown 漏洞

采取页表隔离 pti（Page Table Isolation）来控制内核页表隔离功能，此功能适用于 CVE-2017-5754。

- Spectre 漏洞

采取间接分支限制预测 ibrs（Indirect Branch Restricted Speculation）控制 SPEC_CTRL 模型特定寄存器（MSR）中的 IBRS 功能，结合 retpoline，及间接分支预测障碍 ibpb（Indirect Branch Prediction Barriers）控制 PRED_CMD 模型特定寄存器（MSR）中的 IBPB 功能，此功能适用于 CVE-2017-5715。

说明

CVE-2017-5753 漏洞是通过内核补丁修复的，它无法禁用，并且它在 Red Hat 的性能测试中没有显示出任何可见的影响。

- 关闭 Meltdown 安全漏洞补丁

如果您认为 Meltdown 漏洞修复对系统的性能影响不可接受，或者您有更好的保护机制，可根据以下步骤操作：

1.根据不同的操作系统修改内核参数：

a. CentOS、EulerOS、Ubuntu、Fedora、Red Hat：添加内核参数 nopti

b. Debian、OpenSUSE：添加内核参数 pti=off

2.重启云主机。

- 关闭 Spectre 安全漏洞补丁

如果您认为 Spectre 漏洞修复对系统的性能影响不可接受，或者您有更好的保护机制，可根据以下步骤操作：

1.根据不同的操作系统修改内核参数：

a. CentOS、EulerOS、Fedora、Debian、Red Hat、OpenSUSE：添加内核参数 spectre_v2=off

b. Ubuntu：添加内核参数 nospectre_v2=off

2.重启云主机。

如果您使用的是第三方操作系统，可以前往对应官网查询更多信息。

7.8.2.2 如何开启 CentOS 操作系统的 SELinux 功能？

操作步骤

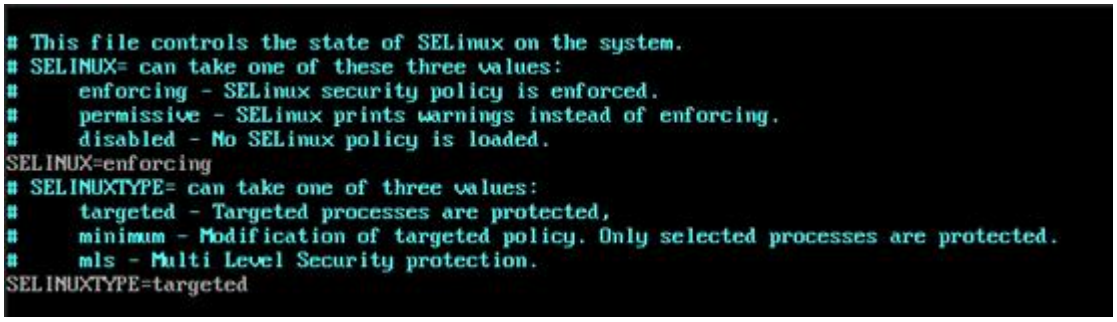
本文操作以 CentOS7.6 操作系统弹性云主机为例。

1.登录弹性云主机。

2.输入 sestatus 命令以检查当前 SELinux 功能是否开启。



3. 执行 vim/etc/selinux/config 命令，将 SELinux 配置文件中的“SELINUX=disabled”修改为“SELINUX=enforcing”。



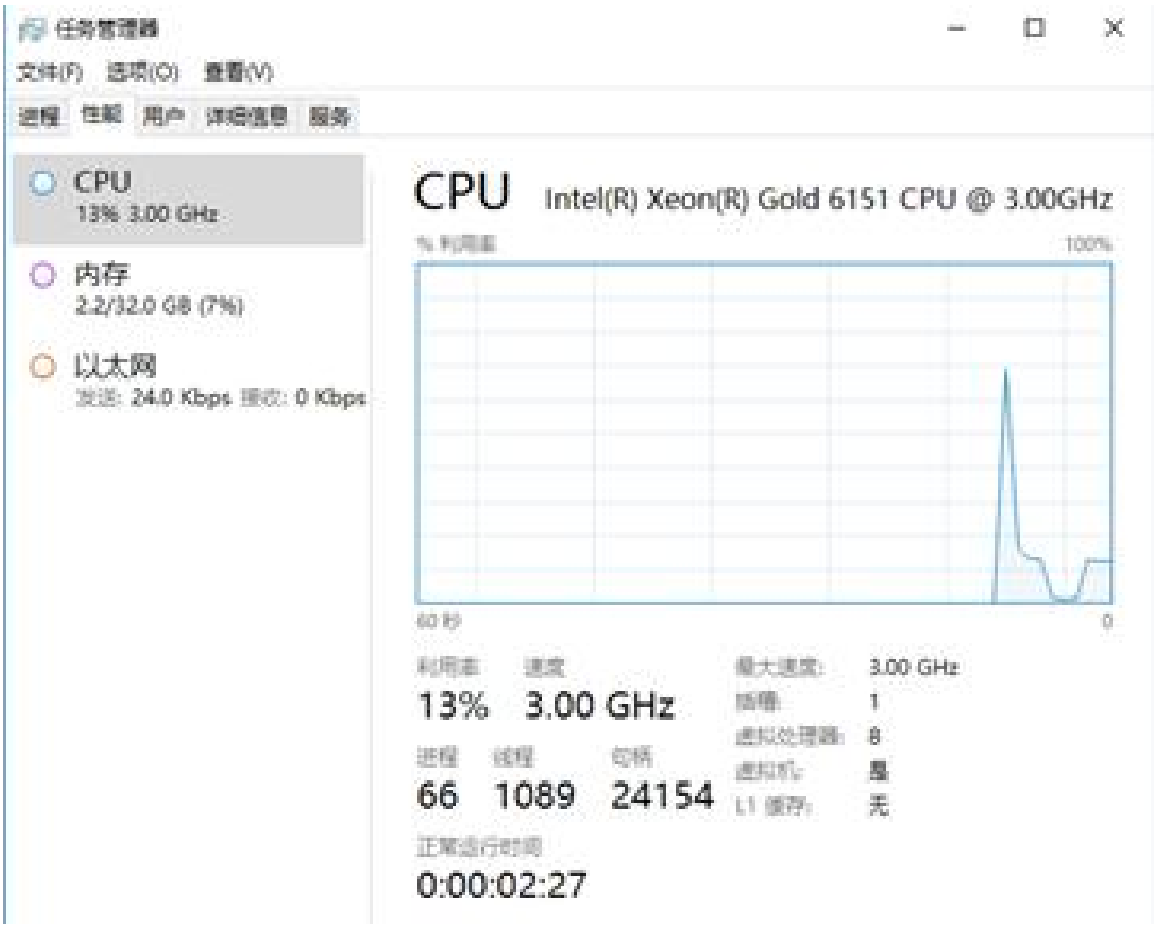
4.执行 touch /.autorelabel 命令，使系统重新启动时自动为文件系统打上 selinux 标签。

5. 执行 reboot 命令，重启弹性云主机，使配置生效。

8.8.2.3 怎样查看 GPU 加速型云主机的 GPU 使用率？

问题描述

Windows Server 2012 和 Windows Server 2016 操作系统的 GPU 加速型云主机无法从任务管理器查看 GPU 使用率。



本节以介绍两种查看 GPU 使用率的方法：

1. 在 cmd 窗口执行 nvidia-smi 命令查看 GPU 使用率。
2. 通过安装 GPU-Z 工具查看 GPU 使用率。

前提条件

GPU 加速型云主机已安装 NVIDIA 驱动。

方法一

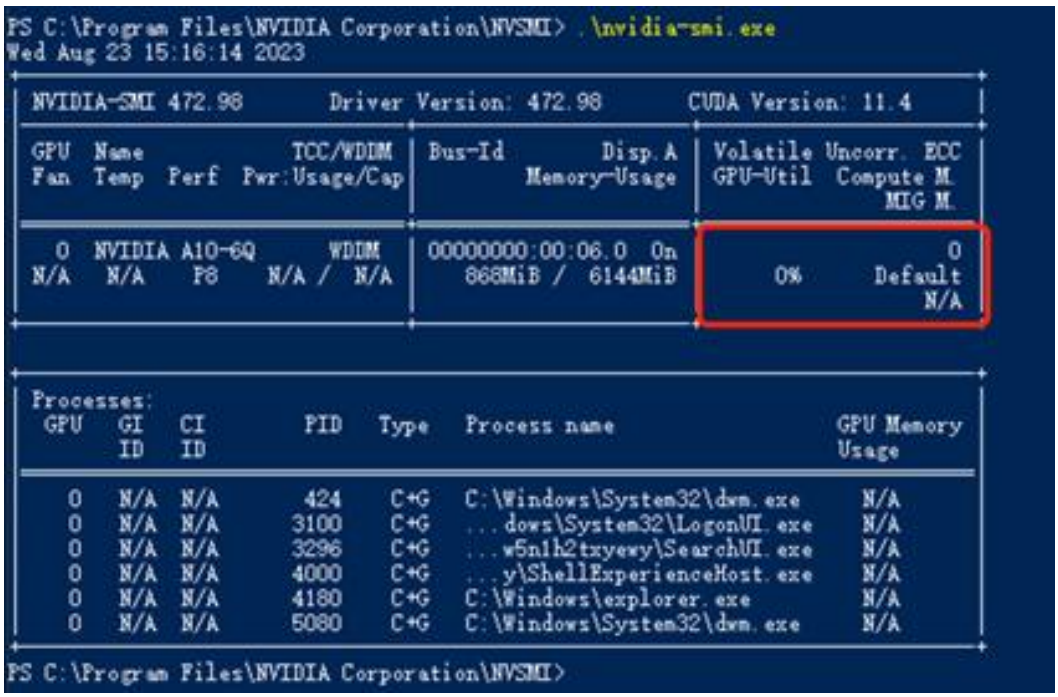
1. 登录 GPU 加速型云主机。
2. 打开 cmd 窗口或 powershell 窗口。
3. 执行下列命令，查看 GPU 使用率。

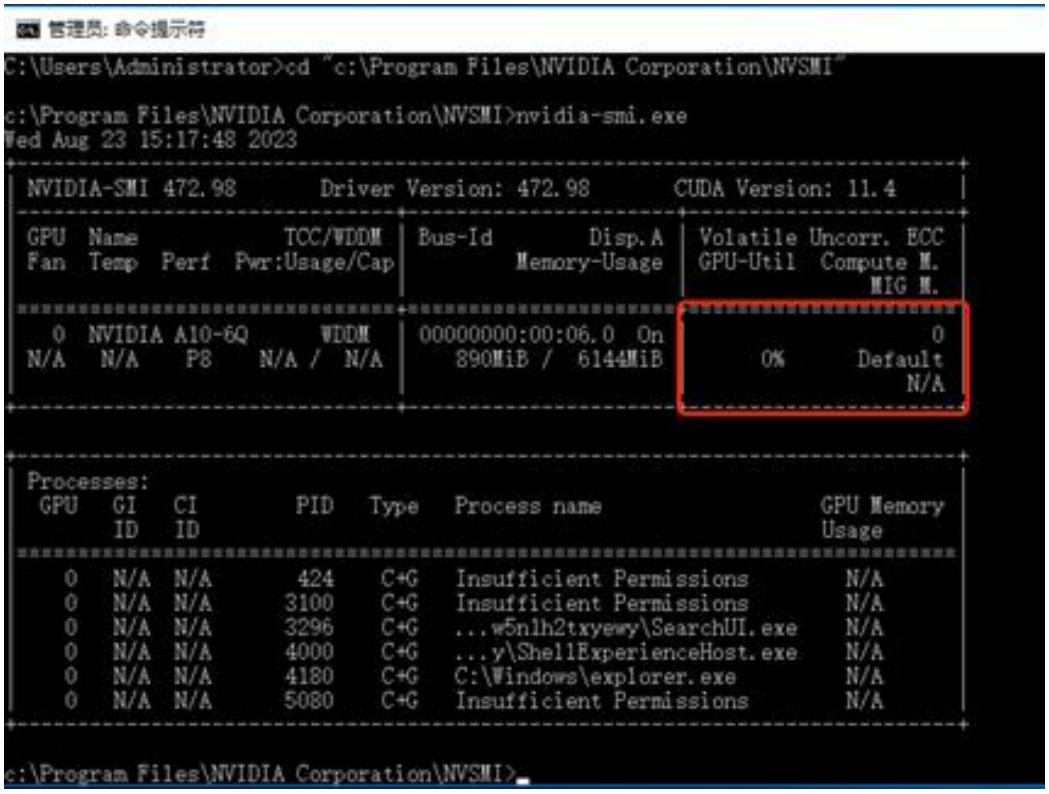
```
cd C:\Program Files\NVIDIA Corporation\NVSMI

nvidia-smi
```

如果需要持续查看 GPU 使用情况，可以执行以下命令。

```
nvidia-smi -l 1
```



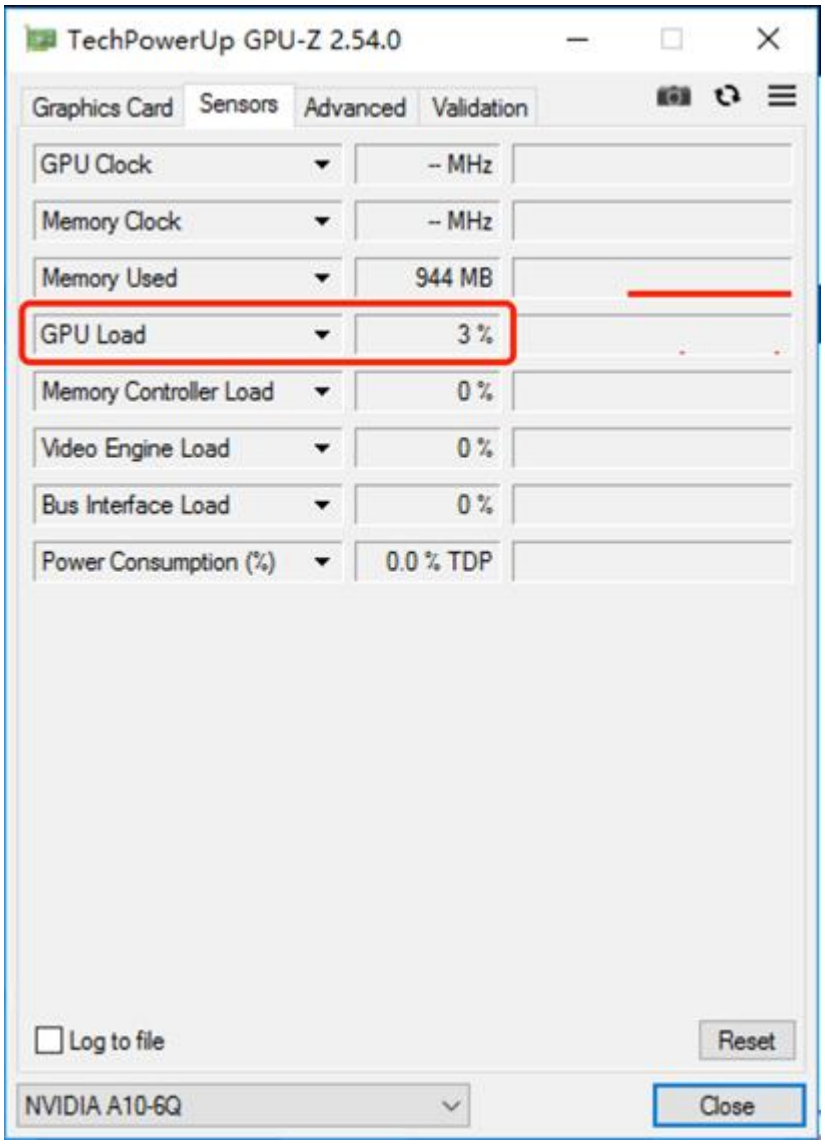


说明

- NVIDIA GPU 可以配置为 TCC（Tesla Compute Cluster）模式或 WDDM（Windows Display Driver Model）模式。
- TCC 模式下，GPU 完全用于计算。
- WDDM 模式下，GPU 同时用于计算和图形。
- 仅在 GPU 加速型云主机安装了 GRID 驱动时才可以切换至 WDDM 模式。

方法二

1. 登录 GPU 加速型云主机
2. 下载 GPU-Z 工具并安装
3. 打开 GPU-Z，选择“Sensors”即可查看 GPU 使用情况



8.8.2.4 Linux 弹性云主机如何升级内核？

备工作

由于这个操作可能会影响系统稳定性，强烈建议在升级内核之前确保您有有效的备份。

操作步骤

本文操作以 CentOS 7.0 操作系统弹性云主机为例。

1. 登录弹性云主机。
2. 安装 epel-release 软件包，执行以下命令：

```
sudo yum install epel-release -y
```

3. 升级内核。

- 升级内核版本为 yum 仓库最新版本，执行以下命令：

```
yum install kernel
```

- 升级内核到指定版本，执行以下命令：

```
yum install kernel-3.10.0
```

4. 重启弹性云主机后，查看升级后内核版本，执行以下命令：

```
uname -r
```

5. 其他升级内核的具体方式由用户自己决定。

8.8.2.5 用户自持镜像导致的弹性云主机操作系统无法正常启动是什么原因？

故障原因

1. repo 源无法使用；
2. 系统中配置文件锁死（如 chattr 修改的问题）；

3.OS 磁盘分区问题，OS 中分区大于购买云主机上的磁盘；

4.自启动服务在云环境下不兼容，开机自启动配置文件 /etc/rc.local 等文件中的配置问题；

解决方法

1.repo 源无法使用

建议检测源是否可用，并尝试更换其他 repo 源。

2.系统中配置文件锁死（如因 chattr 修改的问题）

chattr 命令用于更改系统中文件或者目录的属性，来对系统中关键文件进行锁定保护防止误删或者修改。如果对系统中配置文件锁死可能导致操作系统无法正常启动。建议检查系统中相关配置文件是否被 chattr 修改导致锁死。可以通过 lsattr 命令来查看是否锁死。

使用方式示例：



输出结果中包含 i 属性时说明文件被锁死。

使用 chattr 命令来去除属性 i：



再次使用 lsattr 查看属性：

```
[root@ecs-yhr-0830560 home]# lsattr file
-------------e-- file
```

此时文件已经被解锁。

3.OS 磁盘分区问题，OS 中分区大于购买云主机上的磁盘

建议查看 OS 分区大小，排查 OS 分区大小是否大于购买云主机磁盘。可通过 df 命令来查看 OS 分区大小。

使用方式示例：

```
[root@ecs-yhr-0830560 home]# df -h
Filesystem      Size  Used Avail Use% Mounted on
devtmpfs        3.9G     0  3.9G   0% /dev
tmpfs           3.9G     0  3.9G   0% /dev/shm
tmpfs           3.9G   25M  3.8G   1% /run
tmpfs           3.9G     0  3.9G   0% /sys/fs/cgroup
/dev/vda2        36G  2.6G   31G   8% /
tmpfs           783M     0  783M   0% /run/user/0
```

4.自启动服务在云环境下不兼容，开机自启动配置文件 /etc/rc.local 等文件中的配置问题

检查开机自启动配置文件 /etc/rc.local 等文件中的相关配置是否正确设置。可通过 cat 命令来查看相关配置文件内容，如果存在问题可考虑重新进行配置。

8.8.2.6 文件系统损坏，Linux 弹性云主机启动失败

故障描述

Linux 弹性云主机的文件系统损坏，再次启动 Linux 弹性云主机时失败。

故障原因

在 Linux 弹性云主机中，强制关机、强制重启、物理机宕机疏散、设备读写异常等原因都会小概率导致 Linux 弹性云主机文件系统损坏。

解决步骤

通过 Linux 操作系统中的磁盘修复工具 fsck 修复已损坏的文件系统，操作步骤如下：

1.根据启动页面提示，输入 Linux 弹性云主机的 root 账号与密码。

2.以/dev/vdb 为例，执行命令 mount | grep vdb 查看问题磁盘分区是否已经挂载至 Linux 弹性云主机。如果已经挂载，执行第 3 步，反之执行第 4 步。

3.执行 umount /dev/vdb 卸载问题磁盘分区。

4.执行 fsck -y /dev/vdb 修复问题磁盘分区的文件系统。

5.修复完成后，执行 reboot 命令，重启 Linux 弹性云主机。

8.8.2.7 文件系统 checkfs 使用 fsck 工具检查并修复文件系统

本文主要介绍如何使用 fsck 工具检查并修复 Linux 实例的文件系统

Linux 中的 fsck 命令用于检查和修复文件系统中的错误。fsck 是文件系统检查(File System Check)的缩写。它可以在系统启动时自动运行，也可以手动执行

fsck 命令的基本语法如下：

- -a：自动修复文件系统中的错误。
- -f：强制检查即使文件系统看起来干净。
- -n：模拟运行，不实际执行修复操作。
- -r：交互式检查，并询问用户是否修复错误。
- -y：自动回答“是”修复错误的提示。

设备是指需要检查的文件系统所在的设备，如硬盘分区（/dev/sda1）或逻辑卷（/dev/mapper/vg-root）等。前提条件

已针对系统盘或数据盘创建快照，创建快照请参见云硬盘快照（公测）

操作步骤

本文以/dev/vda1 为例，对文件系统的检查和修复操作进行介绍。

1. 远程连接 Linux 实例

数据盘文件系统问题：请参见登录 Linux 弹性云主机

系统盘文件系统问题：您需要将系统盘挂载到其他实例中成为数据盘，然后再登录实例进行以下操作。

2. 执行以下命令，确认文件系统没有挂载到系统上。

```
df -h
```

3. 执行 blkid 命令，查看文件系统类型，从而针对不同的文件系统类型，执行不同的检查和修复命令。

```
对于 EXT3、EXT4 文件系统：fsck -y /dev/vda1
对于 XFS 文件系统：xfs_repair /dev/vda1
```

4. 确认修复成功后，若您是系统盘文件系统问题，需要将磁盘挂载到之前的实例中。若您是数据盘文件系统问题，请忽略此步骤。

5. 执行以下命令，挂载文件系统。

```
mount /dev/vda1 /mnt
```

6. 若您再次执行以下命令，检查已挂载的文件系统，依然会提示文件系统异常，此时您可以忽略相关告警信息,更多信息，请参见 Redhat issue

```
对于 EXT3、EXT4 文件系统：fsck -y /dev/vda1
对于 XFS 文件系统：xfs_repair /dev/vda1
```

8.8.3 重装操作系统

8.8.3.1 重装弹性云主机时是否可以选择其他操作系统

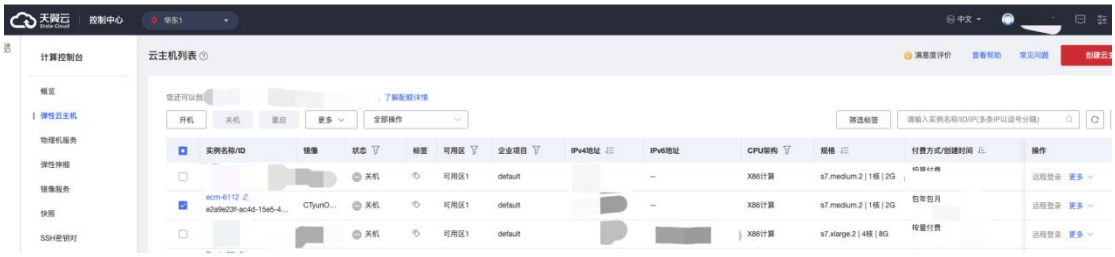
注意事项

1. 重装后 IP 地址、MAC 地址不变，但系统盘会清空，请预先做好云硬盘备份。重装后云主机备份将无法用于恢复数据、云主机快照为错误状态，不支持申请云主机、恢复数据。
2. 重装后系统盘所在的云硬盘备份策略无法继续使用。请在重装前将系统盘从所在备份策略中移出，重装后重新加入；若未提前移出，请在重装后及时新建云盘备份策略。
3. Linux 系统下使用了 LVM 的用户，LVM 如果加入了数据盘，请先从 LVM 卸载掉数据盘，再一键重装，否则会导致主机重装后无法使用。
4. 多次批量一键重装操作建议间隔 120s 以上，否则可能造成重装失败。

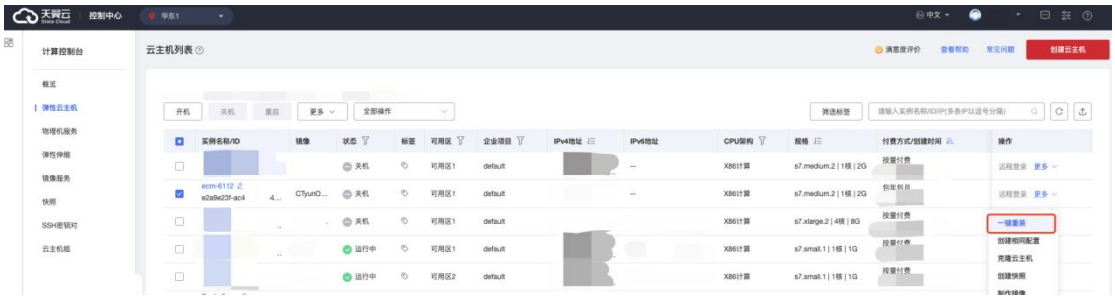
5. 使用加密系统盘重装大约需要 10~25 分钟，容量越大加密耗时越久。

操作步骤

1. 登录云主机控制台，选中需要重装系统的弹性云主机，重装前需要先对云主机进行关机。



2. 点击在云主机右侧操作栏的更多，选择“一键重装”。



3. 在一键重装界面“重装选择”下拉框选择要重装的操作系统，输入“密码”等参数后点击“确认”按钮。



8.8.3.2 重装操作系统需要多长时间？

将云主机关机，单击“更多 >一键重装”，选择需要安装的操作系统，输入密码，点击确定，一般需等待几分钟，即可完成重装操作系统。

重装系统过程中，云主机会显示任务状态为“重建中”。重建完成显示关机，选定云主机，点击开机按钮，即可打开云主机进行使用。

8.8.3.3 用户能否自己安装或者升级操作系统

弹性云主机所使用的是云平台提供的操作系统，支持用户对操作系统进行系统更新或补丁升级。

若用户需要更改操作系统时，首先需要在控制台将弹性云主机关机，然后再进行一键重装更改操作系统。例如升级操作系统，从 CentOS 7.6 升级到 Cent OS 7.8 就可以通过切换操作系统的功能来实现。

8.8.3.4 重装/切换操作系统/变更规格对磁盘数据有影响吗

对系统盘数据的影响

重装系统（linux）会清除系统盘数据，包括系统盘的系统分区和所有其他分区挂载点配置，请做好数据备份；

重装系统（windows）会清除系统盘数据，包括系统盘的系统分区和所有其他分区挂载点配置，软件安装后的注册表信息，请做好数据备份。

变更规格对系统盘没有影响。

对数据盘的影响

重装系统（windows），相关软件需要试用可能造成程序不可用，需要重新安装，但是之前程序在数据盘的数据会保留。

变更规格对数据盘没有影响。

8.9 文件上传/数据传输

8.9.1 使用 FTP 上传文件时写入失败、文件传输失败

故障描述

使用 FTP 上传文件时，写入文件或者传输文件失败。

约束与限制

本文档适用于服务端为 Windows 系统上的 FTP 服务。

故障原因

弹性公网 IP 与内网 IP 是通过 NAT 方式绑定的，因此本地访问弹性云主机时，客户端是通过被动模式连接服务端的，在这种情况下，服务端的 IP 地址无法从路由器外部访问，所以需要在服务端的对外 IP 中填写此弹性云主机分配的公网 IP，同时设置数据传输端口范围来限制需要通过路由器转发的端口数量。

解决步骤

1.检查 FTP 服务的地址是否填写正确

点击“ 服务器管理器 -> 工具 -> Internet Information Services(IIS)管理器” ，然后选择左边栏点击“ 网站” 。首先查看建立的 FTP 绑定的 IP 是否填写的是内网 IP，类型为 FTP。

图 1 查看 FTP 所绑定的 IP



2.检查服务端的 FTP 防火墙支持配置

展开“ 网站” ，会显示建立的 FTP 服务，双击右侧的 FTP 防火墙支持。

图 2 FTP 防火墙支持



数据通道端口范围需要填写 1024-65535 范围内的，比如 20000-20045。

防火墙的外部 IP 地址这一栏需要填写分配到的 公网 IP 。

3.检查安全组设置

点击“ 控制中心 -> 服务列表 -> 弹性云主机” ，选择对应的可用区，然后再展示出来的弹性云主机列表里选择需要的弹性云主机，然后双击这个弹性云主机，即可打开详情页面。在下方选择“ 安全组” ，然后查看详细的安全

组配置，是否在入方向里已经正确配置了端口 21 和上面配置的数据通道端口这两条。

图 3 安全组配置



方向	类型	优先级	授权策略	协议	端口范围/ICMP类型	远端	描述	操作
入方向	IPv4	1	允许	TCP	20000-20045	0.0.0.0/0		删除 修改
入方向	IPv4	1	允许	TCP	21	0.0.0.0/0	ftp端口	删除 修改

4.在客户端测试

一种方法是使用 FileZilla 填写弹性云主机的公网 ip,FTP 的用户名和密码，端口填写 21。

另一种方法是访问 ftp://弹性云主机公网 ip:服务端的端口（如果不填为 21），然后在弹出的对话框里输入 FTP 的用户名和密码。

8.9.2 WinSCP 无法连接到 Linux 云主机

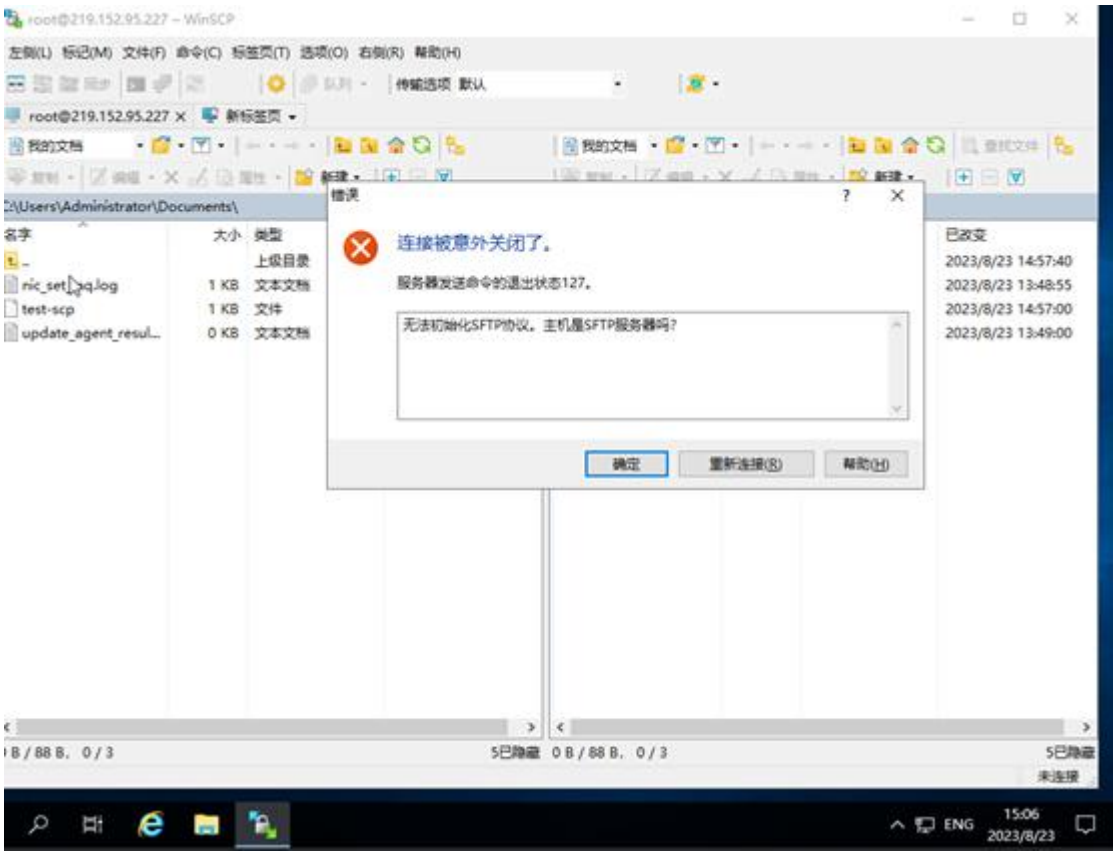
前提条件

Windows 云主机内已下载并安装 WinSCP 客户端（建议从官方网站获取最新版本），默认使用的端口号是 22，请确保双方云主机的 22 端口开放，连接 Linux 云主机的文件协议选 SFTP 或 SCP。

故障描述

通过 Windows 云主机内的 WinSCP 连接 Linux 云主机失败，报错如图 1。

图 1 通过 winscp 连接 linux 云主机报异常



故障排查

WinSCP 是通过 SFTP 协议进行的主机连接，所以需要判断 Linux 云主机内的 SSH 配置文件中是否有 sftp 相关的配置，查看 /etc/ssh/sshd_config 中关于 sftp 的配置描述。

图 2 检查 linux 云主机 sshd 是否开启 sftp 相关配置



上图 2 说明 SFTP 关联的配置文件是/usr/libexec/openssh/sftp-server，如果此文件缺失或者权限不对将会导致 WinSCP 连接异常，此文件默认权限是 755。

解决步骤

如果是 sftp-server 文件缺失，可以从其他云主机拷贝一份到故障云主机的对应目录下。

如果是权限问题，通过 chmod 命令修改文件权限。命令如下。

```
chmod 755 -R /usr/libexec/openssh/sftp-server
```

8.9.3 Windows 云主机搭建了 FTP，通过公网无法访问

问题描述

本文档解决的是在天翼云创建的 windows 操作系统的云主机，其上通过 IIS(信息服务管理器)搭建 FTP 站点后，通过公网无法访问的问题。

造成问题的原因

1、windows 云主机内防火墙拦截了 FTP 进程。

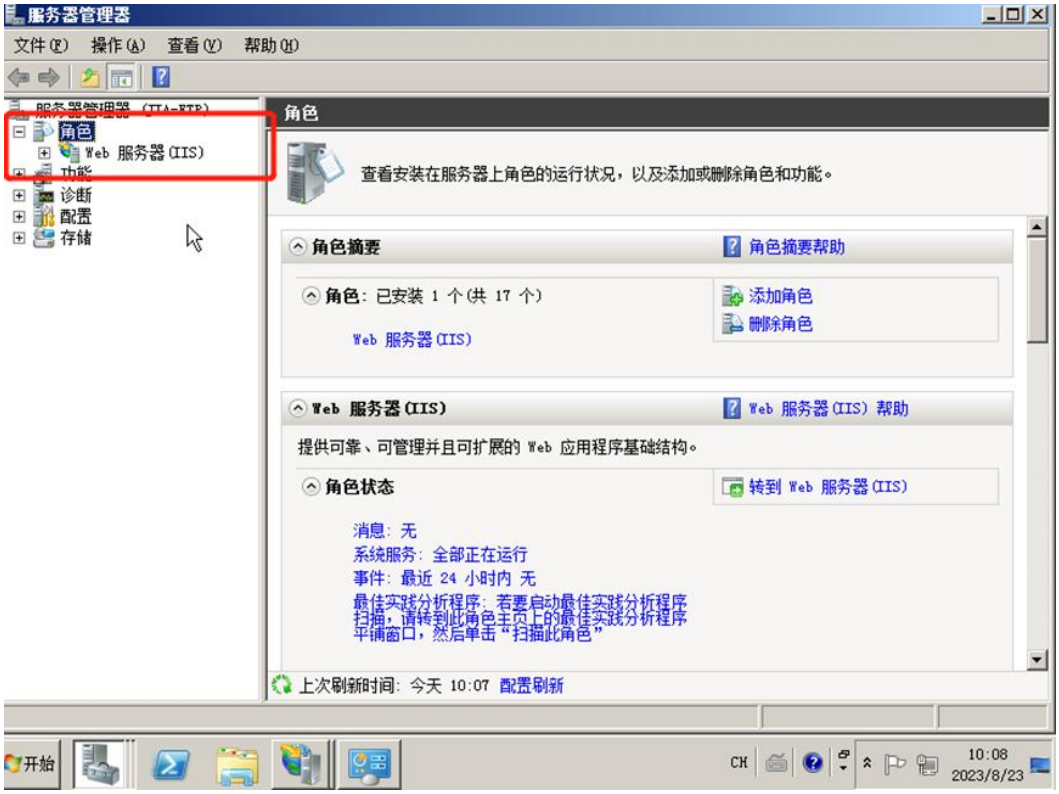
2、云主机相关的安全组拦截了外网的连接和访问。

windows 云主机配置 FTP 防火墙

这里演示的是 windows2008 的操作系统，其他系统的可自行搜索开启方式。具体操作步骤如下。

1. 判断云主机是否开启 IIS。在云主机内点击“开始->控制面板->打开或关闭 windows 功能进入到服务器管理器”，查看角色下是否有 IIS，如无执行步骤 2 进行开启，若已经开启请执行步骤 3。

2. 图 1 检查云主机是否开启 IIS 信息服务管理器



2.在服务器管理器界面，单击角色，并添加角色，按添加角色的向导，在服务器角色中 **选择 web 服务器 IIS** ，点击下一步进行角色服务设置，在 web 服务器的角色服务中**选择 FTP** 后点击安装。

图 2 开启 IIS 信息管理服务器搭建 FTP 站点步骤 1

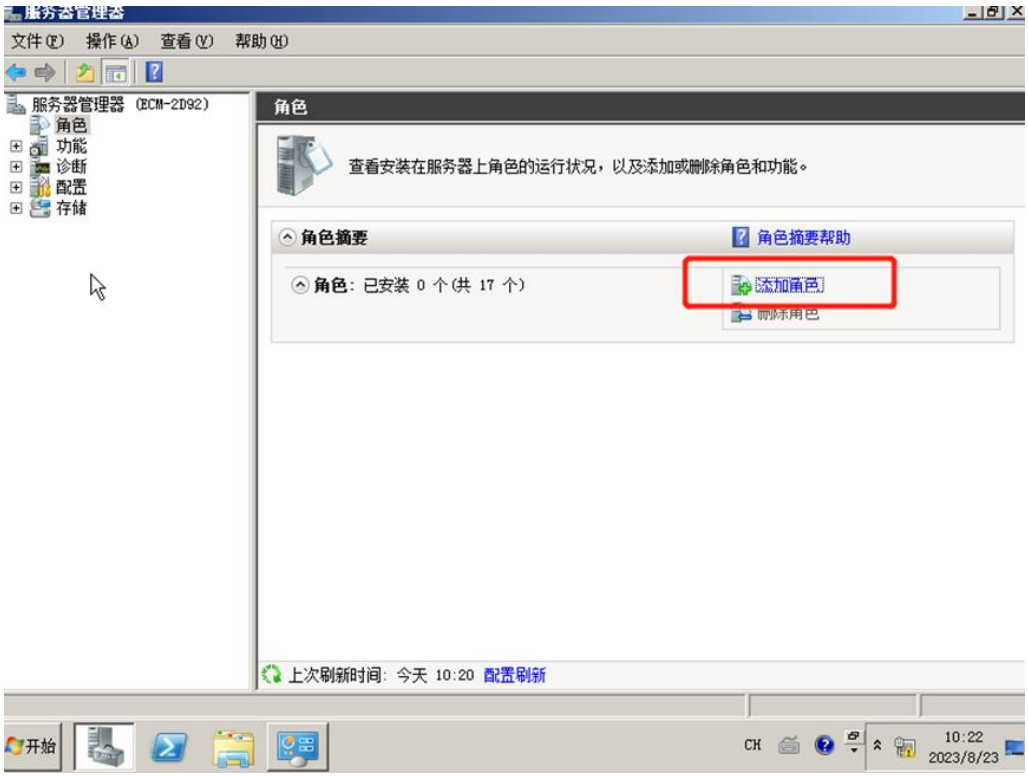


图 3 开启 IIS 信息管理服务器搭建 FTP 站点步骤 2

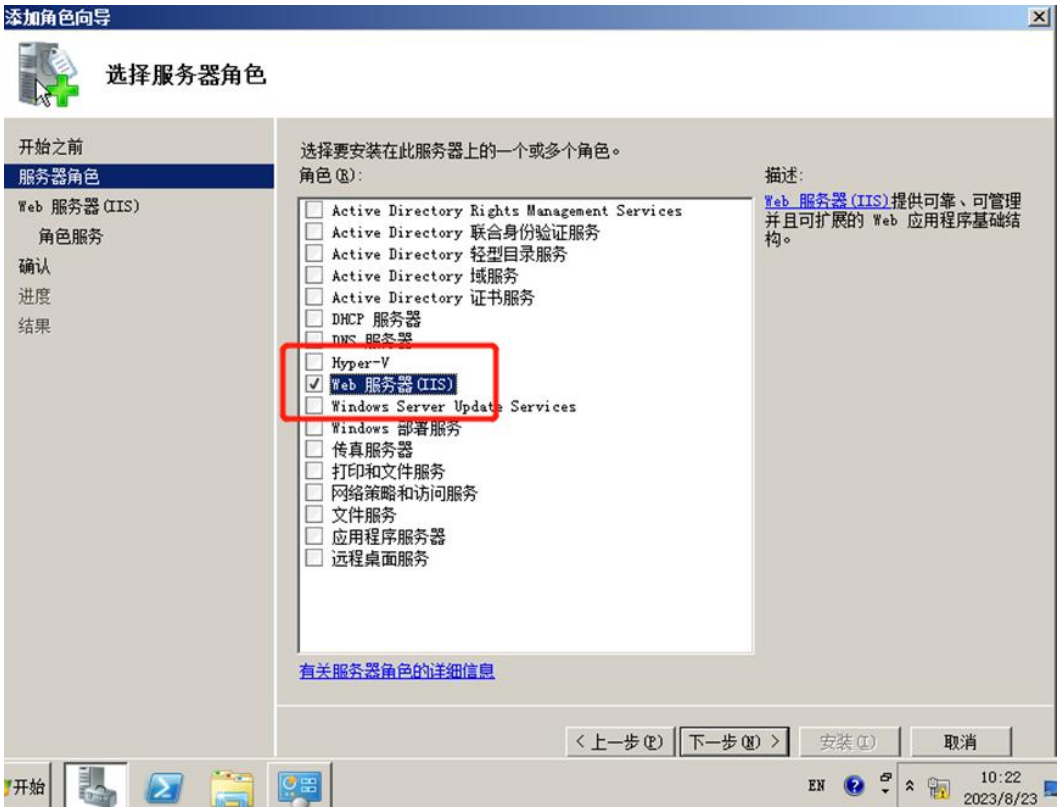
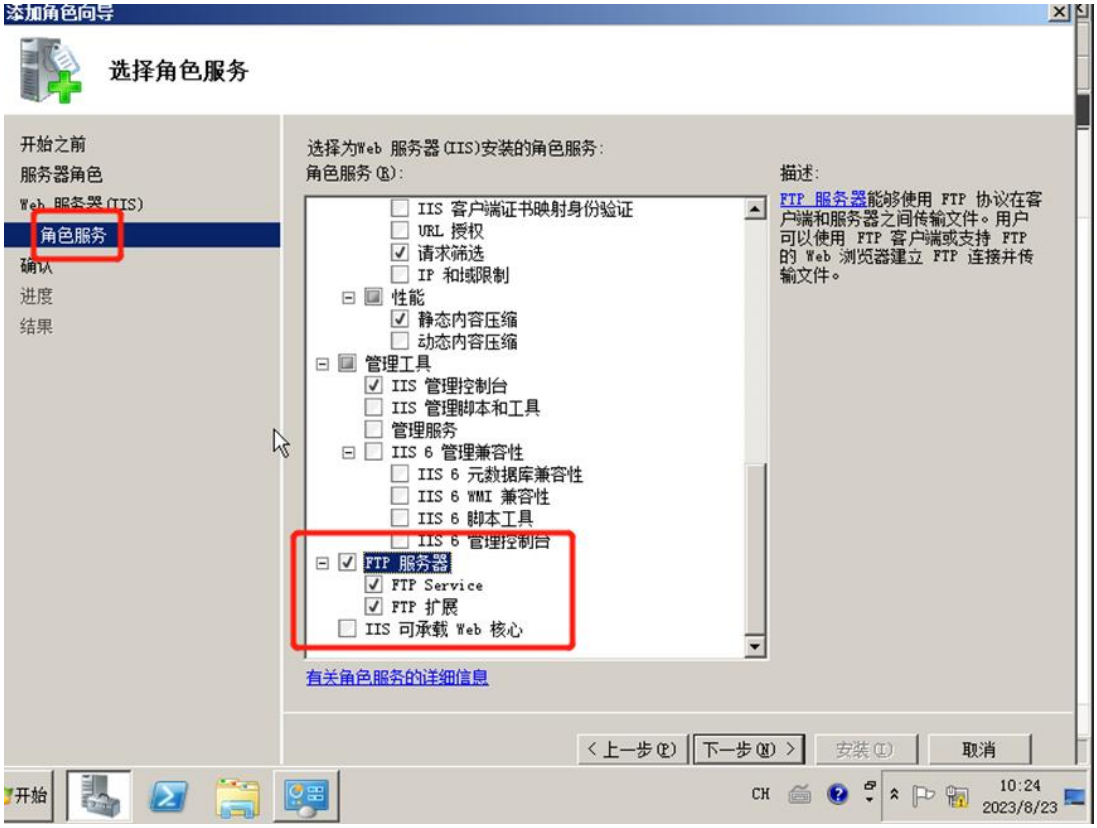


图 4 开启 IIS 信息管理服务器搭建 FTP 站点步骤 3



FTP 防火墙故障的解决步骤

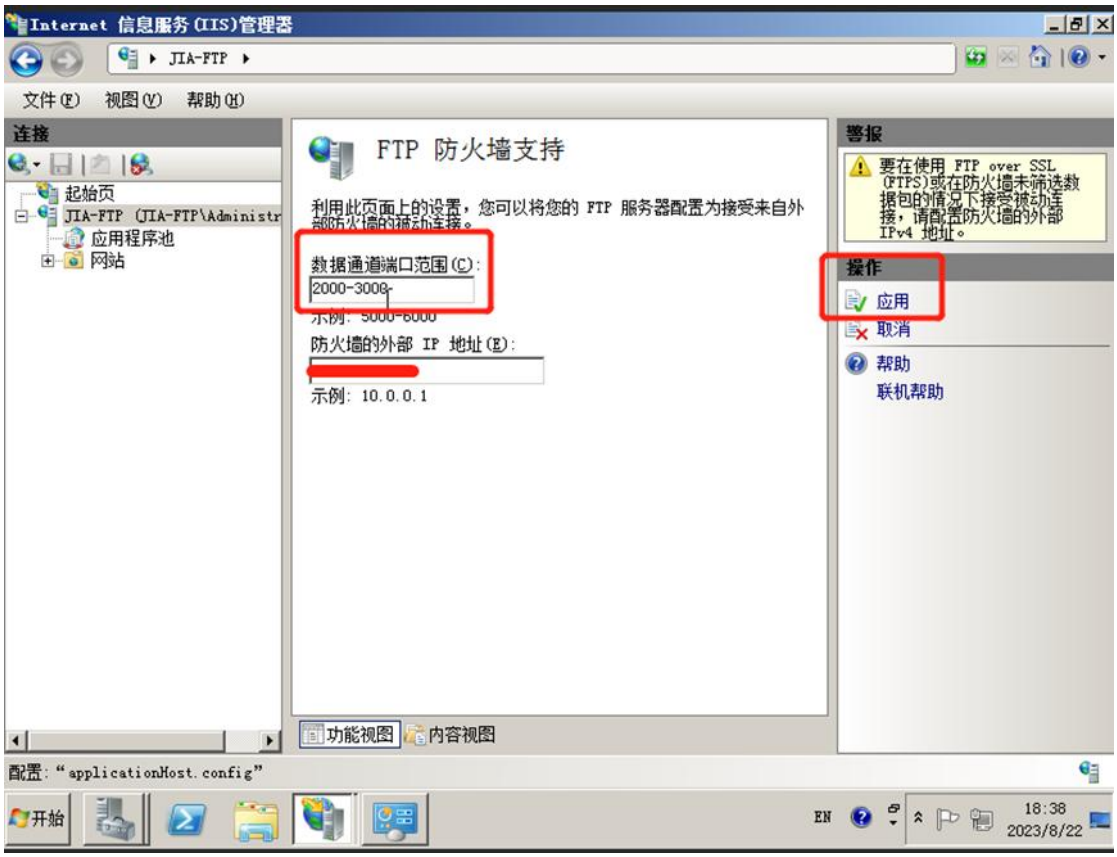
a. 在开始中搜索 IIS，在 Internet 信息服务（IIS）管理器界面中找到 FTP 防火墙设置。

图 5 打开 FTP 站点的防火墙



b. 双击 FTP 防火墙支持，设置数据通道端口范围（可指定的有效端口范围是 1025-65535），防火墙的外部 IP 地址中输入该云主机的公网 IP 地址，设置完成后点击右侧的应用。

图 6 调整 FTP 防火墙中的数据通道端口范围



c. 重启云主机使 FTP 防火墙配置生效。

云主机设置安全组及防火墙

在云主机内搭建好 FTP 站点后，需要在云主机安全组的入方向添加一条放行 FTP 端口的规则，请按以下步骤进行操作。

1. 进入弹性云主机的管理控制台页面，进入云主机控制台，在顶端菜单栏 处选择云主机所在的地理区域，在云主机列表中找到内部搭建了 FTP 的云主机，点击云主机的名称进入详情页面。

图 7 打开云主机的安全组

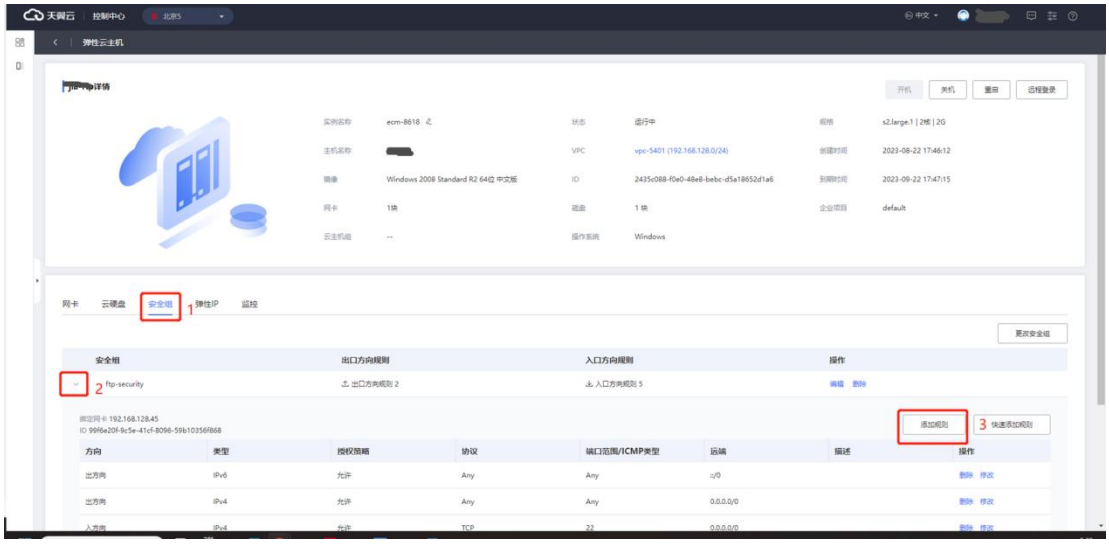


3. 在云主机详情页面，选择“安全组”，点击云主机当前的安全组，确定安全组入口规则中有 TCP 协议的开放端口中涵盖了本文中“FTP 防火墙故障的解决步骤”中步骤 2 指定的端口范围。如无请按本文中“云主机安全组规则添加的解决步骤”进行处理。

云主机安全组规则添加的解决步骤

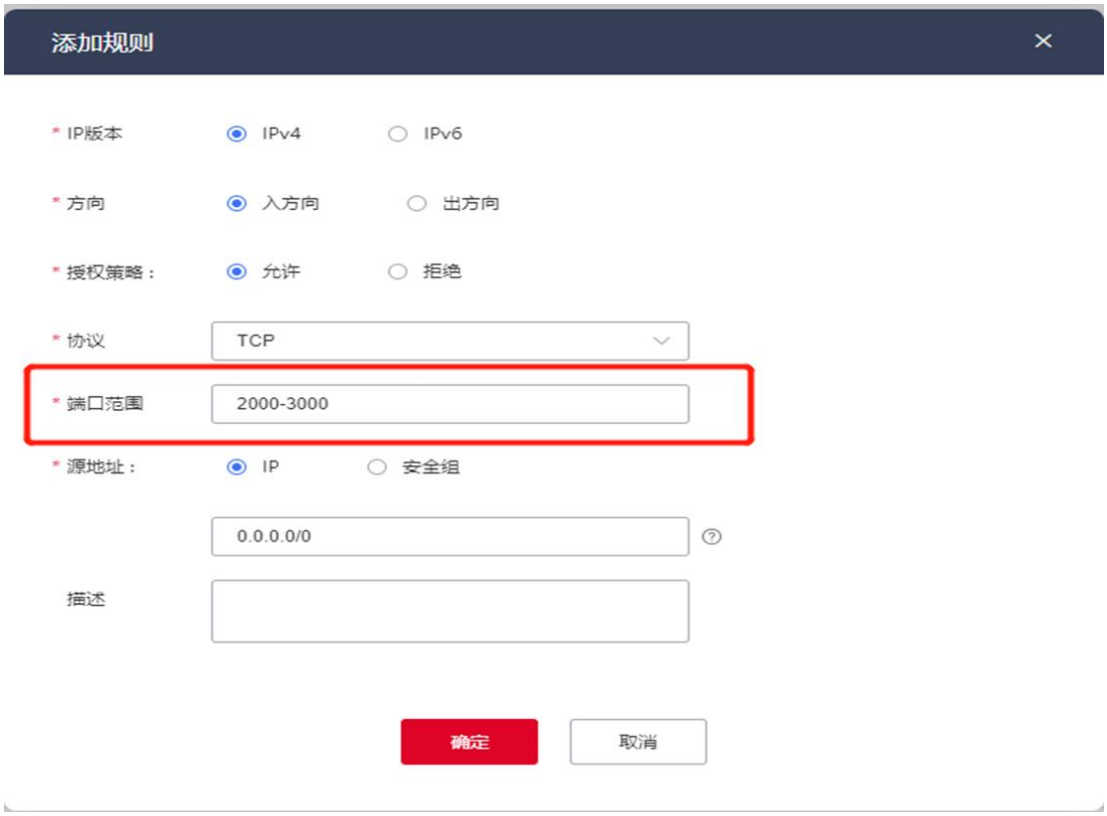
a. 对故障云主机的安全组添加端口开放规则。

图 8 云主机安全组中添加端口开放规则



b. 添加规则时选择“入方向”，端口范围需涵盖“FTP 防火墙故障的解决步骤”中步骤 2 指定的端口范围，例如此处配置的数据端口范围为 2000-3000。

图 9 云主机安全组中添加端口开放规则



8.9.4 使用 FTP 上传文件时客户端连接服务端超时

故障描述

使用客户端连接服务端时，提示连接服务端地址失败或者连接超时。

约束与限制

本文档适用于本地主机为 windows 系统上的 FTP 服务。

故障原因

1. 服务端 FTP 防火墙或者安全组配置错误导致请求被拦截。
2. FTP 身份验证选择的类型与实际访问方式不匹配。

解决步骤

服务端为 windows 系统（以 windows Server 2012 版本为例）

1. 检查服务端的 FTP 防火墙支持配置

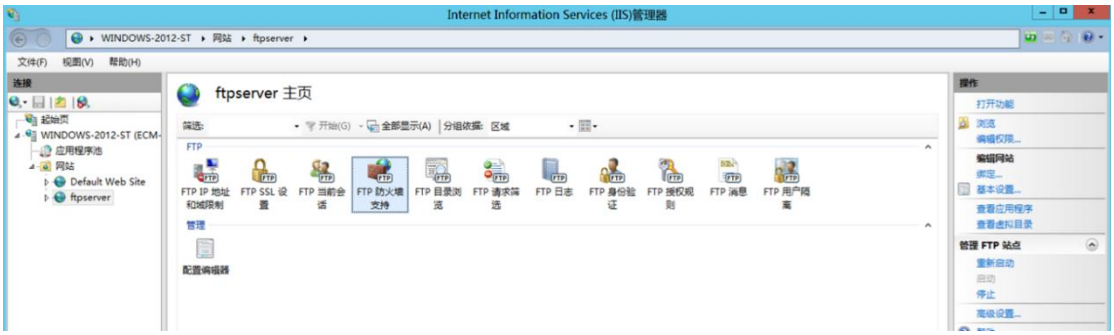
点击" 服务器管理器 -> 工具 -> Internet Information Services(IIS)管理器"，然后选择左边栏点击 网站 。首先查看建立的 FTP 绑定的 ip 是否填写的是内网 ip，类型为 FTP。

图 1 查看 FTP 绑定 IP



展开 网站 ，会显示建立的 ftp 服务，双击右侧的 FTP 防火墙支持。

图 2 FTP 防火墙支持



数据通道端口范围需要填写 1024-65535 范围内的，比如 20000-20045。

防火墙的外部 IP 地址这一栏需要填写给弹性云主机分配到的 公网 IP 。

1. 检查 FTP 身份验证类型

双击 FTP 身份验证 ，这里需要注意，如果基本身份验证开启，那么是需要填写用户名和密码的。如果开启的是 匿名身份验证 ，那么连接时不能填写用户名和密码，否则会报连接失败。

2. 检查安全组设置

点击" 控制中心->服务列表->弹性云主机" ，选择对应的可用区，然后再展示出来的弹性云主机列表里选择需要的弹性云主机，然后双击这个弹性云主机，

即可打开详情页面。在下方选择" 安全组" ，然后查看详细的安全组配置，是否在入方向里已经正确配置了端口 21 和上面配置的数据通道端口这两条。

图 3 安全组配置



服务端为 Linux 系统(以 Centos7.2 64 位操作系统为例)

1.检查配置文件

执行以下命令，检查配置文件。

```
cat /etc/vsftpd/vsftpd.conf
```

确保下面的内容已经正确填入。

```
listen=YES  listen_ipv6=NO  pasv_address=xx.xx.xx.xx          #弹性云主机的公网 IP 地址 pasv_min_port=20000          #被动模式下的最小端口 pasv_max_port=20045          #被动模式下的最大端口
```

2.检查安全组配置

点击 控制中心 -> 服务列表 -> 弹性云主机 ，选择对应的可用区，然后再展示出来的弹性云主机列表里选择需要的弹性云主机，然后双击这个弹性云主机，即可打开详情页面。在下方选择 安全组 ，然后查看详细的安全组配置，是否在入方向里已经正确配置了端口 21 和配置文件“/etc/vsftpd/vsftpd.conf”中参数“pasv_min_port”和“pasv_max_port”之间的所有端口。

3.重新进行客户端测试

一种方法是使用 FileZilla 填写弹性云主机的公网 ip，ftp 的用户名和密码，端口填写 21。

另一种方法是在浏览器或者文件夹访问栏输入 ftp://弹性云主机公网 ip:服务端的端口（如果不填为 21），然后在弹出的对话框里输入 ftp 的用户名和密码。

8.9.5 本地 Windows 主机通过 MSTSC 上传文件到 Windows 云主机

前提条件

- Windows 云主机已绑定弹性 IP。
- Windows 云主机已经放通 MSTSC 安全组接口（默认 3389）。

操作步骤

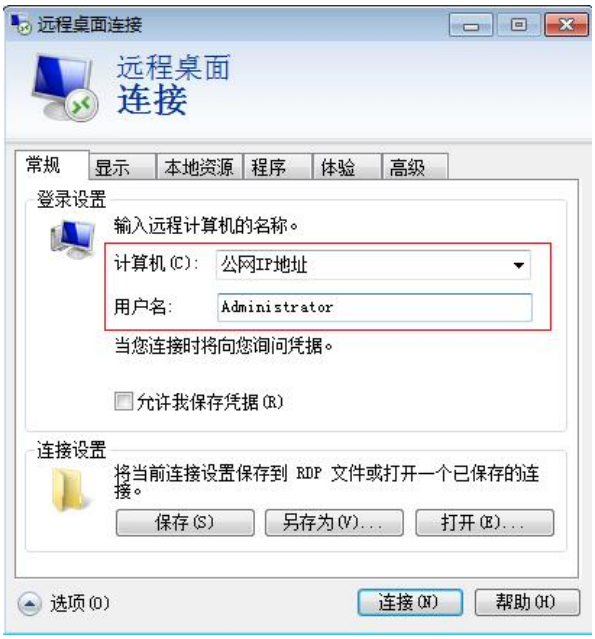
1. 在本地 Windows 计算机上，单击“开始”，在出现的“搜索程序和文件”输入框中输入 mstsc。

弹出远程桌面连接对话框。

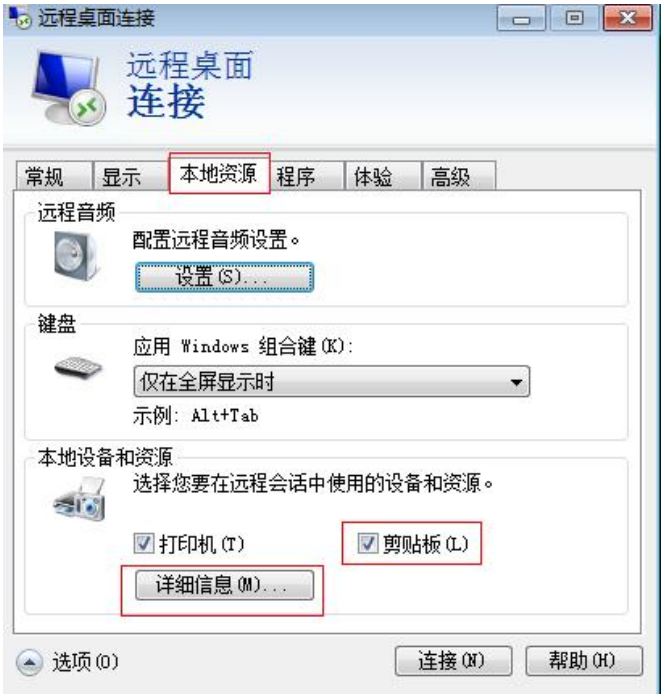
2. 单击“选项”。



3. 在“常规”页签中，输入云主机的公网 IP 地址和用户名“Administrator”。

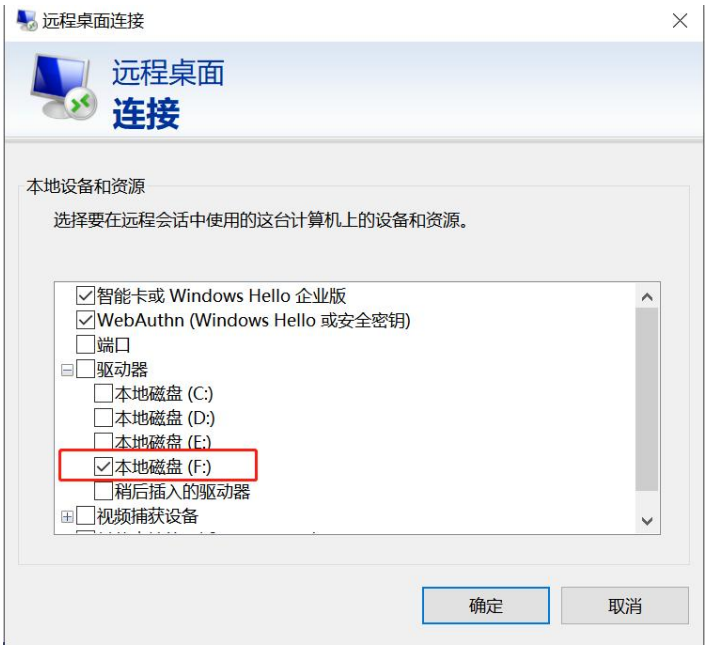


4. 选择“本地资源”页签，确认“本地设备和资源”栏的“剪切板”处于勾选状态。



5. 单击“详细信息”。

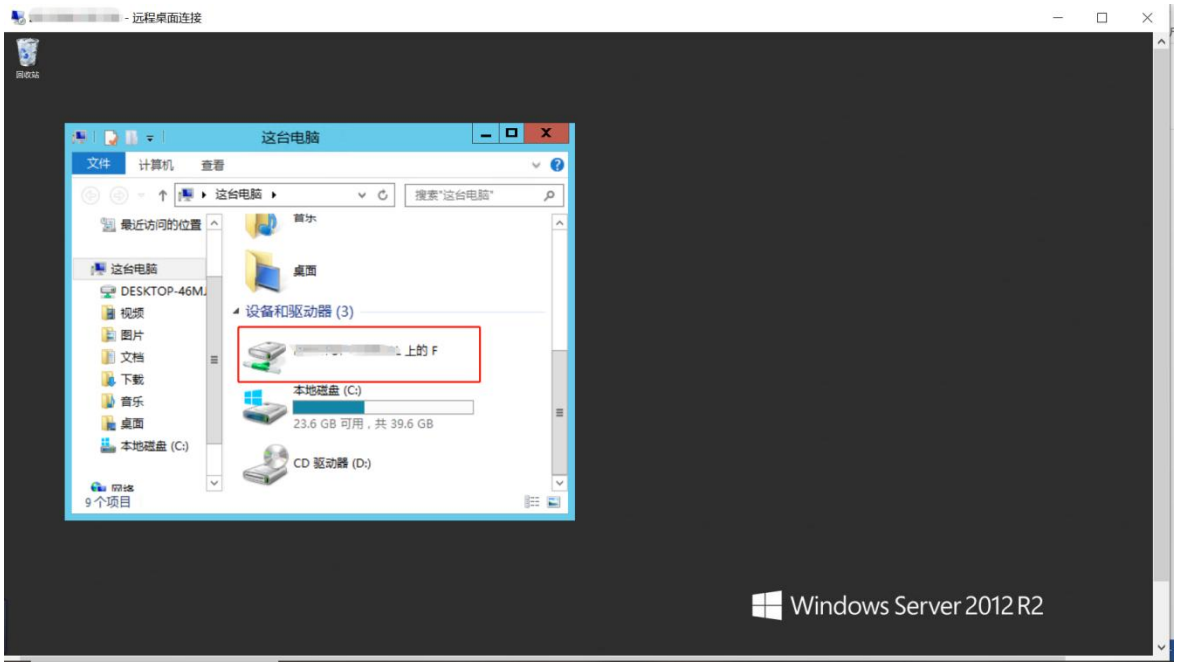
6. 在“驱动器”多选框列表，勾选要上传到 Windows 云主机上的文件所在的本地磁盘。



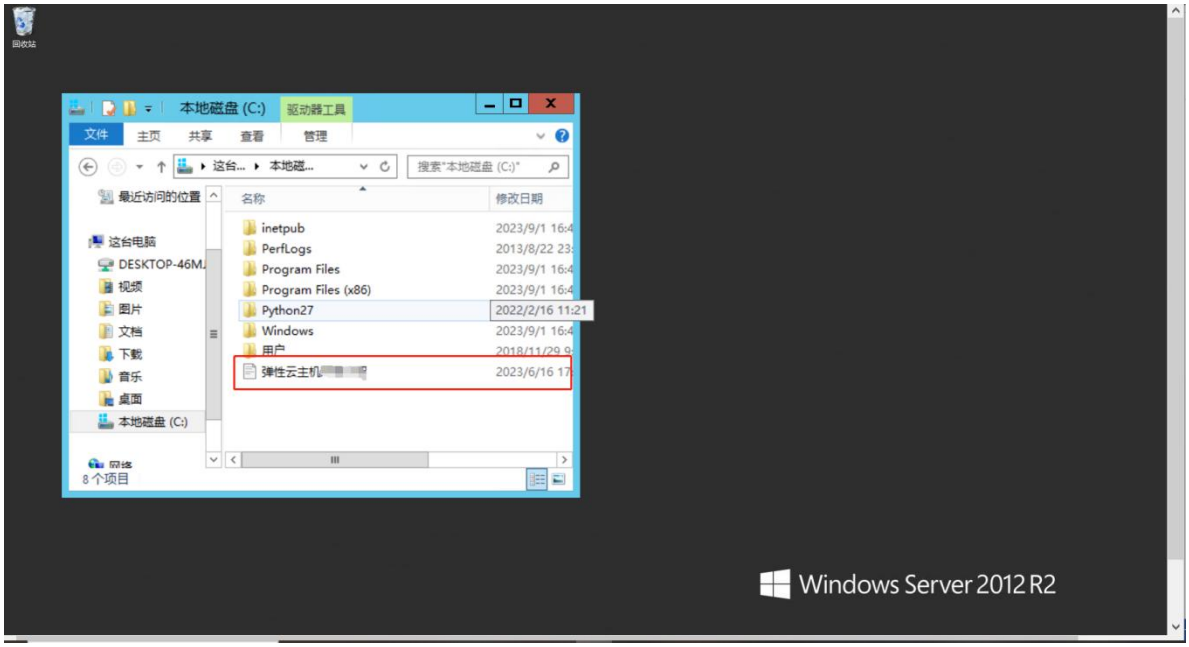
7. 打开“确定”，登录 Window 云主机

8. 单击“开始 > 这台电脑”。

在出现的 Windows 云主机上可看到本地盘的信息。



9. 在云主机中，双击进入本地磁盘，将需要上传的文件复制到 Windows 云主机。



8.9.6 本地 Windows 主机使用对象存储上传文件到 Windows 云主机

云主机支持通过内网访问对象存储，对象存储可供用户存储任意类型的数据。将图片、视频等数据存储至对象存储后，在弹性云主机上可以访问对象存储，下载桶中的图片或视频等数据。

- 步骤 1：创建桶

在对象存储（原生版）I型页面，单击“创建 Bucket”，填写桶信息，完成桶的创建。



步骤二：上传文件

点击创建完成的桶名称，进入桶详情页面，单击“文件管理”页签，进入文件管理页签，单击“上传文件”上传所需的文件至对象存储桶中。



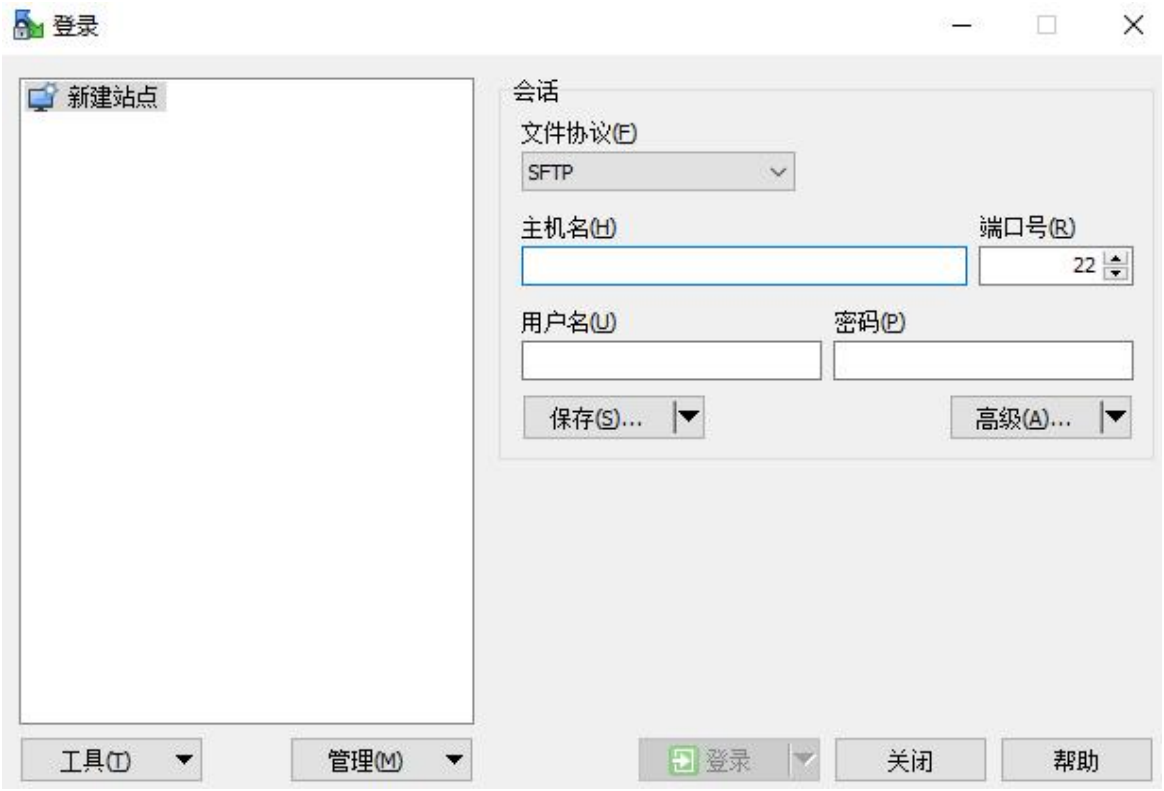
- 步骤三：在 Windows 云主机上通过内网访问对象存储，详细内容参见对象存储（原生版）I型-云主机通过内网访问对象存储。

8.9.7 本地 Windows 主机使用 WinSCP 上传文件到 Linux 云主机

WinSCP 工具可以实现在本地与远程计算机之间安全地复制文件。

1. 下载 WinSCP 客户端并安装。下载 WinSCP。

2. 启动 WinSCP，启动后界面如下：

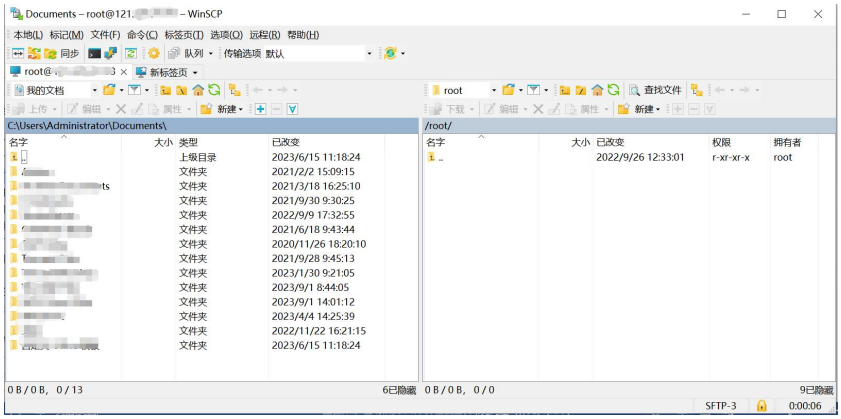


填写说明：

- 协议：选填 SFTP 或者 SCP 均可。
- 主机名：云主机的公网 IP。登录管理控制台即可查看对应云主机的公网 IP。
- 端口：默认 22。
- 用户名：root。
- 密码：购买云主机设置的密码。

3. 单击“登录”，进入 “WinSCP” 文件传输界面。

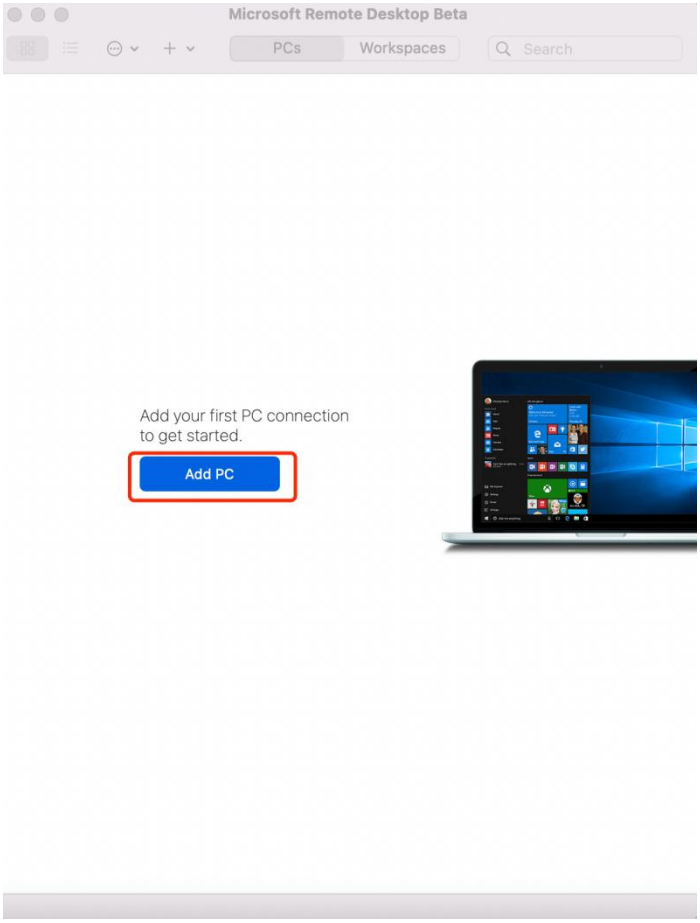
4. 登录成功之后，您可以选择左侧本地计算机的文件，拖拽到右侧的远程云主机，完成文件上传到云主机。



8.9.8 本地 MacOS 系统主机上传文件到 Windows 云主机

本次我们采用 Microsoft Remote Desktop for Mac 作为远程连接工具。下载 Microsoft Remote Desktop for Mac。

1. 启动 Microsoft Remote Desktop。单击“Add Desktop”。



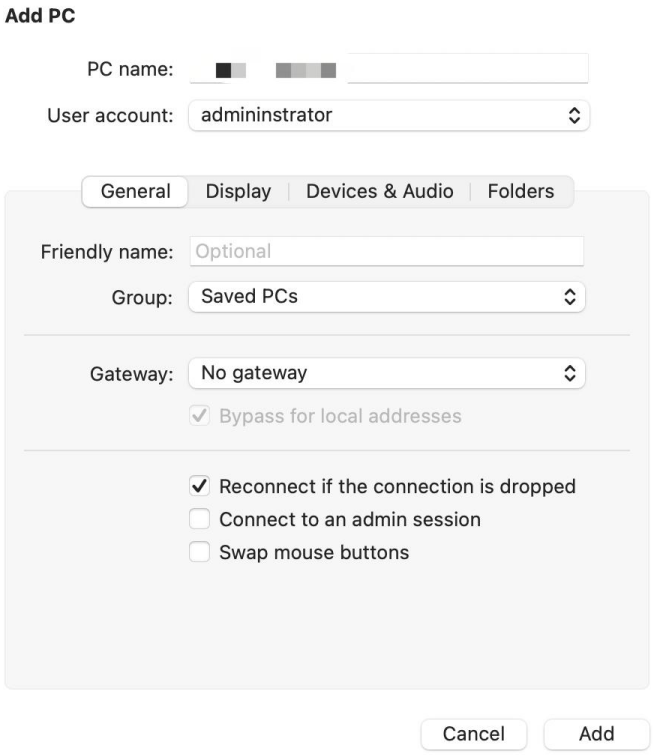
2. 设置登录信息。

PC name：输入需要登录的 Windows 实例的弹性公网 IP 地址。

User account：在下拉列表中选择“Add user account”。

弹出“Add user account”对话框。

输入 Windows 实例帐号“administrator”，并输入实例的登录密码，单击“Add”。

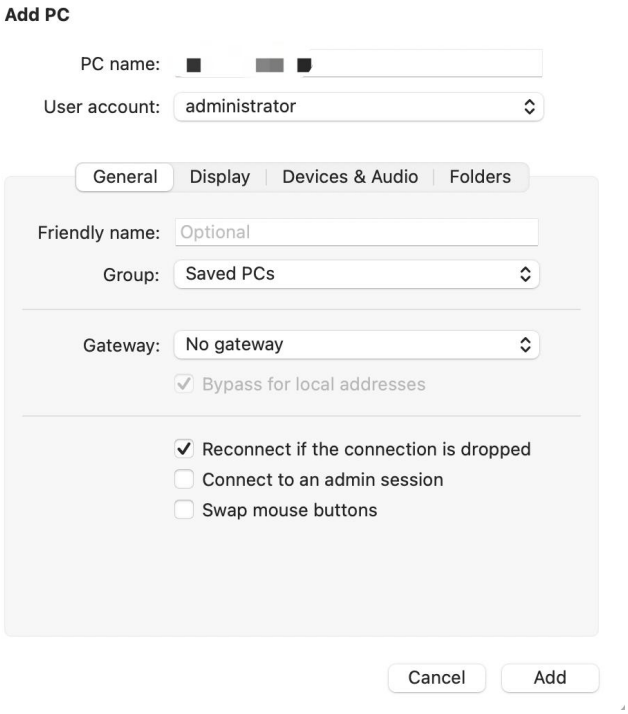


3 选择待上传的文件夹。

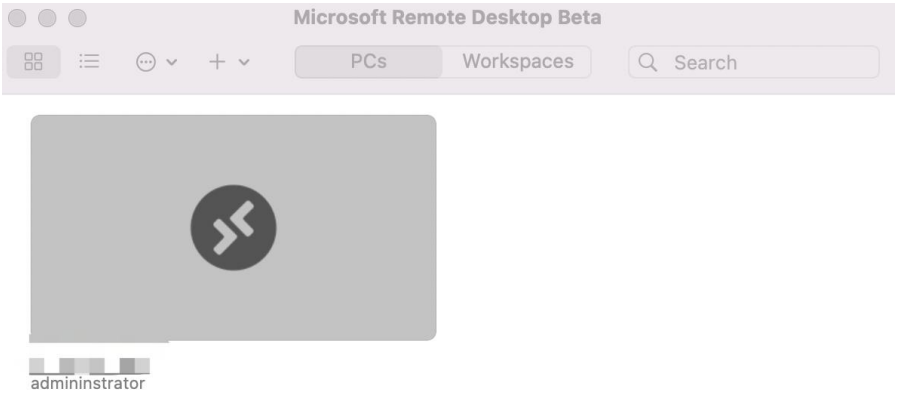
单击“Folders”切换至选择文件夹列表。

勾选“Redirect folders”。

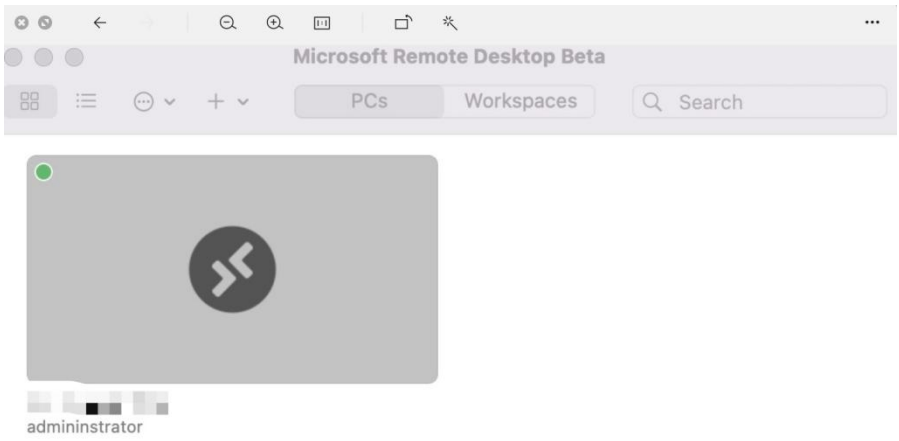
单击左下角　　，选择需上传的文件夹，并单击“Add”。

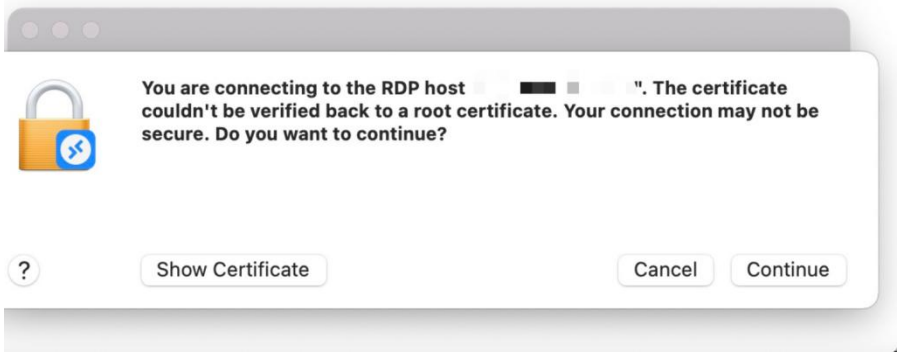


在“Remote Desktop”页面，双击需要登录的 Windows 实例图标。



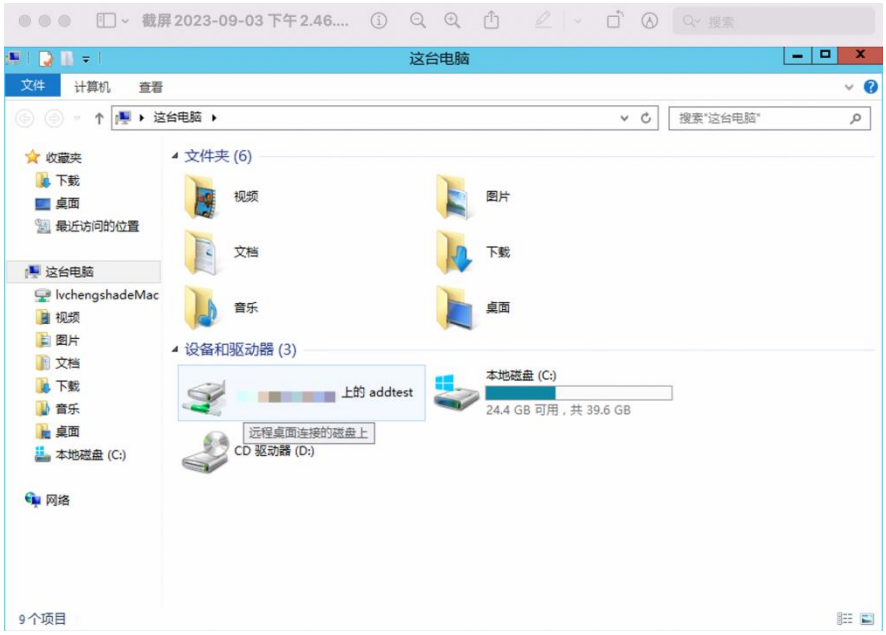
确认登录信息后，选择“Continue”。





至此，您已经链接 Windows 实例。

4.打开 Windows 云主机，查看共享的文件夹。



将需要上传的文件复制到 Windows 云主机。或将 Windows 云主机的文件下载到本地主机。

8.9.9 本地 Linux 主机使用 SCP 上传文件到 Linux 云主机

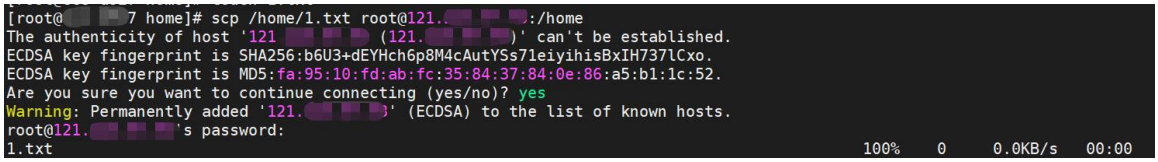
在本地 Linux 操作系统主机上执行以下命令，传输文件到 Linux 操作系统云主机。

```
scp 本地主机文件地址 用户名@弹性公网 IP:云主机文件地址
```

例如：将本地文件 /home/1.txt 上传至弹性公网 IP 地址为 121.x.x.x 的云主机对应目录下，命令如下：

```
scp /home/1t.txt root@121.x.x.x:/home
```

根据提示输入登录密码，即可完成上传。



8.9.10 本地 Linux 主机使用 SFTP 上传文件到 Linux 云主机

1.登录弹性云主机，创建用户和组，以 test 用户为例。

```
groupadd sftp

useradd -g sftp -s /sbin/nologin test
```

2.设置用户密码。

```
passwd test
```

3.设置目录权限。

```
chown  root:sftp /home/testchmod 755 -R /home/testmkdir
/home/test/uploadchown -R test:sftp /home/test/uploadchmod -R 755
/home/test/upload
```

4.执行以下命令，编辑 sshd_config 文件。

```
vim /etc/ssh/sshd_config
```

注释掉如下信息

```
#Subsystem sftp /usr/libexec/openssh/sftp-server
```

补充如下内容：

```
Subsystem sftp internal-sftp

Match Group sftp

ChrootDirectory /home/%u

ForceCommand internal-sftp

AllowTcpForwarding no

X11Forwarding no
```

5.重启云主机，或执行以下命令重启 sshd 服务。

```
systemctl restart sshd
```

6.在本地主机执行以下命令，远程连接到服务器。

```
sftp root@ IP 地址
```

连接成功后，您可以使用交互式的 sftp 命令。



7.执行以下命令，上传或下载文件、文件夹。

上传文件：put -r

```
sftp> put -r /home/test/upload
Uploading /home/test/upload/ to /root/upload
Entering /home/test/upload/
/home/test/upload/1.txt
/home/test/upload/2.txt
sftp> _
```

8.9.11 本地 Windows 主机使用 FTP 上传文件到 Windows/Linux 云主机

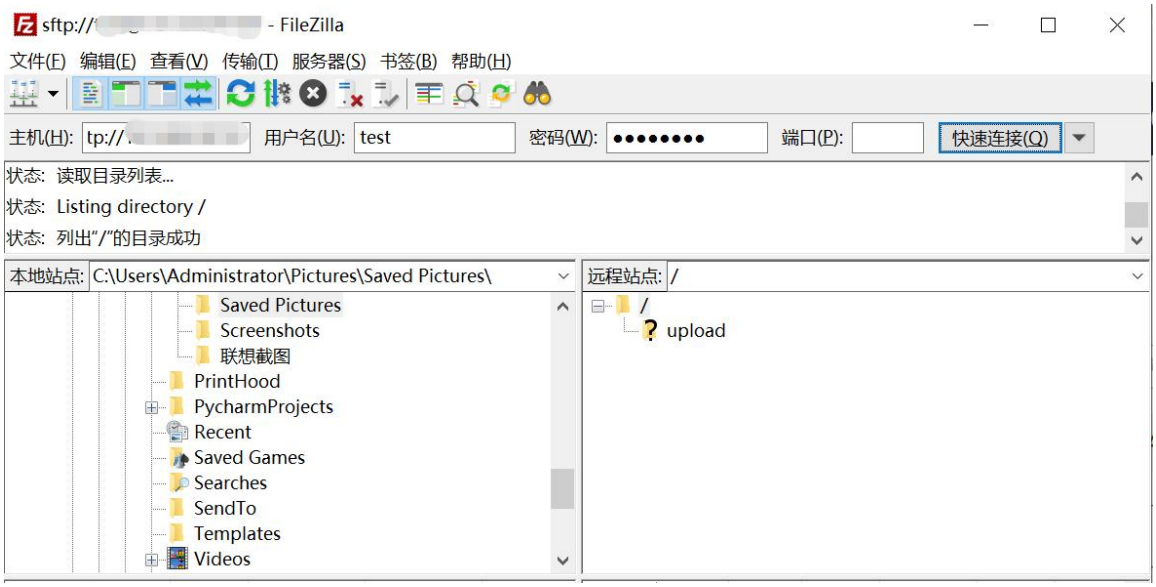
本次我们以 FileZilla 为例演示上传文件。下载 FileZilla。

1.首先需要在弹性云主机上搭建 FTP 站点。具体操作参见弹性云主机－搭建 FTP。

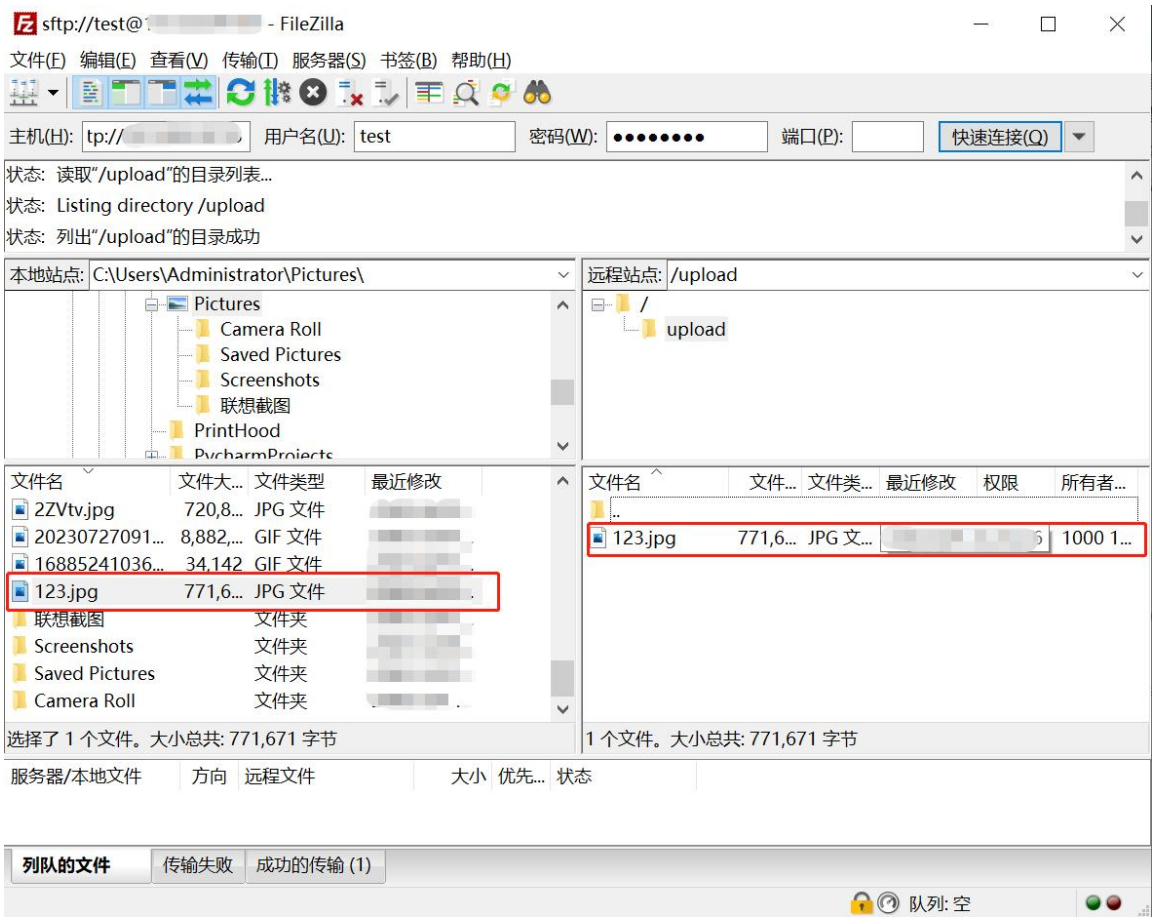
2.在 Windows 本地主机打开 FileZilla，填写待连接的云主机信息，单击“快速连接”。



- 主机：云主机的弹性公网 IP。
- 用户名：搭建 FTP 时设置的用户名。
- 密码：搭建 FTP 时设置的用户对应的密码。
- 端口：FTP 链接端口，默认使用 21 端口。



3. 您可以选择左侧本地计算机的文件，拖拽到右侧的远程云主机，完成文件上传到云主机。



8.9.12 本地 Linux 主机使用 FTP 上传文件到 Linux 云主机、

1. 首先需要在弹性云主机上搭建 FTP 站点。具体操作参见弹性云主机－搭建 FTP。

2. 以 CentOS 7.6 操作系统为例，执行以下命令安装 ftp。

```
yum -y install ftp
```

3. 执行以下命令连接云主机。

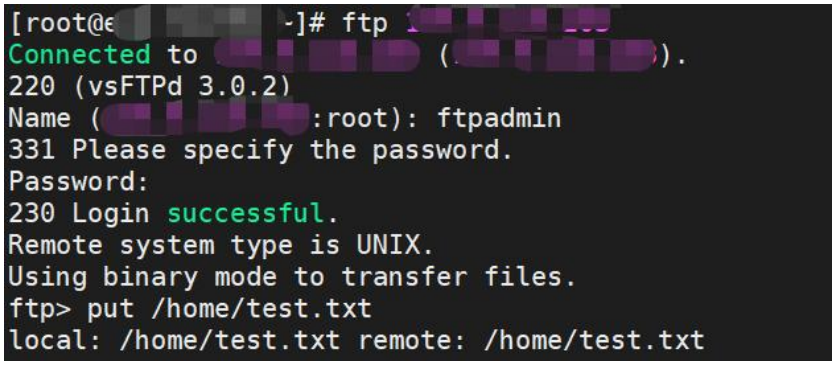
```
ftp 云主机弹性公网 IP
```

并根据提示，输入 FTP 服务的用户名和密码。

4. 上传文件

执行以下命令，将本地文件上传至云主机中。

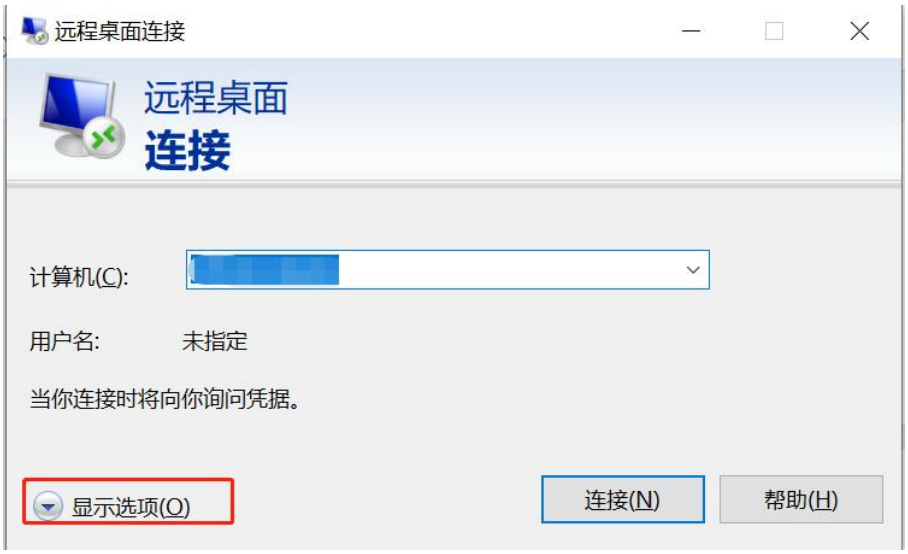
```
put 本地主机文件地址
```



8.9.13 怎样在本地主机和 Windows 云主机之间互传数据？

除了前面问题中使用的 FTP 传输方式，我们本次介绍本地磁盘映射方式实现互传数据。

1. 在本地主机上，在“打开”栏，输入“mstsc”，打开远程桌面连接。

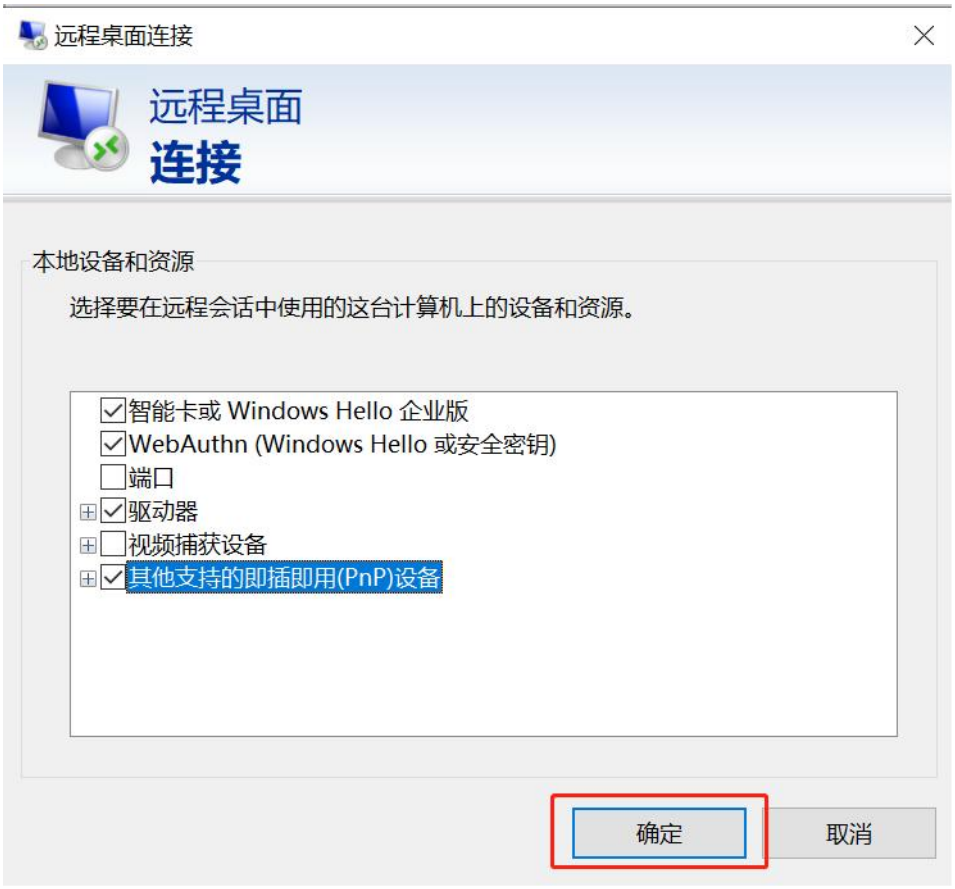


2. 单击左下角的，展开选项菜单。

3. 选择“本地资源”页签，并单击“本地设备和资源”栏的“详细信息”。



4.勾选“驱动器”和“其他支持的即插即用（PnP）设备”，并单击“确定”，将本地主机上的所有磁盘映射到 Windows 云主机。



5.如果只需要映射部分本地主机上的磁盘到 Windows 云主机，请展开“驱动器”，勾选待映射的磁盘设备。

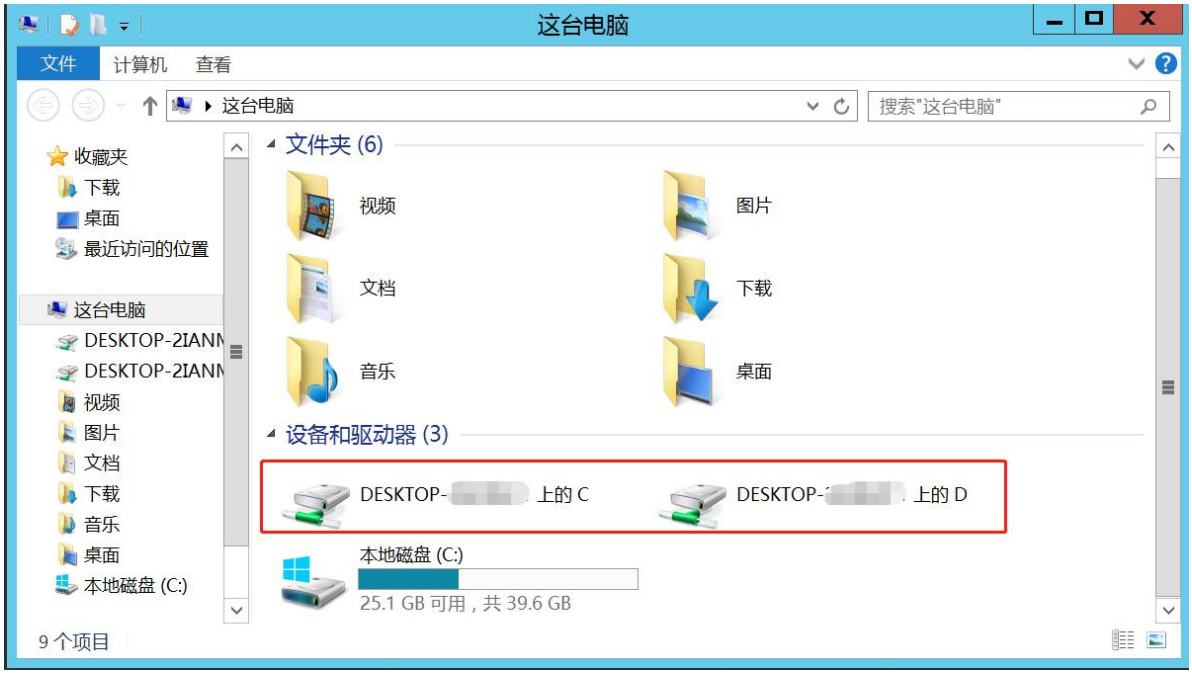


6. 再次打开远程桌面连接窗口，并在“计算机”栏输入 Windows 云主机的弹性公网 IP 地址。



7. 单击“连接”，登录 Windows 云主机。

8. 查看 Windows 云主机的磁盘信息，如果有显示本地主机的磁盘设备，则表示您的本地主机与 Windows 云主机之间可以互传数据了。



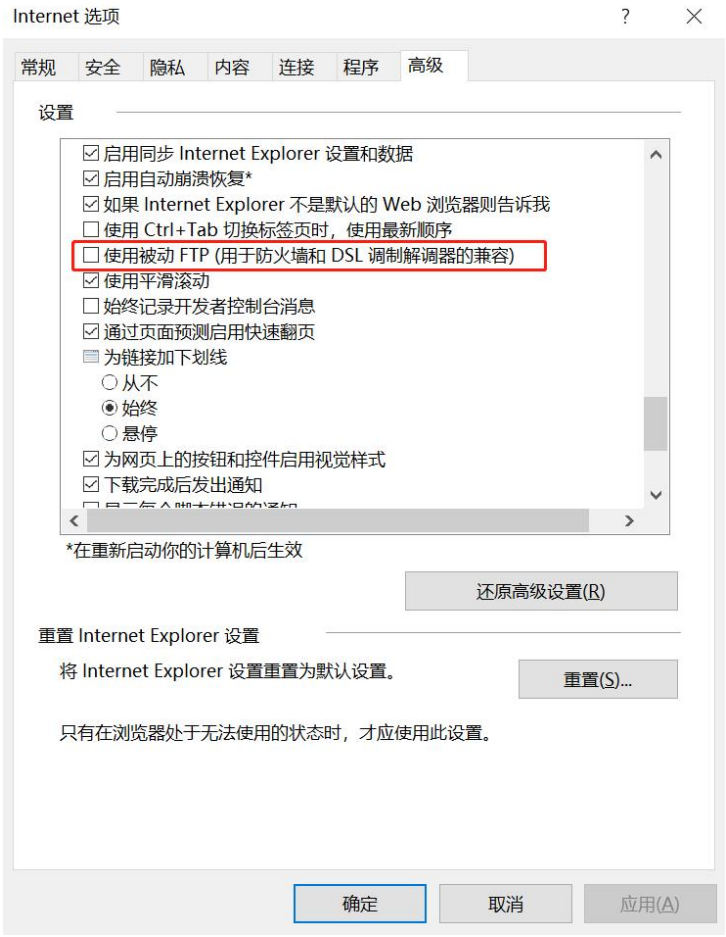
8.9.14FTP 服务器上的文件夹因权限访问出错怎么办？

当您打开 FTP 服务器上的文件夹时，遇到弹窗提示需检查访问权限，那可能是由于您的浏览器设置了 FTP 防火墙，您可以通过以下步骤来解决。

1. 单击浏览器菜单"工具 > Internet 选项"。

2. 选择“高级”标签卡。

3. 取消勾选“使用被动 FTP（用于防火墙和 DSL 调制解调器兼容）”。



4. 单击“确定”，重启 IE 浏览器，重试打开 FTP 服务器上的文件夹。

8.10 迁移云服务器

8.10.1 弹性云主机可以迁移到其他地域/帐号吗？

您可以通过镜像迁移方式实现云主机的跨帐号跨地域迁移。

- 将云主机制作整机镜像

在弹性云主机页面选中需要迁移的弹性云主机点击关机，关机完成后，点击右侧更多-制作镜像，选择云主机系统盘或数据盘，填写镜像名称，点击确认下单，等待镜像创建完成。

第一步：将镜像共享给其他账号，其他账号接收了共享镜像后，即可根据镜像创建新的云主机。

第二步：导出镜像文件，在待迁移地域，通过镜像文件制作镜像，待镜像完成，可以通过镜像创建新的云主机。

8.11 镜像源管理

8.11.1 内网 yum 源及 NTP 配置

内网 YUM 源的使用

操作场景

更新弹性云主机的系统或者软件时，可以连接 Internet，通过外部镜像源提供相关服务。但是，如果弹性云主机无法访问 Internet，或者外部镜像源提供的服务不稳定时，可以使用天翼云提供的内网 yum 源服务完成系统或软件的更新。

约束与限制

- 当前支持的资源池

  华东 1、广州 6、北京 5、内蒙 6、贵州 3、合肥 3、芜湖 2、西安 3、西安 4、杭州 2、上海 7、成都 4、重庆 2、温州 X 节点、广州 X 节点、南京 2、中卫 2 、拉萨 3、昆明 2、福州 4、十堰、南京 4。
- 当前支持的源

  Ubuntu 系列。

说明

- CentOS：2024 年 6 月 30 日开始，Redhat 官方已停止对 CentOS 7 发行版操作系统维护更新，官方该版本操作系统 rpm 包软件仓库已停用，天翼云也同步停用内网 yum 源的提供，至此 CentOS 全面停维，CentOS 内网 yum 停止提供。建议更换系统为其他官方仍维护中系统进行需求软件包安装。
- Ubuntu：当前仅支持公共镜像 Ubuntu 20.04 及以下版本的内网 yum 源。

操作步骤

部分资源池（华东 1）适用以下配置：yumserver 的地址配置为

http://169.254.169.253:10080。

部分资源池（广州 6、北京 5、内蒙 6、贵州 3、合肥 3、芜湖 2、西安 3、西安 4、杭州 2、上海 7、成都 4、重庆 2、温州 X 节点、广州 X 节点、南京 2、中卫 2 、拉萨 3、昆明 2、福州 4、十堰、南京 4）适用以下配置：yumserver 的地址配置为 http://100.126.0.130:10080。

配置中的 url 路径使用对应操作系统默认路径即可。

1. Ubuntu 系列操作步骤如下：

进入 apt 配置路径

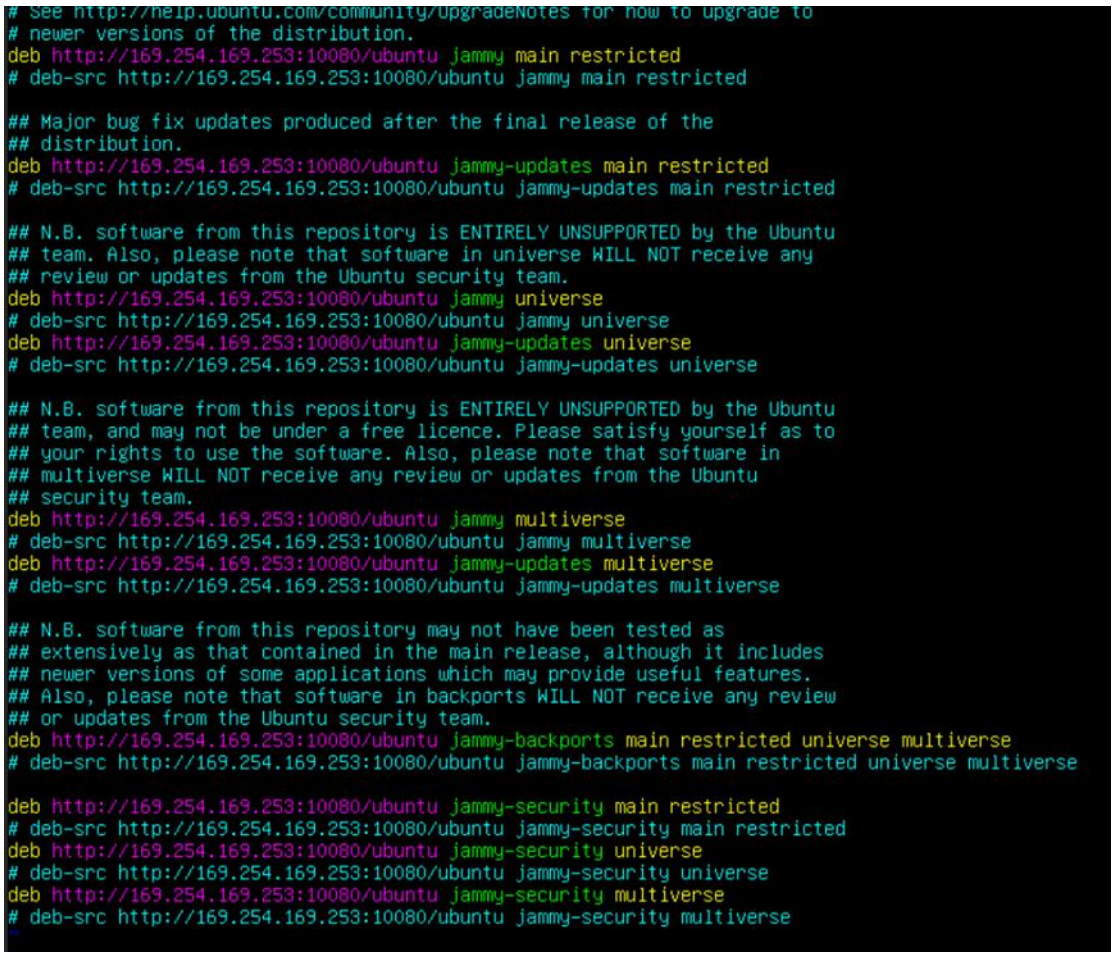
#cd /etc/apt

#ls

查看目录下的文件如下图：



2. 将 sources.list 做好备份，进入 sources.list 对源地址进行修改，如下图：



3. 修改完成后即可使用内网 apt 源。

NTP 的使用

操作场景

当用户想要进行时间同步时，可以使用 NTP 服务。

约束与限制

当前支持的资源池包括：

以下资源池使用 NTP Server 地址：169.254.169.254

华东 1、华北 2、华南 2、青岛 20、西南 1、南昌 5、长沙 42、上海 36、武汉 41、上海 15、西安 7、太原 4、芜湖 4、上海 17、郑州 5、呼和浩特 3、西南 2-贵州、杭州 7、庆阳 2。

以下资源池使用 NTP Server 地址：100.126.0.162

北京 5、成都 4、拉萨 3、上海 7、广州 6、郴州 2、兰州 2、内蒙 6、南京 2、南京 4、九江、西安 3、西安 4、西安 5、福州 4、重庆 2、芜湖 2、合肥 2、贵州 3、

杭州 2、中卫 2、昆明 2、乌鲁木齐 4。

操作步骤

1. 登录弹性云主机；
2. 部分资源池（如华东 1）适用以下配置：通过 ntpdate 169.254.169.254 同步即可；部分资源池（如北京 5）适用以下配置：通过 ntpdate 100.126.0.162 同步即可。

8.11.2Windows 或 Linux 操作系统镜像怎么选

Windows 操作系统与 Linux 操作系统没有好坏之分，只有适合不适合。请根据业务需求选择 Windows 或 Linux 的操作系统。

Windows2012、Windows2016、Windows2019 操作系统网络优化和系统兼容性比较好，建议选择最高版本 Windows2019。

Linux 提供了多种操作系统和版本，不同操作系统之间没有好坏之分，请根据实际业务选择适合自己的操作系统版本。

购买弹性云主机选错操作系统，可以通过控制台的“一键重装”功能更换操作系统。

1.选择待切换操作系统的弹性云主机，单击弹性云主机列表左上角的“关机”。
2.在待切换操作系统的弹性云主机的“操作”列下，单击“更多 > 一键重装”。
3.系统进入“切换操作系统”界面。根据需求选择需要更换的“镜像类型”和“镜像”。
4.设置登录方式，目前重装时登录凭证仅支持使用密码的方式。
5.配置完成后单击“确定”，提交申请切换操作系统。
6.提交切换操作系统的申请后，弹性云主机的状态变为“切换中”，当该状态消失后，表示切换结束。

8.11.3 常见 docker 源配置指导

配置使用清华大学 docker 源

Debian/Ubuntu 用户

如果你过去安装过 docker，为避免产生冲突，可以先删除再重新安装：

```
for pkg in docker.io docker-doc docker-compose podman-docker containerd runc; do apt-get remove $pkg; done
```

首先安装依赖：

```
apt-get update

apt-get install ca-certificates curl gnupg
```

信任 Docker 的 GPG 公钥并添加仓库：

发行版：Debian

```
install -m 0755 -d /etc/apt/keyrings

curl -fsSL https://download.docker.com/linux/debian/gpg | gpg --dearmor -o /etc/apt/keyrings/docker.gpg

sudo chmod a+r /etc/apt/keyrings/docker.gpg

echo \

  "deb [arch=$(dpkg --print-architecture) signed-by=/etc/apt/keyrings/docker.gpg] https://mirrors.tuna.tsinghua.edu.cn/docker-ce/linux/debian \

  "$(. /etc/os-release && echo "$VERSION_CODENAME")" stable" | \

  tee /etc/apt/sources.list.d/docker.list > /dev/null
```

发行版：Ubuntu

```
install -m 0755 -d /etc/apt/keyrings

curl -fsSL https://download.docker.com/linux/ubuntu/gpg | gpg --dearmor -o /etc/apt/keyrings/docker.gpg

sudo chmod a+r /etc/apt/keyrings/docker.gpg

echo \
```

```
  "deb [arch=$(dpkg --print-architecture)
signed-by=/etc/apt/keyrings/docker.gpg]
https://mirrors.tuna.tsinghua.edu.cn/docker-ce/linux/ubuntu \
  "$(. /etc/os-release && echo "$VERSION_CODENAME")" stable" | \
  tee /etc/apt/sources.list.d/docker.list > /dev/null
```

最后安装:

```
apt-get update

apt-get install docker-ce docker-ce-cli containerd.io
docker-buildx-plugin docker-compose-plugin
```

Fedora 用户

如果你过去安装过 docker，为避免产生冲突，可以先删除再重新安装：

```
dnf remove docker \
                  docker-client \
                  docker-client-latest \
                  docker-common \
                  docker-latest \
                  docker-latest-logrotate \
                  docker-logrotate \
                  docker-selinux \
                  docker-engine-selinux \
                  docker-engine
```

安装依赖，下载 repo 文件，并把软件仓库地址替换为镜像站：

```
dnf -y install dnf-plugins-core
```

```
dnf config-manager --add-repo
https://download.docker.com/linux/fedora/docker-ce.repo

sed -i
's+https://download.docker.com+https://mirrors.tuna.tsinghua.edu.cn/docker-ce+' /etc/yum.repos.d/docker-ce.repo
```

最后安装：

```
dnf install docker-ce docker-ce-cli containerd.io docker-buildx-plugin
docker-compose-plugin
```

CentOS 用户

如果你过去安装过 docker，为避免产生冲突，可以先删除再重新安装：

```
yum remove docker \
                  docker-client \
                  docker-client-latest \
                  docker-common \
                  docker-latest \
                  docker-latest-logrotate \
                  docker-logrotate \
                  docker-engine
```

安装依赖，下载 repo 文件，并把软件仓库地址替换为镜像站：

```
yum install -y yum-utils

yum-config-manager --add-repo
https://download.docker.com/linux/centos/docker-ce.repo
```

```
sed -i
's+https://download.docker.com+https://mirrors.tuna.tsinghua.edu.cn/d
ocker-ce+' /etc/yum.repos.d/docker-ce.repo
```

最后安装：

```
yum install docker-ce docker-ce-cli containerd.io docker-buildx-plugin
docker-compose-plugin
```

RHEL 用户

如果你过去安装过 docker，为避免产生冲突，可以先删除再重新安装：

```
yum remove docker \
                  docker-client \
                  docker-client-latest \
                  docker-common \
                  docker-latest \
                  docker-latest-logrotate \
                  docker-logrotate \
                  docker-engine
```

安装依赖，下载 repo 文件，并把软件仓库地址替换为镜像站：

```
yum install -y yum-utils
yum-config-manager --add-repo
https://download.docker.com/linux/rhel/docker-ce.repo
sed -i
's+https://download.docker.com+https://mirrors.tuna.tsinghua.edu.cn/d
ocker-ce+' /etc/yum.repos.d/docker-ce.repo
```

最后安装：

```
yum install docker-ce docker-ce-cli containerd.io docker-buildx-plugin
docker-compose-plugin
```

配置使用阿里云 docker 源

Ubuntu14.04/16.04 用户

```
# step 1: 安装必要的一些系统工具

sudo apt-get update

sudo apt-get -y install apt-transport-https ca-certificates curl
software-properties-common

# step 2: 安装 GPG 证书

curl -fsSL https://mirrors.aliyun.com/docker-ce/linux/ubuntu/gpg |
sudo apt-key add -

# Step 3: 写入软件源信息

sudo add-apt-repository "deb [arch=amd64]
https://mirrors.aliyun.com/docker-ce/linux/ubuntu $(lsb_release -cs)
stable"

# Step 4: 更新并安装 Docker-CE

sudo apt-get -y update

sudo apt-get -y install docker-ce


# 安装指定版本的 Docker-CE:

# Step 1: 查找 Docker-CE 的版本:

# apt-cache madison docker-ce
```

```
#   docker-ce | 17.03.1~ce-0~ubuntu-xenial |
https://mirrors.aliyun.com/docker-ce/linux/ubuntu xenial/stable amd64
Packages

#   docker-ce | 17.03.0~ce-0~ubuntu-xenial |
https://mirrors.aliyun.com/docker-ce/linux/ubuntu xenial/stable amd64
Packages

# Step 2: 安装指定版本的 Docker-CE: (VERSION 例如上面的
17.03.1~ce-0~ubuntu-xenial)

# sudo apt-get -y install docker-ce=[VERSION]
```

CentOS7 用户

```
# step 1: 安装必要的一些系统工具

sudo yum install -y yum-utils device-mapper-persistent-data lvm2

# Step 2: 添加软件源信息

sudo yum-config-manager --add-repo
https://mirrors.aliyun.com/docker-ce/linux/centos/docker-ce.repo

# Step 3

sudo sed -i 's+download.docker.com+mirrors.aliyun.com/docker-ce+'
/etc/yum.repos.d/docker-ce.repo

# Step 4: 更新并安装 Docker-CE

sudo yum makecache fast

sudo yum -y install docker-ce

# Step 4: 开启 Docker 服务

sudo service docker start
```

```
# 注意：
# 官方软件源默认启用了最新的软件，您可以通过编辑软件源的方式获取各个版本的软件包。例如官方并没有将测试版本的软件源置为可用，您可以通过以下方式开启。同理可以开启各种测试版本等。
# vim /etc/yum.repos.d/docker-ce.repo
#   将[docker-ce-test]下方的 enabled=0 修改为 enabled=1
#
# 安装指定版本的 Docker-CE:
# Step 1: 查找 Docker-CE 的版本:
# yum list docker-ce.x86_64 --showduplicates | sort -r
#   Loading mirror speeds from cached hostfile
#   Loaded plugins: branch, fastestmirror, langpacks
#   docker-ce.x86_64            17.03.1.ce-1.el7.centos
docker-ce-stable
#   docker-ce.x86_64            17.03.1.ce-1.el7.centos
@docker-ce-stable
#   docker-ce.x86_64            17.03.0.ce-1.el7.centos
docker-ce-stable
#   Available Packages
# Step2: 安装指定版本的 Docker-CE: (VERSION 例如上面的 17.03.0.ce.1-1.el7.centos)
# sudo yum -y install docker-ce-[VERSION]
```

校验：

```
docker version
```

具体支持的 Guest OS 镜像参考天翼云镜像服务。

8.12 磁盘管理类

8.12.1 磁盘分区与虚拟内存类

Windows 操作系统云主机看不到挂载的数据盘怎么办？

Windows 操作系统云主机有时会看不到挂载的数据盘，出现这种情况时，可能是因为新添加的数据盘未进行分区和初始化。

新的数据盘默认是没有分区和文件系统的，因此在资源管理器中不会显示，需要先手动执行初始化操作。具体方法请参见初始化数据盘。

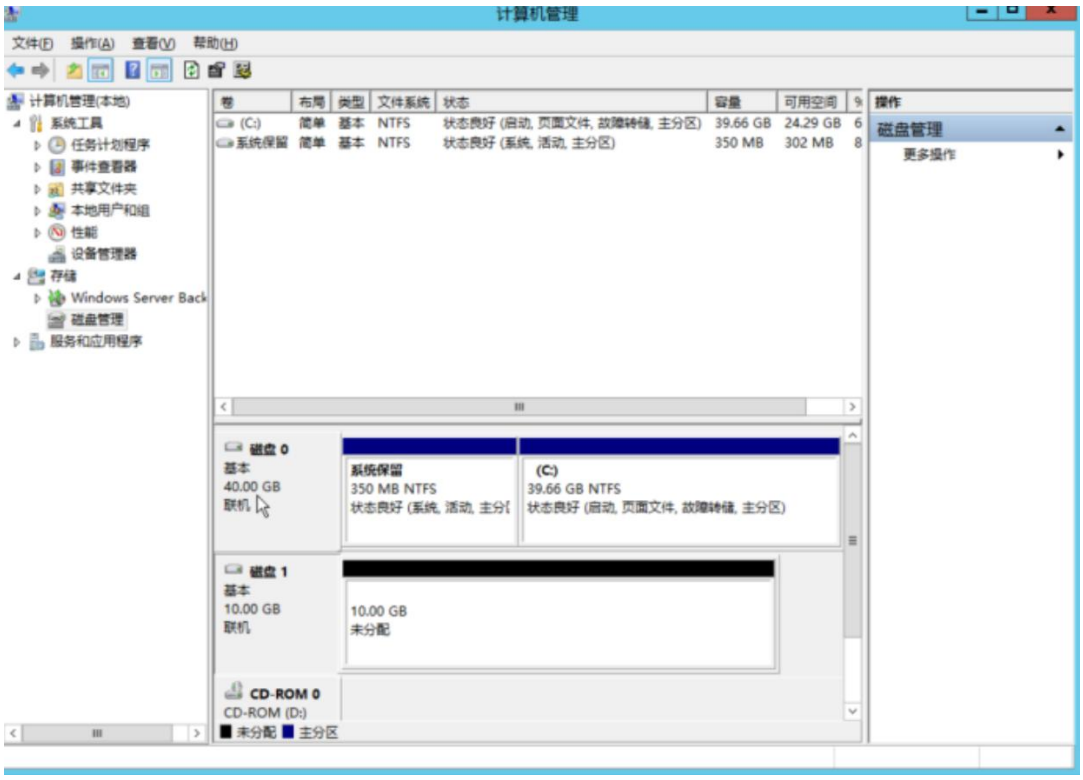
怎样调整系统盘分区？

通常我们在进行系统盘扩容的时候，会用到调整系统盘分区，Windows 云主机和 Linux 云主机的调整方式不同，具体步骤分别参见扩容云硬盘中针对 Windows 云主机和 Linux 云主机的不同调整方式。

怎样查询 Windows 弹性云主机磁盘分区与磁盘设备的对应关系？

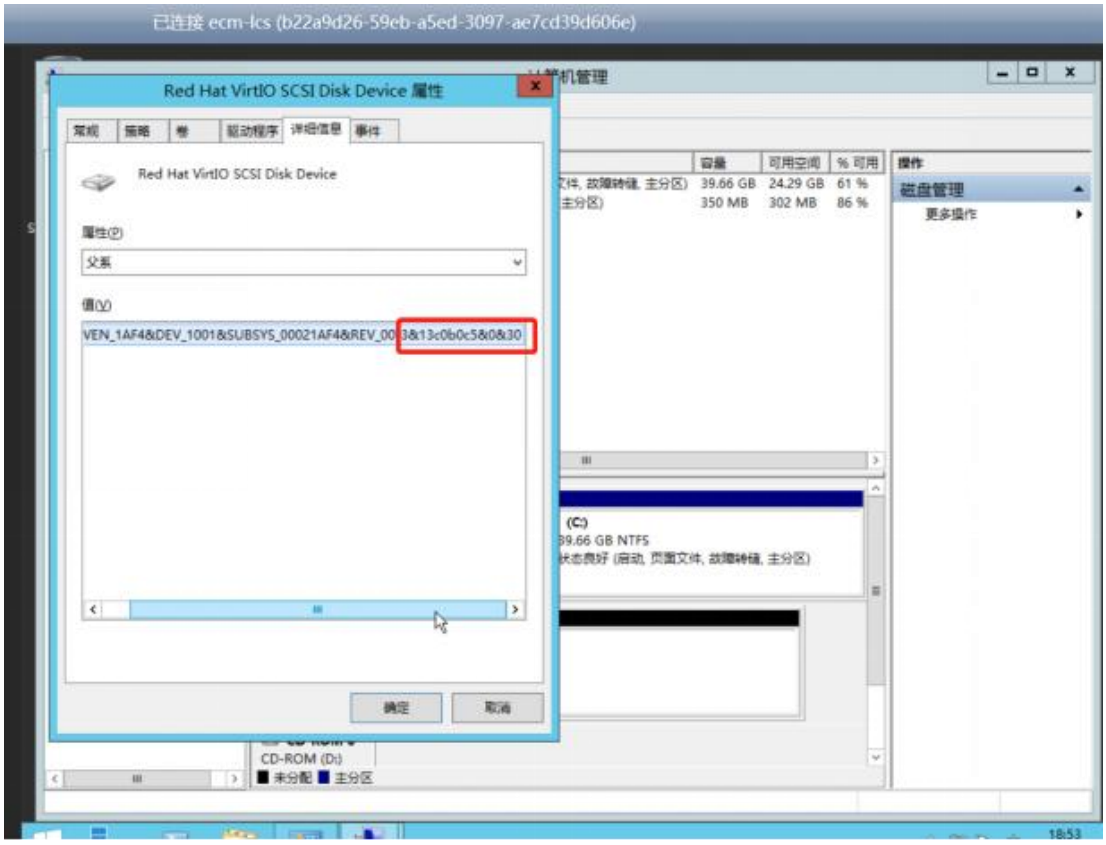
以 Windows 2012 为例，介绍查询 Windows 弹性云主机磁盘分区与磁盘设备对应关系的方法。

1. 登录 Windows 弹性云主机。单击桌面左下角的“开始”菜单。
2. 选择“管理工具 > 计算机管理”。
3. 在左侧导航栏，选择“存储 > 磁盘管理”。



4. 右键单击“磁盘 1”所在灰色地域，单击“属性”。系统打开磁盘属性窗口。

5. 选择“详细信息”页签，并选择参数“属性”的值为“父系”。

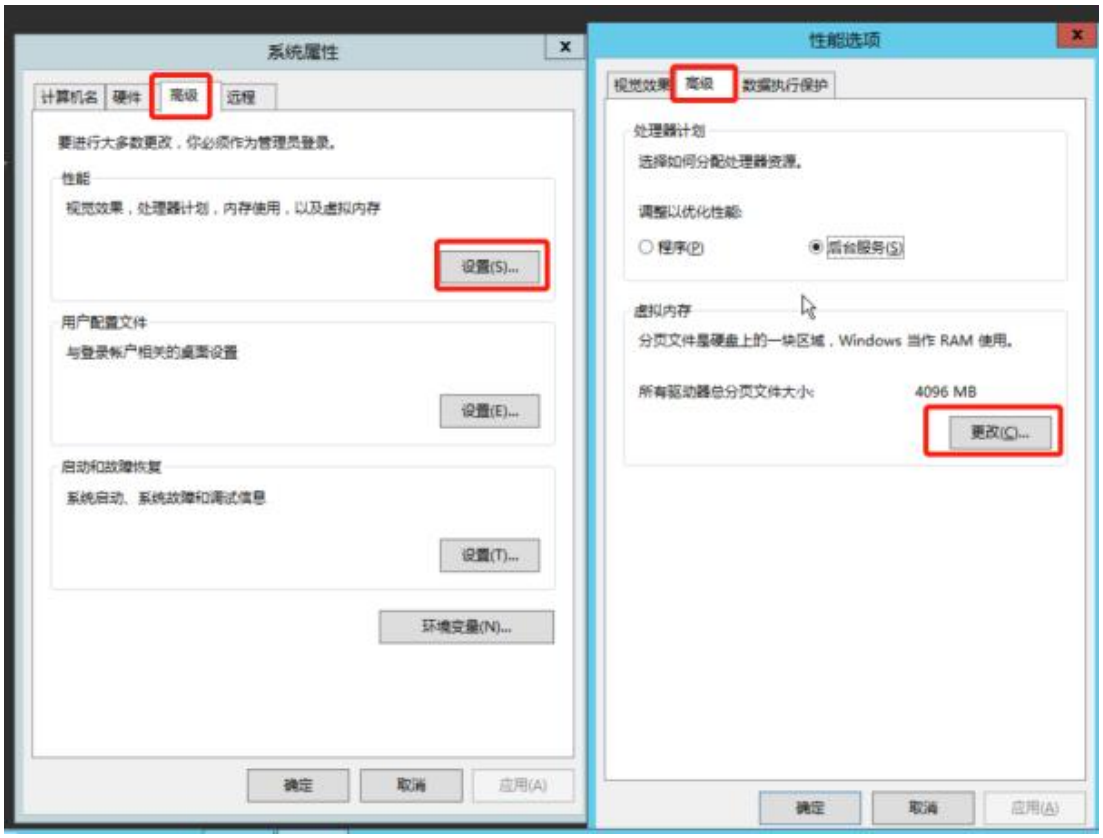


6.记录参数“值”中符号“&”后的几位数字，该值为磁盘分区对应的主从设备号。

怎样配置 Windows 弹性云主机的虚拟内存？

当弹性云主机内存不足时，建议通过变更规格操作来扩大内存。若因业务需要，必须开启虚拟内存，请参见本节内容进行配置。

1.登录弹性云主机，单击开始，选择计算机，右键选择属性，点击高级系统设置。
2.点击“高级”页签，在此页签内，点击“性能”的“设置(S)”按钮，弹出“性能选项”对话框。
3.点击“高级”页签，在此页签内，点击“虚拟内存”的“更改(C)”按钮。



4.单击“设置”，然后单击右下角的“确定”完成虚拟内存配置。



5. 重启弹性云主机，使配置生效。

8.12.2 磁盘扩容类

扩容类常见问题

8.12.3 卸载云硬盘操作指南及故障排除

问题场景

在使用弹性云主机+云硬盘的服务时，您可能需要将磁盘属性为数据盘的云硬盘从当前云主机卸载，以便挂载至另一台弹性云主机继续使用。请注意，卸载云硬

盘的操作不会清除硬盘上的数据。

支持范围

适用对象：仅支持数据盘的卸载，系统盘不支持卸载操作。

前提条件与注意事项

1. 为保证数据完整性，在卸载前，建议暂停所有针对该磁盘的文件系统读写操作，以防止数据丢失。如果有服务正在运行并使用该云硬盘，请先停止这些服务，确认所有依赖于该云硬盘的服务都已停止后再卸载云硬盘。
2. 卸载云硬盘前建议先执行脱机（Windows）操作或者 umount（Linux）操作，否则可能会导致云主机卸载云硬盘失败或者云主机再次挂载云硬盘时无法识别等异常情况。
3. 若出现卸载失败等异常情况，可以先检查进程占用，例如使用任务管理器（Windows）查找并终止占用磁盘的进程，然后重新尝试卸载。或者可以先对云主机关机后再重新尝试卸载。

如遇复杂问题，建议联系技术支持获取进一步协助。

8.12.4 其它

包年　/包月的云硬盘怎么删除或者退订？

随云主机购买的包年/包月的云硬盘不支持单独退订。需要和云主机一同退订。

单独购买的包年/包月云硬盘支持单独退订。

退订云主机时，云硬盘会一起退订吗？

单独购买的云硬盘：单独购买的云硬盘如果已挂载至云主机，在删除云主机时，系统会提示是否同步删除挂载的云硬盘，您可以根据实际情况进行选择。

随云主机购买的云硬盘：退订云主机时，云硬盘同步退订。

系统盘和数据盘有什么不同？

系统盘 ：云主机中用于安装操作系统、存放系统配置文件的云硬盘。类似于个人传统 PC 中的 C 盘。用户在购买云主机时，必须同时购买系统盘，购买成功后，系统盘会自动挂载至云主机。系统盘的最大容量为 2048GB。

数据盘 ：在云主机与物理机中存放数据的云硬盘，类似于个人传统 PC 中的 D 盘、E 盘、F 盘。数据盘有两种购买方式，一种是在购买云主机时购买，此时数据盘将会自动挂载至云主机。第二种购买方式为在云硬盘管理控制台单独购买，单独购买之后用户需要将云硬盘手动挂载至云主机。数据盘的最大容量为 32TB（32768 GB）。

用户可根据实际的业务需求与成本预算来决定系统盘的具体容量以及是否购买数据盘。当云主机承载业务数据量较小，且系统盘可以满足需求，则可以不购买数据盘，若数据量比较大，建议您购买数据盘。

系统盘可以单独购买吗？

系统盘不能单独购买。

系统盘是弹性云主机实例的绑定资源，只能在创建弹性云主机实例的过程中进行设置。相应的，系统盘的计费模式与生命周期与对应的弹性云主机实例完全一致。

虽然不能单独购买，但系统盘资源仍然按云硬盘的标准资费独立收费。详细资费请参见云硬盘计费说明：云硬盘计费说明

怎样为云主机增加数据盘？

云主机增加数据盘有两种方式，一种是在购买云主机的过程中直接添加数据盘。另一种是在云硬盘控制台中创建云硬盘，然后将其挂载至云主机中。用户可在弹性云主机详情页面云硬盘页签中选择云硬盘挂载，也可以在云硬盘列表中选择云主机进行挂载。详情请参见创建云硬盘以及挂载云硬盘。

为什么无法卸载云硬盘？

云硬盘挂载至弹性云主机时有两种形态，一种是系统盘，一种是数据盘。系统盘是无法卸载的，数据盘可以卸载。

当卸载数据盘时，支持离线卸载或在线卸载，即可在挂载该数据盘的云主机处于“运行中”或“关机”状态时进行卸载。

云硬盘容量不足了怎么办？

当您云硬盘容量不足且可能影响到自身业务时，可以参考以下内容来解决此问题：

单独购买一块云硬盘做数据盘，并挂载至弹性云主机，挂载成功之后进行初始化。具体操作步骤请参见云硬盘快速入门。

扩容当前已有云硬盘，系统盘和数据盘均可进行扩容。操作步骤请参见扩容云硬盘。

清理当前云硬盘上不需要的程序、文件来清理磁盘空间，具体步骤可参见：

解决 Linux 云主机磁盘空间不足的问题

解决 Windows 弹性云主机磁盘空间不足的问题

本地盘实例怎么添加到云主机组？

本地盘实例　创建后无法加入云主机组　，如需使用云主机组功能，请在购买时选择云主机组。

云主机组是对云主机的一种逻辑划分，云主机组中的弹性云主机遵从同一策略，如强制反亲和性、强制亲和性、反亲和性、亲和性等。

如何只对系统盘进行快照？

操作方法如下：

1.在云主机控制台选中该云主机，点击进入详情页

2.点击云硬盘标签，在磁盘属性中找到“系统盘”，复制该云盘名称。

3.在“控制中心-所有服务”页面，单击“存储>云硬盘”，进入云硬盘主页面

4.在云硬盘列表页搜索第 2 步复制好的云盘名称，找到该系统盘。

5.点击该系统盘“操作>创建快照”，进入“创建快照”页面，创建快照。

更多操作信息可参考创建云硬盘快照。

8.13 密码与密钥对类

8.13.1 密码

8.13.1.1 云主机默认密码是多少？

Linux 操作系统默认密码是多少？

Linux 系统的弹性云主机，默认用户名是“root”，密码是您购买弹性云主机时自行创建的密码。

如忘记登录密码或创建云主机时未设置密码，可通过 “ 重置密码 ” 功能设置新密码。

具体步骤可参考云主机重置密码。

Windows 操作系统默认密码是多少？

Windows 操作系统没有默认密码，密码是您购买弹性云主机时自行创建的密码。

如忘记登录密码或创建云主机时未设置密码，可通过 “ 重置密码 ” 功能设置新密码。

具体步骤可参考云主机重置密码。

8.13.1.2 忘记天翼云账号密码怎么办？

忘记天翼云账号密码怎么办？

当用户忘记密码时，可通过登录页面点击“忘记密码”进行重置密码。

1. 进入天翼云官网，点击右上角“登录”，在登录页面点击右下角的“忘记密码”。
2. 填写需要找回的账号，并进行验证。填写邮箱，验证码会通过邮箱发送。填写手机号，验证码会通过手机发送。

说明

邮箱需在管理中心-基本信息完成邮箱验证后，方可接收验证码。

3. 验证成功后，设置新的登录密码。

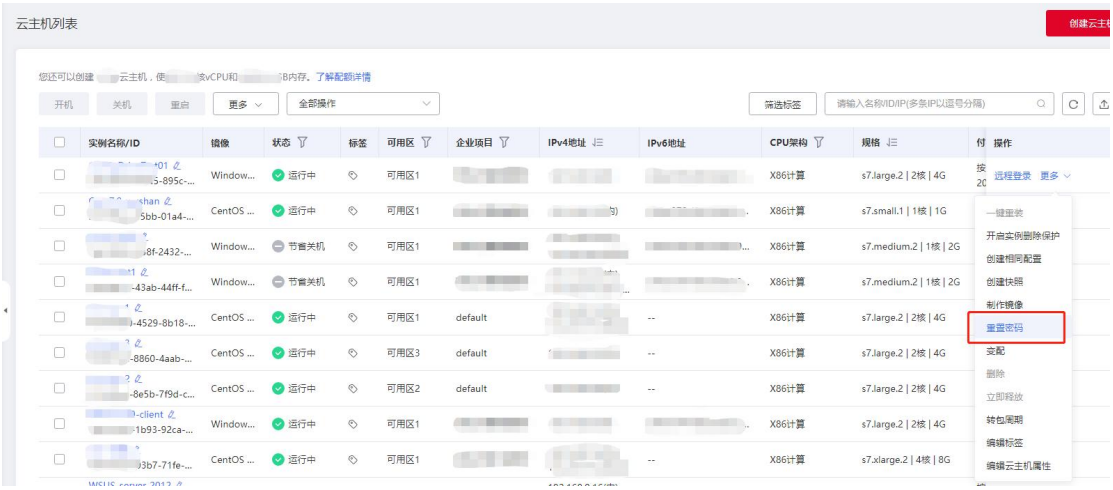
8.13.1.3 如何在控制台重置弹性云主机密码？

问题描述

云主机密码遗忘，需要在控制台快速重置密码。

操作步骤

1. 登录管理控制台。
2. 单击管理控制台左上角的  ，选择区域和项目。
3. 选择“产品服务列表 > 弹性云主机”。
4. 选择需重置密码云主机右侧的更多，选择下拉选项中的“重置密码”。



5. 按照提示的密码要求输入符合要求的密码。

8.13.1.4 如何在操作系统内部修改云主机密码？

- Windows 云主机

1. 登录 Windows 弹性云主机。
2. 使用快捷键“Win+R”打开“运行”页面。
3. 输入命令行“cmd”打开命令行窗口。
4. 执行以下命令，修改密码。

```
net user Administrator 新密码
```

- Linux 云主机

1. 以 root 用户登录 Linux 弹性云主机。
2. 执行以下命令，重置 root 的用户密码。

```
passwd
```

根据系统回显信息，输入新密码。

系统显示如下回显信息时，表示密码重置成功。

```
passwd: all authentication tokens updates successfully
```

8.13.1.5 用户创建弹性云主机成功后远程登录提示密码错误？

如远程登录提示密码错误，可通过 ″重置密码″ 功能设置新密码。

您可以通过弹性云主机控制台重置密码，点击更多-重置密码更改密码，使用新密码再进行登录。具体操作步骤如下：

1. 打开弹性云主机控制台页面。
2. 选择您需要重置密码的弹性云主机，点击“更多”按钮。
3. 在弹出的下拉菜单中，选择“重置密码”选项。
4. 在弹出的重置密码窗口中，您需要输入一个新的密码，并确认一遍输入正确的密码。请注意密码的规则为 8～30 个字符，必须同时包含三项（大写字母、小写字母、数字、 ()`~!@#$%^&*_-+=|{}[]:;'<>,.?/ 中的特殊符号），且不能以斜线号（/）开头，不能包含连续字符。

5. 点击“确定”按钮，系统会提示您，重置密码成功。

6. 使用新密码登录弹性云主机控制台，进行后续操作。

更多可以参考在控制台重置密码。

8.13.1.6 一键重置密码后无法使用新密码登录弹性云主机

问题描述

用户在一键重置密码操作后，出现无法使用新密码登录弹性云主机的情况。

解决办法

1. 登录控制台，点击弹性云主机页面。

2. 选择对应弹性云主机，点击更多，选择重置密码。



3. 重置密码后，点击重启。



4. 若以上方法无法解决问题，请联系工单进行处理。

8.13.1.7 用户未修改密码，正常使用弹性云主机一段时间后无法登录

问题描述

用户没有进行修改密码操作，但在正常使用云主机一段时间后出现无法登录的情况。

处理方法

确认弹性云主机远程连接是否能进入登录界面。

1.如果不能进入登录界面，可能是云主机操作系统进程异常，请联系客服进行定位。
2.如果可以进入登录界面，需要进入操作系统检查是否可以修改密码。
3. 若以上方法无法解决问题，请联系工单进行处理。

8.13.1.8 如何关闭 Selinux？

问题描述

用户需要关闭 Selinux。

操作步骤

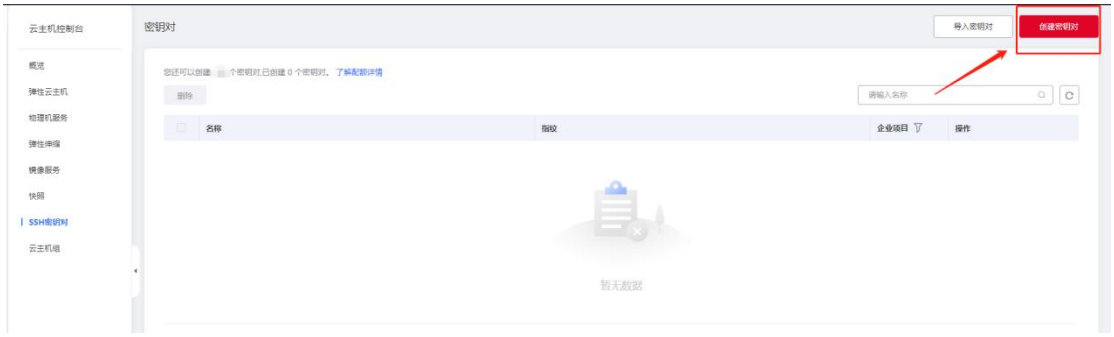
1.远程登录弹性云主机，打开命令行。
2.执行命令 vi /etc/selinux/config。
3.输入“i”进入编辑模式，修改 SELINUX=disabled。
4.修改完成后，键入“ESC”，输入 :wq 保存并退出文件编辑。

8.13.2 密钥对

8.13.2.1 如何创建密钥对？

处理方法

1.登录管理控制台，左侧边栏选择“SSH 密钥对”。
2.单击“创建密钥对”。



3. 在“创建密钥对”弹窗中，填写密钥对名称、选择企业项目。单击“确定”将密钥对保存到本地。

8.13.2.2 如何使用密钥对？

在弹性云主机控制台，点击左侧导航栏的 SSH 密钥对，点击创建密钥对，即可生成密钥对，同时将密钥对下载到本地。

您可以在创建弹性云主机时，登录方式选择密钥对，绑定已创建的密钥对用于弹性云主机登录。

SSH 密钥对登录方式详见弹性云主机-登录 Linux 弹性云主机。

8.13.2.3 密钥对无法下载

密钥对的私钥文件只能下载一次。 如果密钥对私钥文件已丢失，建议您重新创建并下载保存。

处理方法

1. 登录云主机控制台，选择“密钥对”；
2. 单击“创建密钥对”；
3. 单击“确定”将密钥对保存到本地。

8.13.2.4 无法导入密钥对

修改浏览器默认属性后重试，步骤如下：

1. 在浏览器主界面，点击设置。

2. 选择“Internet 选项”。

3. 单击选择“安全”页签。



4. 单击“Internet”。

5. 单击“默认级别”按钮，把设置还原为默认级别。



6. 把安全级别调到“中”级别，单击“应用”按钮。

7. 选择“自定义级别”。

8.将“对未标记为可安全执行脚本的 ActiveX 控件初始化并执行脚本”设置为“提示”。



9.单击“确定”。

8.13.2.5 怎样查询弹性云主机使用的密钥对是哪个？

问题描述

当用户创建了多个密钥对时，可能会混淆登录云主机使用的密钥对，因此需要查询弹性云主机使用的密钥对是哪个。

操作步骤

1.打开天翼云官网，登录并点击“控制中心”，进入云主机控制台。

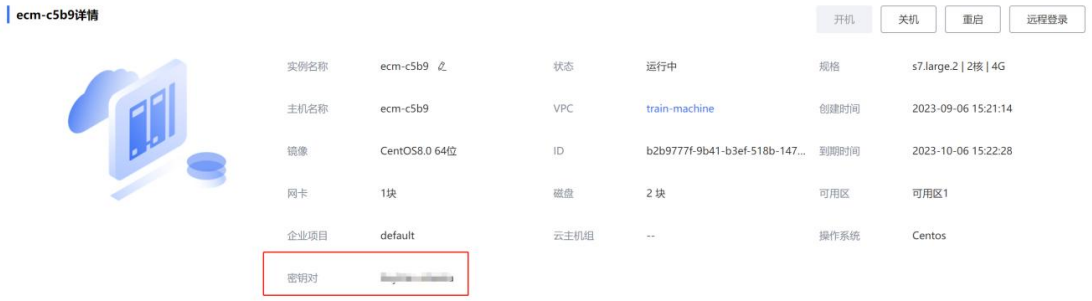
2. 选择区域，进入云主机列表页。

图 1 区域选择



3. 在云主机列表页中，选择需要查看密钥对的云主机，点击“实例名称”一列，进入云主机详情页即可查看该云主机所使用的密钥对。

图 2 云主机详情页



8.13.2.6 已开云主机是否还支持使用密钥对？

- 若创建云主机时登录方式选择“密钥对”，则支持使用。
- 若创建云主机时登录方式选择“密码”或想更改原绑定的密钥对，操作步骤请参考如何更改密钥对。

8.13.2.7 如何在云主机上使用安装 USB 类型的 Ukey 秘钥？

目前云主机不支持使用 USB key，USB key 只支持插在宿主机上，不支持插在笔记本电脑上。

8.13.2.8 弹性云主机中使用 cloudbase-init 是什么？

Cloudbase-Init 是一个开源的系统初始化工具，用于云环境中的虚拟机和物理机初始化。它支持在不同的云平台上进行自动配置和自定义操作，包括创建用户、安装软件包、配置网络和安全等。

在 Windows 上使用 Cloudbase-Init

如果你需要在 Windows 上使用 Cloudbase-Init，可以按照以下步骤进行安装和配置：

1、下载最新版本的 Cloudbase-Init 安装包（msi 文件）。

2、执行安装包，按照提示完成安装过程。

3、在安装目录下的配置文件（cloudbase-init.conf）中，配置需要的参数，如用户名、密码、网络配置等。你可以根据需要进行自定义配置。

4、在 Windows 服务中，启动并设置 Cloudbase-Init 服务为自动启动。

5、 重启系统，Cloudbase-Init 将会在系统启动时自动运行，并根据配置文件进行初始化。

在 Linux 上使用 Cloud-Init

如果你需要在 Linux 上使用 Cloud-Init，可以按照以下步骤进行安装和配置：

1、使用相应的包管理工具，如 yum 或 apt-get，安装 Cloud-Init 软件包。

2、在配置文件（cloud-init.conf）中，进行必要的参数配置，如用户名、密码、网络配置等。你可以根据需要进行自定义配置。

3、启动 Cloud-Init 服务，可以使用命令 systemctl start cloud-init（适用于 systemd 系统）或 service cloud-init start（适用于 init 系统）。

4、重启系统，Cloud-Init 将会在系统启动时自动运行，并根据配置文件进行初始化。

8.13.2.9Python 升级导致 Cloud-init 不工作

问题描述

Python 升级后导致 Cloud-init 无法正常工作。

可能原因

安装 Cloud-Init 时，Python 解释器默认使用 Python（即 /usr/bin/python 与 /bin/python 这两个软链接指向 Python2），当用户业务有需要时，可能会在实例内部把 Python 的默认解释器改为 Python3（即修改 /usr/bin/python 与 /bin/python 这两个软链接，使其指向 Python3）。由于兼容性问题，导致在开机启动执行 Cloud-Init 时报错。

解决办法

修改 /usr/bin/cloud-init 文件中指定的 Python 解释器，将 #!/usr/bin/python 或 #!/bin/python 修改为#!/usr/bin/python2.7。

8.13.2.10 使用密钥文件无法正常登录 Linux 弹性云主机

问题描述

使用密钥文件无法正常登录 Linux 弹性云主机。

可能原因

Linux 弹性云主机在创建的时时候，获取密钥失败。

处理方法

1.选择需要获取密钥的弹性云主机，点击“更多”，选择“重启”。

2.重启成功后，再次选择使用密钥文件登录弹性云主机，确认是否能够正常登录弹性云主机。

3.若以上方法无法解决问题，请联系工单进行处理。

8.13.2.11 通过 puttygen.exe 工具创建的密钥对，导入管理控制台失败

问题描述

通过 puttygen.exe 工具创建的密钥对，导入天翼云管理控制台失败。

可能原因

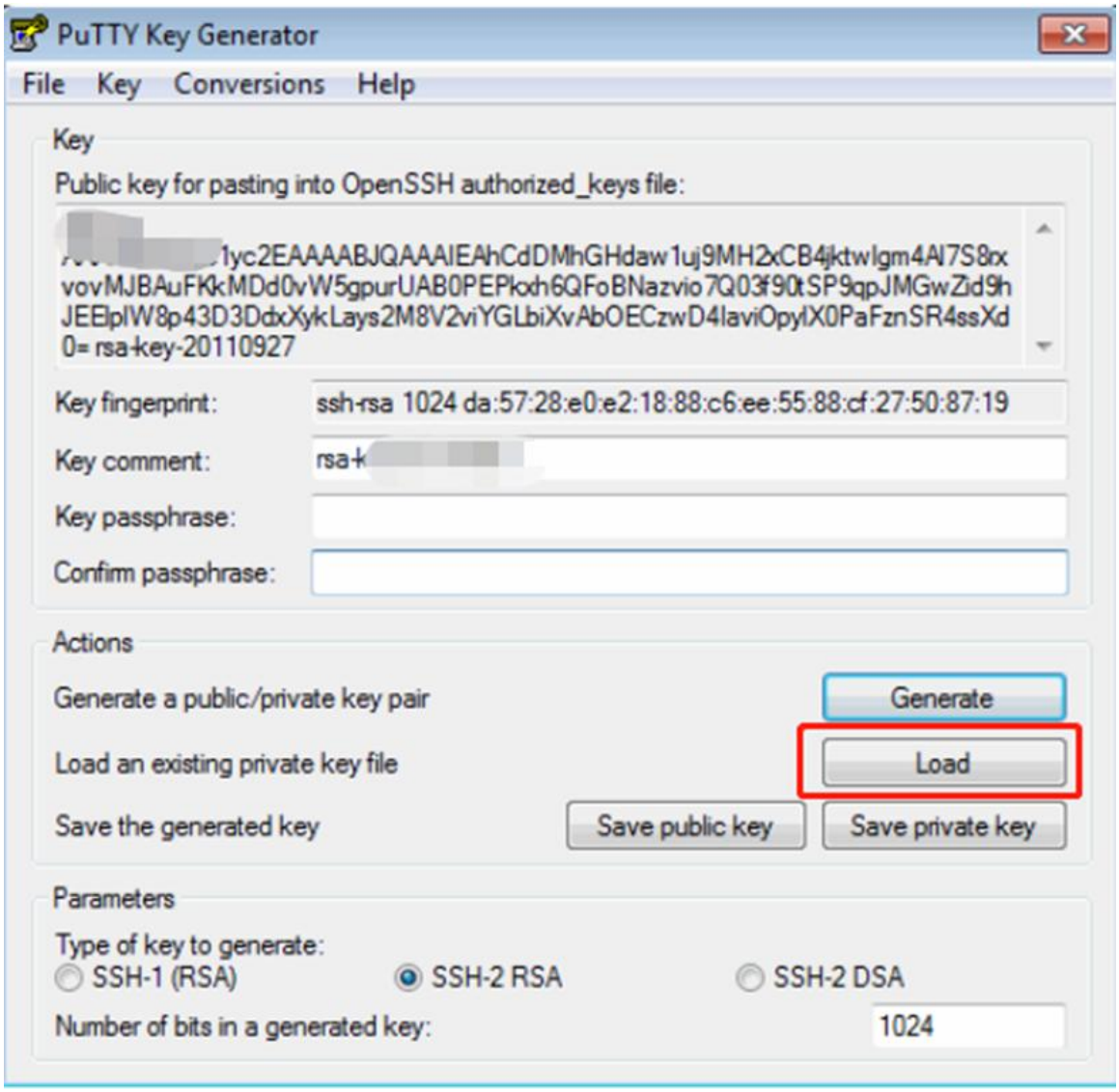
公钥格式不符合系统要求，如果使用 puttygen 工具中的“Save public key”保存公钥，公钥内容的格式会发生变化，不能直接导入天翼云控制台使用。

处理方法

需要将本地保存的私钥文件用“PuTTY Key Generator”恢复成格式正确的公钥文件，然后再将该公钥文件导入天翼云控制台。

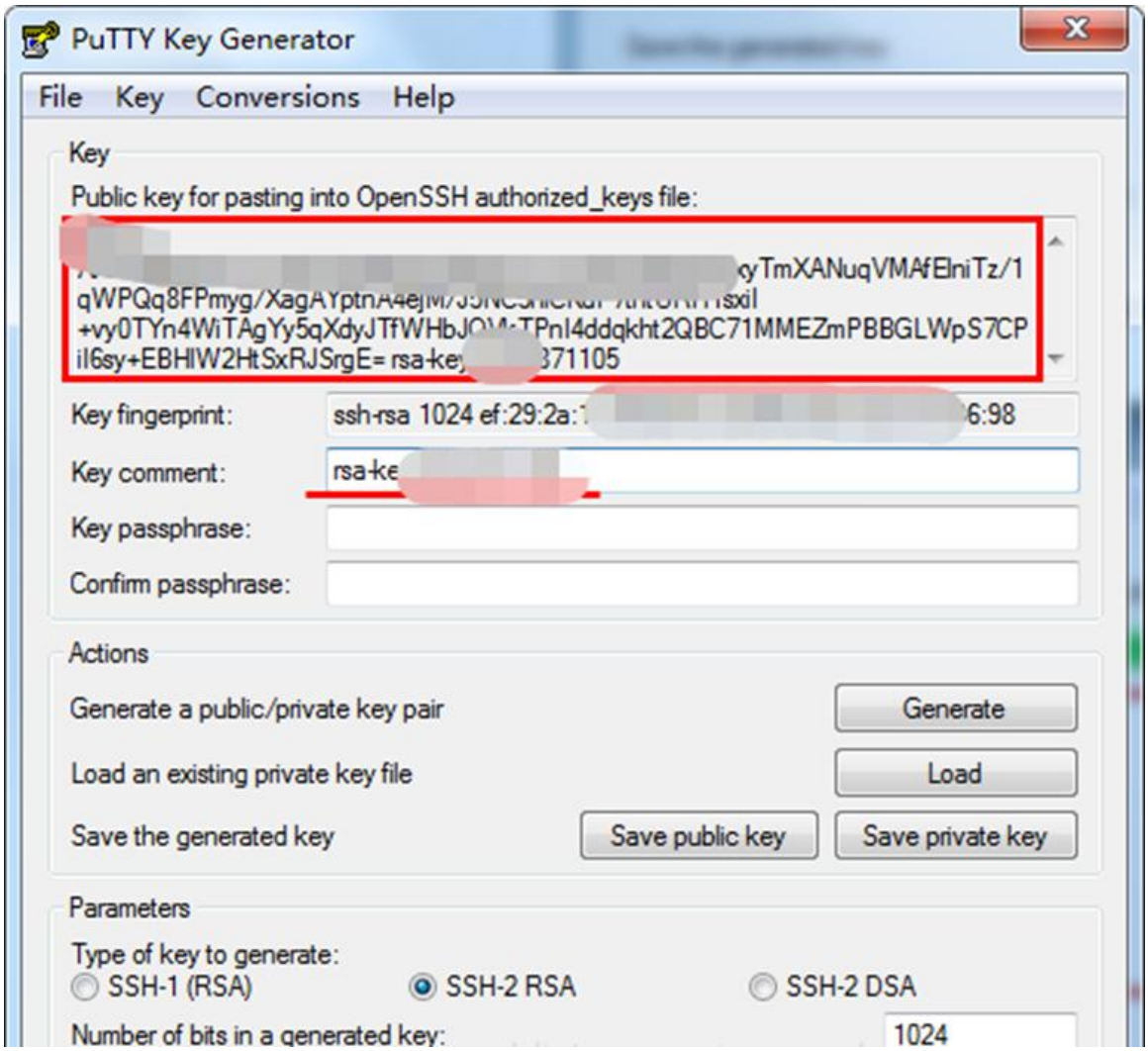
1、点击“puttygen”工具，打开“PuTTY Key Generator”。

图 1 打开 PuTTY Key Generator 页面



2、点击“Load”，并在本地选择对应密钥对的私钥文件。系统会自动加载该私钥文件，并在恢复成格式正确的公钥文件。

图 2 选择私钥文件



3、复制红框中的公钥内容，并将其粘贴在文本文档中，保存为.txt 格式的本地公钥文件。

4、将公钥文件导入天翼云控制台。

8.14 网络配置类

8.14.1 弹性 IP 类

一台弹性云主机是否可以绑定多个弹性公网 IP？

一台弹性云主机可以绑定多个弹性公网 IP，但是不建议您这样操作。

当云主机绑定多个公网 IP 时，需要在云主机内部配置策略路由，才可以确保这些弹性 IP 可以正常进行外部通信。

未绑定弹性公网 IP 的弹性云主机能否访问外网？

能。

您可以使用天翼云的 NAT 网关产品。NAT 网关（NAT Gateway）是一种支持 IP 地址转换的网络云服务，能够为虚拟私有云（Virtual Private Cloud，VPC）内的计算实例提供网络地址转换（Network Address Translation），分为 SNAT 和 DNAT 两个功能。通过 SNAT 可使多个弹性云主机共享使用弹性 IP 访问 Internet。

为什么弹性云主机可以远程连接，但是无法 ping 通？

弹性云主机需要确保安全组的入方向规则添加了“ICMP 协议”。

登录弹性云主机控制台，单击弹性云主机实例名称，进入详情页，单击安全组页签，点击安全组名称，下拉详情，点击添加规则：

方向：入方向

授权策略：允许

协议：ICMP

源地址：IP

地址 0.0.0.0/0

单击确定，完成 ICMP 协议添加，再次尝试是否能够 ping 通。

如何查询云平台上服务器的出口公网 IP 地址？

操作场景

当云主机绑定弹性公网 IP（对于可用区资源池配置了指向 IPv4 网关的路由；对于地域资源池绑定弹性公网 IP 即可），通常通过弹性公网 IP 访问 Internet。

您可以登录天翼云控制台，在弹性云主机列表中查看弹性云主机绑定的弹性公网 IP，具体请参见查看云主机详细信息。

您也可以直接在弹性云主机上通过命令查询 IP 地址，本文以 Ubantu20.04 系统的弹性云服务器为例进行介绍。

操作步骤

1.登录弹性云主机。
2.执行以下任意一条命令，查询弹性云主机绑定的的弹性公网 IP 地址。

```
curl icanhazip.com
curl ifconfig.me
curl ipinfo.io/ip
curl ipecho.net/plain
```

云主机不想使用公网 IP，但是主机是包周期主机，无法解绑 EIP，请问还有什么方法限制主机 IP 在互联网上暴露？

可使用安全组功能限制 IP 的保留，安全组具体介绍可参考安全组概述。

具体操作方法为：

1.在云主机控制台选中该云主机，点击进入详情页。
2.选择安全组页签，在安全组列表中，点击安全组名称左侧的展开符，将安全组详情信息展开。确认安全组-入方向源地址只放行客户本地 IP，即可限制主机 IP 在互联网上暴露。安全组具体配置规则可参考添加安全组操作

如何绑定虚拟 IP，单虚拟 IP 支持绑定几个主机？

虚拟 IP 是虚拟私有云（VPC）产品下的功能，通常用于后端云主机之间部署业务高可用。虚拟 IP 的详细说明请见：虚拟私有云-虚拟 IP 概述

虚拟 IP 仅支持绑定 5 台主机。详细绑定操作与操作限制，请见：虚拟 IP 绑定服务器。

如何修改内网 IP 地址？

在云主机控制台列表里选择要修改内网 IP 的云主机，操作关机。点击“实例/主机名”进入该云主机详情页面，在“网卡”tap 页面选择“修改内网 IP”，在弹出的“修改内网 IP”对话框修改内网 IP 地址。

创建好子网，子网网关设置成 254，但主机 IP 无法设置成 1，提示已占用？

IP 地址 1 已经被其他服务使用，在虚拟私有云中，IP 为 1 的地址被保留用于网络设备（如路由器）或服务（如 DHCP 服务器）使用，因此系统无法将这个 IP 分配给您的云主机。另外还请注意，IP 地址最后一位除去不能为 1 以外，也不可以为 0 或者 255。

更多信息可参考创建子网说明部分。

是否支持弹性 IP 批量从共享带宽中移出？

目前暂不支持此功能的批量操作，仅支持从共享带宽中一次移出一个弹性 IP。

详细操作请见共享带宽帮助手册：共享带宽-从共享带宽中移出弹性 IP 管理添加的弹性 IP。

云主机的弹性 IP ping 不通怎么办？

弹性 IP ping 不通可以用以下步骤进行排查

1.检查安全组规则。检查弹性云主机网卡对应的安全组是否放通了“入方向”的“ICMP”规则。满足下图中任意一条规则即可。



2.检查“ 网络 ACL”规则。查看“网络 ACL”状态，下图分别表示网络 ACL 当前为开启和关闭状态状态。





检查“弹性 IP”绑定的网卡是否在“网络 ACL”关联的子网下，请参考下图进行操作检查。

关联子网：



入方向：



出方向：



若“网络 ACL”为“开启”状态，需要添加 ICMP 放通规则进行流量放通。

注意

需要注意“网络 ACL”的默认规则是丢弃所有出入方向的包，若关闭“网络 ACL”后，其默认规则仍然生效。

3.确认弹性云主机内默认路由是否存在。

- 登录弹性云主机，执行如下命令，查看是否存在默认路由。

```
# ip route
```

```
[root@do-not-del-scy ~]# ip route
default via 192.168.2.1 dev eth0
169.254.0.0/16 dev eth0  scope link  metric 1002
169.254.169.254 via 192.168.2.1 dev eth0  proto static
192.168.2.0/24 dev eth0  proto kernel  scope link  src 192.168.2.112
```

- 若没有该路由，执行如下命令，添加默认路由。

```
# ip route add default via XXXX dev eth0
```

说明

XXXX 表示网关 IP。

4.相同区域主机进行 ping 测试。

在相同区域的弹性云主机去 ping 没有 ping 通的弹性公网 IP，如果可以正常 ping 通说明虚拟网络正常，请联系基础网络技术支持进行处理。

5.使用站长工具进行 ping 测试。

- 登录站长工具网站 http://ping.chinaz.com/。
- 输入弹性 IP，单击“Ping 检测”，以“122.112.197.21”为例。



- 如果大部分监测点能 ping 通，请客户检查本地网络。

监测点	响应IP	IP归属地	响应时间	TTL	赞助商
江苏徐州[电信]	122.112.197.21	上海市 电信	14ms	48	枫信科技徐州独家机房
湖南长沙[电信]	超时(重试)	-	-	-	香港站群服务器
福建福州[电信]	122.112.197.21	上海市 电信	25ms	45	【云彩】100G高防独服450元
福建福州[电信]	122.112.197.21	上海市 电信	21ms	45	【云彩】100G高防独服450元
北京[电信]	122.112.197.21	上海市 电信	31ms	47	特价云服务器15元/月
山东济南[电信]	122.112.197.21	上海市 电信	22ms	47	济南速通-电信大带宽特价
北京[电信]	122.112.197.21	上海市 电信	29ms	48	特价云服务器15元/月

6. 在云监控中查看该弹性 IP 的流量监控。

进入云监控控制台查看该弹性 IP 的流量，排查 IP 的带宽是否打满。

7. 弹性云主机内部排查。

- 确认弹性云主机是否为多网卡，如果为多网卡且弹性 IP 绑定在非主网卡上，需要在弹性云主机内部配置策略路由。
- 登录弹性云主机，查看网卡是否创建，且网卡是否获取私有 IP 地址。如果无法通过用户设置的密码登录弹性云主机，请提交工单联系天翼云的技术支持寻求帮助。

```
[root@ecs-acl ~]# ifconfig
eth0      Link encap:Ethernet  HWaddr FA:16:3E:BC:B7:81
          inet addr:192.168.72.209  Bcast:192.168.72.255  Mask:255.255.255.0
          inet6 addr: fe80::f816:3eff:febc:b781/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:801 errors:0 dropped:0 overruns:0 frame:0
          TX packets:547 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:49604 (48.4 KiB)  TX bytes:44454 (43.4 KiB)
          Interrupt:46
```

- 如果网卡没有正确创建，或者网卡已经创建但没有获取私有 IP 地址，请提交工单联系天翼云的技术支持寻求帮助。
- 执行命令"top"，查看弹性云主机的 CPU 占用率是否冲高。

- 如果弹性云主机的操作系统是 Linux 系统，执行命令"iptables-save"，查看弹性云主机内部是否有安全规则所限制。
- 如果弹性云主机的操作系统是 Windows 系统，请排查弹性云主机的防火墙是否有限制。
- 端口不通排查。
  - 如果弹性云主机的 ping 测试没有问题，但是访问特定端口不通，且端口在图中所示范围内， 由于以下端口被运营商封堵，用户需要更换服务端口。

```
ip access-list extended ACL_FOR_bt_deny+virus_worm

 20 deny tcp any any eq 135
 30 deny udp any any eq 135
 40 deny udp any any eq netbios-ns
 50 deny udp any any eq netbios-dgm
 60 deny tcp any any eq 139
 70 deny tcp any any eq 445
 80 deny udp any any eq 445
 90 deny tcp any any eq 539
 100 deny udp any any eq 539
 110 deny tcp any any eq 593
 120 deny udp any any eq 593
 130 deny udp any any eq 1434
 140 deny tcp any any eq 4444
 150 deny tcp any any eq 5554
 160 deny tcp any any eq 9996
 170 deny udp any any eq 6667
 180 deny tcp any any eq 5444
 190 deny udp any any eq 5444
 200 deny tcp any any eq 5800
 210 deny tcp any any eq 6667
 220 deny tcp any any eq 5900
 230 deny udp any any eq 9996
```

  - 如果不能访问弹性云主机的特定端口，需要排查安全组规则以及“网络 ACL”是否对服务端口放行。
  - 登录弹性云主机，执行以下命令，查看弹性云主机内部是否在该端口进行监听。

```
# netstat - na | grep XXX
```

说明

XXX 表示端口号。

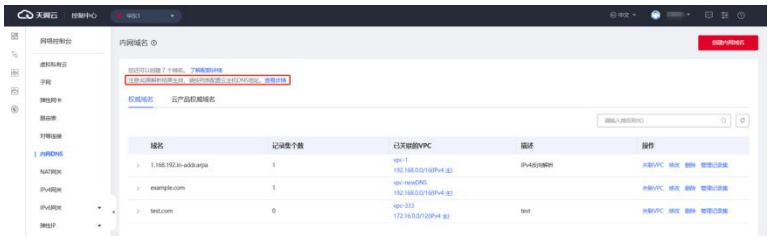
8.14.2DNS 与 NTP 配置类

8.14.2.1 配置 DNS

为实现内网域名在 VPC 上的正常解析，用户需要将云主机内的 DNS 改成内网 DNS 服务地址。

查看内网 DNS 地址

1. 在内网 DNS 的列表中，可以看到如下红框内的提示。



2. 点击“查看列表”，可以看到内网 DNS 地址。



修改 Linux 云主机内 DNS 地址

1. 登录 Linux 弹性云主机。
2. 执行以下命令，打开 resolv.conf 文件。

```
vi /etc/resolv.conf
```



3. 按“i”键进入编辑模式。



4. 将“nameserver”修改为当前云主机所在 VPC 的内网 DNS IP 地址,按“ESC”后输入“：wq”保存配置。



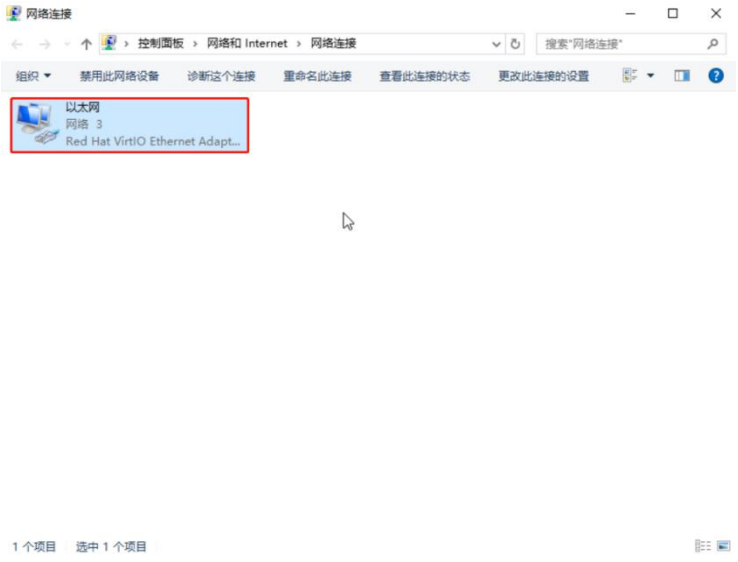
修改 Windows 云主机内 DNS 地址

以 Windows server 2016 操作系统为例，以用户名 Administrator，登录 Windows 弹性云主机。

1. 在任务栏的右下角，右击网络连接的图标，单击“打开网络和共享中心”。
2. 在左侧导航栏，单击“更改适配器设置”。



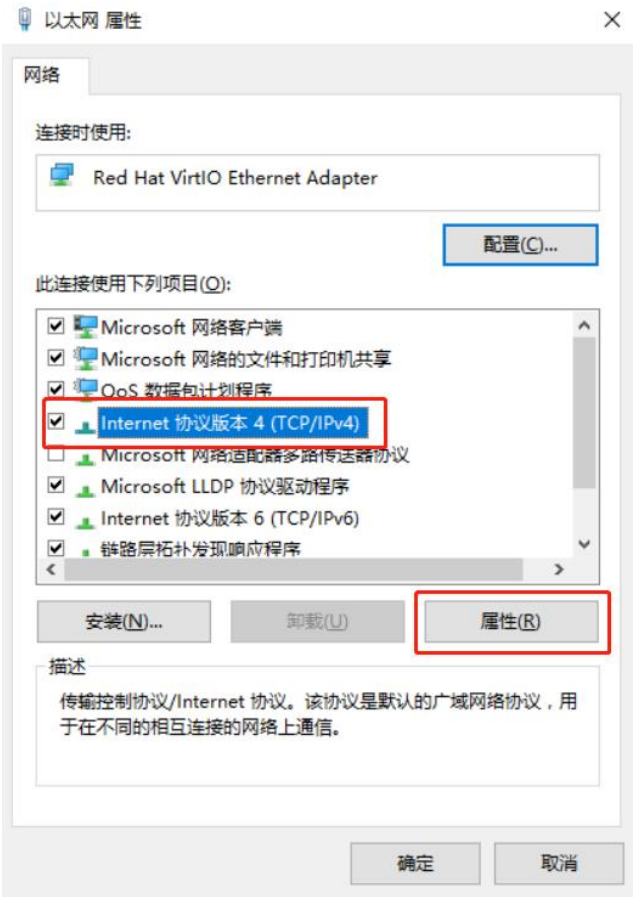
3. 双击“本地连接”。



4. 单击左下角的“属性”。



5. 选择“Internet 协议版本（TCP/IP）”，并单击“属性”。



6. 选择“使用下面的 DNS 服务器地址”，并根据界面提示填写当前云主机所在 VPC 的内网 DNS IP 地址。



8.14.2.2 怎样配置弹性云主机的 DNS 和 NTP 信息？

Linux 操作系统

操作系统均为 CentOs 7.6 为例：

步骤一：给弹性云主机配置 NTP 服务器

1. 登录 Linux 弹性云主机。执行以下命令，编辑 ntp.conf 文件。

```
vim /etc/ntp.conf
```

2. 添加以下语句，配置 NTP 服务器。

```
server NTP IP 地址
```

3. 执行以下命令，系统重新启动时启动服务。

```
service ntp restart
```

4. 执行以下命令，检查 NTP 服务器的状态。

```
service ntp status
```

步骤二：给弹性云主机配置 DNS 服务器。

1. 登录 Linux 弹性云主机。执行以下命令，编辑 resolv.conf 文件。

```
vi /etc/resolv.conf
```

2. 添加如下语句，配置 DNS 服务器。

```
nameserver = DNS 服务器的 IP 地址
```

3. 执行以下命令，重启网络。

```
rcnetwork restart
```

```
service network restart
```

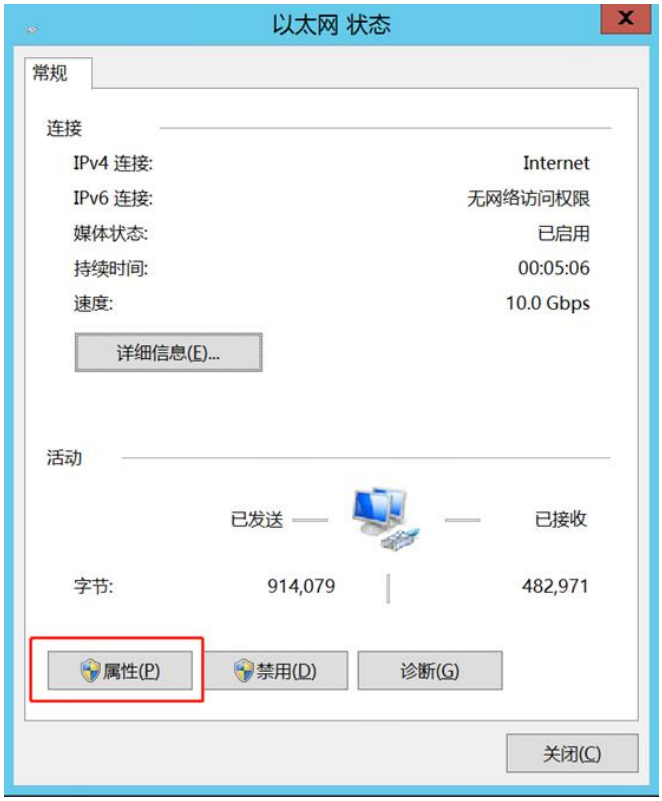
```
/etc/init.d/network restart
```

Windows 操作系统

以 Windows 2012 操作系统为例：

步骤一：配置 DNS 信息

1. 以用户名 Administrator，登录 Windows 弹性云主机。
2. 在任务栏的右下角，右键单击网络连接的图标。单击“打开网络和共享中心”。
3. 在左侧导航栏，单击“更改适配器设置”。
4. 给弹性云主机配置 DNS 服务器。双击网络连接。



5. 选择“Internet 协议版本 4（TCP/IPv4）”，并单击“属性”。



6. 选择“使用下面的 DNS 服务器地址”，并根据界面提示填写 DNS 服务器的 IP 地址，如图 4 所示。



步骤二：给弹性云主机配置 NTP 服务器。

1. 打开“运行”窗口。输入“regedit”，并单击“确定”，打开注册表编辑器。
2. 修改如下注册表信息：

- 打开文件“\W32Time \ TimeProviders \ NtpClient”，修改“Enabled”的“数值数据”为“1”，表示该设备为 NTP 客户端。



- 打开文件“\ W32Time \ TimeProviders \ NtpServer”，修改“Enabled”的“数值数据”为“0”，关闭 NTP 服务器功能。
- 打开文件“\W32Time \Parameters”，设置“NtpServer”的“数值数据”。例如设置为“ntp.myctyun.com”。设置“TYPE”的“数值数据”为“NTP”。





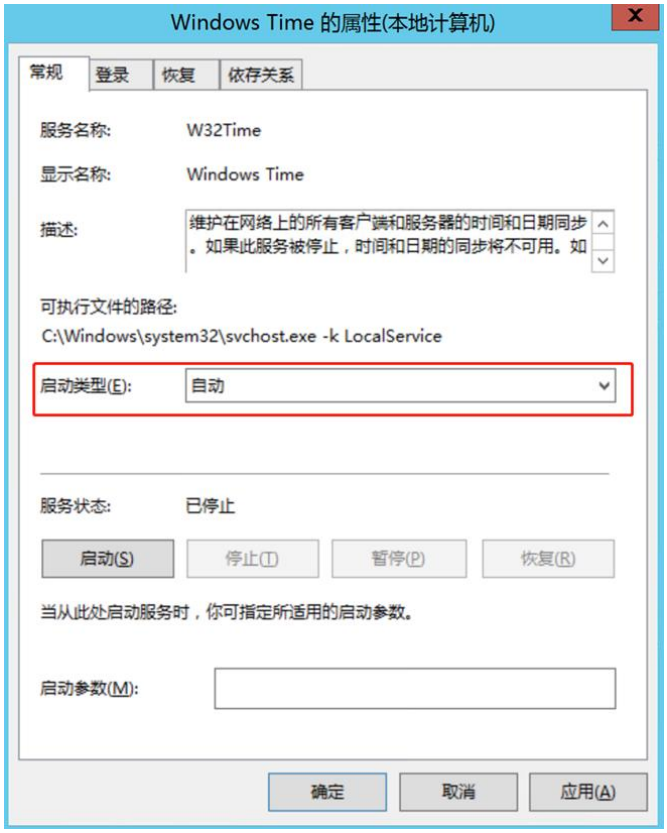
- 打开文件“\ W32Time \ TimeProviders \ NtpClient”，设置“SpecialPollInterval”的“数值数据”为“60”，“基数”为“十进制”，表示时钟同步周期为60s。



- 打开文件“\ W32Time \ config”，设置“MaxPosPhaseCorrection”和“MaxNegPhaseCorrection”的“数值数据”为“ffffffff”，“基数”为“十六进制”。

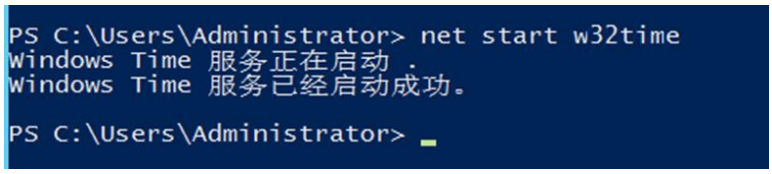


3. 打开“运行”窗口，输入“services.msc”，并单击“确定”，打开“服务”窗口。
4. 查找名称为“Windows Time”的服务，右键选择属性并设置“启动类型”为“自动”，从NTP服务中同步时间。



5. 打开“运行”窗口，依次执行以下命令，重启 Windows Time 服务。

```
net stop w32time
net start w32time
```



6. 手动修改客户端时间，使得客户端时间与服务器时间一致。一分钟后，检查客户端时间与服务器时间是否一致，如果一致，则表示时间同步配置成功。

8.14.3 网卡类

8.14.3.1 弹性云主机添加网卡后是否会自动启动？

操作场景

目前给云主机添加网卡后，默认是会自动启动。

如果用户发现网卡没有自启动，可以手动启动网卡或者咨询客服。

操作步骤

以 Centos7 为例，手动启动网卡的命令如下：

```
# ifup ens33
```

ens33 为需要启动的网卡名字，根据实际情况进行替换。

或者用户也可以用以下命令重启所有网卡：

```
# systemctl restart network
```

8.14.3.2 怎样修改云主机的 VPC 网段、子网的网段？

操作场景

VPC、子网创建好后，网段是不支持修改的。

如需修改 VPC 网段可以通过切换 VPC 来实现，如需修改子网网段可以通过更换子网来实现。

VPC 网段、子网的网段一旦创建，不能进行修改。

操作步骤

您可以通过更换子网或者更换 VPC 来实现，具体参考如何修改内网 IP、切换 VPC。

8.14.4 路由类

CentOS7.6 系统如何添加静态路由？

操作场景

为系统添加静态路由，避免重启系统后路由丢失而影响到网络连通性。

操作步骤

以 CentOS 7.6 操作系统为例。

1.登录弹性云服务器。

2.创建或修改静态路由配置文件。

在“/etc/sysconfig/”目录下创建或修改静态路由配置文件 static-routes 文件，当系统没有此文件时，可以新建此文件。向文件中添加一行静态路由记录，例如：any net 192.168.3.0 netmask 255.255.255.0 gw 192.168.2.3。

3.重启网络服务，使静态路由生效。

```
# service network restart
```

4.执行以下命令，查看路由信息。

```
# route -n
```

8.14.5 访问网站或应用故障类

Windows 云主机无法访问外网怎么办？

当 Windows 云主机状态正常，但是无法访问外网时，我们可以通过以下步骤进行排查：

1.查看云主机是否负载过高

可以通过云监控查看云主机负载情况，您可以创建告警任务，当 CPU 或带宽利用率高时，系统会自动发送告警给您。

如果确定云主机负载过高，可以使用进程管理查找负载过大的进程，Kill 掉过高负载的进程，再尝试进行访问外网。

2.查看云主机的安全组设置

检查弹性云主机安全组是否有配置错误，安全组出方向规则为允许，放通出方向网络流量。

3.查看 VPC 的网络 ACL 设置

查看云主机对应的子网是否关联了网络 ACL。单击子网名称查看网络 ACL 的信息。点击 ACL 名称，进入修改 ACL 规则，或者将子网内的 ACL 删除。

4.检查云主机的防火墙设置

查看并禁用 Windows 弹性云主机防火墙策略，禁用后检查是否可以连通网络。

登录 Windows 弹性云主机。单击桌面左下角的开始图标，选择“控制面板 > 系统和安全 > Windows 防火墙”。 单击“检查防火墙状态 > 关闭 Windows 防火墙”。

再次尝试访问公网。

5.使用 Ping 命令检查网关通信是否正常

使用 ping 命令检查云主机与网关通信是否正常。如果可以 ping 通就是其他原因，无法 ping 通需查看 DNS 设置情况。

如果上述排查都没能成功访问公网，请提交工单，联系技术支持人员帮助解决。

Linux 云主机无法访问外网怎么办？

当 Linux 云主机状态正常，但是无法访问外网时，我们可以通过以下步骤进行排查：

1.查看云主机是否负载过高

可以通过云监控查看云主机负载情况，您可以创建告警任务，当 CPU 或带宽利用率高时，系统会自动发送告警给您。

如果确定云主机负载过高，可以使用进程管理查找负载过大的进程，Kill 掉过高负载的进程，再尝试进行访问外网。

2.查看云主机的安全组设置

检查弹性云主机安全组是否有配置错误，安全组出方向规则为允许，放通出方向网络流量。

3.查看 VPC 的网络 ACL 设置

查看云主机对应的子网是否关联了网络 ACL。单击子网名称查看网络 ACL 的信息。点击 ACL 名称，进入修改 ACL 规则，或者将子网内的 ACL 删除。

4.检查云主机的防火墙设置

查看 Linux 弹性云主机防火墙状态，禁用后检查是否可以连通网络。

使用以下命令查看防火墙状态。

```
firewall-cmd  --state
```

如果回显是 running，即防火墙开启，使用以下命令关闭防火墙

```
systemctl stop firewalld.service
```

再次尝试访问公网。

5.使用 Ping 命令检查网关通信是否正常

使用 ping 命令检查云主机与网关通信是否正常。如果可以 ping 通就是其他原因，无法 ping 通需查看 DNS 设置情况。

如果上述排查都没能成功访问公网，请提交工单，联系技术支持人员帮助解决。

网站无法访问怎么办？

无法访问弹性云主机实例中运行的网站的原因较多，此处列举较为常见的问题原因，具体原因以现场的排查结果为准。

- TCP 80 端口不可用。
- Web 服务不可用。
- 网站未备案。

- 网站资源或后端服务存在异常。

以 CentOS 7.6 系统为例，为大家介绍排查思路：

1.检查 80 端口配置

```
netstat -ntulp |grep 80
```

如果端口被正常监听，请执行检查安全组规则。

如果端口没有被正常监听，请检查 Web 服务进程是否启动或者正常配置。

2.检查安全组配置

如果安全组没有网站访问使用的端口，需要在云主机实例对应的安全组中添加放行该端口的规则。

a. 在弹性云主机控制，单击主机名称。

b. 选择“安全组”页签，展开安全组规则。

c. 单击“更改安全组规则”。

d. 根据网站使用的端口配置新的安全组规则，放行网站使用的端口。

3.检查防火墙配置

a. 使用 iptables -nvL --line-number 命令查看已配置的防火墙策略。

b. 依次执行以下命令放行 80 端口。

c. iptables -A INPUT -p tcp --dport 80 -j ACCEPT

d. iptables -A OUTPUT -p tcp --sport 80 -j ACCEPT

e. 使用 service iptables save 命令保存添加的规则。

f. 使用 service iptables restart 命令重启 iptables。

g. 使用 iptables -nvL --line-number 命令查看增加的规则是否生效。

h. 关闭防火墙后，重新测试网站访问是否正常。

4.检查云主机 CPU 利用率

可以通过云监控查看云主机负载情况，您可以创建告警任务，当 CPU 或带宽利用率高时，系统会自动发送告警给您。使用 Top 命令查看当前系统的运行状态。排查进程是否正常，结束掉异常进程，再次尝试访问。

5.检查网站备案及域名解析是否正常

无论网站是通过 IP 地址还是通过域名对外提供服务，都需要进行备案。请检查您的备案信息是否正常。

如果域名已备案，但未正确配置域名解析也可能会导致域名无法 Ping 通。您可以在域名提供商的控制台查看解析详情。

如果上述排查都没能成功访问网站，请提交工单，联系技术支持人员帮助解决。

弹性云主机访问中国大陆外网站时加载缓慢怎么办？

一、为什么访问中国大陆外网站卡顿？

1.购买的弹性云主机可能会出现访问中国大陆外（包括中国港澳台及其他国家、地区）网站卡顿的问题。
2.这是由于云主机在中国大陆内，访问的网站在中国大陆外，国际带宽线路则是首选，国际带宽线路节点分布世界各地，访问过程中不可避免的出现绕节的情况，延时增加，访问卡顿，因此延迟高于国内云主机访问国内网站的情况。

二、处理方法

1.重新购买中国大陆外的弹性云主机，例如“香港 1”区域的云主机。
从物理距离与网络基础设施等因素考虑方面，如果您有访问中国大陆外网站的需求，我们建议您购买中国大陆外的弹性云主机。例如您可以在购买弹性云主机时选择“香港 1”区域的弹性云主机。



2. 修改 DNS 配置。

修改 DNS 服务器地址为公共 DNS 服务器，例如：101.226.4.6 、1.1.1.1。

(1) 以用户名 Administrator，登录 Windows 弹性云主机。

(2) 打开本地连接,在任务栏的右下角，右键单击网络连接的图标。

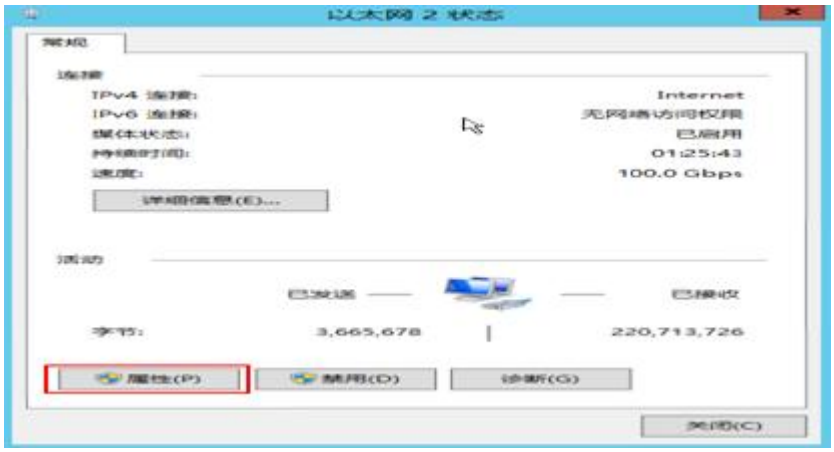
(3) 单击“打开网络和共享中心”,打开网络共享中心。



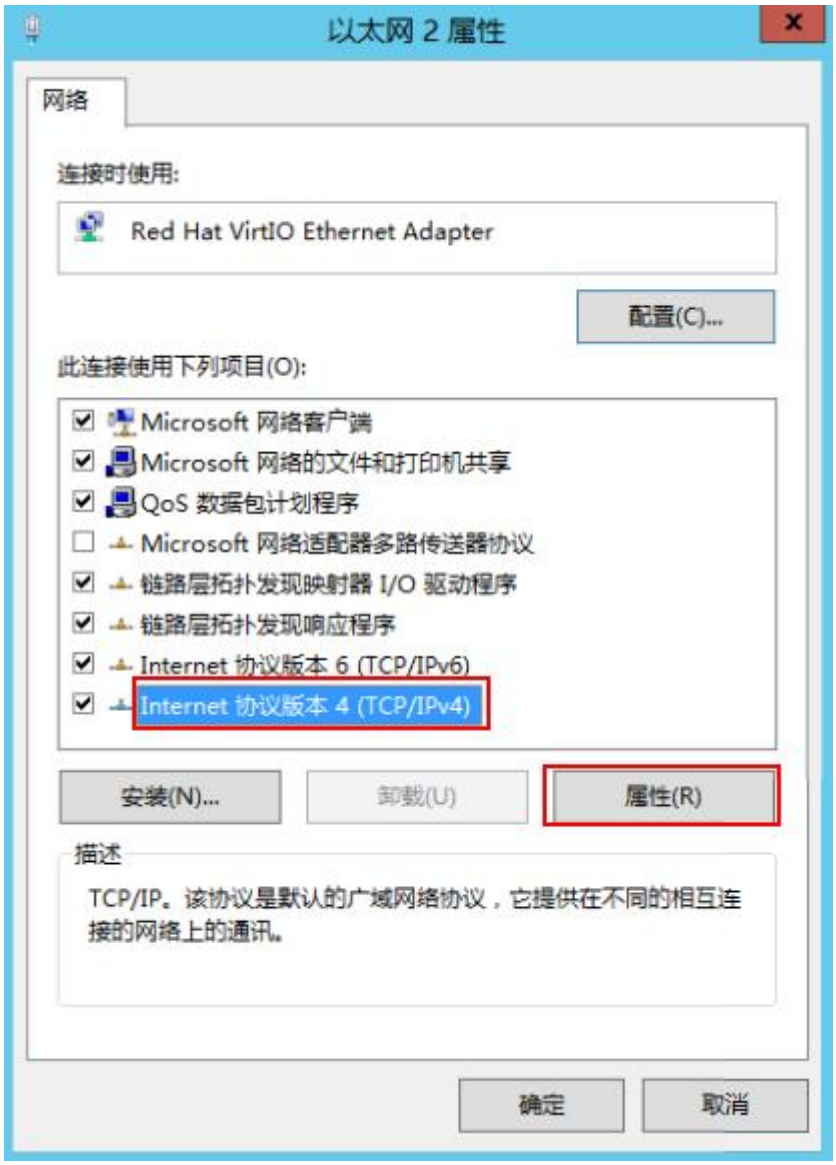
(4) 在左侧导航栏，单击“更改适配器设置”。

(5) 给弹性云主机配置 DNS 服务器，双击网络连接。

(6) 单击左下角的“属性”。



(7) 选择“Internet 协议版本 4（TCP/IPv4）”，并单击“属性”。



(8) 选择“使用下面的 DNS 服务器地址”，并根据界面提示填写 DNS 服务器的 IP 地址。

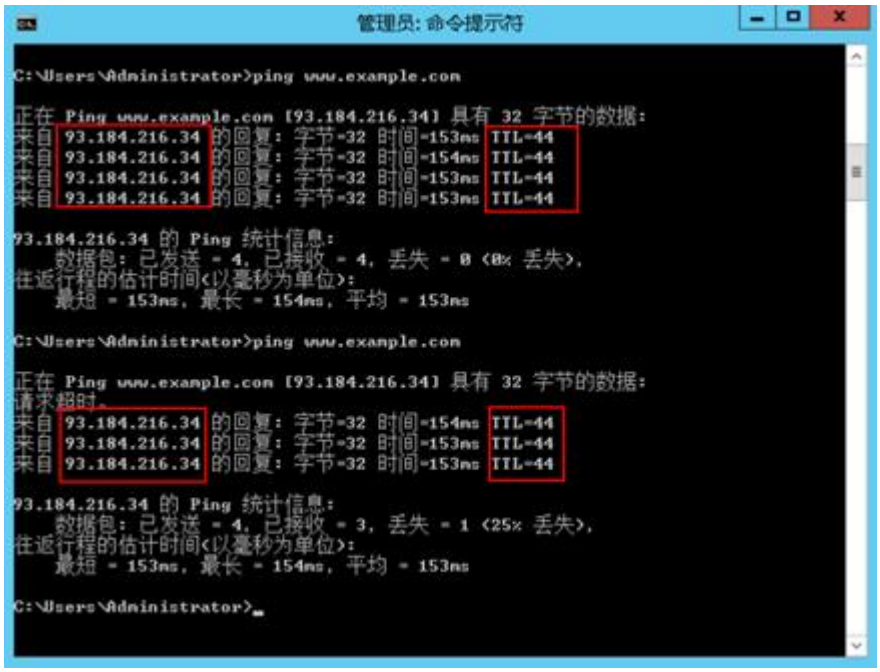


3. 修改 hosts 文件来优化访问速度。

选择访问速度最快的服务器，并将其 IP 地址和域名写入 hosts 文件来优化访问速度。

方法一：使用 ping 命令判断访问速度最快的服务器 IP 地址。

(1) 以访问 www.example.com 为例，在命令行 ping www.example.com，查询 ping 结果。



(2) 重复执行多次 ping www.example.com，记录下一条 TTL 值最小、且稳定的 IP 地址。

(3) 修改 hosts 文件。

打开 C:\Windows\System32\drivers\etc\，将之前复制的 IP 地址以如下方式写入 hosts 文件末行。例如复制的 IP 地址为 93.184.216.34，则将 93.184.216.34 www.example.com 写入到 hosts 文件的末行，保存后关闭。

(4)重新访问中国大陆外网站，则卡顿或无法访问的问题会有所好转。

修改 hosts 文件只能优化访问速度，如果重试后问题仍未解决，我们建议您购买“香港 1”区域的云主机。

8.14.6iptables 如何配置、建议配置规则指导

CentOS7 系统默认的防火墙是 Firewalld。但是，仍有大量用户习惯于在 CentOS7 系统中使用 iptables。本文以 CentOS7.9 为例，说明在 CentOS7 系统中如何安装并使用 iptables。

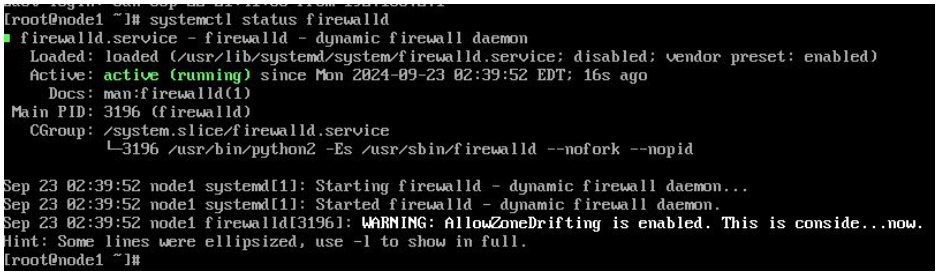
禁止 firewalld 开机启动

为了防止与 iptables 产生冲突，您需要先禁用 Firewalld 开机自启

1. 连接 Centos7.9 实例，关于如何连接实例可参考登录 Linux 弹性云主机

2. 执行如下命令查看服务状态

```
systemctl status firewalld
```



3. 当服务处于 active 状态，运行以下命令关闭 Firewalld 服务。

```
systemctl disable firewalld --now
```

安装 iptables

执行如下命令，安装 iptables：

```
yum install -y iptables-services
```

启动 iptables 并开启开机自启：

```
systemctl enable iptables.service --now
```

验证 iptables 是否启动成功

```
systemctl status iptables.service
```

查看并修改 iptables 默认规则

执行 iptables -L 命令，查看 iptables 默认规则，发现在默认规则下，INTPUT 链允许来自任何主机的访问，可以参考如下步骤修改默认规则。

1. 如果之前已经设置过规则，建议执行如下命令，备份原有的 iptables 文件，避免之前设置的规则丢失。

```
cp -a /etc/sysconfig/iptables /etc/sysconfig/iptables.bak
```

2. 执行如下命令，清空所有规则。

```
iptables -F
```

3. 根据业务需求添加规则，放行或者禁用端口。示例：依次执行如下命令，放行 80 端口和 22 端口。

```
iptables -I INPUT -p tcp --dport 80 -m state --state NEW -j ACCEPT
iptables -I INPUT -p tcp --dport 22 -m state --state NEW -j ACCEPT
```

示例：依次执行如下命令，添加规则，使 INPUT 链拒绝所有请求，即 ECS 实例会拒绝所有请求。如果是线上业务请勿直接操作，会直接中断业务。

```
iptables -P INPUT DROP
```

4. 执行如下命令，确认新规则生效。

```
iptables -L
```

5. 执行如下命令，保存添加的规则。

```
iptables-save > /etc/sysconfig/iptables
```

6. 批量导入规则

```
iptables-restore < 文件名称
```

8.14.7 其它类

8.14.7.1 云主机 mac 是否为固定的？

云主机网卡分为主网卡和辅助网卡。云主机所挂载的主网卡无法更换，因此 mac 地址固定。云主机可以添加或更换辅助网卡，不同的辅助网卡对应的 mac 地址不同，挂载不同的辅助网卡会获得不同的 mac 地址。

8.14.7.2DHCP 无法正常获取内网 IP？

Linux 云主机（CentOS）

1.登录 Linux 云主机，执行如下命令，查看是否存在 dhclient 进程。

```
ps -ef | grep dhclient
```

2.若 dhclient 进程不存在，可以重启网卡或执行以下命令主动发起 DHCP 请求。

```
dhclient eth0 或 ifdown eth0 + ifup eth0 或 dhcpcd eth0
```

3.对于 DHCP Client 长期不发起请求的情况（例如重启网卡后又复现），可以参照以下操作：执行如下命令，配置静态 IP。

```
vi
/etc/sysconfig/network-scripts/ifcfg-eth0BOOTPROTO=staticIPADDR=192.168.1.200 #IP 地址 NETMASK=255.255.255.1 #掩码值 GATEWAY=192.168.1.0 #网关地址
```

4.重启云主机使网络配置生效。选取有 DHCP 持续获取能力的镜像。

5.若依然无法解决，请联系技术支持。

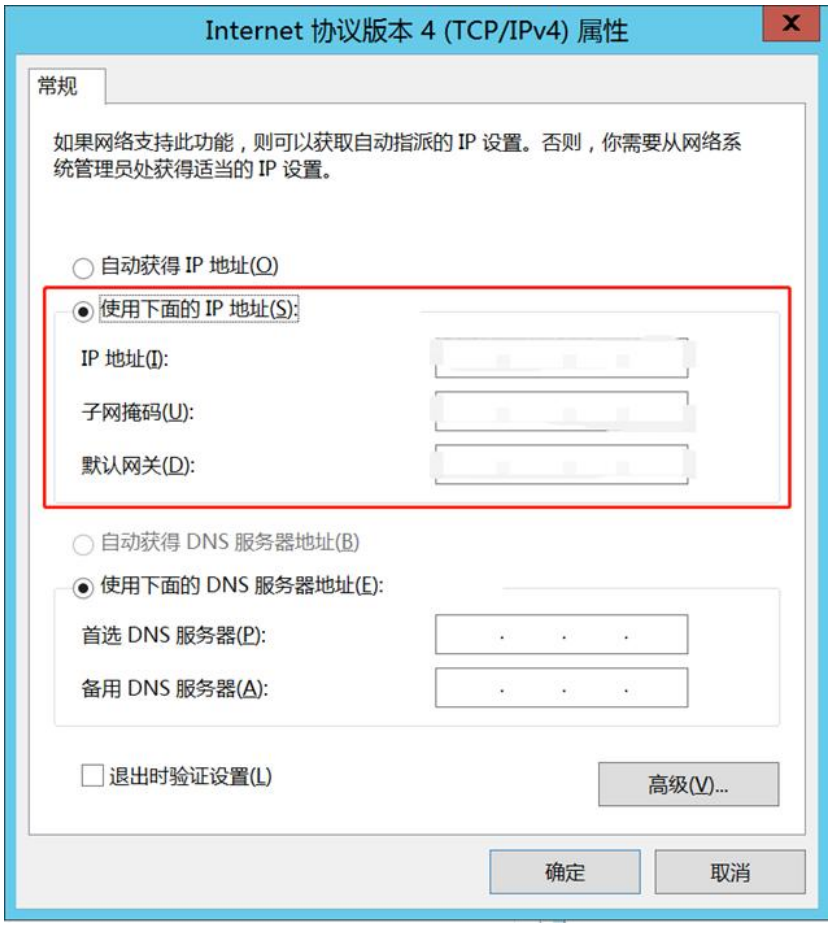
Windows 云主机（2012）

1.在网络连接中选择其中一个网络设备，单击鼠标右键，选择“禁用”，再单击“启用”。



2. 对于 DHCP Client 长期不发起请求的情况（例如重启网卡后又复现），可以参照以下操作：在网络连接中选择“属性”。

3. 在弹出的窗口中选择“Internet 协议版本 4”，单击“属性”，修改相关配置。



4. 重启云主机使网络配置生效。

5. 若依然无法解决，请联系技术支持。

8.14.7.3 如何查看、修改 Linux 弹性云主机的内核参数？

如果您的业务未受到影响，建议不要执行修改内核参数的操作。

查看内核参数

通过“/etc/sysctl.conf”文件进行查看。执行以下命令，查看当前系统中生效的所有参数。

```
/usr/sbin/sysctl -a
```

修改内核参数

通过“/etc/sysctl.conf”文件进行修改。该方法修改的参数值，永久生效。

1.执行以下命令，根据需求修改指定的参数值。示例：

```
/sbin/sysctl -w kernel.domainname="myctyun.com"

sysctl -w net.ipv4.tcp_tw_recycle="0"
```

2.执行以下命令，修改“/etc/sysctl.conf”文件中的参数。

```
vi /etc/sysctl.conf
```

3.执行以下命令，使配置生效。

```
/sbin/sysctl -p
```

8.14.7.4 端口映射配置

做端口映射时，请确保安全组、防火墙已放通对应的端口。

Windows 云主机（2012）

打开 cmd 窗口执行命令。

netsh interface portproxy add v4tov4 listenaddress=本机内网 IP

listenport=源端口

connectaddress=映射主机内网 IP

connectport=映射端口

Linux 云主机（CentOS 7.6）

以登录为例，登录“弹性云主机 1”的 1080 端口自动跳转访问“弹性云主机 2”的 22 端口。

1.登录 Linux 弹性云主机。执行如下命令，修改文件。

```
vi /etc/sysctl.conf
```

在文件中添加“net.ipv4.ip_forward = 1”。

执行如下命令，完成修改。

```
sysctl -p /etc/sysctl.conf
```

```
[root@ ]# vi /etc/sysctl.conf
[root@ -]# sysctl -p /etc/sysctl.conf
net.ipv6.conf.all.accept_ra = 1
net.ipv6.conf.default.accept_ra = 1
net.ipv6.conf.eth0.accept_ra = 1
net.ipv4.ip_forward = 1
```

2.在“iptables”的“nat”表中添加规则，执行如下命令，通过弹性云主机 1 的 1080 端口映射到弹性云主机 2 的 22 端口。

iptables -t nat -A PREROUTING -d 弹性云主机 1 私网 IP

-p tcp --dport 1080 -j DNAT --to-destination 弹性云主机 2 私网 IP:22

iptables -t nat -A POSTROUTING -d 弹性云主机 2 私网 IP

-p tcp --dport 22 -j SNAT --to 弹性云主机 1 私网 IP

```
[root@ ]# iptables -t nat -A PREROUTING -d  -p tcp --dport 1080 -j DNAT --to-destination :
22
[root@ -]# iptables -t nat -A POSTROUTING -d  -p tcp --dport 22 -j SNAT --to
```

3.验证配置是否生效，执行如下命令，登录弹性云主机 1 的 1080 端口。

ssh -p 1080 弹性云主机 1 公网 IP

4.输入 yes 及弹性云主机 2 的密码后，登录到弹性云主机 2。

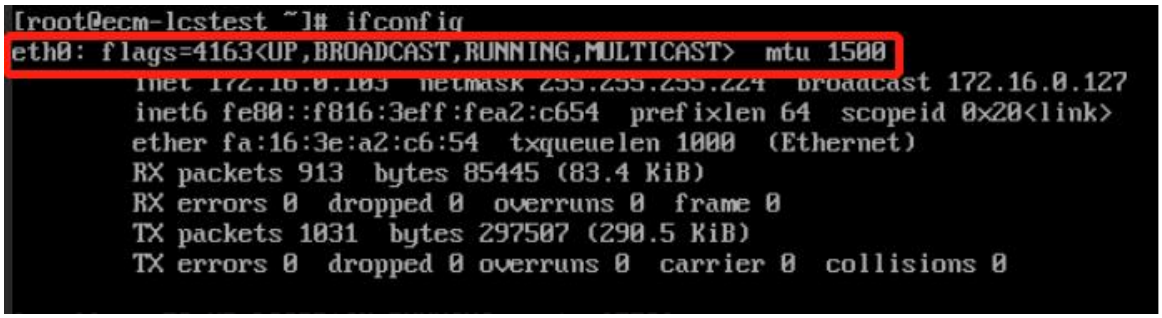
```
[root@ -]# ssh -p 1080
The authenticity of host '[ ]:1080 ([ ]:1080)' can't be established.
ECDSA key fingerprint is SHA256:cZEgJiEaBGsa2wQtREmLgGZW3dU3tM6DDQQWaWnSG+U.
ECDSA key fingerprint is MD5:8e:4a:d8:33:84:7d:03:36:f5:3b:45:94:2c:24:a2:f6.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '[ ]:1080' (ECDSA) to the list of known hosts.
root@ 's password:
[root@ ]#
```

8.14.7.5 如何查看云主机的 mac 地址？

Linux 操作系统（CentOS 7.6）

登录 Linux 云主机，使用如下命令查看网卡信息。

```
Ifconfig
```



使用如下命令查看 mac 地址。

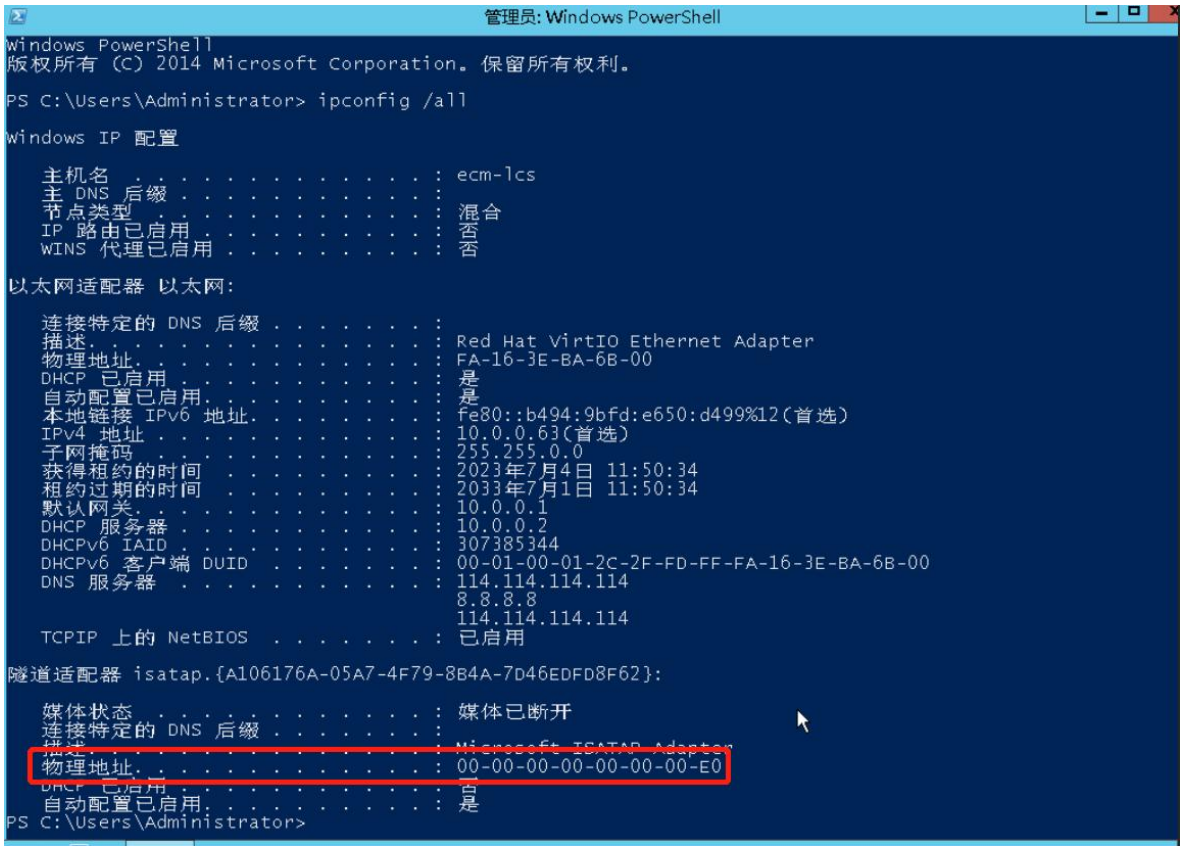
```
ifconfig eth0 |egrep "ether"
```



Windows 操作系统（2012）

打开任务栏的 cmd 窗口给，执行以下命令，查看云主机的 mac 地址。

```
ipconfig /all
```



8.14.7.6 网络性能测试方法

基础信息

- 被测机：被压力测试网络性能的弹性云服务器，可作为 iperf3 测试中的 client 端（发送端）。
- 辅助云服务器：作为 iperf3 测试中的 server 端（接收端）。
- 测试工具 iperf3 常用参数说明如表 1 所示。

表 1 iperf3 工具 client 端常用参数说明

client 端参数	参数说明
-p	端口号，iperf3 的默认端口号是 5201

client 端参数	参数说明
-c	接收端（server 端）IP 地址
-u	如果发送 UDP 报文，则使用该参数
-b	发送带宽
-t	测试时长
-l	数据包大小
-R	以反向模式运行。（服务器发送数据，客户端接收数据）

表 2 iperf3 工具 server 端常用参数说明

server 端参数	参数说明
-s	以 Server 模式运行
-D	在后台以守护进程运行

测试准备

1.准备弹性云服务器。

要求：创建 5 台弹性云服务器作为被测机，创建 5 台弹性云服务器作为辅助云服务器，保证被测机和辅助云服务器之间能够正常通信。

表 3 测试准备

类型	数量	镜像	规格
被测机	5 台	CentOS 7.6 64bit（推荐）	4C8G（推荐）
辅助云服务器	5 台	CentOS 7.6 64bit（推荐）	4C8G（推荐）

2.准备测试工具。

要求：需分别在被测机和辅助云服务器上安装测试工具 iperf3，将被测机和辅助云服务器分别绑定弹性 EIP，使用如下安装命令安装测试工具：

```
yum -y install iperf3
```

3.在辅助云服务器上开启 iperf3 服务

执行以下命令，在辅助云服务器上开启 iperf3 服务，开启的端口号为 xxx。

```
iperf3 -s -D -p xxx
```

TCP 带宽测试

1.测试 TCP 发送带宽。

a. 分别在“所有辅助云服务器”中执行以下命令，启动 iperf3 server 进程。

```
iperf3 -s -D -p xxxx
```

其中，-p 用于指定监听端口。

b. 在“被测机”中执行以下命令，启动 iperf3 打流进程，分别指定到不同辅助云服务器的监听端口。其中，iperf3 工具的常用参数说明请参见表 1 和表 2。

IP 地址对应第 1 台辅助云服务器

iperf3 -c 辅助云服务器 1 的 ip 地址 -p xxx -t 60

IP 地址对应第 2 台辅助云服务器

iperf3 -c 辅助云服务器 2 的 ip 地址 -p xxx -t 60

IP 地址对应第 3 台辅助云服务器

iperf3 -c 辅助云服务器 3 的 ip 地址 -p xxx -t 60

IP 地址对应第 4 台辅助云服务器

iperf3 -c 辅助云服务器 4 的 ip 地址 -p xxx -t 60

IP 地址对应第 5 台辅助云服务器

iperf3 -c 辅助云服务器 5 的 ip 地址 -p xxx -t 60

2.测试 TCP 接收带宽。

a. 分别在“所有辅助云服务器”中执行以下命令，启动 iperf3 server 进程。

```
iperf3 -s -D -p xxxx
```

其中，-p 用于指定监听端口。

b. 在“被测机”中执行以下命令，启动 iperf3 打流进程，分别指定到不同辅助云服务器的监听端口。其中，iperf3 工具的常用参数说明请参见表 1 和表 2。

IP 地址对应第 1 台辅助云服务器

iperf3 -c 辅助云服务器 1 的 ip 地址 -p xxx -t 60 -R

IP 地址对应第 2 台辅助云服务器

iperf3 -c 辅助云服务器 2 的 ip 地址 -p xxx -t 60 -R

IP 地址对应第 3 台辅助云服务器

iperf3 -c 辅助云服务器 3 的 ip 地址 -p xxx -t 60 -R

IP 地址对应第 4 台辅助云服务器

iperf3 -c 辅助云服务器 4 的 ip 地址 -p xxx -t 60 -R

IP 地址对应第 5 台辅助云服务器

iperf3 -c 辅助云服务器 5 的 ip 地址 -p xxx -t 60 -R

3.解析 TCP 测试结果。

测试结束后，发送端 iperf3 进程输出结果如图所示：



最终结果为所有 iperf3 进程测试结果之和。

UDP PPS 测试

1.测试 UDP 发送 PPS。

a. 登录辅助云服务器。

b. 分别在所有“辅助云服务器”中执行以下命令，启动 server 进程。

iperf3 -s -p xxxx -D

其中，-p 用于指定监听端口。

c. 在“被测机”中执行如下命令，启动 client 进程。其中，iperf3 工具的常用参数说明请参见表 1 和表 2。

IP 地址对应第 1 台辅助云服务器

iperf3 -c 辅助云服务器 1 的 ip 地址 -p xxx -t 60 -u -b 30M

IP 地址对应第 2 台辅助云服务器

iperf3 -c 辅助云服务器 2 的 ip 地址 -p xxx -t 60 -u -b 30M

IP 地址对应第 3 台辅助云服务器

iperf3 -c 辅助云服务器 3 的 ip 地址 -p xxx -t 60 -u -b 30M

IP 地址对应第 4 台辅助云服务器

iperf3 -c 辅助云服务器 4 的 ip 地址 -p xxx -t 60 -u -b 30M

IP 地址对应第 5 台辅助云服务器

iperf3 -c 辅助云服务器 5 的 ip 地址 -p xxx -t 60 -u -b 30M

2.测试 UDP 接收 PPS。

a. 执行以下命令，在“被测机”中启动 server 进程。其中，iperf3 工具的常用参数说明请参见表 1 和表 2。

IP 地址对应第 1 台辅助云服务器

iperf3 -c 辅助云服务器 1 的 ip 地址 -p xxx -t 60 -u -b 30M -R

IP 地址对应第 2 台辅助云服务器

iperf3 -c 辅助云服务器 2 的 ip 地址 -p xxx -t 60 -u -b 30M -R

IP 地址对应第 3 台辅助云服务器

iperf3 -c 辅助云服务器 3 的 ip 地址 -p xxx -t 60 -u -b 30M -R

IP 地址对应第 4 台辅助云服务器

iperf3 -c 辅助云服务器 4 的 ip 地址 -p xxx -t 60 -u -b 30M -R

IP 地址对应第 5 台辅助云服务器

iperf3 -c 辅助云服务器 5 的 ip 地址 -p xxx -t 60 -u -b 30M -R

3.解析 UDP PPS 测试结果。

UDP PPS 测试结果示例如图所示：



8.14.7.7 不同帐号下弹性云主机内网是否可以互通？

不同帐号下的弹性云主机内网是不通的。

不同的账号和内网地址之间存在隔离机制，这种隔离机制可以保护用户的数据安全和隐私，避免不同用户之间的数据泄露或恶意攻击。同时也可以避免不同用户之间的网络冲突和干扰，确保每个用户的应用程序和数据在网络传输过程中的安全性和稳定性。

8.14.7.8 我购买的云主机是否在同一子网？

如果在创建不同云主机时选择了同一 VPC 下同一子网，则云主机在同一子网；如果创建不同云主机时选择了不同子网，则云主机不在同一子网。针对不同业务需求，可将云主机设在同一子网，或不同子网下面，此操作需要用户根据业务需求自行选择。

若您需要确认创建完成的云主机属于哪一子网，可点开云主机详情页，找到“弹性网卡”标签，点击“展开详情”，查看云主机所在子网。

8.14.7.9 云主机网络优化方案

简单组网

对于相对简单的业务场景，可以仅使用一个 VPC。VPC 下创建不同子网，对应不同功能或安全区域。

比如：用户搭建一个 web 网站，业务简单，无需和其他网络互通。

所有服务器全部放到一个个 VPC 内。

业务子网分为 web/APP/DB 三个区域，分别放置 Web 业务经典三层架构的对应服务器。

如有运维和远程接入需求，可再创建相应子网，运维和接入区子网用于部署堡垒机或管理鉴权设备，方便客户远程接入部署业务和运维。

方便后续通过子网 ACL 的访问权限控制能力实现各个功能区域间的隔离和放通。

复杂组网

针对业务相对复杂的场景，一个 VPC 对应一个相对独立的业务应用系统或部门网络。

比如：一家大型车企业务上云后，所有业务共享同 1 根专线连接线下数据中心。

不同业务或部门分别承载在不同的 VPC。

不同 VPC 分别通过对等连接连接到整个系统的公共资源 VPC，共享专线访问。

8.14.7.10 如何排查带宽超过限制

问题现象

云主机部署对公网开放的服务后，如果公网带宽不足，则可能因为随机丢包，导致公网服务卡顿。

解决思路

云主机绑定的公网带宽，为独享带宽或是共享带宽，其带宽均需要在购买时设置上限。

对于公网带宽不足情况，建议根据您的业务流量，设置合理的公网带宽告警策略，以便及时发现此类问题，提前干预。

如果是随着运行服务需要使用更高带宽的正常运行情况，可以通过对带宽进行扩容解决此问题。

扩容公网带宽，请参考修改弹性 IP 带宽。

设置合理的公网带宽告警策略

您需要根据购买的带宽以及业务运行情况设置合理的告警策略。

假设您购买的公网带宽大小为 10Mbps，基于自定义告警模板，设定告警触发规则，用于系统发送告警通知。

1. 登录控制中心，在产品服务列表内，点击“云监控”，跳转至云监控产品控制台。
2. 左侧导航栏选择“告警服务”>“告警模板”>“创建自定义告警模版”，进入到自定义告警模版管理界面。



3. 创建用于监控公网带宽的自定义告警模版。

以此配置好“选择类型”>“添加规则”>“模版信息”后，点击“确定”后提示“创建告警模版成功”。

选择类型：

- 云服务：选择“弹性 IP 和共享带宽”。
- 维度：请根据您实际使用的带宽类型，选择“弹性 IP”或“共享带宽”。
- 添加规则：

- 监控指标：根据实际业务监控需求，可选择“入网流量”、“入网带宽”、“出网流量”、“出网带宽”。
- 聚合方式：根据实际业务监控需求，可选择“原始值”、“平均值”、“最大值”、“最小值”。
- 阈值方式：根据实际业务监控需求，可选择“=”、“>”、“<”、“≥”、“≤”。
- 告警阈值：根据实际业务监控需求，用户自定义配置（不能超过最大带宽值）。
- 聚合周期：根据实际业务监控需求，可选择“1 分钟”、“5 分钟”、“20 分钟”、“1 小时”、“4 小时”、“12 小时”和“24 小时”。
- 出现次数：根据实际业务监控需求，可选 1-5，最小为 1 次，最大为 5 次。
- 操作：可删除添加的规则，最多可添加 20 条告警规则
- 模版信息：
- 名称：用户自定义，建议根据实际业务规划命名。
- 描述：可选项，建议根据实际业务场景，做简述。

4.创建告警模板后，可参考云监控服务操作说明，创建告警规则并启用告警规则。启用后，您即可收到对应的告警通知。

创建告警规则请参考云监控-创建告警规则

启用告警规则请参考云监控-启用告警规则

8.15 安全类

8.15.1 安全组和安全规则配额是多少

安全组相关配额限制请参考文档：VPC 及安全组相关使用限制

如果当前配额数量不满足您的需求，请您提交工单申请配额扩大。

8.15.2 安全组加规则报错：参数错误，请检查你的参数

请您从以下两个方面考虑问题的出现

1、问题现象：安全组规则配额不足导致；

2、问题现象：系统组件故障或其它未知问题；

对于 1，您可以提交工单申请提升配额；针对 2，您均可提交工单请求后台技术支持。

8.15.3 怎么设置云主机密码可以在一定程度上提升云主机安全？

在设置密码时，用户应尽量将密码长度设置在 8 到 30 位，建议加入数字、大小写字母、特殊符号组成。用户还可以选择购买天翼云安全防护产品，进一步加强弹性云主机的安全防护。关于弹性云主机的相关安全问题，用户还可以通过提交工单的方式联系客服，我们会尽快给予用户相应技术支持。

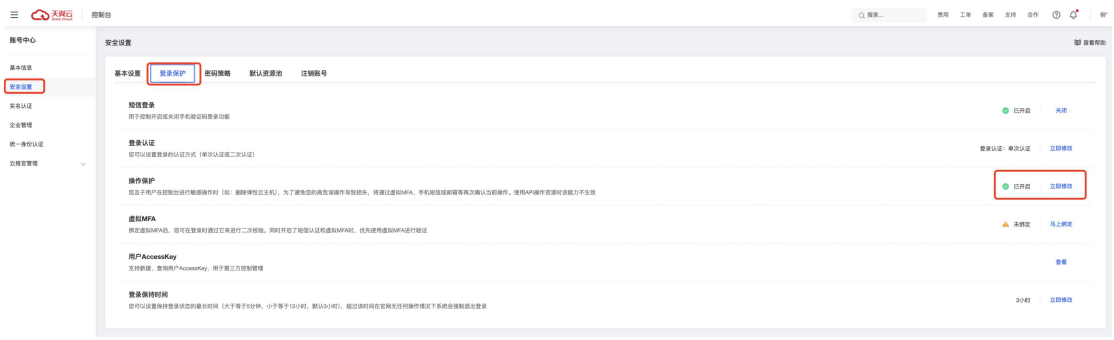
密码设置规则

参数	规则	样例
密码	1、密码长度范围为 8 到 30 位。 2、密码至少包含以下 4 种字符中的 3 种：大写字母、小写字母、数字、 ()`~!@#$%^&*_-+=。 3、密码不能以斜线号（/）开头。	

8.15.4 如何关闭重点操作短信验证

重点操作验证，是天翼云平台为了保障安全，在用户进行重点操作前增加的二次认证，二次认证通过后系统才能执行用户操作。可避免因误操作引起严重后果，提供更高的安全性。您可以按照以下方式，关闭重点操作短信验证功能：

1. 登录天翼云账号，点击右上角头像下方的【个人中心】进入基本信息，下拉点击“安全设置”再点击“登录保护”如下图所示。



2. 点击【立即修改】跳转至概览/设置敏感操作保护，如下图所示。



3. 选择关闭，如下图所示。



4. 点击【保存】，即可关闭重点操作保护。

8.15.5 弹性云主机上是否可以搭建数据库？

云主机作为由 CPU、内存、镜像、云硬盘等资源组成的计算服务，支持使用云主机搭建数据库。为保证数据库读写速度与性能要求，推荐您使用内存优化型云主机搭建数据库。

更多内存优化型云主机信息请参考内存优化型。

8.15.6 如何保护弹性云主机安全

天翼云为您提供安全的弹性云主机产品，通过多种技术手段来保障您的主机安全。但是如果云主机没有进行相应的安全设置，仍然可能收到外部的攻击或者遭到病毒入侵。

您可以从以下方面对云主机进行保护：提高账号安全、提高主机密码安全、使用密钥对、使用云监控、备份云主机、配置访问策略、定期升级操作系统。

详细内容参见提升云主机安全的方法。

8.15.7 如何防止实例被暴力破解

以下是一些有效的防护措施：

1）修改公网 SSH 登陆的端口：不要使用系统默认的 22 端口，建议改成一个较大不常见的端口号，以减少被暴力攻击的可能性；或者使用 NAT 网关，配置 DNAT 规则，使用一个较大不常见的端口号，以减少被暴力攻击的可能性。

2）关闭 SSH 密码登录：使用密钥登录代替密码登录，关闭 SSH 密码登录功能和 root 登录账号，并设置密钥登录。

3）设置强密码：密码应包含大小写字母、数字和特殊符号的组合，并避免使用常见的单词或短语。

4）更新系统和应用程序：及时更新系统和应用程序，安装最新的安全补丁和更新，以修复已知漏洞。

5）启用安全组和云防火墙：限制 SSH 登录的 IP 范围，仅允许从可信的 IP 地址进行访问，并配置和使用防火墙，限制对服务器的入站和出站流量，只允许必要的网络流量通过。

6）定期检查和更新服务器：设置定期的更新和安全检查任务，及时修补已知的安全漏洞，防止攻击者利用旧版软件中的漏洞进行攻击。

7）开通主机安全防护产品：服务器安全卫士（原生版）提供暴力破解能力，实时监控暴力破解口令行为，发现爆破行为自动拦截。

8.16 云主机卡顿

8.16.1Linux 云主机卡顿

问题原因

当您发现云主机的运行速度变慢或云主机突然出现网络断开的现象，则可能是由以下原因导致的：

1. 云主机使用共享资源型实例：共享资源型实例在云计算中是指多个虚拟机共享同一台物理服务器的资源，这会导致资源争用，可能导致服务器卡顿问题。
2. 云主机受到恶意软件或病毒感染：恶意软件可能执行各种恶意任务，占用大量计算资源，增加云主机网络带宽和资源消耗，影响正常的服务运行等，可能导致服务器卡顿问题。
3. 云主机系统资源利用率异常：资源包括 CPU、内存、磁盘和网络等，当其中一个或多个资源受到过度使用或不合理分配时，可能导致服务器卡顿问题。

问题处理

云主机使用资源共享型实例

步骤一：问题定位

检查当前云主机的规格类型，共享资源型和独享资源型实例的说明请参见规格类型。

步骤二：问题处理

独享资源型实例在云计算中独享物理服务器的资源，可以提供更稳定和可预测的性能。如果您对业务稳定性有较高要求，建议您将共享型资源型实例变更为独享资源型实例，请参见变更规格。

云主机受到恶意软件或病毒的感染

步骤一：问题定位

定位云主机是否受到恶意软件感染，您可以运行受信任的杀毒软件或安全扫描工具，对服务器进行全面扫描。此外，您还可以检查操作系统、应用程序和安全软件是否为最新版本，并已应用最新的安全更新和补丁。

步骤二：问题处理

如果服务器受到恶意软件感染，及时采取行动非常重要。隔离服务器、运行安全扫描工具、清除恶意软件、更新系统等都是确保服务器安全的关键步骤。

云主机系统资源利用率异常

步骤一：问题定位

定位云主机中 CPU 利用率高、带宽流量大、内存使用量高的进程。查看网络问题时，您可以检查网线、路由器等硬件设备是否正常工作，之后使用命令行工具 ping 命令来测试与其他计算机的网络连接，检查延迟和丢包情况。

步骤二：问题定位步骤

本部分详细介绍了问题定位步骤，您可按照如下流程进行排查。以 CentOS 7 64 位操作系统为例，其他版本的操作命令请查阅相应操作系统的官方文档。

本文以 CentOS 7 64 位操作系统为例，其他版本的操作命令请查阅相应操作系统的官方文档。

CPU 占用率高

使用 top、htop 和 sar 来监控系统的资源利用率是诊断服务器性能问题的常用方法。其中，top 用于实时监控系统的进程和资源利用率，可以查看 CPU 使用率、内存使用率、运行中的进程列表、负载情况等；htop 是 top 的增强版本，提供更多的交互式功能和更友好的界面；sar 是系统性能分析工具，用于监控和报告系统资源利用率的历史数据。

本文以 top 命令为例，诊断服务器性能问题。

1.执行 top 命令查看当前系统的运行状态。

```
top - 17:39:20 up 7 min,  2 users,  load average: 0,22, 0,57, 0,37
Tasks: 221 total,   1 running, 220 sleeping,   0 stopped,   0 zombie
%Cpu(s):  3,3 us,  2,9 sy,  0,0 ni, 93,6 id,  0,0 wa,  0,0 hi,  0,2 si,  0,0 st
KiB Mem :  1863032 total,    85136 free,   814768 used,   963128 buff/cache
KiB Swap:  2097148 total,  2096628 free,      520 used.   861004 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
 1454 root      20   0  361708  68552  34840 S   7,0  3,7   0:03,81 X
 2428 zycento+  20   0 3562384 225992  80356 S   6,0 12,1   0:15,96 gnome-shell
 3137 zycento+  20   0  688972  31344  17732 S   3,0  1,7   0:00,84 gnome-terminal-
  737 root      20   0  295564   5168   3924 S   0,7  0,3   0:03,43 vmtoolsd
 2399 zycento+  20   0  233104   3932   3172 S   0,7  0,2   0:00,06 at-spi2-registr
   15 root      20   0       0      0      0 S   0,3  0,0   0:01,06 kworker/1:0
  724 root      20   0   21684   1300    992 S   0,3  0,1   0:00,10 irqbalance
 2751 zycento+  20   0  608964  27636  18692 S   0,3  1,5   0:02,71 vmtoolsd
 3228 root      20   0  162252   2340   1580 R   0,3  0,1   0:00,04 top
    1 root      20   0  128304   5588   2760 S   0,0  0,3   0:05,06 systemd
    2 root      20   0       0      0      0 S   0,0  0,0   0:00,02 kthreadd
    3 root      20   0       0      0      0 S   0,0  0,0   0:00,00 kworker/0:0
    4 root       0 -20       0      0      0 S   0,0  0,0   0:00,00 kworker/0:0H
    5 root      20   0       0      0      0 S   0,0  0,0   0:00,79 kworker/u256:0
    6 root      20   0       0      0      0 S   0,0  0,0   0:00,37 ksoftirqd/0
    7 root      rt   0       0      0      0 S   0,0  0,0   0:00,09 migration/0
    8 root      20   0       0      0      0 S   0,0  0,0   0:00,00 rcu_bh
    9 root      20   0       0      0      0 S   0,0  0,0   0:02,14 rcu_sched
   10 root       0 -20       0      0      0 S   0,0  0,0   0:00,00 lru-add-drain
   11 root      rt   0       0      0      0 S   0,0  0,0   0:03,56 watchdog/0
   12 root      rt   0       0      0      0 S   0,0  0,0   0:01,23 watchdog/1
   13 root      rt   0       0      0      0 S   0,0  0,0   0:00,14 migration/1
```

2.分析回显结果。

- 命令回显第一行：17:39:20 up 7 min, 2 users, load average: 0.22, 0.57, 0.37 表示系统当前时间为 17:39:20，该服务器已运行 7 分钟，当前共有 2 个用户登录，最近 1 分钟、最近 5 分钟和最近 15 分钟的 CPU 平均负载分别为 0.22，0.57 和 0.37。
- 命令回显第二行:任务总体情况,Tasks 下的总数(total)、运行中(running)、休眠中（sleeping）、已停止（stopped）、僵尸进程（zombie）表示不同状态下的进程数量。
- 命令回显第三行：CPU 资源总体使用情况，%Cpu(s) 下的用户（us）、系统（sy）、等待（wa）和空闲（id）等表示不同类型的 CPU 使用情况。
- 命令回显第四行：内存资源总体使用情况，KiB Mem 下的总内存（total）、使用内存（used）、空闲内存（free）和缓冲区/缓存（buff/cache）表示不同的内存使用情况。
- 命令回显最下方：进程列表描述了各进程的资源占用情况，显示了系统中运行的进程，包括进程 ID（PID）、CPU 利用率、内存利用率、进程状态等。
- 按下小写“q”键退出 top 命令。

3.通过“ll /proc/PID/exe”命令查看每个进程 ID 对应的程序文件，判断该进程为正常程序或异常程序。

```
[root@elb-mq01 sysconfig]# ll /proc/4243/exe
lrwxrwxrwx 1 root root 0 Mar 18 11:46 /proc/4243/exe -> /CloudResetPwdUpdateAgent/depend/jre1.8.0_131/bin/java
```

4.按下大写“P”键以 CPU 占用率大小的顺序排列进程列表，找到占用 CPU 资源最多的进程，记录其进程 ID 和进程名称。使用工具 strace、perf 等进行进程的性能分析，找出造成高 CPU 使用率的具体原因。

5.如果是您自己的应用程序占用 CPU 较高，可以分析和优化代码，减少 CPU 使用，比如优化算法、减少循环次数、缓存数据等。

6.如果是系统资源不足引起 CPU 占用较高，您可以考虑升级硬件服务器，比如增加 CPU 核数或内存，请参见变更实例的 CPU 和内存规格。

CPU 使用率高

1.如果是异常程序导致 CPU 使用率高，可以直接通过 top 命令终止进程。

- 在 top 命令运行同时，按下小写“k”键，输入要终止进程的 PID 后回车。例如，本文要终止 PID 为 2 的进程。

```
top - 17:51:54 up 20 min,  2 users,  load average: 0.01, 0.11, 0.21
Tasks: 220 total,   1 running, 219 sleeping,   0 stopped,   0 zombie
%Cpu(s):  8.0 us, 20.0 sy,  0.0 ni, 72.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
KiB Mem :  1863032 total,    86288 free,   810836 used,   965908 buff/cache
KiB Swap:  2097148 total,  2096628 free,      520 used.   868164 avail Mem
PID to signal/kill [default pid = 3495] 2
  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
 3495 root      20   0  162100   2212   1528 R   8.7  0.1   0:00.10 top
    1 root      20   0  128304   6404   3576 S   0.0  0.3   0:05.35 systemd
    2 root      20   0       0      0      0 S   0.0  0.0   0:00.02 kthreadd
    4 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kworker/0:0H
    5 root      20   0       0      0      0 S   0.0  0.0   0:00.80 kworker/u256:0
    6 root      20   0       0      0      0 S   0.0  0.0   0:00.41 ksoftirqd/0
    7 root      rt   0       0      0      0 S   0.0  0.0   0:00.09 migration/0
    8 root      20   0       0      0      0 S   0.0  0.0   0:00.00 rcu_bh
    9 root      20   0       0      0      0 S   0.0  0.0   0:02.47 rcu_sched
   10 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 lru-add-drain
   11 root      rt   0       0      0      0 S   0.0  0.0   0:03.56 watchdog/0
   12 root      rt   0       0      0      0 S   0.0  0.0   0:01.23 watchdog/1
   13 root      rt   0       0      0      0 S   0.0  0.0   0:00.21 migration/1
   14 root      20   0       0      0      0 S   0.0  0.0   0:00.57 ksoftirqd/1
   15 root      20   0       0      0      0 S   0.0  0.0   0:01.35 kworker/1:0
   16 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kworker/1:0H
   18 root      20   0       0      0      0 S   0.0  0.0   0:00.02 kdevtmpfs
```

- 终止进程成功后，会显示如下类似的信息，按回车确认。

```
文件(F)  编辑(E)  查看(V)  搜索(S)  终端(T)  帮助(H)
top - 17:52:50 up 21 min,  2 users,  load average: 0.09, 0.12, 0.21
Tasks: 220 total,   1 running, 219 sleeping,   0 stopped,   0 zombie
%Cpu(s):  5.0 us,  2.9 sy,  0.0 ni, 91.9 id,  0.0 wa,  0.0 hi,  0.2 si,  0.0 st
KiB Mem :  1863032 total,    84732 free,   812188 used,   966112 buff/cache
KiB Swap:  2097148 total,  2096628 free,      520 used.   866812 avail Mem
Send pid 2 signal [15/sigterm]
  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
 1454 root      20   0  356148  62984  31544 S   7.6  3.4   0:12.64 X
 2428 zycento+  20   0 3553120 224028  78044 S   7.6 12.0   0:33.81 gnome-shell
 3137 zycento+  20   0  688972  29912  18020 S   3.3  1.6   0:03.96 gnome-terminal-
  729 root      20   0   90568   3184   2320 S   0.3  0.2   0:02.73 rngd
  737 root      20   0  295564   5168   3924 S   0.3  0.3   0:04.94 vmtoolsd
 2751 zycento+  20   0  610168  28772  18940 S   0.3  1.5   0:04.28 vmtoolsd
 3495 root      20   0  162100   2336   1580 R   0.3  0.1   0:00.14 top
    1 root      20   0  128304   6404   3576 S   0.0  0.3   0:05.35 systemd
    2 root      20   0       0      0      0 S   0.0  0.0   0:00.02 kthreadd
    4 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kworker/0:0H
    5 root      20   0       0      0      0 S   0.0  0.0   0:00.80 kworker/u256:0
    6 root      20   0       0      0      0 S   0.0  0.0   0:00.41 ksoftirqd/0
    7 root      rt   0       0      0      0 S   0.0  0.0   0:00.09 migration/0
    8 root      20   0       0      0      0 S   0.0  0.0   0:00.00 rcu_bh
    9 root      20   0       0      0      0 S   0.0  0.0   0:02.49 rcu_sched
   10 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 lru-add-drain
   11 root      rt   0       0      0      0 S   0.0  0.0   0:03.56 watchdog/0
```

2. kswapd0 是 Linux 内核中的后台进程，是内核内存管理子系统中的一部分，负责处理内存交换操作。它的主要目的是将不经常使用的内存页面交换到磁盘空间，以释放物理内存供其他进程使用。使用 top 命令查看 kswapd0 进程的资源使用情况。

```
                                                          root@zycentos7:~
文件(F)  编辑(E)  查看(V)  搜索(S)  终端(T)  帮助(H)
top - 18:03:44 up 32 min,  2 users,  load average: 0.05, 0.04, 0.12
Tasks: 223 total,   2 running, 221 sleeping,   0 stopped,   0 zombie
%Cpu(s):  3.0 us,  2.2 sy,  0.0 ni, 94.8 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
KiB Mem :  1863032 total,    82496 free,   812772 used,   967764 buff/cache
KiB Swap:  2097148 total,  2096628 free,      520 used.   866156 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
   25 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 bioset
   26 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kblockd
   27 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 md
   28 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 edac-poller
   29 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 watchdogd
   35 root      20   0       0      0      0 S   0.0  0.0   0:00.39 kswapd0
   36 root      25   5       0      0      0 S   0.0  0.0   0:00.00 ksmd
   37 root      39  19       0      0      0 S   0.0  0.0   0:00.70 khugepaged
   38 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 crypto
   46 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kthrotld
   48 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kmpath_rdacd
   49 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kaluad
   51 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kpsmoused
   53 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 ipv6_addrconf
   66 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 deferwq
  102 root      20   0       0      0      0 S   0.0  0.0   0:00.05 kauditd
  285 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 ata_sff
```

- 如果 kswapd0 处于睡眠状态，这表示系统当前的内存使用情况相对较好。
- 如果 kswapd0 处于激活状态（不处于睡眠状态），且运行时间较长，这可能暗示系统内存不足，初步判断正在进行持续换页操作，您可以增加内存进行系统规格的升级，或者对应用程序进行优化。

3.使用 vmstat 命令进一步查看系统虚拟内存的使用情况，如果 si 和 so 的值较低，说明系统内存和交换空间的平衡状态良好；如果 si 和 so 的值较高，说明系统正在进行频繁换页操作，您可以增加内存进行系统规格的升级，或者对应用程序进行优化。

```
[root@zycentos7 ~]# vmstat
procs -----------memory---------- ---swap-- -----io---- -system-- ------cpu-----
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
 2  0    520  83084     40 967760    0    0   349    87  125  149  3  2 94  0  0
```

4.当内存不足时，您可以关闭或减少大量内存的应用程序或进程，或者使用 free、ps 等命令进一步查询系统及系统进程的内存占用情况，再做进一步排查分析。

5.您可通过在业务空闲期重启应用或者系统来释放内存，或者优化应用程序、配置大页内存来暂时缓解 CPU 使用率高的情况。如果要从根本上解决内存不足问题，建议您对服务器内存进行扩容。

带宽使用率高

1.如果是正常业务访问和正常应用程序导致带宽使用率高，您可以升级服务器的带宽进行解决。

2.如果是异常应用程序导致带宽使用率高，您可以通过流量监控工具 nethogs 来实时监测统计各进程的带宽使用情况，进一步定位异常进程。

- 使用命令 yum install nethogs -y 安装 nethogs 工具，安装完成后可以查看网络带宽的使用情况。执行命令时指定参数：-d 设置刷新的时间间隔，默认为 1s；-t 开启跟踪模式；-c 设置更新次数；device 设置要检测的网卡，默认是 eth0。
- 例如：使用 nethogs eth1 查看 eth1 网卡上每个进程的网络带宽使用情况。回显参数说明：PID 表示进程 ID；USER 表示运行该程序的用户；PROGRAM 表示进程或连接双方的 IP 地址和端口，第一个是服务器 IP 和端口，第二个是客户端 IP 和端口；DEV 表示流量要去往的网络端口；SENT 表示进程每秒发送的数据量；RECEIVED 表示进程每秒接收的数据量。

```
NetHogs version 0.8.5

  PID USER     PROGRAM                                    DEV        SENT     RECEIVED
 4596 nginx    nginx: worker process                      eth1      34.360      3.267 KB/sec
    ? root     192.168.0.92:98-180.125.68.19:17873                   0.179      0.246 KB/sec
    ? root     192.168.0.92:11211-213.32.10.149:44945                0.000      0.000 KB/sec
    ? root     192.168.0.92:28101-185.176.26.66:43488                0.000      0.000 KB/sec
    ? root     unknown TCP                                           0.000      0.000 KB/sec

  TOTAL                                                             34.540      3.512 KB/sec
```

3.如果是恶意应用程序导致带宽使用率高，您可以使用 kill PID 命令终止恶意进程。

4.如果是非正常业务访问导致带宽使用率高，比如某个 IP 恶意访问、遭受 DDoS 攻击或 CC 攻击，您可以采取以下措施。

- 使用 iptables 服务对指定 IP 地址进行处理，比如屏蔽 IP 地址或限速。
- 查看 Anti-DDoS 攻击是否开启，并检查防护策略是否配置合适，请参考开启 Anti-DDoS 防护。
- 查看 CC 防护策略是否开启，并检查防护策略是否配置合适，请参考 CC 防护。
- 步骤三：问题处理
- 正常程序：优化程序，或变更云主机配置，请参见实例配置变更。
- 异常程序：确保异常程序不会影响到系统组件或其他用户，您可以手动关闭进程或隔离异常程序。根据异常原因，您还可以采用适当的措施来修复问题，比如：更新程序代码、修复配置、应用补丁等。

8.16.2 Windows 云主机卡顿

问题原因

当您发现云主机的运行速度变慢或云主机突然出现网络断开的现象，则可能是由以下原因导致的：

1.云主机使用共享资源型实例：共享资源型实例在云计算中是指多个虚拟机共享同一台物理服务器的资源，这会导致资源争用，可能导致服务器卡顿问题。

2.云主机受到恶意软件或病毒的感染：恶意软件可能执行各种恶意任务，占用大量计算资源，增加云主机网络带宽和资源消耗，影响正常的服务运行等，可能导致服务器卡顿问题。

3.云主机系统资源利用率异常：资源包括 CPU、内存、磁盘和网络等，当其中一个或多个资源受到过度使用或不合理分配时，可能导致服务器卡顿问题。

问题处理

云主机使用资源共享型实例

步骤一：问题定位

检查当前云主机的规格类型，共享资源型和独享资源型实例的说明请参见规格类型。

步骤二：问题处理

独享资源型实例在云计算中独享物理服务器的资源，可以提供更稳定和可预测的性能。如果您对业务稳定性有较高要求，建议您将共享型资源型实例变更为独享资源型实例，请参见变更规格。

云主机受到恶意软件或病毒的感染

步骤一：问题定位

定位云主机是否受到恶意软件感染，您可以运行受信任的杀毒软件或安全扫描工具，对服务器进行全面扫描。此外，您还可以检查操作系统、应用程序和安全软件是否为最新版本，并已应用最新的安全更新和补丁。

步骤二：问题处理

如果服务器受到恶意软件感染，及时采取行动非常重要。隔离服务器、运行安全扫描工具、清除恶意软件、更新系统等都是确保服务器安全的关键步骤。

云主机系统资源利用率异常

步骤一：问题定位

定位云主机是否受到恶意软件感染，您可以运行受信任的杀毒软件或安全扫描工具，对云主机进行全面扫描。此外，您还可以检查操作系统、应用程序和安全软件是否为最新版本，并已应用最新的安全更新和补丁。

步骤二：问题处理

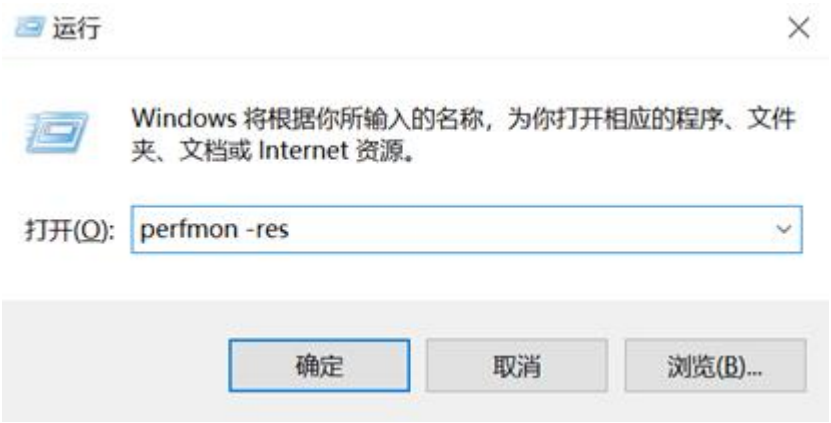
如果云主机受到恶意软件感染，及时采取行动非常重要。隔离云主机、运行安全扫描工具、清除恶意软件、更新系统等都是确保云主机安全的关键步骤。

云主机系统资源利用率异常

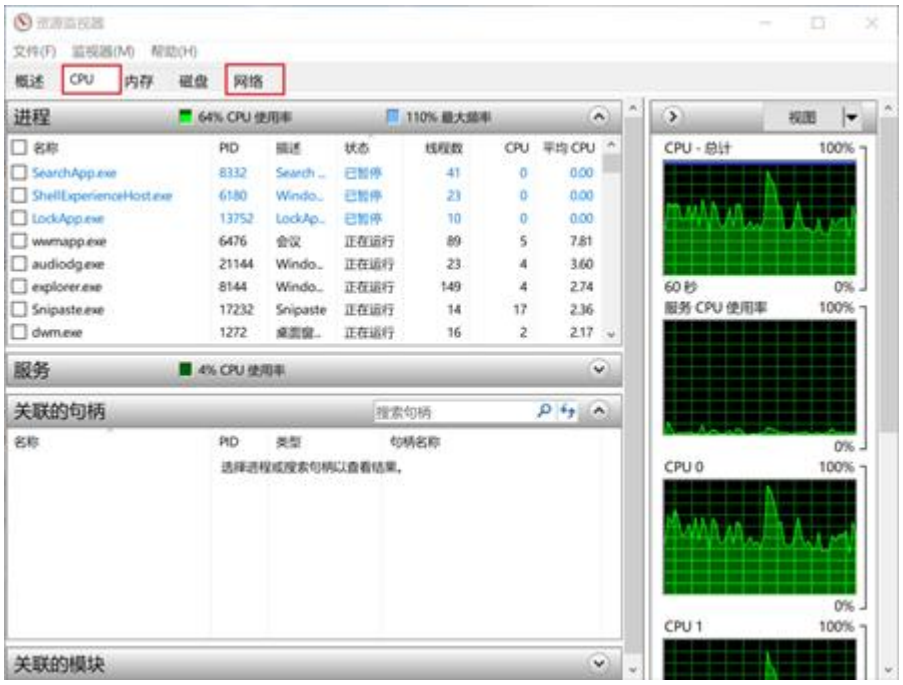
问题定位：定位云主机中 CPU 利用率高、带宽流量大、内存使用量高的进程。Windows 系统提供多个工具用于定位系统资源利用率异常问题，包括任务管理器、性能监视器、资源监视器、事件查看器等。在带宽流量大的情况下，您可以使用 Wireshark 抓取一段时间内的网络包，分析带宽流量的使用情况。查看网络问题时，您可以检查网线、路由器等硬件设备是否正常工作，之后使用命令行工具 ping 命令来测试与其他计算机的网络连接，检查延迟和丢包情况。

问题定位步骤：

1.本部分详细介绍了问题定位步骤，您可按照如下流程进行排查。
2.打开“运行”窗口，输入“perfmon-res”，打开“资源监视器”。

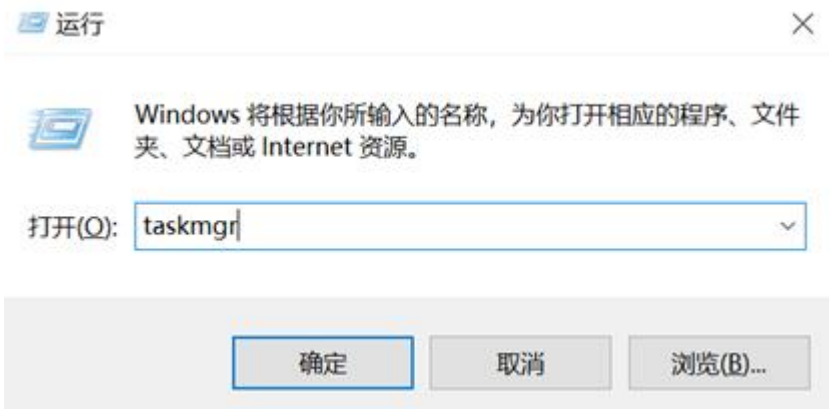


3.在“资源监视器”中，单击“CPU”或“网络”，查看 CPU 占用率或带宽使用情况。

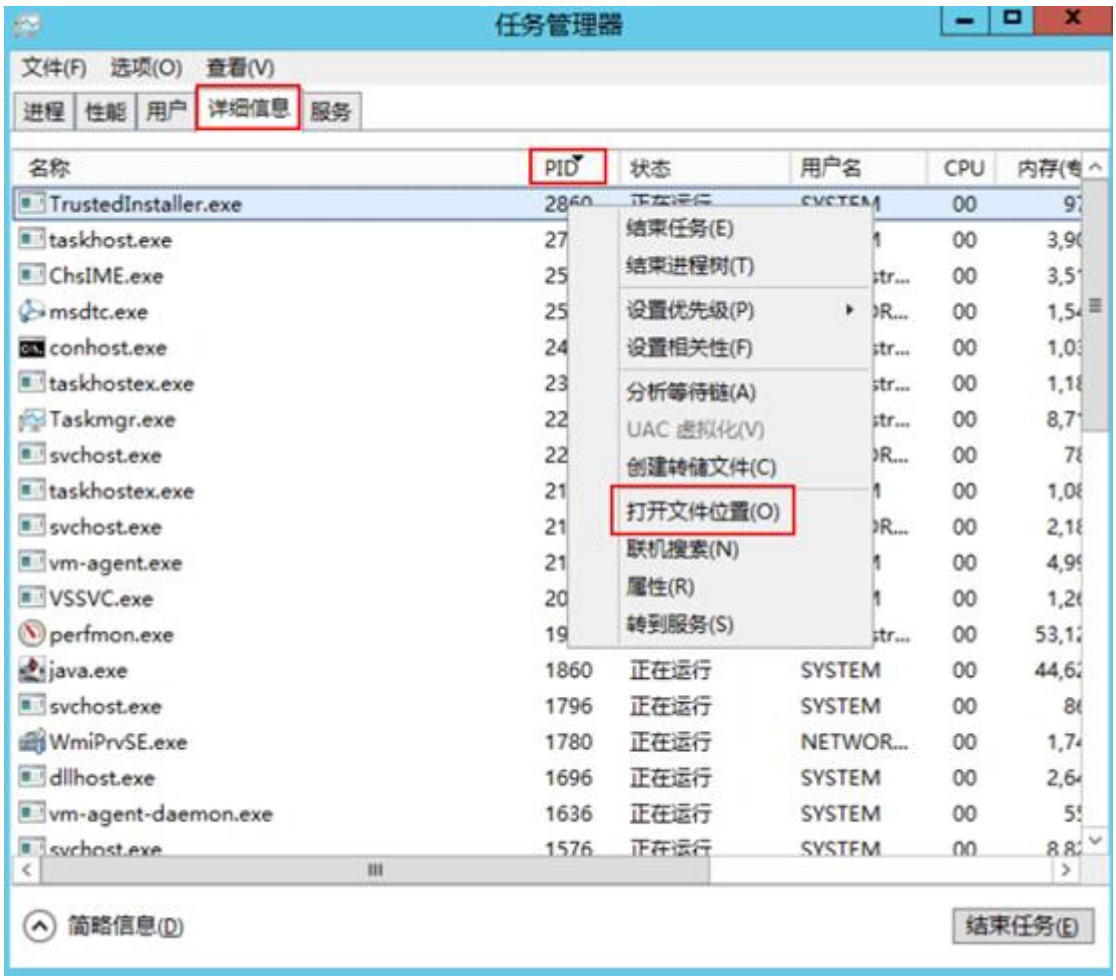


4. 查看 CPU 和带宽占用率较高的进程 ID 和进程名。

5. 打开“运行”窗口，输入“taskmgr”，打开“Windows 任务管理器”。



6. 在“任务管理器”中，选择“详细信息”选项卡，单击“PID”进行排序。在查找到的 CPU 或带宽占用率较高的进程上单击右键，选择“打开文件位置”，定位进程是否为正常程序或恶意程序。



问题处理

排查进程是否正常，进行分类处理。

1. 正常程序：优化程序，或变更云主机配置，请参见实例配置变更。
2. 异常程序：确保异常程序不会影响到系统组件或其他用户，您可以手动关闭进程或隔离异常程序。根据异常原因，您还可以采用适当的措施来修复

分析处理

正常程序处理建议

1. 如果您正在使用 Windows 2008 或 Windows 2012 操作系统，您可以增加物理内存或升级硬件组件，确保您的内存配置在 2GB 或更高。

2.请更新或升级相关的驱动程序，禁用或卸载可能引起冲突的驱动程序，然后观察系统是否仍然卡顿。

3.运行杀毒软件或安全扫描程序，对系统进行全面扫描，查找和清除任何潜在的恶意软件。

4.请检查后台是否有正在执行 Windows Update 的操作，或者杀毒软件是否正在后台执行扫描操作。

5.如果您在云主机中运行的应用程序对系统资源有较高要求，建议您变更云主机配置，请参见实例配置变更或者修改带宽，请参见修改共享带宽。

6.如果您的云主机配置已经比较高，或者您不想修改云主机配置，您可以考虑更新或升级相关的程序，或者在云主机上应用场景分离部署。

异常程序处理

1.确保异常程序不会影响到系统组件或其他用户，您可以手动关闭进程或隔离异常程序。通过右键单击任务栏中程序图标，选择“关闭”或“退出”实现。如果异常程序无反应，通过任务管理器来终止该异常程序的进程。

2.如果异常程序是由于病毒感染、木马入侵导致的，您可以执行以下操作：

- 运行受信任的杀毒软件或安全扫描工具，对服务器进行全面扫描。
- 检查操作系统、应用程序和安全软件是否为最新版本，并已应用最新的安全更新和补丁。
- 使用 MSconfig 禁用所有非微软自带服务驱动，检查问题是否再次发生。

3.如果异常程序让服务器或站点遭受 DDoS 攻击或 CC 攻击，短期内产生大量的访问需求，您可以执行以下操作：

- 查看 Anti-DDoS 攻击是否开启，并检查防护策略是否配置合适，请参考开启 Anti-DDoS 防护。
- 查看 CC 防护策略是否开启，并检查防护策略是否配置合适，请参考 CC 防护。

8.17 数据库应用

8.17.1 弹性云主机是否支持 Oracle 数据库？

弹性云主机能力上支持 Oracle 数据库，但是由于 Oracle 数据库本身可提供高性能与高并发等场景支持，所以在性能方面考虑更推荐使用物理机搭建 Oracle 数据库。建议您在正式使用前，先做性能测试看是否能满足您的需要。

更多关于物理机产品的介绍请参考物理机产品定义。

8.18 云主机快照常见问题

云主机重装系统，云主机快照是否还支持使用？

不支持。云主机重装前需要先删除云主机快照。

云主机重装系统，云主机快照是否还支持使用？

不支持。云主机重装前需要先删除云主机快照。

云主机退订，删除，云主机快照是否还支持使用？

不支持。云主机退订，删除前需要先删除云主机快照。

本地盘云主机支持创建快照吗？

不支持。建议您在应用层做好数据冗余处理，或者为集群创建部署集，提高应用的高可用性。

一个云主机可以绑定几个快照策略？

一个云主机只支持绑定一个快照策略。当云主机已经绑定云主机快照策略 A 时，再次绑定到云主机快照策略 B，云主机将会自动解绑策略 A。

云主机快照能否下载或导出到本地？

云主机快照不能下载或导出到本地。有下载需求，请参考镜像相关功能。

文件系统一致性快照和非一致性快照的区别？

云主机快照支持非一致性快照和文件系统一致性快照，在创建快照和配置快照策略时，可以根据用户需求选择文件系统一致性快照选项，选中时创建文件系统一致性快照。云主机快照不支持应用一致性快照，应用一致性需要根据用户自身应用特点从应用架构和实现去保障，比如停用应用，应用一致性同步，分布式，主备，异地多活等手段。

文件系统一致性快照会先同步操作系统内存数据，然后冻结文件系统，快照完成后解冻文件系统。 非一致性不会同步数据，冻结文件操作，不能保证所有云硬盘的数据一致性，快照有时间差。

文件系统一致性快照，会对业务系统造成影响，建议用户在业务低峰或主机空闲时段创建快照时。高 IO 的等场景可能会影响客户业务，存在文件系统长时间冻结等情况，建议做非一致性快照。

云主机快照策略绑定云主机数量限制问题怎么解决？

一条云主机快照策略默认只能绑定 10 个云主机，一个池子下默认只能创建 20 条策略。用户可以通过创建多条相同的策略实现绑定更多云主机。

或用户可以提工单提升云主机快照策略配额，云主机快照策略绑定的云主机配额最大能调整到 50，创建策略数量最大能调整到 100。

8.19 产品咨询类

8.19.1 怎样注册天翼云

怎样注册天翼云 ？

1.登录天翼云官网 www.ctyun.cn，点击右上角“免费注册”按钮。

2.填写注册信息。填写邮箱后，设置登录密码，并通过手机验证。注册邮箱不可更改，请确保邮箱的准确性。一个手机号最多可注册 5 个天翼云账号，超出则不再支持注册新的天翼云账号。

3.勾选协议，并点击“同意协议并提交”，即可完成账号注册。

8.19.2 弹性云主机可以做什么

弹性云主机可以用作多种用途，如下：

- 搭建网站

弹性云主机最常见的应用场景就是用来搭建网站，一般有企业类型网站、个人网站、论坛网站等。比如小配置的云主机非常适合用来搭建个人博客，经济又实用。

- 数据库应用

在互联网行业中，有些企业选择给自己的网站做站库分离，这样可以减轻主机的业务压力，用两台云主机合作共同为网站服务，其中一台云主机用来专门储存和管理数据库，可以发挥云主机更大的性能。

- 大数据分析和处理。

用户可以使用弹性云主机提供的高性能计算资源来处理大规模的数据集，洞察和提取有价值的信息。

云主机还有多种用途，可以供大家进行探索使用。

8.19.3 私有数据是否支持物理邮寄

出于安全性考虑，云上的服务与数据基于物理存储介质的传递与交互。因此私有数据只能通过网络上传下载，不支持物理邮寄。

您可以通过以下两种方式，实现私有数据的上传或下载：

1、自建 FTP 服务，通过公网实现云主机实例与您本地服务器的数据传输。

Linux 云主机搭建 FTP 方式请见：快速构建 FTP 站点（Linux）。

Windows 云主机搭建 FTP 方式请见：快速构建 FTP 站点（Windows）。

2、通过私有镜像导入与导出的功能，将数据随私有镜像导入到云平台，或者从云平台下载。

私有镜像导入到云平台的操作步骤请见：通过镜像文件创建系统盘镜像。

私有镜像从云平台导出的操作步骤请见：导出镜像。

8.19.4 升级 CPU、内存、带宽时是否需要关机

升级 CPU 和内存配置，是对云主机规格进行升配操作。

所有规格均支持关机后升配，通常成为冷变配。只有部分规格与镜像支持“运行中”状态下升配。云主机变配规则与说明请见云主机帮助中心：云主机变更配置概述。

云主机变配操作手册请见：云主机变更规格。

升级带宽，通常指的是对公网带宽进行升级，通过升级可以获得更大的公网带宽，同时也需要支付更高的带宽费用。弹性 IP-用户指南-修改带宽。

8.19.5Linux 虚机 product_uuid 参数是否可以修改

product_uuid 是计算机唯一标识符（UUID），是一种用于标识计算机硬件的数字代码。

在云平台上，Linux 云主机的 product_uuid 参数是由虚拟化层模拟的硬件参数，通常情况下不同云主机之间的 product_uuid 不会重复，也无法修改。

8.19.6 香港节点的云主机是否可以实现与国内云主机互通

可以实现。

香港节点与国内节点，网络涉及跨境。因此当前互通方案中，仅提供公网互通与跨境 SD-WAN 互通两种方式。香港节点不支持对等连接和云间高速互通。

公网互通：将国内节点的云主机与香港节点的云主机分别绑定弹性公网 IP，公网直连互通。

SD-WAN 互通：通过 SD-WAN 产品，实现跨资源池之间的网络互通，详细请见：天翼云 SD-WAN 产品手册。

8.19.7 云主机 windows 2019 操作系统对应普通电脑哪个操作系统

基于 Microsoft 官方文档给出的说明中，Windows Server 2019 对应的大概是 Windows 10 版本 1809。

如果您需要了解关于具体的版本特性，建议您联系微软进行咨询。

相关说明请参考微软官方文档：Windows 10 版本 1809 和 Windows Server 2019 | Microsoft

8.20 其它类

天翼云主机的默认 grub 密码是多少？

默认 grub 没有密码。

服务器上安装 Web 服务，外网可以直接访问吗？

如搭建的 web 服务使用的端口为 80 端口，需要在天翼云进行备案才可以访问。

any 端口入方向是否需要开放？

出方向 any 是全部协助，这个必须要放通，不然主机无法上网，入方向只需要您放通您所需访问主机的端口即可，不建议放通 any，放通 any 主机可能会有安全隐患。

香港节点主机能否访问外网，如谷歌，推特？

可以访问外网。谷歌、推特可以访问。

使用云资源如何优化成本？

相比自建数据中心，使用云资源时无须投入硬件、物理环境人力等成本，单位资源成本相对线性，所有资源按需取用，交付便利。除此之外，云资源支持多种付费模式，方便进一步优化成本。　针对使用云资源如何优化成本，我们给出以下建议：

追踪成本

从费用账单了解消费情况，追踪成本并确定如何优化；　使用标签等功能分类资源，以便统计相应成本。

优化资源

首先，监控资源使用情况 ，监控资源利用率，评估当前配置是否过高。例如 CPU、内存、云盘、带宽等资源的利用率。　监控闲置的资源，避免浪费。例如升配但未挂载的云盘、未关联的弹性 IP 等。　监控资源使用周期。如果长期使用按量付费实例、云盘等资源，考虑以更实惠的方式购买，例如包年包月、资源包等。　监控资源生命周期，了解包年包月资源的到期日，及时续费。例如包年包月实例、存储容量单位包等。 其次，选择合适规格， 根据业务场景选择最佳性价比的实例规格，并调整合适的数量，在满足业务需求的同时追求高资源利用率，可参考选型最佳实践。第三，组合付费模式 ，不同类型的业务对资源使

用周期有不同要求。为每一类业务确定合适的付费模式，灵活组合达到最优效果。如业务负载稳定使用包年包月方式，业务负载动态变化使用按量付费方式。

树立节约意识

按需取用是云计算最大的一个特点，您可以将成本优化融入到日常工作中，如：定期盘点资源使用情况，明确闲置资源的通知和处置流程；定期和成本相关方（例如财务、研发等团队）评审预算执行情况，改进优化策略；按时续费，提前申请包周期预算，避免到期释放后重新购买部署增加额外成本。

自研池的 AS 号是多少？

自研池的 AS 号的取值范围为：64513---65081。

云主机是否支持 vrrp 协议？

部分支持（不支持广播的 vrrp，只支持单播的 vrrp）。

云主机是否具备组播功能、是否支持修改 MTU 值、是否支持 Oracle rac 集群？

不具备组播功能、不支持修改 MTU 值、不支持 Oracle rac 集群；如果要搭建 Oracle rac 集群，可以使用安装了 FC san 存储的物理机。

目前不支持 VPC 切换的资源池有哪些？

长沙 3 、 武汉 3、 内蒙 5、 上海 6、 佛山 2 、 北京 4、 上海 5 、佛山 1、沈阳 2 、保定、 拉萨 2 、拉萨、 武威、 郴州、 泉州 1、 中卫 2、成都 4、 宜昌 、杭州、武林、边缘、 随州、 广州 6、南京 5 不支持切换 VPC，其他资源池均支持。如果不能切换，派单后端开通权限（以上地域随着实际迭代会有变化，已实际线上情况为准）。

最大支持创建多少个快照?

单台云主机最大支持创建 7 个快照，每个账号可以创建 30 个快照，支持调整配额。

天翼云主机服务器可以装声卡和扬声器吗？

暂时不支持安装声卡和扬声器。

自研池中 s2 型云主机的 CPU 基频/睿频 CPU 型号的具体参数？

5118（2.3 GHz/3.2 GHz ）/5218（2.3 GHz/3.9 GHz）/5218R（2.1 GHz/4.0 GHz ）/ E5-2650 v2（2.6 GHz/3.4 GHz） / E5-2630 v4（2.2 GHz/3.1 GHz）

linux 系统 VNC 登录界面如何翻页查看命令回显记录？

shift+pgup 是向上翻页，shift+pgdn 是向下翻页。

天翼云服务器能否更改 BIOS？

天翼云弹性云主机无法更改 BIOS。

云主机能否安装在手机上面使用？

主机不支持安装在手机使用，但可以使用手机进行远程。

VNC 连接可选的分辨率太少，是否支持选择多种分辨率？

无法更改，若对图像质量有硬性要求建议购买 GPU 云主机。

云主机 Windows 操作系统，哪个版本占用空间最小？

系统占用空间相差不多。

如何注册天翼云账号？

1.登录天翼云官网 www.ctyun.cn，点击右上角“免费注册”按钮。2.填写注册信息。填写邮箱后，设置登录密码，并通过手机验证。 注册邮箱不可更改，请确保邮箱的准确性。一个手机号最多可注册 5 个天翼云账号，超出则不再支持注册新的天翼云账号。3.勾选协议，并点击“同意协议并提交”，即可完成账号注册。

能否将一个账号下的云主机转移到另一个账号下？

暂时无法提供账号资源转移的服务。

忘记天翼云账号密码怎么办？

当用户忘记密码时，可通过登录页面点击“忘记密码”进行重置密码。1.进入天翼云官网，点击右上角“登录”，在登录页面点击右下角的“忘记密码”。2.填写需要找回的账号，并进行验证。填写邮箱，验证码会通过邮箱发送。填写手机号，验证码会通过手机发送。说明：邮箱需在管理中心-基本信息完成邮箱验证后，方可接收验证码。3.验证成功后，设置新的登录密码。

一个手机号能注册几个天翼云账号？

一个手机号最多可注册 5 个天翼云账号，若该手机号注册天翼云账号已达上限，则不再支持注册新的天翼云账号。

一个证件号能认证几个云帐号？

同一证件最多可认证 5 个天翼云账号（包含个人账号、企业账号），同一证件指身份证、企业营业执照。

无法收到验证短信怎么办？

手机收不到短信验证码的几种情况：手机装了拦截软件；手机关机、暂无信号、欠费或者停机；手机欠费停机后充值复机第一天无法接收 SP 短信；用户自己在运营商那边设置了黑名单；手机无法接收长度超过 70 个字的短信；手机在境外使用或者使用境外手机；手机短信息存储满了；个别地区有手机网络、运营商故障等。

8.21 云助手常见问题

什么是云助手？

天翼云云助手（CT-CA， Cloud Assistant）是专门为云服务器打造的原生自动化运维工具，免密码、免登录、无需使用跳板机，即可批量执行命令（Shell、PowerShell、Bat、Python 等），完成运行自动化运维脚本、轮询进程、安装卸载软件、启动或停止服务、安装补丁或安装安全更新等任务。

如何使用云助手？

您可以通过弹性云主机、物理机控制台创建及执行命令，详情请参见 云助手产品介绍 。

可以修改已经创建的命令吗？

可以。云助手支持修改命令基本信息中的所有内容，具体步骤如下：

1. 登录 弹性云主机控制台 或 物理机控制台 ，选择左侧导航栏中的运维工具。
2. 展开运维工具下云助手标签页，选择我的命令标签页。
3. 在命令列表中选择需要修改的命令，点击修改命令按钮。
4. 在弹出的“修改命令”窗口中，编辑后单击保存即可。

为什么返回结果中显示乱码？

云助手支持中文。出现乱码可能有以下原因：

- 您输入的命令内字符编码方式非 UTF-8 编码，请确保采用 UTF-8 编码。
- 您执行的命令本身输出了二进制内容等非 UTF-8 编码字符，该情况与您直接在终端执行命令是同样的效果。例如，tail -n4 /usr/bin/ls 命令的输出一定是包含乱码的。

为什么某些脚本在本机可以执行，但通过云助手执行会提示 command not

found?

可能是由于 PATH 环境变量设置导致，建议您使用绝对路径来执行命令，或确保您需要执行的命令路径已经在 PATH 环境变量中。例如，使用 /usr/local/bin/python test.py 替代 python test.py，以获取更加明确的执行预期。

为什么在控制台执行命令时，提示实例未安装云助手？

可能是实例初始化时未安装、未成功安装云助手、云助手 Agent 未初始化完成或安装后被删除或停用引起。

云助手客户端的默认安装路径在哪里？

默认安装路径为：

- Linux 云主机：/usr/bin/qemu-ga
- Windows 云主机：C:\Program Files\Qemu-ga\Qemu-ga.exe
- Linux 物理机：/usr/local/bin/ca-agent

如何查看云助手客户端状态？

- Linux 云主机：

查看是否安装了 QGA：执行 ps -ef | grep qemu-ga 查看进程是否存在，或者 systemctl status qemu-guest-agent 服务为 active。

- Windows 云主机：

查看是否安装了 QGA：打开任务管理器，选择服务，查看 QEMU Guest Agent 服务是否正在运行。

- Linux 物理机：

查看是否安装了 ca-agent：执行 ps -ef | grep ca-agent 查看进程是否存在，或者 systemctl status ca-agent 服务为 active。

如何安装云助手客户端？

通过下载安装包，解压安装。详情请参见安装云助手客户端。

云助手命令内容是明文传输还是加密传输？

创建命令、执行命令 接口中命令内容是加密的，加密方式参考：命令内容加密

9 故障修复

9.1 高频故障案例

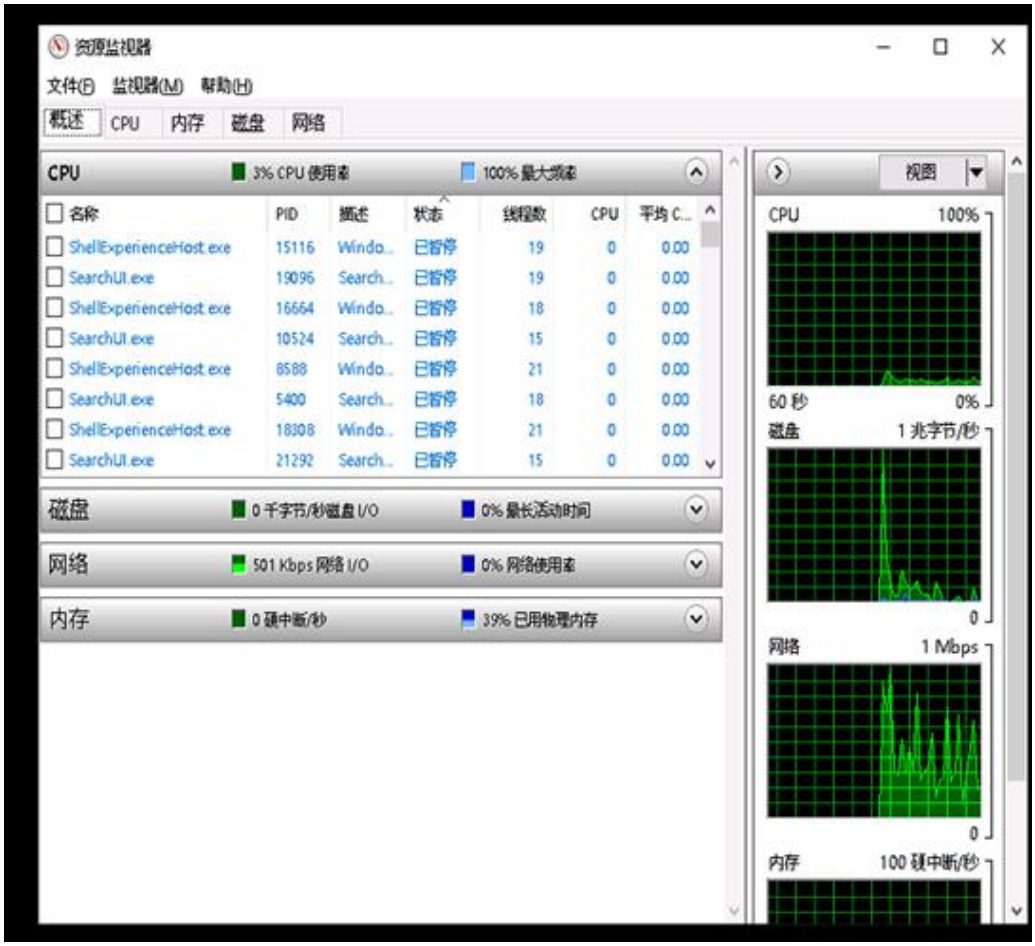
9.1.1 云主机带宽占用高怎么办

操作场景

如果云主机操作卡顿或无法连接，可能是由于云主机带宽占用过高导致的，下面内容介绍了排查云主机带宽占用高的方法及相应的解决方案。

Windows 操作系统云主机

1. 首先登录云主机，以 Windows server 操作系统云主机为例。
2. 在云主机中打开“运行”窗口，输入“perfmon -res”，打开资源监视器。



3. 在“资源监视器”中，单击“CPU”或“网络”，查看 CPU 占用率或带宽使用情况。

4.在监视器中查看带宽占用率较高的进程名，一般会有以下情况。

（1）如果消耗带宽较多的进程为业务进程，表示云主机规格不满足业务运行需求，建议变更云主机的配置。

（2）如果消耗带宽较多的进程为异常进程，可能是病毒或木马导致，建议自行终止进程或者使用安全软件进行查杀。必要时候建议重装系统确保系统正常运行。

Linux 操作系统云主机

1. 通过管理控制台登录云主机，以 CentOS 7.6 操作系统云主机为例。

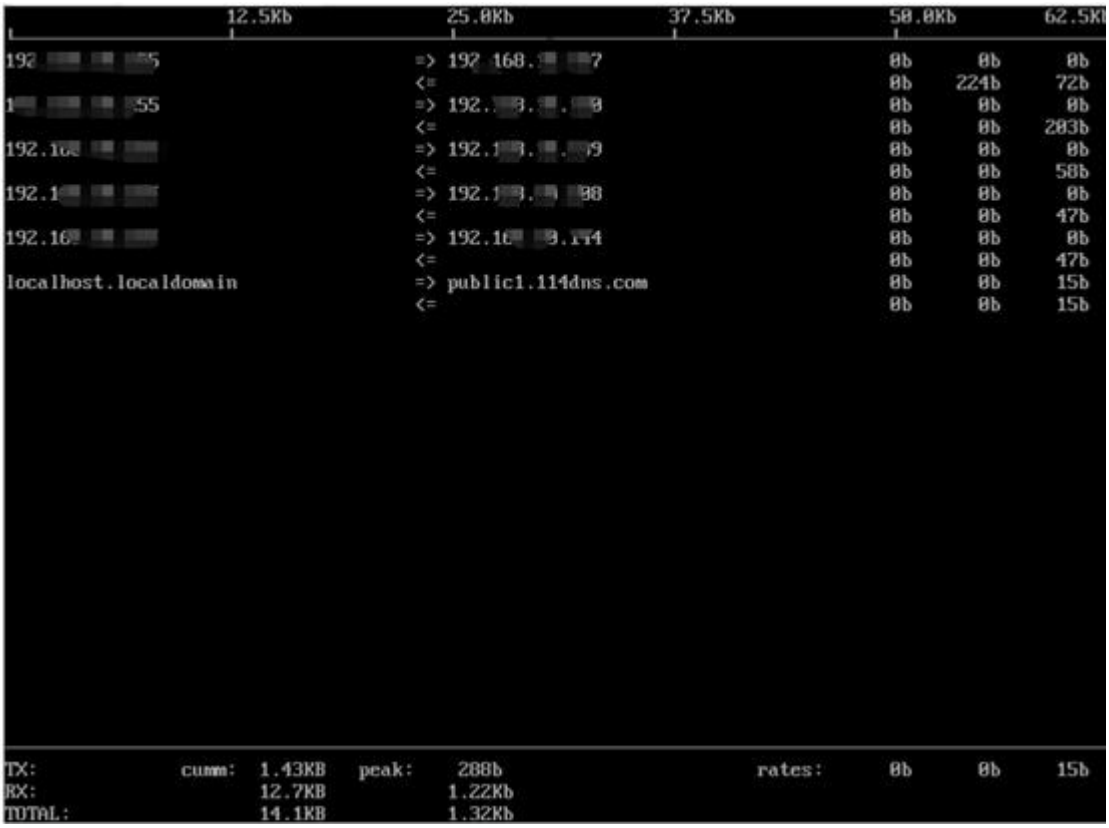
2. 执行以下命令安装 Linux 流量监控工具 iftop。

```
yum install iftop -y
```



3.执行如下命令，查看导致流量较高的端口与消耗流量的 IP，以 eth0 端口为例。

```
iftop -i eth0 -P
```



4.执行如下命令查看端口对应的进程，以 38366 端口为例。

```
netstat -tunlp |grep 38366
```

TX 表示发送流量，RX 表示接收流量，TOTAL 表示总流量。

cum：表示第一列各种情况的总流量。

peak：表示第一列各种情况的流量峰值。

rates：表示第一列各种情况 2 秒、10 秒、40 秒内的平均流量。

5.查看带宽占用率较高的进程名。

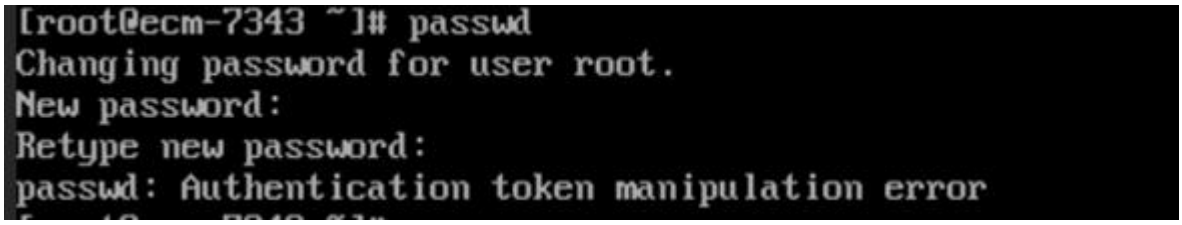
（1）如果消耗带宽较多的进程为业务进程，表示云主机规格不满足业务运行需求，建议变更云主机的配置。

（2）如果消耗带宽较多的进程为异常进程，可能是病毒或木马导致，建议自行终止进程或者使用安全软件进行查杀。必要时候建议重装系统确保系统正常运行。

9.1.2Linux 操作系统执行 passwd 命令重置密码失败提示：Authentication token manipulation error

问题现象

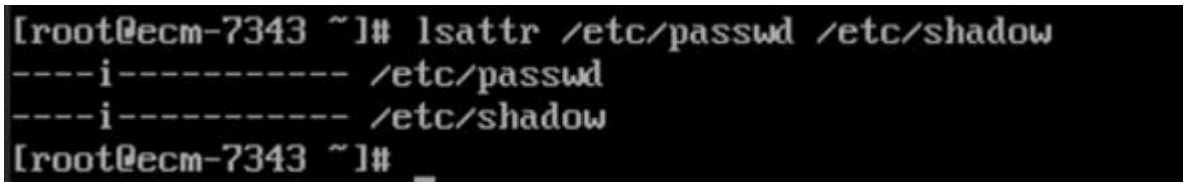
root 用户用 passwd 命令修改管理员用户以及普通用户的密码时失败，提示 passwd:Authentication token manipulation error。



根因分析

出现该问题通常是密码文件的属性的问题导致，但也有可能是根目录空间满。执行以下命令查看存放用户和密码的文件（/etc/passwd 和/etc/shadow）属性。

```
# lsattr/etc/passwd/etc/shadow
```



如上图所示，/etc/passwd 和/etc/shadow 文件中有 i 属性，"i"的文件属性表示该文件不能修改：它不能被删除或重命名，无法为此创建任何链接文件，不能将任何数据写入该文件。只有管理员用户可以设置或清除此属性。

处理方法

用 chattr 命令将 i 权限撤销，然后再修改密码。

如果文件属性为"i"，执行以下命令。

```
# chattr -i /etc/passwd /etc/shadow
```

如果文件属性为"a"，执行以下命令。

```
# chattr -a /etc/passwd /etc/shadow
```



如上图所示，/etc/passwd 和/etc/shadow 文件中无 i 属性，此时可以正常修改密码。

9.1.3 云主机错误状态及解决方案

一、引言

随着云计算技术的快速发展，云主机已经成为企业、个人和开发者们进行互联网应用的首选。然而，使用云主机的过程并非完全顺利，可能会遇到各种错误状态。本文将详细探讨云主机错误状态的类型、原因以及解决方案，帮助用户更好地管理和维护自己的云主机。

二、云主机错误状态类型

1.网络错误

网络错误是云主机使用过程中最常见的错误类型之一。网络错误可能包括网络连接超时、无法连接到远程服务器或网络连接中断等。这些错误通常是由于网络配置问题、网络设备故障或网络拥堵等原因引起的。

2.服务器错误

服务器错误通常包括服务器启动失败、运行异常、过载等。这类错误可能是由于服务器硬件故障、软件问题、操作系统故障或服务器资源不足等原因引起的。

3.应用程序错误

应用程序错误是指运行在云主机上的应用程序出现错误。这类错误可能包括应用程序崩溃、运行缓慢、内存泄漏等。这些错误可能是由于应用程序本身的问题，如代码错误、不兼容性或配置问题等引起的。

4.安全错误

安全错误是指与安全相关的错误，如身份验证失败、权限问题或安全策略冲突等。这些错误可能是由于用户名或密码错误、权限配置错误或安全策略设置不当等原因引起的。

三、云主机错误状态原因分析

1，网络问题

网络问题可能是由于网络设备故障、网络连接问题或网络配置不当等原因引起的。例如，如果云主机的网络设备出现故障，或者网络连接被中断，那么就会出现网络错误。此外，如果网络配置不正确，例如 DNS 解析不正确，也可能会导致无法连接到远程服务器。

2.服务器硬件故障

服务器硬件故障可能是由于服务器硬件设备出现故障或服务器资源不足等原因引起的。例如，如果服务器的硬盘出现故障，那么服务器可能无法启动或运行异常。此外，如果服务器的资源不足，例如内存不足或 CPU 过载，也可能会导致服务器运行异常。

3.应用程序问题

应用程序问题可能是由于应用程序本身的问题，如代码错误、不兼容性或配置问题等引起的。例如，如果应用程序的代码存在错误，那么应用程序可能无法正常运行。此外，如果应用程序的配置不正确，例如数据库连接不正确或文件路径错误，也可能会导致应用程序运行异常。

4.安全问题

安全问题可能是由于身份验证失败、权限问题或安全策略冲突等原因引起的。例如，如果用户名或密码不正确，那么身份验证可能会失败。此外，如果权限配置不正确，例如权限设置过于宽松或过于严格，也可能会导致权限问题。另外，如

果安全策略设置不当，例如过于宽松的安全策略或过于严格的安全策略，也可能会导致安全问题。

四、云主机错误状态解决方案

1.网络错误的解决方案

对于网络错误，可以通过以下步骤来解决：首先检查网络设备是否正常工作；其次检查网络连接是否正常；最后检查网络配置是否正确。如果以上步骤都无法解决问题，可以考虑联系网络管理员或运营商寻求帮助。

2.服务器错误的解决方案

对于服务器错误，可以通过以下步骤来解决：首先检查服务器硬件设备是否正常工作；其次检查服务器软件是否正常工作；最后检查服务器资源是否充足。如果以上步骤都无法解决问题，可以考虑联系服务商或技术支持团队寻求帮助。

3.应用程序错误的解决方案

对于应用程序错误，可以通过以下步骤来解决：首先检查应用程序是否存在代码错误；其次检查应用程序的配置是否正确；最后检查应用程序是否与操作系统或其他软件兼容。如果以上步骤都无法解决问题，可以考虑联系应用程序开发者或技术支持团队寻求帮助。

9.1.4 如何解决云平台创建云主机状态错误

在使用云平台创建云主机时，有时候可能会遇到云主机状态错误的情况。这个错误通常表示在创建云主机的过程中出现了一些问题，导致云主机无法正常启动或运行。本篇博客文章将介绍如何解决云平台创建云主机状态错误的问题，帮助读者快速解决该错误并成功创建云主机。

第一部分：检查云主机配置

1.确保已正确配置云主机的基本参数，包括名称、镜像、规格、网络等。
2.检查是否选择了适合的镜像和规格，确保云主机的配置满足需求。

3.确保云主机所属的网络和子网配置正确，并且与其他网络设备兼容。

第二部分：检查资源配额

1.登录云平台控制台。
2.在控制台的菜单中，找到“云主机”或类似的选项，并点击进入云主机管理页面。
1.在云主机管理页面中，找到目标项目，并点击进入项目详情页面。
2.在项目详情页面中，查找并点击“资源配额”或类似的选项。
3.检查是否已达到了某些资源的配额限制，如CPU、内存、磁盘等。
4.如果资源配额不足，可以联系系统管理员或云平台的技术支持团队，申请增加相应资源的配额。

第三部分：检查日志和错误信息

1.在云平台控制台中，找到云主机管理页面。
2.在云主机管理页面中，找到目标云主机，并点击进入云主机详情页面。
3.在云主机详情页面中，查找并点击“日志”或类似的选项。
4.注意查找是否有报错信息或异常情况的记录。
5.根据日志和错误信息，进行相应的故障排查和修复操作。

第四部分：联系技术支持

1.如果以上方法都无法解决云主机状态错误的问题，可以联系云平台的技术支持团队，寻求进一步的帮助和支持。
2.提供详细的错误信息、日志和操作步骤，以便技术支持团队更好地理解和解决问题。

总结

通过以上步骤，我们可以尝试解决云平台创建云主机状态错误的问题。在遇到问题时，及时检查云主机配置、资源配额和日志信息，同时与技术支持团队保持沟通和协作，可以更快地解决问题并成功创建云主机。

9.1.5 云主机端口不通怎样排查？

问题现象

云主机端口不通，会导致云主机之间无法通信或云主机无法对外服务，会影响应用或服务间的调用，或者用户对服务的访问。

根因分析

可能有以下几种原因导致端口不通：

1.在控制台界面，云主机所在安全组未放行该端口；
2.应用服务配置存在问题，对外服务端口未被正常监听；
3.操作系统内，防火墙配置策略异常，如防火墙未放行端口访问。

问题定位步骤

本文分别对 Windows 和 Linux 操作系统的远程连接端口进行排查，以下为操作步骤。

Windows 操作系统

此部分以 Windows Server 2012 操作系统的云主机为示例，对云主机无法远程连接问题的定位步骤。

步骤 1：排查安全组是否放通 Windows 远程连接端口 3389。

1.登录控制中心，选择云主机所在区域，点击“弹性云主机”进入云主机控制台。
2.在云主机列表，点击需要远程连接的云主机实例名称。
3.点击“安全组”页签，查看安全组规则。

4. 检查云主机所在的安全组是否已添加 3389 入方向的安全组规则。

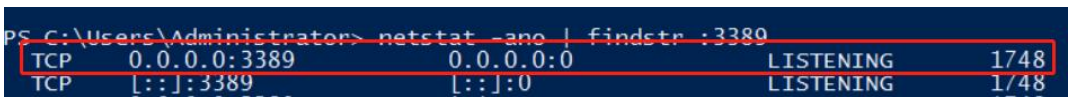


步骤 2 ：排查端口是否正常被监听。

在 Windows 操作系统中打开 cmd 窗口，执行如下命令。

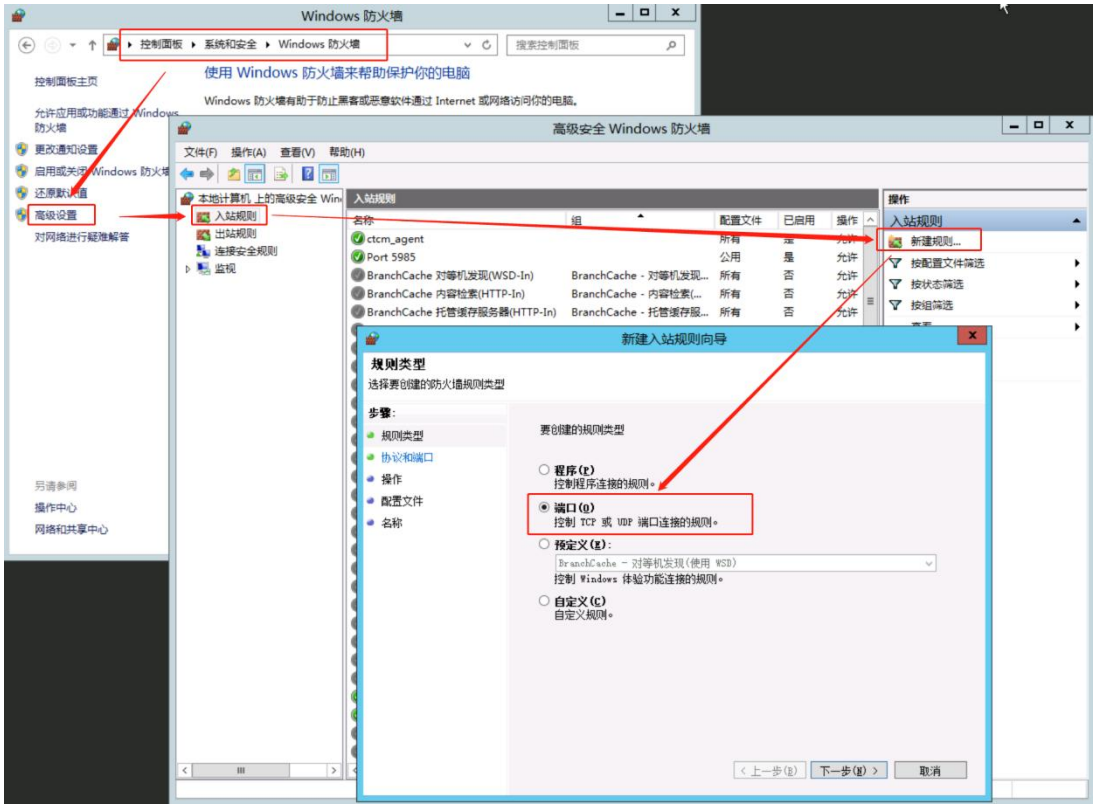
```
netstat -ano | findstr :3389
```

如果回显信息如下则说明 3389 端口正常全网监听。否则，请开启监听。



步骤 3 ：排查防火墙已经放行服务。

1. 单击“控制面板”>“Windows 防火墙”。
2. 根据防火墙状态，如果防火墙处于关闭状态，且您不需要使用防火墙，则无需再做其他处理。
3. 如果防火墙处于开启状态，则单击“高级设置”。
4. 在弹出窗口的左侧导航栏中，单击“入站规则”。
5. 右键“入站规则”>新建入站规则>选择“端口”>填写“需要放开的端口”完成配置。

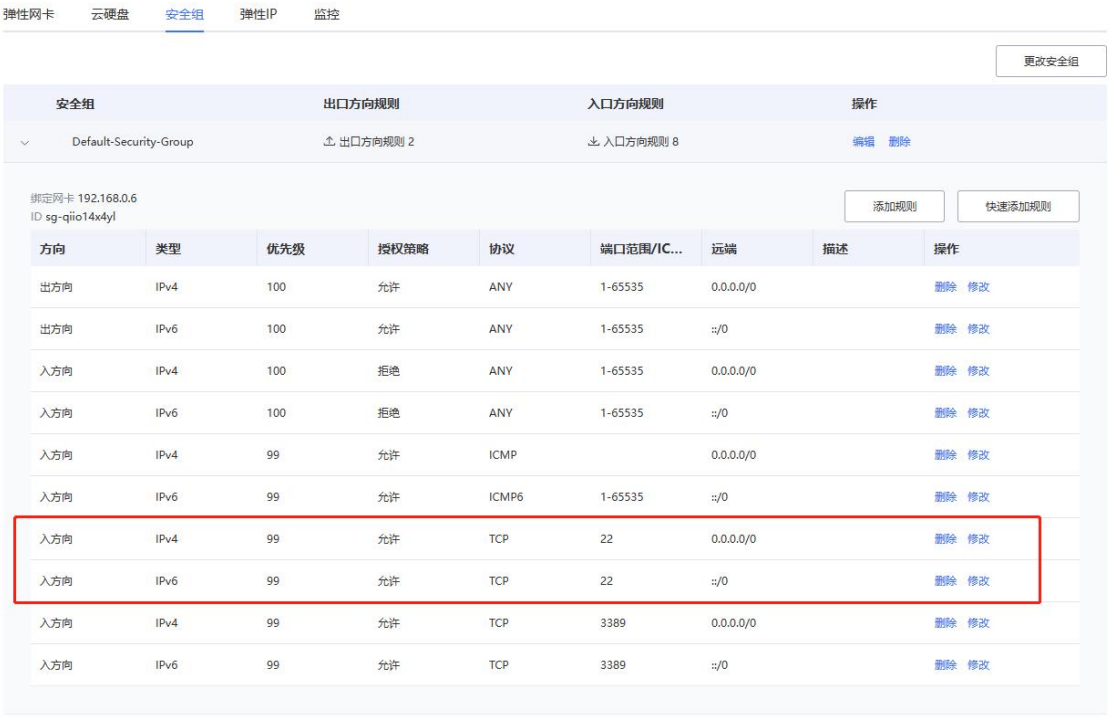


Linux 操作系统

此部分以 CentOS 7.8 操作系统的云主机为示例，对云主机无法远程连接问题的定位步骤。

步骤 1 ：确认安全组已经放行 22 端口。

1. 登录控制中心，选择云主机所在区域，点击“弹性云主机”进入云主机控制台。
2. 在云主机列表，点击需要远程连接的云主机实例名称。
3. 点击“安全组”页签，查看安全组规则。
4. 检查云主机所在的安全组是否已添加 22 入方向的安全组规则。

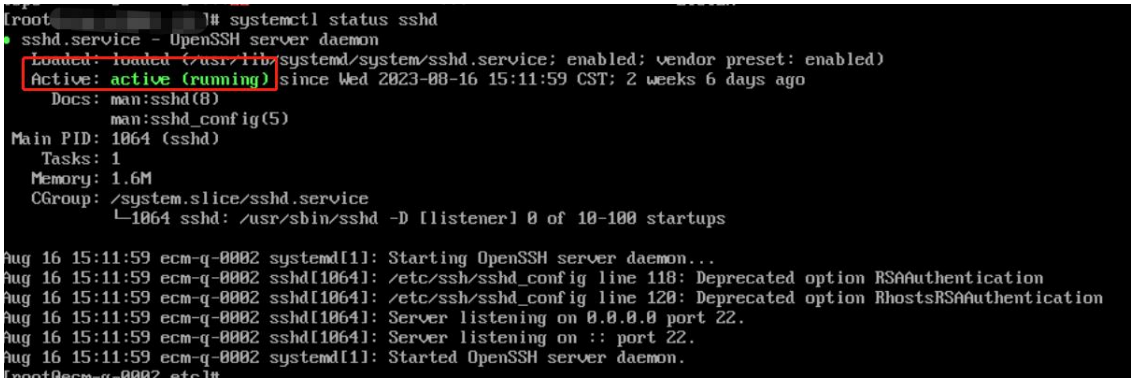


步骤 2 ：可通过控制台登录云主机操作系统，确认远程连接服务已经开启。

执行如下命令，检查 sshd 服务是否已经开启。

```
systemctl status sshd
```

回显信息如下所示，则说明 sshd 已经启动。



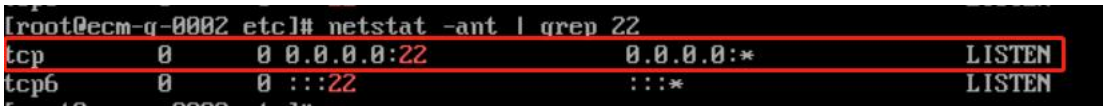
如果未开启，则执行以下命令启动 sshd 服务。

```
systemctl start sshd
```

步骤 3 ：执行如下命令，查看 22 端口在实例中是否正常被监听。

```
netstat -ant | grep 22
```

回显信息则说明 22 端口正常全网监听。如果返回的不是上述结果，请开启监听。



步骤 4 ：查看防火墙 iptables 规则列表，排查防火墙是否已经放行服务。

1. 执行以下命令，查看防火墙状态。

```
systemctl status firewalld 或 firewall-cmd --state
```

2. 根据防火墙状态，如果防火墙处于关闭状态，且您不需要使用防火墙，则无需再做其他处理。如果您需要使用防火墙，且防火墙处于关闭状态，可以执行以下命令开启。

```
systemctl start firewalld
```

- 如果开启命令执行后提示“Failed to start firewalld.service: Unit is masked.”请执行以下命令后再重新执行开启防火墙的命令。

```
systemctl unmask firewalld
```

3. 查看打开的端口。（如果没有已打开的端口则无返回结果）

```
firewall-cmd --zone=public --list-ports
```

4. 放行 TCP 22 端口。

```
firewall-cmd --zone=public --add-port=22/tcp --permanent
```

5. 更新防火墙规则。再次执行查看打开的端口命令，将会返回已打开的接口。

```
firewall-cmd --reload
```



9.1.6 云主机网络延迟和丢包，如何定位

问题概述

当您的云主机出现网络延迟和丢包的问题，可能出现以下三种情况。

- 云主机被入侵：云主机如果发生被黑客入侵或者中毒情况，那么可能病毒程序可能会频繁发送数据包造成带宽超出上限，从而导致丢包现象，严重的情况下可能会使远程登录服务无法使用。
- 云主机本地网络故障：系统中可能存在业务进程传输 io 过大导致网络丢包，比如云主机在大量传输文件等。
- 云主机应用运行异常：应用运行异常会导致网络丢包，如某个应用占用很高的 CPU、内存，造成云主机资源不足。

解决方案

云主机被入侵的排查方式：

可以通过 netstat -anpt 命令查看系统是否存在异常连接。

可以通过查询图中 Foreign Address 字段中的 IP 地址归属判断云主机是否存异常连接，如果 IP 地址归属为国外地址或者非业务访问连接地址，那么云主机有可能被攻击或中毒。这种情况下建议在云主机中安装专业杀毒软件或者重装云主机系统，并对云主机做相关基于业务场景的安全加固，确保不会再次发生中毒现象。



本地网络故障的排查方式

以 CentOS 7.6 为例，执行以下命令，安装 iotop 工具。

```
yum install -y iotop
```

```
Total DISK READ :       0.00 B/s | Total DISK WRITE :       0.00 B/s
Actual DISK READ:       0.00 B/s | Actual DISK WRITE:       0.00 B/s
  TID  PRIO  USER     DISK READ  DISK WRITE  SWAPIN     IO>    COMMAND
 1948 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.01 % [kworker/1:0]
    1 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % systemd --switched-root --system --deserialize 22
    2 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [kthreadd]
    4 be/0 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [kworker/0:0H]
    5 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [kworker/u256:0]
    6 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [ksoftirqd/0]
    7 rt/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [migration/0]
    8 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcu_bh]
    9 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcu_sched]
   10 be/0 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [lru-add-drain]
   11 rt/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [watchdog/0]
   12 rt/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [watchdog/1]
```

可通过图中 DISK READ 和 DISK WRITE 字段以及 COMMAND 字段判断某个进程的读写情况。若存在 IO 流量值很大的情况，需要考虑对相应进程进行整改以确保网络稳定性。比如关闭正在传输的进程，或者分批次进行文件传输等。

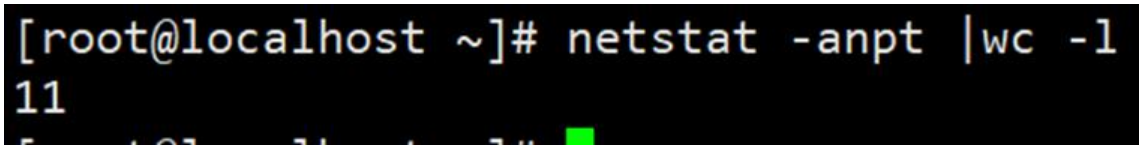
云主机应用运行异常的排查方式

1.可以通过 top 命令查看对应进程对 CPU 资源消耗情况。若%CPU 字段值很大则需要确认，对应的进程运行是否正常。如果是非正常的业务结果，请考虑排查应用程序的 Bug 或中毒现象。

```
top - 05:38:35 up  3:08,  2 users,  load average: 0.00, 0.01, 0.05
Tasks: 120 total,   2 running, 118 sleeping,   0 stopped,   0 zombie
%Cpu(s):  0.0 us,  0.0 sy,  0.0 ni,100.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
KiB Mem :  1863100 total,  1189796 free,   179368 used,   493936 buff/cache
KiB Swap:  2097148 total,  2097148 free,        0 used.  1517760 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
  782 root      20   0  305180   6380   5004 S   0.9  0.3   0:09.53 vmtoolsd
    1 root      20   0  127992   6684   4176 S   0.0  0.4   0:01.61 systemd
    2 root      20   0       0      0      0 S   0.0  0.0   0:00.01 kthreadd
    4 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 kworker/0:0H
    5 root      20   0       0      0      0 S   0.0  0.0   0:00.29 kworker/u256:0
    6 root      20   0       0      0      0 S   0.0  0.0   0:00.05 ksoftirqd/0
    7 root      rt   0       0      0      0 S   0.0  0.0   0:00.03 migration/0
    8 root      20   0       0      0      0 S   0.0  0.0   0:00.00 rcu_bh
    9 root      20   0       0      0      0 S   0.0  0.0   0:00.57 rcu_sched
   10 root       0 -20       0      0      0 S   0.0  0.0   0:00.00 lru-add-drain
   11 root      rt   0       0      0      0 S   0.0  0.0   0:00.07 watchdog/0
   12 root      rt   0       0      0      0 S   0.0  0.0   0:00.05 watchdog/1
```

2.可以通过 netstat -anpt|wc -l 命令查看系统进程链接数数量。若存在系统进程链接数量很多的情况下，需要查看具体是哪个进程引发的问题，进而判断是否存在业务链接负载高或中毒现象。



9.1.7 弹性云主机启动缓慢

问题现象

弹性云服务器启动时间较长。

处理方法

可以通过修改默认等待时间，提高启动速度。

配置调整方式

1.首先登录弹性云服务器。

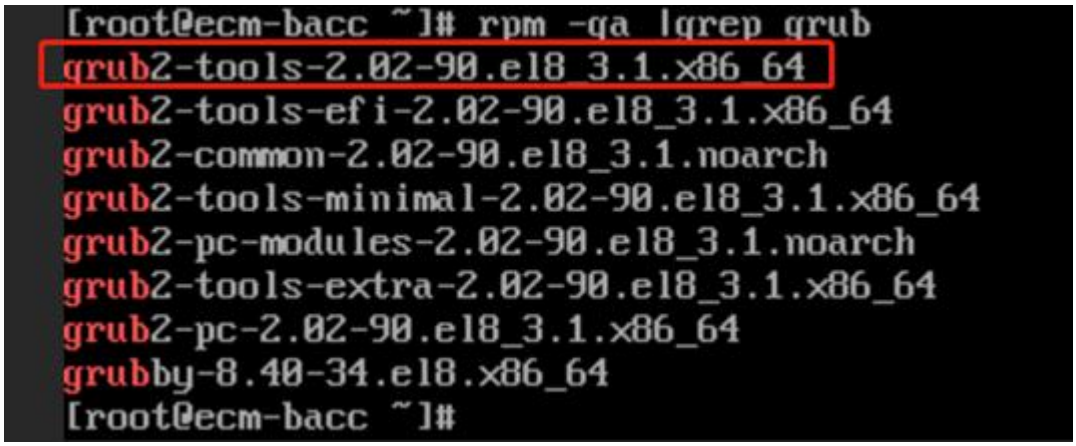
2.执行以下命令，切换至 root 用户。

```
sudo su -
```

3.执行以下命令，查询 grub 文件的版本。

```
rpm -qa | grep grub
```

可以看到查询到的 grub 版本如图。



4.将 grub 文件中 timeout 时间修改为 0s。

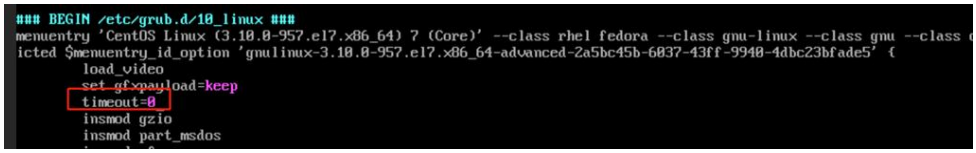
（1）对于 grub 版本小于 2 的：

打开文件“/boot/grub/grub.cfg”或“/boot/grub/menu.lst”，并修改等待时间“timeout=0”（打开方式可通过 vi 或者 vim）。

（2）对于 grub 版本为 2 的：

打开文件/boot/grub2/grub.cfg，并修改等待时间“timeout=0”（打开方式可通过 vi 或者 vim）。

修改 timeou 的值为 0，添加内容如图中。



8.1.7ping 不通或丢包时如何进行链路测试？

问题描述

云主机对外访问网络资源时出现网络卡顿的情况。通过执行 ping 命令，发现网络存在丢包或时延较高的情况。

原因分析

出现网络访问丢包或时延较高的问题，可能存在多方面原因，比如链路拥塞、链路节点故障、服务器负载高、系统设置问题等。

解决方法

在出现访问丢包或者时延高的情况下，可以优先对云主机自身原因进行排查。

如果云主机运行状态正常（如不存在负载异常等情况），可以使用 Tracert 或 MTR 工具进行进一步诊断。

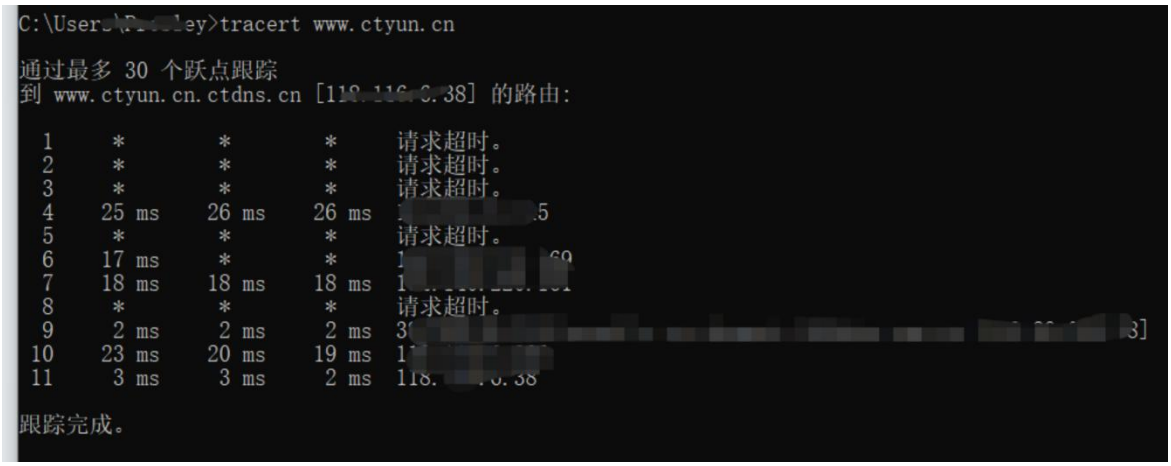
Tracert 和 MTR 工具使用方式

Windows 操作系统使用 Tracert 命令方式

tracert 是路由跟踪命令，用来跟踪数据包到达目的地址过程中的跳转路径，以及每一跳消耗的时间。tracert 命令功能与 ping 命令类似，但可以获得更详细的路径信息，包括数据包所走的全部路径、节点 IP 以及消耗时间。

1.登录 Windows 云主机。

2.打开 cmd 命令窗，执行命令　tracert IP 地址/域名　跟踪路径，例如：tracert www.ctyun.cn。



可以根据回显信息，对数据节点进行分析：

1）tracert 默认最大跳数 30，第 1 列为起跳顺序号。

2）tracert 每次会发送三个数据包，第 2、3、4 列为对应三个数据包的返回时间，如果包丢失则返显为“*” 。

3)第5列为跳转的目标IP节点,如果某一行的此列出现了“请求超时”或“request timed out”，请您根据实际情况对该节点问题进行排查。

Windows操作系统使用WinMTR工具介绍

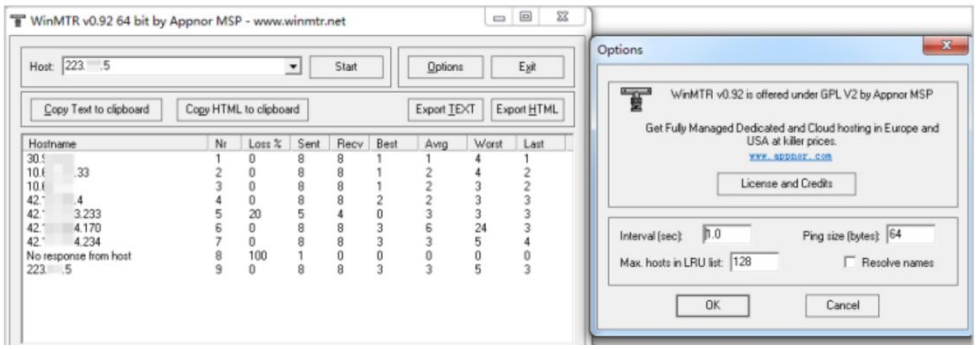
1.登录Windows云主机，WinMTR需要访问公网下载，因此请确认云主机可访问公网。

2.请您通过浏览器寻找正规途径搜索并下载WinMTR软件包。此软件无需安装，解压即可运行。

3.双击WinMTR.exe，运行WinMTR。

4.在WinMTR窗口的Host处，输入目的地址，可以为IP地址或者域名。完成后单击“Start”。

5.根据链路情况不同，运行时间不同，如果一段时间未运行完成可单击“Stop”以结束测试。结果如下图所示。



测试结果的主要信息如下：

1）第一列-Hostname：到目的地址经过的每个节点IP或域名。

2）第二列-Nr：经过的节点数量。

3）第三列-Loss%：对应节点的丢包率。

4）第四列-Sent：已发送的数据包数量。

5）第五列-Recv：已接收到的响应数量。

6）第六列-Best：最短响应时间。

7）第七列-Avrg：平均响应时间。

8）第八列-Worst：最长响应时间。

9）第九列-Last：最近一次响应时间。

Linux 操作系统 MTR 介绍和使用

1.您可以执行以下命令进行安装 MTR，安装前请确保您的云主机可访问公网。如果已安装可跳过此步骤。

1）CentOS 操作系统执行以下命令安装：

```
yum install mtr
```

2）Ubuntu 操作系统执行以下命令安装：

```
sudo apt-get install mtr
```

2.MTR 相关参数说明，请根据您的实际情况选择参数使用：

-h/--help：显示帮助菜单。

-v/--version：显示 MTR 版本信息。

-r/--report：结果以报告形式输出。

-p/--split：与 --report 相对，分别列出每次追踪的结果。

-c/--report-cycles：指定每次探测发送的数据包数量，默认值是 10。

-s/--psize：设置数据包的大小。

-n/--no-dns：不对 IP 地址做域名解析。

-a/--address：用户设置发送数据包的 IP 地址，主要用户单一主机多个 IP 地址的场景。

-4：IPv4。

-6：IPv6。

3.以本机到目的地址 www.ctyun.cn 为例，执行以下命令，以报告形式输出 MTR 的诊断结果。

```
mtr www.ctyun.cn --report
```

回显信息如下：

```
[root@localhost ~]# mtr www.ctyun.cn --report
Start: Wed Aug 30 02:31:11 2023
HOST: localhost.localdomain       Loss%   Snt   Last   Avg  Best  Wrst StDev
  1.|-- gateway                    0.0%    10    0.2   0.4   0.1   1.6   0.3
  2.|-- ???                      100.0     10    0.0   0.0   0.0   0.0   0.0
  3.|-- ???                      100.0     10    0.0   0.0   0.0   0.0   0.0
  4.|-- ???                      100.0     10    0.0   0.0   0.0   0.0   0.0
  5.|-- 10.132.24.145              0.0%    10   19.5  18.9  18.0  19.9   0.3
  6.|-- 182.140.214.249            0.0%    10   14.2   5.3   3.1  14.2   3.1
  7.|-- 182.140.128.173            0.0%    10   18.9  19.1  17.8  20.6   0.6
  8.|-- 139.206.135.161           40.0%    10    3.4   3.2   2.2   4.9   0.9
  9.|-- ???                      100.0     10    0.0   0.0   0.0   0.0   0.0
 10.|-- 38.200.89.218.broad.ls.sc  0.0%    10    3.6   3.7   2.6   4.8   0.3
 11.|-- 118.116.6.190              0.0%    10    5.7   5.9   4.6   9.8   1.4
 12.|-- 118.116.6.42               0.0%    10    2.5   3.8   2.4   5.9   0.9
```

主要输出的信息如下：

1）第一列-HOST：到目的地址经过的每个节点 IP 或域名。

2）第二列-Loss%：对应节点的丢包率。

3）第三列-Snt：已发送的数据包数量。

4）第四列-Last：最近一次响应时间。

5）第五列-Avg：平均响应时间。

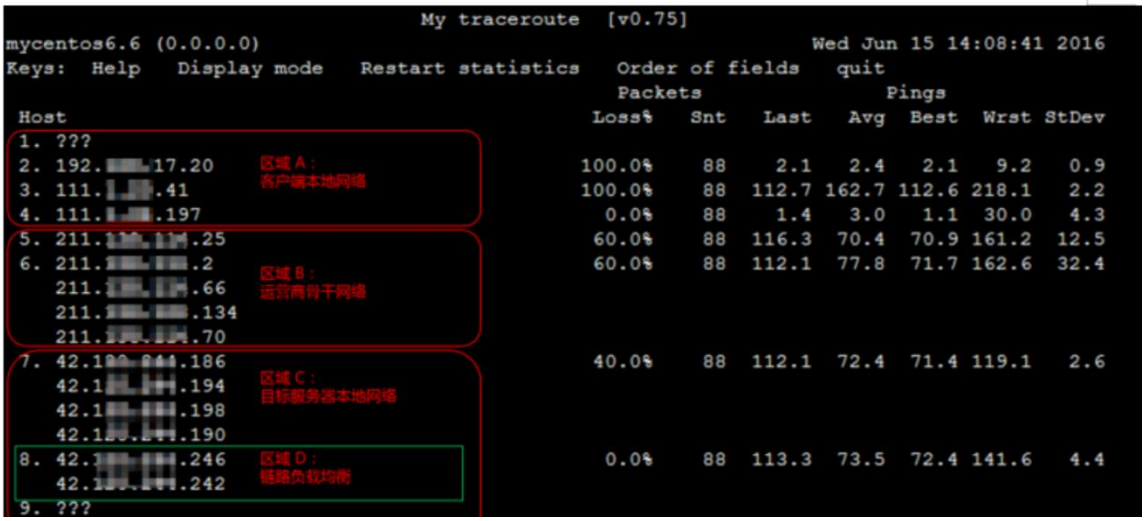
6）第六列-Best：最短响应时间。

7）第七列-Wrst：最长响应时间。

8）第八列-StDev：标准偏差。偏差值越高，说明各个数据包在该节点的响应时间相差越大。

WinMTR 和 MTR 的报告分析处理

以下图为例，对 MTR 报告进行分析。



1.客户端本地网络：即图中 A 区域，代表本地局域网和本地网络提供商网络。

1）如果客户端本地网络中的节点出现异常，则需要对本地网络进行相应的排查分析。

2）如果本地网络提供商网络出现异常，则需要向当地运营商反馈问题。

2.运营商骨干网络：即图中 B 区域，如果该区域出现异常，可以根据异常节点的 IP 查询其所属的运营商，向对应运营商进行反馈。

3.目标服务器本地网络：即图中 C 区域，即目标服务器所属提供商的网络。

1）如果丢包发生在目标服务器，则可能是目标服务器的网络配置原因，请检查目的服务器的防火墙配置。

2）如果丢包发生在接近目标服务器的几跳，则可能是目标服务器所属提供商的网络问题。

4.链路负载均衡：即图中的 D 区域，如果中间链路某些部分启用了链路负载均衡，则 mtr 命令只会对首尾节点进行编号和探测统计，中间节点只会显示相应的 IP 或域名信息。

常见的链路异常案例

1.目标主机配置不当。

如下示例所示，数据包在第 5 条的目标地址出现了 80%的丢包，问题可能是目标服务器网络配置原因，需检查目标服务器的访问配置。

```
t@mycentos6 ~]# mtr -no-dns www.[illegible]comMy traceroute  [v0.75]mycentos6.6 (0.0.0.0)
Wed Jun 15 19:06:29 2016Keys:  Help   Display mode   Restart statistics   Order of fields   quit
Packets               Pings Host                Loss%   Snt   Last   Avg  Best  Wrst StDev
1. ???
2. ???
3. 111.[illegible]                              0.0%     10  521.3  90.1   2.7 521.3 211.3
4. 111.[illegible]9                             0.0%     10    2.9   4.7   1.6  10.6   3.9
5. 211.[illegible].29                          80.0%     10    3.0   3.0   3.0   3.0   0.0
```

2. ICMP 限速。

如下图所示，在第 6、7 跳出现丢包，但后续节点均未见异常。推断是该节点 ICMP 限速所致。该场景对最终客户端到目标服务器的数据传输不会有影响，分析时可以忽略此种场景。

```
Start: 2018-11-29T14:52:30+0800
HOST: laptop                      Loss%   Snt   Last   Avg  Best  Wrst StDev
  1.|-- 30.[illegible]54           0.0%    10    1.9   2.1   1.6   3.9   0.7
  2.|-- 30.[illegible]             0.0%    10    1.5   2.0   1.3   3.2   0.7
  3.|-- 30.[illegible]3            0.0%    10    3.5   4.7   2.7  10.8   2.6
  4.|-- 30.[illegible]             0.0%    10    4.5   3.3   2.3   4.6   0.9
  5.|-- 30.[illegible]             0.0%    10    2.2   2.4   2.0   3.8   0.5
  6.|-- 30.[illegible]             0.0%    10    3.3   3.2   2.5   4.6   0.6
  7.|-- ???                       100.0     10    0.0   0.0   0.0   0.0   0.0
  8.|-- ???                       100.0     10    0.0   0.0   0.0   0.0   0.0
  9.|-- 220[illegible]0.89         50.0%    10    8.7   9.9   5.3  12.2   2.9
 10.|-- 202[illegible]129          0.0%    10    7.5   8.6   7.5  11.0   1.1
 11.|-- 202[illegible]134         40.0%    10   10.8   8.9   7.6  10.8   1.4
 12.|-- 202[illegible]29          30.0%    10    7.9   9.4   7.9  11.1   1.0
 13.|-- 202[illegible]218          0.0%    10  153.0 158.1 139.8 188.7  17.2
 14.|-- 202[illegible]54           0.0%    10  139.7 140.6 137.5 143.8   2.2
 15.|-- 218[illegible]101          0.0%    10  174.4 158.3 150.1 187.3  12.4
 16.|-- 89.[illegible].205         0.0%    10  188.6 184.8 168.6 198.7  10.5
```

3. 环路。

如下图所示，数据包在第 5 跳之后出现了循环跳转，导致最终无法到达目标服务器。出现此场景是由于运营商相关节点路由配置异常所致，需联系相应节点归属运营商处理。

```
[root@mycentos6 ~]# mtr -no-dns www.google.com
My traceroute  [v0.75]mycentos6.6 (0.0.0.0)       Wed Jun 15 19:06:29 2016
Keys:  Help   Display mode   Restart statistics   Order of fields   quit
                                 Packets                Pings
Host                          Loss%   Snt   Last   Avg  Best  Wrst StDev
1. 63.[illegible].43           0.0%    10    0.3   0.6   0.3   1.2   0.3
2. 63.[illegible].157          0.0%    10    0.4   1.0   0.4   6.1   1.8
3. 209[illegible]0.213         0.0%    10    0.8   2.7   0.8  19.0   5.7
4. aix[illegible]tl.google.com 0.0%    10    6.7   6.8   6.7   6.9   0.1
5. 72.[illegible].56           0.0%    10    0.0   0.0   0.0   0.0   0.0
6. 72.[illegible].57           0.0%    10    0.0   0.0   0.0   0.0   0.0
7. 72.[illegible].56           0.0%    10    0.0   0.0   0.0   0.0   0.0
8. 72.[illegible].57           0.0%    10    0.0   0.0   0.0   0.0   0.0
```

4. 链路中断。

如下图所示，数据包在第 4 跳之后就无法收到任何反馈。这通常是由于相应节点中断所致。建议结合反向链路测试做进一步确认。该场景需要联系相应节点归属运营商处理。



9.1.8ping 不通或丢包时如何进行链路测试？

问题描述

云主机对外访问网络资源时出现网络卡顿的情况。通过执行 ping 命令，发现网络存在丢包或时延较高的情况。

原因分析

出现网络访问丢包或时延较高的问题，可能存在多方面原因，比如链路拥塞、链路节点故障、服务器负载高、系统设置问题等。

解决方法

在出现访问丢包或者时延高的情况下，可以优先对云主机自身原因进行排查。如果云主机运行状态正常（如不存在负载异常等情况），可以使用 Tracert 或 MTR 工具进行进一步诊断。

Tracert 和 MTR 工具使用方式

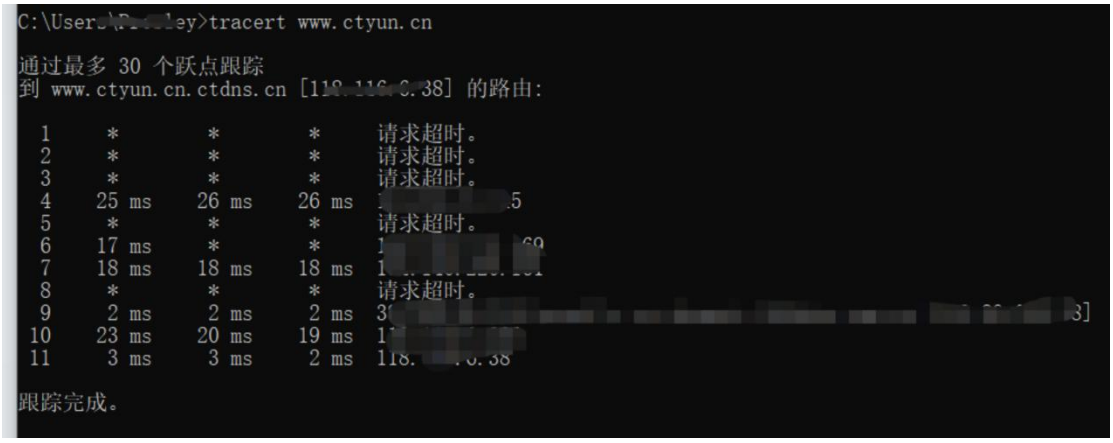
Windows 操作系统使用 Tracert 命令方式

tracert 是路由跟踪命令，用来跟踪数据包到达目的地址过程中的跳转路径，以及每一跳消耗的时间。tracert 命令功能与 ping 命令类似，但可以获得更详细的路径信息，包括数据包所走的全部路径、节点 IP 以及消耗时间。

1. 登录 Windows 云主机。

2. 打开 cmd 命令窗，执行命令 tracert IP 地址/域名 跟踪路径，例如：tracert

www.ctyun.cn。



可以根据回显信息，对数据节点进行分析：

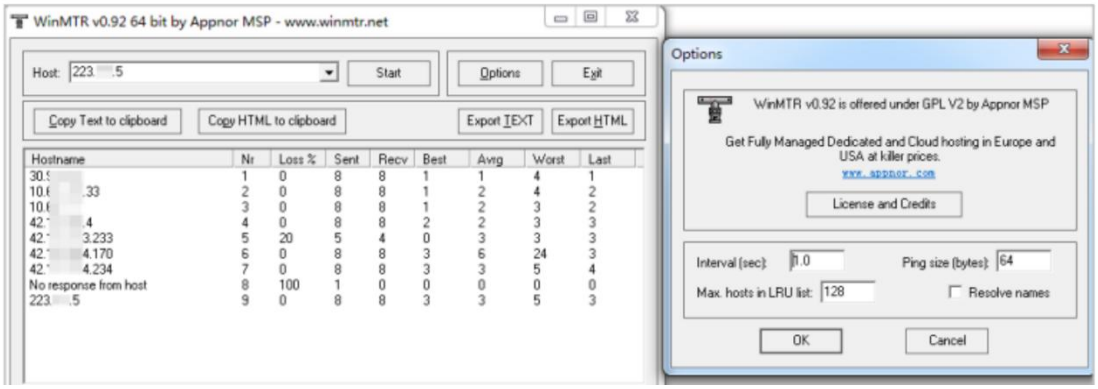
1）tracert 默认最大跳数 30，第 1 列为起跳顺序号。

2）tracert 每次会发送三个数据包，第 2、3、4 列为对应三个数据包的返回时间，如果包丢失则返显为“*” 。

3）第 5 列为跳转的目标 IP 节点，如果某一行的此列出现了“请求超时”或“request timed out”，请您根据实际情况对该节点问题进行排查。

Windows 操作系统使用 WinMTR 工具介绍

1. 登录 Windows 云主机，WinMTR 需要访问公网下载，因此请确认云主机可访问公网。
2. 请您通过浏览器寻找正规途径搜索并下载 WinMTR 软件包。此软件无需安装，解压即可运行。
3. 双击 WinMTR.exe，运行 WinMTR。
4. 在 WinMTR 窗口的 Host 处，输入目的地址，可以为 IP 地址或者域名。完成后单击“Start”。
5. 根据链路情况不同，运行时间不同，如果一段时间未运行完成可单击“Stop”以结束测试。结果如下图所示。



测试结果的主要信息如下：

1）第一列-Hostname：到目的地址经过的每个节点 IP 或域名。

2）第二列-Nr：经过的节点数量。

3）第三列-Loss%：对应节点的丢包率。

4）第四列-Sent：已发送的数据包数量。

5）第五列-Recv：已接收到的响应数量。

6）第六列-Best：最短响应时间。

7）第七列-Avrg：平均响应时间。

8）第八列-Worst：最长响应时间。

9）第九列-Last：最近一次响应时间。

Linux 操作系统 MTR 介绍和使用

1. 您可以执行以下命令进行安装 MTR，安装前请确保您的云主机可访问公网。如果已安装可跳过此步骤。

1）CentOS 操作系统执行以下命令安装：`yum install mtr`

2）Ubuntu 操作系统执行以下命令安装：

```
sudo apt-get install mtr
```

2. MTR 相关参数说明，请根据您的实际情况选择参数使用：

-h/--help：显示帮助菜单。

-v/--version：显示 MTR 版本信息。

-r/--report：结果以报告形式输出。

-p/--split：与 --report 相对，分别列出每次追踪的结果。

-c/--report-cycles：指定每次探测发送的数据包数量，默认值是 10。

-s/--psize：设置数据包的大小。

-n/--no-dns：不对 IP 地址做域名解析。

-a/--address：用户设置发送数据包的 IP 地址，主要用户单一主机多个 IP 地址的场景。

-4：IPv4。

-6：IPv6。

3. 以本机到目的地址 www.ctyun.cn 为例，执行以下命令，以报告形式输出 MTR 的诊断结果。

```
mtr www.ctyun.cn --report
```

回显信息如下：

```
[root@localhost ~]# mtr www.ctyun.cn --report
Start: Wed Aug 30 02:31:11 2023
HOST: localhost.localdomain       Loss%   Snt   Last   Avg  Best  Wrst StDev
  1.|-- gateway                    0.0%    10    0.2   0.4   0.1   1.6   0.3
  2.|-- ???                      100.0    10    0.0   0.0   0.0   0.0   0.0
  3.|-- ???                      100.0    10    0.0   0.0   0.0   0.0   0.0
  4.|-- ???                      100.0    10    0.0   0.0   0.0   0.0   0.0
  5.|-- 10.132.24.145              0.0%    10   19.5  18.9  18.0  19.9   0.3
  6.|-- 182.140.214.249            0.0%    10   14.2   5.3   3.1  14.2   3.1
  7.|-- 182.140.128.173            0.0%    10   18.9  19.1  17.8  20.6   0.6
  8.|-- 139.206.135.161           40.0%    10    3.4   3.2   2.2   4.9   0.9
  9.|-- ???                      100.0    10    0.0   0.0   0.0   0.0   0.0
 10.|-- 38.200.89.218.broad.ls.sc  0.0%    10    3.6   3.7   2.6   4.8   0.3
 11.|-- 118.116.6.190              0.0%    10    5.7   5.9   4.6   9.8   1.4
 12.|-- 118.116.6.42               0.0%    10    2.5   3.8   2.4   5.9   0.9
```

主要输出的信息如下：

1）第一列-HOST：到目的地址经过的每个节点 IP 或域名。

2）第二列-Loss%：对应节点的丢包率。

3）第三列-Snt：已发送的数据包数量。

4）第四列-Last：最近一次响应时间。

5）第五列-Avg：平均响应时间。

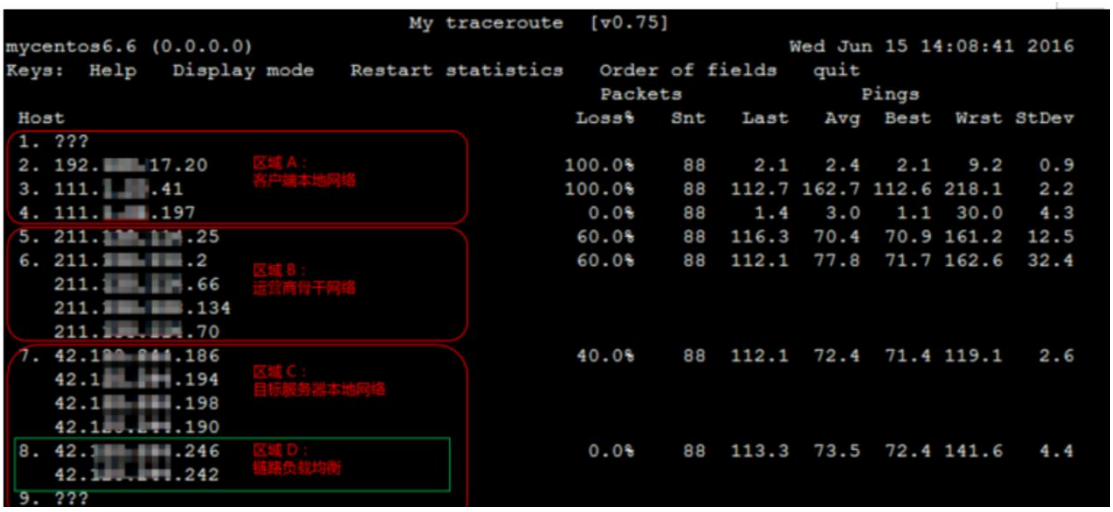
6）第六列-Best：最短响应时间。

7）第七列-Wrst：最长响应时间。

8）第八列-StDev：标准偏差。偏差值越高，说明各个数据包在该节点的响应时间相差越大。

WinMTR 和 MTR 的报告分析处理

以下图为例，对 MTR 报告进行分析。



1. 客户端本地网络：即图中 A 区域，代表本地局域网和本地网络提供商网络。

1）如果客户端本地网络中的节点出现异常，则需要对本地网络进行相应的排查分析。

2）如果本地网络提供商网络出现异常，则需要向当地运营商反馈问题。运营商骨干网络：即图中 B 区域，如果该区域出现异常，可以根据异常节点的 IP 查询其所属的运营商，向对应运营商进行反馈。

2. 目标服务器本地网络：即图中 C 区域，即目标服务器所属提供商的网络。

1）如果丢包发生在目标服务器，则可能是目标服务器的网络配置原因，请检查

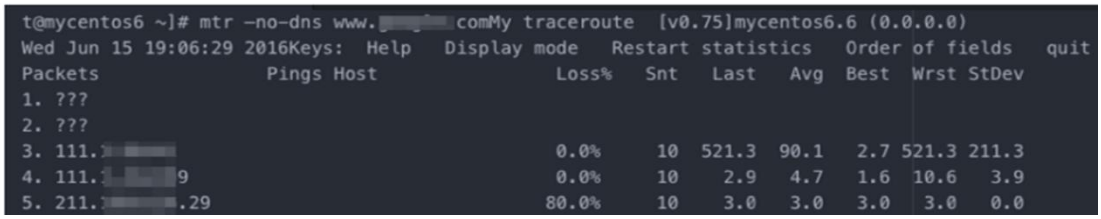
目的服务器的防火墙配置。

2）如果丢包发生在接近目标服务器的几跳，则可能是目标服务器所属提供商的网络问题。链路负载均衡：即图中的 D 区域，如果中间链路某些部分启用了链路负载均衡，则 mtr 命令只会对首尾节点进行编号和探测统计，中间节点只会显示相应的 IP 或域名信息。

常见的链路异常案例

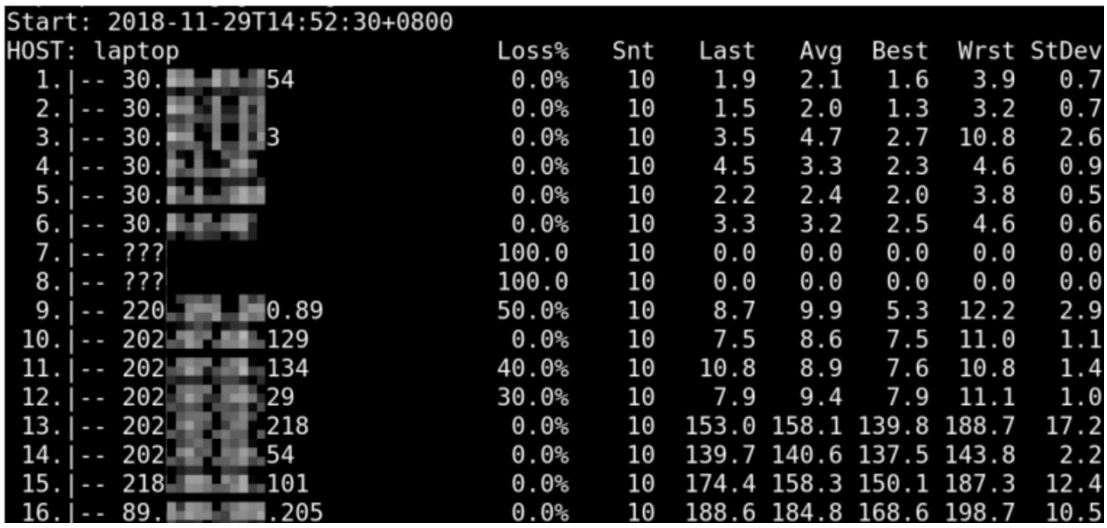
1. 目标主机配置不当。

如下示例所示，数据包在第 5 条的目标地址出现了 80%的丢包，问题可能是目标服务器网络配置原因，需检查目标服务器的访问配置。



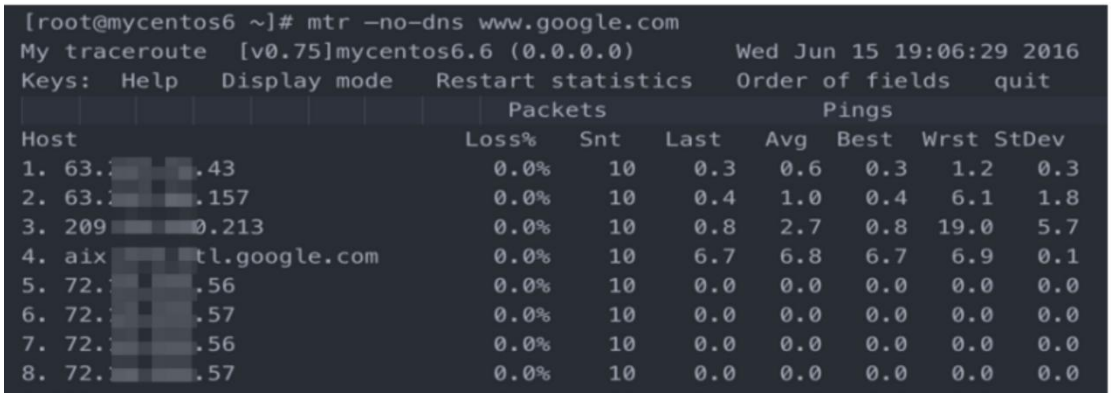
2. ICMP 限速。

如下图所示，在第 6、7 跳出现丢包，但后续节点均未见异常。推断是该节点 ICMP 限速所致。该场景对最终客户端到目标服务器的数据传输不会有影响，分析时可以忽略此种场景。



3. 环路。

如下图所示，数据包在第 5 跳之后出现了循环跳转，导致最终无法到达目标服务器。出现此场景是由于运营商相关节点路由配置异常所致，需联系相应节点归属运营商处理。



4. 链路中断。

如下图所示，数据包在第 4 跳之后就无法收到任何反馈。这通常是由于相应节点中断所致。建议结合反向链路测试做进一步确认。该场景需要联系相应节点归属运营商处理。



9.1.9Linux 系统云主机如何自查病毒及中病毒后处理建议

一，排查思路

1.排查系统运行情况

结合用户反映的故障现象，查看进程运行情况（例如，故障现象为资源占用率高，则排查占用资源较多的进程；故障现象为网络带宽占用较多，则排查网络连接较

多的进程）、网络连接情况。进一步排查可疑进程所关联的文件、启动进程的账户等信息。

2.排查系统运行环境

通过排查系统账户、启动项、计划任务、系统服务启动等情况，排查可疑进程自动拉起的信息，为阻止恶意程序自运行提供依据。

3.排查攻击来源

查看登录日志、运行日志，排查是否存在远程暴力破解行为、是否有不明 IP 登录系统，系统内是否出现不明账户，分析 passwd、shadow 文件完整性和修改时间是否正常，同时对安全事件发生事件进行判定。

4.排查非法入侵系统后的操作

查看历史操作命令，据此对非法入侵系统后的操作进行分析，对非法操作造成的破坏进行针对性地恢复。

5.确定解决方法

根据排查结果，研究解决安全事件的方法。如无法彻底清除恶意程序，则考虑在备份重要数据后重装。

6.系统加固

安全事件处理完毕后，应分析入侵的方式，有针对地对系统进行加固，防止被再次攻击。

二，常见安全入侵排查

(一)挖矿病毒

1.现象

CPU 占用率接近 100%

2.排查思路

确认占用 CPU 资源的进程，分析该进程对应的程序，查找可疑文件，并根据文件内容确认是否为挖矿程序。

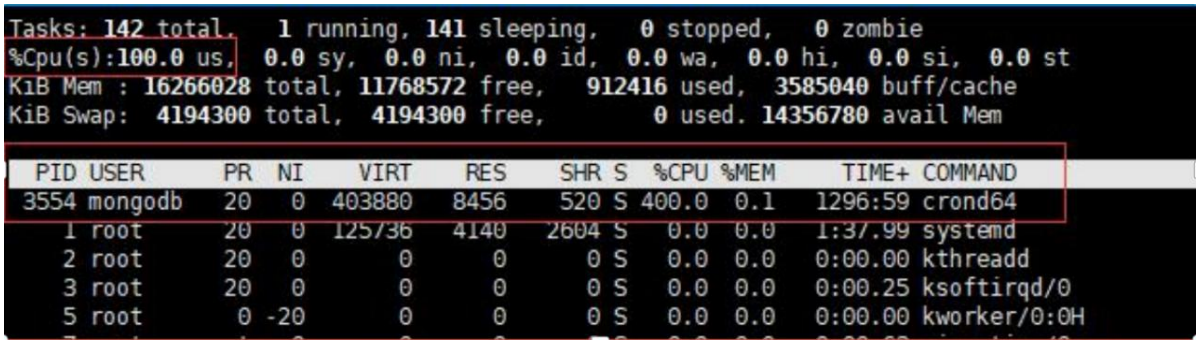
3.排查过程

对于用户反馈云主机性能卡顿，CPU 和内存占用较高的情况：

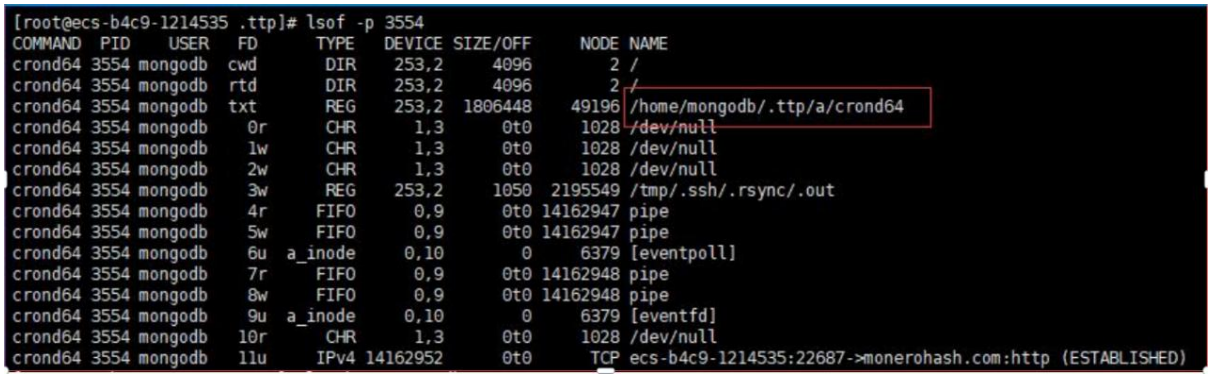
执行 TOP 命令，查看占用 CPU 较高的异常进程，一般多为 80%以上，有个别病毒占用 CPU 可能不高，但是从 command 名称明显不是系统进程或用户进程，如 TSM，Lixsyn，ynlyvtpxia 等

对于用户反馈连接云主机卡顿，但 CPU 和内存占用不高的情况：

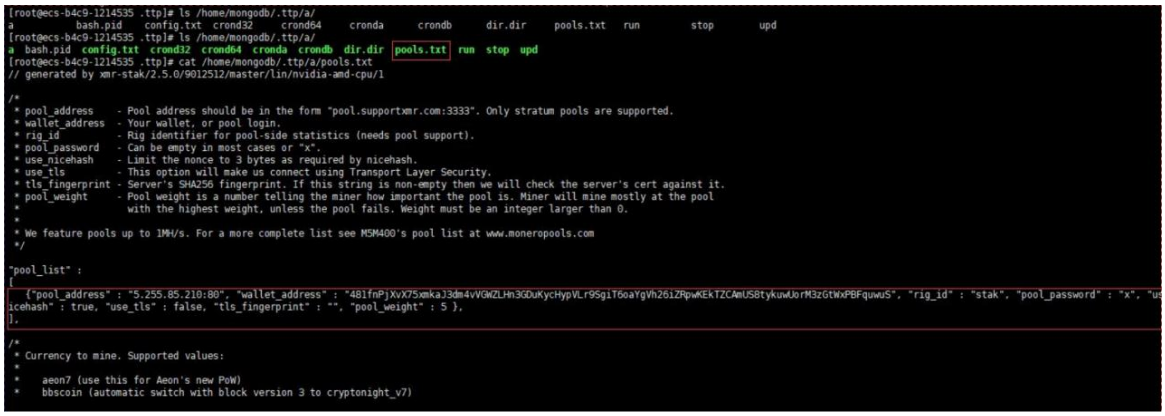
可能是云主机出网带宽跑满导致。此时 ping 云主机公网 IP 会出现大量丢包



通过该进程 PID，查看其关联的程序。



经查看，发现该进程关联隐藏路径，进一步排查该路径下的文件情况。





查看该路径下 pools.txt 文件，发现内有挖矿文件关联的钱包信息，且关联的 IP 地址 5.255.85.210 为荷兰 IP 地址，因此可确认该主机被植入挖矿进程。

解决办法：

1.杀掉占用 CPU 资源的病毒进程。

2.删除恶意程序。

3.停用与恶意程序关联的系统账户并修改系统账户密码，限制该账户远程登录权限。

4.检查定时任务、开机自启动文件等，彻底清理挖矿程序痕迹。

（二）十字符病毒

1.故障现象

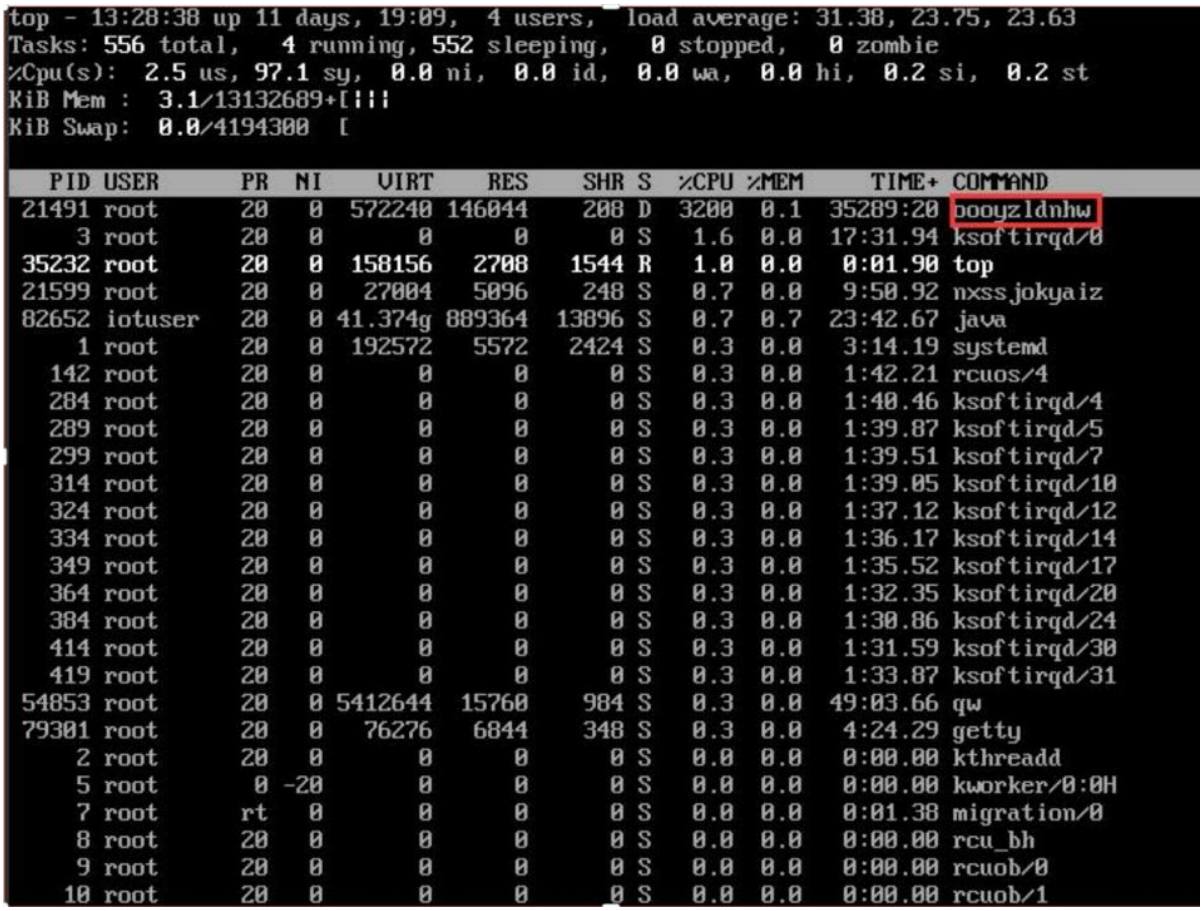
名称由随机 10 字符组成的进程，运行时占用大量 CPU 资源，伴有网络流量较大、占满带宽的现象，无法通过 kill 命令终止该进程。

2.排查思路

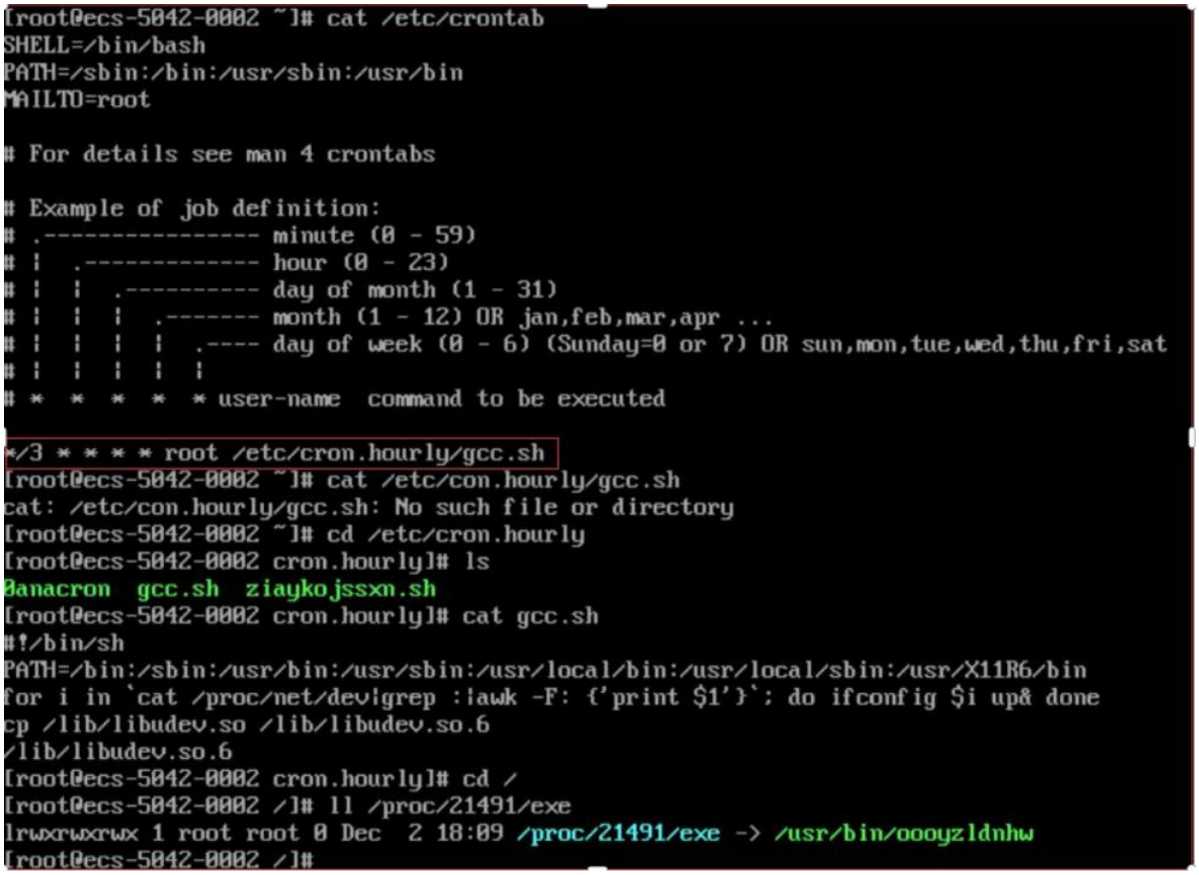
确认占用 CPU 资源的进程，分析该进程对应的程序，同时查看计划任务、启动脚本等，排查是否存在可疑文件，并根据文件内容确认恶意脚本程序。

3.排查过程

查看进程列表，确认占用 CPU 资源的主要进程。



由于病毒通常具有自动运行机制，因此常常藏身于计划任务中。查系统计划任务中是否存在可疑脚本，并对脚本进行进一步分析核查，可确认系统被感染病毒，病毒本体是/lib/libudev.so



4.解决方法

1）暂停病毒进程（强制终止该进程后，病毒程序还会自动拉起一个新的进程运行。

2）删除恶意程序脚本。

3）杀掉病毒进程。

4）清理计划任务内容。

5）删除病毒本体。

6）检查开机自启动项等内容，彻底清理病毒痕迹。

三，系统加固操作建议

针对 Linux 云主机，建议您绑定 Linux 开放 22 端口安全组，移除默认安全组，单独开放需要的端口，使用复杂密码并定期更换，同时定期对您的云主机制作私有镜像，操作方法请参考下列文档：

安全组概述

实例加入/移出安全组

配置安全组规则

修改云主机默认远程端口

通过云主机创建 Linux 系统盘镜像

9.1.10Windows 系统云主机如何自查病毒及中病毒后处理建议

一，排查思路

1.排查系统运行情况

结合用户反映的故障现象，排查相关进程（例如，故障现象为资源占用率高，则排查占用资源较多的进程；故障现象为网络带宽占用较多，则排查网络连接较多的进程）。重点排查进程所关联的文件、启动进程的账户等信息。

2.排查系统运行环境

通过排查系统启动项、计划任务、系统服务启动等情况，排查可疑程序自动拉起的信息，为阻止恶意程序自运行提供依据。

3.排查攻击来源

查看登录记录，排查是否存在远程暴力破解行为、是否有不明 IP 登录系统，系统内是否出现不明账户

二，常见安全入侵排查

(一)挖矿病毒

1.故障现象

CPU 占用率接近 100%。

2. 排查思路

确认占用 CPU 资源的进程，分析该进程对应的程序，查找可疑文件，并根据文件内容确认是否为挖矿程序。

3. 排查过程

在任务管理器中查看进程列表，确认占用 CPU 资源的主要进程。



查找该进程关联的程序。



xmrig 是典型的门罗币挖矿程序，可通过其同目录下的“start.cmd”文件查看挖矿参数。其参数说明如下：-o 矿池地址和端口号；-u 账户地址；-p 密码。

4.解决方法

停用恶意进程、删除恶意程序、检查自启动项目和计划任务、检查系统内可疑账户、为远程登录账户设置强口令、通过安全组对远程访问的 IP 地址进行限制。

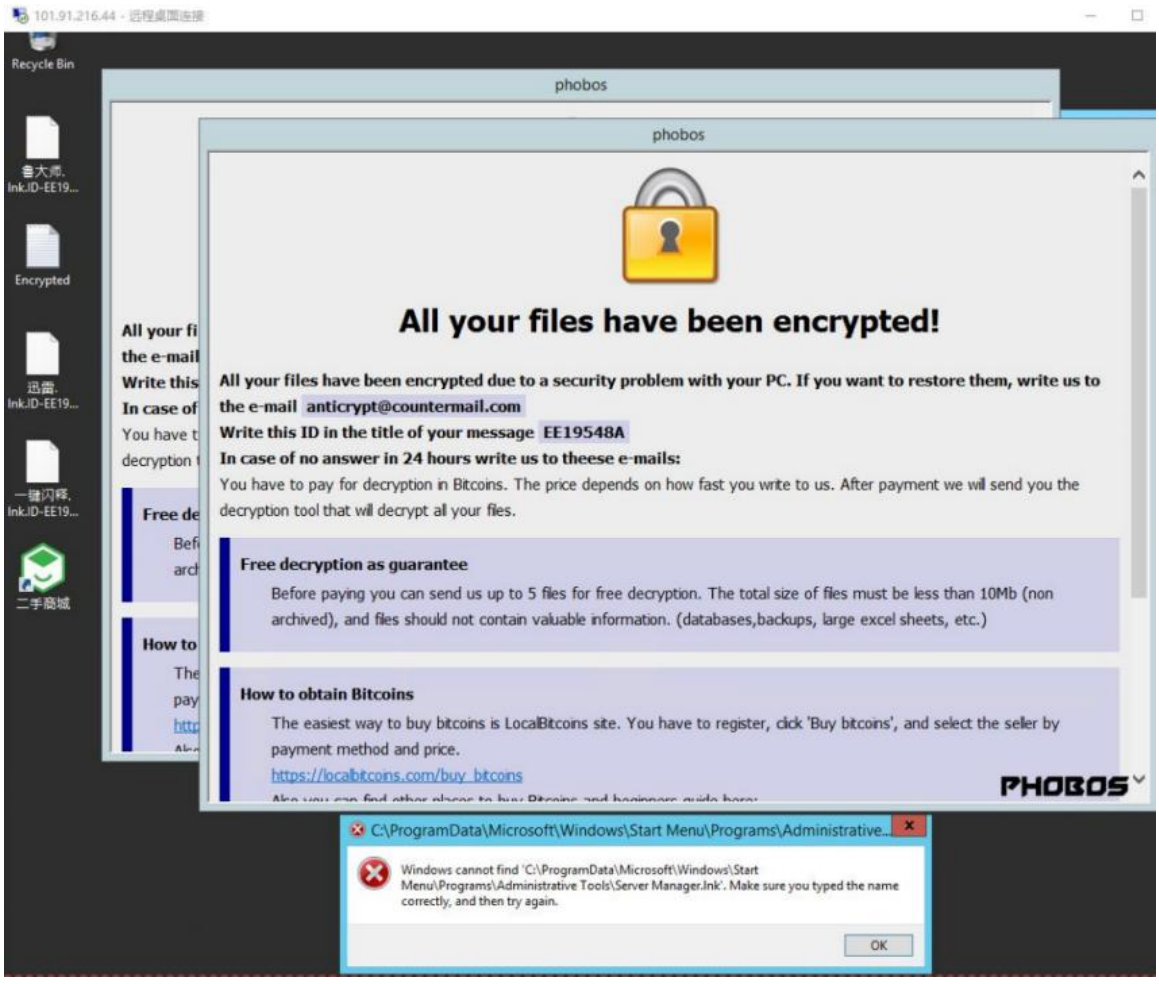
(二)，勒索病毒

1.故障现象

大量文件被加密，无法正常打开。

2.排查过程

登录主机侧，确认故障现象。



3.解决方法

寻求第三方专业解密公司进行数据恢复，或直接重装系统，并进行加固后，再部署业务系统。

(三)系统加固操作建议

1.为账户设置高强度口令

在“控制面板”->“用户账户”中选择需要设置/修改口令的账户，选择“更改密码”，然后在弹出的界面中输入新口令即可。

【密码设置原则】

1）不要使用弱密码、默认密码。

2）不要在多主机、多系统使用同一个密码。

3）不要长期使用固定密码，定期或者不定期修改密码，安全更有保障。

4）最好不要把密码保存在电脑、U 盘、笔记本、书籍等上面，如果保存要采取安全保护措施。

【高强度口令需满足如下要求】

1）密码长度不少于 10 位，且至少包含大写字母、小写字符、数字、特殊字符 4 类中的 3 种；

2）不要使用有一定特征和规律容易被破解的常用口令的密码，如在常用彩虹表中的密码、滚键盘密码、含账户名的密码、含连续数字或字母的密码等。

（如：!QAZxsw2，qazxsw，1qaz@WSX，1q2w3e，123456789，password……）

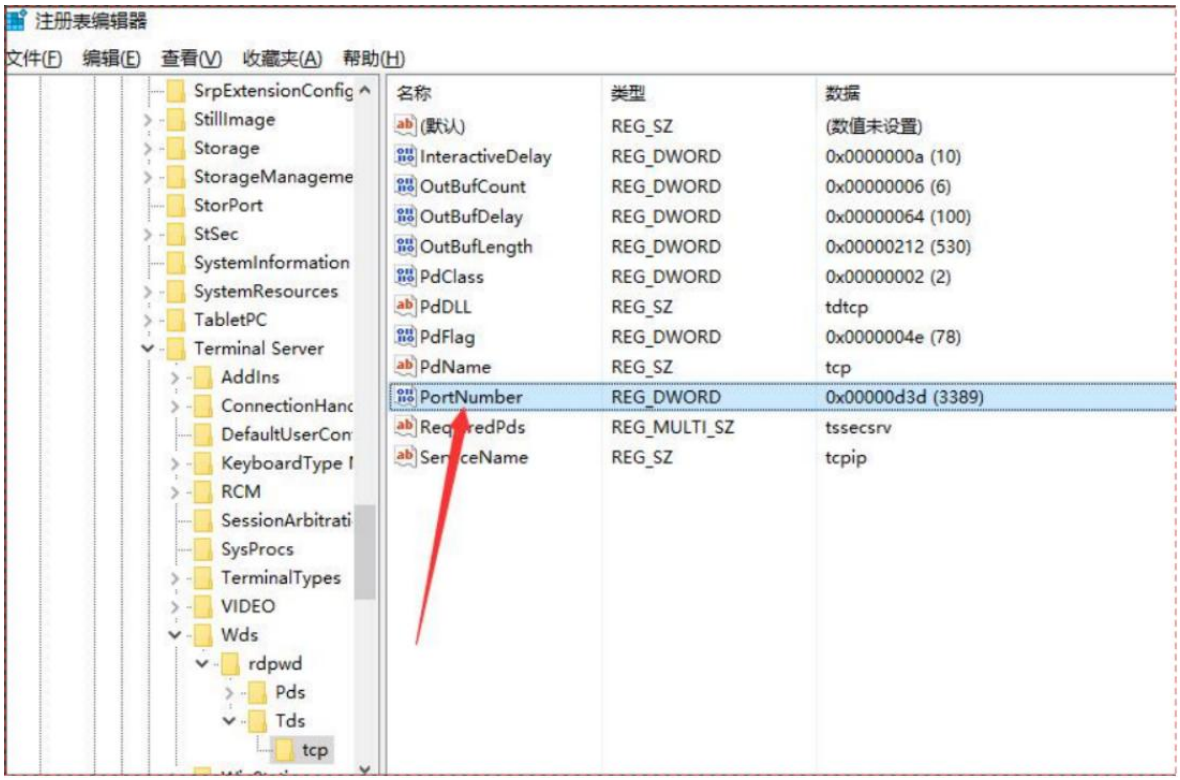
3）避免使用生日、电话号码、身份证号码、QQ、邮箱等与个人、单位信息有明显联系的数据。

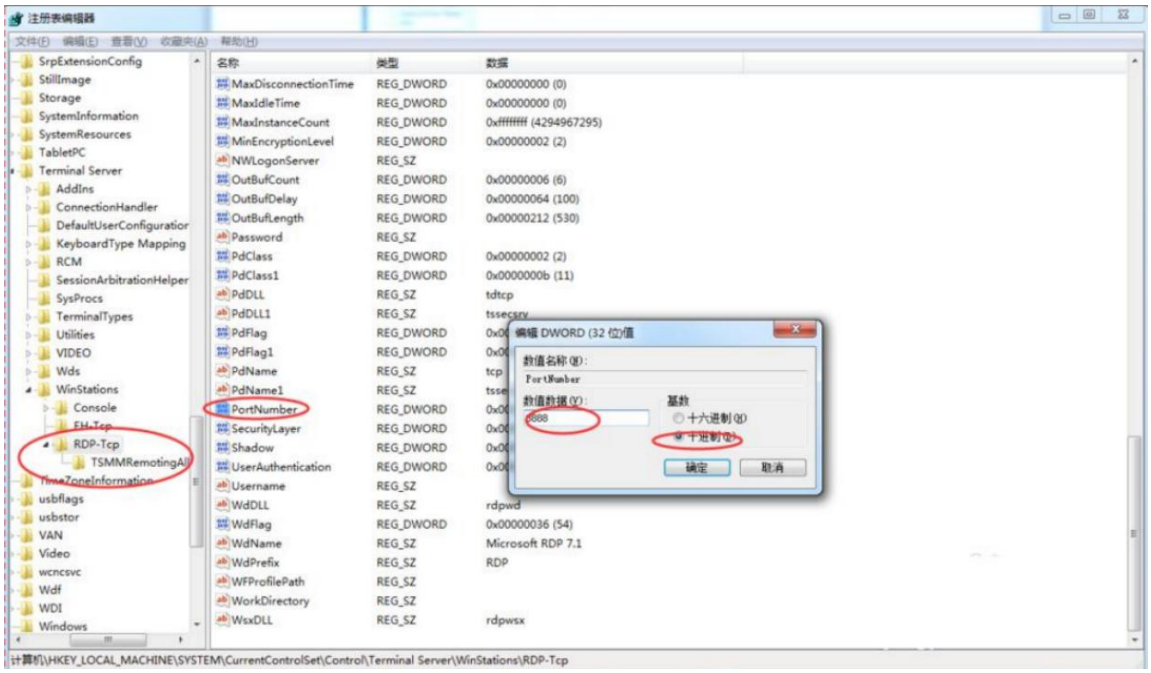
2.修改远程登录端口

在注册表编辑器

“HKEYLOCALMACHINE\SYSTEM\CurrentControlSet\ontrol\TerminalServer\Wds\rdpwd\Tds\tcp\”路径下找到 PortNamber 键，改为要使用的远程端口。建议

修改为 4 位数端口号。（说明：在 win10 系统中 CurrentControlSet 已经变更为 CurrentControlSet001），修改完毕后，重启云主机，配置即可生效。





3. 通过本机防火墙进行外联控制

1）限制允许远程访问本机的 IP 地址范围，在 Windows 系统本机防火墙的“入站

规则”中，找到 RPC 相关条目，在编辑条目的界面中可以添加允许访问的 IP 地址。

2）限制本机开放的端口，在 Windows 系统本机防火墙中，可通过新建规则，设定允许开放的端口范围。

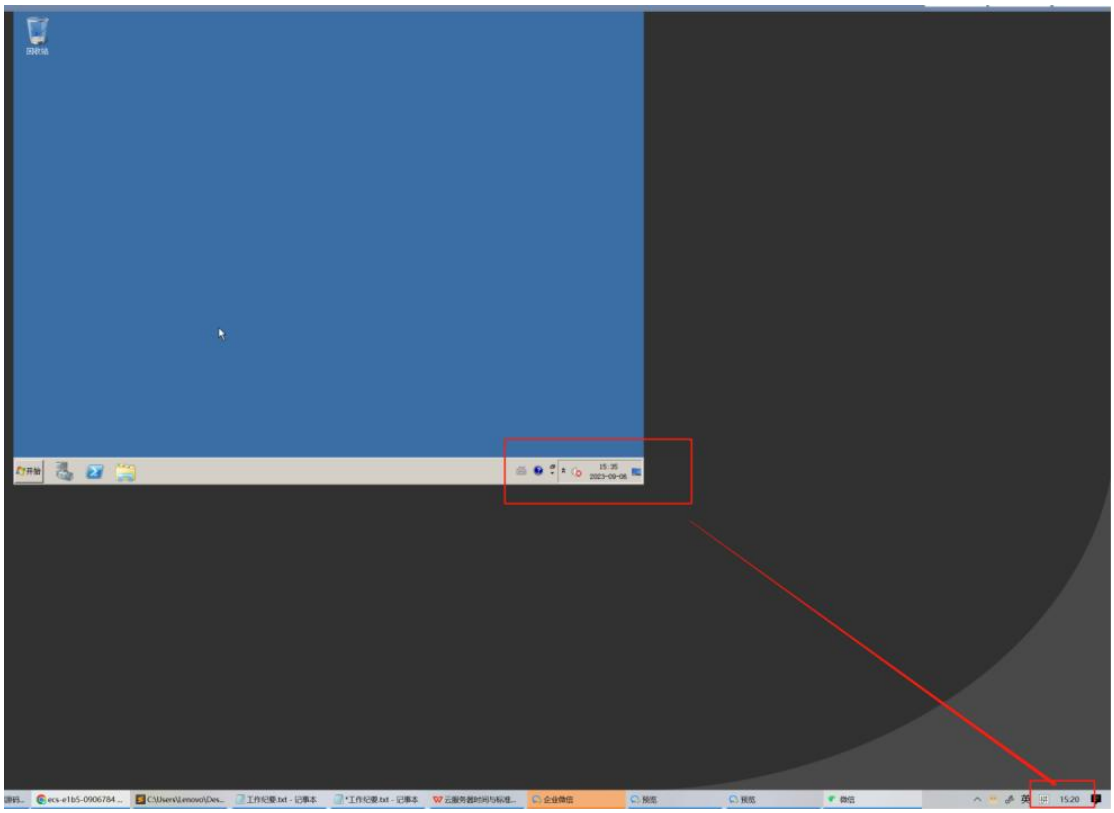
9.2 操作系统类（Windows）

9.2.1 云主机时间与标准时间不一致

问题描述

云主机显示的 Windows 操作系统时间与本地标准时间不一致。

如图操作系统时间与本地标准时间不一致。



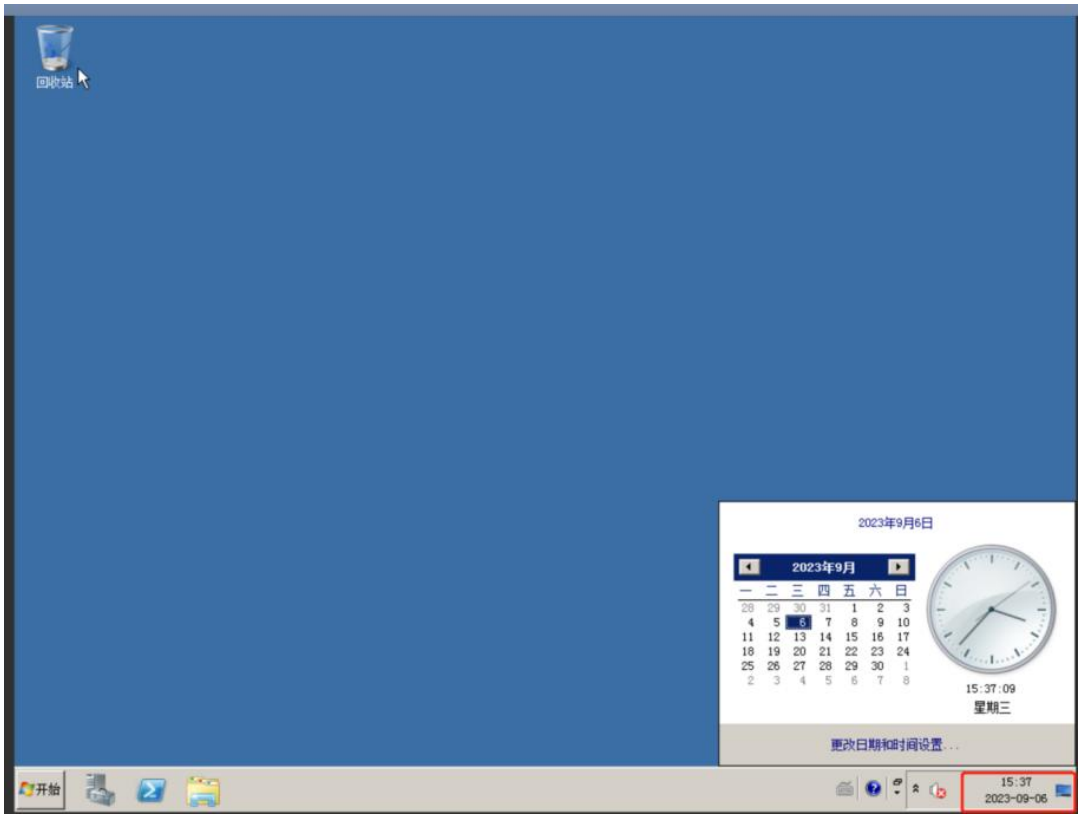
可能原因

因网络问题或某些进程的干扰，系统时间可能会发生偏差，与标准时间不一致。

处理方式（一）

手动同步系统时间

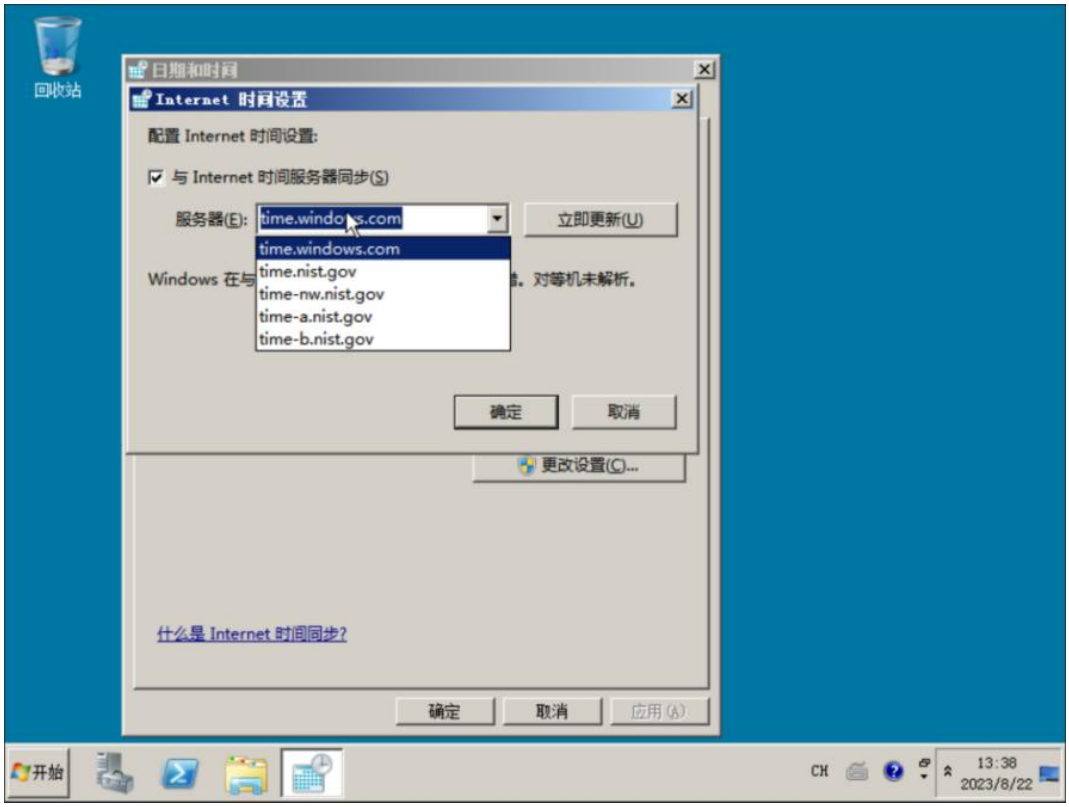
1.单击桌面右下角的“更改日期和时间设置”，打开“日期和时间”窗口。



2.点击“Internet 时间”页签。

3.单击“更改设置”，并选择时间服务器源。默认时间服务器源为:“time.windows.com”。

4.依次点击“立即更新 > 确定”按钮。



1. 对比 Windows 系统时间与本地标准时间是否一致。

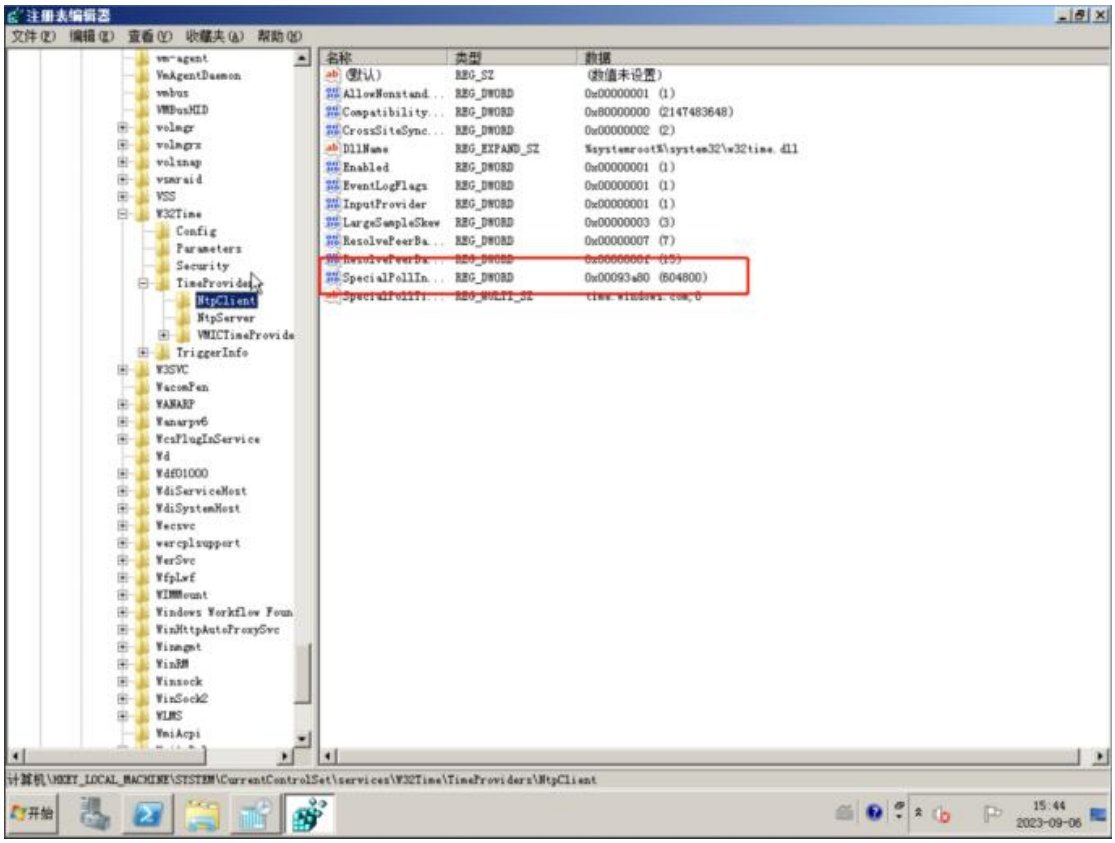
2. 处理方式（二）

您可以通过修改注册表来调整 Windows 系统的时间更新频率，具体步骤如下。

1.快捷键“win+r”打开“运行”对话框，输入“regedit”进入注册表编辑器。

2. 在注册表编辑器中，依次展开以下路径：HKEY_LOCAL_MACHINE > SYSTEM >CurrentControlSet > Services > W32Time > TimeProviders >

NtpClient

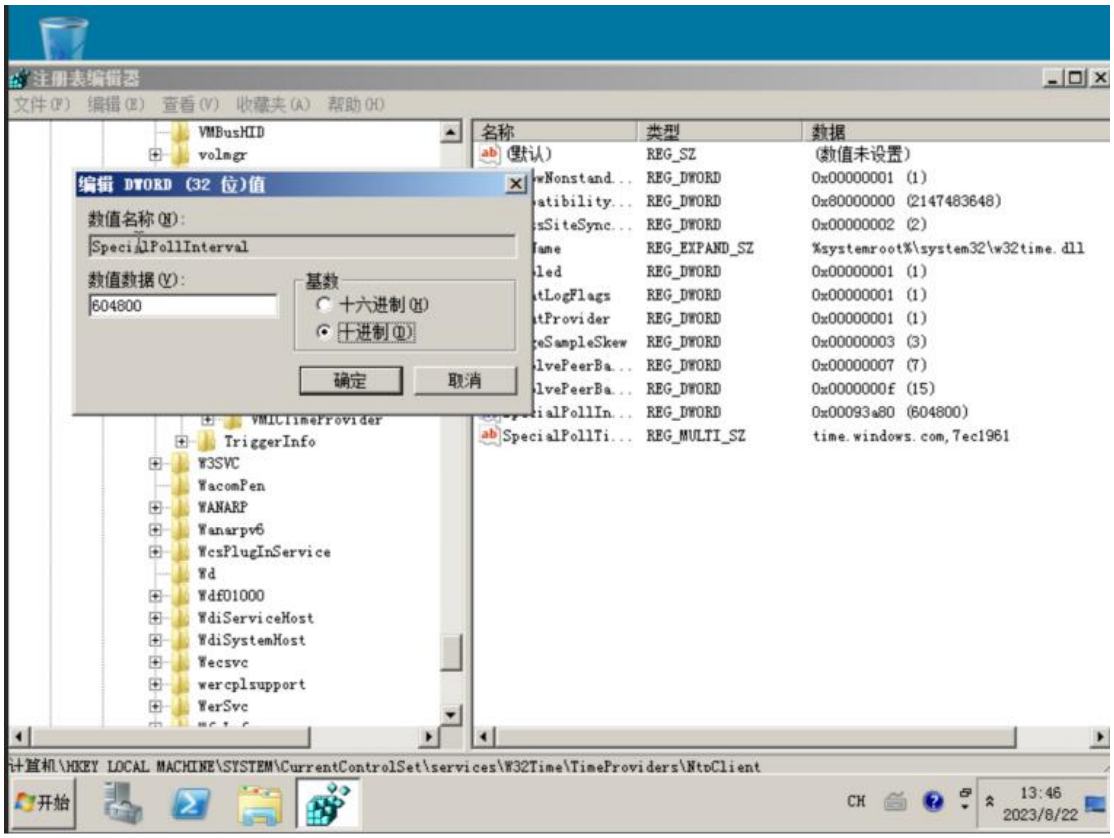


3.在"NtpClient" 下找到"SpecialPollInterval" 键值。

4.设定时间同步周期。

“数值数据”中显示的数字正是自动对时的间隔（以秒为单位），输入您想要的更新频率（单位是秒）。例如，如果您希望系统每隔 3600 秒（1 小时）更新一

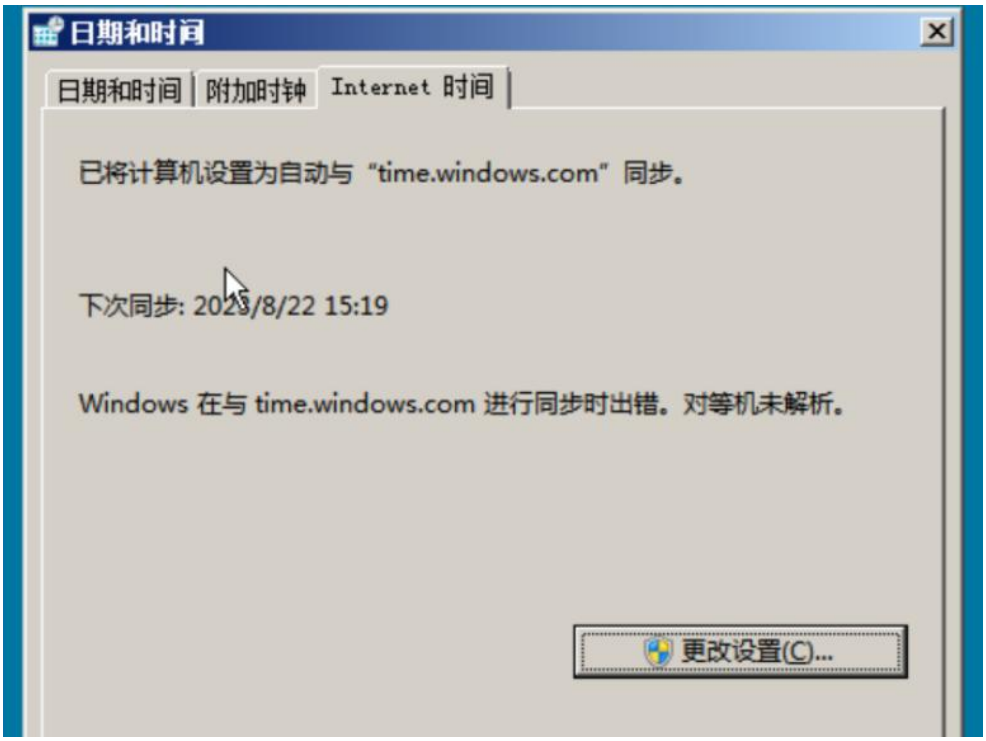
次时间，您可以输入"3600"。



5. 单击“确定”保存关闭对话框。

6. 配置完成后，打开“命令提示符”窗口（CMD），执行以下命令以刷新组策略并使更改生效： Gpupdate 。

7. 在 Internet 时间里查看，如下图所示，可见时间同步频率已经每变成 15 分钟次。



9.2.2windows 操作系统下卸载云硬盘操作指南及故障排除

问题场景

在使用弹性云主机+云硬盘的服务时，您可能需要将磁盘属性为数据盘的云硬盘从当前云主机卸载，以便挂载至另一台弹性云主机继续使用。请注意，卸载云硬盘的操作不会清除硬盘上的数据。

支持范围

适用对象：仅支持数据盘的卸载，系统盘不支持卸载操作。

前提条件与注意事项

1、为保证数据完整性，在卸载前，建议暂停所有针对该磁盘的文件系统读写操作，以防止数据丢失。如果有服务正在运行并使用该云硬盘，请先停止这些服务，确认所有依赖于该云硬盘的服务都已停止后再卸载云硬盘。

2、卸载云硬盘前建议先执行脱机（Windows）操作或者 umount（Linux）操作，否则可能会导致云主机卸载云硬盘失败或者云主机再次挂载云硬盘时无法识别等异常情况。

3、若出现卸载失败等异常情况，可以先检查进程占用，例如使用 fuser -k /mnt/point（Linux）或任务管理器（Windows）查找并终止占用磁盘的进程，然后重新尝试卸载。或者可以先对云主机关机后再重新尝试卸载。

如遇复杂问题，建议联系技术支持获取进一步协助。

9.2.3Windows 弹性云主机蓝屏如何处理

问题描述

Windows 操作系统弹性云主机蓝屏。

可能原因

1.驱动程序冲突导致。
2.使用了来源不明的第三方软件导致。
3.病毒感染，导致系统文件、注册表损坏。
4.CPU 占用过高导致。

处理方法

1.确保服务器上的所有驱动程序都是最新版本，并且与操作系统兼容。可以尝试更新或重新安装相关的驱动程序，特别是与硬件设备相关的驱动程序。
2.建议使用正版软件，不要安装来源不明的第三方软件。
3.如果排查是因为 CPU 占用过高导致的蓝屏，可以参考以下操作：

- 可以通过把一些暂时不使用的进程关掉后再尝试。
- 可以尝试重启服务器。
- 如果重装系统，请先备份重要数据。

- 如果服务器有重要数据不能重装，可以通过挂载磁盘方式拷贝数据，需要先备份，再卸载磁盘，然后挂载盘拷贝数据。

4.如果需要分析蓝屏原因，需要确认是否有配置内存转储文件(crash dump)的收集，系统会自动生成蓝屏 dump 日志到指定的目录。由于蓝屏日志的分析非常耗时，建议客户在遇到蓝屏的情况，重启弹性云主机后，参考如上的可能原因进行排查。一般第三方杀毒软件、系统故障和 CPU 占用过高是导致弹性云主机蓝屏最常见的原因。

9.2.4Windows 弹性云主机配置双网卡外网访问

约束限制

操作过程中，请不要修改主网卡的配置。

操作步骤

1.登录控制中心，并选择“网络-虚拟私有云”。



2.在网络控制台中选择“弹性网卡“中”创建弹性网卡”按钮。



3. 在“创建弹性网卡”页面中，根据界面提示添加弹性网卡。



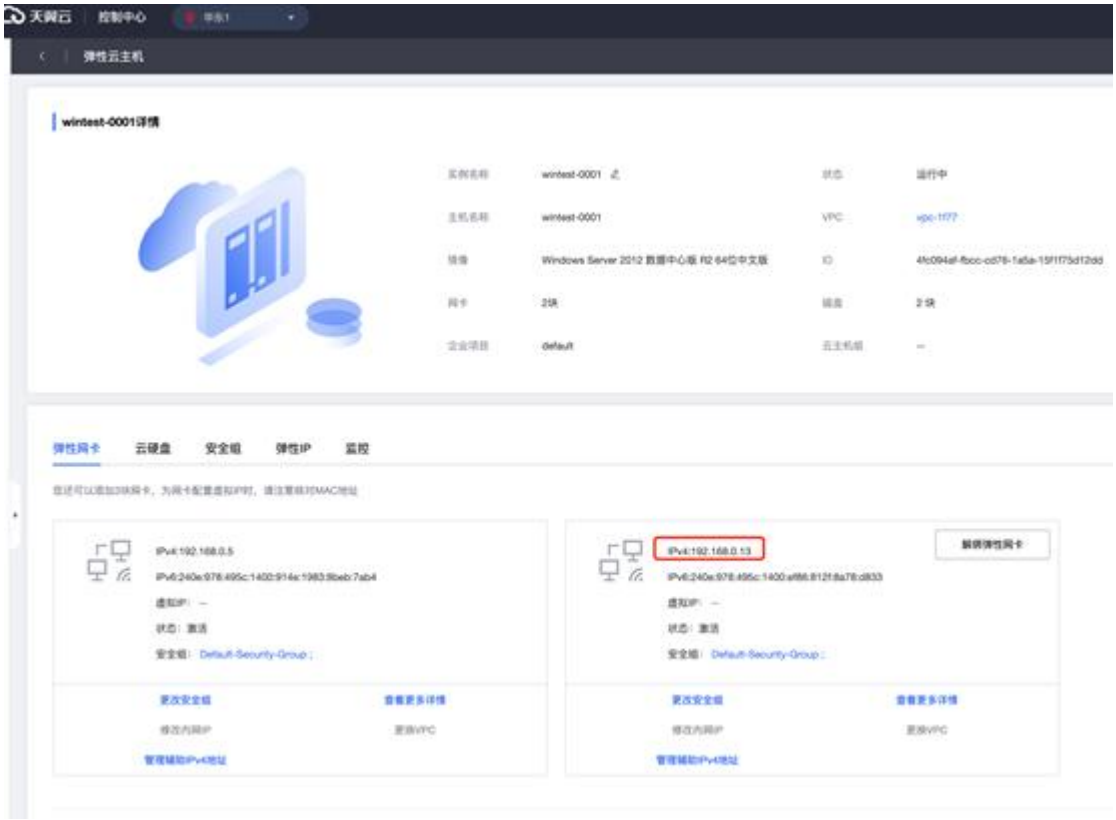


4. 在弹性网卡列表页选中弹性网卡，点击“操作”列中的“绑定实例”按钮，进行弹性网卡绑定。





5. 点击绑定实例，进入弹性云主机详情页面，可以看到网卡信息。



6. 远程登录弹性云主机。

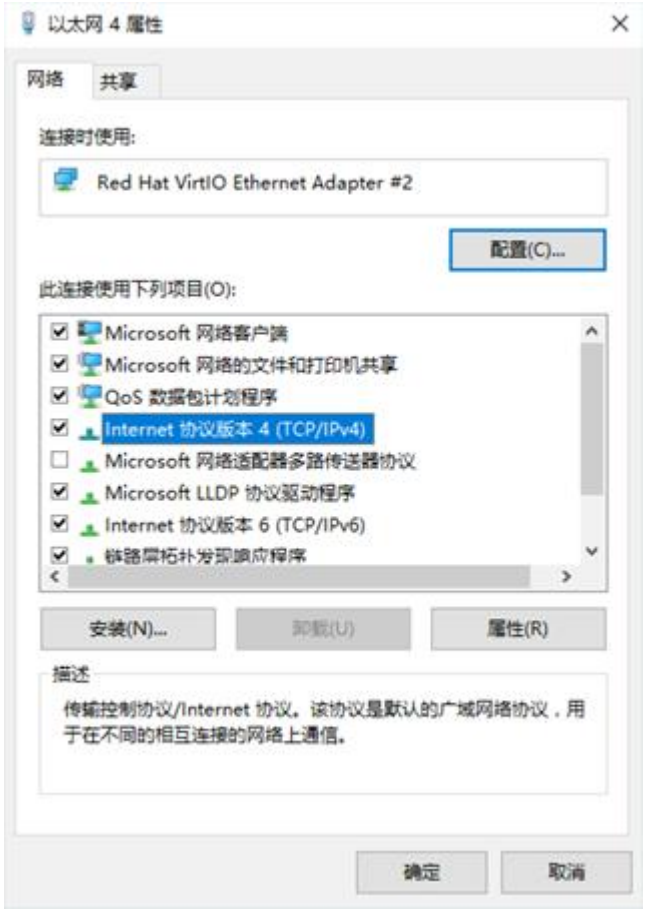
7. 找到右下角网络配置图标，再点击右侧“更改适配器设置”（或者“控制面板”-“网络和 Internet”-“网络连接“），找到新添加的网卡。





8. 选择新添加的网卡，然后右击图标，选择“属性”。

9. 在网络连接属性面板中“网络”选项下，选中“Internet 协议版本 4(TCP/IPv4)”，然后单击下面的“属性”。



10.在属性面板中，切换到常规选项卡下，将步骤 5 中自动获取的私有 IP 地址配置给新添加的网卡。

示例：

IP 地址：192.168.0.13

子网掩码：255.255.255.0

默认网关：192.168.0.1

首选 DNS 服务器：114.114.114.114

备用 DNS 服务器：114.114.115.115

11.配置完成后，勾选“退出时验证设置”。



12.返回控制台，然后绑定弹性 IP 即可使用。

9.2.5Windows 无法正常启动时怎样恢复数据？

故障描述

登录 Windows 操作系统云主机时，无法正常进入系统。自启动系统修复模式，但选择修复选项后报错，无法继续进行系统修复。

可能原因

Windows 文件已损坏。

处理方法

1.登录云主机控制台，选择“计算->弹性云主机”。
2.卸载云主机挂载的数据盘。
在云主机的详情页，单击“云硬盘”页签，单击数据盘后的“卸载”。
3.重装故障云主机的操作系统。
选择待重装操作系统的云主机，并点击“关机”按钮，待云主机关机完成之后，在云主机的“操作”列下，单击“更多>一键重装”，在“一键重装”弹窗选择镜像及密码相关信息，并点击“确定”。

说明
在执行任何更改或操作之前，建议备份弹性云主机数据，并确保具备恢复机制。
在遇到复杂问题或无法解决的情况下，寻求天翼云专业支持和建议。

4.将数据盘重新挂载至云主机。
在云主机的详情页，点击“云硬盘”页签，选择“挂载磁盘”，选择步骤 2 卸载的磁盘并进行挂载。
5.云主机开机后，重新验证云主机是否可以正常进入系统。

9.2.6 如何查看 Windows 云主机的登录日志？

操作场景

本文介绍查看 Windows 云主机的登录日志方法。

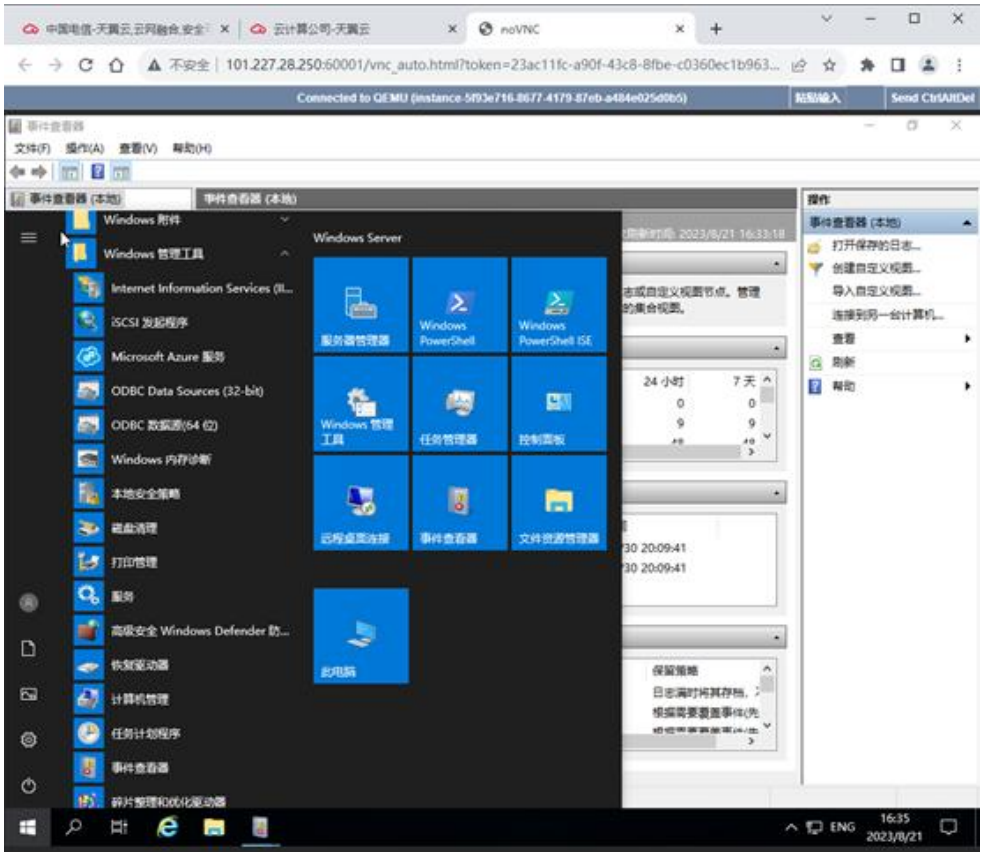
操作步骤

本文操作以 Windows2019-DC 操作系统云主机为例。

1. 登录弹性云主机。

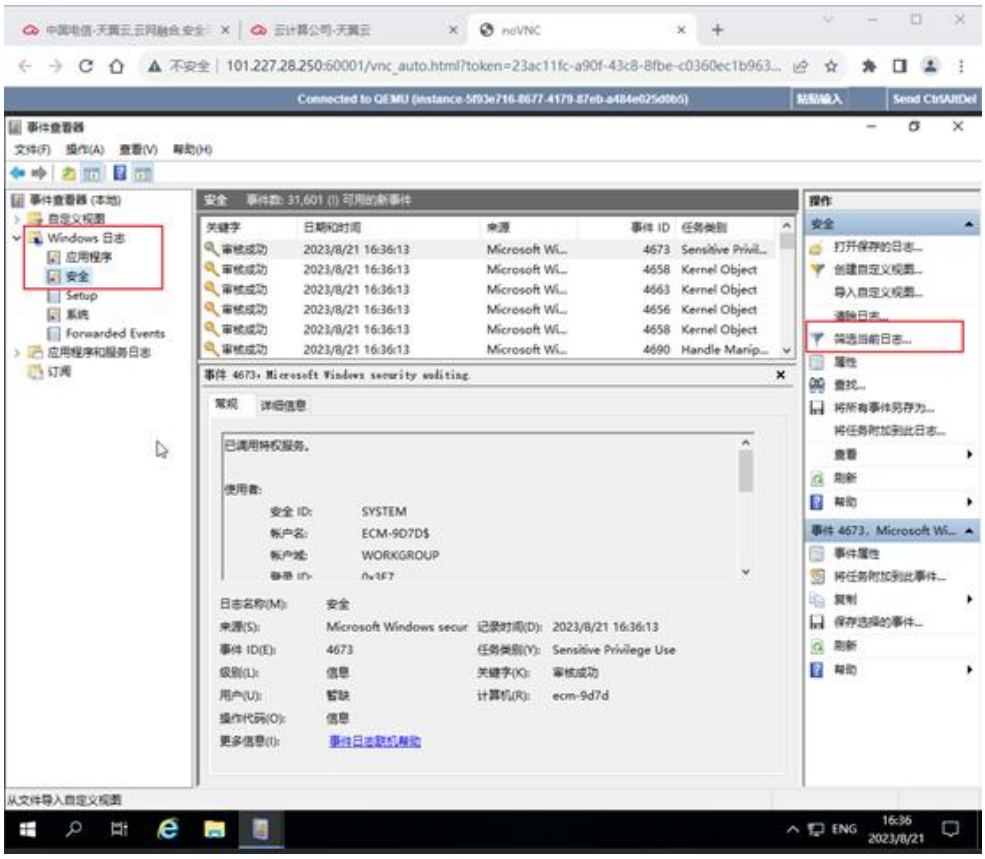
2. 选择“开始 > Windows>管理工具 > 事件查看器”。

图 1 事件查看器



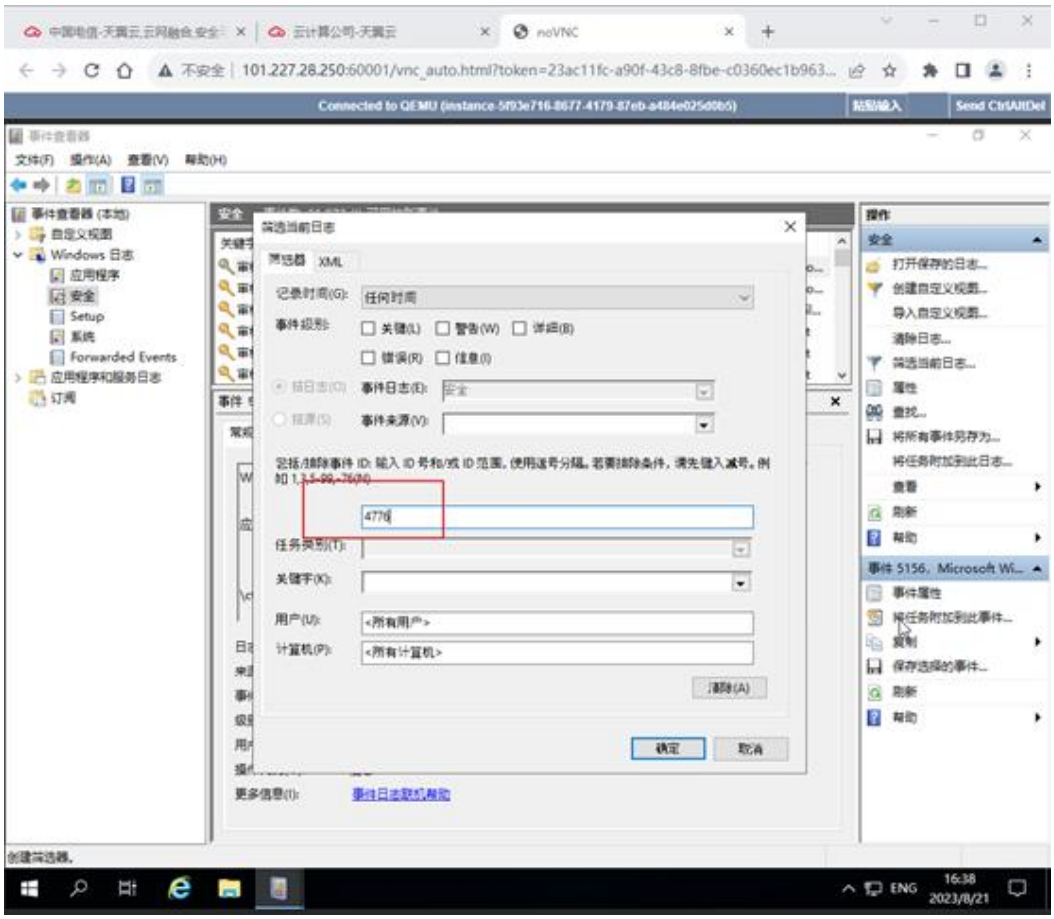
3. 打开“Windows 日志 > 安全 > 筛选当前日志”。

图 2 筛选当前日志



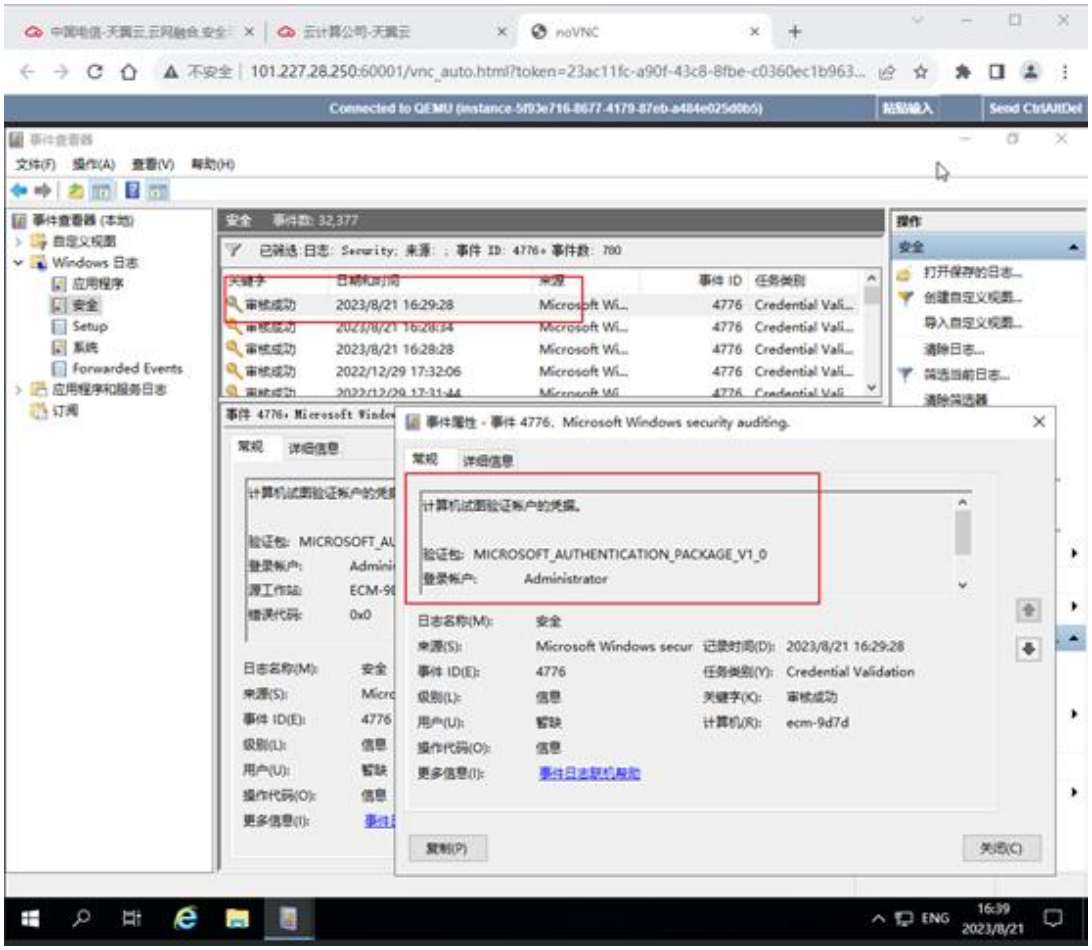
4. 筛选事件 ID 为 4776 的事件即为远程登录日志。

图 3 筛选远程登陆日志



5. 单击右键查看“事件属性”，即可看到该事件的登录账户。

图 4 查看“事件属性”



9.2.7CentOS 设置系统时区

操作场景

本节操作介绍在 CentOS 操作系统的弹性云主机里如何设置时区。

约束与限制

- 该文档已在 CentOS 6.8 及 CentOS 7.5 操作系统云服务器上验证，其他版本可能存在差异。
- 档中以中国上海时区为例。

CentOS 6/RHEL 6 操作系统（以中国上海时区为例）

1.修改系统时区（永久），请执行：

```
# cp -vf /usr/share/zoneinfo/Asia/Shanghai /etc/localtime
```

2.确认当前时区是否已经为 CST，重新登录后执行以下命令。

```
# date |awk '{print $5}'
```

CentOS 7 操作系统（以中国上海时区为例）

1.查看所需的时区，请执行：

```
# timedatectl list-timezones
```

2.更改时区，请执行：

```
# timedatectl set-timezone Asia/Shanghai
```

3.验证时区状态，请执行：

```
# timedatectl status
```

9.2.8 启动 Tomcat 时报错，提示 80 端口被占用怎么办？

问题描述

当 Windows 系统弹性云主机启动 Tomcat 时报错，提示 80 端口被占用。

本节操作以 Windows Server 2019Datacenter 操作系统的弹性云主机为例介绍排查 80 端口占用的操作步骤。

可能原因

提示 80 端口被占用，由提示可知，该端口可能是被其他程序占用，包括被木马病毒占用，可以按照下面步骤，查看占用该端口的进程 ID，并再根据进程 ID 查看具体进程。

1.使用 netstat -ano | find “80”查看占用 80 端口的进程 ID。

图 1 占用端口的进程 ID



2.根据上面步骤得到的进 ID，通过命令 tasklist /svc | find “4”进一步查看具体的进程。



通过检查发现，系统 System 进程占用了 80 端口。

处理方法

说明

该操作可能需要关闭正在运行的业务或者重启弹性云主机。

方法一

1.以管理员身份运行 cmd，执行命令 net stop http 。

2.如果需要停止相应的服务，则选择 Y 。

3.执行命令 sc config http start= disabled 。

方法二

1.在 cmd 下执行命令 regedit， 打开注册表。

2.找到注册表

HKEY_LOCAL_MACHINE>SYSTEM > CurrentControlSet > services > HTTP，修改 Start 值为 0。

1. 重启弹性云主机。

9.2.9 怎样实现 Windows 云主机文件共享？

操作场景

本文介绍实现同一个子网下 Windows 云主机之间文件共享的方法。

约束和限制

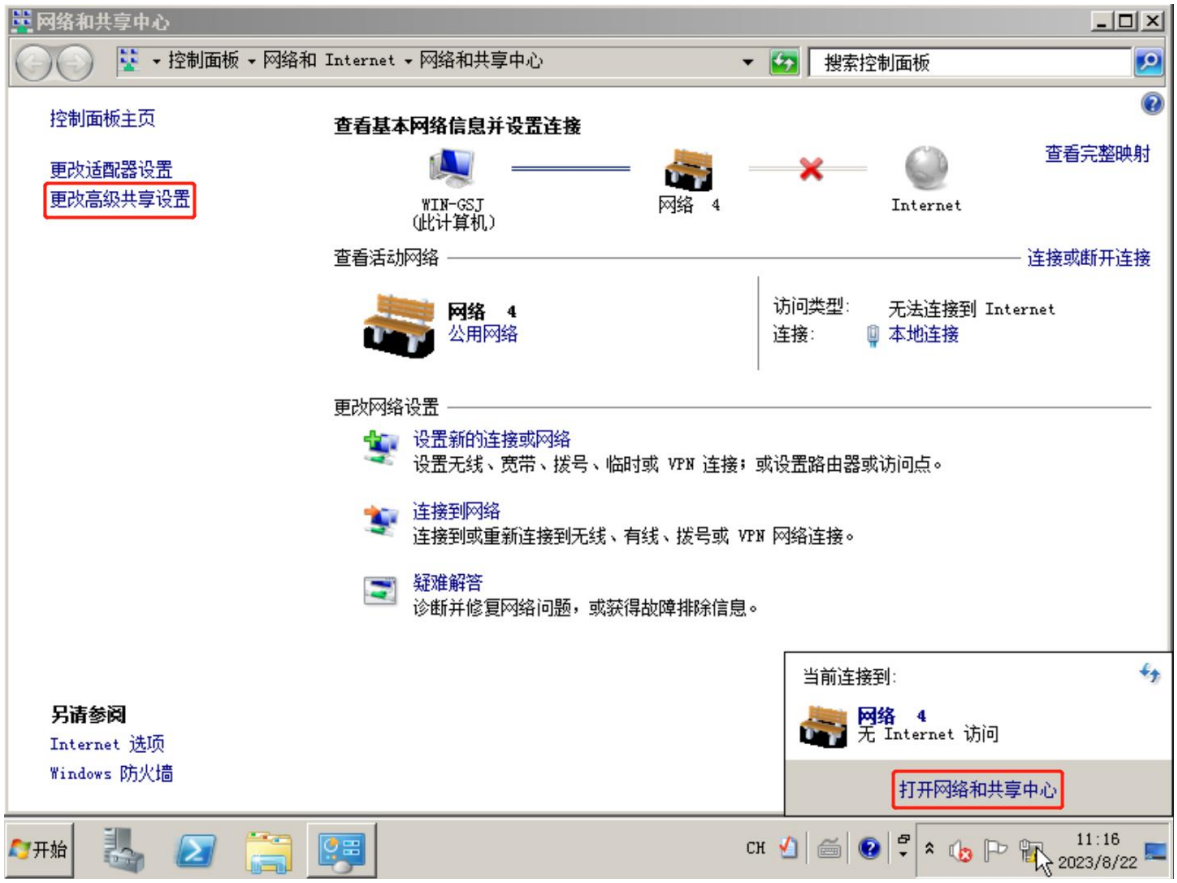
共享文件的云主机在同一个子网下，且网络互通。

操作步骤

本文以 Windows Server 2008 标准版 64 位中文版操作系统云主机为例。

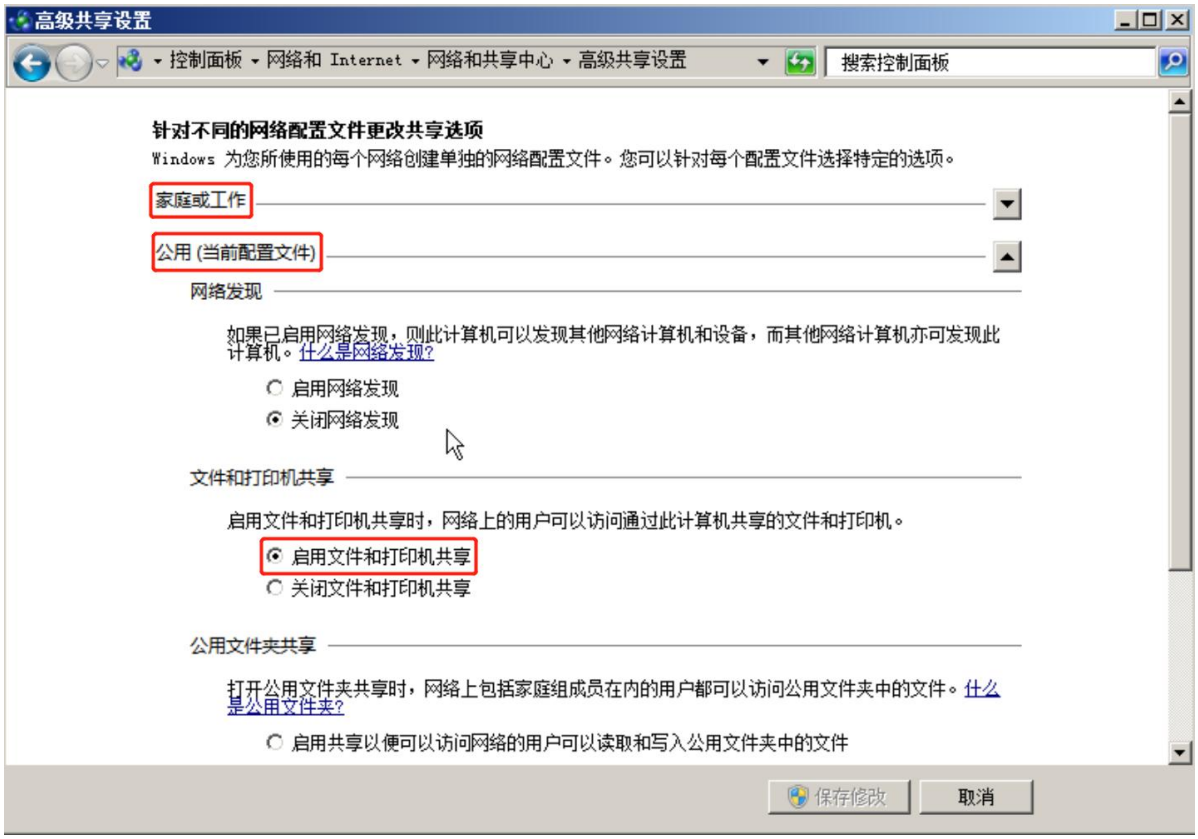
- 登录弹性云主机。
- 在云主机右下方的网络图标处，右键点击“打开网络和共享中心”。

图 1 打开网络和共享中心



- 点击“更改高级共享设置”，分别开启“家庭或工作”和“公用”下的“启用文件和打印机共享”，点击“保存修改”

图 2 启用或关闭共享



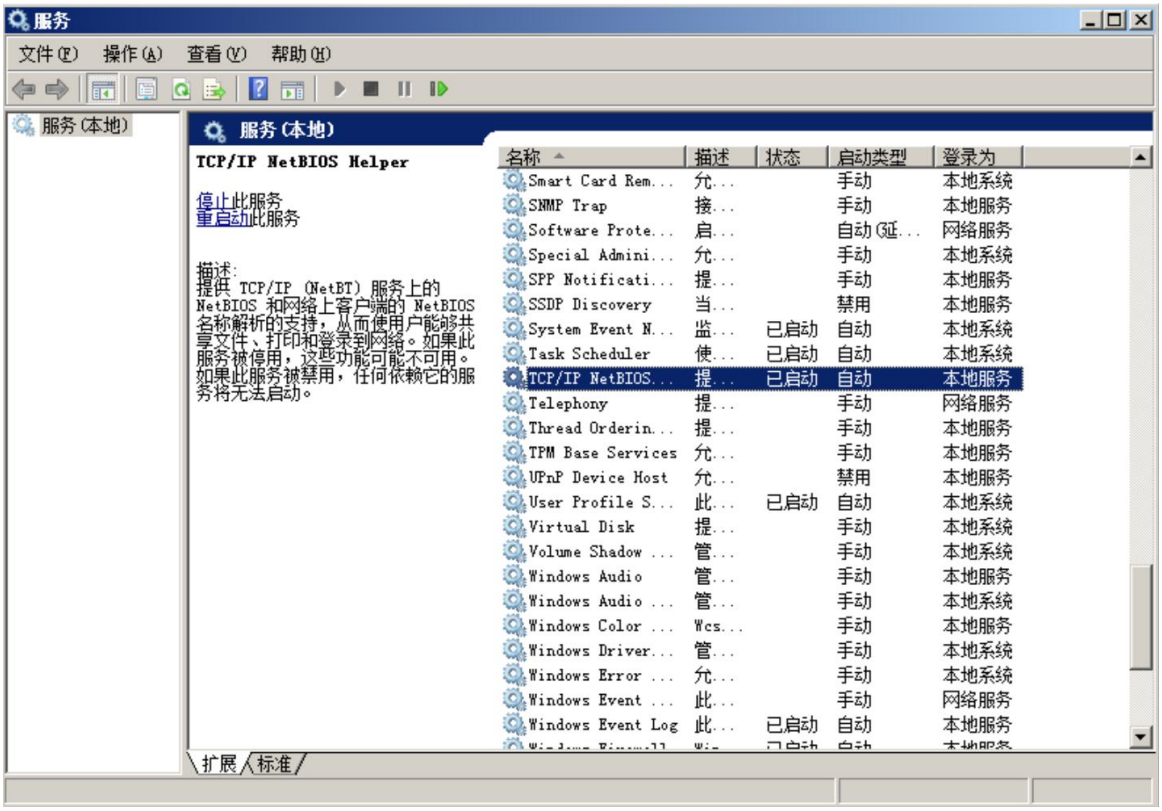
- 在 CMD 窗口执行命令 services.msc，找到 TCP/IP NetBIOS Helper 服务，并设置为开启。

若服务处于“禁用”状态，则右键属性，将服务设为“自动”或者“手动”，然后开启服务。

图 3 CMD 窗口执行命令



图 4 开启 TCP/IP NetBIOS Helper



9.2.10 输入法无法使用怎么办？

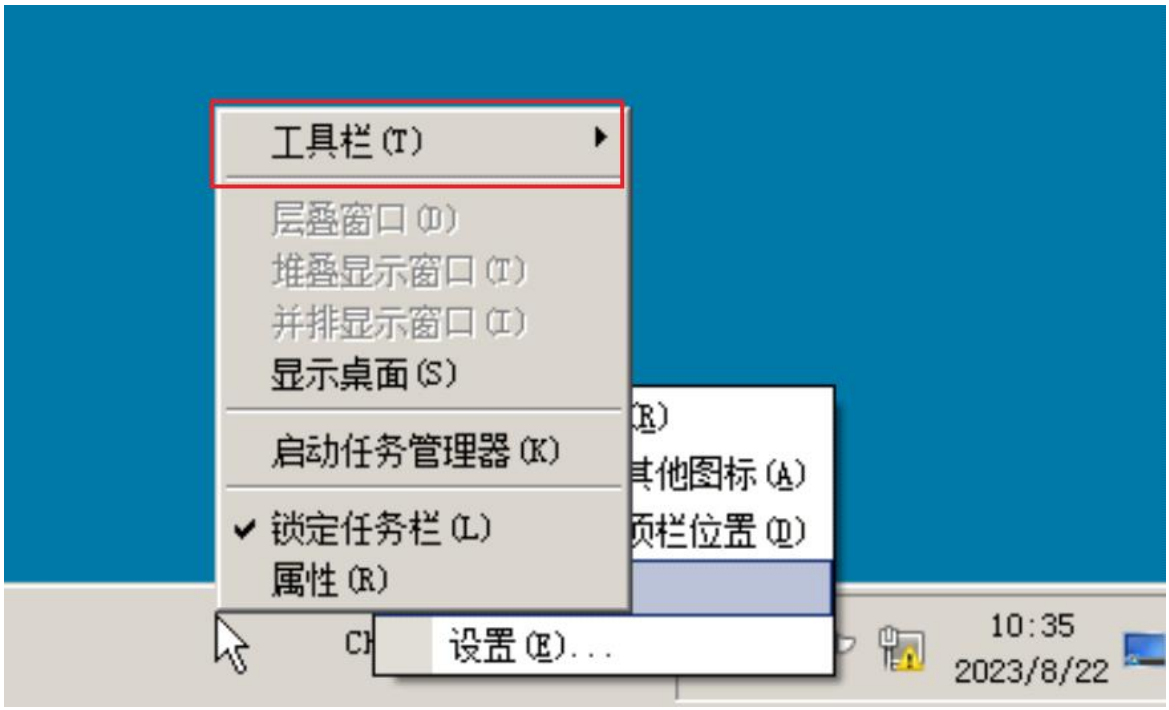
故障原因

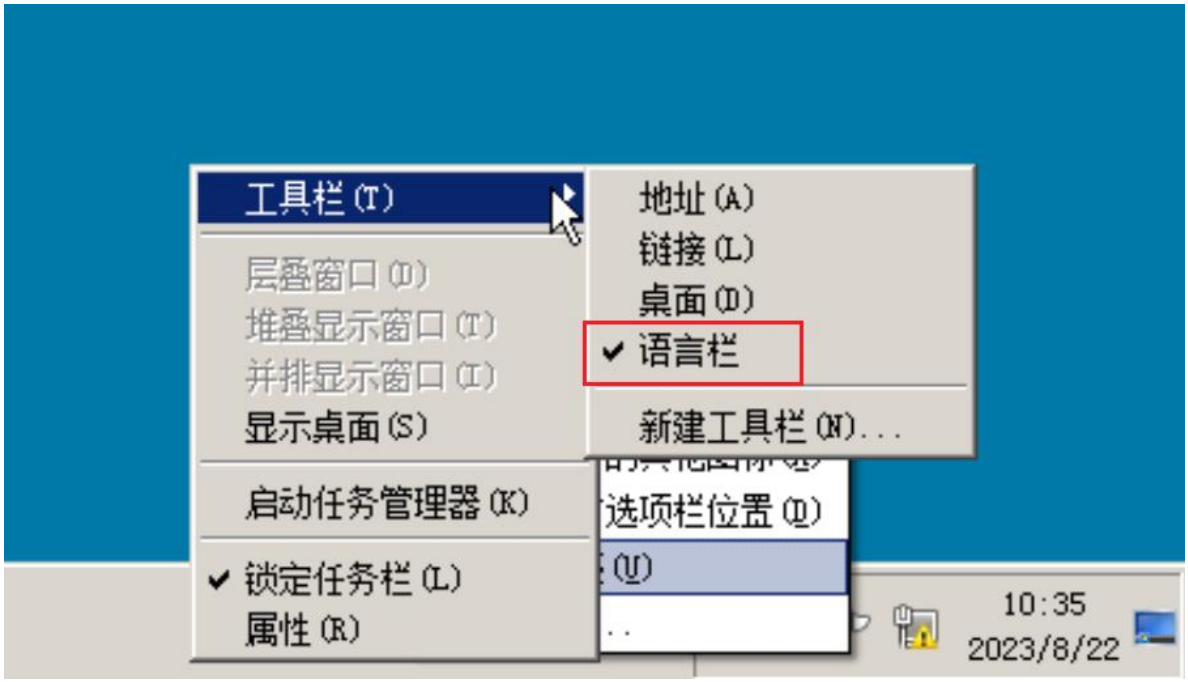
- 工具栏中未开启语言栏
- 病毒感染，导致系统文件、注册表损坏。

操作步骤

方式一

1. 登录弹性云服务器。
2. 检查右下角任务栏中是否存在输入法的图标。
3. 若不存在输入法图标，请尝试使用 Ctrl+Shift 查看是否能进行输入法切换，若失败请进行下面操作。
4. 在任务栏的空白处右键后选择：工具栏 -->语言栏。

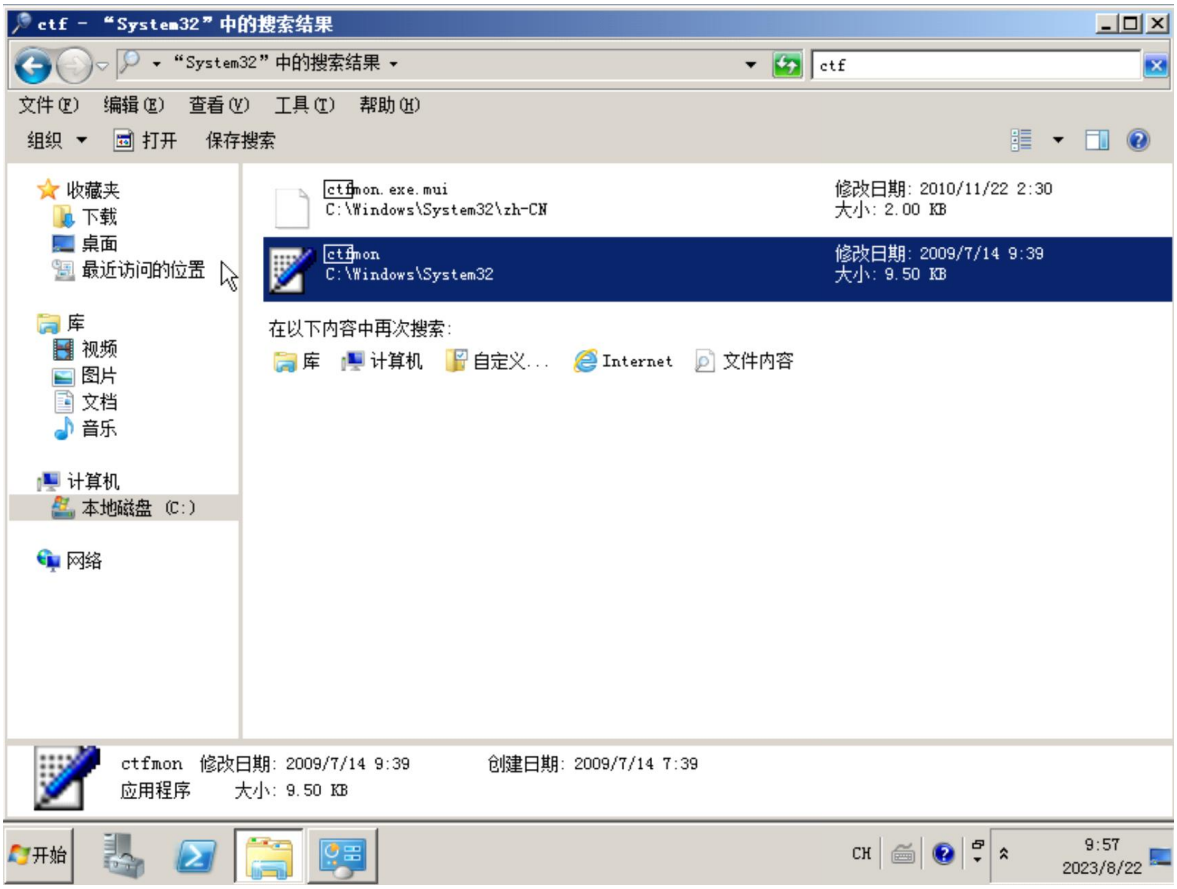




方式二

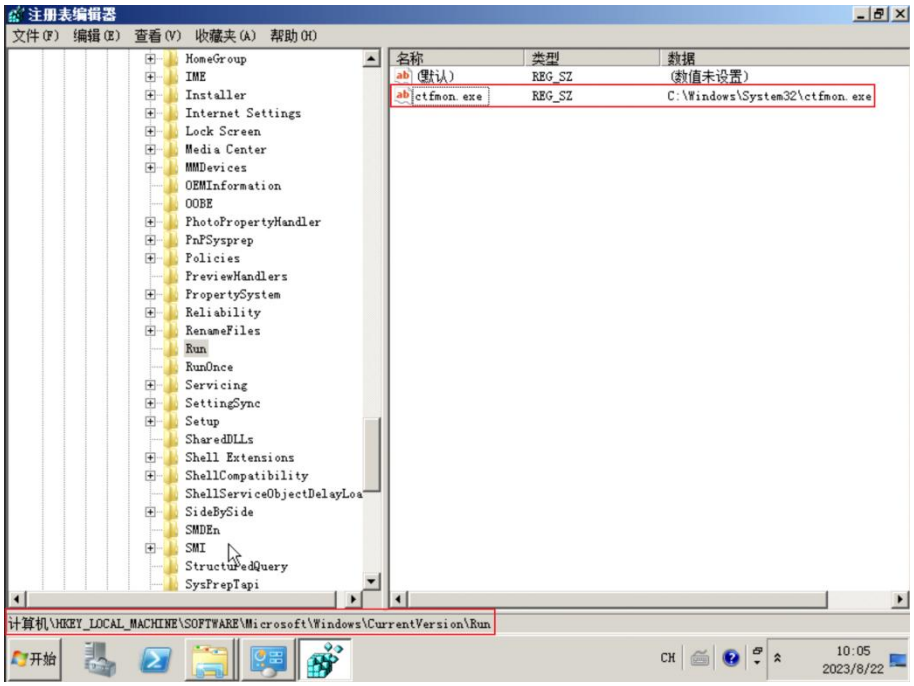
1. 登录弹性云服务器。

2. 在 C:\Windows\System32 中找到 ctfmon.exe。

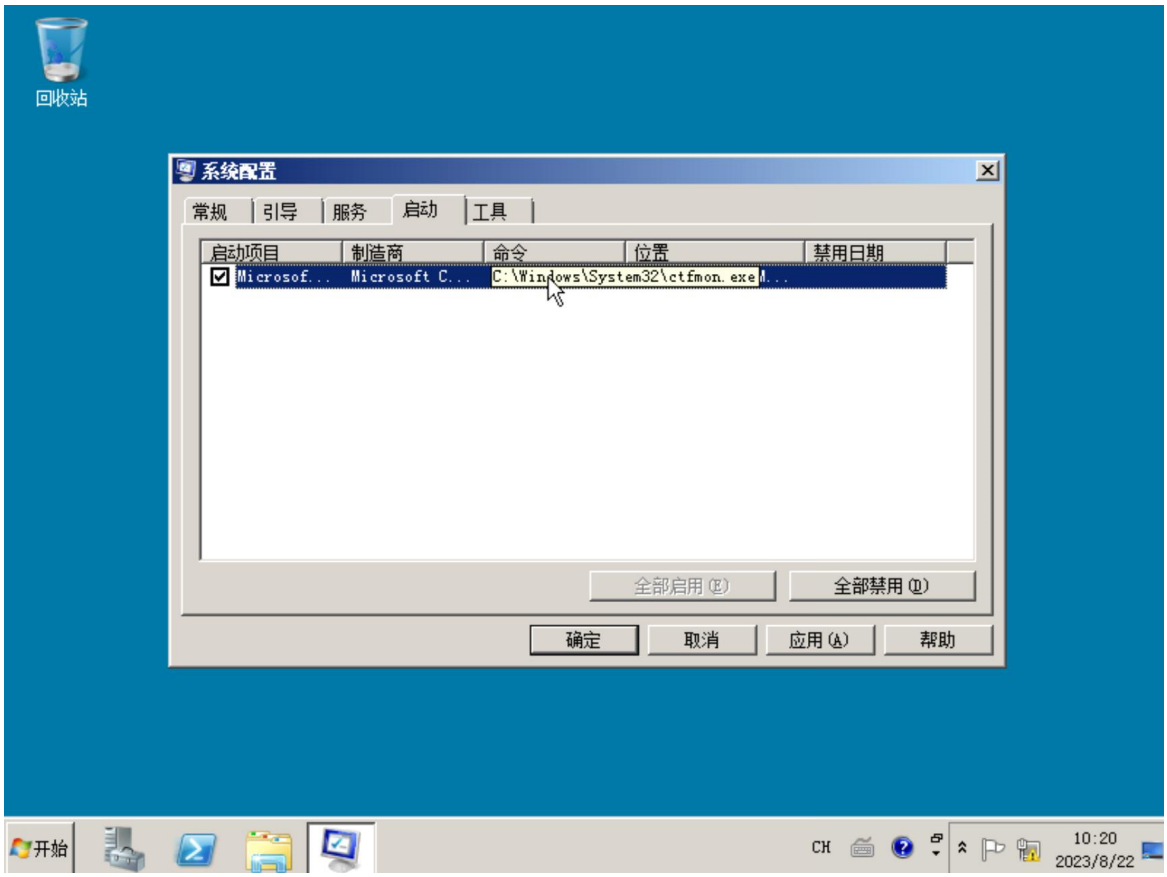


3.双击执行后查看是否显示输入法图标。

4.将 ctfmon.exe 添加到注册表中：在运行窗口中输入 regedit 进入注册表编辑器，在下图路径 HKEY_LOCAL_MACHINE > SOFTWARE > Microsoft > Windows > CurrentVersion > Run 新建一个注册表文件—ctfmon.exe。



5. 将 ctfmon.exe 添加到开机自启动中：在运行窗口执行 msconfig，查看“启动”中 ctfmon 项是否已经勾选。



6. 点击应用后重启弹性云服务器可生效。

9.2.11 Windows 云主机配置文件共享和网络磁盘映射方法

约束限制

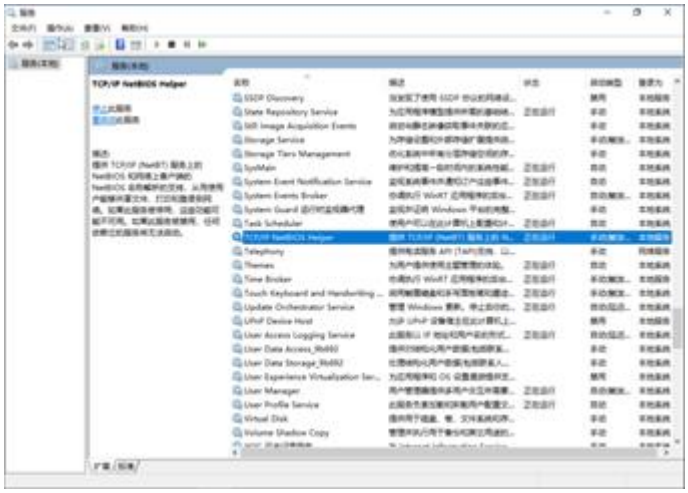
- 有些网络服务供应商可能会限制 139、445 端口的访问，这会导致外部网络无法访问共享文件夹。因此，建议将 Windows 云主机文件共享方案仅用于内部环境中。
- 注意安全组的设置，放行对应的端口号。

操作步骤

1.确保两台需要共享文件夹的 Windows 云主机的设置是正确的，包括以下几个方面：

- 确认已开启名为“TCP/IP NetBIOS Helper”的服务。你可以通过打开命令提示符窗口（cmd）并输入 services.msc，然后在服务清单中找到“TCP/IP NetBIOS Helper”服务，检查该服务的运行状态。

图 1 TCP/IP NetBIOS Helper



- 在网络适配器设置中开启 TCP/IP 上的 NetBIOS 功能。你可以通过以下步骤完成：右键点击任务栏中的网络图标并选择“打开网络和共享中心”；然后点击“更改适配器设置”；在适配器列表中选择你的以太网连接，右键点击并选择“属性”；双击“Internet 协议版本 4（TCP/IPv4）”项目；点击“高

级”按钮，然后切换到“WINS”选项卡；在这里，选择“启用通过 TCP/IP 传输 NetBIOS（N）”。

图 2 启用 TCP/IP 上的 NetBIOS 服务



- Windows 云主机的防火墙已经设置了入站访问策略 139 和 445 端口。

2.修改网络和共享中心。

进入网络和共享中心”，点击“更改高级共享设置”。

图 3 更改高级共享设置



a. 确保两台 Windows 云主机所属的网络应一致，例如均为“公用”或均为“专用”的网络，并勾选“启用网络发现”、“启用共享，以便可以访问网络的用户可以读取和写入公用文件夹中的文件”。

图 4 启动网络发现



图 5 启用共享，以便可以访问网络的用户可以读取和写入公用文件夹中的文件。



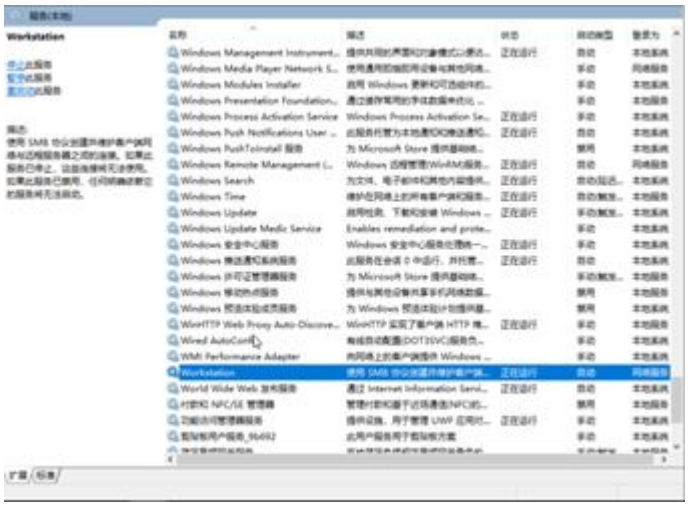
如果无法启用网络发现功能，你可以通过打开“服务”管理器（运行 services.msc 命令）来解决，启动“服务”管理器。请检查网络发现功能依赖的服务是否均已启用。网络发现功能依赖的服务包括：

Function Discovery Resource Publication、SSDP Discovery Resource Publication、UPnP Device Host。

b. 在“服务”管理器中，开启 Workstation 服务。

c. 这个服务需要依赖其他部件的支持，开启 workstation 服务时，必须启动依赖组件 Computer Browser 和 Remote Desktop Configuration。

图 6 开启 Workstation 服务



3. 在需要允许磁盘共享访问的云主机上，进行磁盘共享的设置，点击鼠标右键磁盘，然后选择“属性”选项。切换到“共享”选项卡，接着点击“高级共享”按钮。

图 7 设置磁盘共享



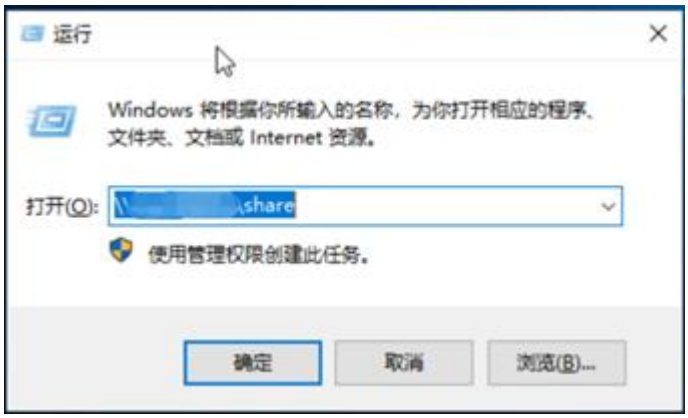
设置“共享名”，并单击“确定”。给共享文件夹自定义一个名称。

图 8 设置共享文件夹名称



4. 使用位于相同地域的另一台云主机，通过内部网络访问共享文件夹。打开“运行”窗口。输入"\内网 IP 地址\share"，并点击确定，即可打开共享文件夹.

图 9 访问共享文件夹



5.（可选）为了更加方便的访问共享文件夹，您可以将共享路径文件，创建一个网络驱动器映射。

a. 选择共享的文件夹，右键单击“映射网络驱动器”。

b. 创建映射网络驱动器。

图 10 映射网络驱动器



c. 点击选择要映射的网络驱动器名称，接着点击“创建快捷方式”并将其发送到桌面。这样，你就能通过双击桌面上的新快捷方式，方便快速地访问共享文件夹。

图 11 创建快捷方式



9.2.12Windows 云主机如何保持会话连接长时间不断开？

操作场景

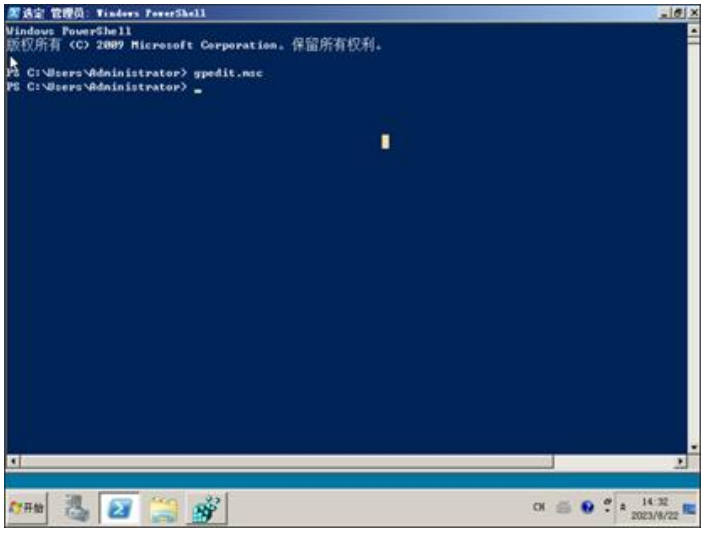
本节操作介绍如何设置 Windows 服务器以防止远程桌面连接在长时间不活动后自动断开。

操作步骤

以下操作以 Windows 2008 操作系统为例。

1.打开“开始”菜单，在“运行”对话框中，输入 gpedit.msc 并按回车键，以打开“本地组策略编辑器”。

图 1 gpedit.msc



2.在“本地组策略编辑器”窗口中，依次展开以下路径：计算机配置 > 管理模板 > Windows 组件 > 远程桌面服务 > 远程桌面会话主机 > 会话时间限制。

3.设置已中断会话的时间限制。

a. 选择“已启用”。

b. 结束已断开连接的会话：从不。



4.设置活动但空闲的远程桌面服务会话的时间限制。

a. 选择”已启用“。

b. 空闲会话限制：从不。



9.2.13Windows 弹性云主机不能复制粘贴内容？

故障描述

当本地通过客户端软件远程连接 Windows 弹性云主机后，不能复制、粘贴内容，右键单击菜单栏，“粘贴”选项置灰。

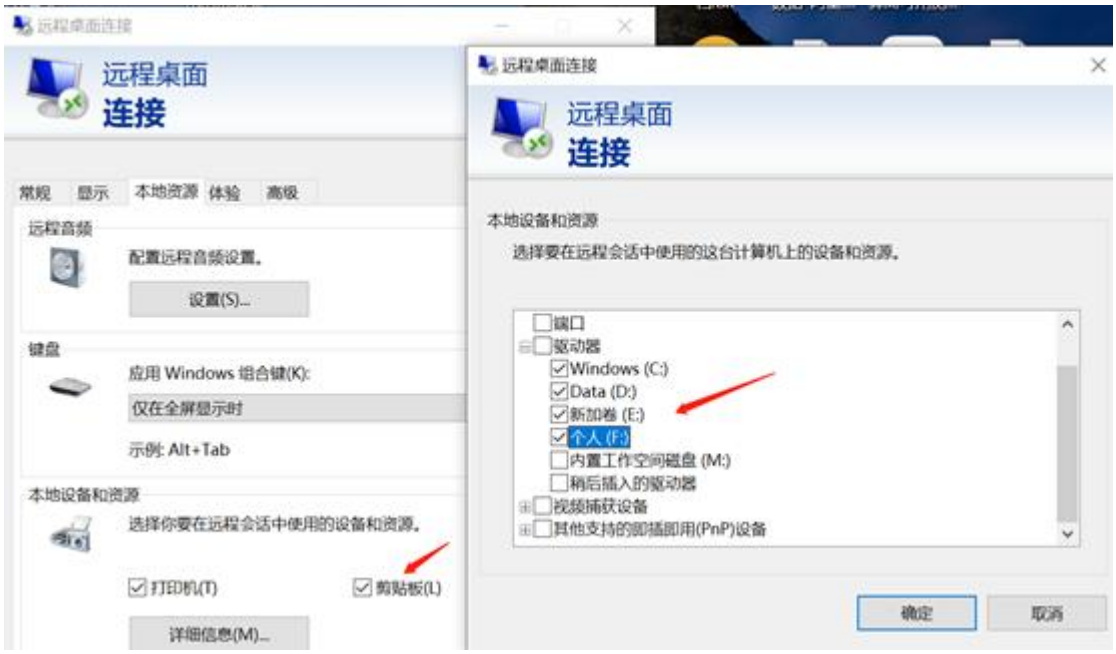
故障原因

- 本地远程桌面配置未进行本地驱动器映射。
- 远程服务器 rdpclip.exe 进程异常。
- 远程服务器系统禁止弹性云主机和本地主机之间进行文件的复制粘贴功能。

解决步骤

故障 1：未进行本地驱动器映射，图 1 以使用本地的 C 盘、D 盘、E 盘、F 盘映射为例。

1.在本地运行窗口输入 mstsc 回车打开“远程桌面连接”。
2.选择“本地资源”页签，并勾选“本地设备和资源”栏的“剪切板”、“驱动器”。
3.单击“确定”，检查复制、粘贴功能是否恢复正常。



故障 2：服务器 rdpclip.exe 进程异常，需要重启 rdpclip.exe 进程。

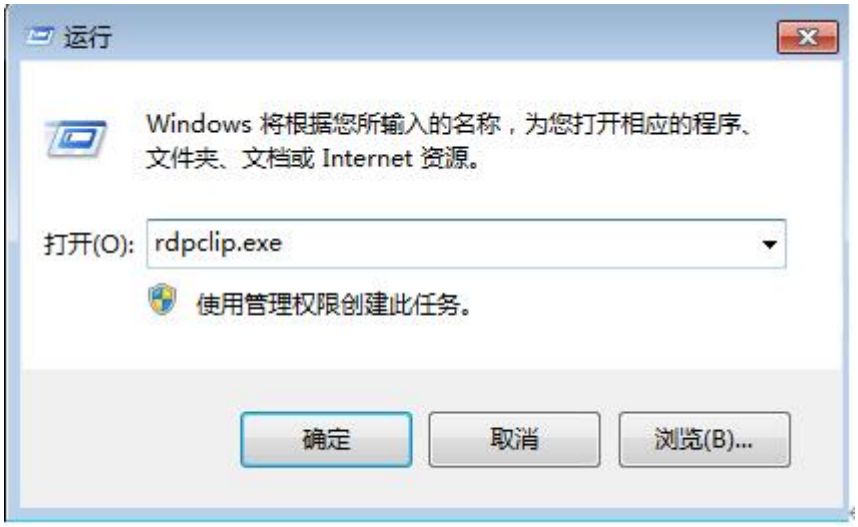
1.远程桌面连接服务器，启动 Windows 任务管理器，在进程页签下结束 rdpclip.exe 进程。

图 2 结束 rdpclip.exe 进程



2.打开“开始”菜单中的“运行”窗口，输入 rdpclip.exe，重新启动 rdpclip.exe。

图 3 启动 rdpclip.exe 进程



3.检查复制、粘贴功能是否恢复正常。

故障 3：系统禁止弹性云主机和本地主机之间进行文件的复制粘贴功能，需重新配置弹性云主机的本地组策略编辑器，开启复制粘贴功能。

1.打开“开始 > 运行”，输入 gpedit.msc，打开“本地组策略编辑器”。

2.依次找到：“计算机配置 > 管理模板>Windows 组件 > 远程桌面服务 >远程桌面会话主机 > 设备和资源重定向”。

3.将“不允许剪切板重定向”、“不允许 COM 端口重定向”、“不允许驱动器重定向”、“不允许 LPT 端口重定向”几个配置都设置为“已禁用”。

图 4 设备和资源重定向



4.设置完成之后重启弹性云主机使配置生效，检查复制、粘贴功能是否恢复正常。

9.2.14Windows 弹性云主机不能复制粘贴内容？

故障描述

当本地通过客户端软件远程连接 Windows 弹性云主机后，不能复制、粘贴内容，右键单击菜单栏，“粘贴”选项置灰。

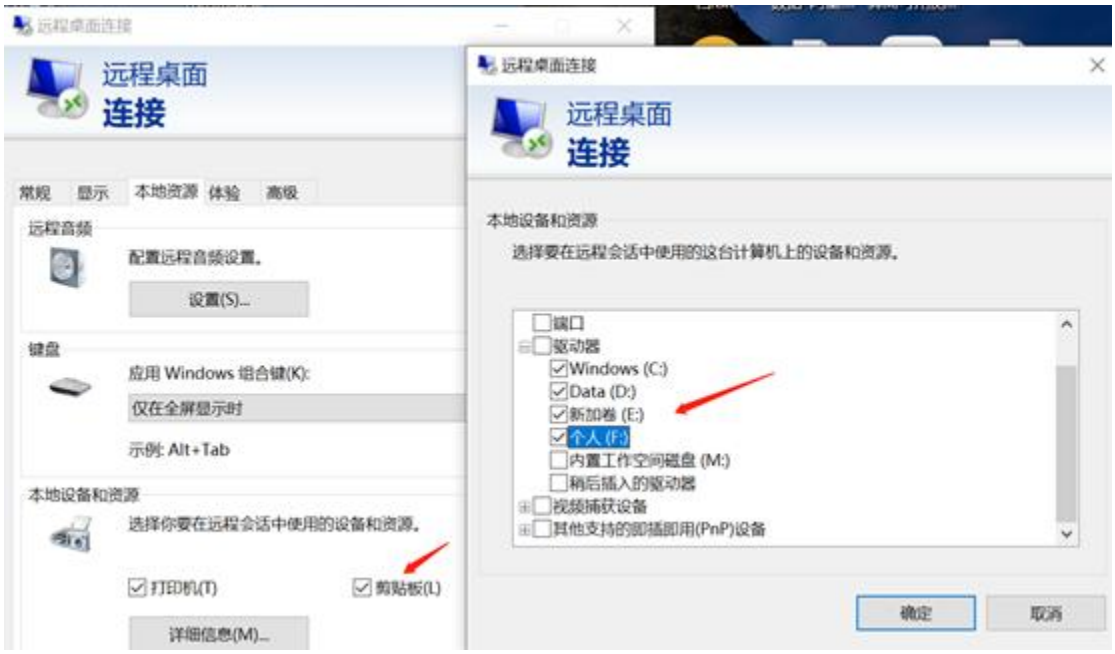
故障原因

- 本地远程桌面配置未进行本地驱动器映射。
- 远程服务器 rdpclip.exe 进程异常。
- 远程服务器系统禁止弹性云主机和本地主机之间进行文件的复制粘贴功能。

解决步骤

故障 1：未进行本地驱动器映射，图 1 以使用本地的 C 盘、D 盘、E 盘、F 盘映射为例。

1. 在本地运行窗口输入 mstsc 回车打开“远程桌面连接”。
2. 选择“本地资源”页签，并勾选“本地设备和资源”栏的“剪切板”、“驱动器”。
3. 单击“确定”，检查复制、粘贴功能是否恢复正常。



故障 2：服务器 rdpclip.exe 进程异常，需要重启 rdpclip.exe 进程。

1. 远程桌面连接服务器，启动 Windows 任务管理器，在进程页签下结束

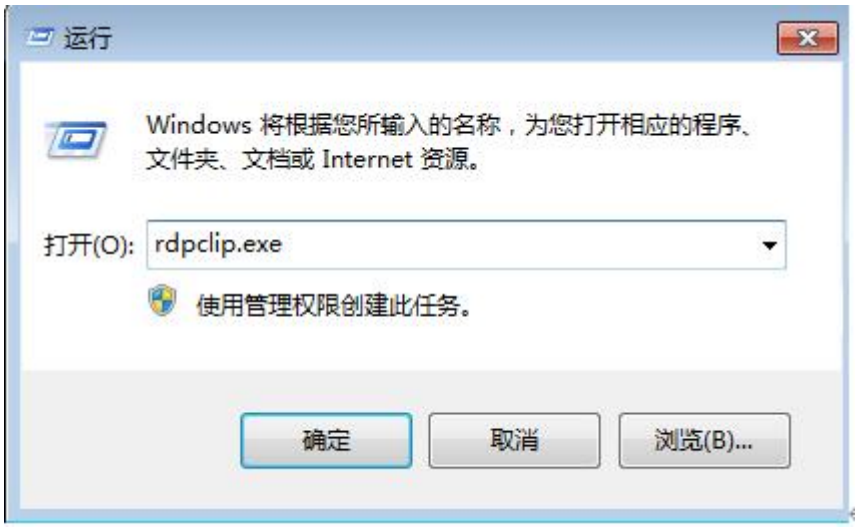
rdpclip.exe 进程。

图 2 结束 rdpclip.exe 进程



2. 打开“开始”菜单中的“运行”窗口，输入 rdpclip.exe，重新启动 rdpclip.exe。

图 3 启动 rdpclip.exe 进程



3. 检查复制、粘贴功能是否恢复正常。

故障 3：系统禁止弹性云主机和本地主机之间进行文件的复制粘贴功能，需重新配置弹性云主机的本地组策略编辑器，开启复制粘贴功能。

1. 打开“开始 > 运行”，输入 gpedit.msc，打开“本地组策略编辑器”。

2. 依次找到：“计算机配置 > 管理模板>Windows 组件 > 远程桌面服务 >远程桌面会话主机 > 设备和资源重定向”。

3. 将“不允许剪切板重定向”、“不允许 COM 端口重定向”、“不允许驱动器重定向”、“不允许 LPT 端口重定向”几个配置都设置为“已禁用”。

图 4 设备和资源重定向



4. 设置完成之后重启弹性云主机使配置生效，检查复制、粘贴功能是否恢复正常。

9.2.15 怎样设置 Windows 弹性云主机首选语言

操作步骤

本文以 Windows 2008 操作系统弹性云主机，切换日语输入法为例：

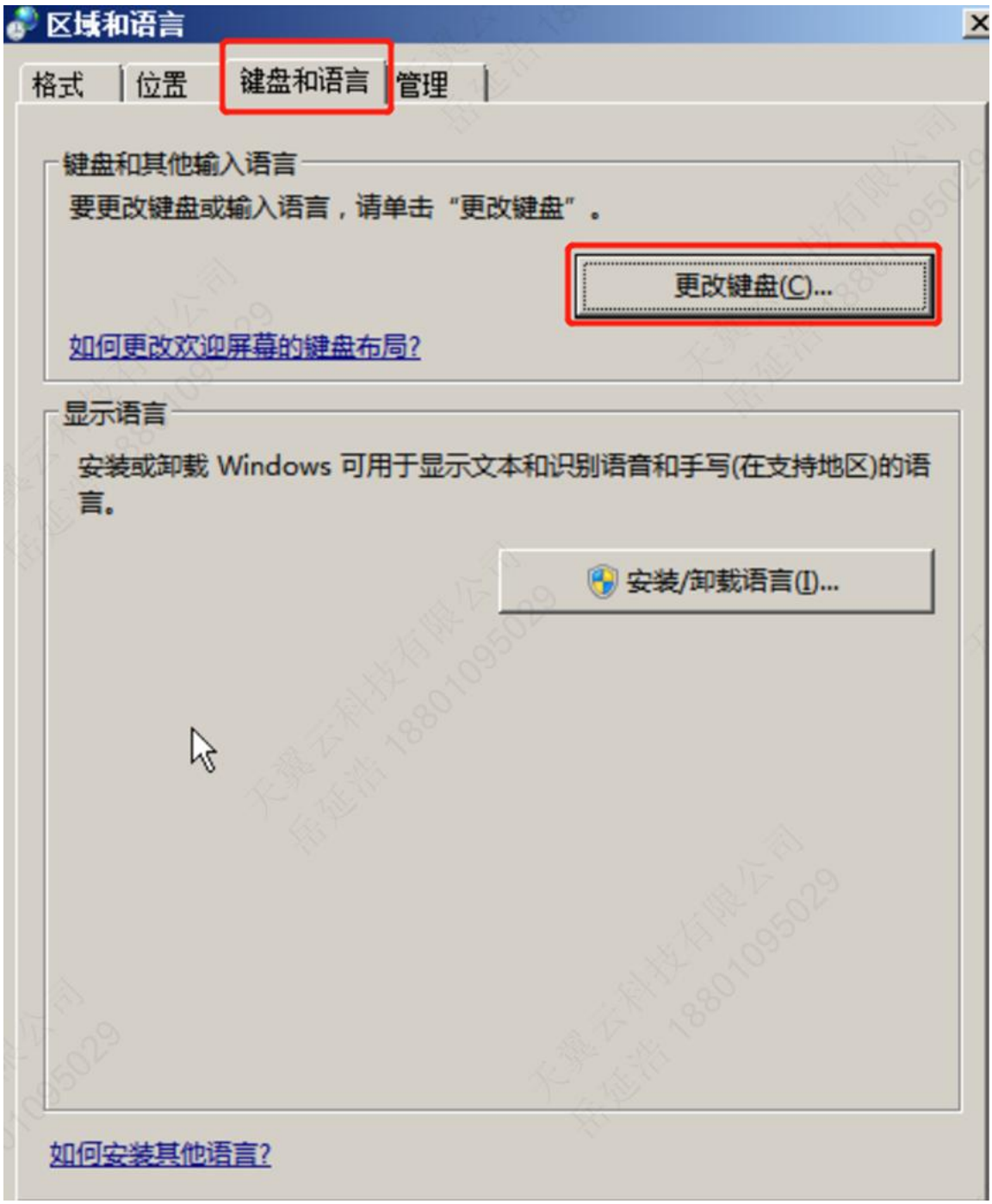
- 登录云主机，打开“控制面板”。
- 单击”时钟、语言和区域“下的“更改键盘或其他输入法”选项。

图 1 更改键盘或其他输入法



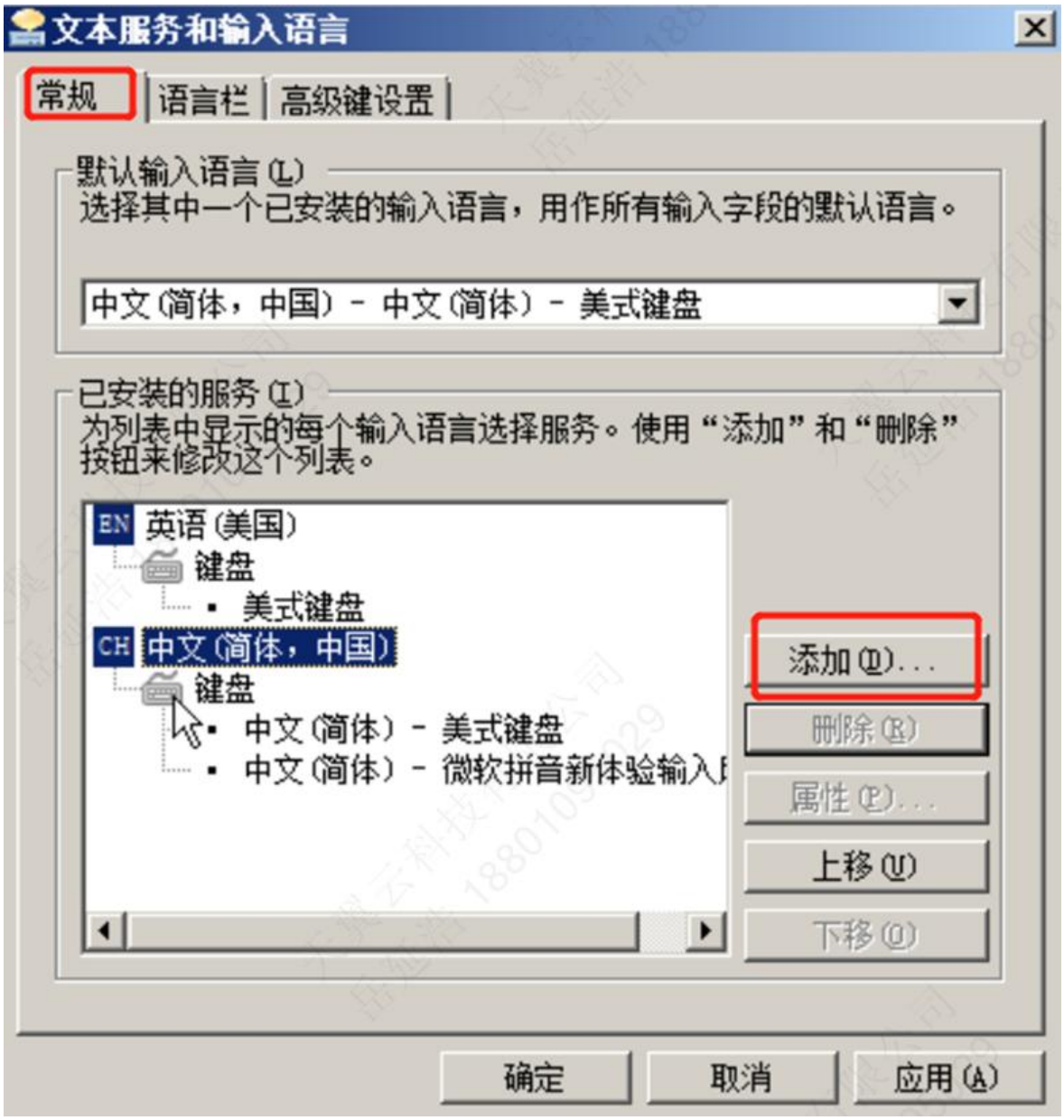
- 选择“键盘和语言”标签，并单击“更改键盘”。

图 2 键盘和语言



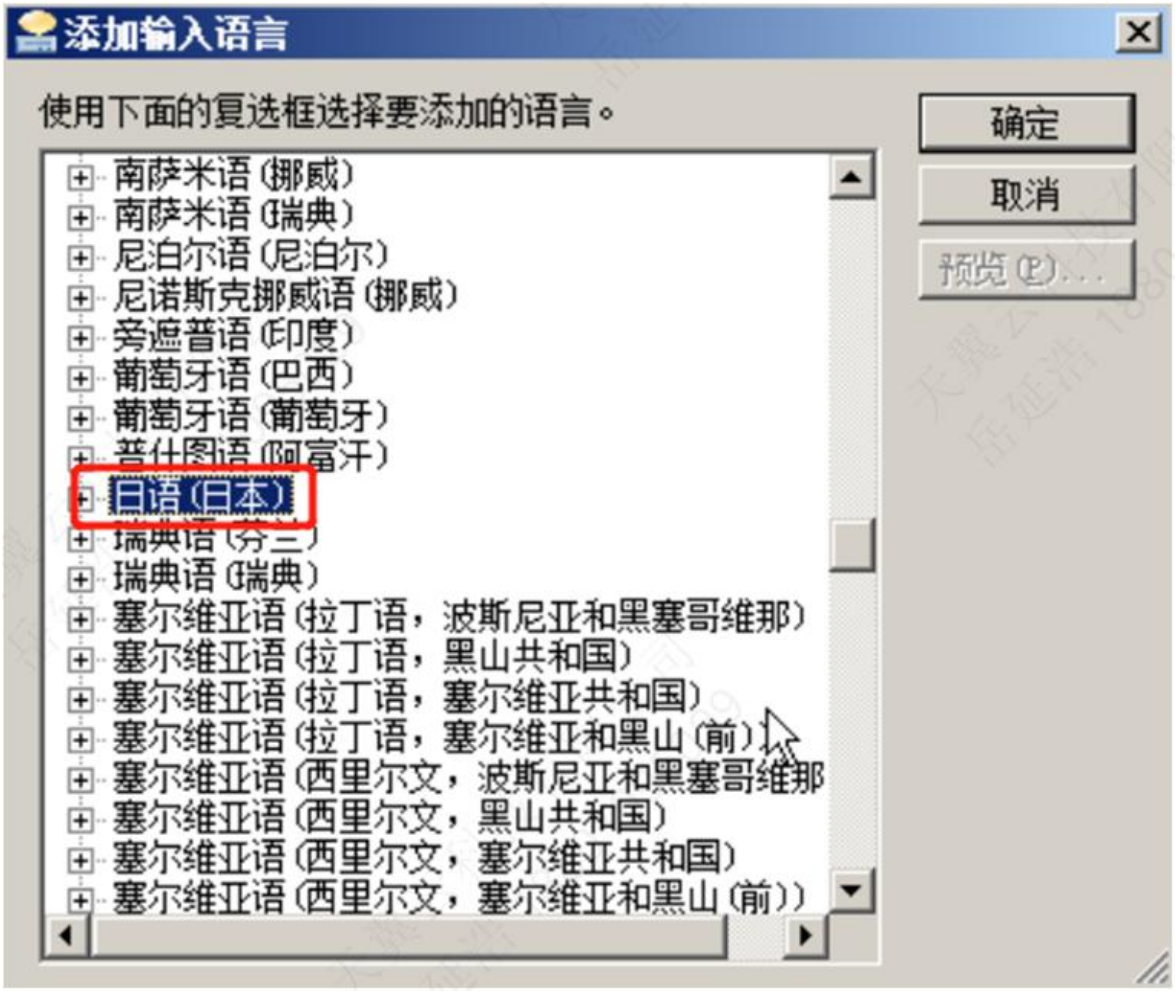
- 打开“文本服务和输入语言”对话框，在“常规”标签中单击 “添加”。

图 3 文本服务和输入语言



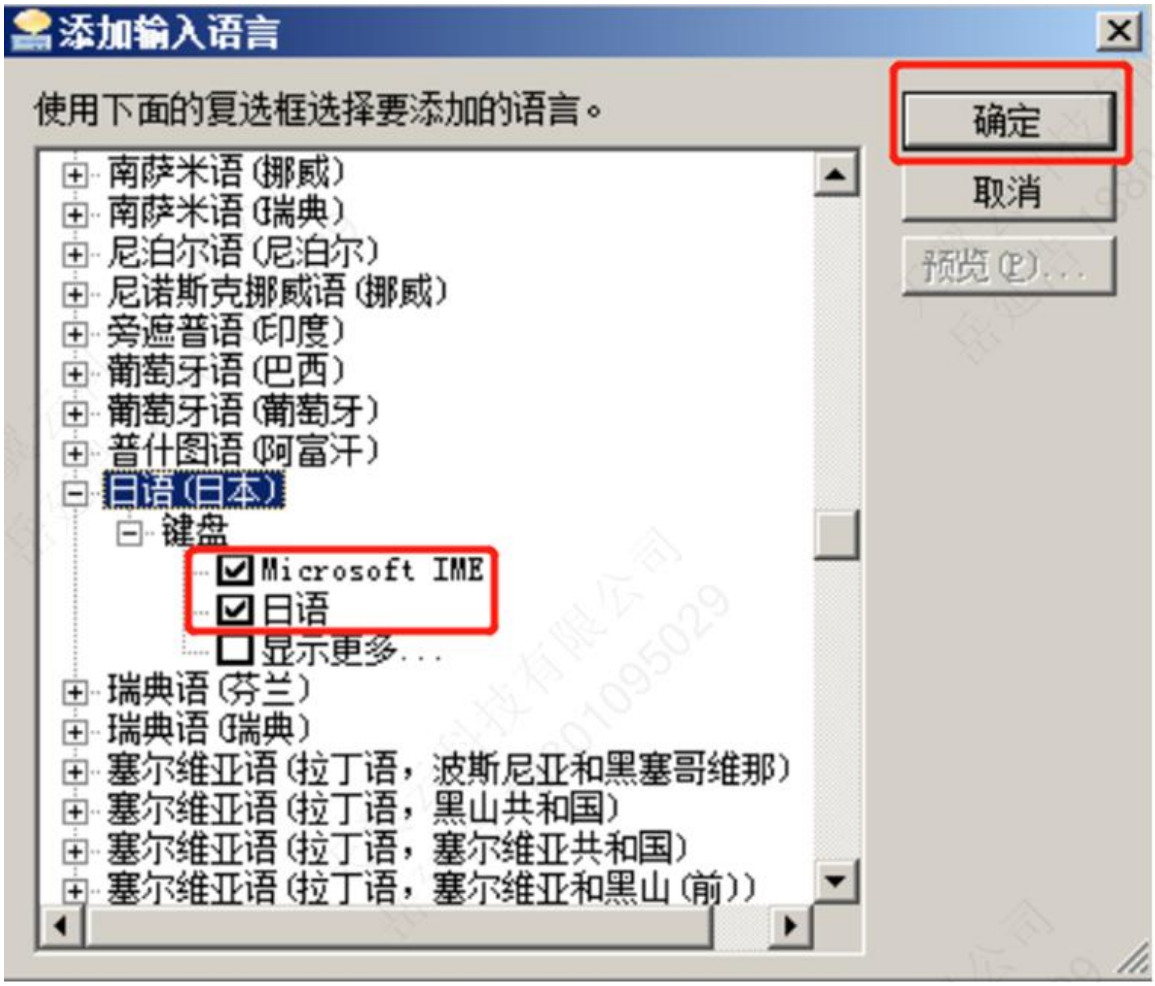
- 在弹出的“添加输入语言”对话框中，找到“日语（日本）”。

图 4 添加输入语言



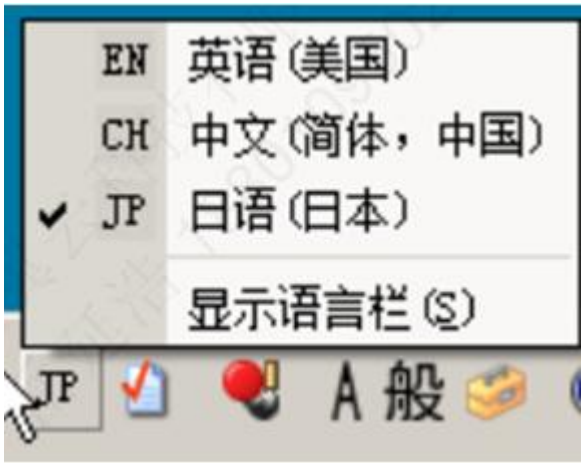
- 双击“日语（日本）”展开，双击“键盘”子项展开，勾选“Microsoft IME”和“日语”，单击“确定”，分别再单击“应 用”和“确定”保存修改后退出。

图 5 选择日语



- 点击云主机右下角任务栏输入法图标，输入法已切换为日语。

图 6 显示栏输入法列表



9.3 操作系统类（Linux）

9.3.1 弹性云主机新内核启动失败如何设置使用第二内核启动

操作场景

本节操作介绍云主机新内核启动失败时如何设置使用第二内核启动。

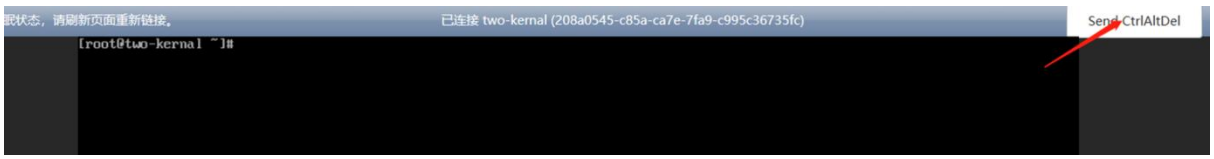
本节操作适用于 CentOS、CTyunOS 操作系统，且系统内安装至少两个内核。

以 CTyunOS 为例操作

1. 选择对应的云主机，点击“远程登录”。



2. 点击“Send CtrlAltDel”重启云主机。



3. 出现内核选择界面后，按下键，移动光标至第二内核，按回车键。

这里以救援模式，来当成第二内核。实际操作中，选择实际的内核版本即可。

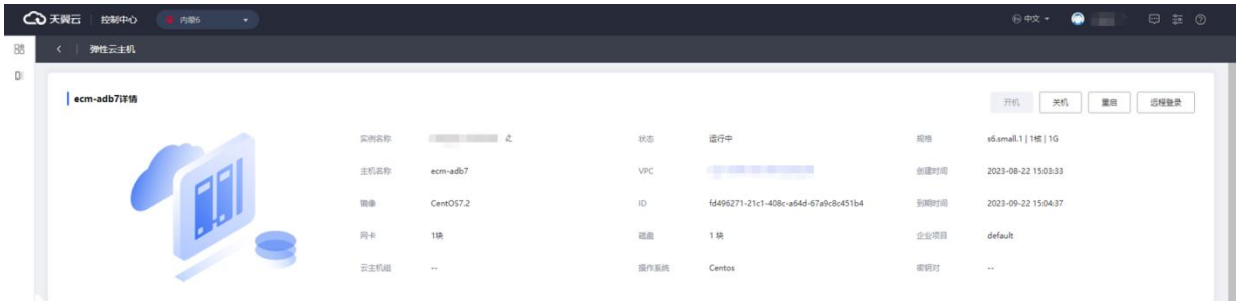


3. 待进入系统后，执行下面命令来设置第二内核为默认启动内核。

9.3.2 为什么操作系统实际版本和购买时的镜像版本不一致？

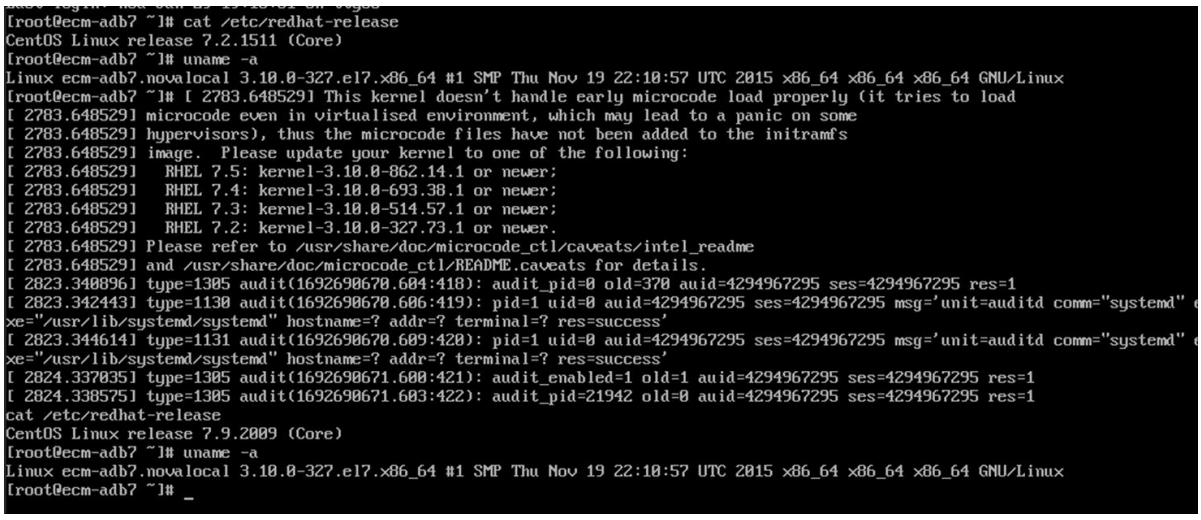
问题现象

购买时的镜像版本是 CentOS7.2，控制台 console 上显示镜像信息是 CentOS7.2。但是实际登录弹性云主机后查看到的操作系统版本是比 7.2 更高的版本，此问题常见于 linux 操作系统。



根因分析

原因是因为有做过系统内核更新的操作，例如 yum -y update 操作，会使弹性云主机从旧的操作系统版本更新成新的操作系统版本。



如果之前没有清理过历史命令的话，可以通过 yum history 查看 yum 命令记录。

```
[root@ecm-adb7 ~]# yum history
Loaded plugins: fastestmirror
ID     | Login user               | Date and time    | Action(s)      | Altered
-------------------------------------------------------------------------------
     8 | root <root>              | 2023-08-22 15:49 | I, O, U        |  247 E<
     7 | root <root>              | 2022-02-10 10:38 | Update         |    1 >
     6 | root <root>              | 2021-03-03 09:15 | Install        |    1  <
     5 | root <root>              | 2021-01-27 15:51 | I, U           |   82 >
     4 | root <root>              | 2021-01-27 15:51 | I, U           |   37
     3 | root <root>              | 2020-05-11 22:18 | I, O, U        |  103 EE
     2 | root <root>              | 2020-05-11 22:17 | Install        |    5
     1 | System <unset>           | 2020-05-11 14:06 | Install        |  284
history list
[root@ecm-adb7 ~]#
```

可以看到 2023 年 8 月 22 日有 yum 的 install 和 update 操作，用 yum history info 进一步查看操作的详细信息。

```
[root@ecm-adb7 ~]# yum history info 8
Loaded plugins: fastestmirror
Transaction ID : 8
Begin time     : Tue Aug 22 15:49:27 2023
Begin rpmdb    : 410:a7d99c16ca16b0f690810a25dbdf3e33947ddec0
End time       :            15:52:48 2023 (201 seconds)
End rpmdb      : 426:0c5d994c4aa209480f7cd8e2f3655737d77b132b
User           : root <root>
Return-Code    : Success
Command Line   : -y update
```

9.3.3 怎样配置 Linux 分析工具：atop 和 kdump

atop 配置

atop 简介

- atop 是一款功能非常强大的 Linux 服务器监控工具。它能定期记录系统运行状态，并采集系统资源（如 CPU、内存、磁盘、网络）的使用情况，同时还会监控进程的运行情况。通过将这些数据保存为日志文件的方式，我们可以在服务器出现问题时获取相应的 atop 日志文件进行详细分析。无论是查看系统资源的使用情况，还是追踪特定进程的运行状况，atop 都能提供有用的信息。
- 使用 atop 工具，系统管理员可以及时发现服务器的性能瓶颈、异常情况以及资源利用率等问题，从而更好地管理和优化服务器的运行。

CentOS 8 系列操作系统 atop 配置方法

1. 通过远程登录或本地登录的方式登录弹性云主机。

2. 下载 atop 安装包。命令如下。

```
# wget https://www.atoptool.nl/download/atop-2.6.0-1.el8.x86_64.rpm
```

3. 执行以下命令安装 atop。

```
# rpm -ivh atop-2.6.0-1.el8.x86_64.rpm
```

4. 编辑配置文件，修改采样周期，atop 配置文件为/etc/default/atop。

```
# vi /etc/default/atop
```

可以根据自己的实际需求修改如下配置参数，修改后保存并退出。

```
LOGINTERVAL=15LOGGENERATIONS=7LOGPATH=/var/log/atop/
```

参数说明：

LOGINTERVAL：监控周期，单位为秒，默认 600s 采集一次数据，建议修改为 15s。

LOGGENERATIONS：日志保留时间，单位为天，默认保留时间为 28 天，建议修改为 7 天，避免大量占用磁盘空间。

LOGPATH：日志保存路径。默认路径为/var/log/atop/。

5. 重启 atop 服务。

```
# systemctl restart atop
```

6. 可以设置 atop 服务自启动（根据实际情况设置）。

```
# systemctl enable atop
```

7. 检查是否启动成功，active(running)表示运行正常。

```
# systemctl status atop
```

CTyunOS 、CentOS 7、EulerOS、Anolis、KylinOS 系列操作系统 atop 配置方法

1. 通过远程登录或本地登录的方式登录弹性云主机。

下载 atop 安装包。

```
# wget https://www.atoptool.nl/download/atop-2.6.0-1.el7.x86_64.rpm
```

1.

3.执行以下命令安装 atop。

```
# rpm -ivh atop-2.6.0-1.el7.x86_64.rpm --nodeps
```

4.编辑配置文件，修改采样周期，atop 配置文件为/etc/default/atop。

```
# vi /etc/default/atop
```

可以根据自己的实际需求修改如下配置参数，修改后保存并退出。

```
LOGINTERVAL=15LOGGENERATIONS=7LOGPATH=/var/log/atop/
```

参数说明：

LOGINTERVAL：监控周期，单位为秒，默认 600s 采集一次数据，建议修改为 15s。

LOGGENERATIONS：日志保留时间，单位为天，默认保留时间为 28 天，建议修改为 7 天，避免大量占用磁盘空间。

LOGPATH：日志保存路径。默认路径为/var/log/atop/。

5.重启 atop 服务。

```
# systemctl restart atop
```

6.可以设置 atop 服务自启动（根据实际情况设置）。

```
# systemctl enable atop
```

7.检查是否启动成功，active(running)表示运行正常。

```
# systemctl status atop
```

使用源码方式安装（适用于 Fedora、Debian、Ubuntu 等系列操作系

1.下载 atop 源码。

```
# wget https://www.atoptool.nl/download/atop-2.6.0.tar.gz
```

2.执行以下命令解压源码 atop。

```
# tar -zxvf atop-2.6.0.tar.gz
```

3.执行以下命令查看 systemctl 版本。

```
# systemctl  -version
```

如果版本大于等于 220，直接进行下一步。

否则需要修改 atop 的 Makefile 文件，删除--now 参数。

```
# vi atop-2.6.0/Makefile
```

删除 systemctl 命令后的--now 参数。

4.安装编译 atop 依赖软件包。

Fedora 系列操作系统执行以下命令安装：

```
# yum install make gcc zlib-devel ncurses-devel -y
```

Debian9、Debian10、Ubuntu 系列操作系统执行以下命令安装：

```
# apt install make gcc zlib1g-dev libncurses5-dev libncursesw5-dev  -
y
```

5.执行以下命令编译并安装 atop。

```
# cd atop-2.6.0# make systemdinstall
```

6.编辑配置文件，修改采样周期，atop 配置文件为/etc/default/atop。

```
# vi /etc/default/atop
```

7.可以根据自己的实际需求修改如下配置参数，修改后保存并退出。

```
LOGINTERVAL=15LOGGENERATIONS=7LOGPATH=/var/log/atop/
```

8.重启 atop 服务。

```
# systemctl restart atop
```

可以设置 atop 服务自启动（根据实际情况设置）。

```
# systemctl enable atop
```

9.检查是否启动成功，active(running)表示运行正常。

```
# systemctl status atop
```

atop 结果分析

1.atop 启动后，会将采集到的记录默认存放在/var/log/atop 目录中，通过执行如下命令，查看日志文件。

```
atop -r /var/log/atop/atop_20230821
```

备注：可通过 ls /var/log/atop/查看所有生成的日志文件，atop_20230821 为其中的一个日志文件。

atop 常用指令如下所示：

c：按照进程 CPU 使用率进行降序筛选。

m：按照进程内存使用率进行降序筛选。

d：按照进程磁盘使用率进行降序筛选。

a：按照进程资源综合使用率进行降序筛选。

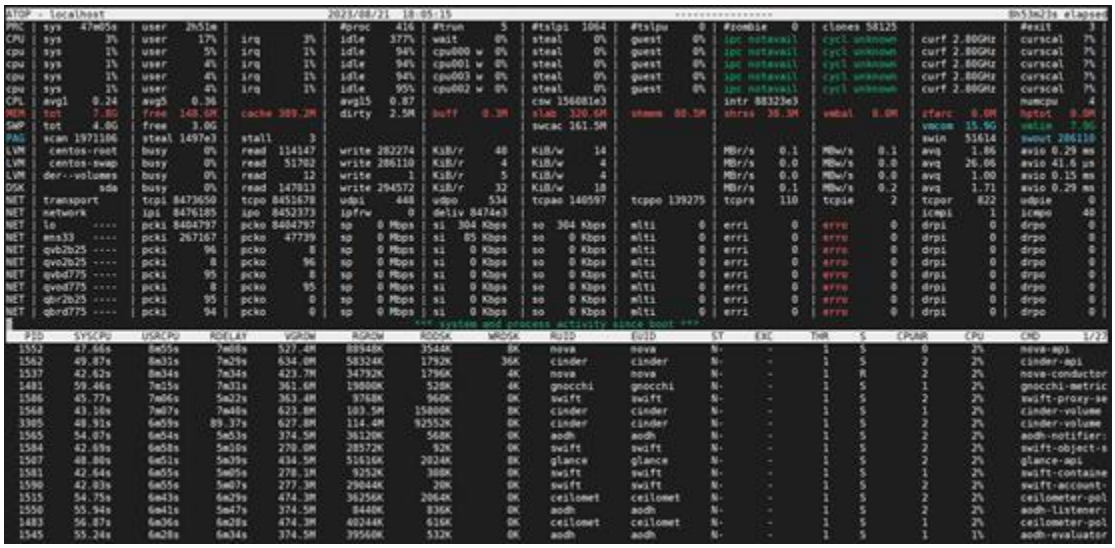
n：按照进程网络使用率进行降序筛选，需要额外安装内核模块才支持，默认不支持。

t：跳转到下一个监控采集点。

T：跳转到上一个监控采集点。

b：指定时间点。

2.系统资源监控字段说明。



ATOP 行：主机名、采样日期和时间点。

PRC 行：进程整体运行情况。

sys 及 user：进程在内核态及用户态所占 CPU 的时间比例。

#proc：进程总数。

#zombie：僵死进程的数量。

#exit：atop 采样周期期间退出的进程数量。

CPU 行：CPU 整体（即多核 CPU 作为一个整体 CPU 资源）的使用情况。【各字段数值相加结果为 N00%，N 为 CPU 的核数】

sys 及 user：CPU 被用于处理进程时，进程在内核态、用户态所占 CPU 的时间比例。

irq：CPU 被用于处理中断的时间比例。

idle：CPU 处在完全空闲状态的时间比例。

wait：CPU 处在“进程等待磁盘 IO 导致 CPU 空闲”状态的时间比例。

CPL 行：CPU 负载情况。

avg1、avg5 和 avg15：过去 1 分钟、5 分钟和 15 分钟内运行队列中的平均进程数量。

csw：指示上下文交换次数。

intr：指示中断发生次数。

MEM 行：内存的使用情况。

tot：物理内存总量。

cache：用于页缓存的内存大小。

buff：用于文件缓存的内存大小。

slab：系统内核占用的内存大小。

SWP 行：交换空间的使用情况。

tot：交换区总量。

free：空闲交换空间大小。

PAG 行：虚拟内存分页情况

swin 及 swout：换入和换出内存页数。

DSK 行：磁盘使用情况，每一个磁盘设备对应一列。如果有 sdb 设备，那么增加一行 DSK 信息。

sda：磁盘设备标识。

busy：磁盘忙时比例。

read 及 write：读、写请求数量。

NET 行：多列 NET 展示了网络状况，包括传输层（TCP 和 UDP）、IP 层以及各活动的网口信息。

XXXi：各层或活动网口收包数目。

XXXo：各层或活动网口发包数目。

kdump 配置

kdump 简介

- kdump 是 Linux 系统中一种用于保存系统崩溃时的内存转储（crash dump），以便进行故障诊断和调试的工具。
- 在 Linux 系统中，如果遇到严重的内核崩溃或死锁等问题，通常会导致系统无法继续正常运行。为了解决这种情况，kdump 允许在系统崩溃时将当前内核的内存转储保存到硬盘上。kdump 通过一些特定的配置和设置，可以在系统崩溃时触发一个独立的内核（用作专用的 crash 内核），该内核能够将系统的内存转储保存到预定义的目录中。这个转储文件可以记录包括内核代码、进程状态、内核堆栈跟踪等信息，使开发人员和系统管理员能够在崩溃发生时进行故障分析和调试。
- kdump 是一种非常有用的工具，它可以帮助我们更有效地诊断系统崩溃问题，定位和解决导致系统异常的根本原因。

- kdump 使用须知

1. 系统设置

要使用 kdump 功能，您的系统必须满足一些硬件和配置要求：

- 必须有足够的磁盘空间来存储转储文件。
- 必须有足够的内存来容纳专用的 crash 内核。
- CPU 必须支持物理地址扩展（PAE）或 64 位寻址模式。

2. 内核配置：您的 Linux 内核必须正确配置以启用 kdump 功能。通常，许多发行版都已经为您设置了合适的内核配置。

3. 专用的 crash 内核：为了执行转储过程，kdump 需要一个独立的 crash 内核。这个 crash 内核是一个精简的内核，只包含用于内存转储和调试的必要组件。

4. 存储位置：您需要选择一个足够大且可靠的存储位置来保存转储文件。这可能是本地磁盘、NFS 共享目录或 iSCSI 设备等。

5. 网络配置（可选）：如果您选择将转储文件发送到远程服务器进行分析，则需要配置网络连接和远程服务器的接收端。

kdump 配置方法

以 EulerOS 以及 CentOS 7/8 系列 Linux 产品为例：

1. 查看是否已安装 kexec-tools。

```
# rpm -qa | grep kexec-tools
```

如果没有安装，执行下面命令安装 kexec-tools。

```
# yum install -y kexec-tools
```

2. 设置 kdump 开机启动。

```
# systemctl enable kdump
```

3. 设置 crashkernel 参数，设置这个参数的目的是预留内存给捕获内核（capture kernel）。

首先查看该参数是否已经设置。

```
# grep crashkernel /proc/cmdline
```

如果有显示，则表示已经设置，如果没有显示，则需要重新设置。

编辑/etc/default/grub 文件，此文件通常包含如下内容：

```
GRUB_TIMEOUT=5GRUB_DEFAULT=savedGRUB_DISABLE_SUBMENU=trueGRUB_TERMINAL_OUTPUT="console"GRUB_CMDLINE_LINUX="crashkernel=auto
```

```
rd.lvm.lv=rhel00/root rd.lvm.lv=rhel00/swap rhgb
quiet"GRUB_DISABLE_RECOVERY="true"
```

在 GRUB_CMDLINE_LINUX 一行添加 crashkernel=auto。

4.更新 grub，执行更新 grub 命令，使配置生效。命令如下。

```
# grub2-mkconfig -o /boot/grub2/grub.cfg
```

5.打开/etc/kdump.conf 文件中找到“path”参数，添加以下内容。

```
path /var/crash
```

默认是保存在/var/crash 目录下，如果要保存到其他目录，则改成对应的目录，例如保存在/home/kdump 下，则改成：

```
path /home/kdump## 请确保指定的路径有足够的空间保存 vmcore，建议剩余空间不小于物理内存（RAM）的大小，也可以保存在 SAN，NFS 等共享设备上。
```

6.设置转存 vmcore 级别，查看/etc/kdump.conf 文件，是否存在以下设置，如果存在则无需添加。

```
core_collector makedumpfile -d 31 -c
```

-c：表示压缩 vmcore 文件，

-d：表示过滤掉部分无效的内存数据，可以根据需要调整，一般 31 即可，31 是由如下的值与计算而成。

zero pages= 1

cache pages = 2

cache private = 4

user pages= 8

free pages= 16

7.设置内核参数（可选）。

为了控制在哪些场景下触发 kdump，内核提供了一些参数，建议设置以下参数：

kernel.hardlockup_panic=1

kernel.panic=5

kernel.panic_on_oops=1

kernel.softlockup_panic=1

kernel.unknown_nmi_panic=1

kernel.nmi_watchdog=1

将以上配置参数写入/etc/sysctl.conf 文件并保存。另外，可以选择添加如下几个参数到/etc/sysctl.conf 文件中。

kernel.panic_on_io_nmi=1

kernel.panic_on_warn=1

8.进入天翼云控制台-云主机详情页，对当前云主机进行重启，使以上配置生效。

Kdump 配置生效验证

1.执行以下命令，确认回显信息中 crashkernel=auto。

```
# cat /proc/cmdline |grep crashkernel
```

2.执行以下命令，并确认回显信息中的配置信息正确。

```
# grep core_collector /etc/kdump.conf |grep -v ^"#"

core_collector makedumpfile -l --message-level 1 -d 31
```

3.执行以下命令，并确认回显信息中的配置信息正确。

```
# grep path /etc/kdump.conf |grep -v ^"#"

path /var/crash
```

4.执行以下命令，并确认回显信息中的 Active 的状态为 active (exited)。

```
# systemctl status kdump
```

5.执行测试命令。

```
# echo c > /proc/sysrq-trigger
```

这会触发 kdump，重新启动，并将生成的 vmcore 文件保存的 path 参数指定的位置。

6.检查 vmcore 是否生成。

到所在环境 path 参数所指定的路径查看是否有 vmcore 文件生成，例如/var/crash/目录。

```
# ll /var/crash/
```

可以看到生成了一个文件夹，里面有 vmcore 文件

9.3.4Linux 操作系统云主机安装 Docker 支持的操作系统版本

Docker 可以运行在多种 Linux 操作系统云主机上，但 Docker 的支持情况随时间和版本变化有所不同。无论选择哪个操作系统版本，需要根据具体的操作系统版本和 Docker 版本来确保满足安装要求，并在安装 Docker 之前查阅 Docker 官方文档获取最新的安装指南和版本要求。下表仅列举一些常见的 Linux 云主机操作系统，以及最新版 Docker 安装所要求的操作系统版本，如下表所示。

操作系统	操作系统版本	Docker 官方安装指南地址
CentOS	CentOS 7 CentOS 8 (stream) CentOS 9 (stream)	https://docs.docker.com/install/linux/docker-ce/centos/
Debian	Debian Bookworm 12 (stable) Debian Bullseye 11 (oldstable)	https://docs.docker.com/install/linux/docker-ce/debian/

操作系统	操作系统版本	Docker 官方安装指南地址
Fedora	Fedora 37 Fedora 38	https://docs.docker.com/install/linux/docker-ce/fedora/
Ubuntu	Ubuntu Lunar 23.04 Ubuntu Kinetic 22.10 Ubuntu Jammy 22.04 (LTS) Ubuntu Focal 20.04 (LTS)	https://docs.docker.com/install/linux/docker-ce/ubuntu/

9.3.5Ubuntu 操作系统如何设置默认启动内核

问题描述

1.Ubuntu 操作系统弹性云主机在默认启动时选择的内核版本与期望使用的内核版本不一致；
2.Ubuntu 操作系统弹性云主机中存在比下载的指定内核版本更高的版本，会自动成为默认内核版本，无法启动到指定内核。

操作步骤

1.查看当前内核版本。

```
uname -r
```

```
root@ecm-3216:~# uname -r
5.4.0-148-generic
```

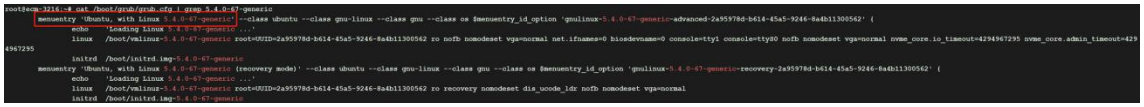
2. 列出系统中已安装的内核列表。

```
dpkg --list | grep linux-image
```

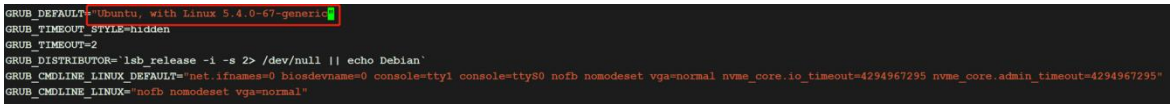
```
root@ecm-3216:~# dpkg --list | grep linux-image
ii  linux-image-5.4.0-148-generic         5.4.0-148.165                   amd64        Signed kernel image generic
ii  linux-image-5.4.0-67-generic          5.4.0-67.75                     amd64        Signed kernel image generic
ii  linux-image-generic                   5.4.0.148.146                   amd64        Generic Linux kernel image
```

3. 选择想要默认启动的内核版本版本号，通过以下命令找到其对应的菜单名称（menuentry）。

```
cat /boot/grub/grub.cfg | grep 5.4.0-67-generic
```

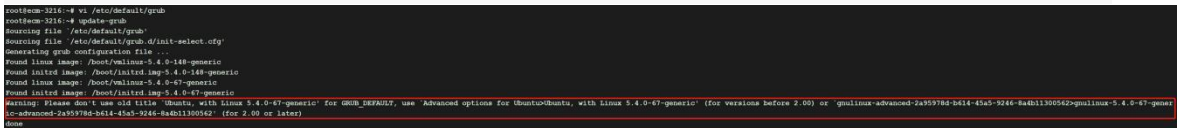


4. 打开/etc/default/grub 文件，将 GRUB_DEFAULT 的值修改为上一步获得的内核名称。



5. 更新 grub 配置，重新生成 grub 配置文件。

```
update-grub
```

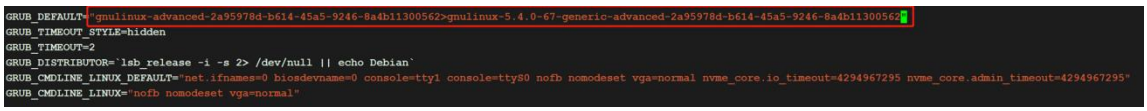


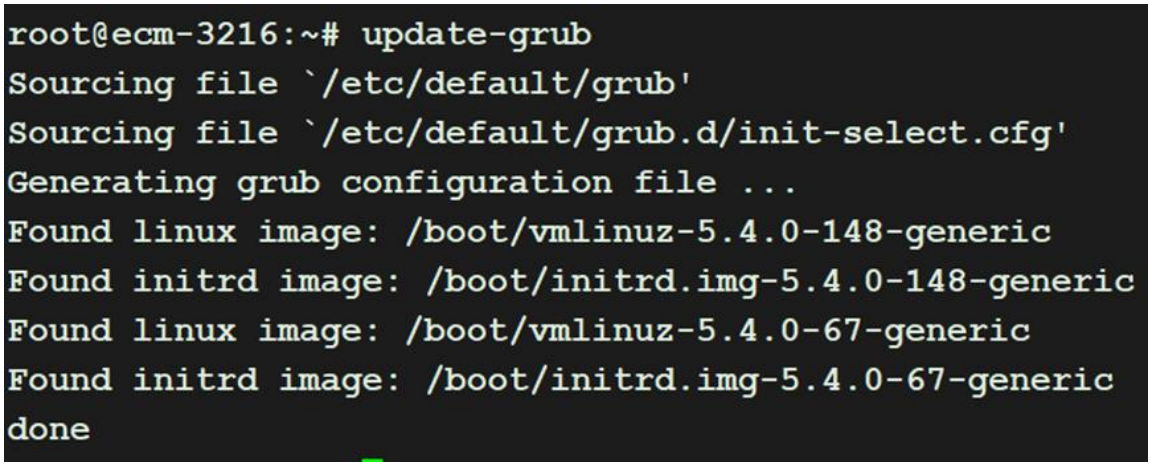
6. 根据 warning 部分提示，使用如下命令查看 grub 版本。

```
grub-install --version
```

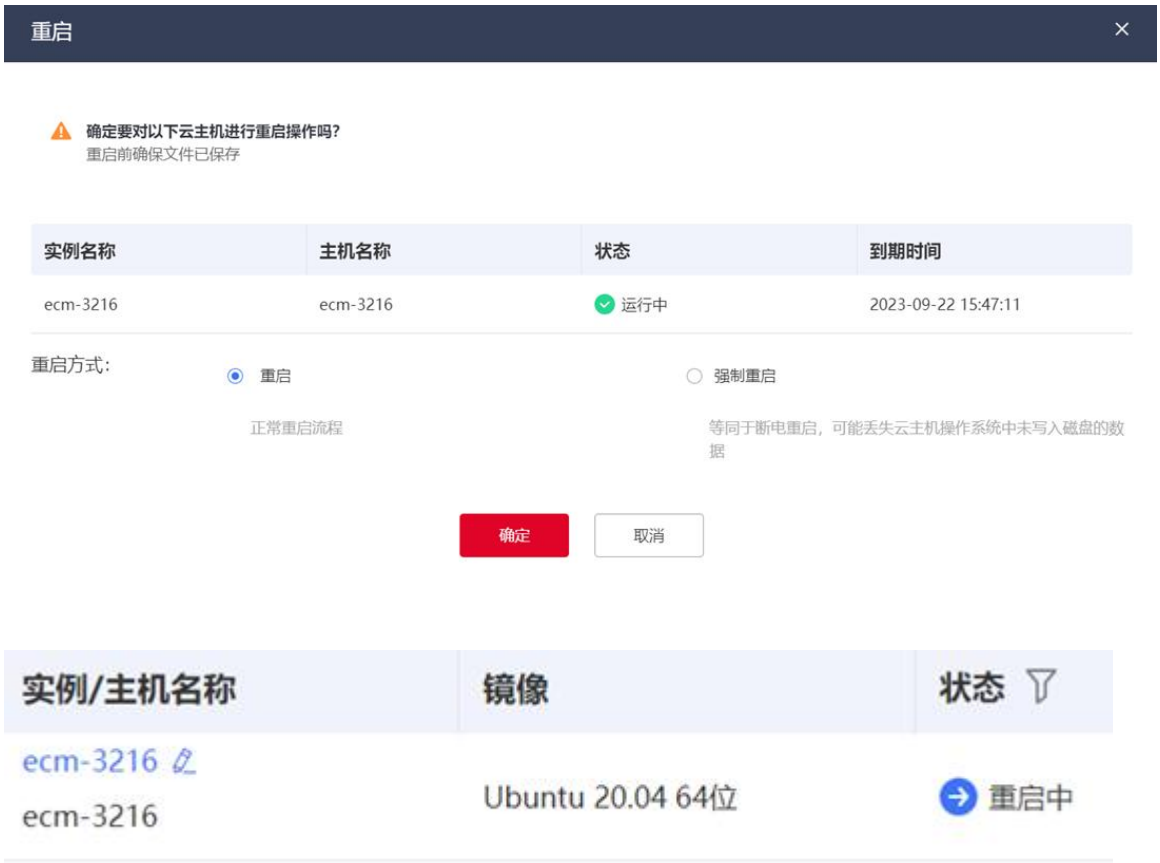
```
root@ecm-3216:~# grub-install --version
grub-install (GRUB) 2.04-1ubuntu26.16
```

7. 根据步骤 6 的版本号，选择步骤 5 warning 部分中适当的内核名称重复步骤 4、5。





8. 重新启动 Ubuntu 系统弹性云主机，查看默认启动内核版本已修改。



```
root@ecm-3216:~# uname -r
5.4.0-67-generic
```

9.3.6 弹性云主机 IP 地址丢失处理方法

异常结果示例如下：

```
ctyun@00bafcjc-nba6lbmx3k:~$ ip ad
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: ens3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1450 qdisc fq_codel state UP group default qlen 1000
    link/ether fa:16:3e:14:0d:c1 brd ff:ff:ff:ff:ff:ff
    inet6 fe80::a4e9:a632:73cb:52a6/64 scope link noprefixroute
       valid_lft forever preferred_lft forever
```

正常结果示例如下：

```
ctyun@00bafcjc-nba6lbmx3k:~$ ip ad
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: ens3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1450 qdisc fq_codel state UP group default qlen 1000
    link/ether fa:16:3e:14:0d:c1 brd ff:ff:ff:ff:ff:ff
    inet 172.16.2.101/24 brd 172.16.2.255 scope global dynamic noprefixroute ens3
       valid_lft 82961sec preferred_lft 82961sec
    inet6 fe80::a4e9:a632:73cb:52a6/64 scope link noprefixroute
       valid_lft forever preferred_lft forever
```

故障原因

1.网络服务故障：由于某些进程影响或用户操作，网络服务可能会崩溃或停止工作，导致 IP 地址丢失。

2.DHCP 问题：弹性云主机可能存在与 DHCP 服务器通信的问题。其中可能包括 DHCP 服务器故障、DHCP 租约到期或 DHCP 配置错误。

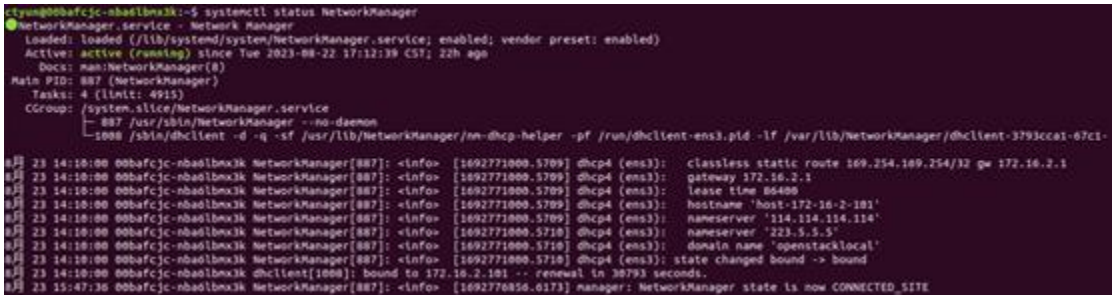
解决步骤

1.重新启动网络服务：尝试重启网络服务以重新获取 IP 地址并设置为开机自启动。

```
systemctl restart NetworkManagersystemctl enable NetworkManager
```

2.查看 NetworkManager 服务状态，确认 NetworkManager 服务重启成功。

```
systemctl status NetworkManager
```



3.执行以下命令，向 DHCP 服务器发送请求以获取新的 IP 地址。

```
dhclient
```

1.

如果重启 NetworkManager 服务后，弹性云主机与 DHCP 服务器通信正常，则会提示 File exists。

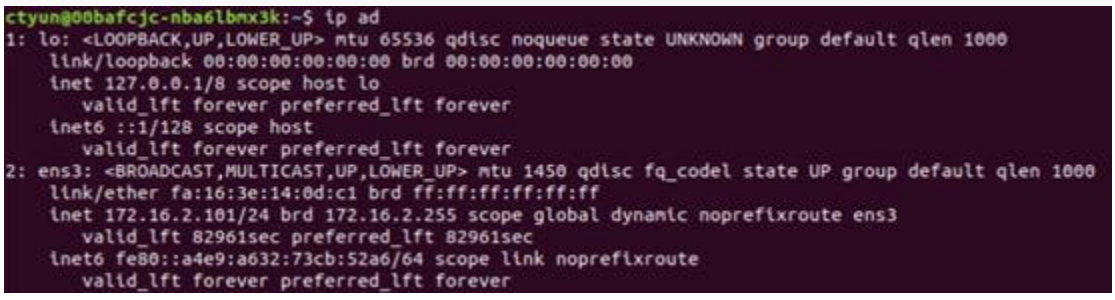
```
ctyun@00bafcjc-nba6lbmx3k:~$ sudo dhclient
RTNETLINK answers: File exists
```

如果重启 NetworkManager 服务后，弹性云主机与 DHCP 服务器通信不正常，则会向 DHCP 服务器通信再次发送请求，正常情况下不再提示。

```
ctyun@00bafcjc-nba6lbmx3k:~$ sudo dhclient
ctyun@00bafcjc-nba6lbmx3k:~$
```

4.执行以下命令，查看网络状态，确认弹性云主机 ip 正常。

```
ip ad
```



9.3.7 内核参数 kernel.unknown_nmi_panic 配置错误导致 Linux 弹性云主机异常重启

问题描述

使用 Linux 操作系统的弹性云主机发生异常重启事件，错误提示如下：

```
Kernel panic - not syncing: NMI: Not continuing
```

同时，内核日志输出如下信息：

```
[…] Uhhuh. NMI received for unknown reason 20 on CPU 1.[…] Do you have a strange power saving mode enabled?[…] Kernel panic - not syncing: NMI: Not continuing
```

可能原因

如果 Linux 操作系统弹性云主机的内核参数 kernel.unknown_nmi_panic 配置为 1，则云主机实例会在系统收到 NMI 中断时主动 panic 导致系统重启。NMI 中断（Non-Maskable Interrupt，非可屏蔽中断）是操作系统中的一种特殊类型的中断，它具有比其他中断更高的优先级，并且通常无法被屏蔽或忽略。因此 NMI 中断通常用于处理一些关键的系统事件，例如硬件故障、系统崩溃、内存校验错误等，这些事件需要立即引起处理，而不能等待其他正在进行的任务完成。内核参数 kernel.unknown_nmi_panic 配置为 1 的目的是为了通过触发 NMI 中断主动通过 panic 系统进行问题定位。不过由于个别型号的 CPU 会在正常业务流程中同样产生 NMI 中断，如果在这种情况下仍将内核参数 kernel.unknown_nmi_panic 配置为将会触发云主机产生非预期的异常重启事件。

解决方案

1.远程登录弹性云主机。
2.执行以下命令，查看该云主机关于内核参数 kernel.unknown_nmi_panic 的配置值

```
sysctl -n kernel.unknown_nmi_panic
```

如果该参数的取值为 1，说明是由于内核参数配置错误引起的云主机异常重启。

```
[root@centos ~]# sysctl -n kernel.unknown_nmi_panic1
```

3.在系统的/etc/sysctl.conf 文件中查看是否存在内核参数 kernel.unknown_nmi_panic 相关配置。如果存在 kernel.unknown_nmi_panic=1 的配置，将该配置改为 kernel.unknown_nmi_panic=0；如果不存在 kernel.unknown_nmi_panic=1 的配置，则在该文件中增加 kernel.unknown_nmi_panic=0 的配置信息。

4.执行 systecl -p 命令使配置生效。

```
[root@centos ~]# sysctl -pkernel.unknown_nmi_panic = 0
```

结果验证

执行以下两个命令，查看结果是否均为 0。如果验证结果符合预期，则说明修改成功。

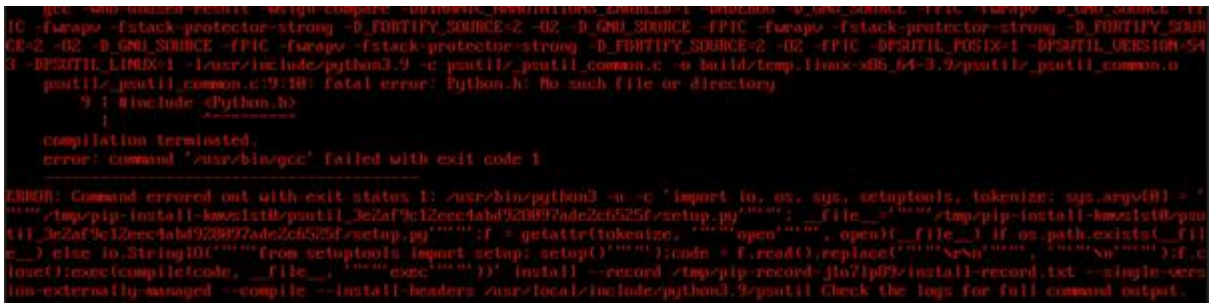
```
[root@centos~]#cat /proc/sys/kernel/panic_on_unrecovered_nmi0

[root@centos ~]# sysctl -n kernel.unknown_nmi_panic0
```

9.3.8pip 安装软件时出现错误：command 'gcc' failed with exit status 1

问题描述

在通过 yum 或 rpm 命令确认已经安装了 gcc 的前提下，通过 pip 安装指定 Python 库时会出现报错：command 'gcc' failed with exit status 1。



问题原因

系统缺少 gcc 编译需要的 openssl-devel、python-devel 等系统包支持。

解决步骤

以安装 Python 库 psutil 的5.4.3 版本为例：

1.通过执行如下命令，安装 openssl-devel 和 python-devel，同时安装相关依赖。

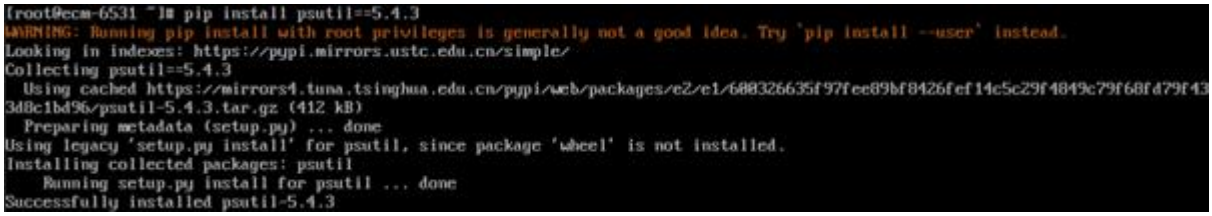
```
yum install -y openssl-devel libffi-devel gcc

yum install -y python-devel
```

若您使用的是 Python3,则可能需要执行如下命令代替上述命令中的第 2 行（即 python3-devel 替换 python-devel）：

```
yum install -y python3-devel
```

2.再次通过 pip 安装，可以看到安装成功，没有报错。



9.3.9CentOS 云主机根目录设置成 777 权限后如何恢复？

问题描述

当您不小心执行了 chmod -R 777 /命令，导致 CentOS 云主机根目录以及下边的所有文件权限均被设置成了对所有用户都是可读、可写和可执行的。本文操作介绍如何修复该问题，但为了避免安全风险，建议您在进行完操作后立即备份数据并重装系统。

操作步骤

以下操作在 CentOS8.0 操作系统中进行了测试验证。

1.不要退出故障云主机，在故障云主机上执行以下命令，修改 ssh 以及 su 相关的文件权限。

```
cd /etcchmod 644 passwd group shadowchmod 400 gshadowcd /etc/sshchmod 600 moduli ssh_*_keychmod 644 ssh_config ssh_*.pubchmod 640 -R sshd_config ssh_config.dchmod 711 /var/empty/sshdchmod u+s 'which su'
```

2.执行 systemctl status sshd 查看 sshd 状态，如果为错误状态，请查看错误原因，修改相关文件权限，直到 sshd 状态正常，然后进入下一步。

3.创建一台权限正常的临时云主机：Linux 操作系统，内核版本与故障云主机相同。

4.执行以下命令，将所有文件的权限记录下来。

```
getfacl -R / >systemp.bak
```

5.使用 scp 命令将 systemp.bak 文件上传到故障云主机中，或者在故障云主机中下载该文件。

6.在故障云主机中进行权限恢复操作。

```
setfacl --restore=systemp.bak
```

7.执行 reboot 命令重启操作系统。绝大多数系统文件的权限已经被恢复，但为了安全，请及时备份数据并重装系统。

9.3.10Linux 弹性云主机执行命令或启动服务时出现错误：Cannot allocate memory

故障描述

在 Linux 操作系统的弹性云主机上执行命令 top, free -m, df -h 等命令时出现错误提示：-bash: fork: Cannot allocate memory 报错。

故障原因

通常是因为内存不足，或者达到最大 pid 数量限制。

故障排查思路

登录 Linux 操作系统弹性云主机。

多次尝试使用 free -m 或者 top 命令查看当前内存使用情况。若可用内存很少，可根据业务需要释放一些非必要的进程，或者增加 swap 空间。

若可用内存充足，则查看操作系统最大 pid 数量限制，命令如下：

```
sysctl -a | grep pid_max
```

或者

```
cat /proc/sys/kernel/pid_max
```

查看当前进程数，命令如下：

```
ps -eLf | wc -l
```

若 pid_max 的值偏小时会影响 pid 的创建，可以根据业务需要调整最大 pid 数量限制，临时调整的命令如下：

```
sysctl -w kernel.pid_max=65535
```

若要弹性云主机重启后调整命令仍然生效，可执行如下命令：

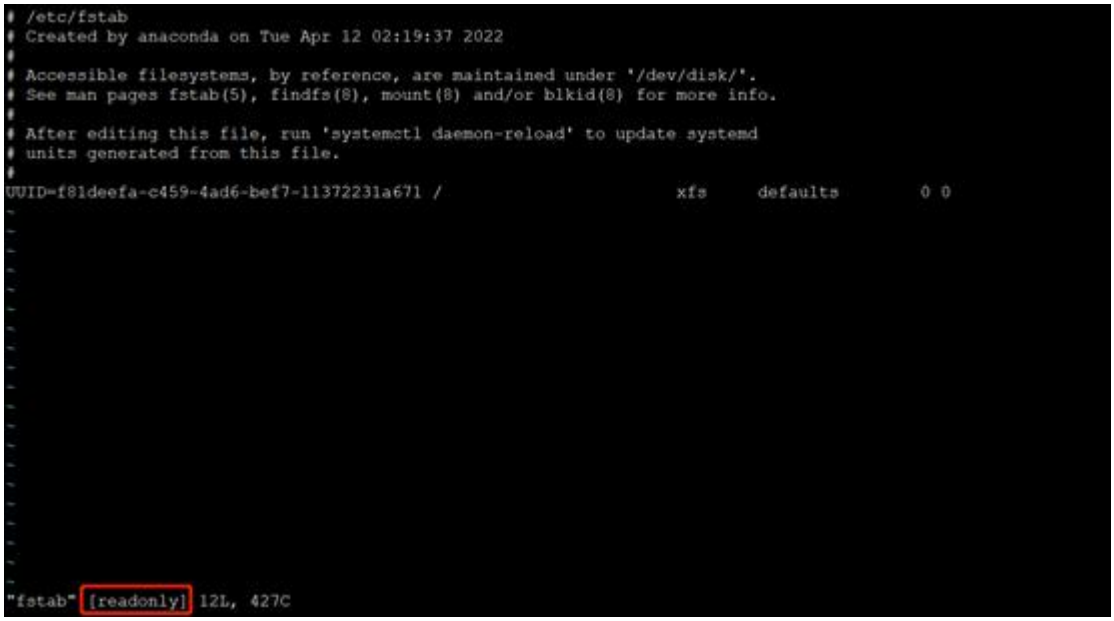
```
echo "kernel.pid_max=65535" >> /etc/sysctl.conf
```

9.3.11 无法编辑 fstab 文件怎么办

问题描述

编辑/etc/fstab 文件，提示“readonly”。

图 1 fstab 文件提示 readonly



处理方法

修复模式下（read-only system）文件是被保护的不能修改，运行下面命令，把系统文件权限改成可读写（rw）。

```
mount -o remount,rw /
```

执行完上述命令，重试编辑 fstab 文件。

9.3.12 非 root 用户切 root 用户时，连接超时怎么办？

问题描述

针对镜像为 CentOS、Ubuntu 的弹性云主机，使用 sudo 命令切换 root 用户时，系统提示连接超时。

操作步骤

r　oot 用户切换超时可能由于 host 配置问题引起，请检查/etc/hosts 文件，正确配置 IP 地址和主机名。

1.登录弹性云主机。

2.编辑 host 文件：

在/etc/hosts 文件最后写入 IP 地址和主机名，格式为：

私有 IP 地址 主机名

比如弹性云主机的主机名为“my_ecs_hostname”，私有 IP 地址为“192.168.0.1”，则在 host 文件最后写入：

192.168.0.1 my_ecs_hostname

3.保存并退出文件。

4.切换 root 用户测试问题是否解决。

故障描述

使用获取弹性云主机 IP 地址的命令获取结果异常。命令如下。

```
ip ad
```

9.3.13 修改/etc/security/limits.conf 文件，重启后不生效怎么办？

可能原因

在 Linux 操作系统云主机中，/etc/security/limits.conf 配置文件主要用于设置系统中用户或用户组的资源限制。可在配置文件中指定用户、用户组、以及资源限制的具体值。该文件中的配置会影响到整个系统范围内的用户或进程。但这些配置会受到/etc/security/limits.d/ 目录中的配置的影响，因为/etc/security/limits.d/目录中的配置优先级高于/etc/security/limits.conf。

如果遇到修改/etc/security/limits.conf 文件，重启后配置项不生效问题，可能原因是/etc/security/limits.d/目录中的配置项覆盖了/etc/security/limits.conf 文件中的值。

解决步骤

1.进入/etc/security/limits.d/目录内，查看是否有配置文件内容与/etc/security/limits.conf 文件中修改的配置项冲突。如果有则尝试修改这些文件中的配置，或者删除冲突的文件。
2.如果想要在/etc/security/limits.d/目录中创建新的配置文件来配置资源限制，确保没有在多个配置文件中重复定义相同的配置项。同时可用数字前缀来指定配置文件的加载顺序，确保正确的加载顺序。
3.在确认配置项已经全部正确设置后，需重新启动系统以确保新的配置生效。
4.如果问题仍然存在，检查位于/var/log 目录下的系统日志，获取关于配置项加载冲突等更多的信息。

9.3.14 使用 taskset 命令让进程运行在指定 CPU 上

操作描述

taskset 命令可用于在 Linux 系统上查看或设定某个进程或线程的 CPU 亲和性。使用此命令可实现让云主机内的某个进程或线程仅在与之绑定的 CPU 核心上运行。此操作指导将以 CtyunOS 3-23.01 64 位操作系统为例，介绍如何使用 taskset 命令让进程运行在指定 CPU 上。

操作步骤

1.通过执行如下命令，查看云主机的 CPU 核数。

```
cat /proc/cpuinfo
```

您会看到类似下图的输出内容。



其中，“processor”参数表示对应 CPU 处理器标号，而“cpu cores”参数表示对应 CPU 处理器的核心数。上图示例展示了标号为 1 且核心数为 1 的 CPU 处理器。

注意：从此图可判断此云主机至少有 2 个 CPU 处理器，因为 CPU 处理器标号是从 0 开始的，了解这个信息对后续步骤的执行很关键。

2. 假设有一个 taskset_test.sh 进程，通过执行如下命令，查看该进程的信息。

```
ps aux | grep taskset_test.sh
```

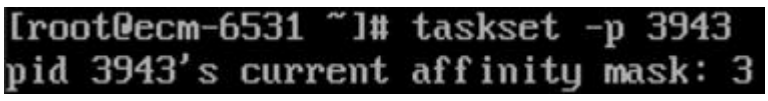


上图示例中 taskset_test.sh 进程的 PID 是 3943。

3. 通过执行如下命令，查看指定进程的 CPU 亲和性。

```
taskset -p <替换为 PID>
```

以第 2 步中获取的 PID 为例：



上图示例中显示的十进制数字 3 转换为二进制数字 11，每个 1 对应一个 CPU，最低两个是 1 表示该进程运行在两个 CPU 上。

4. 通过执行如下命令，设定指定进程运行在第 1 个 CPU（CPU0）上。

```
taskset -pc 0 <替换为 PID>
```

以第 2 步中获取的 PID 为例：

```
[root@ecm-6531 ~]# taskset -pc 0 3943
pid 3943's current affinity list: 0,1
pid 3943's new affinity list: 0
```

可以重复第 3 步来对比变化：

```
[root@ecm-6531 ~]# taskset -p 3943
pid 3943's current affinity mask: 1
```

上图示例中显示的十进制数字 1 转换为二进制数字 01（与第 3 步中对应），最低一个是 1 表示此步设定成功。

此外，仍以设定运行在第 1 个 CPU 上为例，可使用如下命令在启动程序时绑定 CPU 核心：

```
taskset -c 0 bash taskset_test.sh
```

9.3.15Web 访问超时且系统日志打印：nf_conntrack: table full, dropping packet

故障描述

客户端访问 Web 时出现超时(time out)；服务端的系统日志 /var/log/messages 打印“kernel: nf_conntrack: table full, dropping packet”。

故障原因

iptables 的 connection-tracking 模块使用一部分系统内存来跟踪表中的连接。系统日志打印的“nf_conntrack: table full, dropping packet”表明连接跟踪表 conntrack 已满，无法为新连接创建条目并记录跟踪，进而出现“dropping packet”丢包问题。

解决步骤

解决方案是为连接跟踪表增加条目容量。

CentOS 6 系列操作系统解决步骤

1.查看连接跟踪表 conntrack 的容量参数 nf_conntrack_max。执行以下命令：

```
sysctl net.netfilter.nf_conntrack_max
```

2.检查当前跟踪的连接数参数，如果该值达到 nf_conntrack_max 值则会出现丢包现象。执行以下命令：

```
cat /proc/sys/net/netfilter/nf_conntrack_count
```

3. 增加连接跟踪表的容量参数（以内存为 64G，net.netfilter.nf_conntrack_max 值 2097152 为例）。

4.执行以下命令，使配置即时生效。

```
sysctl -w net.netfilter.nf_conntrack_max=2097152
```

5.执行以下命令，确保系统重启后配置仍生效。

```
echo "net.netfilter.nf_conntrack_max = 2097152" >> /etc/sysctl.conf
```

6.如果 conntrack 表中容量显著增加，则还应增加存储 conntrack 的哈希表大小以提高效率。CentOS 6 及以上版本中，计算公式是 hashsize = conntrack_max/4。设置哈希大小，执行以下命令：

```
echo "options nf_conntrack expect_hashsize=524288 hashsize=524288" >> /etc/modprobe.conf
```

7.重启 iptables，执行以下命令：

```
service iptables restart
```

CentOS 7 系列操作系统解决步骤

1.在/etc/modprobe.d/firewalld-sysctls.conf 中设置 conntrack 的哈希值，执行以下命令：

```
echo "options nf_conntrack expect_hashsize=131072 hashsize=131072" >> /etc/modprobe.d/firewalld-sysctls.conf
```

2.重启 firewalld，执行以下命令：

```
systemctl restart firewalld
```

3.确认参数修改成功，执行以下命令：

```
sysctl -a |grep nf_conntrack_max
```

其他说明

- 本文操作适用于 CentOS 系统，且系统已开启主机防火墙。其他 Linux 系统可能存在差异。
- 本文操作涉及修改 CentOS 系统内核参数。在线修改内核参数可能会使内核不稳定，建议修改后重启系统。请评估风险后再操作。

注意

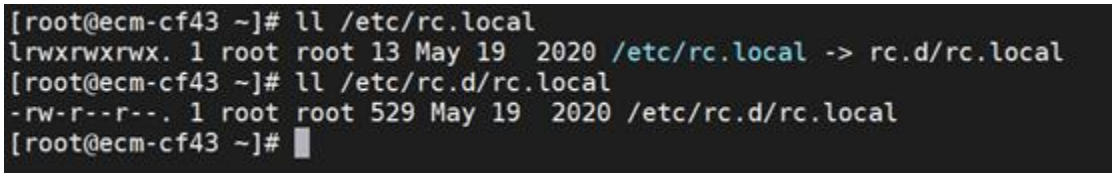
.net.netfilter.nf_conntrack_max 并不是越高越好，通常需要根据内存大小进行设置。nf_conntrack_max 计算公式（64 位）为：CONNTRACK_MAX = RAMSIZE (in bytes)/16384/2。

9.3.16 弹性云主机中/etc/rc.local 开机启动脚本不生效怎么办？

故障原因

查看/etc/rc.d/rc.local 文件是否有执行权限。

如下图可知，/etc/rc.local 是/etc/rc.d/rc.local 文件的软链接，且该文件无执行权限。



解决步骤

使用下面命令为/etc/rc.d/rc.local 添加可执行权限，并查看修改结果。

```
# chmod +x /etc/rc.d/rc.local
```

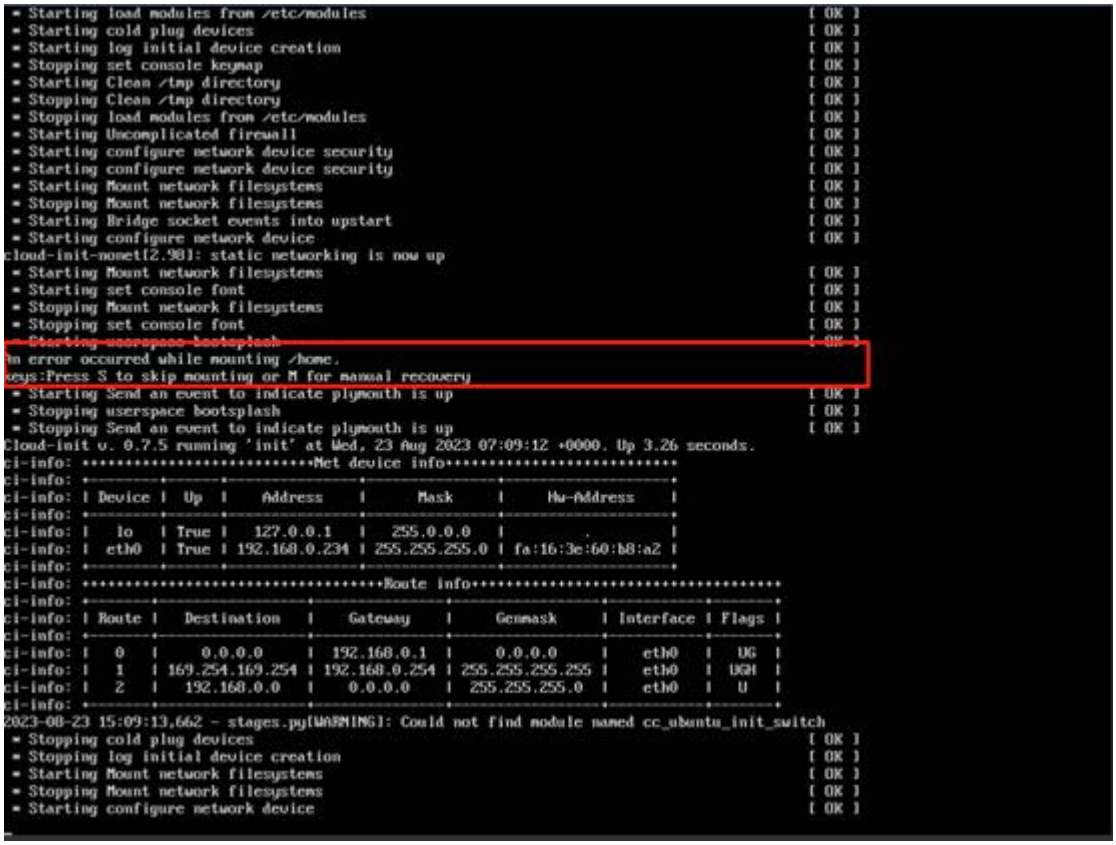
```
[root@ecm-cf43 ~]# chmod +x /etc/rc.d/rc.local
[root@ecm-cf43 ~]# ll /etc/rc.d/rc.local
-rwxr-xr-x. 1 root root 529 May 19  2020 /etc/rc.d/rc.local
[root@ecm-cf43 ~]#
```

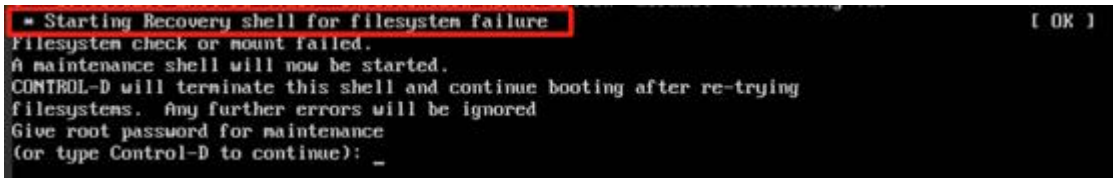
9.3.17emergency mode（紧急模式）问题处理方法

故障描述

操作系统 Ubuntu（18.04 以下版本）

启动时根据如图 1 所示的提示语“keys：Press S to skip mounting or M for manual recovery”，输入 S 直接跳过挂载故障进入弹性云主机登录系统页面，输入 M 进入如图 2 所示的紧急模式登录页面，并提示输入 root 密码进行维护。





操作系统 Centos、OpenEuler、CTyunOS2、CTyunOS3、Ubuntu（18.04 及以上版本）等

启动时进入紧急模式，提示语为“You are in emergency mode”，如图 3 所示，并提示输入 root 密码进入维护。



故障原因

紧急模式提供尽可能最小的环境，即使在系统无法进入救援模式的情况下，您也可以修复系统。在紧急模式下，系统仅安装根文件系统进行读取，不尝试安装任何其他本地文件系统，不激活网络接口，只启动一些基本服务。

进入紧急模式的原因通常是：

1./etc/fstab 文件存在错误导致挂载文件系统时失败。
2.文件系统存在错误导致。

故障排查思路

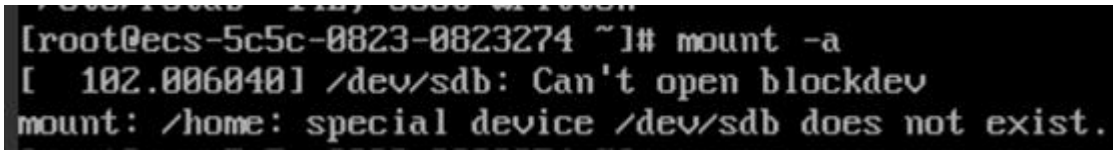
1.输入 root 密码后回车，进入紧急模式。



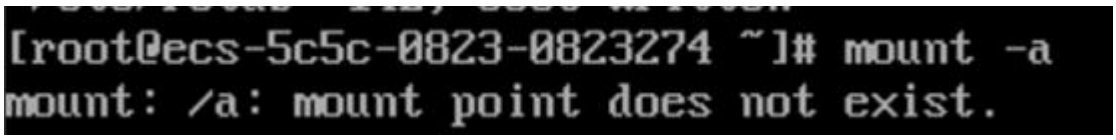
2.请执行以下命令首先检查 fstab 文件是否存在错误，尝试挂载所有未挂载的文件系统。

```
# mount -a
```

a. 如果出现如下图所示的文件系统设备不存在，请注释或删除该挂载行。



b. 如果出现如下图所示的挂载点不存在，请创建对应的挂载点。



c. 如果出现如下图所示的挂载参数错误，请修改为正确的参数。



d. 如果出现如下图所示的文件系统错误，请修复该文件系统。



1.

解决步骤

1.出现故障排查思路里的 a、b、c 现象时：请执行以下命令，打开/etc/fstab 修改相应的错误。

```
# vi /etc/fstab
```

/etc/fstab 文件包含了如下字段，通过空格分隔：

```
[file system] [dir] [type] [options] [dump] [fsck]b
```

表 1 /etc/fstab 参数说明

参数	说明
[file system]	要挂载的分区或存储设备。 [file system]列建议使用 UUID 的方式书写，执行 blkid 命令查询设备文件系统 UUID。 参考格式如下： # UUID=b411dc99-f0a0-4c87-9e05-184977be8539 /home ext4 defaults 0 2 使用 UUID 的好处在于它们与磁盘顺序无关。如果你在 BIOS 中改变了你的存储设备顺序，或是重新拔插了存储设备，或是因为一些 BIOS 可能会随机地改变存储设备的顺序，那么用 UUID 来表示将更有效。
[dir]	[file systems]的挂载位置。
[type]	挂载设备或分区的文件系统类型，支持许多种不同的文件系统：ext2，ext3，ext4，reiserfs，xfs，jfs，smbfs，iso9660，vfat，ntfs，swap 及 auto。 设置成 auto 类型，mount 命令会猜测使用的文件系统类型，对 CDROM 和 DVD 等移动设备是非常有用的。

参数	说明
[options]	挂载时使用的参数，有些参数是特定文件系统才有的。例如：defaults 参数使用文件系统的默认挂载参数，ext4 的默认参数为:rw，suid，dev，exec，auto，nouser，async。 更多参数请执行以下命令查看 man 手册：# man mount
[dump]	dump 工具通过它决定何时作备份。dump 会检查其内容，并用数字来决定是否对这个文件系统进行备份。 取值为 0 和 1。0 表示忽略，1 则进行备份。大部分的用户是没有安装 dump 的，[dump]应设为 0。
[fsck]	fsck 读取[fsck]的数值来决定需要检查的文件系统的检查顺序。 取值为 0，1，和 2。 根目录应当获得最高的优先权 1,其它所有需要被检查的设备设置为 2，0 表示设备不会被 fsck 所检查。

修改完成后，确认修改是否正确，再次执行以下命令检查 fstab 文件。

```
[root@ecs-5c5c-0823-0823274 ~]# mount -a
[root@ecs-5c5c-0823-0823274 ~]#
```

重启弹性云主机即可恢复正常。

2.出现故障排查思路里的 d 现象时

ext 文件系统 ：

请执行以下命令，检查有问题的文件系统。

```
[root@ecs-5c5c-0823-0823274 ~]# mount -a
[  184.446127] EXT4-fs (vdb): VFS: Can't find ext4 filesystem
mount: /home: wrong fs type, bad option, bad superblock on /dev/vdb, missing codepage or helper program, or other error.
[root@ecs-5c5c-0823-0823274 ~]#
```

```
[root@ecs-5c5c-0823-0823274 ~]# fsck -n /dev/vdb|head -n 45
fsck from util-linux 2.32.1
e2fsck 1.45.6 (20-Mar-2020)
ext2fs_open2: Bad magic number in super-block
fsck.ext4: Superblock invalid, trying backup blocks...
Superblock needs_recovery flag is clear, but journal has data.
Recovery flag not set in backup superblock, so running journal anyway.
Clear journal? no

/dev/vdb was not cleanly unmounted, check forced.
Pass 1: Checking inodes, blocks, and sizes
Pass 2: Checking directory structure
Pass 3: Checking directory connectivity
Pass 4: Checking reference counts
Pass 5: Checking group summary information
Padding at end of inode bitmap is not set. Fix? no

Inode bitmap differences: Group 1 inode bitmap does not match checksum.
IGNORED.
Group 2 inode bitmap does not match checksum.
IGNORED.
Group 3 inode bitmap does not match checksum.
IGNORED.
```

```
Group 77 block bitmap does not match checksum.
IGNORED.
Group 78 block bitmap does not match checksum.
IGNORED.

/dev/vdb: ********** WARNING: Filesystem still has errors **********

/dev/vdb: 11/655360 files (0.0% non-contiguous), 66753/2621440 blocks
[root@ecs-5c5c-0823-0823274 ~]#
```

执行以下命令，修复已损坏的文件系统。

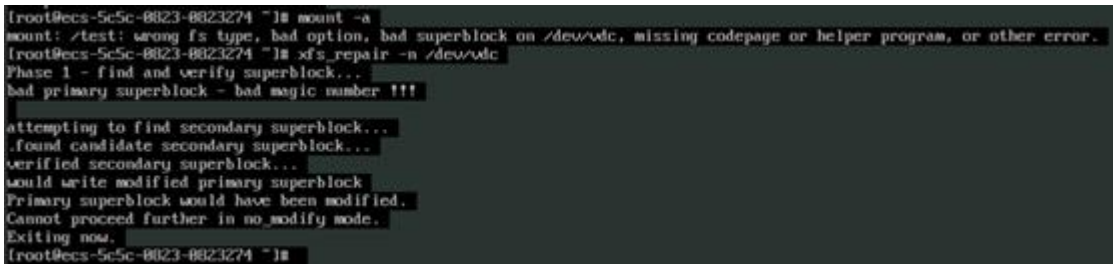
```
[root@ecs-5c5c-0823-0823274 ~]# fsck /dev/vdb -y
fsck from util-linux 2.32.1
e2fsck 1.45.6 (20-Mar-2020)
ext2fs_open2: Bad magic number in super-block
fsck.ext4: Superblock invalid, trying backup blocks...
/dev/vdb was not cleanly unmounted, check forced.
Pass 1: Checking inodes, blocks, and sizes
Pass 2: Checking directory structure
Pass 3: Checking directory connectivity
Pass 4: Checking reference counts
Pass 5: Checking group summary information
Block bitmap differences:  +(32768--33794) +(98304--99330) +(163840--164866) +(229376--230402) +(294912--295938) +(819200--82822
6) +(884736--885762) +(1605632--1606658)
Fix? yes

Padding at end of inode bitmap is not set. Fix? yes


/dev/vdb: ***** FILE SYSTEM WAS MODIFIED *****
/dev/vdb: 11/655360 files (0.0% non-contiguous), 66753/2621440 blocks
[root@ecs-5c5c-0823-0823274 ~]# mount -a
[root@ecs-5c5c-0823-0823274 ~]#
```

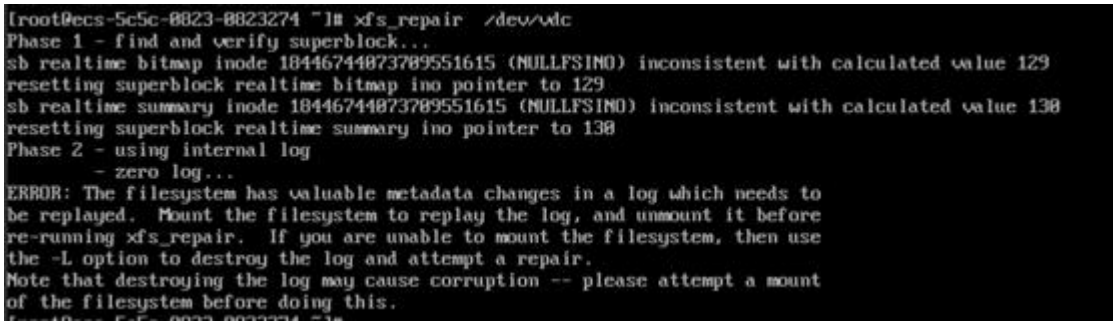
xfs 文件系统：

请执行以下命令，检查有问题的文件系统。



执行以下命令，修复已损坏的文件系统。

```
# xfs_repair 文件系统
```



根据打印消息“ERROR：”，修复失败时：

执行 # xfs_repair -L 文件系统，即 xfs_repair -L /dev/vdc。

注：-L 是修复 xfs 文件系统的最后手段，慎重选择，因为会清空日志，丢失用户数据和文件；另，在执行 xfs_repair 操作前，最好使用 xfs_metadump 工具保存元数据，一旦修复失败，可以恢复到修复之前的状态。

```
agi unlinked bucket 43 is 0 in ag 0 (inode=0)
agi unlinked bucket 44 is 0 in ag 0 (inode=0)
agi unlinked bucket 45 is 0 in ag 0 (inode=0)
agi unlinked bucket 46 is 0 in ag 0 (inode=0)
agi unlinked bucket 47 is 0 in ag 0 (inode=0)
agi unlinked bucket 48 is 0 in ag 0 (inode=0)
agi unlinked bucket 49 is 0 in ag 0 (inode=0)
agi unlinked bucket 50 is 0 in ag 0 (inode=0)
agi unlinked bucket 51 is 0 in ag 0 (inode=0)
agi unlinked bucket 52 is 0 in ag 0 (inode=0)
agi unlinked bucket 53 is 0 in ag 0 (inode=0)
agi unlinked bucket 54 is 0 in ag 0 (inode=0)
agi unlinked bucket 55 is 0 in ag 0 (inode=0)
agi unlinked bucket 56 is 0 in ag 0 (inode=0)
agi unlinked bucket 57 is 0 in ag 0 (inode=0)
agi unlinked bucket 58 is 0 in ag 0 (inode=0)
agi unlinked bucket 59 is 0 in ag 0 (inode=0)
agi unlinked bucket 60 is 0 in ag 0 (inode=0)
agi unlinked bucket 61 is 0 in ag 0 (inode=0)
agi unlinked bucket 62 is 0 in ag 0 (inode=0)
agi unlinked bucket 63 is 0 in ag 0 (inode=0)
sb_fdblocks 2618856, counted 1963502
root inode chunk not found
Phase 3 - for each AG...
        - scan and clear agi unlinked lists...
        - process known inodes and perform inode discovery...
        - agno = 0
        - agno = 1
        - agno = 2
        - agno = 3
        - process newly discovered inodes...
Phase 4 - check for duplicate blocks...
        - setting up duplicate extent list...
        - check for inodes claiming duplicate blocks...
        - agno = 0
        - agno = 1
        - agno = 2
        - agno = 3
Phase 5 - rebuild AG headers and trees...
        - reset superblock...
Phase 6 - check inode connectivity...
        - resetting contents of realtime bitmap and summary inodes
        - traversing filesystem ...
        - traversal finished ...
        - moving disconnected inodes to lost+found ...
Phase 7 - verify and correct link counts...
done
```

再次执行# xfs_repair /dev/vdc。

```
[root@ecs-5c5c-0823-0823274 ~]# xfs_repair  /dev/vdc
Phase 1 - find and verify superblock...
Phase 2 - using internal log
        - zero log...
        - scan filesystem freespace and inode maps...
        - found root inode chunk
Phase 3 - for each AG...
        - scan and clear agi unlinked lists...
        - process known inodes and perform inode discovery...
        - agno = 0
        - agno = 1
        - agno = 2
        - agno = 3
        - process newly discovered inodes...
Phase 4 - check for duplicate blocks...
        - setting up duplicate extent list...
        - check for inodes claiming duplicate blocks...
        - agno = 0
        - agno = 1
        - agno = 2
        - agno = 3
Phase 5 - rebuild AG headers and trees...
        - reset superblock...
Phase 6 - check inode connectivity...
        - resetting contents of realtime bitmap and summary inodes
        - traversing filesystem ...
        - traversal finished ...
        - moving disconnected inodes to lost+found ...
Phase 7 - verify and correct link counts...
done
```

重启弹性云主机即可恢复正常。

9.4 云主机错误状态处理方式

9.4.1 弹性云主机配置的邮件无法正常发送

问题描述

在使用浏览器登录邮箱时直接使用 HTTP 协议，其端口号默认为 80，邮件主机之间还是使用邮件发送协议：SMTP 协议。通过浏览器发送邮件，则需要开放 TCP 出方向端口为 80。

处理方法

1.在弹性云主机页面找到对应的安全组，单击安全组“ID”，进入“安全组”页面。

2.在安全组界面，单击“添加规则”。

3.在弹出的窗口，选择“出方向”，协议为“TCP”，设置端口为“80”单击“确定”完成添加。

在使用邮件客户端收发邮件时，邮件收发使用不同的协议：

- 发件协议一般使用 SMTP 协议，其端 口号为 25；
- 收件协议使用 POP3 协议，其端口号为 110。 具体操作请参见 1～3。

说明 ：需要添加“入方向”：协议为“TCP”； 端口为“110”和“出方向”：协议为“TCP”，端口 为“25”的规则。

9.4.2 如何解决云平台创建云主机状态错误

在使用云平台创建云主机时，有时候可能会遇到云主机状态错误的情况。这个错误通常表示在创建云主机的过程中出现了一些问题，导致云主机无法正常启动或运行。本篇博客文章将介绍如何解决云平台创建云主机状态错误的问题，帮助读者快速解决该错误并成功创建云主机。

第一部分：检查云主机配置

1.确保已正确配置云主机的基本参数，包括名称、镜像、规格、网络等。

2.检查是否选择了适合的镜像和规格，确保云主机的配置满足需求。

3.确保云主机所属的网络和子网配置正确，并且与其他网络设备兼容。

第二部分：检查资源配额

1. 登录云平台控制台。
2. 在控制台的菜单中，找到“云主机”或类似的选项，并点击进入云主机管理页面。
3. 在云主机管理页面中，找到目标项目，并点击进入项目详情页面。

4. 在项目详情页面中，查找并点击“资源配额”或类似的选项。
5. 检查是否已达到了某些资源的配额限制，如 CPU、内存、磁盘等。
6. 如果资源配额不足，可以联系系统管理员或云平台的技术支持团队，申请增加相应资源的配额。

第三部分：检查日志和错误信息

1. 在云平台控制台中，找到云主机管理页面。
2. 在云主机管理页面中，找到目标云主机，并点击进入云主机详情页面。
3. 在云主机详情页面中，查找并点击“日志”或类似的选项。
4. 注意查找是否有报错信息或异常情况的记录。
5. 根据日志和错误信息，进行相应的故障排查和修复操作。

第四部分：联系技术支持

1. 如果以上方法都无法解决云主机状态错误的问题，可以联系云平台的技术支持团队，寻求进一步的帮助和支持。
2. 提供详细的错误信息、日志和操作步骤，以便技术支持团队更好地理解和解决问题。

总结

通过以上步骤，我们可以尝试解决云平台创建云主机状态错误的问题。在遇到问题时，及时检查云主机配置、资源配额和日志信息，同时与技术支持团队保持沟通和协作，可以更快地解决问题并成功创建云主机。

9.4.3 云主机错误状态及解决方案

一、引言

随着云计算技术的快速发展，云主机已经成为企业、个人和开发者们进行互联网应用的首选。然而，使用云主机的过程并非完全顺利，可能会遇到各种错误

状态。本文将详细探讨云主机错误状态的类型、原因以及解决方案，帮助用户更好地管理和维护自己的云主机。

二、云主机错误状态类型

1. 网络错误

网络错误是云主机使用过程中最常见的错误类型之一。网络错误可能包括网络连接超时、无法连接到远程服务器或网络连接中断等。这些错误通常是由于网络配置问题、网络设备故障或网络拥堵等原因引起的。

2. 服务器错误

服务器错误通常包括服务器启动失败、运行异常、过载等。这类错误可能是由于服务器硬件故障、软件问题、操作系统故障或服务器资源不足等原因引起的。

3. 应用程序错误

应用程序错误是指运行在云主机上的应用程序出现错误。这类错误可能包括应用程序崩溃、运行缓慢、内存泄漏等。这些错误可能是由于应用程序本身的问题，如代码错误、不兼容性或配置问题等引起的。

4. 安全错误

安全错误是指与安全相关的错误，如身份验证失败、权限问题或安全策略冲突等。这些错误可能是由于用户名或密码错误、权限配置错误或安全策略设置不当等原因引起的。

三、云主机错误状态原因分析

1. 网络问题

网络问题可能是由于网络设备故障、网络连接问题或网络配置不当等原因引起的。例如，如果云主机的网络设备出现故障，或者网络连接被中断，那么就会出现网

络错误。此外，如果网络配置不正确，例如 DNS 解析不正确，也可能会导致无法连接到远程服务器。

2. 服务器硬件故障

服务器硬件故障可能是由于服务器硬件设备出现故障或服务器资源不足等原因引起的。例如，如果服务器的硬盘出现故障，那么服务器可能无法启动或运行异常。此外，如果服务器的资源不足，例如内存不足或 CPU 过载，也可能会导致服务器运行异常。

3. 应用程序问题

应用程序问题可能是由于应用程序本身的问题，如代码错误、不兼容性或配置问题等引起的。例如，如果应用程序的代码存在错误，那么应用程序可能无法正常运行。此外，如果应用程序的配置不正确，例如数据库连接不正确或文件路径错误，也可能会导致应用程序运行异常。

4. 安全问题

安全问题可能是由于身份验证失败、权限问题或安全策略冲突等原因引起的。例如，如果用户名或密码不正确，那么身份验证可能会失败。此外，如果权限配置不正确，例如权限设置过于宽松或过于严格，也可能会导致权限问题。另外，如果安全策略设置不当，例如过于宽松的安全策略或过于严格的安全策略，也可能会导致安全问题。

四、云主机错误状态解决方案

1. 网络错误的解决方案

对于网络错误，可以通过以下步骤来解决：首先检查网络设备是否正常工作；其次检查网络连接是否正常；最后检查网络配置是否正确。如果以上步骤都无法解决问题，可以考虑联系网络管理员或运营商寻求帮助。

2. 服务器错误的解决方案

对于服务器错误，可以通过以下步骤来解决：首先检查服务器硬件设备是否正常工作；其次检查服务器软件是否正常工作；最后检查服务器资源是否充足。如果以上步骤都无法解决问题，可以考虑联系服务商或技术支持团队寻求帮助。

3.应用程序错误的解决方案

对于应用程序错误，可以通过以下步骤来解决：首先检查应用程序是否存在代码错误；其次检查应用程序的配置是否正确；最后检查应用程序是否与操作系统或其他软件兼容。如果以上步骤都无法解决问题，可以考虑联系应用程序开发者或技术支持团队寻求帮助。

9.5 远程连接报错

9.5.1 远程连接云主机出现蓝屏

问题描述

Windows Server 2012 R2 操作系统云服务器，本地使用远程桌面连接云服务器并启用 redirected drive 功能时，云服务器出现蓝屏。

可能原因

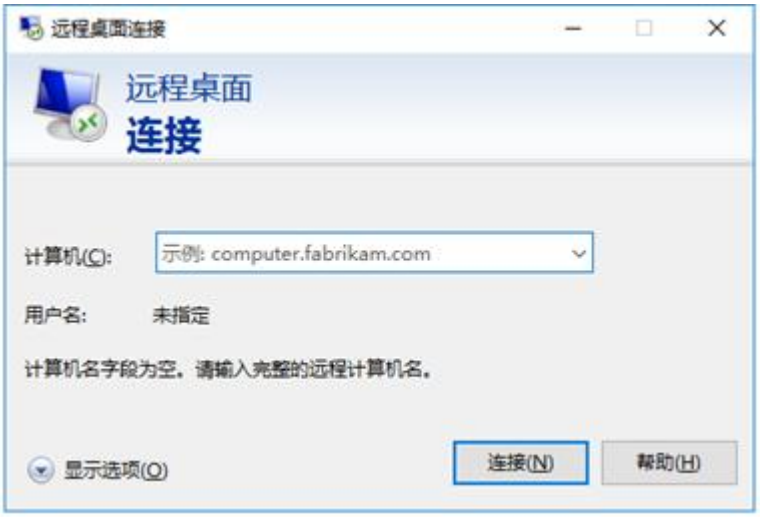
远程桌面连接启用了 redirected drive 功能，同时加载了对应 rdpdr.sys 驱动，该驱动可能会导致云服务器操作系统崩溃，从而导致云服务器蓝屏。

解决方法

1. 打开“运行”窗口。

2. 输入“mstsc”，并单击“确定”，系统打开“远程桌面连接”窗口。





3. 单击左下角的“显示选项”，并选择“本地资源”页签。

4. 在“本地设备和资源”栏，单击“详细信息”。

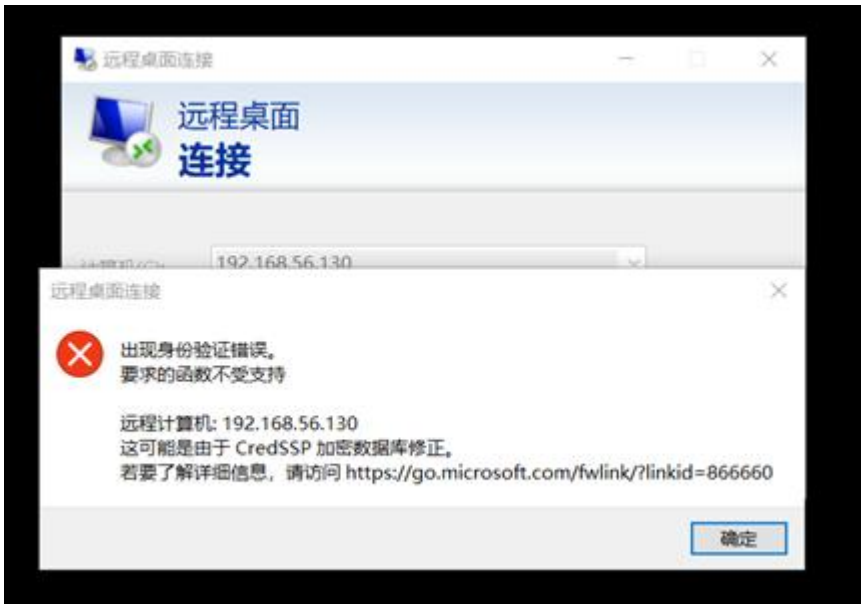
5. 在“本地设备和资源”窗口中，取消勾选“驱动器”，保存修改。

9.5.2 远程连接 Windows 云主机报错：出现身份验证错误，要求的函数不受支持

问题描述

在安装 Windows 操作系统的本地 PC 上，通过远程桌面 RDP 协议（如 MSTSC 方式）连接 Windows 云服务器时，报错提示：出现身份验证错误，要求的函数不受支持。

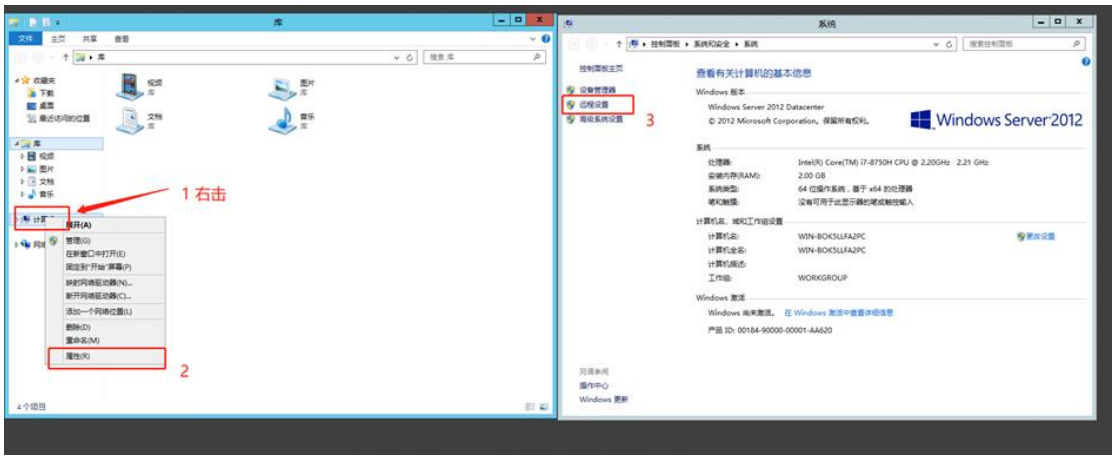
- 如果报错信息中仅提示：出现身份验证错误，要求的函数不受支持。请参考”处理方法“部分解决该问题。
- 如果报错信息中附加提示“这可能是由于 CredSSP 加密数据库修正”，如下图所示，那么可能原因是微软于 2018 年 3 月发布安全补丁，该安全补丁会影响 RDP 连接 CredSSP，从而导致通过 RDP 协议访问云服务器时连接失败。



处理方法

1. 登陆云服务器。

2. 打开“开始”菜单，右键单击“计算机”，选择“属性”，左侧导航栏选择“远程设置”。



3. 在”远程”页签中“远程桌面”栏，选择“允许远程连接到此计算机”。



9.5.3 天翼云 Windows 云主机远程连接时出现报错： 此计算机无法连接到远程计算机

问题描述

使用远程登录方式登录 Windows 云服务器时出现错误：此计算机无法连接到远程计算机。

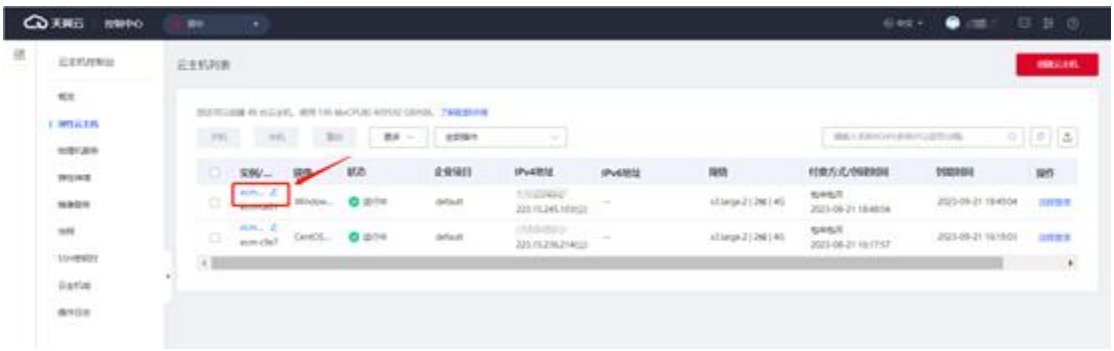
可能原因

1. 云服务器安全组 3389 端口未放行。
2. 远程桌面连接配置错误。
3. 远程桌面会话主机配置错误。

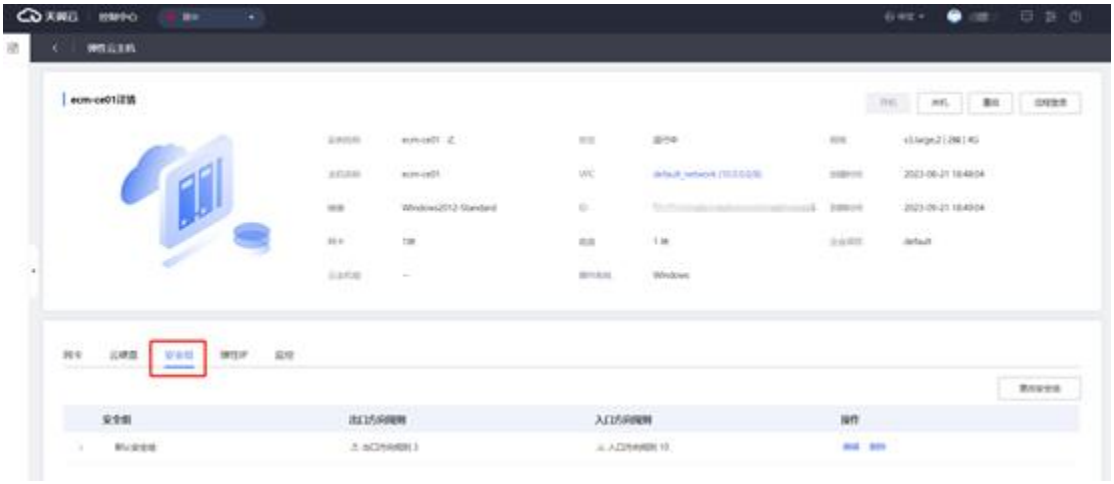
解决方案

检查云服务器安全组端口配置

1. 通过天翼云官网进行登录，通过控制台找到对应云服务器。点击对应云服务实例名进入详情页。



2. 选择安全组。

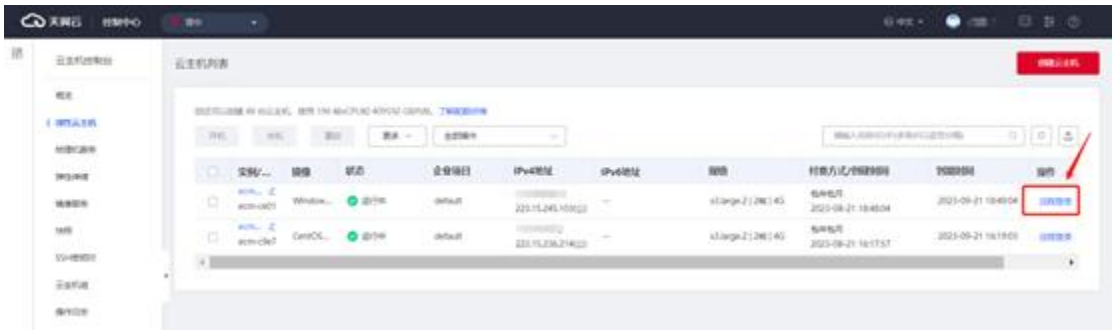


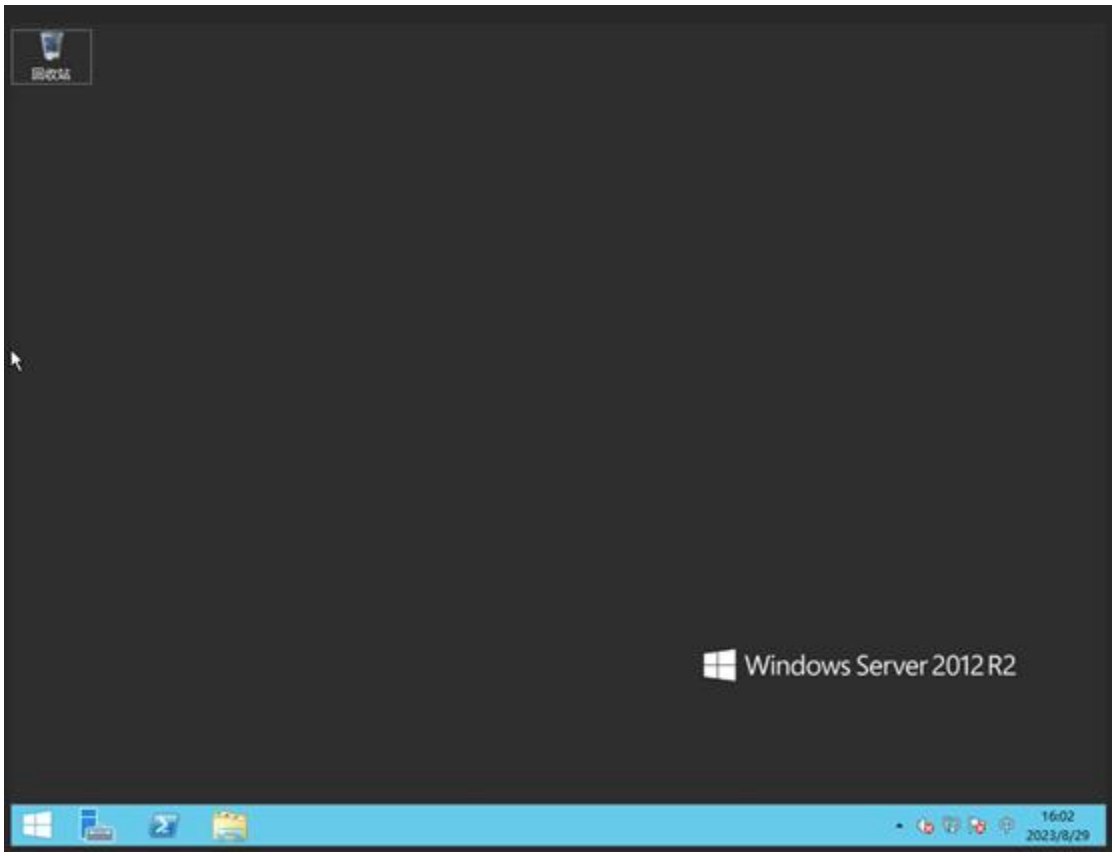
3. 点击默认安全组字样前下拉菜单进入安全组详情页，确认 3389 端口权限是否开启。



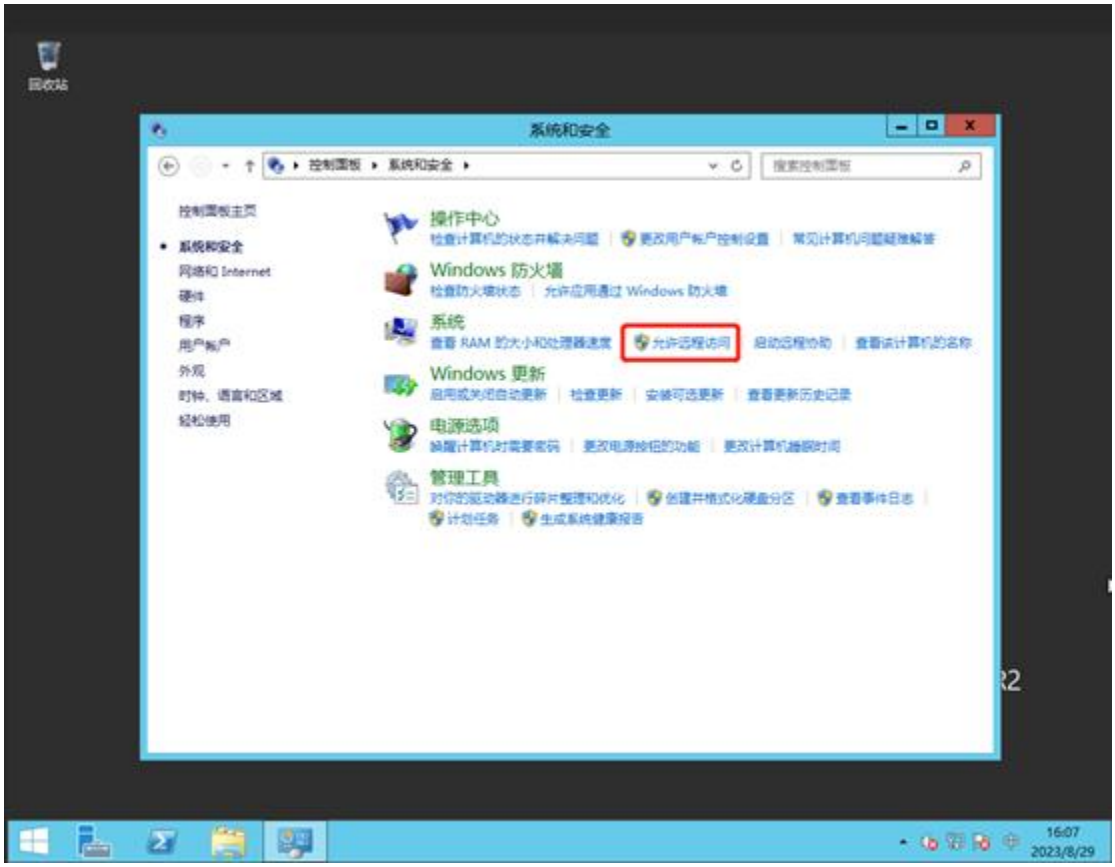
检查远程桌面连接配置

1. 使用管理控制台 VNC 方式登录云服务器。





2. 依次选择选择控制面板>系统和安全>系统>允许远程访问。

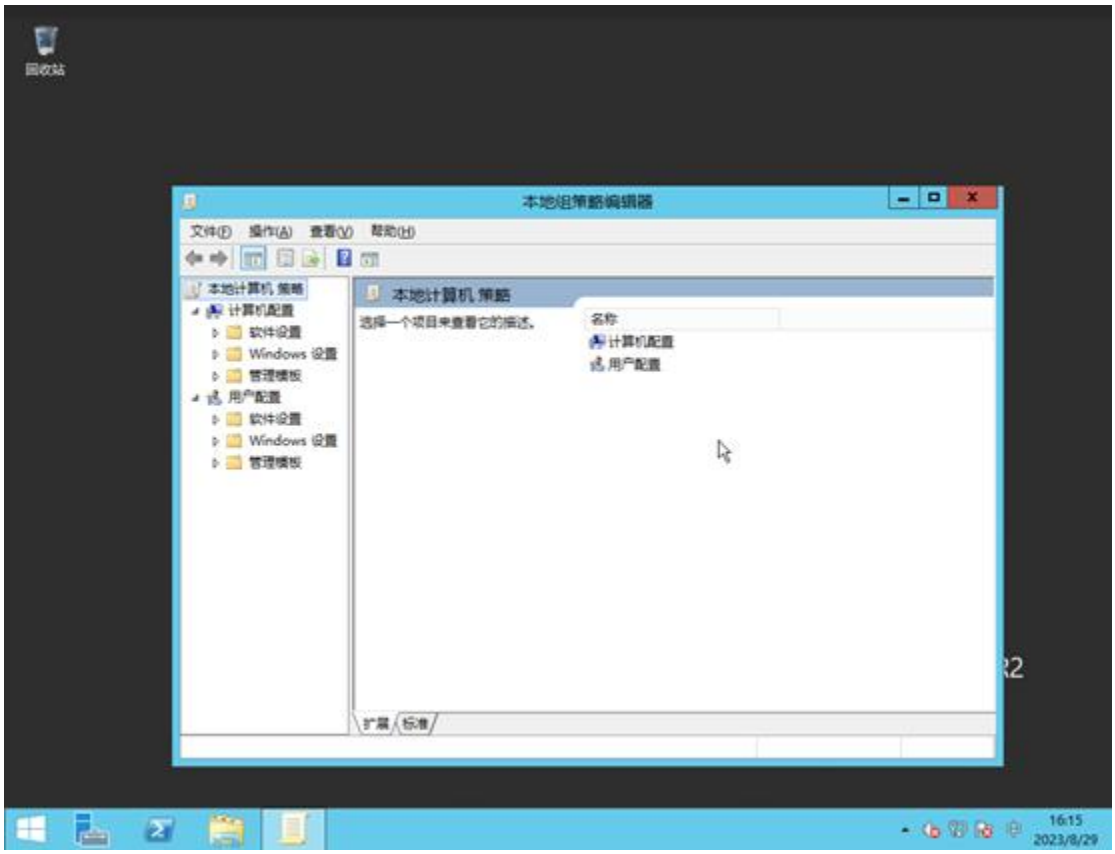


3. 确认允许远程连接到此计算机选项开启。

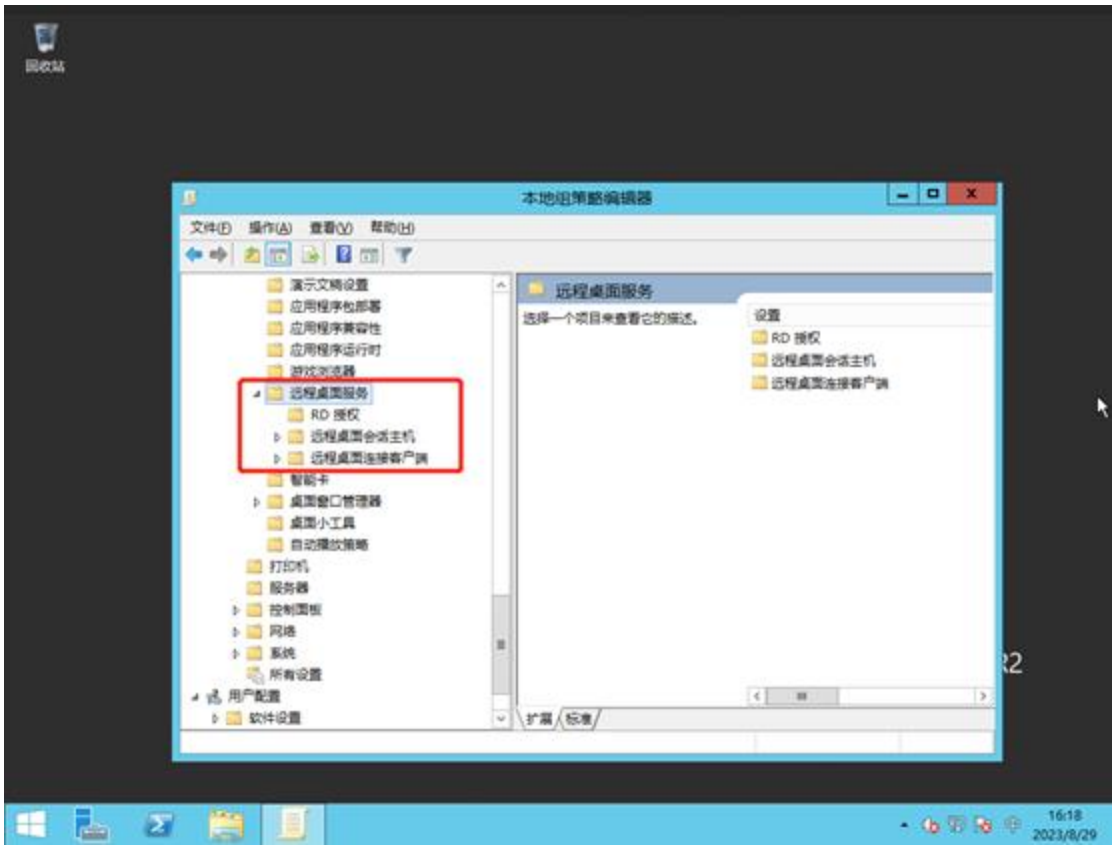


检查远程桌面会话主机配置

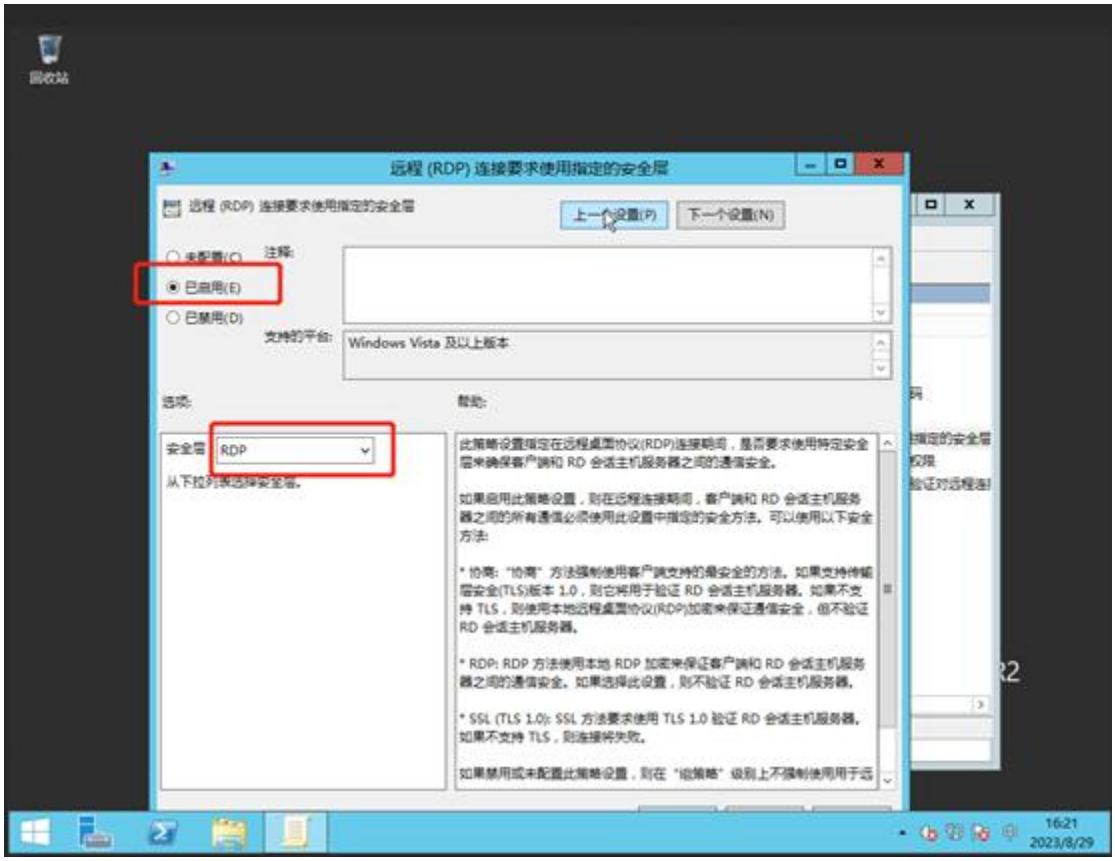
1. 打开运行窗口，输入 gpedit.msc，打开“本地组策略编辑器”。



2. 依次选择“计算机配置 > 管理模板 > Windows 组件”，进入“远程桌面服务”选项。



3. 依次选择“远程桌面会话主机 > 安全>远程（RDP）连接要求使用指定的安全层，选择“已启用”，并将“安全层”设置为“RDP”。



9.5.4 远程连接 Windows 云主机报错：没有远程登录的权限

问题描述

远程连接 windows 云服务器时提示：要远程连接，你需要具有通过远程桌面服务进行登录的权限。



处理方法

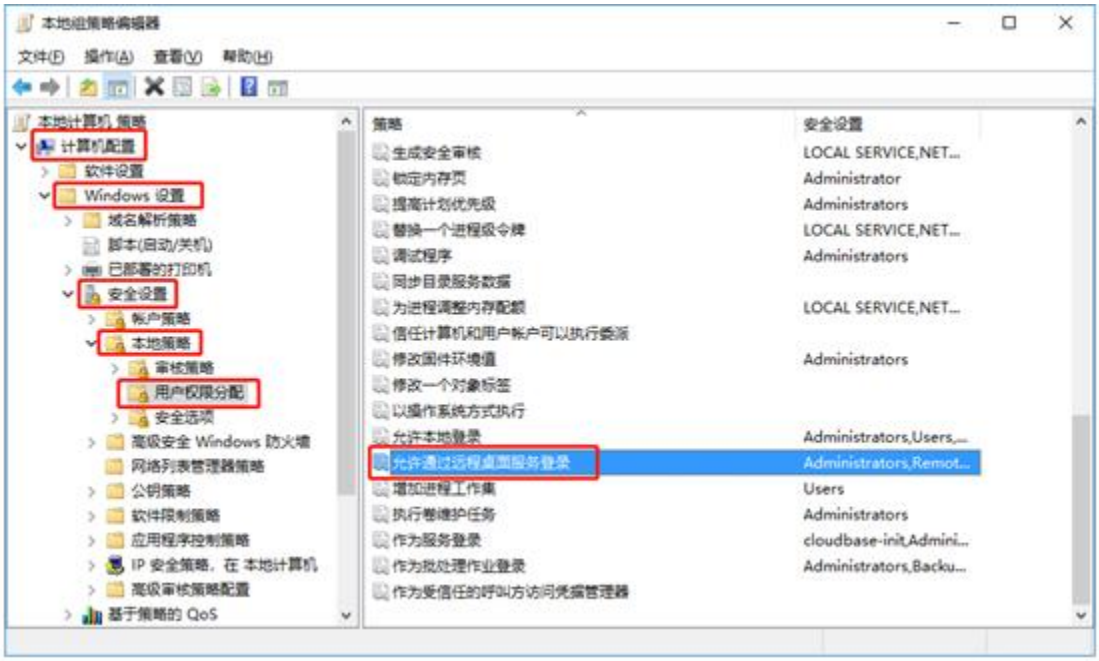
1. 打开 cmd 运行窗口，并输入“gpedit.msc”。



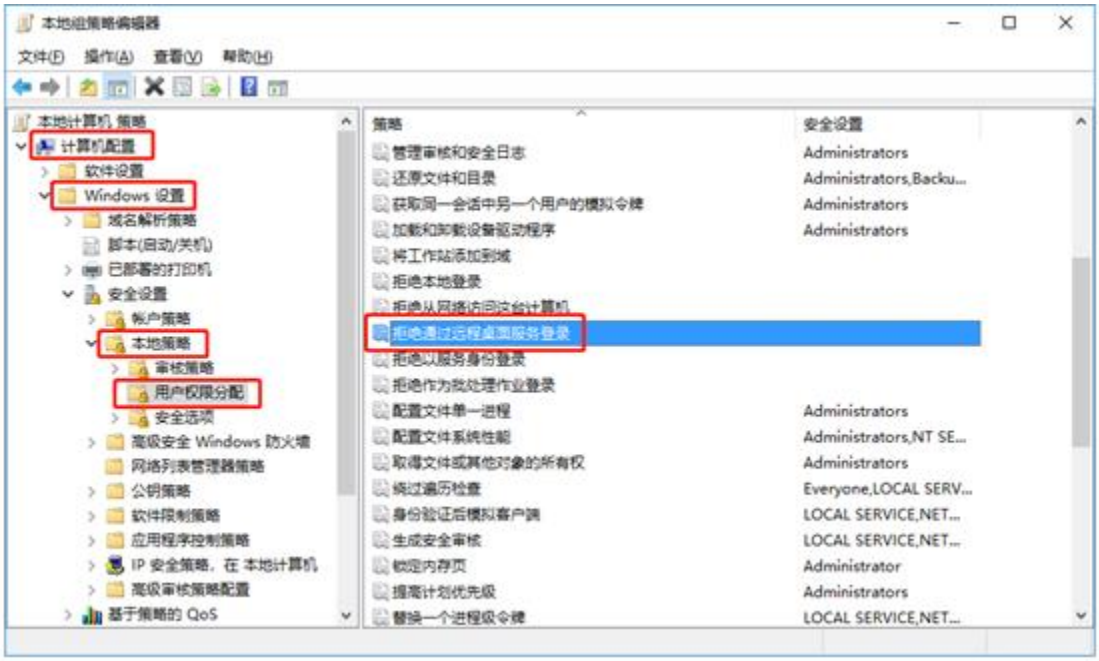
2. 单击“确定”，打开“本地组策略编辑器”。

3. 选择“计算机配置 > windows 设置 > 安全设置 > 本地策略 > 用户权限分配”。

a. 查找并双击“允许通过远程桌面服务登录”。确保已添加“Administrators”和“Remote desktop users”。



b. 查找并双击“拒绝通过远程桌面服务登录”。如果有 Administrator 帐号，则需要删除。



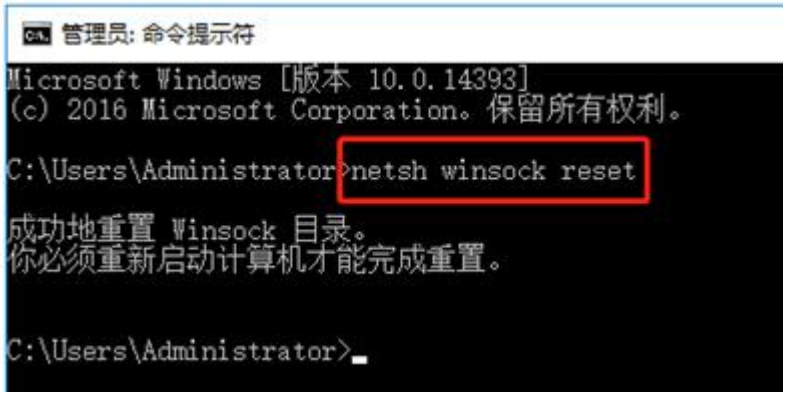
9.5.5 登录 Windows 云主机提示“内部错误”怎么办？

问题描述

使用 mstsc 方式登录 windows 云服务器时，系统报错提示“内部错误”。

处理方法

1. 在本地主机上以管理员身份运行 cmd。
2. 执行 netsh winsock reset。



3. 重启本地主机。
4. 重新通过 mstsc 方式远程登陆。

若您仍无法成功登陆，那么您可以先排查本地的网络是否正常。推荐您更换网络（eg.手机热点）测试是否可以远程登录；如果您使用手机热点的网络可以正常远程登录，那么说明无法成功登陆是由于本地网络异常造成的，你可以尝试重启本地网络（eg.重启路由器）来解决；如果通过上述排查，您仍然无法成功登录

云服务器，请您记录资源信息和问题时间，然后单击管理控制台上方的“工单”，填写工单信息，获取技术支持。



9.5.6 远程连接 Windows 云主机报错：无法验证此远程计算机的身份

问题描述

远程桌面连接 Windows 云服务器时报错提示：″无法验证此远程计算机的身份，是否仍要连接？″



可能原因

云服务器安装了安全软件，防止有未知 IP 登陆云服务器。

解决方法

1. 在提示的对话框中，勾选“ 不再询问我是否连接到此计算机”，点击“是”。

2. 卸载安全软件。

3. 登录安全软件，将登录安全等级修改为系统默认登录方式。

9.5.7Windows 云主机远程连接时出现报错：RDP 连接断开，出现内部错误

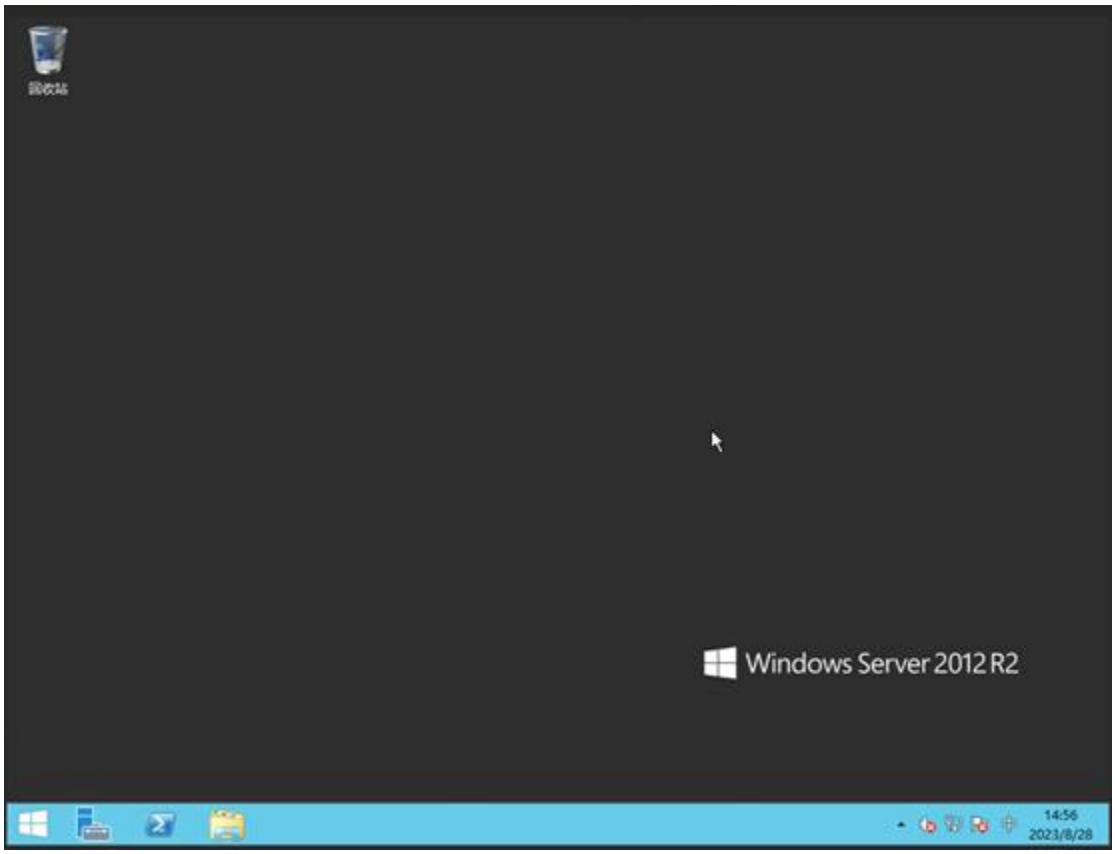
问题描述

远程连接 Windows 操作系统云主机时会遇到“出现了内部错误”提示，无法正常连接到云主机。这种情况一般都是远程服务太忙导致。

解决方法

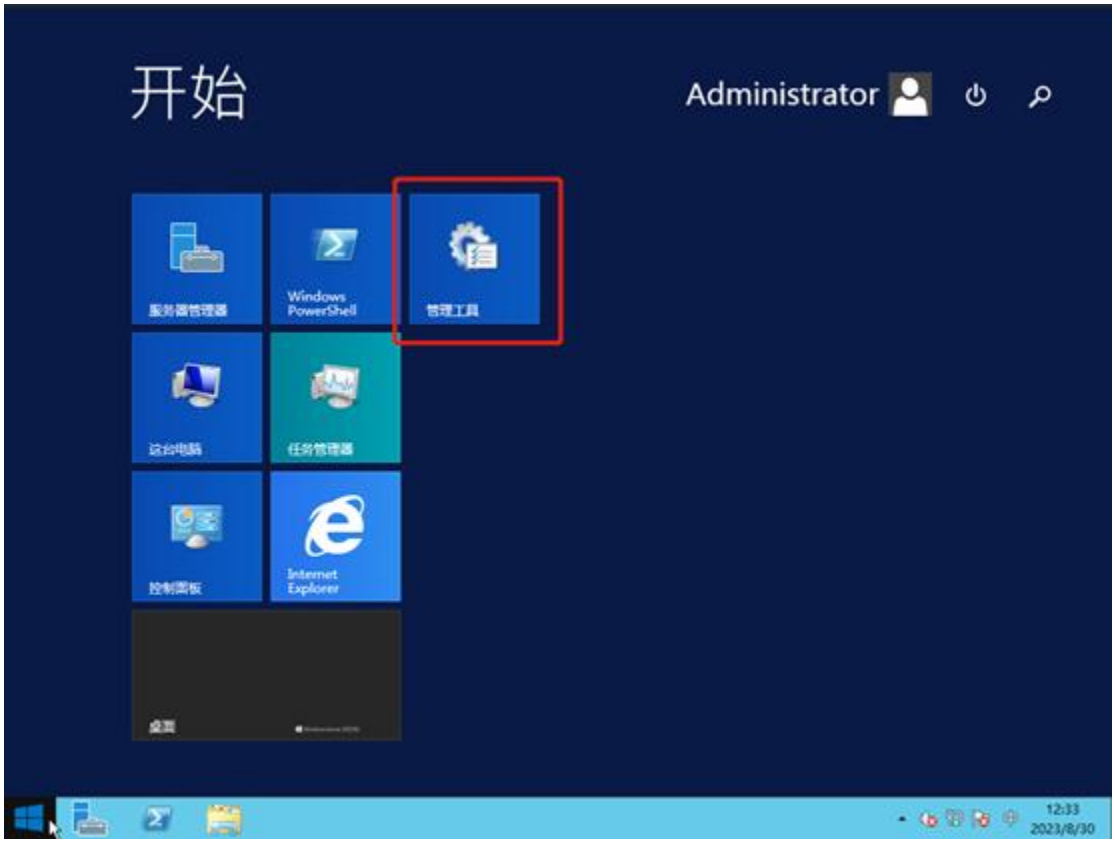
1. 使用管理控制台 VNC 方式登录弹性云主机。



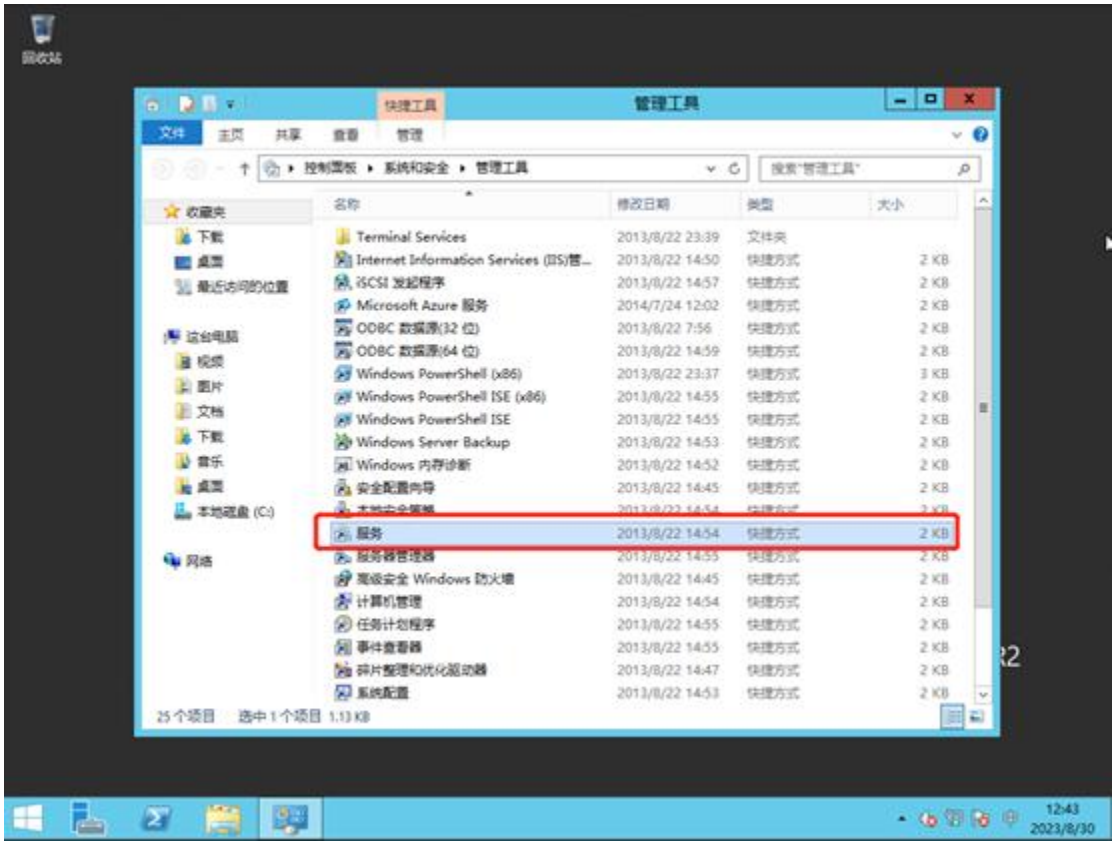


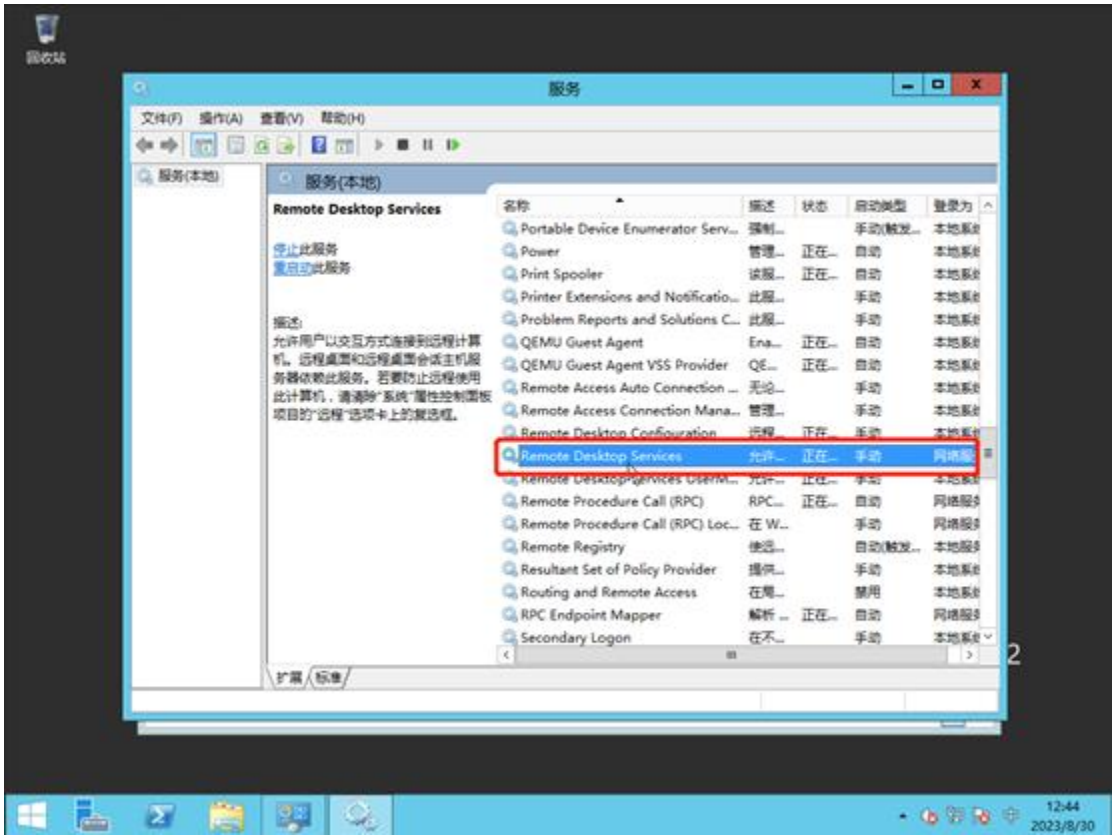
2. 确认弹性云主机 Remote Desktop Services 服务状态。

（1）点击开始>管理工具。



（2）选择服务，找到 Remote Desktop Services，确保其状态为“正在运行”。





3. 如果查询到 Remote Desktop Services 服务没有正在运行，请先重启该服务，然后再尝试连接云主机。

9.5.8Windows 云主机远程连接时出现报错：您的连接已丢失

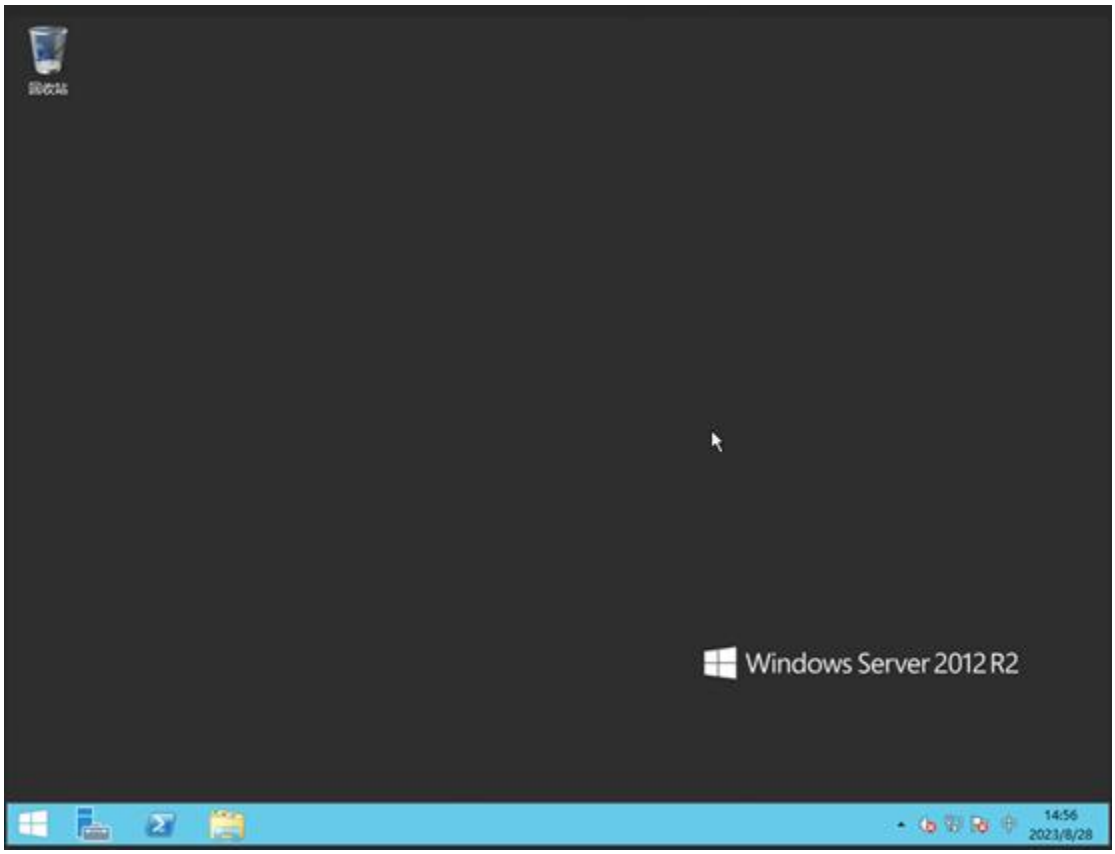
问题描述

远程桌面连接 Windows 云主机时候报错："您的远程桌面会话已结束，另一用户已连接到此远程计算机，因此您的连接已丢失。"

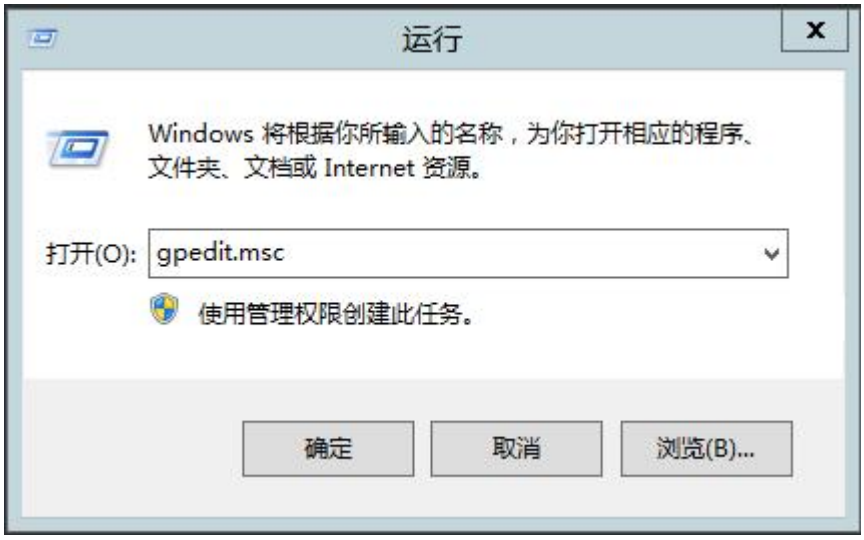
解决方法

1. 使用管理控制台 VNC 方式登录弹性云主机。

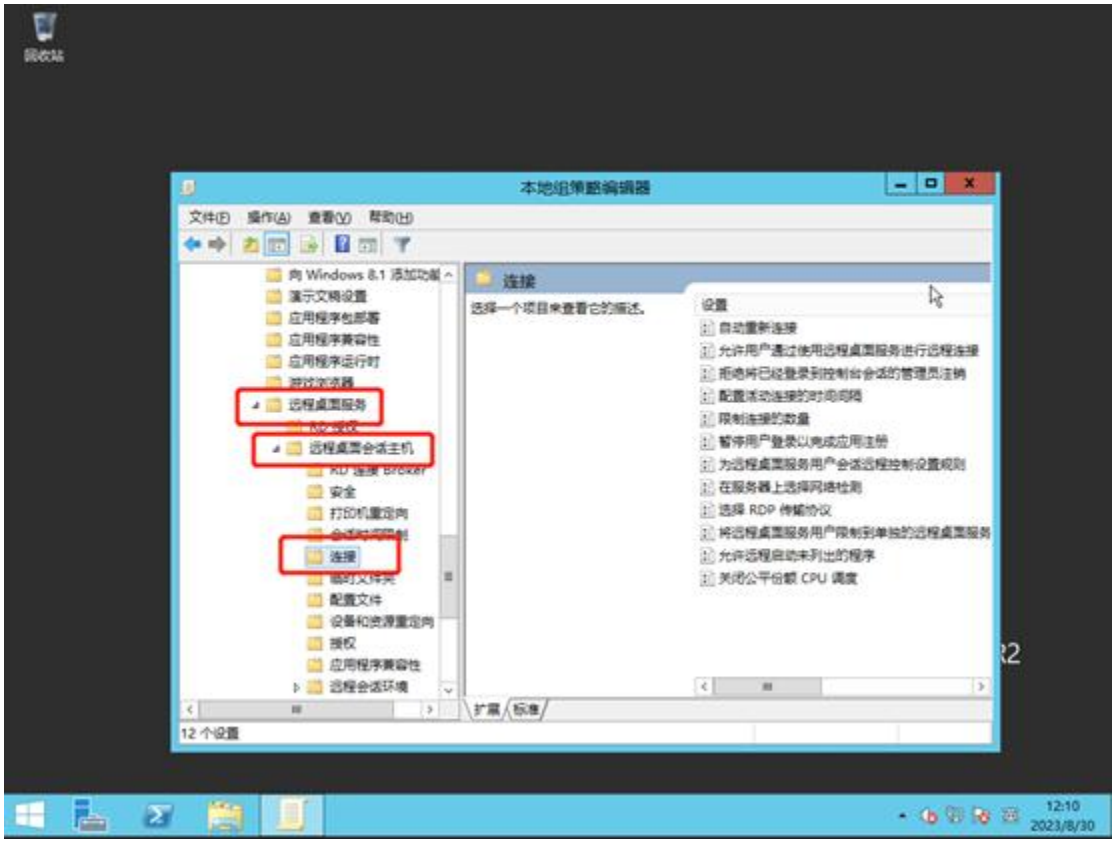




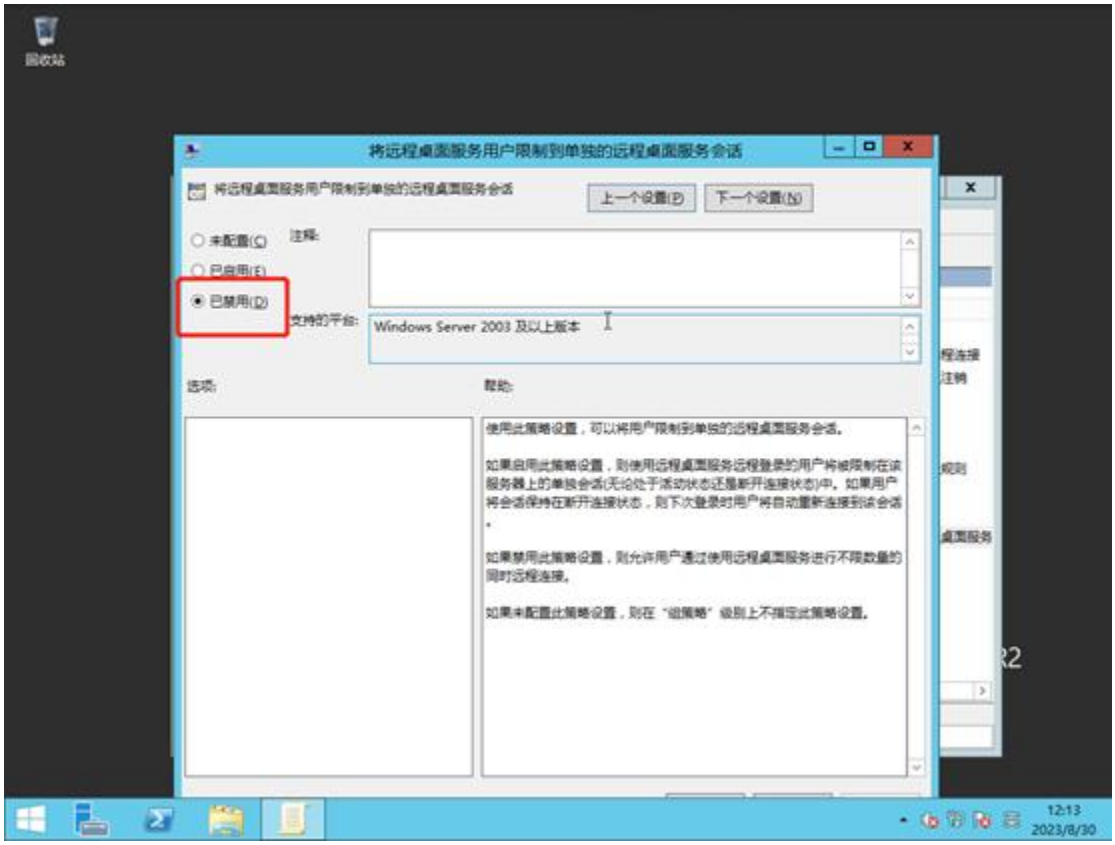
2. 点击“开始 > 运行”，输入“gpedit.msc”，打开“本地组策略编辑器”。



3. 依次选择“计算机配置 > 管理模板>Windows 组件 > 远程桌面服务>远程桌面会话主机 > 连接”。



4. 点击“将远程桌面服务用户限制到单独的远程桌面服务会话”，修改配置为“已禁用”，单击“确定”保存，在命令窗口执行 gpupdate /force，更新组策略。



5. 在命令窗口执行 gpupdate /force，更新组策略。

9.5.9Windows 云主机报错：您的凭据无法工作

问题描述

在天翼云官网开通对应 Windows 云主机后，除控制台 VNC 接入外还有其他 PC 本地进行远程接入场景。当接入时候提示您的凭据无法工作错误，远程登录失败。



已开通云主机信息：远程登录时需要公网 IP 信息、账号名及对应密码。

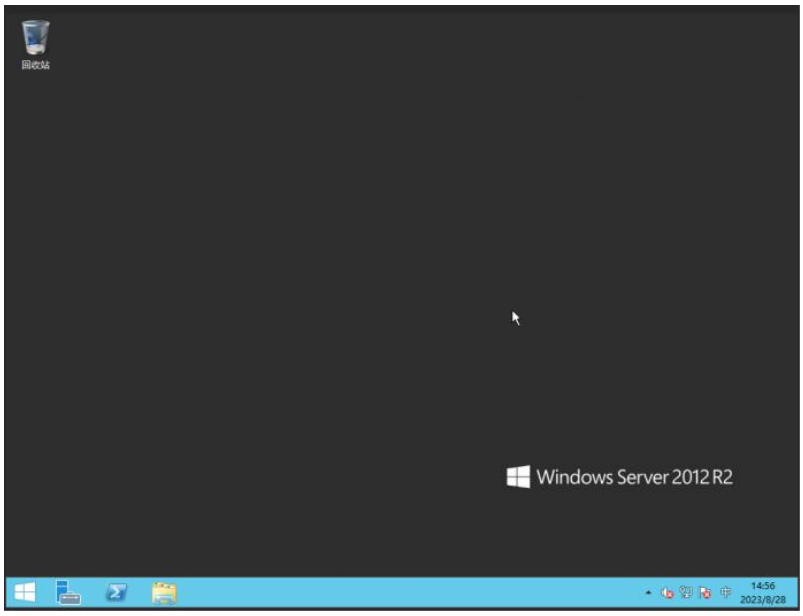


解决方法

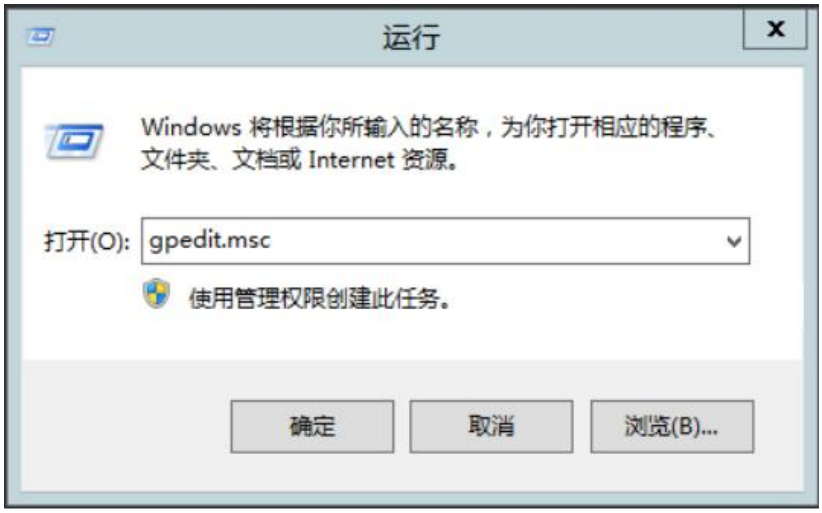
步骤一： 修改 Windows 云主机网络策略

1. 使用管理控制台 VNC 方式登录云服务器。

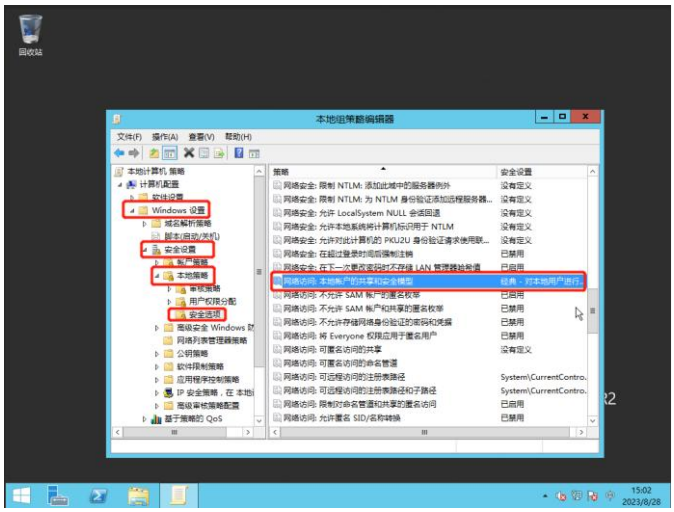




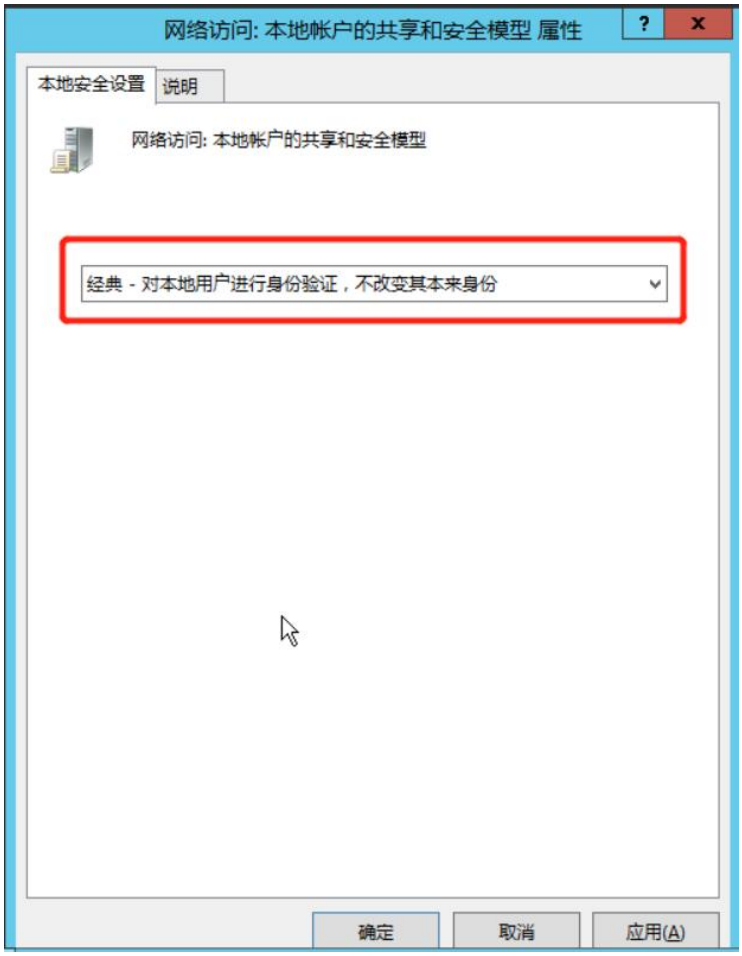
2. 点击“开始 > 运行”，输入“gpedit.msc”，打开“本地组策略编辑器”。



3. 点击“计算机配置 < Windows 设置 < 安全设置 < 本地策略 < 安全选项”，打开“网络访问：本地帐户的共享和安全模型”。



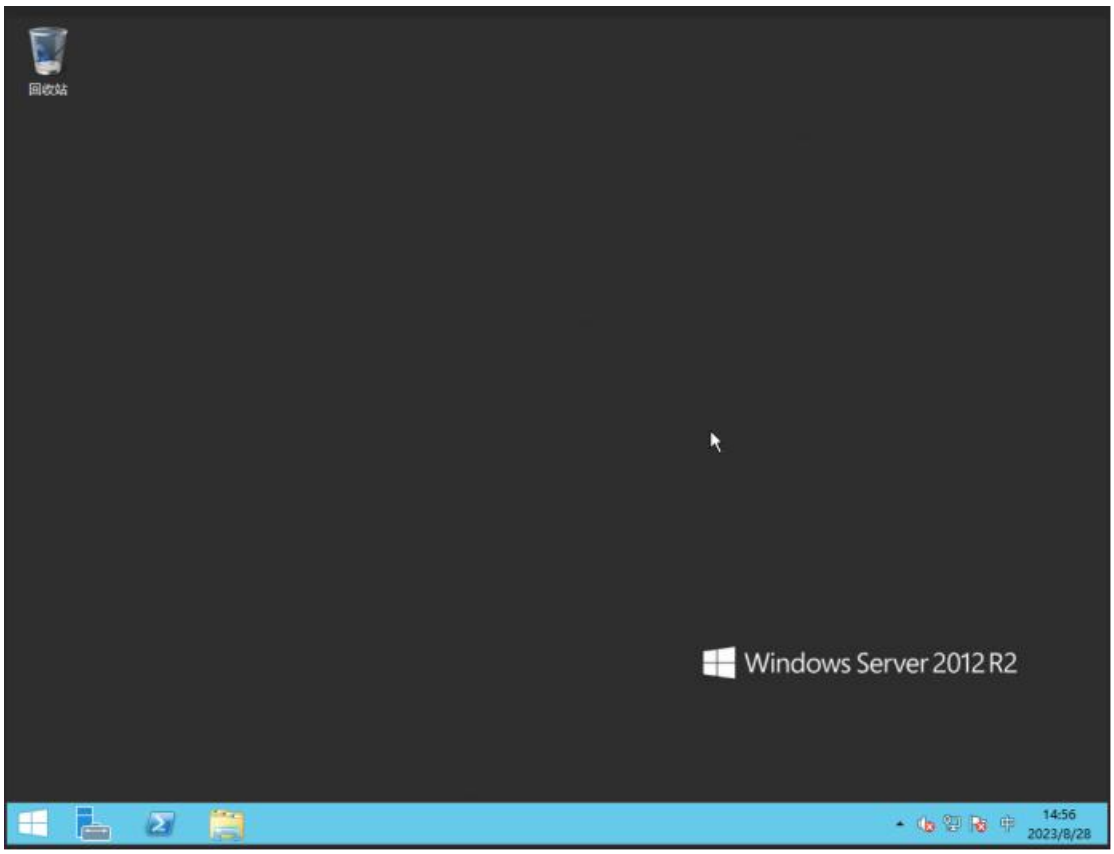
4. 选择“经典 - 对本地用户进行身份验证，不改变其本来身份”，单击“确定”。



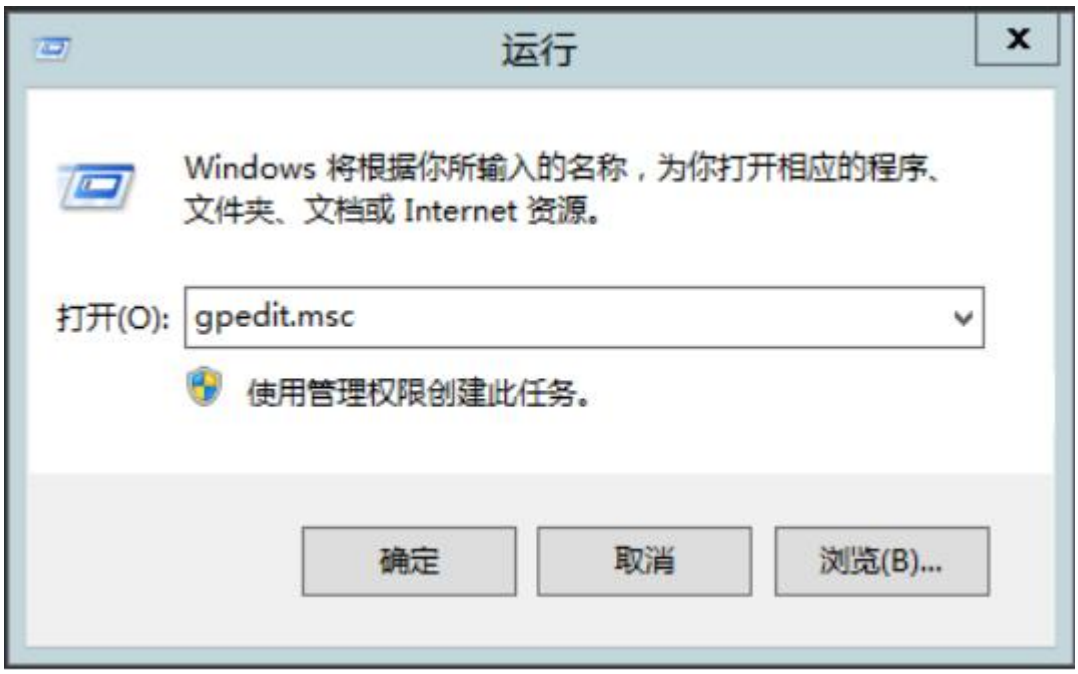
步骤二： 修改 Windows 云主机修改凭据分配

1. 使用管理控制台 VNC 方式登录云服务器。

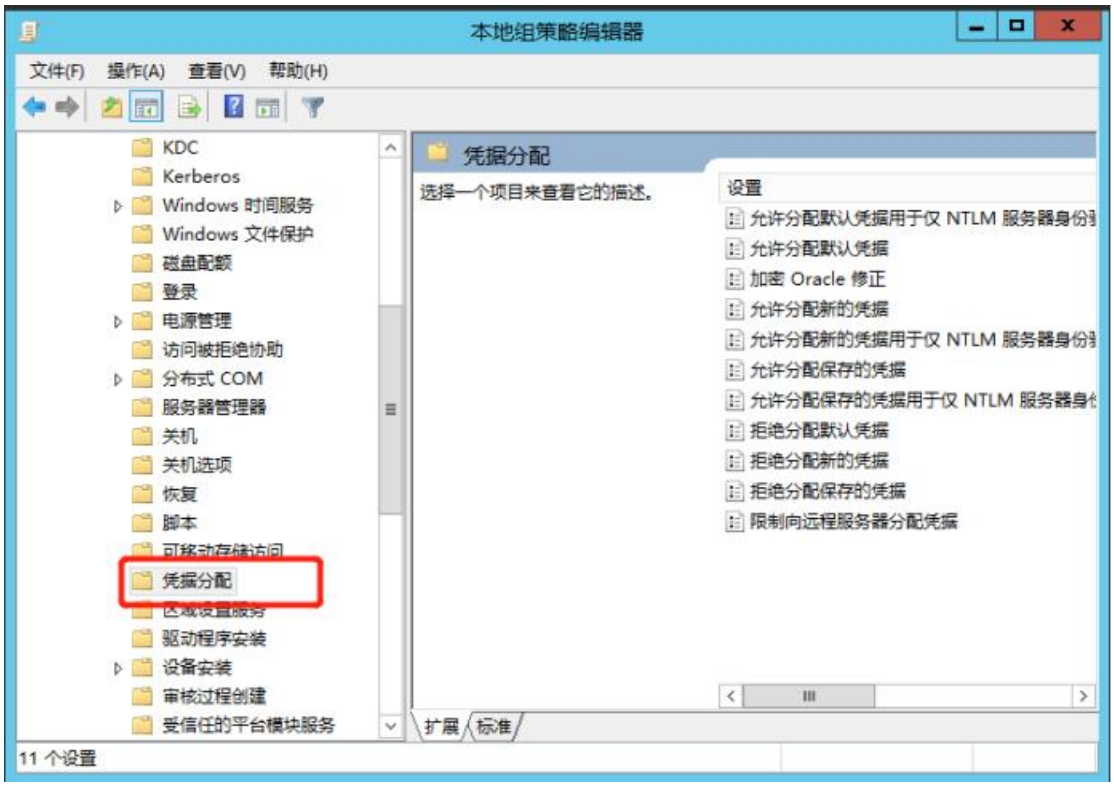




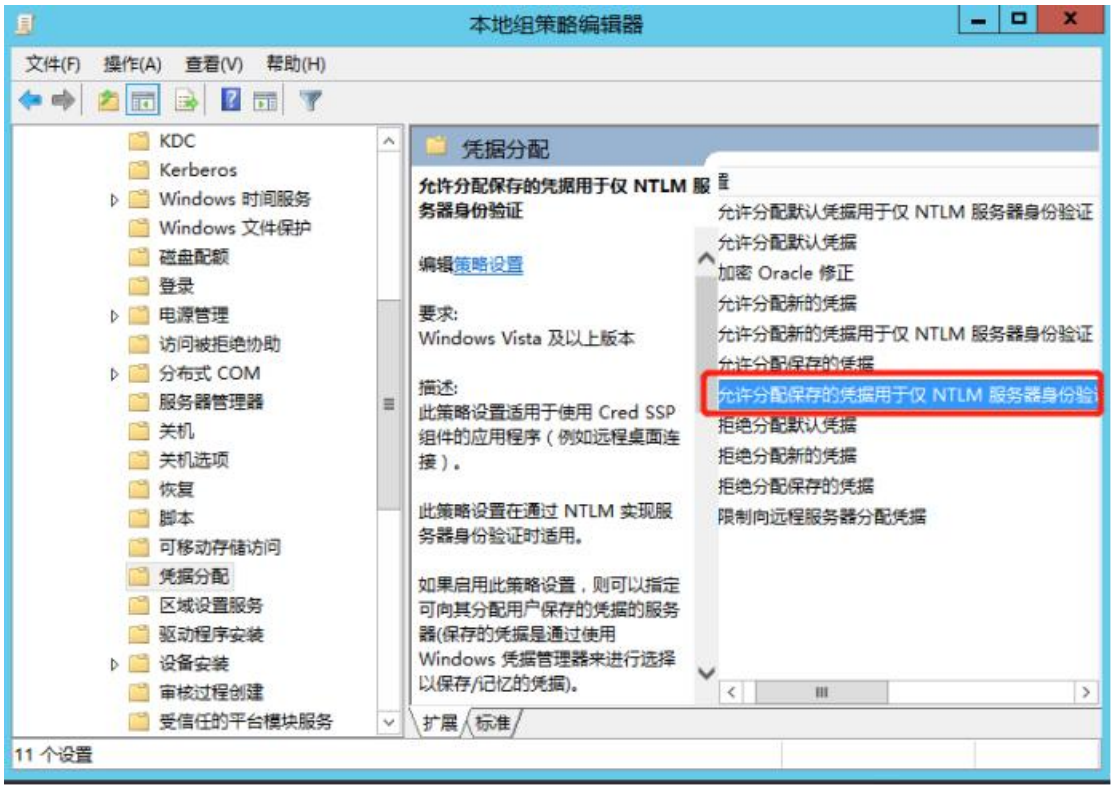
2. 打开“开始 > 运行”，输入“gpedit.msc”，打开“本地组策略编辑器”。



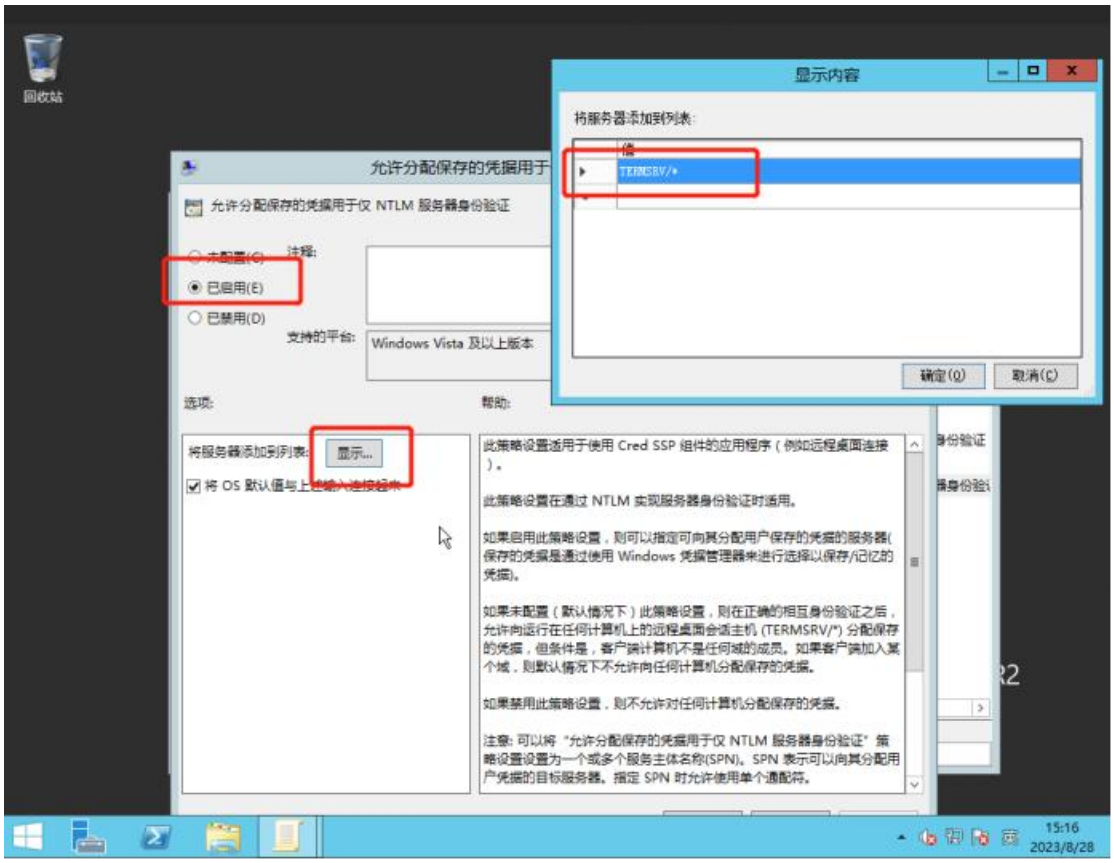
3. 选择“计算机配置 〈 管理模板 〈 系统”，打开“凭据分配”。



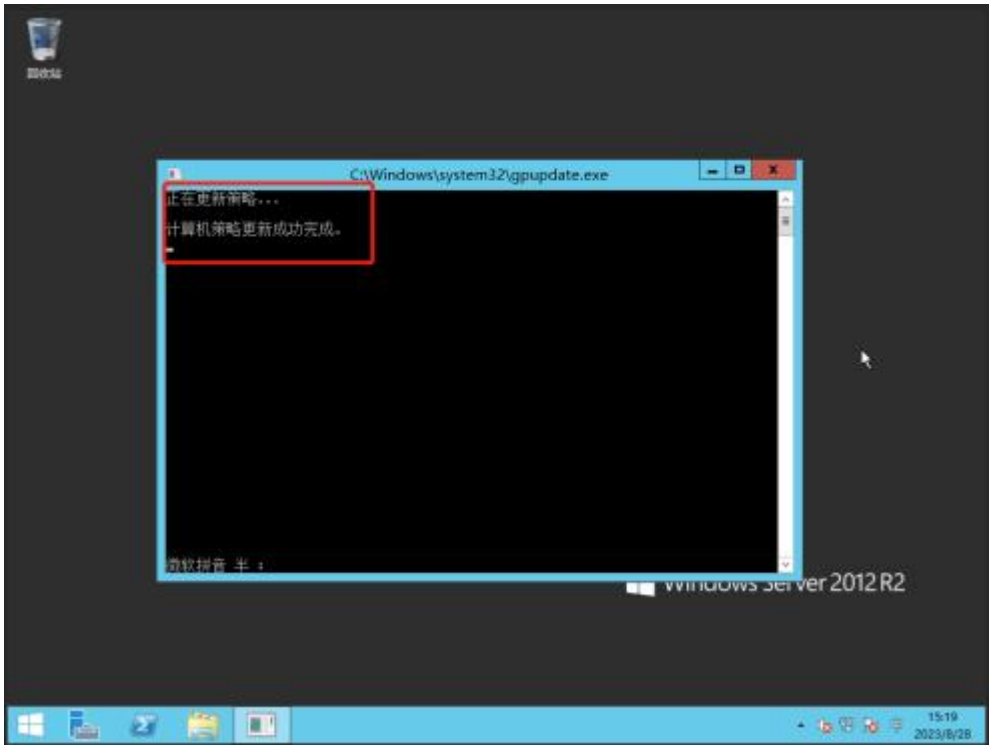
4. 双击打开“允许分配保存的凭据用于仅 NTLM 服务器身份验证”，单击“确定”。



5. 选择“已启用” 在“显示”里面输入 TERMSRV/。TERMSRV/ 表示运行在所有计算机上的终端服务器。



6. 打开“开始 > 运行”输入 gpupdate /force，更新组策略。



步骤三： 设置本地主机的凭据

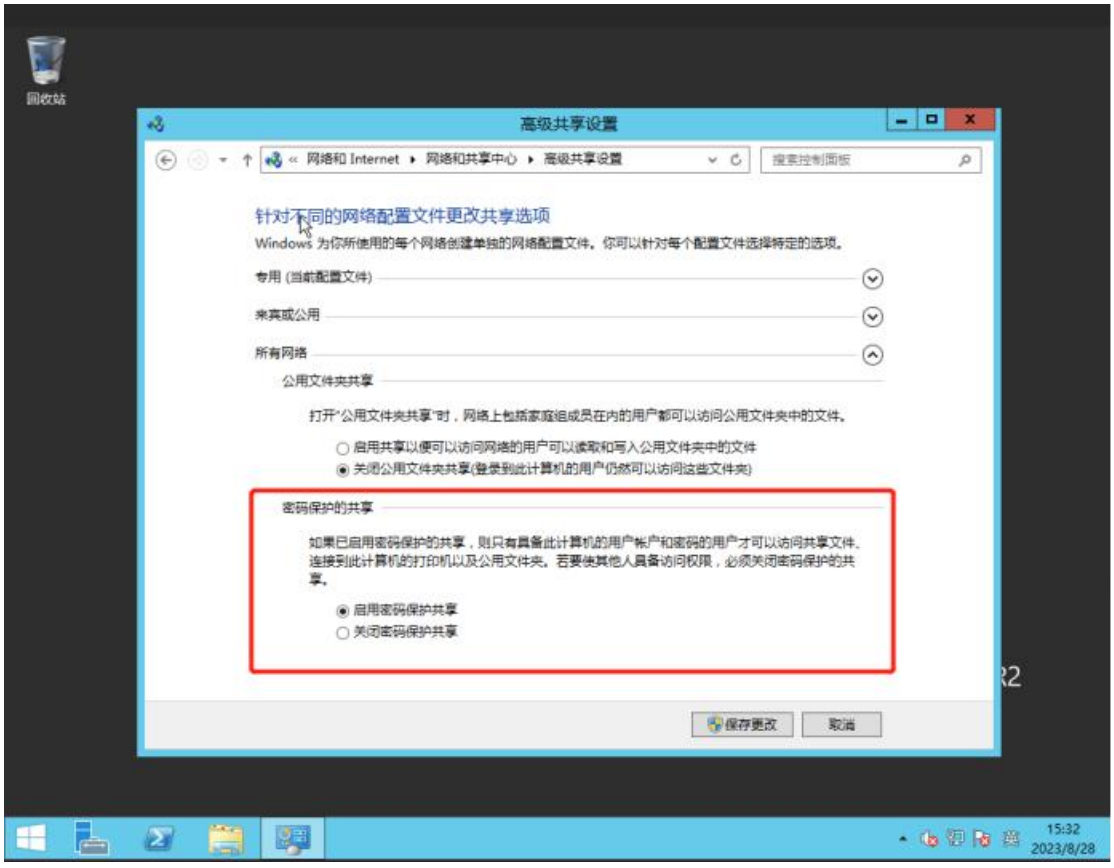
1. 打开本地主机的控制面板，选择“凭据管理器进行 Windows 凭据添加。依次

输入云服务器的 IP 地址、云服务器的登录用户名、云服务器的登录密码。



步骤四： 关闭云服务器密码保护共享

1. 登录 Windows 云服务器。打开左下角的“开始”菜单，选择“控制面板 > 网络和 Internet > 网络和共享中心 > 更改高级共享配置。在“密码保护的共享”页签下选择“关闭密码保护共享”。单击保存更改。



9.6 网络配置

9.6.1resolv 文件被重置怎么办？

问题现象

修改 resolv.conf 文件后重启云主机，发现该文件被重置修改失败。

解决方法

给 resolv 文件添加 i 属性，该命令修改文件属性为只有 root 用户才能修改该文件。执行以下命令：

```
chattr +i /etc/resolv.conf
```

9.6.2Linux 系统重启后/etc/hosts 自动添加主机名解析

问题描述

重启使用 Linux 系统的云主机后发现 /etc/hosts 文件被全部重写或被添加了主机名和回环地址（例：127.0.0.1、127.0.1.1）的解析，导致域名解析不符合预期或出现问题。

问题原因

云平台 Linux 镜像中的一个核心组件是 cloud-init。作为开源的云初始化程序，cloud-init 主要用于对主机名、初始密码等一些自定义信息进行初始化配置。若 cloud-init 配置中存在对 manage_etc_hosts 的配置，则 cloud-init 会根据具体配置值决定如何调整 /etc/hosts 文件。公共镜像默认不启用 cloud-init 对 /etc/hosts 的修改功能，即 manage_etc_hosts 不做配置或配置为 false（默认值）。若您使用了公共镜像并修改了 cloud-init 配置或使用了私有镜像，则可参考此指导来尝试解决问题。

问题排查思路

/etc/hosts 文件被全部重写可能是因为 cloud-init 配置中将 manage_etc_hosts 配置成了 true（若您安装的 cloud-init 版本低于 22.3，则

还可能是配置成了 template）。此时，cloud-init 会根据 /etc/cloud/templates/hosts.tmpl 来重写 /etc/hosts 文件。

而 /etc/hosts 文件被添加主机名解析可能是因为 cloud-init 配置中将 manage_etc_hosts 配置成了 localhost。

解决步骤

1.Cloud-init 配置文件是 /etc/cloud/cloud.cfg，您可能还将自定义配置放于 /etc/cloud/cloud.cfg.d/ 目录下。搜索包含 manage_etc_hosts 配置的文件可使用以下命令：

```
grep -rn ‘manage_etc_hosts’ /etc/cloud/*
```

返回结果示例：

```
[root@localhost ~]# grep -rn 'manage_etc_hosts' /etc/cloud/*
/etc/cloud/cloud.cfg:20:manage_etc_hosts: localhost
```

2.以上述返回结果为例，可以看到在/etc/cloud/cloud.cfg 文件中的第 20 行将 manage_etc_hosts 配置成了 localhost，此时您需要修改 cloud.cfg 中的 manage_etc_hosts 配置，即：

原文：manage_etc_hosts: localhost

修改后：# manage_etc_hosts: localhost

或者也可：manage_etc_hosts: false

3.按您的业务需求修改 /etc/hosts。

9.6.3CentOS 7 重启后 dhclient 未运行，导致无法获取 IP

问题描述

云服务器启动后 dhclient 未运行导致 IP 无法获取。

问题分析

重启后 dhclient 进程未运行的根因通常为：

1.NetworkManager 未开启自启动导致 dhclient 进程未运行。

2.网卡设备未纳入 NetworkManager 管理导致。

约束与限制

本节操作适用于 CentOS 7 系列使用 DHCP 获取 IP。

1.执行以下命令，确认 dhclient 是否运行。

```
ps -ef |grep dhclient |grep -v grep
```

2.如果未找到 dhclient 进程，则确认 dhclient 进程未运行，执行以下命令，继续排查 NetworkManager 是否运行。

```
systemctl status NetworkManager
```

- 如果 NetworkManager 的状态为 Active: inactive (dead)，则 NetworkManager 未启动，执行以下命令，检查该服务是否开机自启。

```
# systemctl is-enabled NetworkManager
```

- 结果为 disabled 则确认为 NetworkManager 未设置开机自启导致，执行以下命令进行恢复。

```
# systemctl enable NetworkManager && systemctl start NetworkManager
```

- 如果 NetworkManager 的状态为 Active: active (running)，执行以下命令查看网卡设备是否被 NetworkManager 管理。

```
# nmcli device statu
```

- 如果显示该网卡为的 STATE 为 unmanaged，则该网卡设备未被 NetworkManager 管理，执行以下命令进行恢复。

```
# nmcli device set eth0 managed yes
```

3.执行以下命令重启 NetworkManager。

```
# systemctl restart NetworkManager
```

4.执行以下命令查看 ip 是否已经获取。

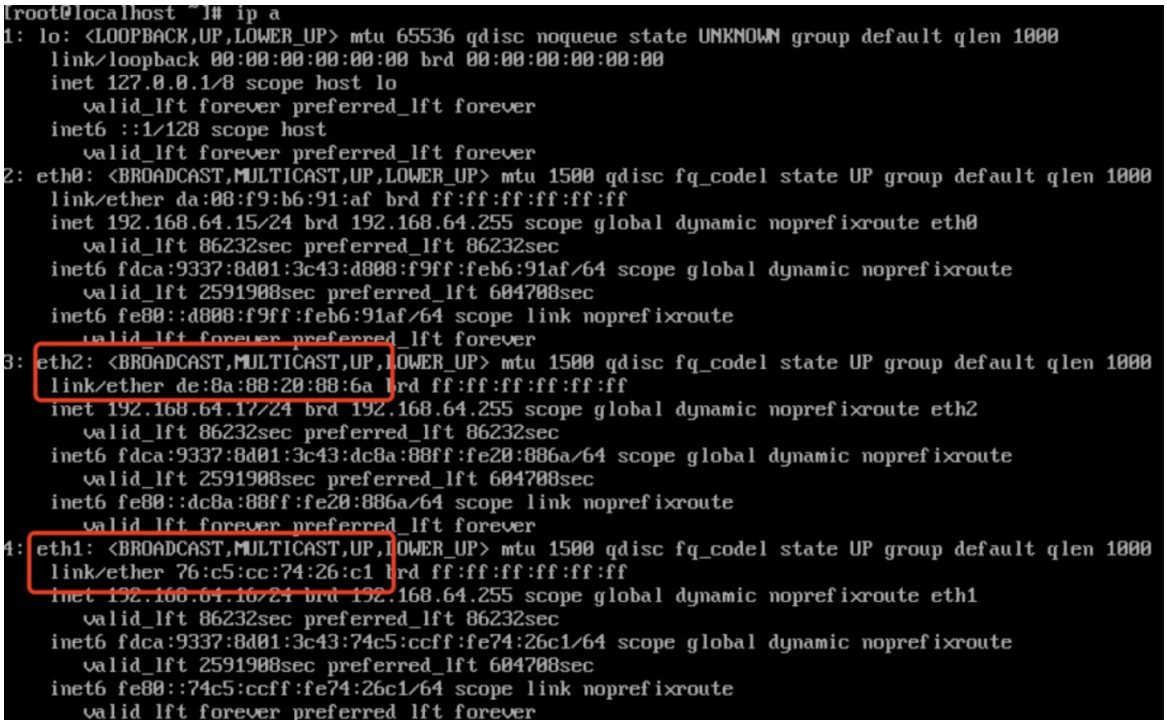
```
# ip add
```

9.6.4Linux 系统使用多网卡时网卡名称出现漂移

问题描述

若使用 Linux 系统的云主机配备了多块网卡，则在重启主机后可能出现网卡名称与网卡 MAC 地址不对应的问题，即网卡漂移问题。

以网卡 eth1 实际对应的 MAC 地址为 de:8a:88:20:88:6a、网卡 eth2 实际对应的 MAC 地址为 76:c5:cc:74:26:c1 为例，若出现网卡漂移问题，则在执行 ip a 命令查看网卡信息时，会发现网卡 eth1 对应的 MAC 地址变为了网卡 eth2 应对应的 MAC 地址。



问题原因

Linux 系统运行后会将网卡相关信息保存在 /etc/udev/rules.d/70-persistent-net.rules 文件中。若此文件不存在或文件内容和 Linux 操作系统实际情况不一致时，则可能导致网卡漂移问题。

解决步骤

注意：此问题的推荐解决方案需重启云主机，请您在评估业务中断影响后谨慎操作。

1.以问题描述中的场景为例，推荐在 /etc/udev/rules.d/70-persistent-net.rules 文件中，将网卡名称与网卡 MAC 地址绑定，使网卡名称与网卡 MAC 地址相对应。每一个网卡可添加相应的规则，每个规则的模板如下：

```
SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="<MAC 地址>", KERNEL=="eth*", NAME="<网卡名称>"
```

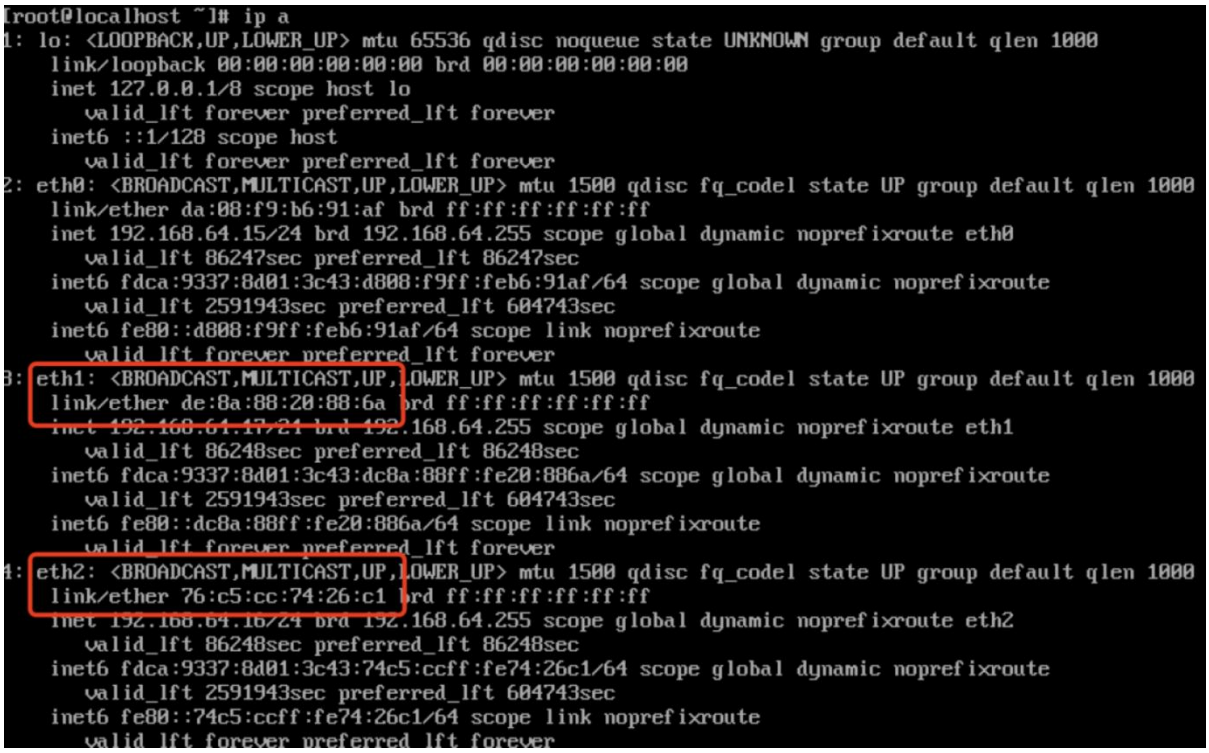
以绑定网卡 eth1 对应的 MAC 地址为 de:8a:88:20:88:6a 为例：

```
SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address de:8a:88:20:88:6a ", KERNEL=="eth*", NAME="eth1"
```

2.保存文件后重启云主机。

```
reboot
```

3.再次通过命令 ip a 查看网卡名称与 MAC 是否对应。



9.6.5 同一子网的两块网卡均绑定弹性公网 IP

问题描述

云主机上两块相同子网的网卡，均绑定了弹性公网 IP。其中，主网卡绑定的弹性公网 IP 可以正常访问，但是扩展网卡的弹性公网 IP 无法访问。

可能原因

CentOS 操作系统的弹性云主机默认开启了反向过滤技术（rpfilter），云主机的默认路由是指向 eth0 的，而扩展网卡弹性公网 IP 的流量从 eth1 进入。系统此时判断，这个报文在发送时应该从 eth0 口送出，而报文实际是从 eth1 进入，即从错误的网卡收到报文，故判定非法而被系统丢弃。

处理方法

通过策略路由让访问扩展网卡的流量从扩展网卡发出，方法如下：

1.执行以下命令，编辑文件 rt_tables。添加一个 route table 的别名，如 test。

```
vi /etc/iproute2/rt_tables
```



2.保存后退出。

3.执行以下命令，在 test 表中添加路由。

```
ip route add default via 扩展网卡网关 dev eth1 table 步骤 1 中添加的表名
```

例如：

```
ip route add default via 192.168.0.X dev eth1 table test
```

注意

如无特殊设置，则网关为子网网段的首个 IP。

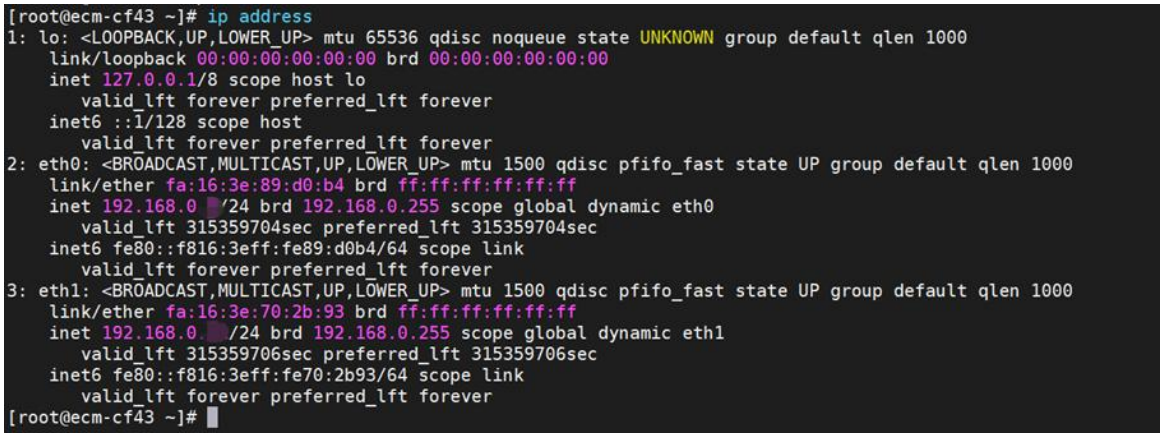
4.执行以下命令，添加策略路由。

```
ip rule add from 扩展网卡 IP 地址 lookup 步骤 1 中添加的表名 prio 低于 32766，优先级高于 main 表
```

例如：

```
ip rule add from 192.168.0.x lookup test prio 32000
```

此时使用扩展网卡的弹性 IP 登录云主机，可以登录，说明两块网卡上的公网 IP 都可以访问了。



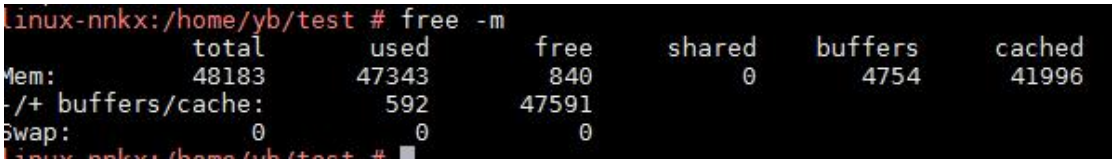
如果要持久化这个规则，可以将上述语句添加到开机脚本“/etc/rc.local”中。

9.7 磁盘空间管理

9.7.1Linux 操作系统云主机中 buffer 和 cache 占用内存怎么办？

问题描述

系统长期运行后，free 命令查看系统内存，发现剩余内存不足，大部分是 buffers 和 cached。



问题分析

在 Linux 操作系统中，缓冲（buffer）和缓存（cache）占用了部分内存。这些是操作系统用于提高文件系统和磁盘访问性能的正常数据结构。然而，如果你发现内存占用过高，可能需要采取一些措施来优化系统。

在 Linux 的内存管理中，buffer 是 Linux 内存中的 Buffer cache。cache 是 Linux 内存中的 Page cache。

- Buffer cache：主要是当系统对块设备进行读写的时候，对块进行数据缓存的系统来使用，即对块的操作会使用 buffer cache 进行缓存。

例如：当对一个文件进行写操作的时候，page cache 的内容会被改变，而 buffer cache 则可以用来将 page 标记为不同的缓冲区，并记录是哪一个缓冲区被修改了。 内核在后续执行脏数据的回写 writeback 时，就不用将整个 page 写回，而只需要写回修改的部分可。

- Page cache：主要用来作为文件系统上的文件数据的缓存来用，尤其是针对当进程文件有 read/write 操作的时候。Linux 默认会将读取的文件内容缓存在内存中，方便后续使用。

Linux 默认使用的是 lazy 模式，即内存如果还够用，则不会主动释放当前的占用的 buffer 和 cache，如果需要内存，则会自动释放 buffer 和 cache，所以正常情况下，cache 占用高不会对系统造成影响。

处理方法

buffer 和 cache 是 Linux 对系统设备的正常环境，占用内存并不会影响系统的运行，反而强制清除 cache 会导致读取磁盘数据时需要重新通过 IO 从磁盘读取，影响系统运行速度。

说明如需清理系统 buffer 和 cache，可以执行以下命令：

```
# echo 3 > /proc/sys/vm/drop_caches
```

根据内存大小可能需要几秒钟，执行完后相应内存被释放。

9.7.2Linux 文件系统提示：Read-only file system

问题现象

Linux 操作系统云主机删除或者修改文件的时候会提示文件系统：Read-only file system，最终导致操作失败。

问题分析

造成这个问题的原因大多数是因为非正常关机后导致文件系统受损引起的，在系统重启之后，受损分区就会被 Linux 自动挂载为只读。解决的方法是通过 fsck 来修复文件系统，然后重启即可。报错 read-only file system 的原因是所在的分区只有读权限，没有写权限(如下图)。



解决方法

使用 fsck 手动修复，具体操作如下：

重启系统后使用 root 进入单用户模式，运行。

```
fsck.ext3 -y /dev/vda3
```

说明

ext3 的文件系统使用 fsck.ext3，ext4 文件系统使用 fsck.etx4。/dev/vda3 是系统/根分区。运行完毕后，reboot 重启系统就恢复正常。fsck.ext3 开始进入扫描、修正文件系统，这个过程有时很快，有时比较长，中间有数次停顿的过程，只需等待即可，千万不要以为死机而重启服务器。修正完文件系统后，如果没有提示重启系统，也需要 reboot 来重启系统。

附 fsck 参数详解：

使用方法：fsck [参数] 设备名

参数：

-t : 给定档案系统的型式，若在/etc/fstab 中已有定义或 kernel 本身已支援的则不需加上此参数。

-s : 依序一个一个地执行 fsck 的指令来检查。

-A ： 对/etc/fstab 中所有列出来的 partition 做检查。

-C ： 显示完整的检查进度。

-d ： 列印 e2fsck 的 debug 结果。

-p ： 同时有-A 条件时，同时有多个 fsck 的检查一起执行。

-R ： 同时有-A 条件时，省略/ 不检查。

-V ： 详细显示模式。

-a ： 如果检查有错则自动修复,所以你不用一直按 y 键。

-y: 与-a 类似，但是某些文件系统仅支持-y 这个参数，所以也可以用-y。

-r ： 如果检查有错则由使用者回答是否修复。

9.7.3Linux 操作系统云主机磁盘分区提示空间不足怎么办？

问题描述

云主机使用过程数据持续增长，容易导致磁盘空间不足的问题，相应的对系统运行造成影响，如系统性能下降、应用程序无法正常运行、系统崩溃等。

问题定位与解决方案

1.查看磁盘空间占用情况

首先您需要了解当前磁盘下各个目录的空间占用情况，以便能够有针对性地进行优化。

执行 df -h 命令查看磁盘空间的占用情况。

执行后可见系统上所有分区与挂载点的空间使用情况，可以判断哪个目录占用了大量的磁盘空间。

2.清理临时文件

在 Linux 系统中，临时文件如果长时间不处理，将可能占用大量磁盘空间。您可以通过定期清理临时文件来释放磁盘空间。

以下是一些常见的临时文件所在的目录。

/tmp 目录是存放临时文件的默认目录，您可以执行 rm -rf /tmp/*命令清理该目录下的全部临时文件，或者执行 rm -rf /tmp/(文件名)命令清理该目录下的指定临时文件。

/var/tmp 目录也是存放临时文件的目录，您可以执行 rm -rf /var/tmp/*命令清理该目录下的全部临时文件，或者执行 rm -rf /var/tmp/(文件名)命令清理该目录下的指定临时文件。

3.清理日志文件

在 Linux 系统中，日志文件如果长时间不处理，也可能占用大量磁盘空间。您可以通过清理不必要的日志文件来释放磁盘空间。

以下是一些常见的日志文件所在的目录：

/var/log 目录存放系统日志文件，您可以执行 rm -rf /var/log/*命令清理该目录下的全部日志文件，或者执行 rm -rf /var/log/(文件名)命令清理该目录下指定日志文件。

因为 Linux 系统支持日志轮转策略，即日志文件保留一定时长后自动清理的策略，所以您可以通过调整日志轮转策略，即调整日志文件自动保留时间，达到日志文件常规管理的目的。

您可以编辑 /etc/logrotate.conf 文件来调整系统日志轮转策略。

4.压缩低频大文件

您可以通过临时压缩使用频率较低的大文件，释放部分磁盘空间。Linux 系统中常用的压缩工具有 gzip、bzip2 以及 zip 等。

5.删除不必要的软件和文件

如果您的系统中存在不使用或非必要的软件，或者无须留存的其它文件，可以通过清理它们来释放磁盘空间。

执行命令 sudo apt-get remove (软件包名) 清理软件。

执行命令 rm -rf (文件路径)/(文件名)清理指定文件。

6.扩容磁盘空间

如果以上方法无法帮助您释放出足够的磁盘空间，请您考虑扩容已有云硬盘或添加新的云硬盘，实现存储空间扩容。

9.7.4 文件已经删除，但空间未释放怎么办？

问题原因

在 Linux 或 UNIX 系统中，通过 rm 命令或者文件管理器对某个文件执行删除操作后，该文件将会从文件系统的目录结构上解除链接(unlink)同时从磁盘上清除数据。如果删除文件时，文件正在被一个或多个进程占用（如文件正在被浏览或读取），将出现解除链接(unlink)成功，而清除数据失败的情况，即空间未释放。

解决方案

查询占用文件的进程，并使用 kill 命令结束该进程。进程结束后，文件空间将自动释放。

操作步骤

1.执行以下命令，查询已删除但仍然被进程占用的文件列表。

```
$ /usr/sbin/lsof | grep deleted
```

执行命令后会获得如下回显信息（此信息为示例，请您以实际回显信息为准）。

```
cfg 27757 data 35u ocd 65,67 4295334545
45555532/oradata/DATA/fcuser.bak(deleted)
```

2.从回显信息中可见 pid 为 27757 的进程占用了刚刚删除的文件 /oradata/DATA/fcuser.bak，文件名后的“(deleted)”为文件已被删除的标记。执行以下命令，通过结束该进程来释放被占用的空间。、

```
kill -9 27757
```

9.7.5Inode 节点耗尽导致无法创建新文件问题处理

问题现象

目录中创建文件提示本地空间不足，但是通过 df -Th 发现本地目录下仍有剩余空间。

释义

在 linux 操作系统中，实际每个文件都是通过 inode 进行标识的。例如使用 cat 命令打开一个文件，对于使用 cat filename 方式打开的操作而言，实际一共经历过三个步骤，首先系统中维护一个 inode table,每个文件都有与之对应的 inode，先根据文件名称找到对应的 Inode 信息，然后将文件的详细信息存放进 inode 中。

根因分析

1.由于每个文件都必须要有一个 inode,因此有可能发生 inode 已经用光,但是硬盘还是未存满的情况。这时,就无法在硬盘上创建新文件。
2 小字节缓存文件过多,虽然占用的块不多,但是占用了大量的 inode,导致无法添加新数据。

处理方法

1.通过 df -i 查看目录下的 inode 数量，可以发现 inode 资源耗尽，所以才导致文件无法创建。

```
# df -i
```

df -i

Filesystem	Inodes	IUsed	IFree	IUse%	Mounted on
tmpfs	497833	1098	496735	1%	/run
/dev/sda2	1277952	216463	1061489	17%	/
tmpfs	497833	1	497832	1%	/dev/shm
tmpfs	497833	4	497829	1%	/run/lock
/dev/sda1	0	0	0	-	/boot/efi
tmpfs	99566	99566	0	100%	/run/user/1000

2.进行问题处理的方式就是通过删除文件，释放 inode 资源即可。
因此在云主机的日常使用中，除了要监控目录本地空间资源外，还需要监控本地 inode 资源。

后续建议

因此在云主机日常使用中除了需要监控本地目录空间资源外，还需要监控本地 inode 资源。

9.7.6Linux 如何创建 swap 分区/swap 文件

适用场景

本节操作以 CentOS 6.8 操作系统云服务器为例，指导用户创建 swap 分区。

约束与限制

操作过程中涉及创建指定大小的文件，请确认系统磁盘空间有足够的空余空间。

场景一：使用块设备创建 swap

1.执行以下命令，新建一个分区（以 2G 为例）。

```
# fdisk /dev/vdb
```

回显信息如下：

```
   Command (m for help): n
Partition type:
   p   primary (0 primary, 0 extended, 4 free)
   e   extended
```

```
Select (default p):

Using default response p

Partition number (1-4, default 1):

First sector (2048-20971519, default 2048):

Using default value 2048

Last sector, +sectors or +size{K,M,G} (2048-20971519, default 20971519):
+2G

Partition 1 of type Linux and of size 2 GiB is set


Command (m for help): p


Disk /dev/vdb: 10.7 GB, 10737418240 bytes, 20971520 sectors

Units = sectors of 1 * 512 = 512 bytes

Sector size (logical/physical): 512 bytes / 512 bytes

I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk label type: dos

Disk identifier: 0x1f02f438


   Device Boot      Start         End      Blocks   Id  System

     /dev/vdb1       2048     4196351     2097152   83   Linux


Command (m for help): w
```

```
The partition table has been altered!



Calling ioctl() to re-read partition table.

Syncing disks.
```

2.执行以下命令，将新建的分区创建为 swap。

```
# mkswap /dev/vdb1
```

```
[root@ecm-cf43 ~]# mkswap /dev/vdb1
Setting up swapspace version 1, size = 10484732 KiB
no label, UUID=
```

3.执行以下命令，激活 swap 分区。

```
# swapon /dev/vdb1
```

4.执行以下命令，查询已启动的 swap。

```
# swapon -s
```

```
[root@ecm-cf43 ~]# swapon /dev/vdb1
[root@ecm-cf43 ~]# swapon -s
Filename                                Type            Size    Used    Priority
/dev/vdb1                               partition       10484732        0       -2
[root@ecm-cf43 ~]#
```

5.执行以下命令，查询 swap 分区 UUID。

```
# blkid |grep swap |awk '{print $2}'
```

```
[root@ecm-cf43 ~]# blkid |grep swap |awk '{print $2}'
UUID="
```

6.执行以下命令，实现 swap 开机自动挂载，将挂载信息写入/etc/fstab。注意：swap 的 UUID 从步骤 5 获取。本例中执行命令如下。

```
# echo "UUID=1ee90e3c-1538-453b-9240-ad430f835f6f swap swap defaults 0 0" >>/etc/fstab
```

7.执行以下命令，挂载 swap。

```
# mount -a
```

```
[root@ecm-cf43 ~]# echo "UUID=                                    swap swap defaults 0 0" >>/etc/fstab
[root@ecm-cf43 ~]# mount -a
[root@ecm-cf43 ~]#
```

场景二：使用文件模拟的块设备做 swap 分区

说明

使用文件模拟的块设备做 swap 性能较之直接使用块设备性能较差。

1、执行以下命令，创建 1G 的 swap 文件。

```
# dd if=/dev/zero of=/swapfile bs=1M count=1000
```

```
[root@ecm-cf43 ~]# dd if=/dev/zero of=/swapfile bs=1M count=1000
1000+0 records in
1000+0 records out
1048576000 bytes (1.0 GB) copied, 7.12398 s, 147 MB/s
```

2. 执行以下命令，更改文件为 swap。

```
# chmod 600 /swapfile
```

3. 执行以下命令，更改文件属性为 swap。

```
# mkswap /swapfile
```

```
[root@ecm-cf43 ~]# chmod 600 /swapfile
[root@ecm-cf43 ~]# mkswap /swapfile
Setting up swapspace version 1, size = 1023996 KiB
no label, UUID=
```

4. 执行以下命令，启用 swap。

执行以下命令，实现 swap 开机自动挂载，将 swap 文件挂载写入/etc/fstab。

```
# echo "/swapfile swap swap defaults 0 0" >>/etc/fstab
```

5. 执行以下命令，挂载 swap。

```
# mount -a
```

```
[root@ecm-cf43 ~]# echo "/swapfile swap swap defaults 0 0" >>/etc/fstab
[root@ecm-cf43 ~]#  mount -a
[root@ecm-cf43 ~]#
```

9. 8GPU 驱动故障

9. 8. 1GPU 驱动异常怎么办？

问题描述

GPU 弹性云主机中,执行 nvidia-smi 命令查看 GPU 设备状态或使用情况时,报错无法执行指定的程序、或文件路径不存在。

可能原因

GPU 弹性云主机的 GPU 驱动程序安装异常。

处理方法

- Windows 系统处理方法:

需要进入 C:\Program Files\NVIDIA Corporation\NVSMI 目录执行。

- 非 Windows 系统的处理方法:

若未安装 GPU 驱动，请重新安装 GPU 驱动。操作指导请参考：安装 GPU 驱动。

若已安装驱动，但是驱动被无意卸载。查看是否有 /var/log/nvidia-uninstall.log 日志文件，如果有说明 GPU 驱动已被卸载，请重新安装 GPU 驱动。

若已安装驱动，但是驱动状态异常。1）执行 nvidia-uninstall 命令，卸载驱动。或者执行 sh NVIDIA-Linux-x86_64-******.run --uninstall，卸载驱动。2）重新安装 GPU 驱动。操作指导请参考：安装 GPU 驱动。

9.8.2GPU 实例启动异常，查看系统日志发现 NVIDIA 驱动空指针访问怎么办？

问题描述

GPU 弹性云主机启动异常，检查系统日志，发现 GPU 驱动提示指针访问错误日志，“BUG: unable to handle

kernel NULL pointer

dereference at **********”。

可能原因

GPU 驱动状态异常。

处理方法

1.卸载驱动。

方法一：执行 nvidia-uninstall 命令，卸载驱动。

方法二：执行 sh NVIDIA-Linux-x86_64-******.run --uninstall，卸载驱动。

2.重装驱动。

操作指导请参考：安装 GPU 驱动。

9.8.3GPU 设备显示异常

问题描述

GPU 弹性云主机中，执行 nvidia-smi 命令查看 GPU 设备状态或使用情况时，有如下问题：

- 单卡 GPU 弹性云主机上，报错“No devices were found”。
- 多卡 GPU 弹性云主机上显示卡数目不全，执行 lspci|grep -i nvidia 显示 GPU 卡数目正常。

可能原因

GPU 设备或者驱动状态异常。

处理方法

请尝试在控制台，执行重启 GPU 弹性云主机操作。再执行 nvidia-smi 查看 GPU 设备信息和状态，确认 GPU 数量是否正常。

如果问题依然存在，请联系客服，由技术支持人员处理。

9.8.4T4 GPU 设备显示异常

问题描述

G6 型或 PI2 型规格的 GPU 弹性云主机中，执行 nvidia-smi 命令查看 GPU 设备状态或使用情况时，有如下问题：

- 单卡 GPU 弹性云主机上，报错“No devices were found”。
- 多卡 GPU 弹性云主机上显示卡数目不全，执行“lspci|grep -i nvidia“显示 GPU 卡数目正常。

可能原因

NVIDIA Tesla T4 GPU 默认使用并开启 GSP Firmware，导致 GPU 无法识别。

处理方法

1.执行以下命令，移除 NVIDIA 内核模块。

```
rmmod nvidia_drm

rmmod nvidi_modeset

rmmod nvidia
```

2.执行以下命令，关闭 GSP Firmware 开关，并载入 NVIDIA 内核模块。

```
modprobe nvidia NVreg_EnableGpuFirmware=0

modprobe nvidia_drm

modprobe nvidia_modeset
```

以上操作只对当前运行的系统生效。如需持久化设置，需要在配置文件 /etc/modprobe.d/nvidia.conf 中增加一行“options nvidia NVreg_EnableGpuFirmware=0“。

3.如果问题依然存在，请联系客服，由技术支持人员处理。

9.8.5G 系列弹性云主机 GPU 驱动故障

问题描述

在 Windows 系统的 G 系列弹性云主机中，打开 NVIDIA 控制面板失败，GPU 驱动显示异常并且无法使用。

可能原因

G 系列弹性云主机的 Grid GPU 驱动状态异常。

处理方法

1.打开设备管理器，在显示适配器选项中查看 GPU 设备驱动状态。

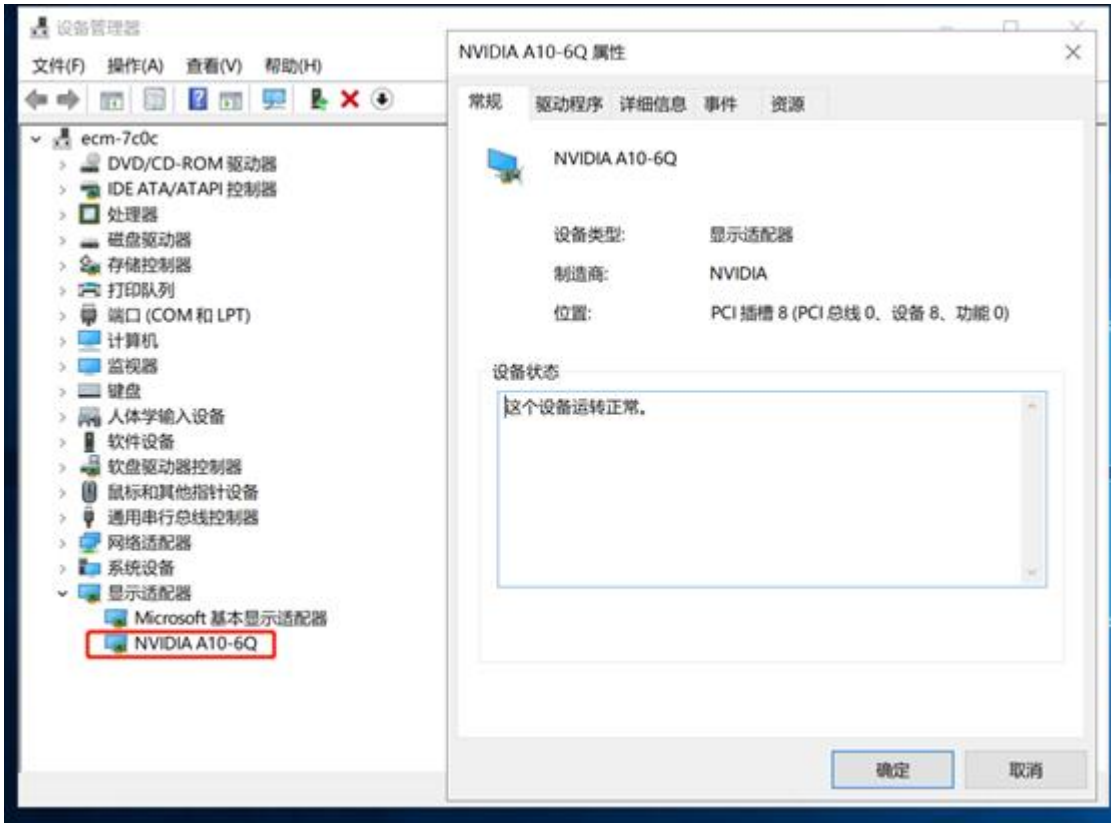
(1) 显示适配器选项中无 GPU 设备信息，如下图所示，执行步骤 2。



(2) 显示适配器选项中，GPU 设备有黄色感叹号，查看属性，显示该设备有问题，执行步骤 2。

(3) 显示适配器选项中，GPU 设备驱动显示正常，查看属性，提示要使这个设备生效，需要重启计算机，执行步骤 2。

重启 GPU 弹性云主机。打开设备管理器，在显示适配器选项中查看 GPU 设备和驱动是否正常。如果设备和驱动状态正常，任务结束。否则，重新安装 Grid 驱动。正常状态，如下图所示：



详细操作指导请参考：安装 Grid GPU 驱动。

9.8.6GPU 驱动不可用

问题描述

GPU 弹性云主机中，执行 nvidia-smi 命令查看 GPU 设备状态或使用情况时，报错无法执行指定的程序。具体显示如下：

```
NVIDIA-SMI has failed because it couldn't communicate with the NVIDIA driver. Make sure that the l
atest NVIDIA driver is installed and running.
```

可能原因

GPU 弹性云主机的 GPU 驱动状态异常。

问题排查

不同的操作系统执行的命令有所不同，以 CentOS 系统为例，查看安装驱动时的内核版本：

```
find /usr/lib/modules -name nvidia.ko
```



执行 uname -r 查看当前内核版本：



如果安装驱动的内核版本跟当前内核版本不一致，则是导致此问题的原因。

处理方法

方法一：重启系统，选择安装 GPU 驱动时的内核版本启动系统。

方法二：在现有内核版本的基础上，重新安装驱动。

(1) 执行 nvidia-uninstall 命令，卸载驱动。或者执行 sh NVIDIA-Linux-x86_64-******.run --uninstall，卸载驱动。

(2) 重新安装 GPU 驱动。操作指导请参考：安装 GPU 驱动.

9.9SSH 连接

9.9.1 怎样长时间保持 SSH 会话连接不断开？

操作场景

使用 SSH 方式登录 CentOS7.6 操作系统的云服务器时，过一段时间就会自动断开连接。本节操作介绍如何保持 SSH 会话持续连接不断开。

该文档适用于 CentOS 系统。

操作方法

1.编辑/etc/ssh/sshd_config 文件设置心跳，保持连接。

编辑/etc/ssh/sshd_config，添加配置项：

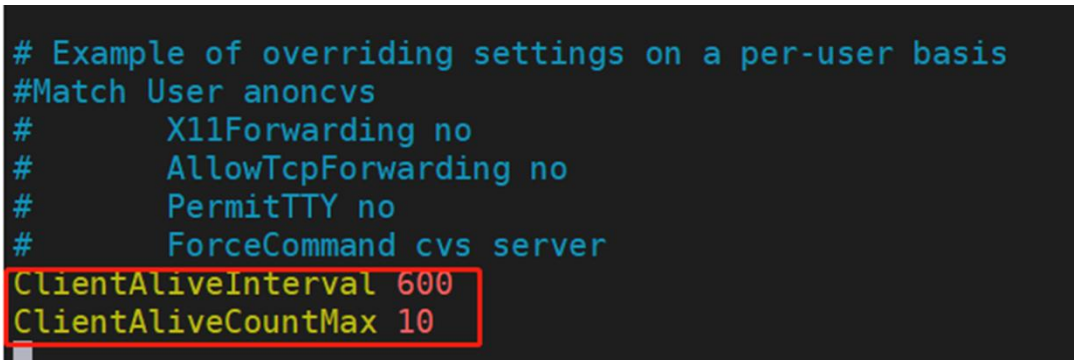
```
vim /etc/ssh/sshd_config#添加配置项 ClientAliveInterval 600
ClientAliveCountMax 10
```

ClientAliveInterval 600 表示每 600 秒发送一次请求，

从而保持连接。

ClientAliveCountMax 10 表示服务器发出请求后客户端没有响应的次数达到 10 次，就自动断开连接。

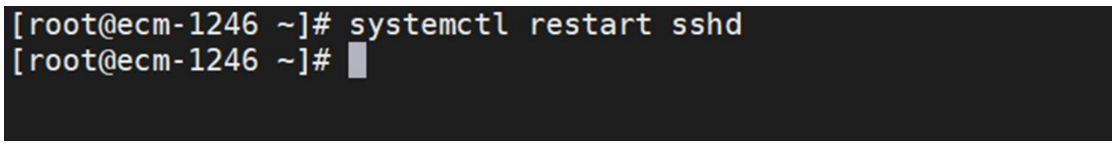
则无响应的 SSH 客户端将在大约 600x10=6000 秒后断开连接。



执行 :wq 保存修改。

2.执行以下命令，重启 sshd 服务，使配置生效。

```
# systemctl restart sshd
```



9.9.2 /etc /passwd 文件损坏导致云主机登录失败怎么办？

操作场景

本节操作适用于 Linux 操作系统云主机，出现因/etc/passwd 文件损坏导致云主机无法登录的问题。

说明

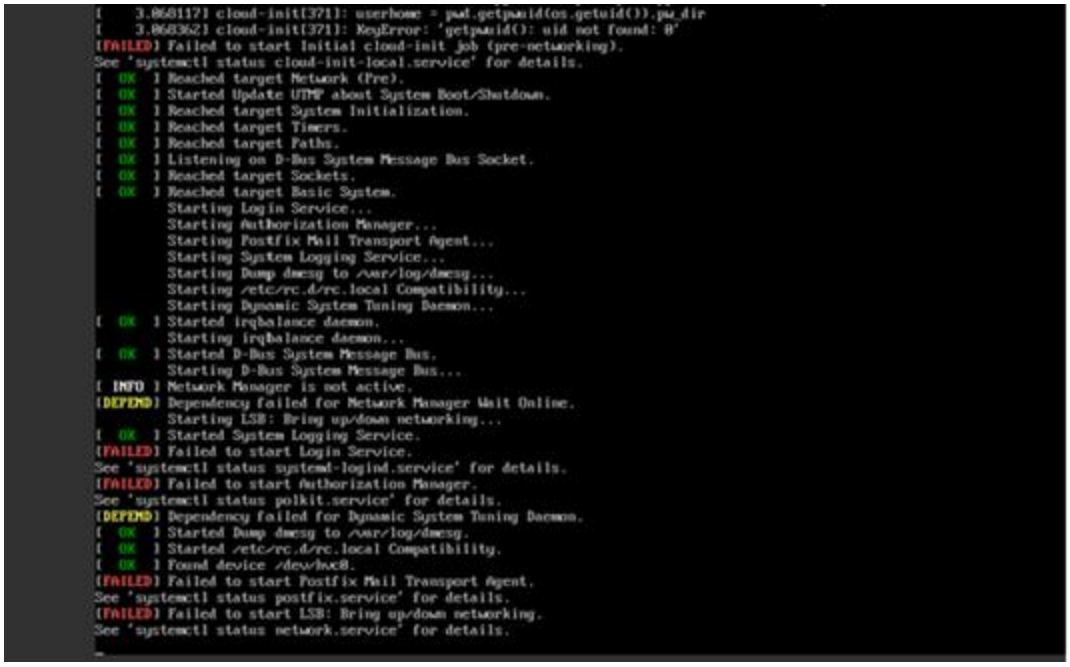
本节操作为紧急恢复系统方法，需要在单用户模式下会将系统备份初始备份/etc/passwd-文件替换已损坏的/etc/passwd 文件，该操作会造成自行添加的用户丢失（包括应用运行的用户，可以参考/etc/shadow 文件添加其他账号）。

本节操作涉及重启云主机操作，重启云主机会造成业务中断，请谨慎操作。

问题描述

Linux 系统中多个服务启动失败：Failed to start Login service 、Filed to start Authorization service。

待系统启动后登录，提示密码错误。



根因分析

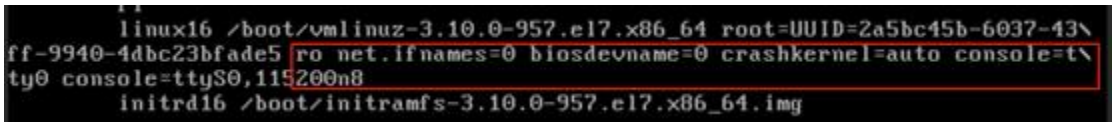
/etc/passwd 和/etc/shadow 文件记录所有的用户信息，每个用户都有一个对应的记录行，如果该文件损坏或者误删除会导致登录服务（systemd-logind.service）启动失败，因此用户无法登录。

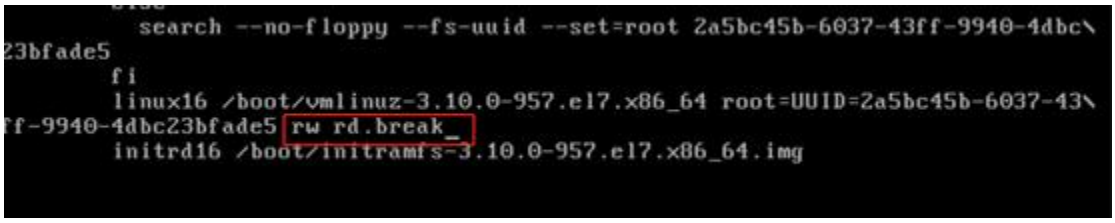
处理方法

1.使用控制台提供的远程登录方式登录云主机。单击远程登录操作面板上方的“发送 CtrlAltDel”按钮，重启虚拟机。

2.确认开始重启后，立即反复单击上下键，阻止系统继续启动，在出现内核选项时按字母键 e 进入内核编辑模式。

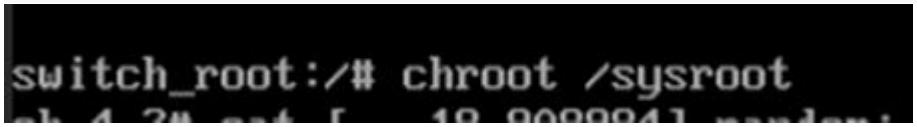
3.找到 linux16 行末尾，删除不需要加载的参数（ro 参数开始到末尾）。修改 ro 为 rw，以读写方式挂载根分区。并添加 rd.break，然后执行 Ctrl+X。





4.执行以下命令切换至/sysroot 目录。

```
# chroot /sysroot
```



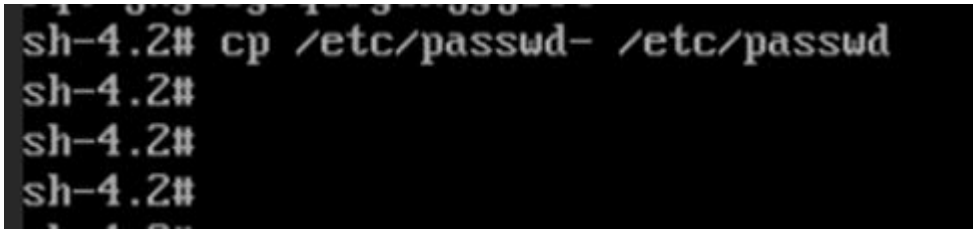
5.执行以下命令检查/etc/passwd 文件。

```
cat /etc/passwd
```



6.确认 passwd 文件已被破坏，执行以下命令使用系统初始的备份 passwd-文件替换损坏的 passwd 文件。

```
# cp /etc/passwd- /etc/passwd
```



说明

该操作会造成自行添加的用户丢失，如果为应用运行的用户会导致应用启动失败，待修复后请自行添加用户。

7. 执行两次以下命令退出当前目录。

```
# exit
```

8. 云主机自动重启，使用 root 可登录云主机，查看/etc/passwd 已正常。



9.9.3 怎样设置允许或禁止用户/IP 通过 SSH 连接云主机？

操作场景

出于对云主机系统安全的考虑，需要对操作系统设置用户/IP 是否可以通过 SSH 方式连接，以深度管理云主机登录权限。

操作方法

您可以选择适用您需求的设置方式，实现限制登录的目的。

1. 通过编辑 sshd 配置文件，可以对指定用户/用户组或 IP 的登录权限设置。

1）编辑 sshd 配置文件中为指定用户设置白名单。

在 /etc/ssh/sshd_config 配置文件中添加 AllowUsers 配置，以下示例中的实际效果为允许用户 user 通过 192.168.8.8 的 IP 地址进行登录。

```
AllowUsers user@192.168.8.8
```

2）编辑 sshd 配置文件中为指定用户设置黑名单。

在 /etc/ssh/sshd_config 配置文件中添加 DenyUsers 配置，以下示例中的实际效果为禁止用户 invaliduser 登录。

```
DenyUsers invaliduser
```

2.通过编辑 host.allow 和 host.deny 文件，可以对指定 IP 或 IP 地址段的登录权限进行设置。通常这两个文件结合设置。

1）编辑 host.allow 文件，添加 sshd: IP 地址或 IP 地址段，允许指定 IP 地址或 IP 地址段进行登录，示例如下：

```
sshd: 192.168.8.8
```

2）编辑 host.deny 文件，添加如下内容，禁止所有 IP 的 SSH 登录权限：

```
sshd: ALL
```

两个文件同时设置规则的时候，hosts.allow 文件中的规则优先级高于 hosts.deny，结合使用可以有针对性地开放 SSH 登录权限。

因此以上示例的实际效果，云主机将只允许 IP 地址为 192.168.8.8 的用户进行 SSH 登录，其它的 IP 都会拒绝。

3.重启 sshd 服务，使以上修订内容生效。

9.9.4Linux 启动 sshd 服务出现/var/empty/sshd 无法访问的解决方案

问题现象

/var/empty/sshd 属主属组异常或目录缺失，导致 Linux 操作系统 sshd 服务启动失败。

适用场景

本节操作以 CentOS7/CtyunOS 系统为示例，其他 Linux 系统可能存在差异。

场景一：/var/empty/sshd 属主非 root 导致 sshd 启动失败

1.执行 systemctl restart sshd 时，服务重启失败，云主机无法正常远程连接。

```
systemctl restart sshd
```

```
systemctl status sshd
● sshd.service - OpenSSH server daemon
   Loaded: loaded (/usr/lib/systemd/system/sshd.service; enabled; vendor preset: enabled)
   Active: activating (auto-restart) (Result: exit-code) since Thu 2023-09-07 10:36:58 CST; 11s ago
     Docs: man:sshd(8)
           man:sshd_config(5)
  Process: 1548 ExecStart=/usr/sbin/sshd -D $OPTIONS (code=exited, status=255)
 Main PID: 1548 (code=exited, status=255)

Sep 07 10:36:58 contos7 systemd[1]: sshd.service: main process exited, code=exited, status=255/n/a
Sep 07 10:36:58 contos7 systemd[1]: Failed to start OpenSSH server daemon.
Sep 07 10:36:58 contos7 systemd[1]: Unit sshd.service entered failed state.
Sep 07 10:36:58 contos7 systemd[1]: sshd.service failed.
```

2.执行以下命令，查看 journal 日志

```
journalctl -xe
```

提示“/var/empty/sshd must be owned by root”，从日志内容分析，由于 /var/empty/sshd 目录属主非 root 用户导致 sshd 服务启动失败。

```
Sep 07 10:41:11 contos7 systemd[1]: Failed to start OpenSSH server daemon.
Sep 07 10:41:11 contos7 systemd[1]: Unit sshd.service entered failed state.
Sep 07 10:41:11 contos7 systemd[1]: sshd.service failed.
Sep 07 10:41:54 contos7 systemd[1]: sshd.service holdoff time over, scheduling restart.
Sep 07 10:41:54 contos7 systemd[1]: Stopped OpenSSH server daemon.
Sep 07 10:41:54 contos7 systemd[1]: Starting OpenSSH server daemon...
Sep 07 10:41:54 contos7 sshd[1603]: /var/empty/sshd must be owned by root and not group or world-writable.
Sep 07 10:41:54 contos7 systemd[1]: sshd.service: main process exited, code=exited, status=255/n/a
```

3.执行以下命令，查看/var/empty/sshd 文件属主信息。

```
# ls -alF /var/empty/sshd
```

```
ls -alF
total 0
drwxr-xr-x.  3 root  root   18 Jul 27 10:44 ./
drwxr-xr-x. 19 root  root  267 Jul 27 10:48 ../
drwx--x--x.  2 linux linux   6 Aug  5 00:00 sshd/
```

4.修改/var/empty/sshd 属主及其权限。

```
chown -R root.root /var/empty/sshd

chmod -R 711 /var/empty/sshd
```

5.执行以下命令重启 sshd 服务：

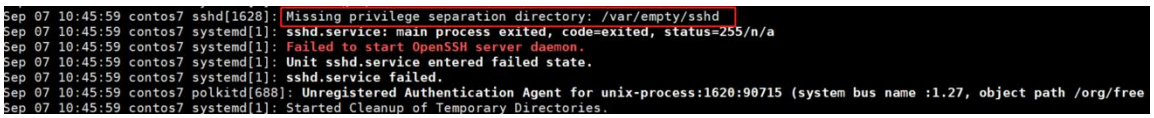
```
systemctl restart sshd
```

场景二：/var/empty/sshd 文件缺失导致 sshd 启动失败

1.执行以下命令，打开 journal 日志，查看服务失败原因。

```
journalctl -xe
```

从下图示例中可知由于/var/empty/sshd 缺失导致 sshd 启动失败。



2.执行以下命令，手动创建/var/empty/sshd 目录。

```
mkdir -p /var/empty/sshd
```

3.重启 sshd 服务。

```
systemctl restart sshd
```

9.9.5CentOS 7 修改 SSH 默认端口后无法连接怎么办？

问题现象

由于默认远程登录端口容易遭受暴力破解，导致一系列使用问题。基于安全性的考虑，修改默认远程登录端口是常用的安全防护手段。但修改后可能会存在无法远程登录的问题。本文以 CentOS 7 为例，介绍此类问题的解决方案。

前提条件

操作系统内已将默认远程登录端口修改为 1234。

解决方案

1、查看安全组是否放通 1234 端口的入方向权限。如果未放通，请添加规则放通该端口。

安全组规则关键信息：

方向：入方向

授权策略：允许

协议：TCP

端口：1234

远端：0.0.0.0/0（如果您的访问地址为固定地址或地址范围，可根据实际情况修订）



2. 查看操作系统是否开启防火墙。执行以下命令查看防火墙状态，可见示例中 active 为开启状态。

```
systemctl status firewalld
```



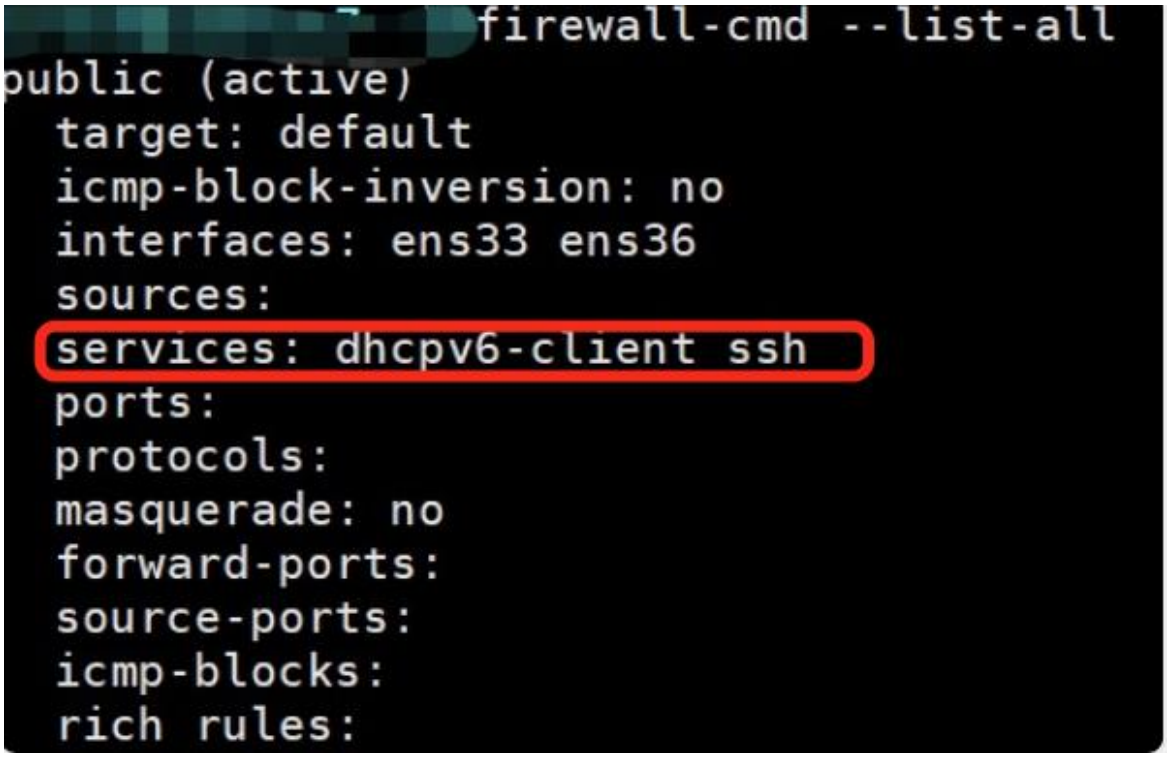
如果您需要使用防火墙，请前进到步骤 3 进一步确认防火墙规则。如果您无需使用，可执行下方命令关闭防火墙，以及关闭防火墙开机自启动；

```
systemctl stop firewalld

systemctl disable firewalld
```

3. 执行以下命令，查看防火墙规则，确认规则中是否放行 1234 端口。示例中，防火墙当前 zone 为 public，该 zone 只放行了 dhcpv6-client 和 ssh 服务，意味着 ssh 服务默认端口 22 被放行，但当 ssh 修改为其它端口将无法访问。

```
# firewall-cmd --list-all
```



4.执行以下命令，在防火墙中添加端口 1234，并重启防火墙服务。

```
firewall-cmd --add-port=1234/tcp --permanent --zone=publicfirewall-cmd
--reload
```

9.9.6 开启 UseDNS 导致 SSH 连接缓慢怎么办？

问题现象

您的操作系统开启了 UseDNS 后，使用 SSH 方式连接云主机，出现停顿等待且需要较长时间才可连接的现象。

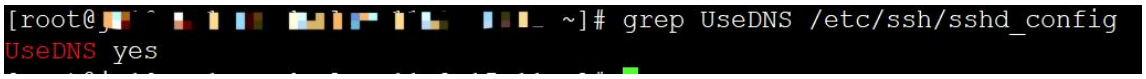
根因分析

UseDNS 特性是 SSH 服务的安全增强特性，登录过程中会进行 DNS 正/反解析，验证客户端的主机名和 IP 是否一致，以增强登录验证的安全性。开启后，由于连接过程会增加 DNS 解析过程，加长了 SSH 连接时间。通常情况下，客户端使用的是动态 IP，没有相应的 PTR 记录。如果开启该特性，不仅无法用于信息比对，

反而由于相关查询操作增加了操作延迟，最终导致客户端连接速度变慢。因此关闭 UseDNS 可加快 SSH 连接速度。

执行如下命令确认是否开启了 UseDNS 选项，如果该选项值为“yes”或者为注释行则说明已开启该选项。

```
# grep UseDNS /etc/ssh/sshd_config
```



处理方法

1.执行以下命令，编辑/etc/ssh/sshd_config 文件。

```
vi /etc/ssh/sshd_config
```

2.修改 UseDNS 选项值为 no（注意配置文件格式）。

```
UseDNS no
```

3.重启 sshd 服务，无报错则为重启成功。

CentOS6/ubuntu 执行：

```
# service sshd restart
```

CentOS7/CtyunOS 执行：

```
# systemctl restart sshd
```

9.9.7 怎样禁用 SSH 密码方式连接云主机？

操作场景

对访问云主机安全性的考虑，SSH 密钥连接方式的安全性高于密码连接方式。选择密钥方式连接云主机后，可以禁用密码连接云主机的方式。

```
说明
```

该设置方法仅适用于 SSH 远程连接操作，控制台登录过程仍然使用密码登录，请妥善保存云主机登录密码，如果密码丢失可执行重置密码操作。

操作步骤

1.登录 Linux 云主机，执行以下命令打开 sshd_config 文件的编辑界面。

```
vi /etc/ssh/sshd_config
```

2.编辑 SSH 连接方式，把 PasswordAuthentication yes 改为 PasswordAuthentication no。

```
[root@ ... mnt]# grep PasswordAuthentication /etc/ssh/sshd_config
PasswordAuthentication no
```

3.重启 sshd 服务后使更新后的配置生效，无报错则为重启成功。

CentOS6/ubuntu 执行：

```
# service sshd restart
```

CentOS7/CtyunOS 执行：

```
# systemctl restart sshd
```

重启该云主机后，使用 SSH 密码方式访问云主机，如果提示“Disconnected:No supported authentication methods available”，说明已成功禁用 SSH 密码连接方式。

9.10 多用户登录

9.10.1 配置多用户登录后，普通用户登录闪屏怎么办？

一 问题描述

Windows 弹性云主机配置多用户登录后，Administrator 登录正常，普通用户登录后出现闪屏，或者打开"我的电脑"出现自动关闭，不能正常使用。

二 处理方法

1.首先使用 Administrator 用户登录弹性云主机，查看系统日志及应用日志，查找异常模板，比如发现 Mglayout64.dll 模块出现异常。

2.打开 C:\Windows\进行搜索模块名称，找到对应模块文件。

3.打开“运行 > cmd”，进入到错误文件所在目录。

执行如下命令：

```
cd C:\Windows\System32
```

4.通过命令进行移除，移除后恢复正常。

执行如下命令：

```
Regsvr32.exe /u Mglayout64.dll
```

9.10.2Windows 云主机如何配置多用户登录？（Windows 2008）

操作场景

本文档将指导您配置多用户登录的 Windows 云主机，以 Windows Server 2008 标准版 R2 中文版为例。

说明

远程桌面授权仅支持 120 天，过期后将因为缺失远程桌面授权服务器许可证而导致多用户登录无法使用，需要重新激活。配置多用户登录后，如果出现多用户登录 Windows 主机时无法打开浏览器的问题，解决方法请参考多用户登录 Windows 主机时无法打开浏览器。

操作步骤

1.安装桌面会话主机和远程桌面授权。

(1) 登录 Windows 云主机。

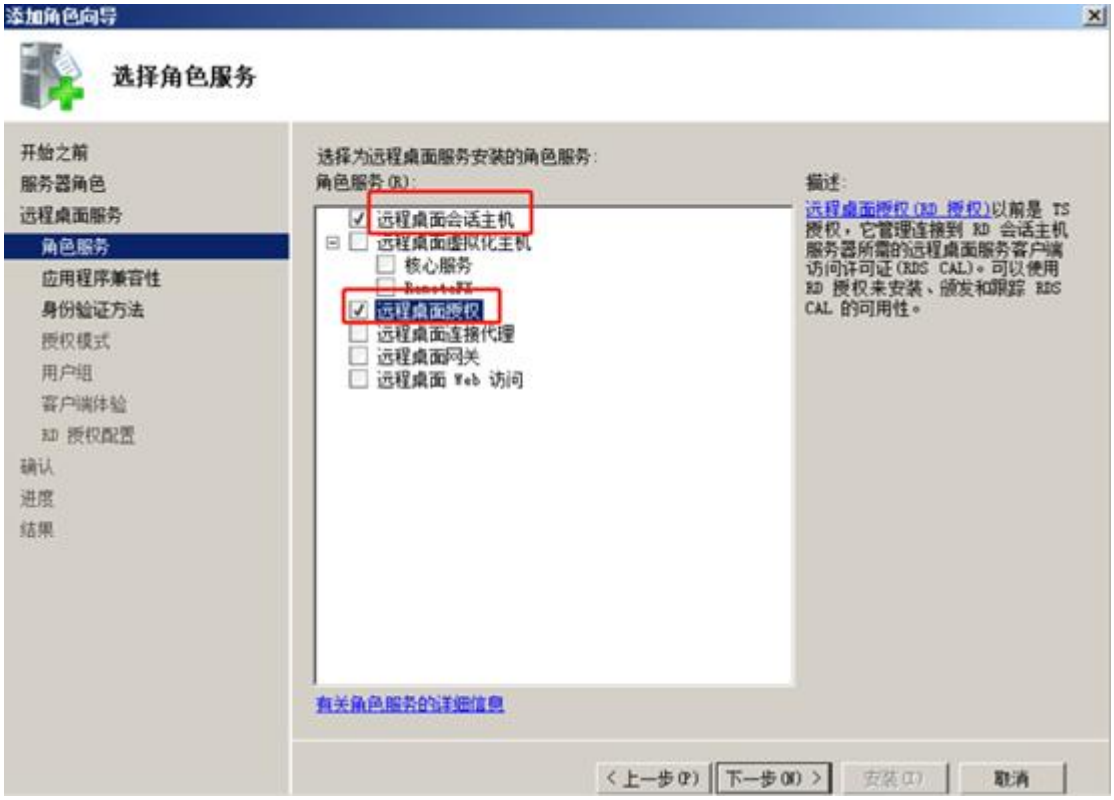
(2) 在操作系统界面，打开“服务器管理器”，单击“角色>操作>添加角色”。

(3) 跳转至左侧导航栏为“服务器角色”，选择“远程桌面服务”，单击“下一步”。



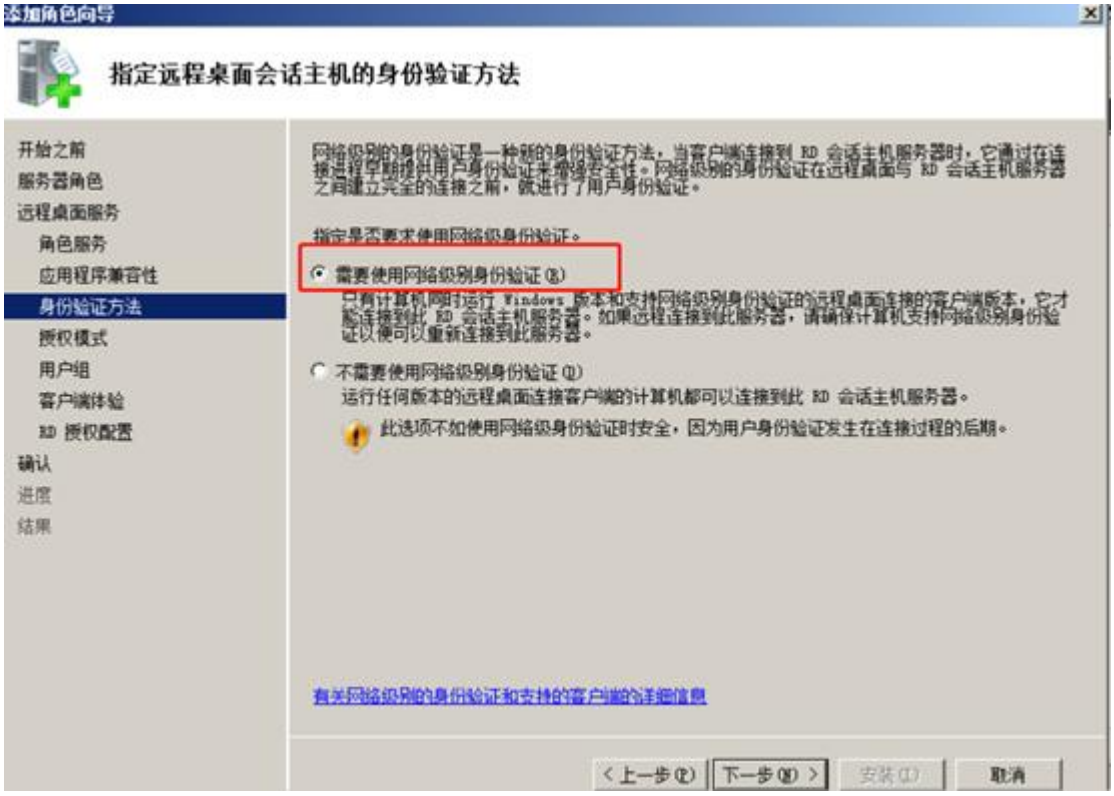
(4) 保持默认参数，单击“下一步”。

(5) 跳转至左侧导航栏为“角色服务”，依次勾选“远程桌面会话主机”和“远程桌面授权”，单击“下一步”。

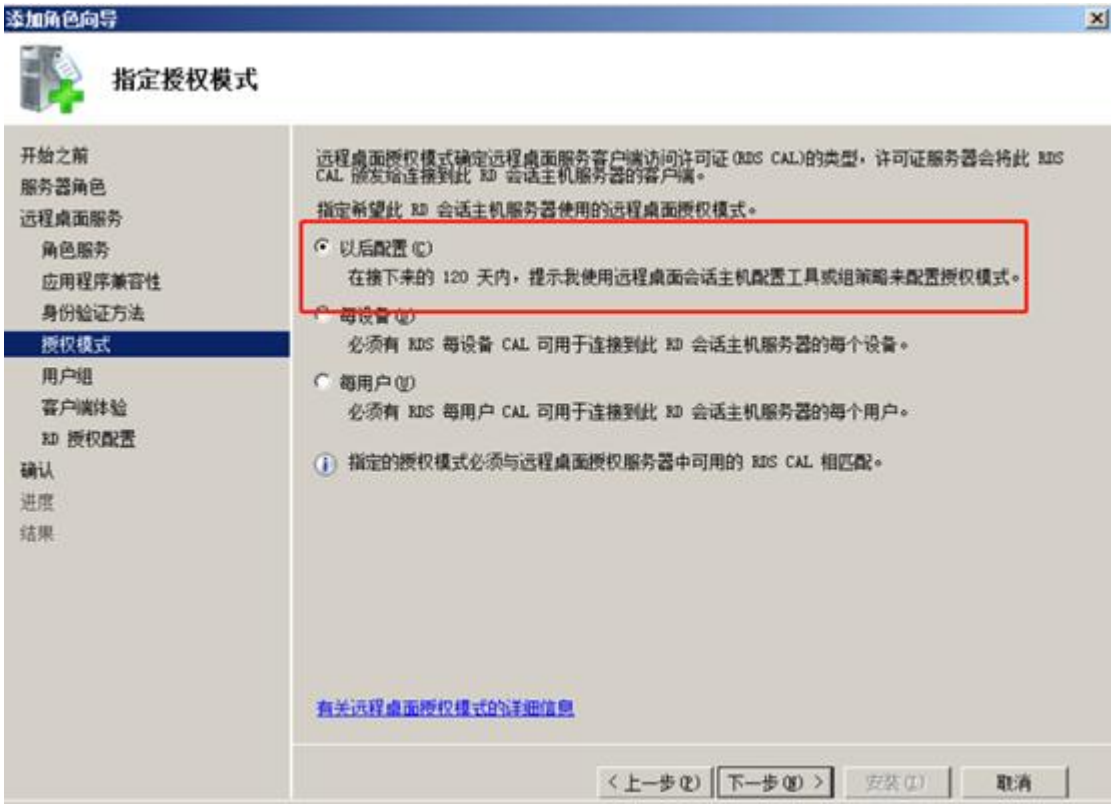


(6) 保持默认参数，单击“下一步”。

(7) 跳转至左侧导航栏为“身份验证方法”，指定远程桌面会话主机的身份验证方法，单击“下一步”。本示例选择的是“需要使用网络级别身份验证”，之后单击“下一步”。



(8) 跳转至左侧导航栏为“授权模式”，指定授权模式。本示例选择“以后配置”，之后单击“下一步”。



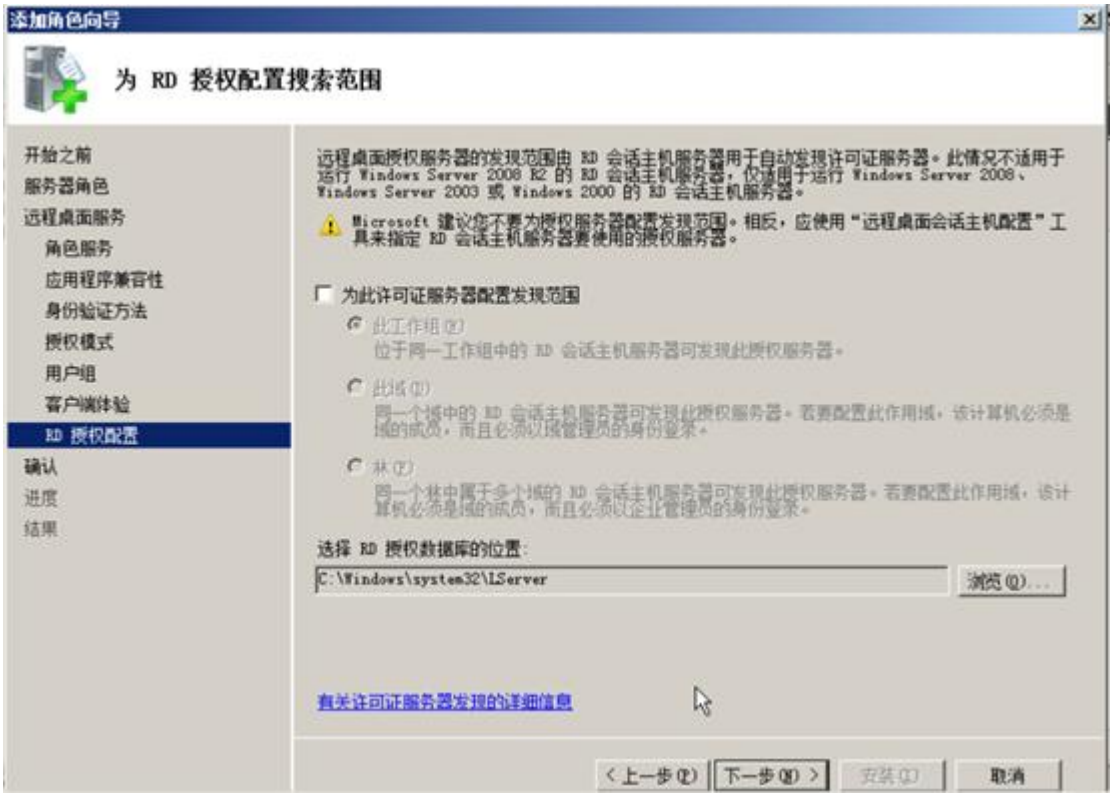
(9) 跳转至左侧导航栏为“用户组”，选择允许访问此 RD 会话主机服务器的用户组，单击“下一步”。



(10) 跳转至左侧导航栏为“客户端体验”。本示例选择的是“音频和视频播放”，之后单击“下一步”。



(11) 跳转至左侧导航栏为“RD 授权配置”，为 RD 授权配置搜索范围，配置完成后单击“下一步”。

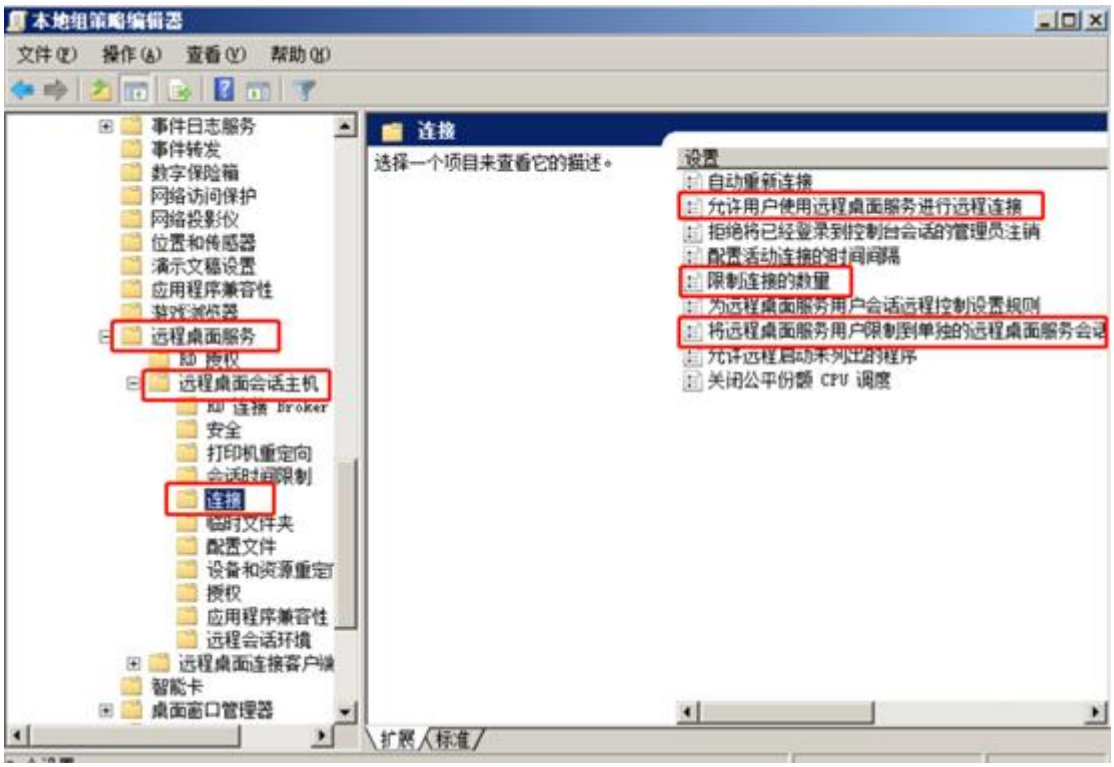


(12) 单击“安装(I)”。

2.允许多用户远程连接云主机。

(1) 在运行里输入 gpedit.msc，打开计算机本地组策略。

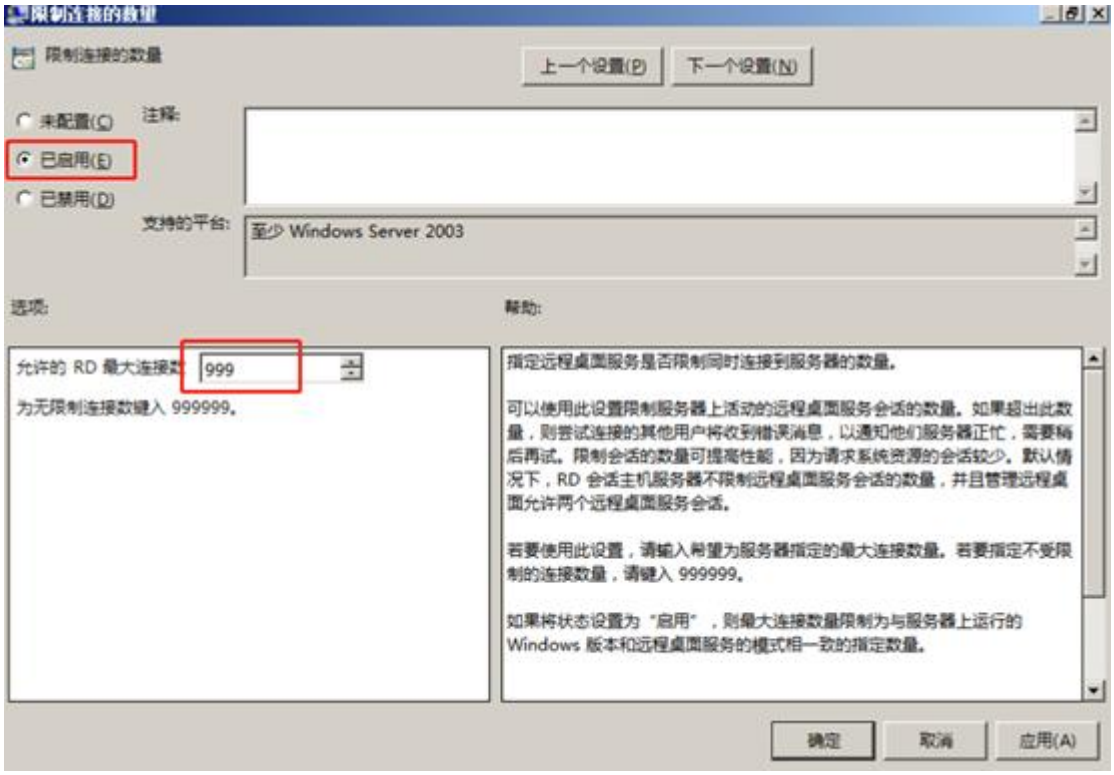
(2) 选择“计算机配置>管理模板>windows 组件>远程桌面服务>远程桌面会话主机>连接”。设置“允许用户通过使用远程桌面服务进行远程连接”、“限制连接数量”和“将远程桌面服务用户限制到单独的远程桌面”。



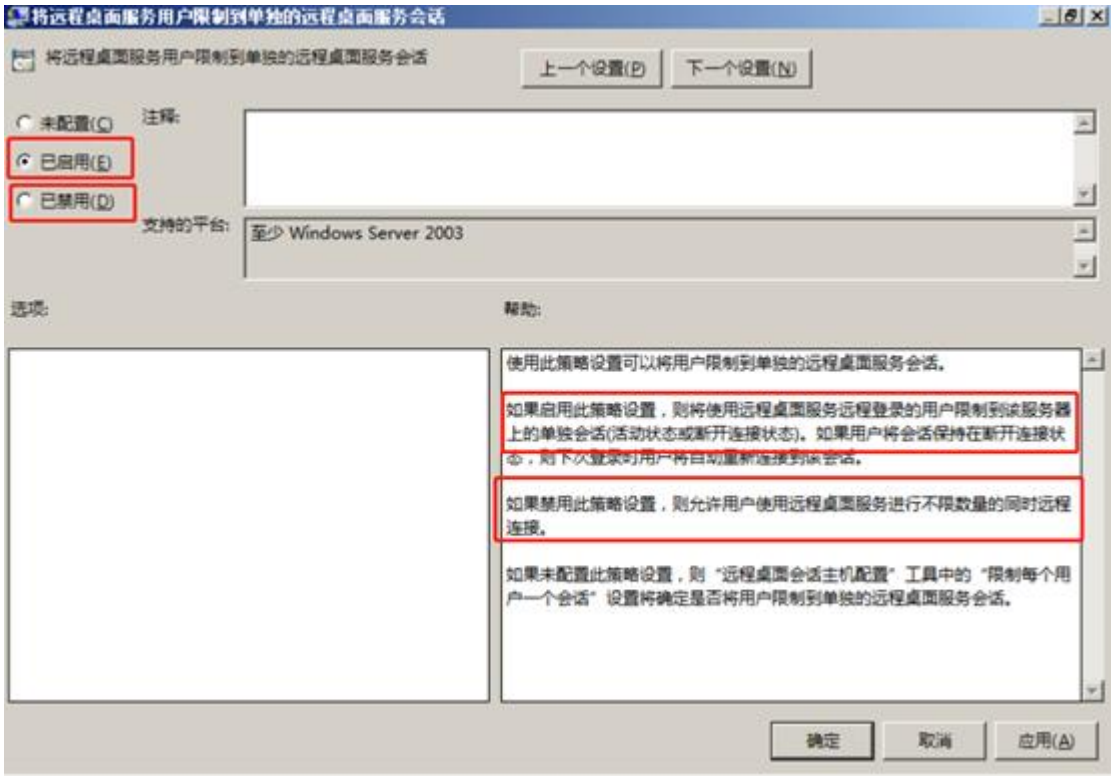
(3) 选中“允许用户使用远程桌面服务进行远程连接”，右键单击“编辑”，将其设置为已启用。



(4) 选中“限制连接数量”，右键单击“编辑”，选择已启用，可根据具体数量设置，本示例设置允许的 RD 最大连接数为 999。



(5) 选中“将远程桌面服务用户限制到单独的远程桌面服务会话”，右键单击“编辑”，选择已启用，或者已禁用。本示例以勾选“已启用”为例。



说明

请注意此处已启用和已禁用的区别：

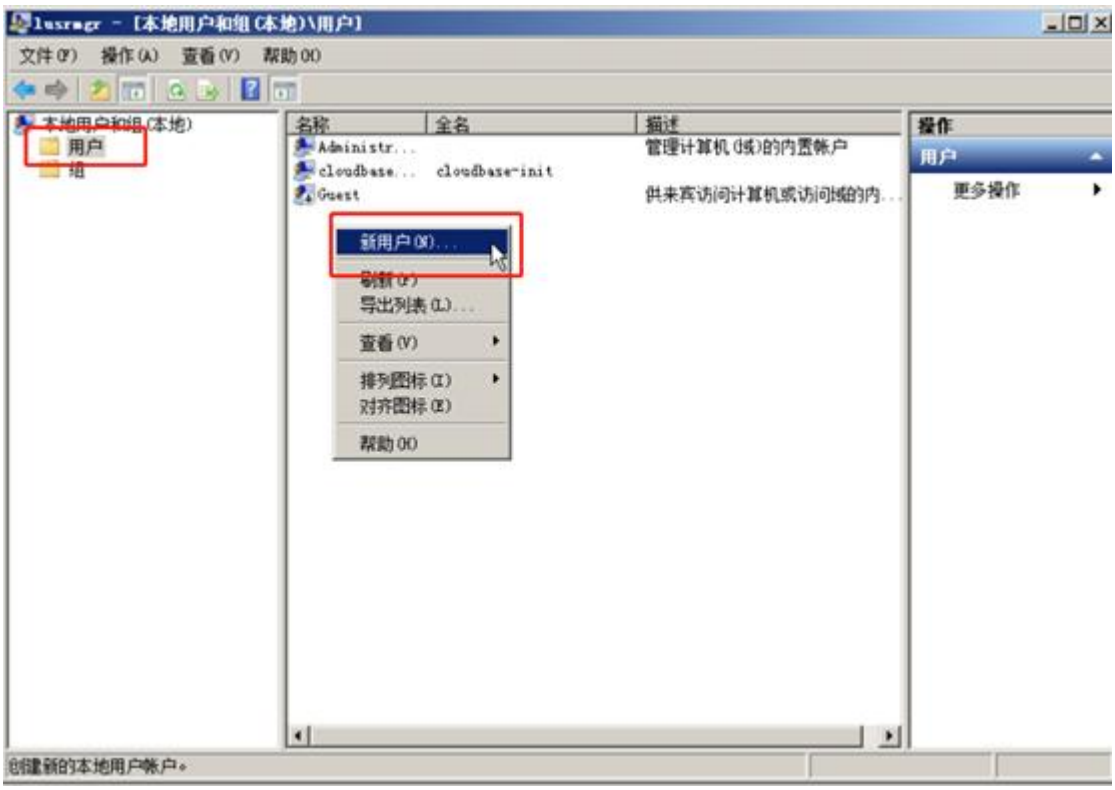
- 1 已启用：允许多个用户同时登录，不能单个用户同时多人登录。 例如：配置为已启用，用户 A、用户 B、用户 C 可以分别使用账号 A、账号 B、账号 C 同时登录云主机，但是用户 A、用户 B、用户 C 无法使用同一账号同时登录云主机。
- 已禁用：允许单个用户同时多人登录。例如：配置为已禁用，用户 A、用户 B、用户 C 可以使用同一个账号同时登录云主机。

(6) 运行 cmd，输入 gpupdate /force，强制执行本地组策略，重启服务器，整个配置过程完成。

3.配置新用户并加入远程桌面用户组。

(1) 在运行中输入 lusrmgr.msc，打开本地用户和组，进行新用户创建。

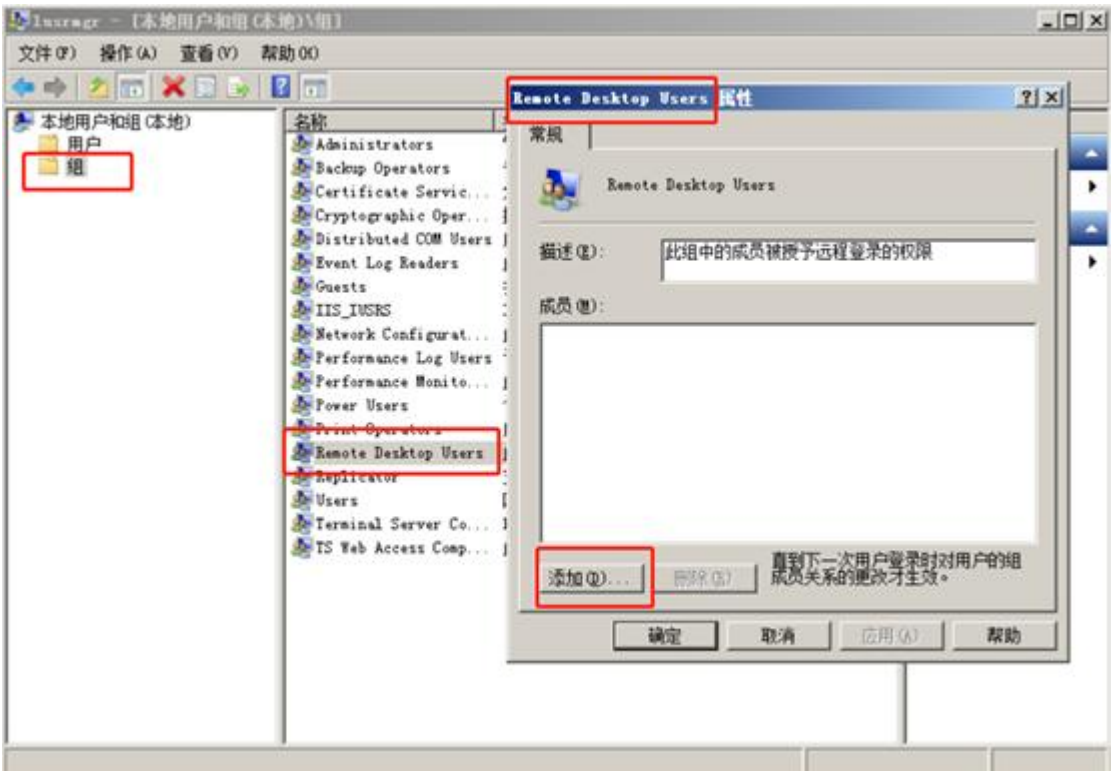
(2) 单击“用户”，在空白处右键选择新用户。



(3) 填写新用户信息，单击“创建”。



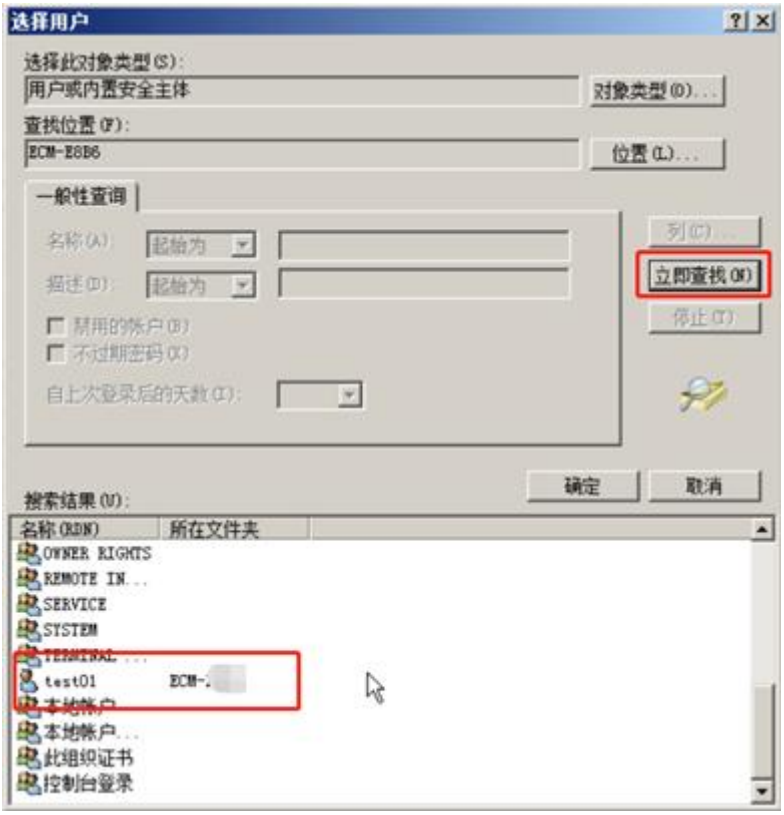
(4) 单击“组”，双击打开 Remote Desktop Users 单击“添加”。



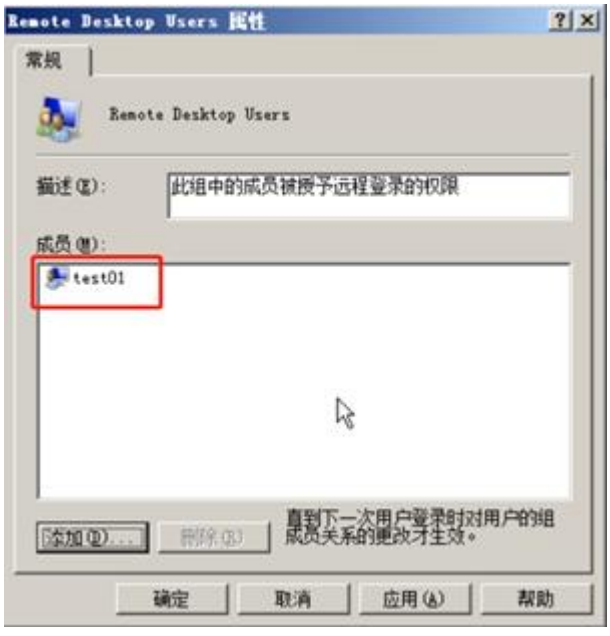
(5) 进入选择用户界面，单击“高级”。



(6) 在新的选择用户界面，单击“立即查找”，在下方搜索结果中选中需要远程登录的用户，并单击“确定”，完成添加，即可远程登录。



(7) 单击“确定”，添加用户 Remote Desktop Users 组。



9.10.3Windows 云主机如何配置多用户登录？（Windows 2012）

操作场景

本文档指导您配置 Windows 云主机实现多用户登录，以 Windows Server 2012 标准版 R2 中文版操作系统的云主机为例。

说明

远程桌面授权仅支持 120 天，过期后将因缺失远程桌面授权服务器许可证而导致多用户登录无法使用，需激活远程桌面授权。

配置多用户登录后，如果出现多用户登录 Windows 主机时无法打开浏览器的问题，请参考多用户登录 Windows 主机时无法打开浏览器

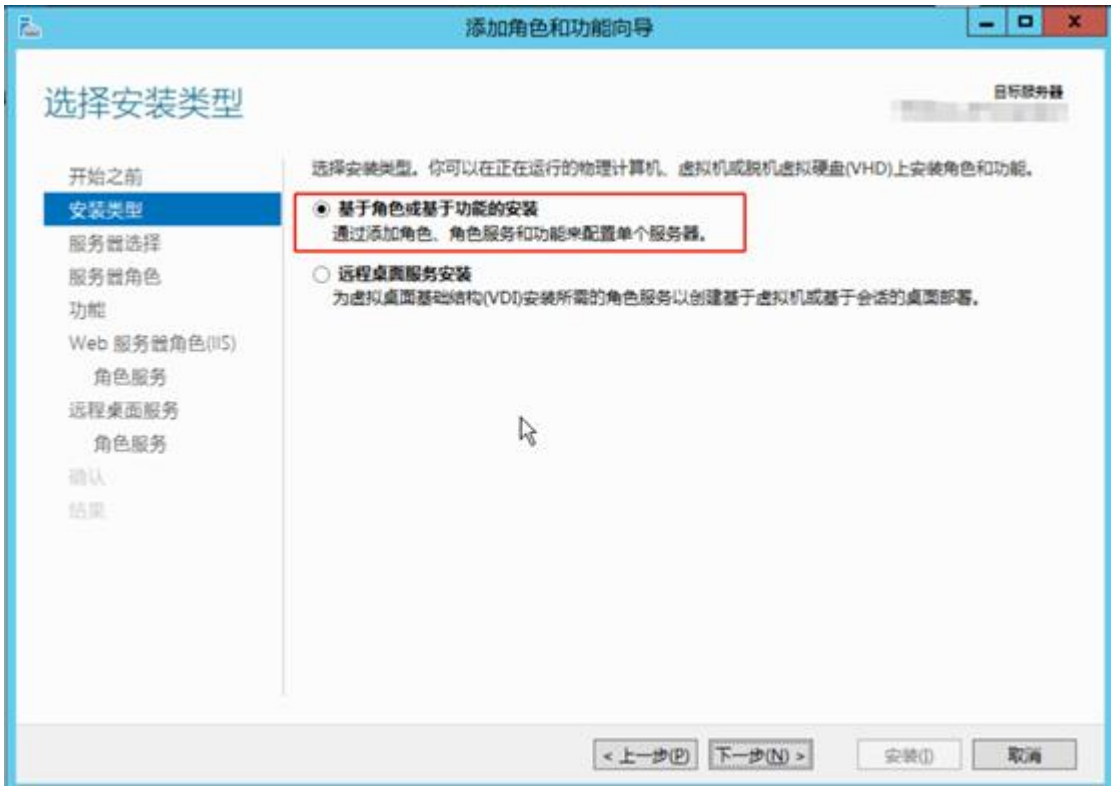
操作须知

- 请确保云主机带宽资源充足，避免由于多用户同时操作负载过高导致云主机卡顿或登录异常。
- 所在安全组入方向已开放云主机登录使用的端口，默认使用 3389 端口。
- 云主机已经绑定弹性公网 IP。
- 配置多用户登录后，不同的用户登录云服务器操作互相之间无影响。
- 配置多用户登录后，如果出现多用户登录 Windows 主机时无法打开浏览器的问题，请参考多用户登录 Windows 主机时无法打开浏览器。

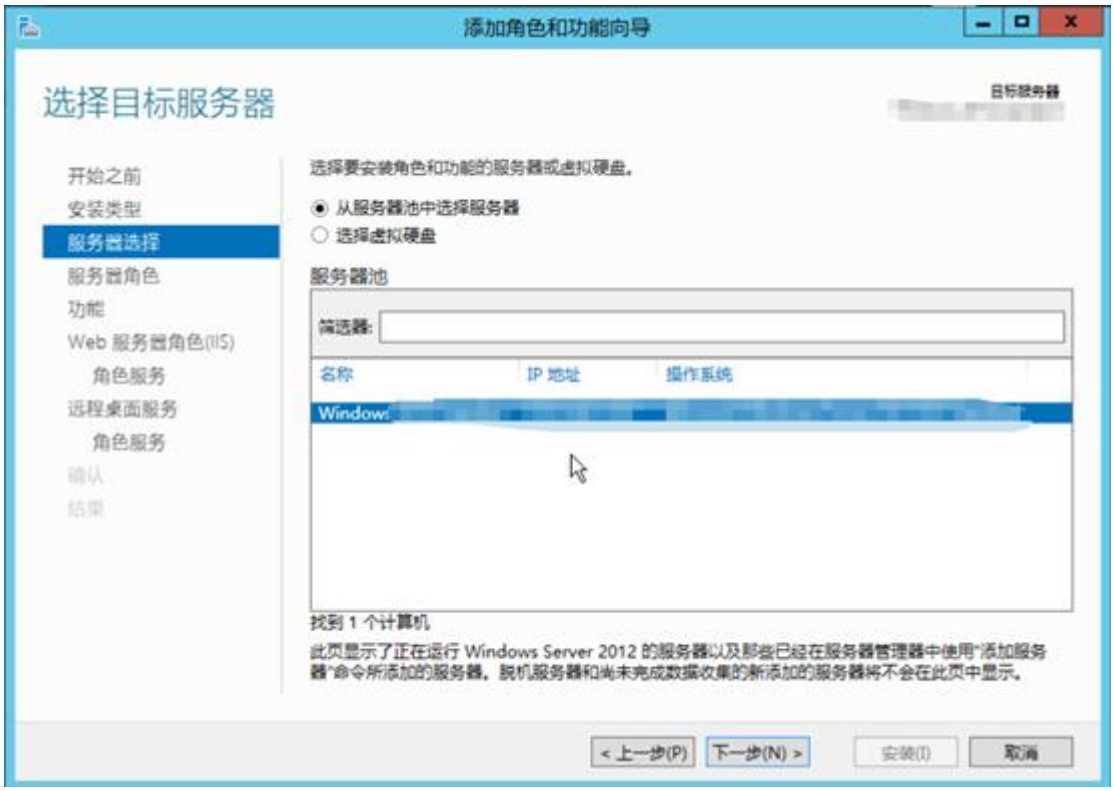
操作步骤

步骤一：安装桌面会话主机和远程桌面授权。

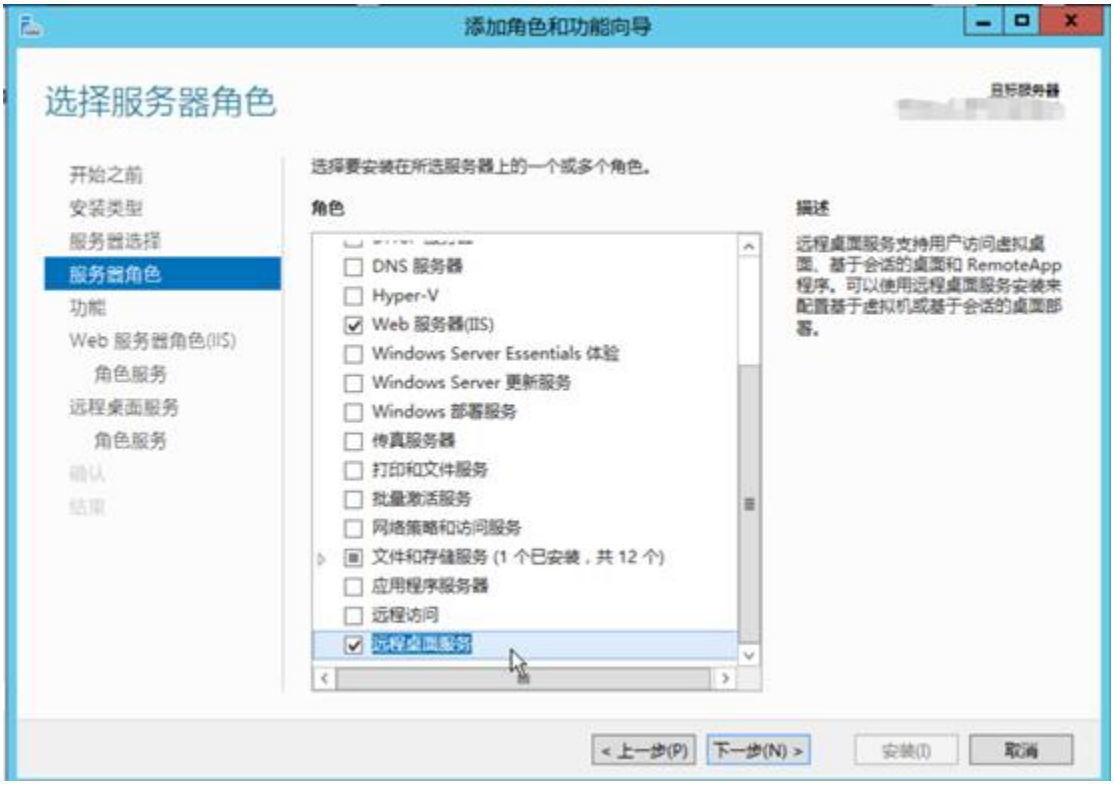
1. 登录弹性云主机。
2. 在操作系统界面， 打开“服务器管理器”，单击“添加角色和功能“。
3. 保持默认参数，单击“下一步”。
4. 左侧导航栏选择“安装类型”，选择“基于角色或基于功能的安装”，单击“下一步”。



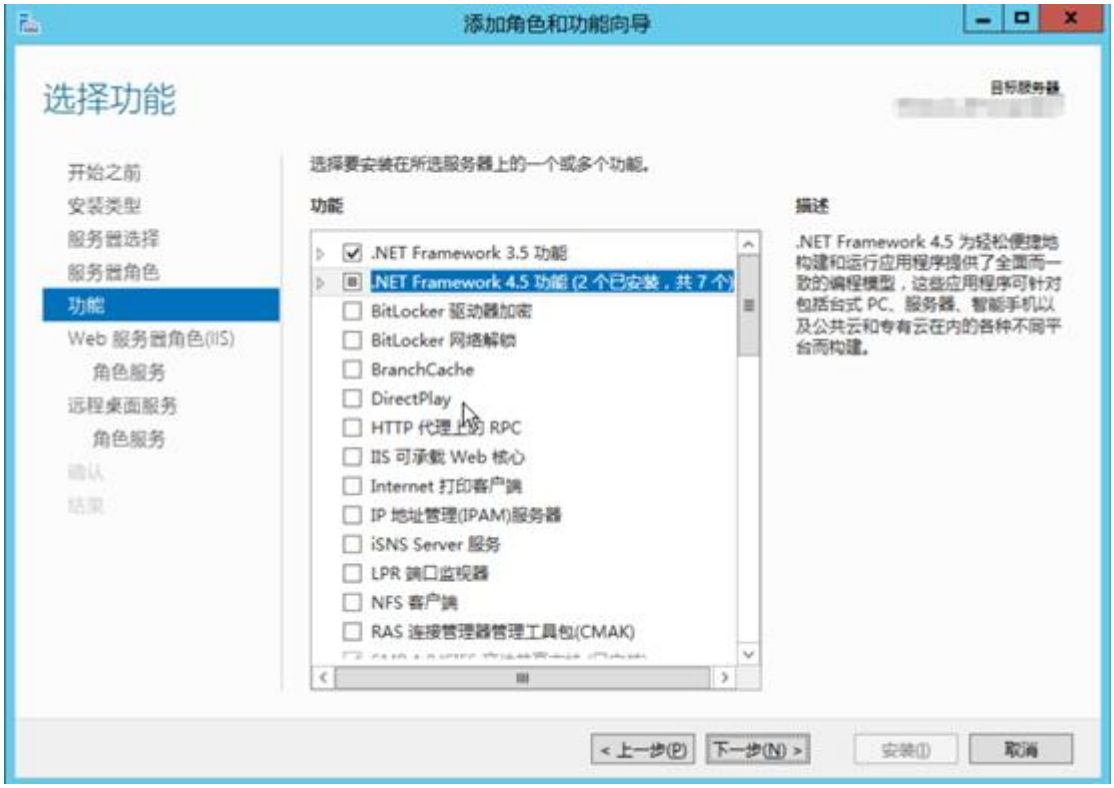
5. 左侧导航栏选择“服务器选择”，选择“从服务器池中选择服务器”，单击“下一步”。



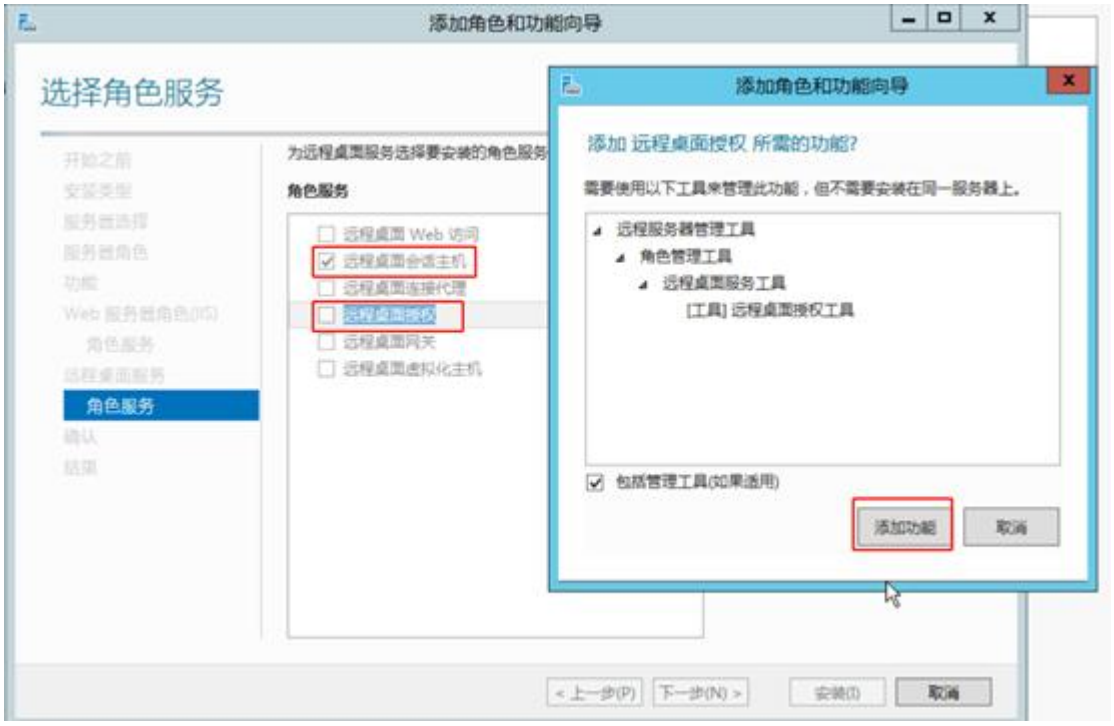
6. 左侧导航栏选择“服务器角色”，选择“远程桌面服务”，单击“下一步”。



7. 左侧导航栏选择“功能”，在此页面保持默认参数，单击两次“下一步”。



8. 左侧导航栏选择“角色服务”，在此界面依次选择“远程桌面会话主机”和“远程桌面授权”，在弹出的窗口单击“添加功能”，单击“下一步”。



9.确认在云主机上安装的角色，单击“安装(I)”。

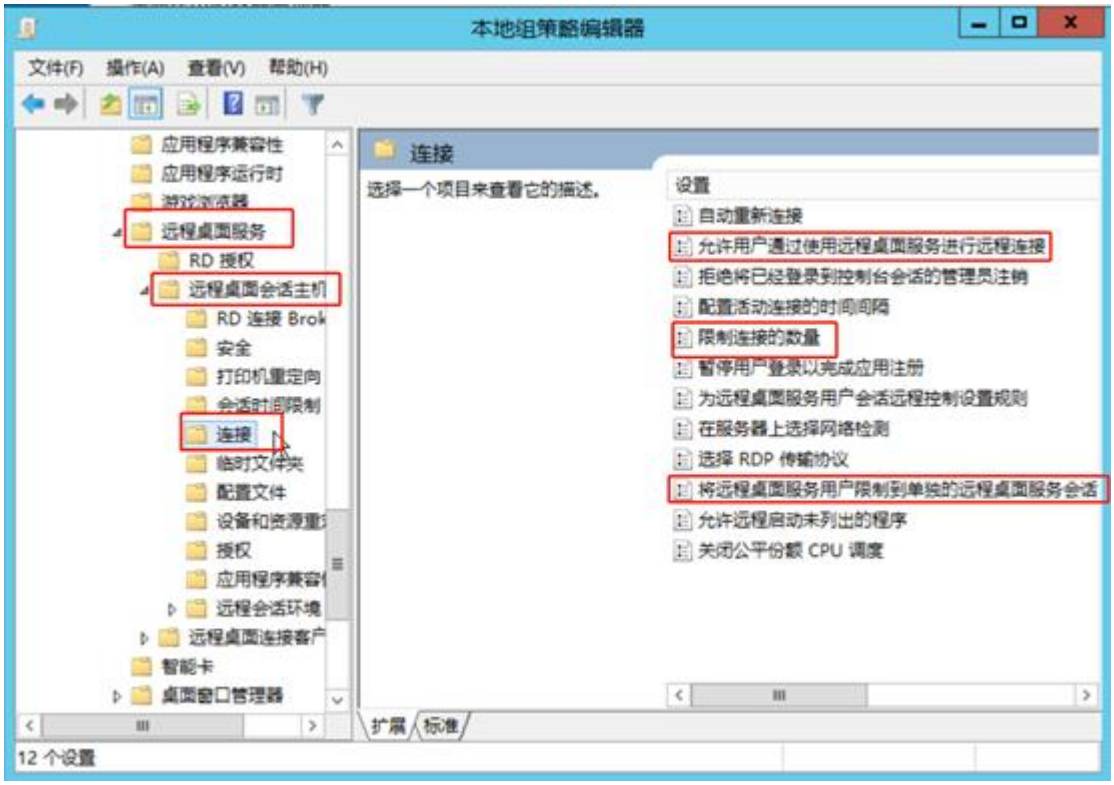


10.安装完成后，重启云主机。

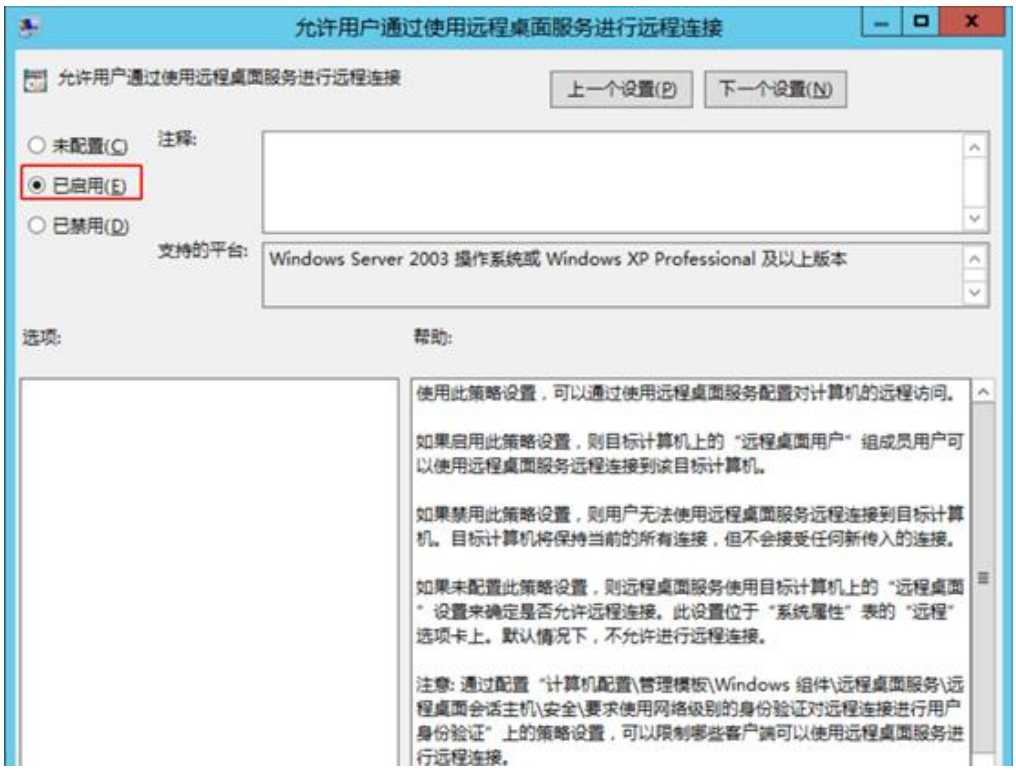
第二步：允许多用户远程连接云主机。

1.在运行里输入　gpedit.msc，打开计算机本地组策略

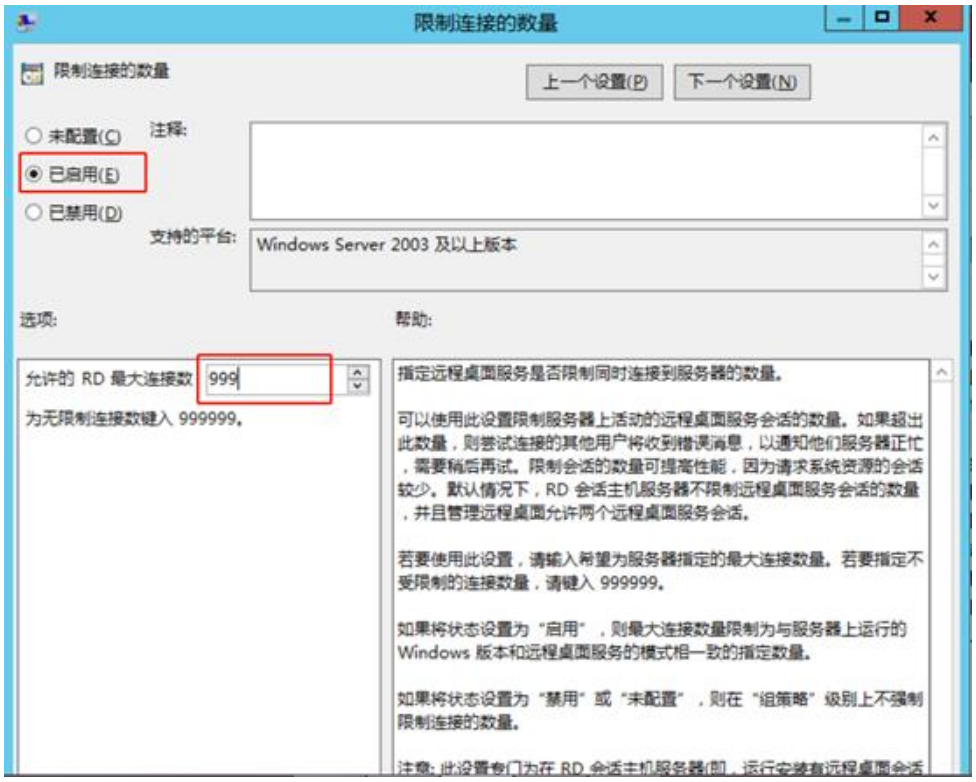
2. 选择“计算机配置>管理模板>windows 组件>远程桌面服务>远程桌面会话主机>连接”，并依次设置“允许用户通过使用远程桌面服务进行远程连接”、“限制连接数量”和“将远程桌面服务用户限制到单独的远程桌面”



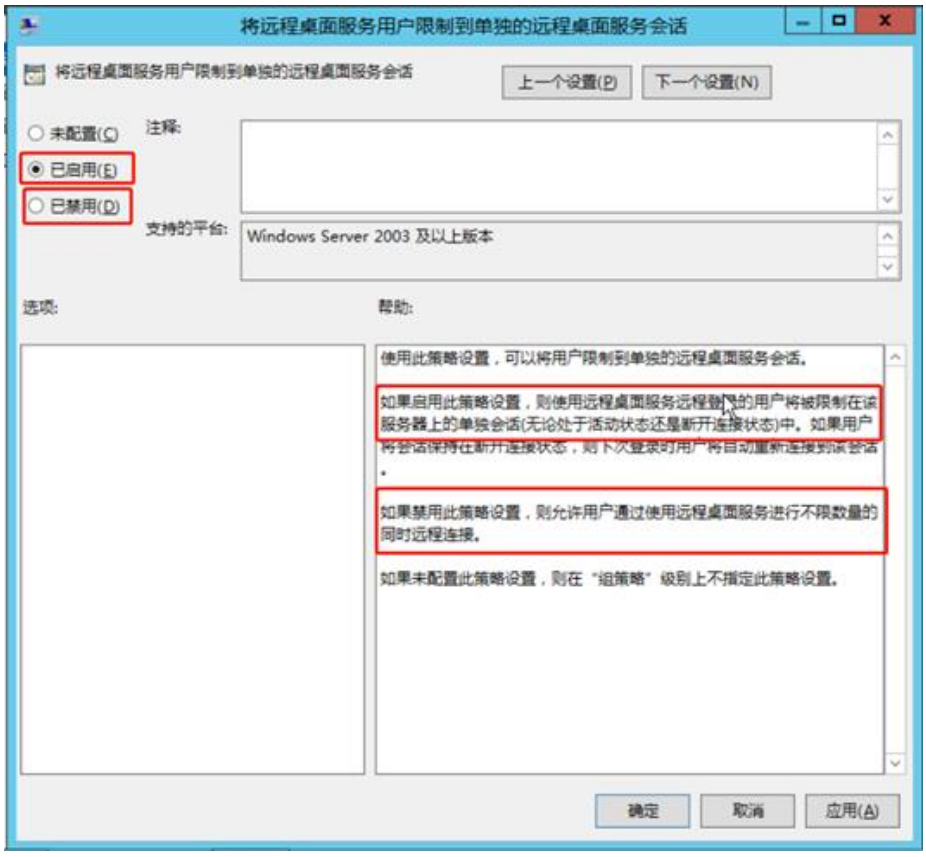
3. 右键单击“编辑”，设置“允许用户通过使用远程桌面服务进行远程连接”设置为“已启用”。



4. 右键单击“编辑”，设置“限制连接数量”选择为已启用，可根据具体数量设置，这里示例将允许的 RD 最大连接数设置为 999。



5. 右键单击“编辑”，“将远程桌面服务用户限制到单独的远程桌面”选择已启用，或者已禁用，请注意此处已启用和已禁用的区别。本示例勾选“已启用”。



说明

1 已启用：多个用户同时登录的多用户登录，不能单个用户多登录

例如：配置为已启用，用户 A、用户 B、用户 C 可以分别使用账号 A、账号 B、账号 C 同时登录云主机，但是不支持用户 A、用户 B、用户 C 使用同一个账号同时登录云主机

2 已禁用：单个用户同时多个登录的多用户登录

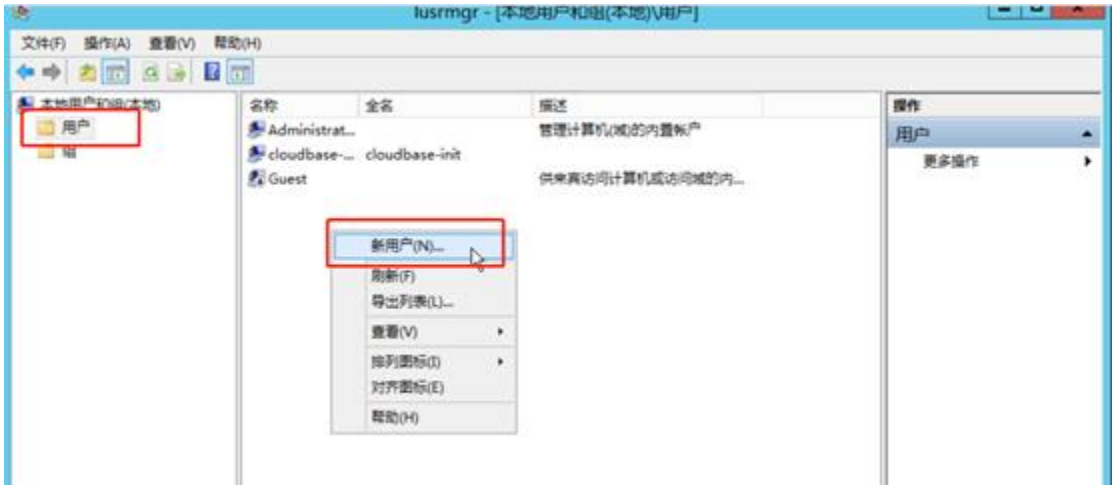
例如：配置为已禁用，用户 A、用户 B、用户 C 可以使用同一个账号同时登录云主机

1.运行 cmd，输入 gpupdate /force，强制执行本地组策略，重启服务器，整个配置过程完成。

第三步：配置新用户并加入远程桌面用户组。

在运行中输入 lusrmgr.msc，打开本地用户和组，进行新用户创建。

2.单击“用户”，在空白处右键选择新用户。



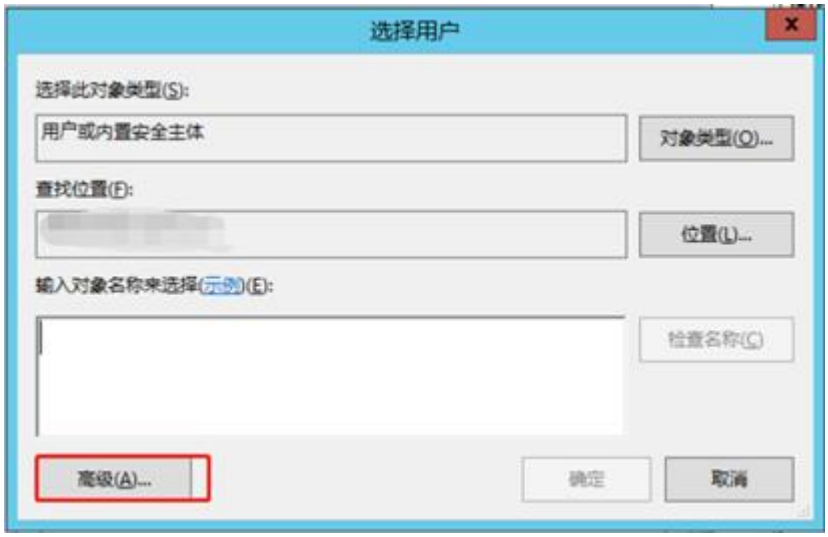
3. 填写新用户信息，单击“创建”。



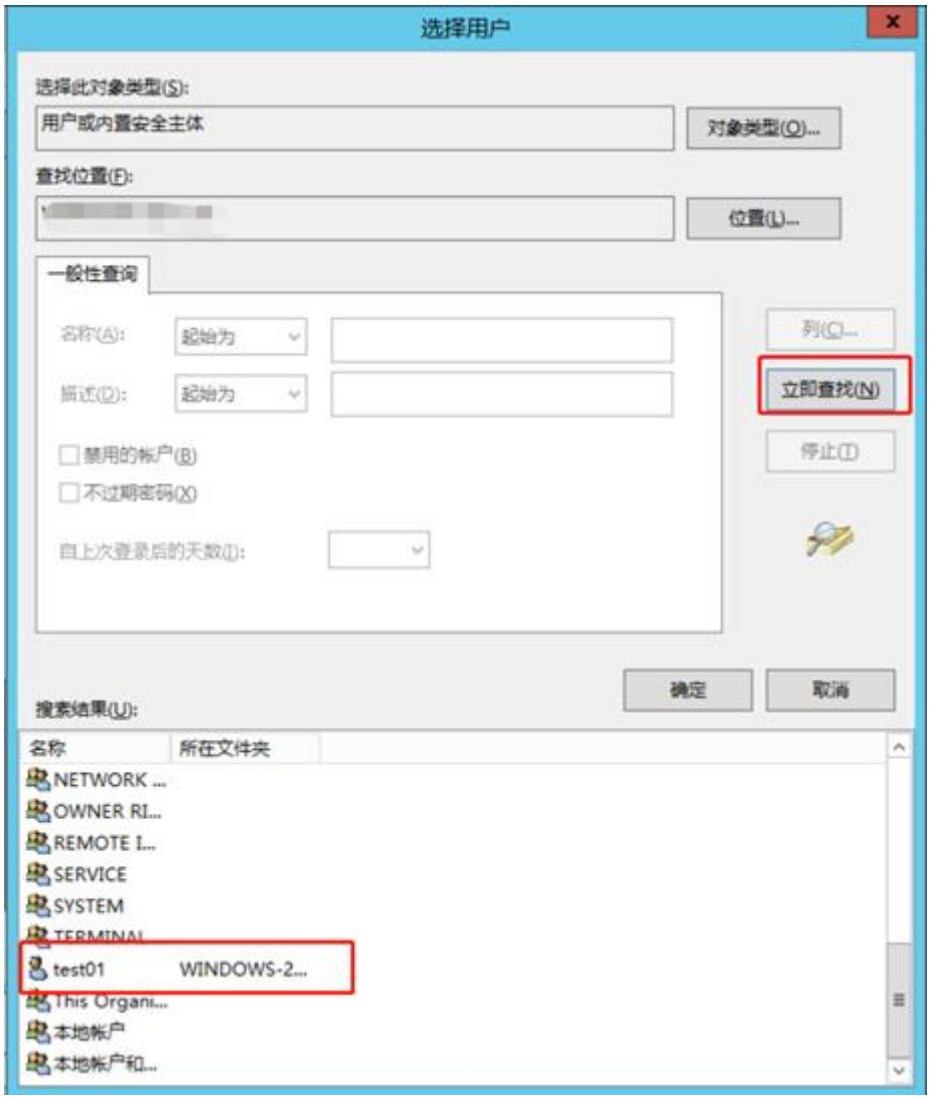
4. 单击“组”，双击打开 Remote Desktop Users 组，单击“添加”。



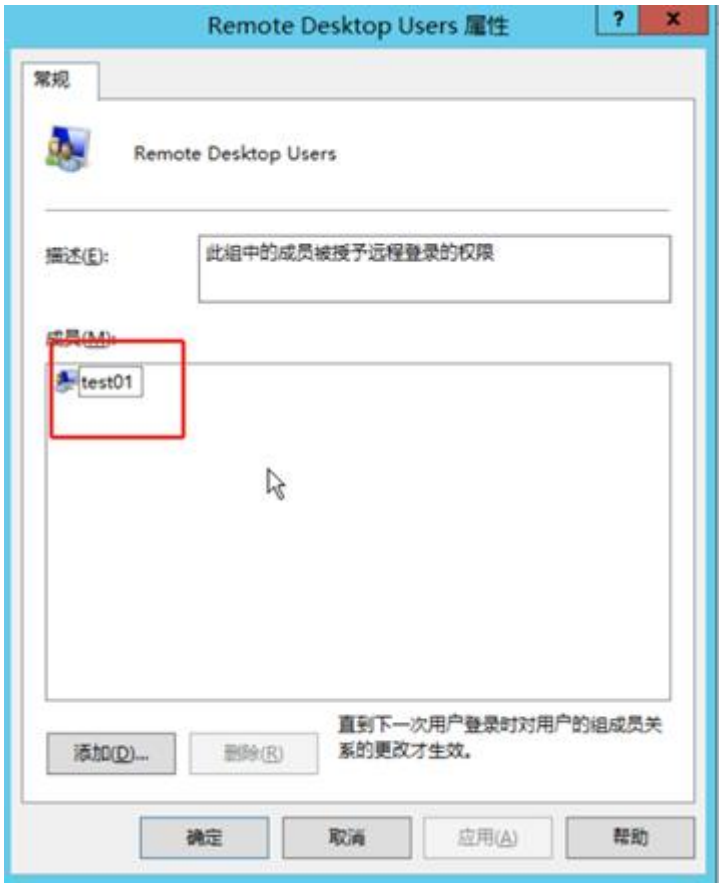
5. 进入选择用户界面，单击“高级”。



6. 在新的选择用户界面，单击“立即查找”，在下方搜索结果中选中需要远程登录的用户，并单击“确定”，完成添加，即可远程登录。



7.单击“确定”，添加用户到 Remote Dsektop Users 组。



9.10.4 多用户登录 Windows 主机时无法打开浏览器

问题描述

存在多用户登录 Windows 弹性云主机时无法打开浏览器的情况。例如：某一用户已经打开了浏览器，但第二个用户打不开浏览器。

处理方法

以 IE 浏览器为例。

1. 在桌面选择浏览器图标，单击右键选择“创建快捷方式”。
2. 选择新创建的快捷方式，单击右键选择“属性”。
3. 选择“快捷方式”页签，找到“目标(I)”，在末尾添加如下内容。

说明

- 在末尾先添加空格再添加以上内容，即.exe"和--user 中间有空格。
- c:\MyInternetExplorerData 表示的是 Internet Explorer 的数据文件存放位置，可以设置为任何有效的文件夹路径，如果此文件夹不存在，Internet Explorer 浏览器会自动创建。

下图是浏览器属性。



4. 经过保存修改后就可以实现多个用户同时打开浏览器。

9.11 密码与密钥对 安装 IIS 服务

9.11.1Linux 云主机怎样切换密钥登录为密码登录？

操作场景

本节操作介绍 Linux 云服务器切换密钥登录为密码登录的操作步骤。

操作步骤

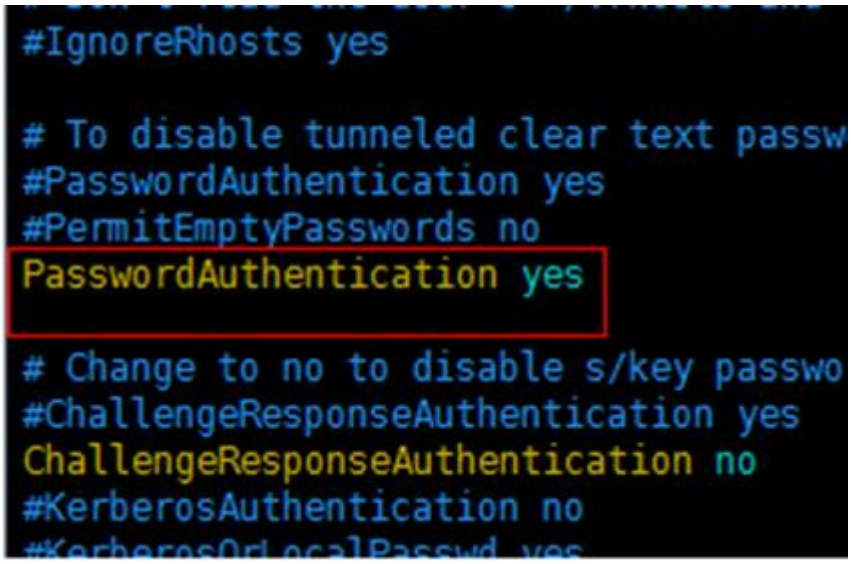
1. 使用 root 身份编辑 Linux 云服务器的 ssh 配置文件。

su root

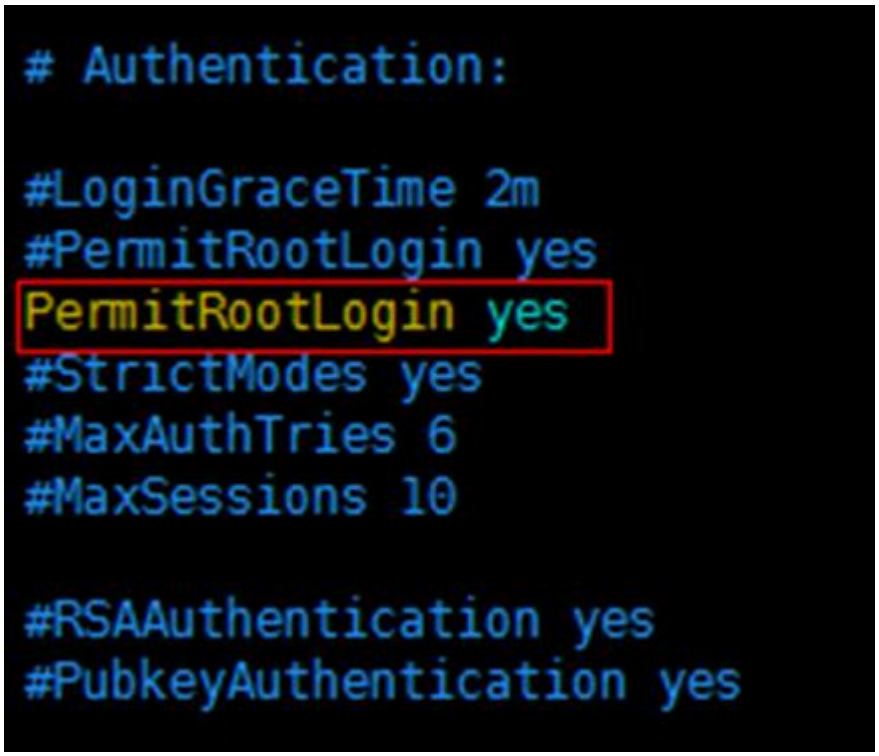
vim /etc/ssh/sshd_config

修改如下配置项：

- 把 PasswordAuthentication no 改为 PasswordAuthentication yes。



- 把 PermitRootLogin no 改为 PermitRootLogin yes。



2. 重启 sshd 服务使配置生效，并查看服务状态。

```
systemctl restart sshd
```

```
systemctl status sshd
```

3. 设置 root 密码，然后就可以通过密码远程登录云主机。

```
passwd root
```

9.11.2 Linux 云主机如何进入单用户模式重置 root 密码

适用场景

本节操作介绍 Linux 操作系统云主机在单用户模式下重置密码的操作步骤。

说明：进入单用户模式下重置 root 密码前请做好数据备份。

请根据操作系统类型选择操作步骤：

- CentOS 7/EulerOS 系列
- CentOS 6/RedHat 6 系列
- Debian/Ubuntu 系列
- SUSE 11 系列

CentOS 7 系列

1. 登录虚拟机。

登录天翼云官网-弹性云主机-选择机器-点击远程登录。



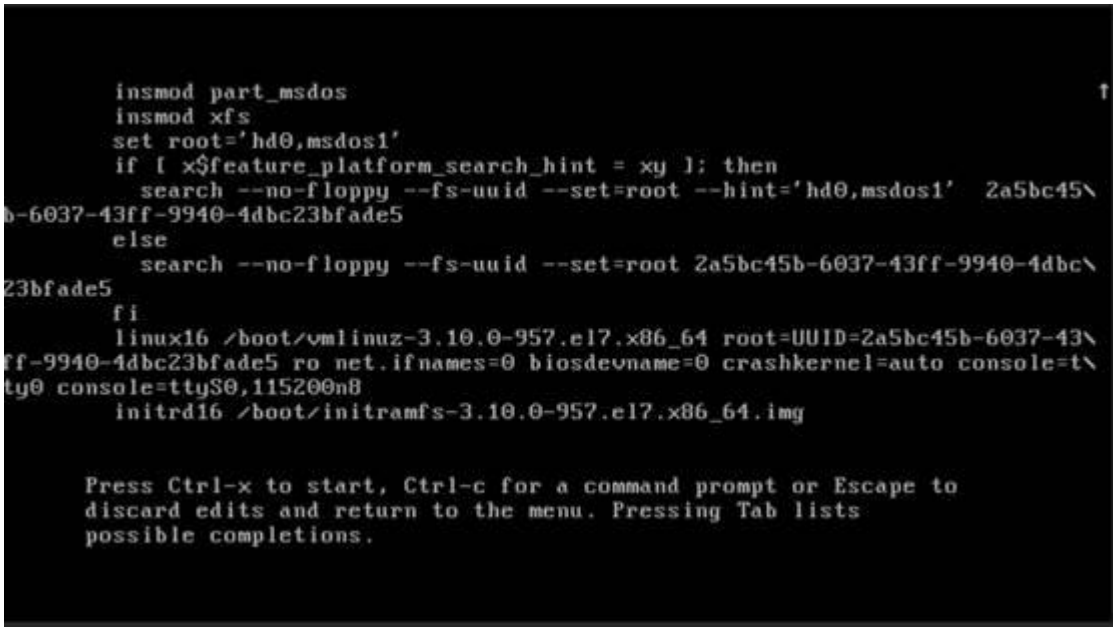
2. 点击 Send ctrlAlt 重启云主机。



3. 开机后，在 GRUB 启动菜单中选择要引导的 CentOS 系统。



4. 按 e 键进入编辑模式。



5. 在编辑模式中，找到以 “linux16” 开头的那一行。在该行的末尾，添加空格和 “init=/bin/bash”。然后按下 “Ctrl+X” 或 “F10”启动进入单用户模式。

6. 此时，系统将进入单用户模式，并且您将获得一个命令行提示符。

7. 输入以下命令来挂载文件系统为可读写模式。

```
mount -o remount,rw /
```

8. 输入以下命令来修改密码（将 “your_username” 替换为您要更改密码的用户名）。

```
passwd your_username
```

9. 系统将要求您输入新密码并进行确认。输入并确认密码后，将会显示密码已成功更改的消息。

10. 输入以下命令以重新挂载文件系统为只读模式。

```
mount -o remount,ro /
```

11. 输入以下命令以重新启动系统。

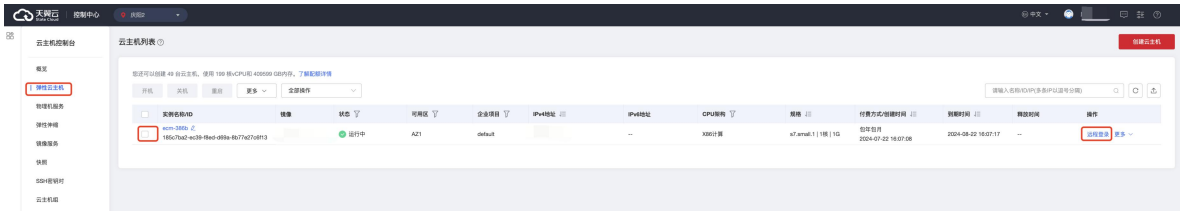
```
Reboot
```

12. 系统将重新启动，并且您可以使用新设置的密码登录。

CentOS 6/RedHat 6 系列

1. 登录虚拟机。

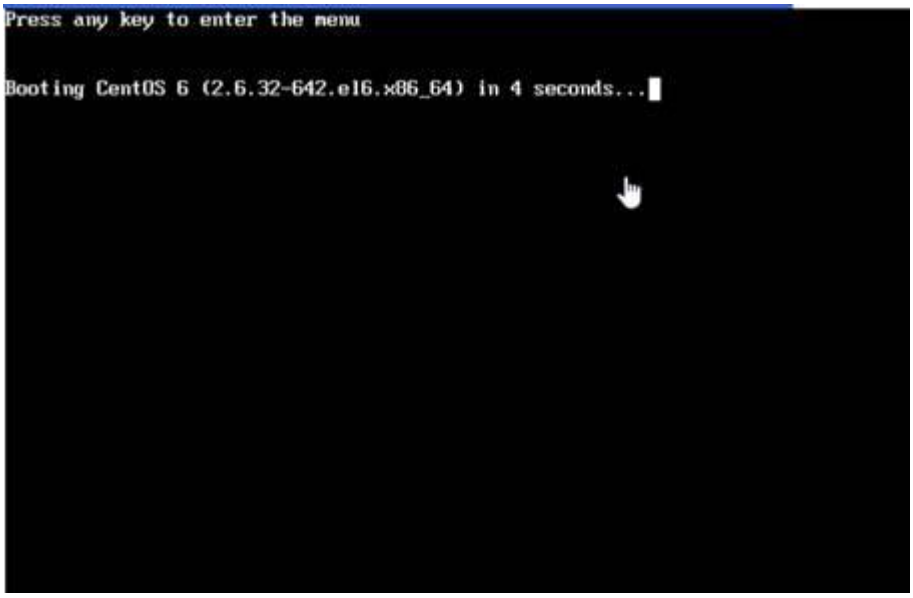
登录天翼云官网-弹性云主机-选择机器-点击远程登录。



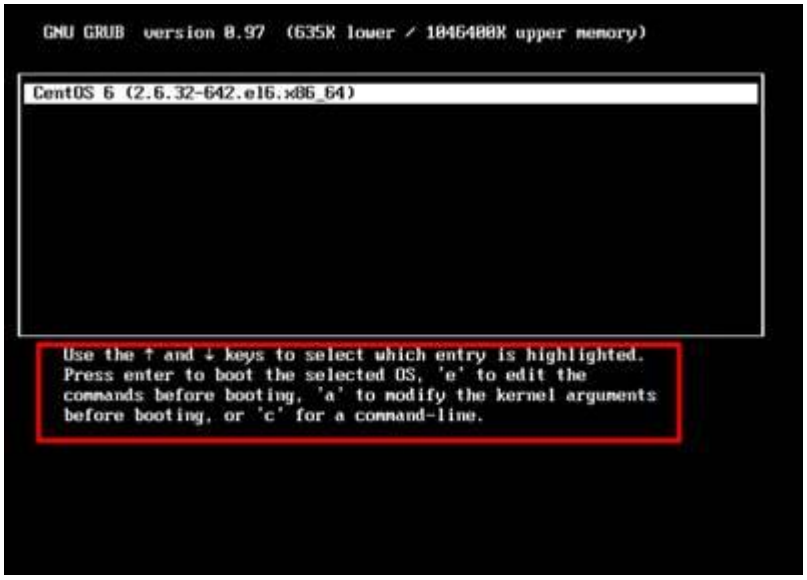
2. 点击 Send ctrlAlt 重启云主机。



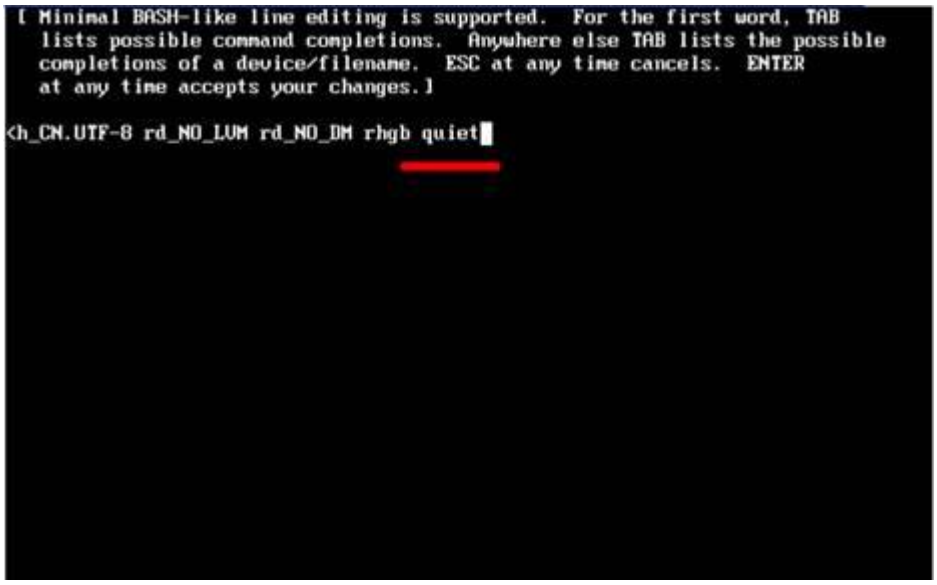
3. 选择进入菜单 menu 界面，在开启系统出现如下界面时，按 Esc 键（只需按一下）。



4. 接下来进入到如下界面。

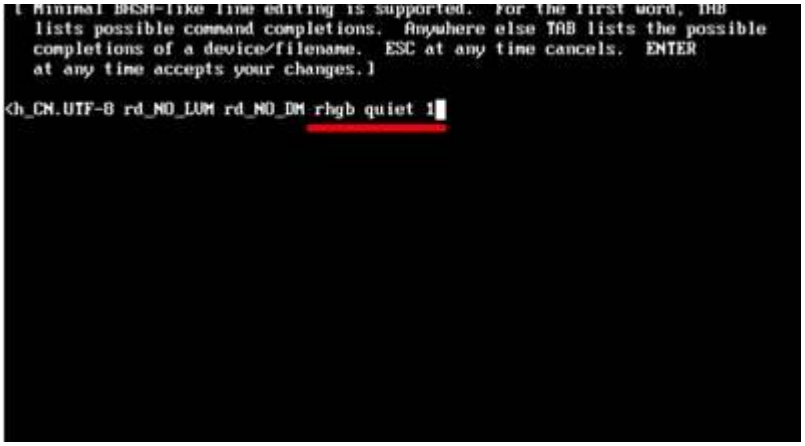


5. 上图中红色矩形类的内容，按“a”键可以修改内核参数，所以按 a 键，然后进入一下界面。

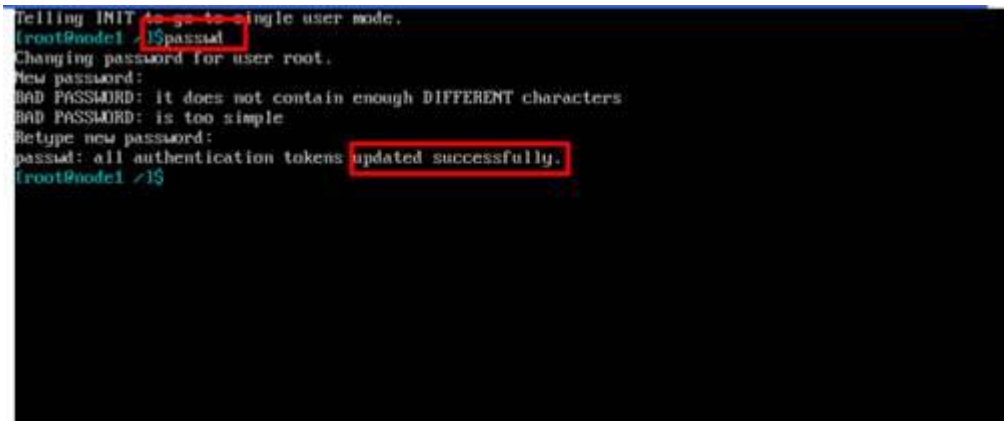


6. 添加“1”，注意“1”前面用空格分开，然后按回车键。

1.



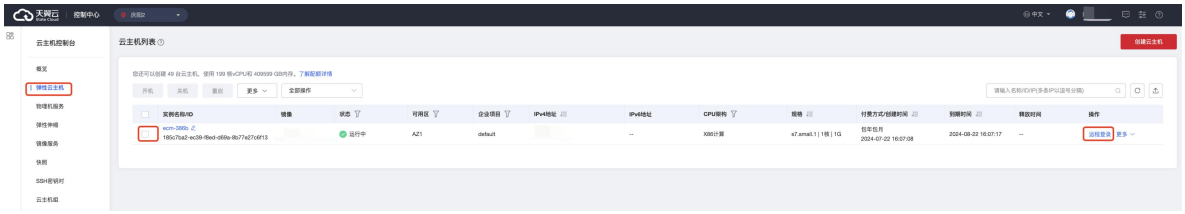
7. 接着进入如下界面，输入 passwd 命令，修改密码，修改密码成功之后，重启之后需用新密码登录。



Debian/Ubuntu 系列

1. 登录虚拟机。

登录天翼云官网-弹性云主机-选择机器-点击远程登录。



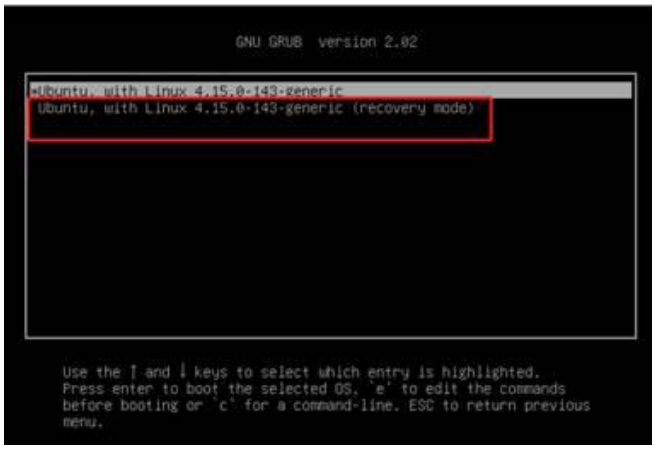
2. 点击 Send ctrlAlt 重启云主机。



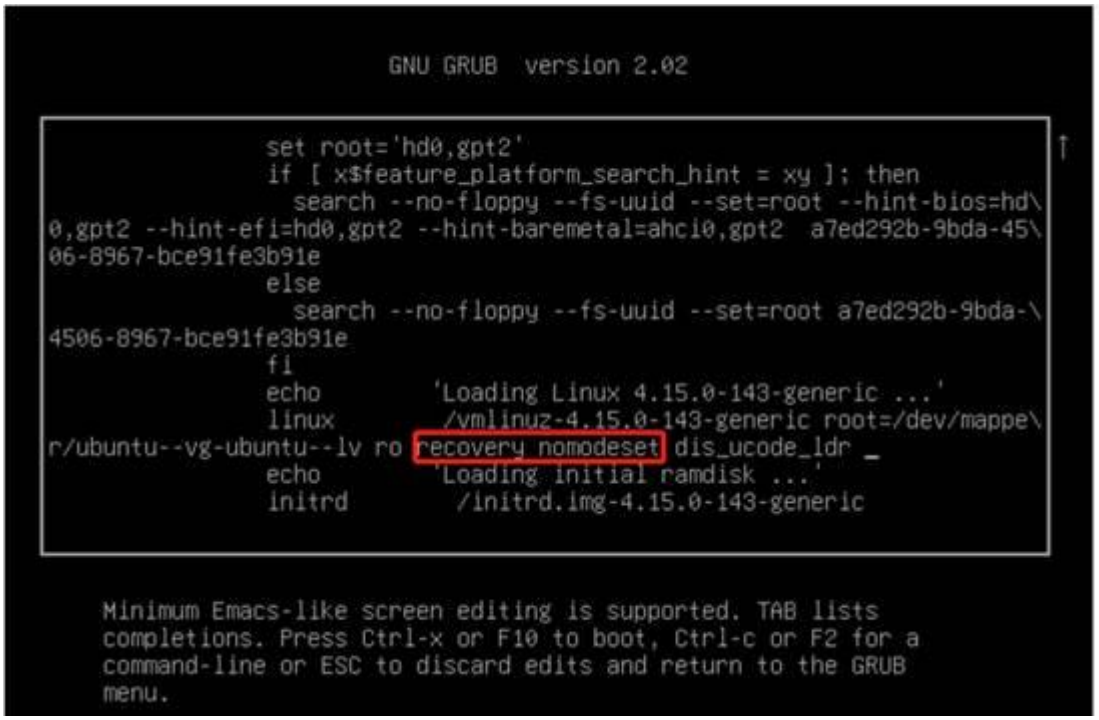
3. 启动 ubuntu18.04 系统的时候连续点按 ESC 键进入选项，选择 Advanced options for Ubuntu 并回车。



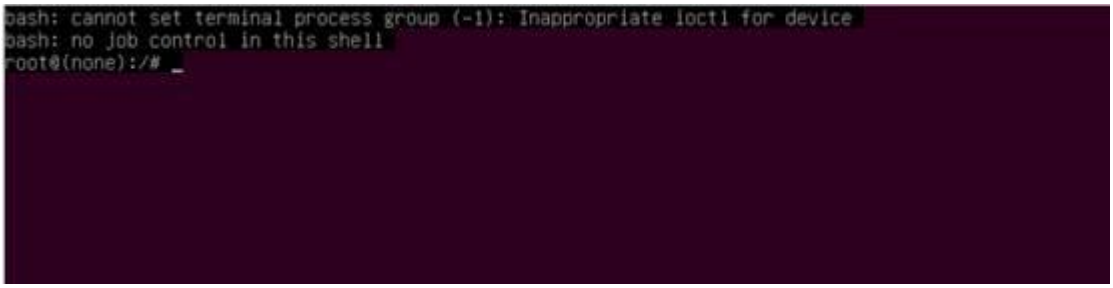
4. 选择 ubuntu, with 4.15.0-143-generic 带有（recovery mode）然后注意不要点回车，按 E 键。



5. 点 E 进入编辑页面，移动光标到 linux 一行，把 recovery nomodeset 删掉，并最后添加 quiet splash rw init=/bin/bash。



6. 修改完成以后，按 ctrl+x 或 F10 启动系统就可以进入 single 模式而不需要密码了。



7. 修改密码。

```
passwd root
```

8. 重启主机，用新密码登录即可。

SUSE 11 系列

1. 在启动主机的时候输入：

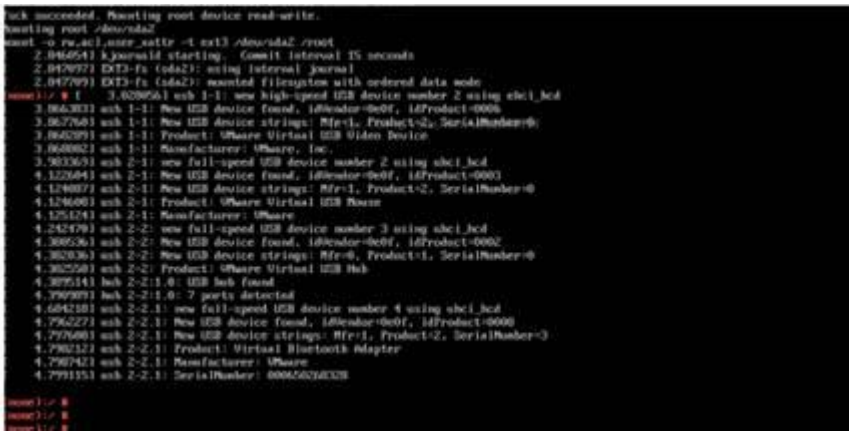
```
vga=0x317init=/bin/bash
```

并按回车。





2. 进入单用户模式。



3. 将根(/)设置为读写，便于修改密码。

```
mount -n / -o remount,ro
```

4.修改用户密码。

```
passwd    用户名字
```

5.将根(/)设置为只读。

```
mount -n / -o remount,ro
```

6. 重启主机。

9.11.3 密钥对方式购买的弹性云主机，使用私钥文件获取登录密码失败

可能原因

1.出现获取密码失败一般原因是弹性云主机 cloudinit 失败，没有通过 cloudinit 注入密码。

2.Cloudinit 注入密码失败原因有：

- 网络原因导致弹性云主机无法连接到 cloudinit 服务器；
- 镜像上 cloudinit 没有获取密码相关配置；
- 弹性云主机上其他问题导致弹性云主机 cloudinit 无法获取。

处理办法

1.检查弹性云主机的网络配置，是否导致弹性云主机 cloudinit 失败，操作如下： 弹性云主机所在安全组 80 端口的“出方向”和“入方向”是否放通。

2.检查镜像上 cloudinit 是否配置了获取密码，当前 HEC 环境中的所有公共镜像均已正确配 置 cloudinit。

- 如果客户使用公共镜像则无需检查此项。
- 如果客户使用私有镜像创建弹性云主机，则需要确认镜像中已正确配置了 cloudinit。

3.其他方面：大规格弹性云主机的虚拟内存导致客户弹性云主机 cloudinit 执行失败。

4.通过查看日志检查原因，cloudinit 日志查看步骤如下：

- 将无法获取密钥的弹性云主机关机后，将系统盘卸载。
- 使用公共镜像创建一个临时 windows 弹性云主机，将上一步骤中卸载的卷挂载在此弹性云主机上。
- 登录临时创建的弹性云主机主机，打开主机管理器，选择“文件和存储服务 > 卷 > 磁盘”，单击鼠标右键选择“脱机”状态的磁盘，单击“联机”。
- 打开新联机的磁盘，查看“/Programe Files/Cloudbase Solution/CloudbaseInit/log”路径下的“cloudbase-init”文件，通过日志查看原因。

9.11.4 如何更换密钥对

问题描述

更换云主机密钥对后，无法使用新的密钥对登录云服务器。

处理方法

1.先登录 Linux 云主机上制作密钥对，登录到云主机，执行以下命令，生成密钥对。

```
[root@host ~]$ ssh-keygen  <== 建立密钥对
Generating public/private rsa key pair.
Enter ile in which to save the key (/root/.ssh/id_rsa): <== 按 Enter
Created directory '/root/.ssh' .
Enter passphrase (empty for no passphrase): <== 输入密钥锁码，或直接按 Enter 留空
Enter same passphrase again: <== 再输入一遍密钥锁码
Your identiication has been saved in /root/.ssh/id_rsa. <== 私钥
Your public key has been saved in /root/.ssh/id_rsa.pub. <== 公钥
```

```
The key ingerprint is:0f:d3:e7:1a:1c:bd:5c:03:f1:19:f1:22:df:9b:cc:08
root@host
```

2.在使用私钥时输入密钥锁码，可以保护私钥不被盗用，在/root/.ssh/下面生成了两个密钥文件：id_rsa 为私钥，id_rsa.pub
为公钥。

3.添加公钥。

```
[root@host ~]$ cd .ssh

[root@host .ssh]$ cat id_rsa.pub >> authorized_keys
```

4.修改.ssh、authorized_keys 文件权限.

```
[root@host .ssh]$ chmod 600 authorized_keys

[root@host .ssh]$ chmod 700 ~/.ssh
```

5.(可选)打开密钥登录功能。

```
RSAAuthentication yes

PubkeyAuthentication yes
```

6.打开允许 root 用户登录配置。

```
PermitRootLogin yes
```

7.完成上述配置后，重启 sshd 服务，以密钥方式登录成功后，再禁用密码登录。

```
PasswordAuthentication no
```

8.重启 ssh 服务。

```
systemctl restart sshd
```

9.11.5 使用密钥文件无法正常登录 Linux 弹性云主机

问题描述

可能原因

Linux 弹性云主机的镜像为用户自己制作的私有镜像，且在创建该私有镜像时用户未安装 Cloud-init 工具。

Linux 弹性云主机的镜像安装了 Cloud-init 工具，但是在创建弹性云主机时，获取密钥失败。

处理方法

针对原因 1

故障排查 115 创建私有镜像时不安装 Cloud-init 工具，将无法对弹性云主机进行自定义配置，此时，用户只能使用镜像原有密码或密钥登录弹性云主机。其中，镜像原有密码或密钥指创建私有镜像时，用户自己设置的操作系统密码或密钥。如果忘记镜像原有密码，或镜像原有密钥丢失，可以通过弹性云主机页面提供的“重置密码”功能， 自助完成弹性云主机的密码重置。

针对原因 2

- 勾选待获取密码的弹性云主机，单击“重启”，重新启动弹性云主机。
- 重启成功后，再次使用密钥文件登录弹性云主机，确认是否能够正常登录。
  - 是，结束。 - 否，请联系客服。

9.11.6 Linux 云主机重置密码

方法一

操作场景

本节操作介绍 Linux 操作系统云主机在控制台修改密码。

操作步骤

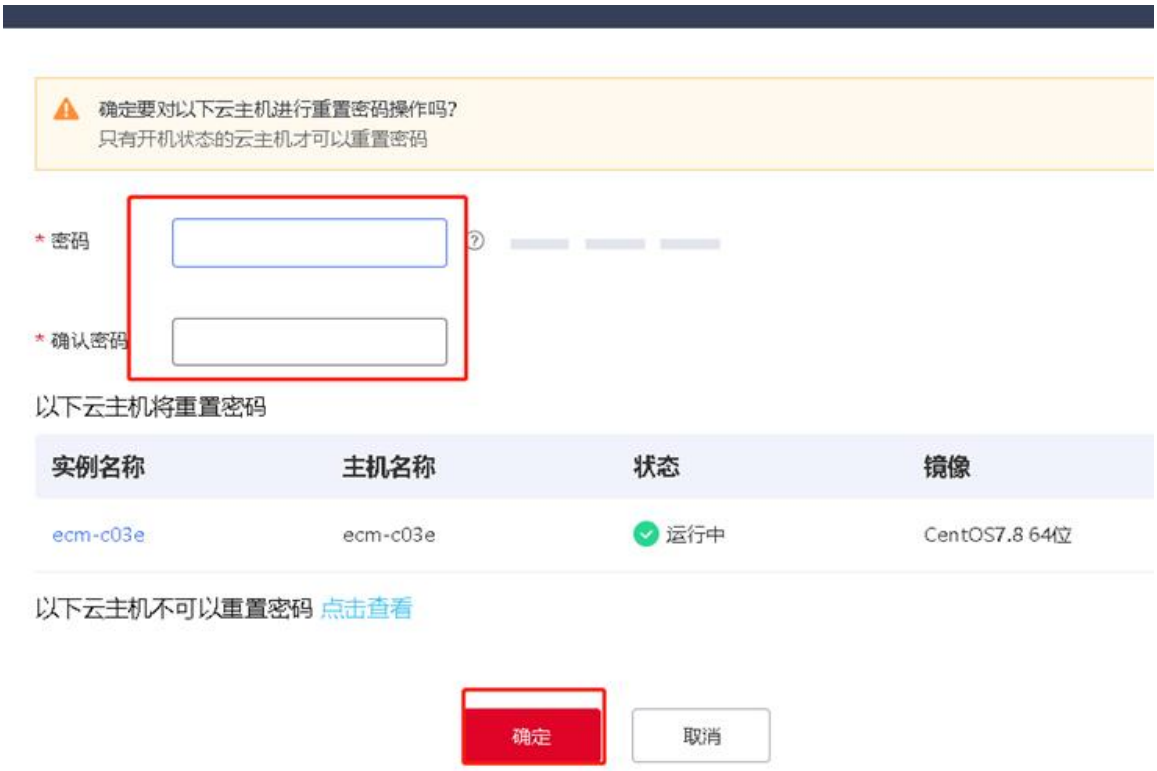
步骤 1：登录云主机控制台



步骤 2：找到修改密码的入口



步骤 3： 重置密码（注意密码为 8～30 个字符，必须同时包 含三项（大写字母、小写字母、数字、()`~!@#$%^&*_-+=|{}[]:;'<>,.?/中的特殊符号，且不能以斜线号（/）开头）



9.11.7 无法导入密钥对

修改浏览器默认属性后重试，步骤如下：

1. 在浏览器主界面，点击设置。
2. 选择“Internet 选项”。
3. 单击选择“安全”页签。



4. 单击“Internet”。

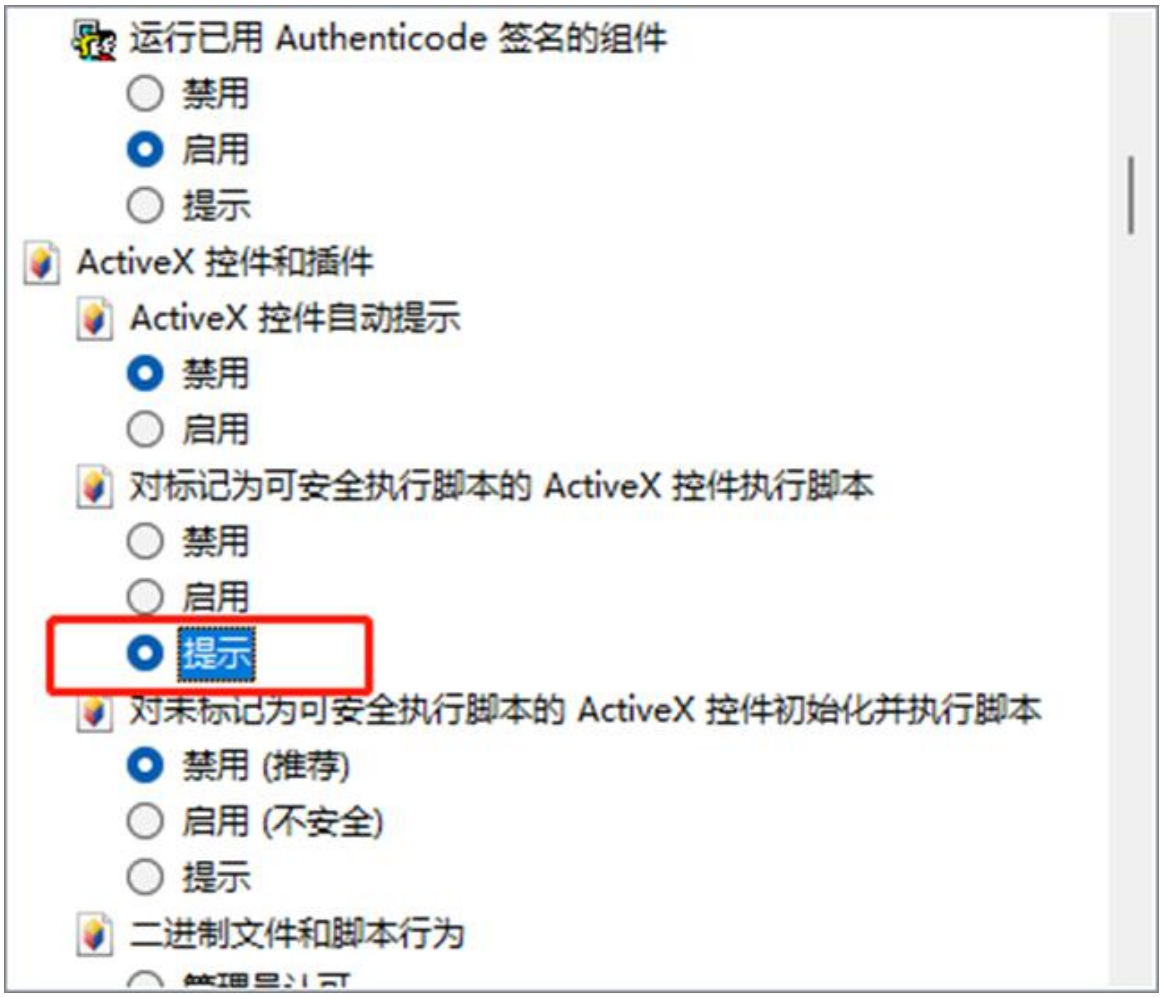
5. 单击“默认级别”按钮，把设置还原为默认级别。



6. 把安全级别调到“中”级别，单击“应用”按钮。

7. 选择“自定义级别”。

8. 将“对未标记为可安全执行脚本的 ActiveX 控件初始化并执行脚本”设置为“提示”。



9. 单击“确定”。

9.12 安装 IIS 服务

9.12.1 怎样做网页定向？

操作场景

本指导适用于用户做网页 301 重定向时参考使用。

前提条件

IIS 服务器中已安装 HTTP 重定向功能。

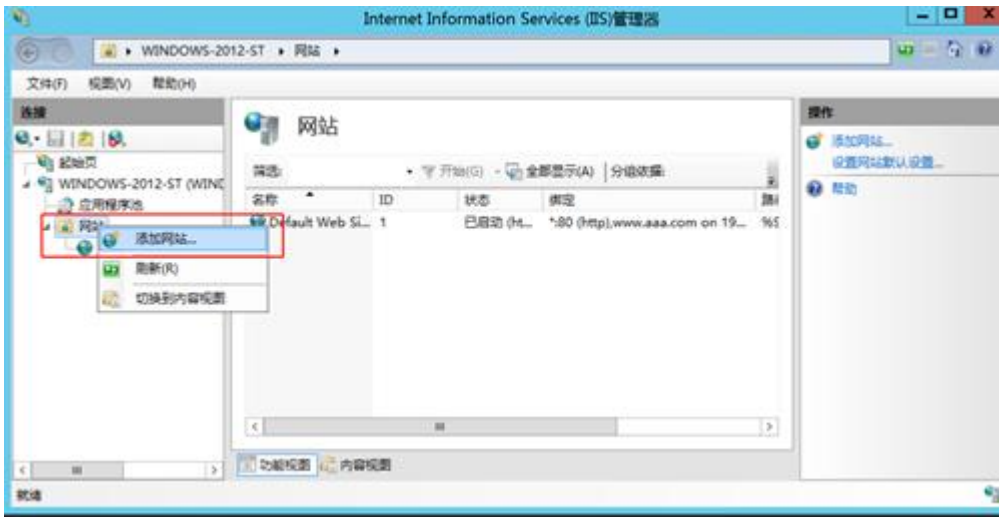
操作步骤

1. 在 IIS 里把网站正常发布，例如域名为 www. aaa. com。

2. 在 C:\web 下建立一个专门供给 301 重定向使用的文件夹，然后在里面建立一个需要重定向域名的文件夹，以 www1.aaa.com 为例。



3. 在 IIS 里建一个网站，例如域名为 aaa.com，指向新建的空文件夹。





4. 在网站的主页，选择“HTTP 重定向”，具体参数设置如下：

- 勾选“将请求重定向到此目标”。
- 重定向地址：www.aaa.com$S$Q

说明

在网址后添加“SQ”的作用是：支持带“？”的网址可以正常跳转。如果未添加“SQ”，带“？”的网址跳转时会出现异常。

- 勾选“将所有请求重定向到确切的目标(而不是相对于目标)”。
- 状态代码：永久(301)。



5.单击“应用”，完成网页 301 重定向。

相关操作

除了通过 IIS 实现重定向外，您还可以通过代码实现 301 重定向，以下为在 PHP 和 Apache 下实现 301 重定向的代码示例：

- PHP 下的 301 重定向。

```
<?phpHeader("HTTP/1.1 301 Moved Permanently");Header("Location:
http://www.***.cn"); //跳转到带 www 的网址?>
```

- Apache 服务器实现 301 重定向。
- WWW 域名的重定向。参考代码如下：

```
deny from all

RewriteEngine on

RewriteCond %{HTTP_HOST}^(****\.com)(:80)?[NC]

RewriteRule ^(.*) http://www.****.com/$1[R=301,L]order deny,allow
```

9.12.2 IIS 服务修改已绑定的网站域名

问题描述

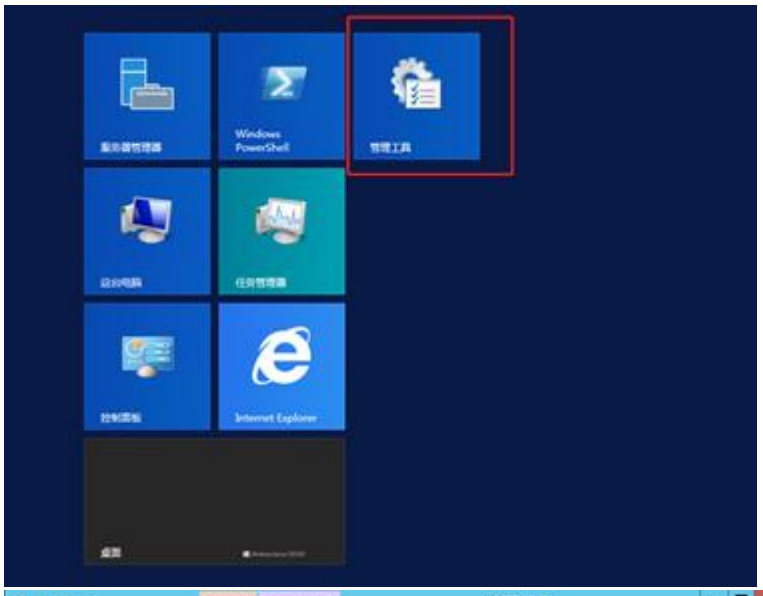
访问 IIS 搭建的网站的时候报错 404。

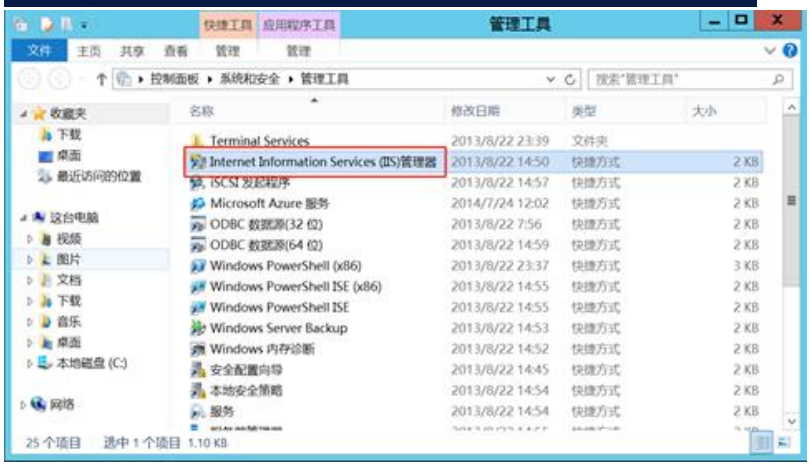
原因分析

IIS 上绑定的域名没有指定 IP 地址。

操作步骤

1.登录弹性云主机，选择“开始 >管理工具 > 信息服务（IIS）管理器”。

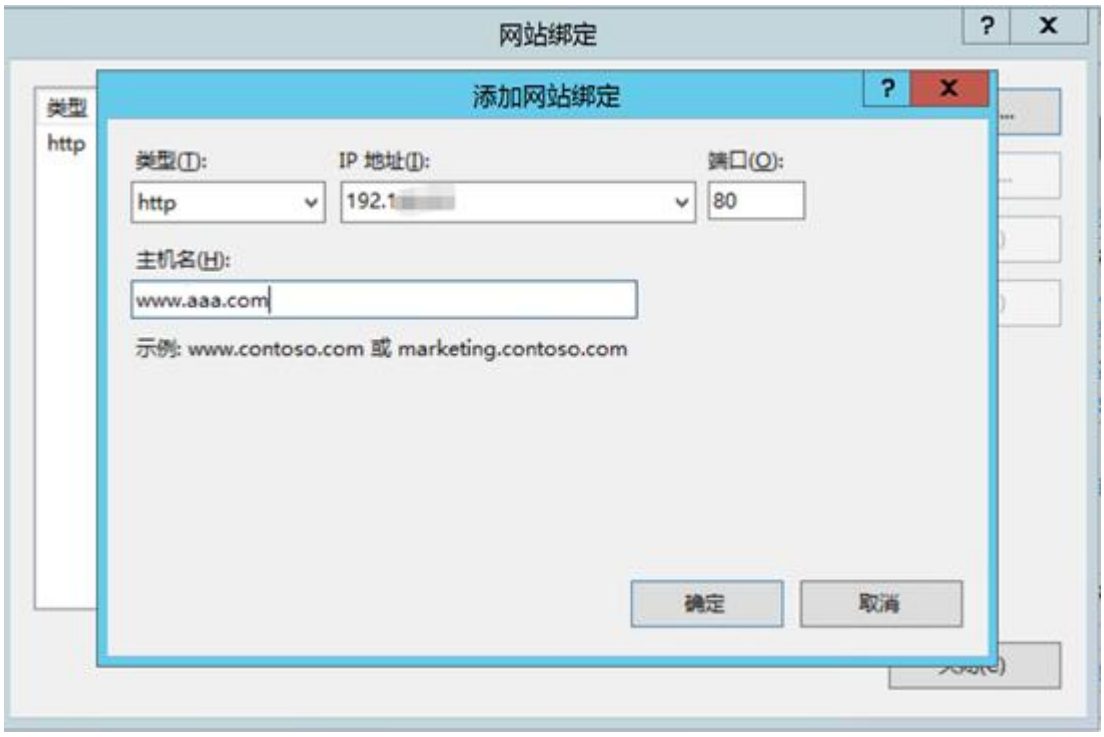




2. 在 IIS 管理器界面，选择需要编辑的网站。单击右键选择“编辑绑定”。



3. 选择待修改的网站，选择相应的域名，单击“编辑”，添加指定弹性云主机的私有 IP 地址。单击确定，完成 IP 绑定。



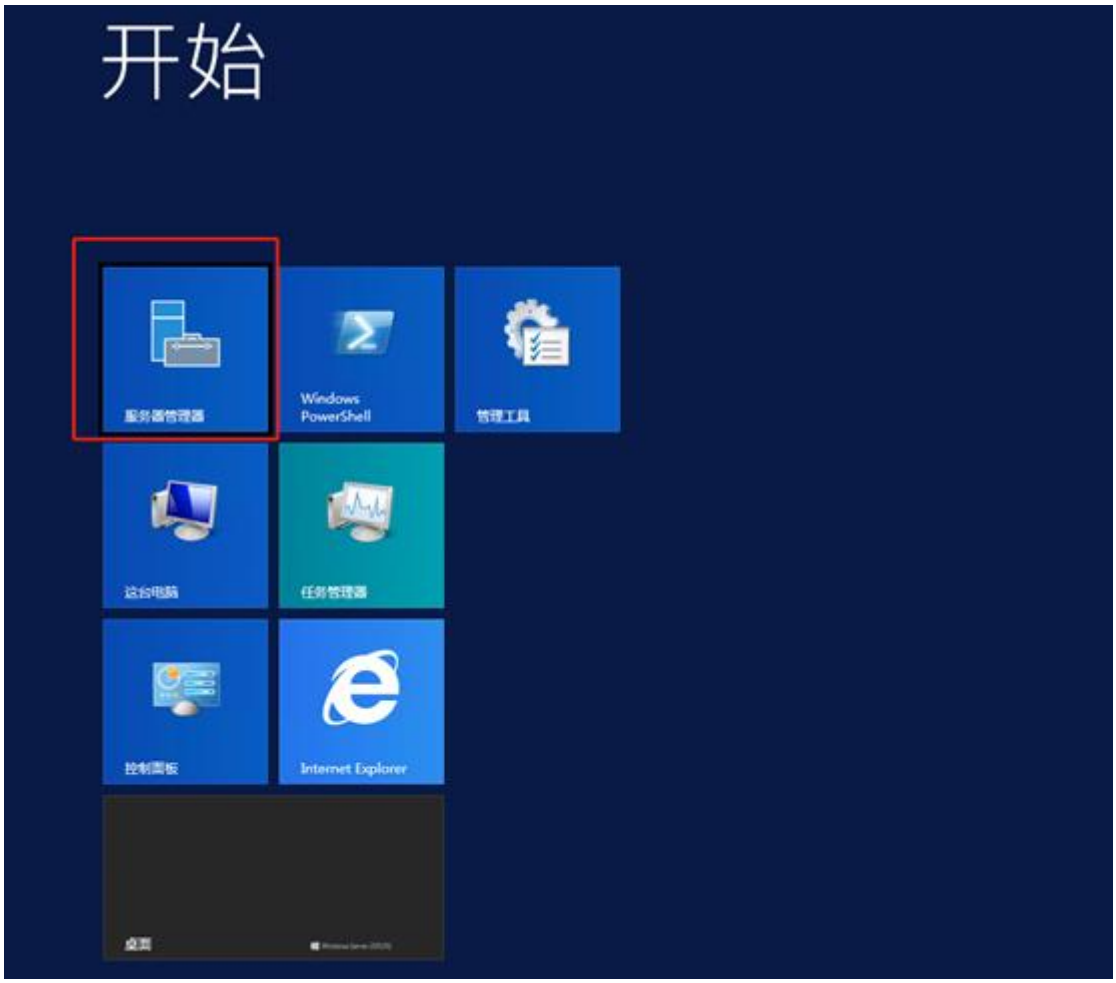
9.12.3Windows 云主机上安装 IIS 服务

操作场景

本文档指导您在 Windows 云主机上安装 IIS 服务，以 Windows Server 2012 标准版 R2 64 位中文版操作系统的云主机为例。

操作步骤

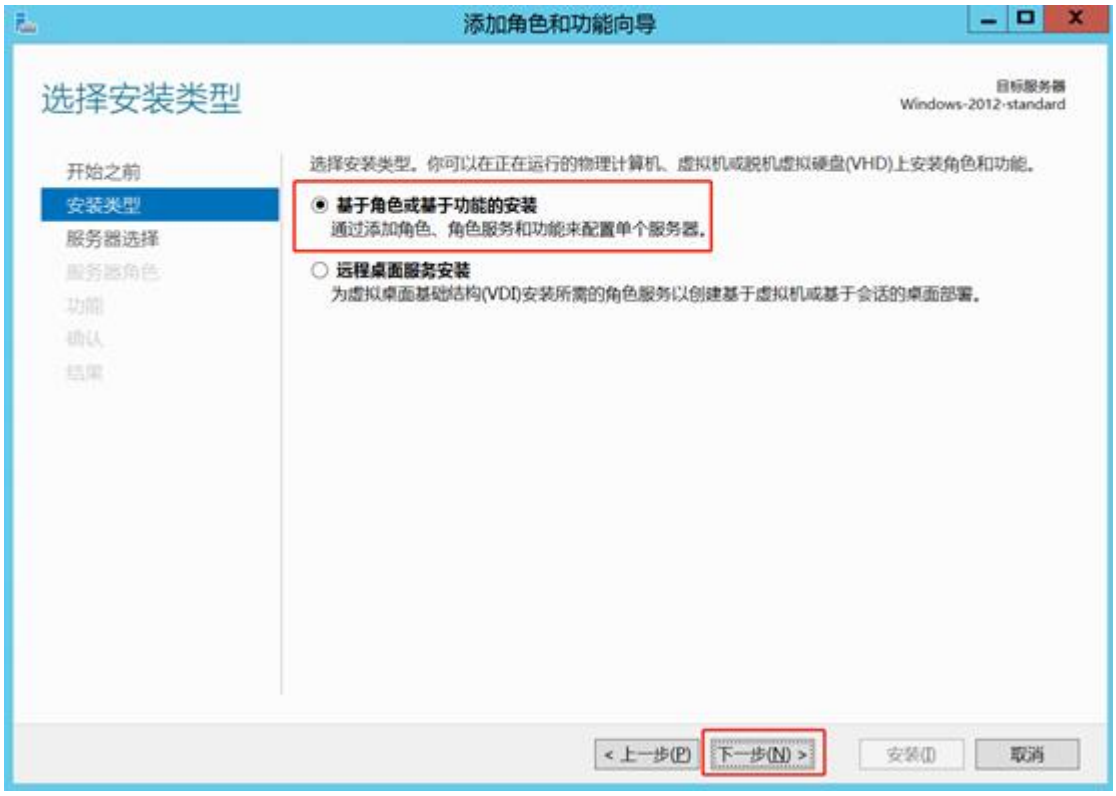
1.打开服务器管理器。



2.在“快速启动”栏，单击“添加角色和功能”。



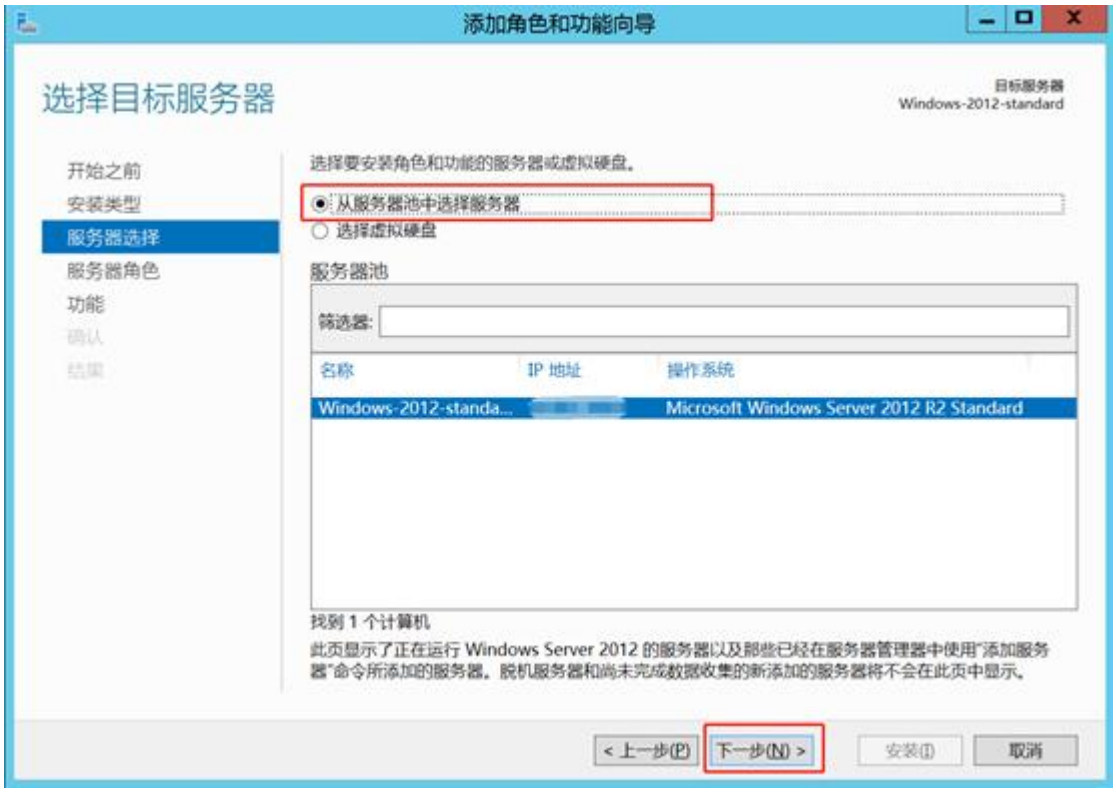
3. 在左侧导航栏，选择“安装类型”。



4. 单击“基于角色或基于功能的安装”，并单击“下一步”。

5. 左侧导航栏选择“服务器角色”。

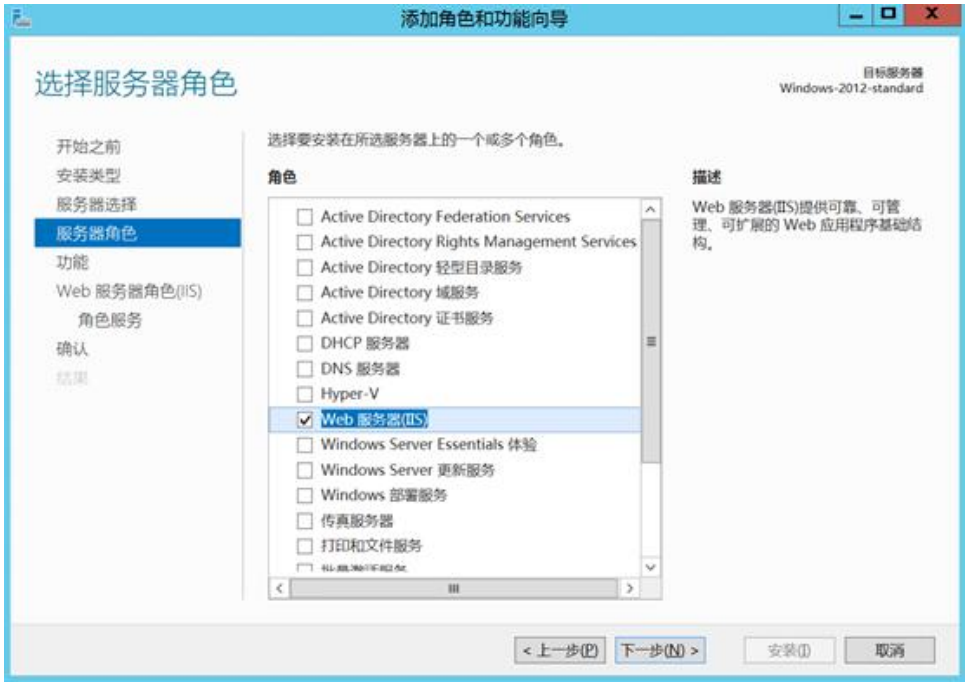
6. 勾选“从服务器池中选择服务器”，并在“服务器池”中选择服务器的计算机名。



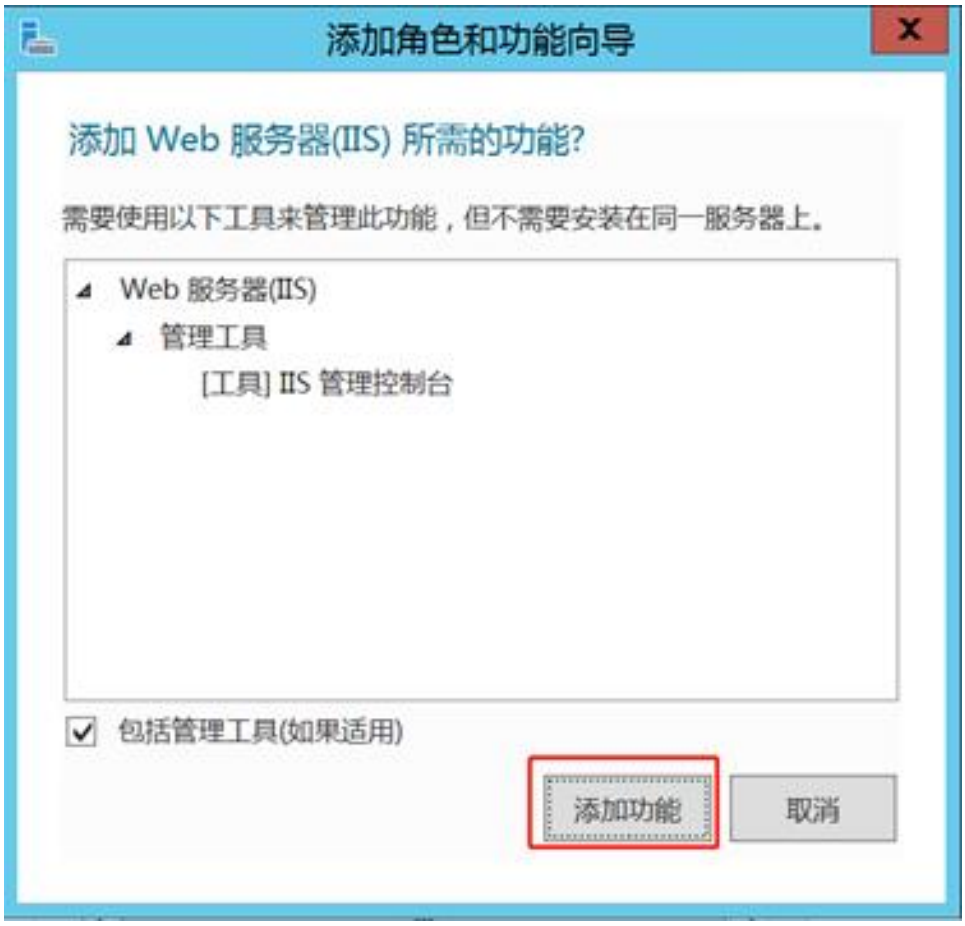
7. 单击“下一步”。

8. 在左侧导航栏，选择“服务器角色”。

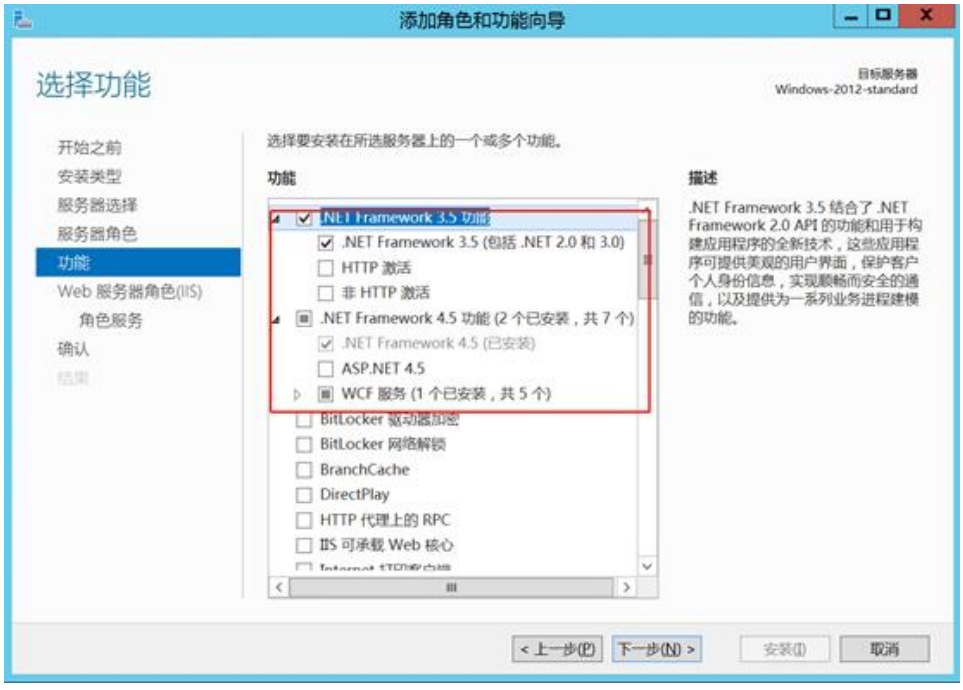
9. 在角色列表内找到“Web 服务器(IIS)”并勾选。



10 在弹窗“添加角色和功能向导”中，单击“添加功能”。



11 单击“下一步”，并勾选“.Net Framework 3.5”和“.Net Framework 4.5”。



12. 两次单击“下一步”，在角色服务列勾选需要安装的项目。

如果您不清楚需要安装哪些项目，除了 FTP 服务器，其他建议全部勾选。



13. 单击“下一步”，确认安装的角色，然后单击“安装(I)”。



14 等待安装完成即可完成 IIS 服务安装。



10 最佳实践

10.1 创建云主机最佳实践

操作场景一

通过控制台开通单台云主机。

操作步骤

本文以南昌 5 资源池为例，创建单台 Linux/Windows 云主机。

计算控制台点击“弹性云主机”，点击右上角“创建云主机”。

“基础配置” 页

1. 计费模式选择按量付费；
2. 地域选择江西-南昌 5；
3. 可用区选择可用区 1；

4. 企业项目选择 default；

5. 虚拟私有云（VPC）选择默认 default-vpc；

6. 实例名称和主机名称按照要求填写即可；

7. 选择所需开通规格，本次以 s7.xlarge.2 为例；

8. 选择所需镜像类型。

“网络配置” 页

1. 网卡会关联上述选择的虚拟私有云（VPC）下的子网，选择相应的子网即可；

2. 安全组可以选择 default 安全组；

3. 弹性 IP 选择“不使用”。

“高级配置” 页

1. 登陆方式默认为密码，创建密码为稍后创建；

2. 用户数据选择“暂不配置”；

3. 左下角购买量选择 1 台。

“确认配置” 页

确认上述选择和填写的配置项符合预期，勾选“已阅读并同意相关协议”，点击右下角的“立即购买”完成订购。

对于单台 Linux/Windows 云主机，通常情况下可在一分钟内完成开通。本次单台 Linux 云主机开通耗时 17 秒，单台 Windows 云主机开通耗时 16 秒。



操作场景二

通过控制台开通多台云主机。

操作步骤

本文以南昌 5 资源池为例，批量创建 20 台 Linux 云主机。

控制台点击“弹性云主机”，点击右上角“创建云主机”。

“基础配置” 页

1. 计费模式选择按量付费；
2. 地域选择江西-南昌 5；
3. 可用区选择可用区 1；
4. 企业项目选择 default；
5. 虚拟私有云（VPC）选择默认自己创建的 VPC；
6. 实例名称和主机名称按照要求填写即可；
7. 选择所需规格，本次以 s7.xlarge.2 为例；
8. 选择所需镜像类型。

“网络配置” 页

1. 网卡会关联上述选择的虚拟私有云（VPC）下的子网，选择相应的子网即可；
2. 安全组可以选择 default 安全组；
3. 弹性 IP 选择“不使用”。

“高级配置” 页

1. 登陆方式默认为密码，创建密码为稍后创建；
2. 用户数据选择“暂不配置”；
3. 左下角购买量选择 20 台。

“确认配置” 页

确认上述选择和填写的配置项符合预期，勾选“已阅读并同意相关协议”，点击右下角的“立即购买”完成订购。

对于多台 Linux 云主机，通常情况下可实现分钟级开通。本次开通 20 台 Linux 云主机耗时 22 秒。



操作场景三

通过 OpenAPI 接口并发开通 1000 台云主机。

接口约束

1. 目前不支持预付费账户创建按需付费类型云主机。
2. 计费模式选择包年包月计费方式时，需要填写订购周期类型与订购时长。
3. 自动分配弹性 IP（extIP="1"）时，需要填写弹性 IP 版本（ipVersion）与带宽大小（bandwidth）；使用已有的弹性 IP（extIP="2"）时，需要填写弹性 IP 的版本（ipVersion），和对应弹性 IP 的 ID（eipID 或 ipv6AddressID）。
4. 挂载网卡时，子网与虚拟私有云存在对应关系，确保子网属于当前虚拟私有云。
5. 云主机绑定多个标签时，标签键（参数 labelKey）不可重复，单台云主机最多可绑定 10 个标签。

操作步骤

1.本文以南昌 5 资源池为例，创建 1000 台按需计费的云主机，其中每批创建 50 台，并发创建 20 批次。

计算控制台点击“镜像服务”，在公共镜像 Tab 选择 CentOS Linux 7.6 64 位镜像，获取镜像 ID。

2.选择需要开通的云主机规格，本次使用 s7.large.2，根据 OpenAPI 的查询云主机规格资源接口获取该规格对应的 flavorID。

3.网络控制台点击“虚拟私有云”，创建虚拟私有云（VPC），同时在该虚拟私有云（VPC）下创建 subnet，保证该 subnet 可使用的 IP 数量超过 1000 个。

4.在官网帮助中心查找 OpenAPI 批量创建云主机接口。

5.批量开通函数封装。

```
@atomic.action_timer("openapi.batch_create_instance")

def _batch_create_instance(self, region, az, flavor, image, vpc, subnet,
ext_ip, disk_type, disk_size,ondemand, order_count, **kwargs):

    """Returns user servers list."""

    body = self.client.servers.batch_create(region, az, flavor, image,
vpc, subnet,              ext_ip, disk_type,
disk_size,ondemand,order_count, **kwargs)

    order = body["returnObj"]["masterOrderID"]     res =
self._query_uuid(order,interval=2, timeout=600)

    return res


def batch_create(self, region, az, flavor, image, vpc, subnet,
ext_ip, disk_type, disk_size, ondemand, order_count,          **kwargs):

name = "s-rally-" + self._random_digits(5)

params = {
```

```
    "clientToken": str(uuid.uuid4()),
    "regionID": region,
    "azName": az,
    "instanceName": name,
    "displayName": name,
    "flavorID": flavor,
    "imageType": 1,
    "imageID": image,
    "bootDiskType": disk_type,    "bootDiskSize": disk_size,
"onDemand": ondemand,
    "orderCount": order_count,    "vpcID": vpc,
    "extIP": ext_ip,    "networkCardList": [
    {
        "subnetID": subnet,        "isMaster": True
    }
    ],
}
params.update(kwargs)
return self._post("/v4/ecs/batch-create-instances", params=params)
```

6.编写 yaml 执行脚本。

- scenario 选择上述编写的 batch_create_instance 方法；
- region、az 对应南昌 5 资源池 ID 和可用区；
- flavor、image 为上述选择的规格和镜像；
- vpc、subnet 是上述创建的虚拟私有云和子网；

- ext_ip 默认为 0，不创建弹性 IP；
- disk_type 选择 SATA 普通 IO，disk_size 选择 40；
- ondemand 选择 true，选择创建按需云主机；
- projectID 为相关的企业项目；
- order_count 选择 50，表示单批次批量创建 50 台云主机；
- runner 中 times 设置 20，代表共批量创建 20 批次；concurrency 设置 20，代表并发 20。
- yaml 可执行脚本如图所示。

```
{% set flavor_name = flavor_name or "m1.tiny" %}
{% set image_name = image_name or "CentOS7.5-test" %}
{% set size = size or 20 %}
{% set volume_type = volume_type or "default_volume_type" %}
{% set external_network = external_network or "ext-net" %}
{% set to_image_name = to_image_name or "CentOS7.5" %}
{% set to_flavor_name = to_flavor_name or "m1.small" %}
{% set volume_size = volume_size or 20 %}
{% set enable_instance_backup = enable_instance_backup or false %}
{% set regionID = regionID or "200000002527" %}

---
  version: 2
  title: Task for SDK test
  description: >
    Test suit for SDK test.
  subtasks:
    -
      title: SDK ComputeFlavors.create_flavor
      scenario:
       #OpenAPI.list_servers_eop:
       OpenAPI.batch_create_instance:
          region: "200000002527"
          az: "cn-jx-nc5-jxnc1A-public-ctcloud"
          flavor: "bf5e0057-6f08-c886-82f6-dd944e5fd531"
          image: "34c74243-dd70-46ed-901d-093965a8b0ec"
          vpc: "vpc-dksq1k6aml"
          subnet: "subnet-vfaps9aszl"
          ext_ip: "0"
          disk_type: "SATA"
          disk_size: 40
          ondemand: true
          projectID: "6d94f8f6396a48858d92e91ac53d4b0d"
          order_count: 50
      runner:
        constant:
          times: 20
          concurrency: 20
```

7. 部署相关脚本到对应的运行机器，执行 yaml 脚本，开始创建云主机。

```
rally task start ***.yaml
```

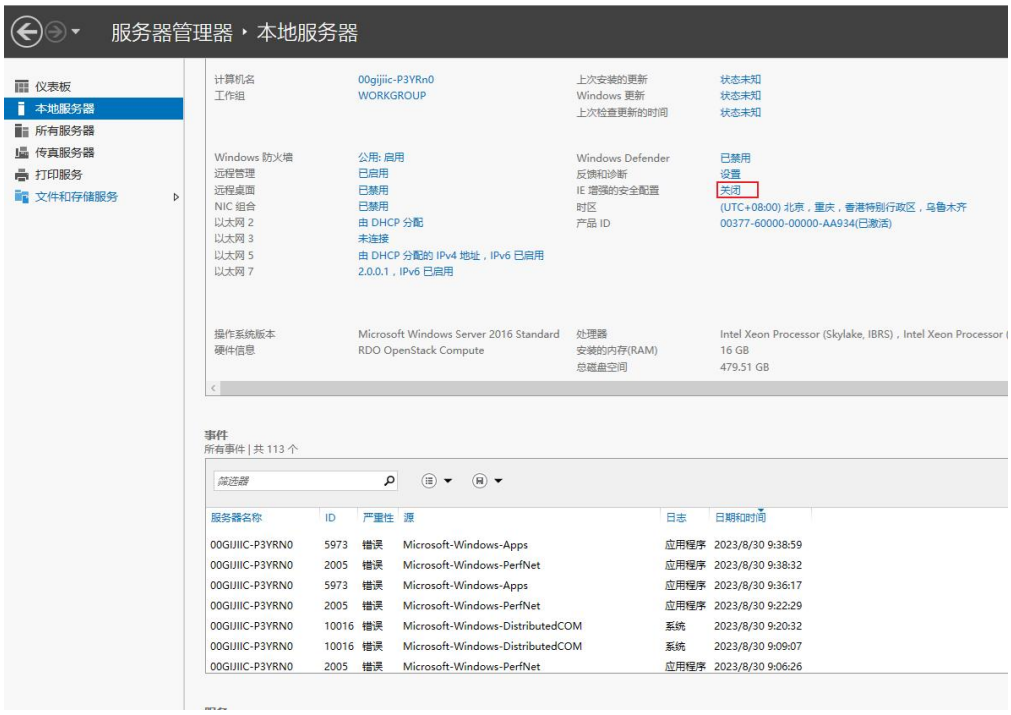
8. 当 20 个订单创建完成后，可在控制台查看“运行中”的云主机数量。对于千台云主机，通常情况下可在十分钟内完成开通。本次批量开通千台云主机耗时 138 秒，开通成功率 100%。



10.2 为 Windows Server 添加 AD、DHCP、DNS、IIS 服务

为 Windows Server 添加 AD、DHCP、DNS、IIS 服务具体操作步骤如下。

1.打开“服务器管理器”，点击“本地服务器“，选择对应的“关闭”选项，之后点击“确定”完成配置。

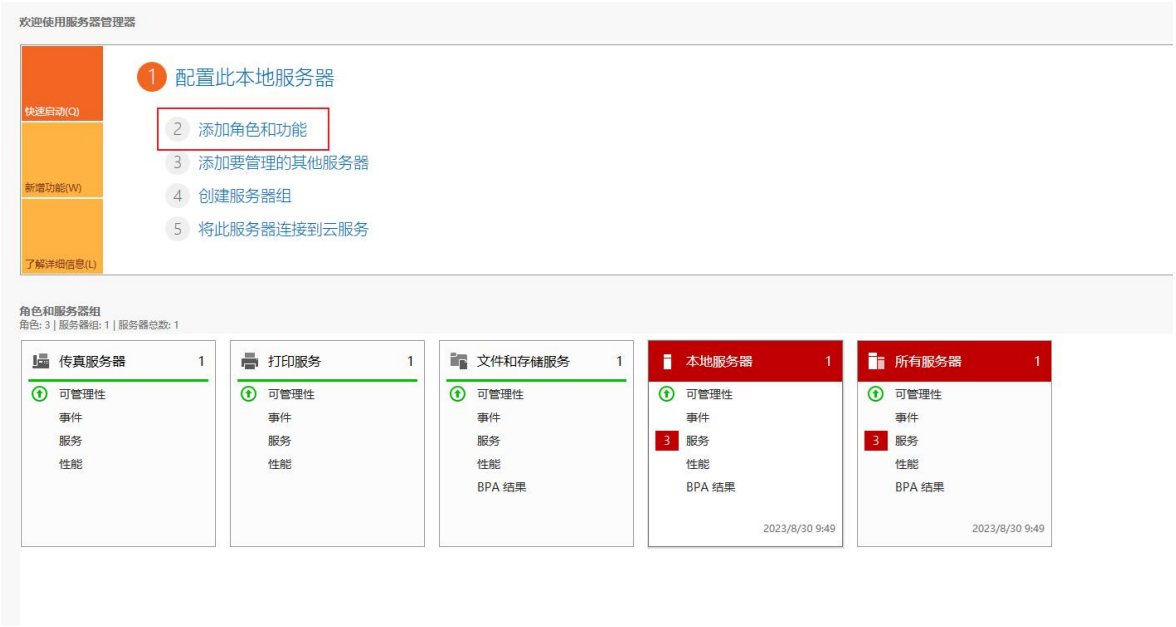




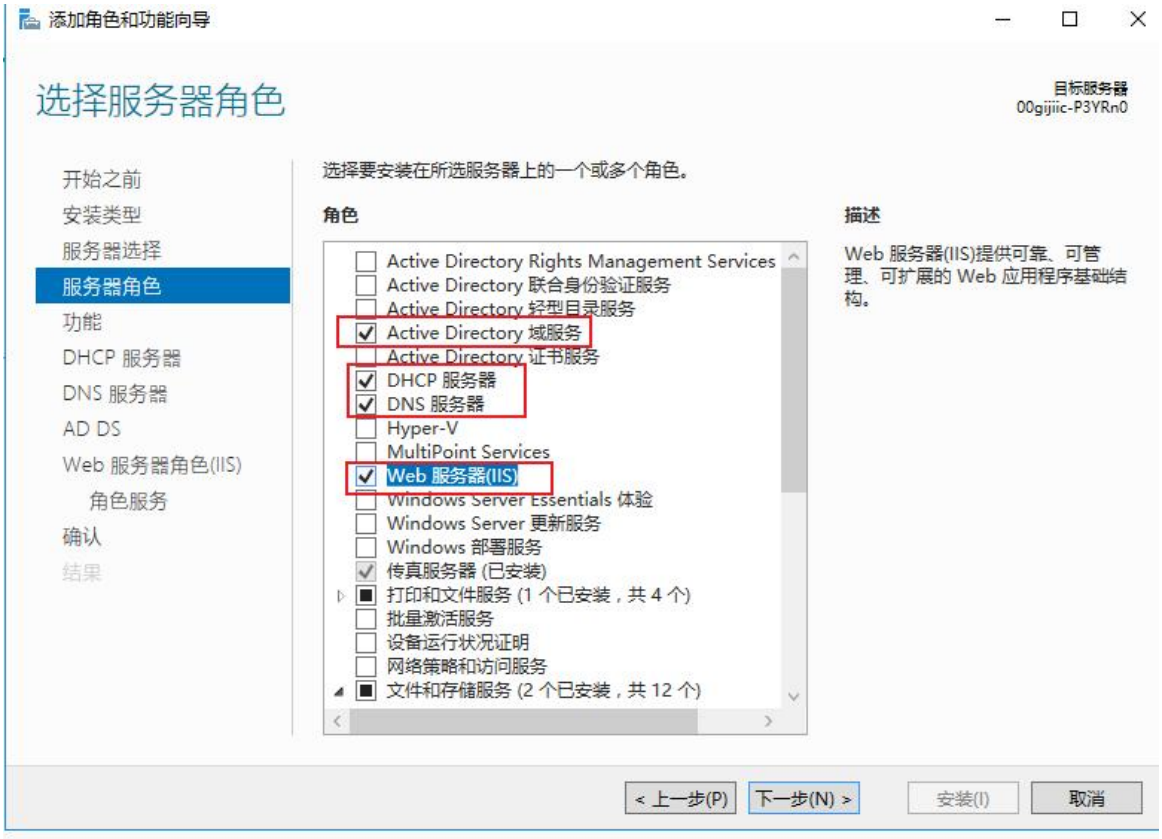
2. 进入“服务器管理器”的页面点击“仪表板”。



3.单击“添加角色和功能”，为服务器添加新的角色和功能，包括 DNS、DHCP、IIS、Net Framework3.5。

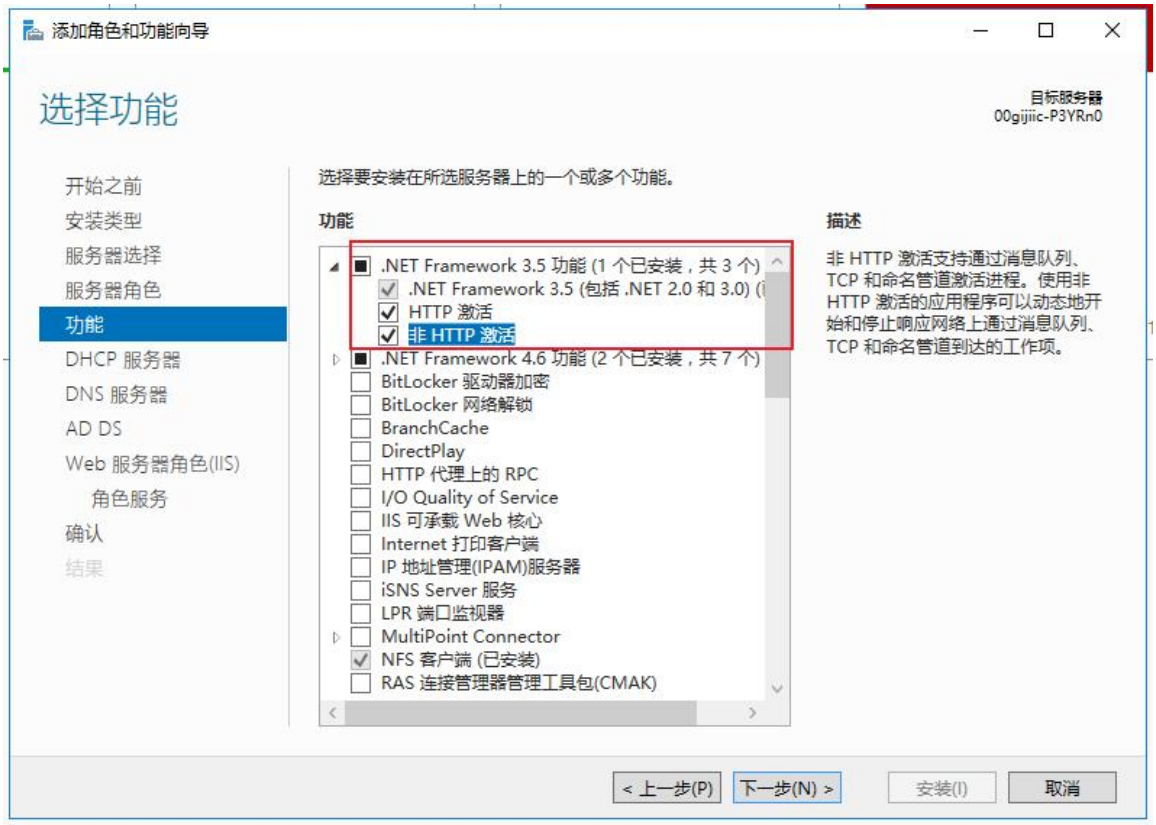


4.在“服务器角色”导航栏中，勾选“Active Directory 域服务”、“DHCP 服务器”、“DNS 服务器”和“Web 服务器(IIS)”。



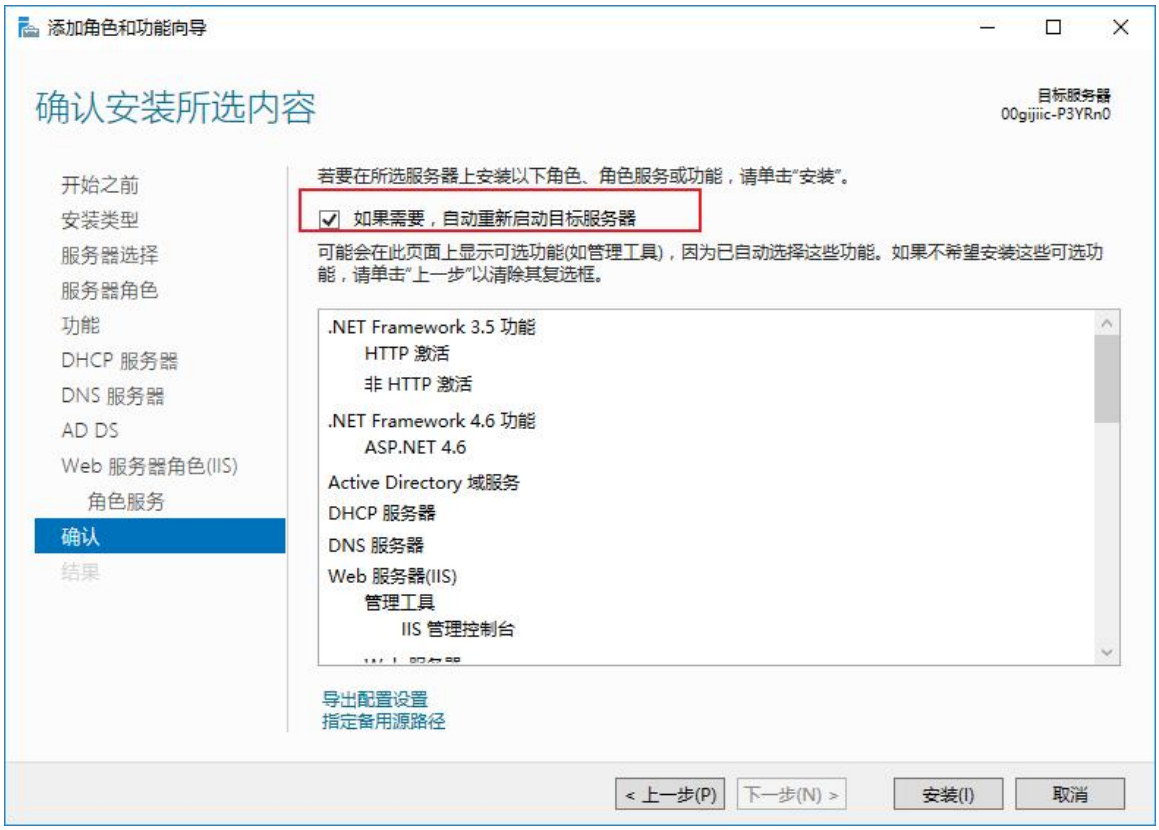
5.单击“下一步”。

6. 在“功能”导航栏中，选中“.NET Framework 3.5 功能”。



7. 单击“下一步”。

8. 在“确认”导航栏中，勾选“如果需要，自动重新启动目标服务器”。



9. 单击“安装”，开始安装。

10. 安装完成后，单击“将此服务器升级为域控制器”，开始配置 AD 服务。

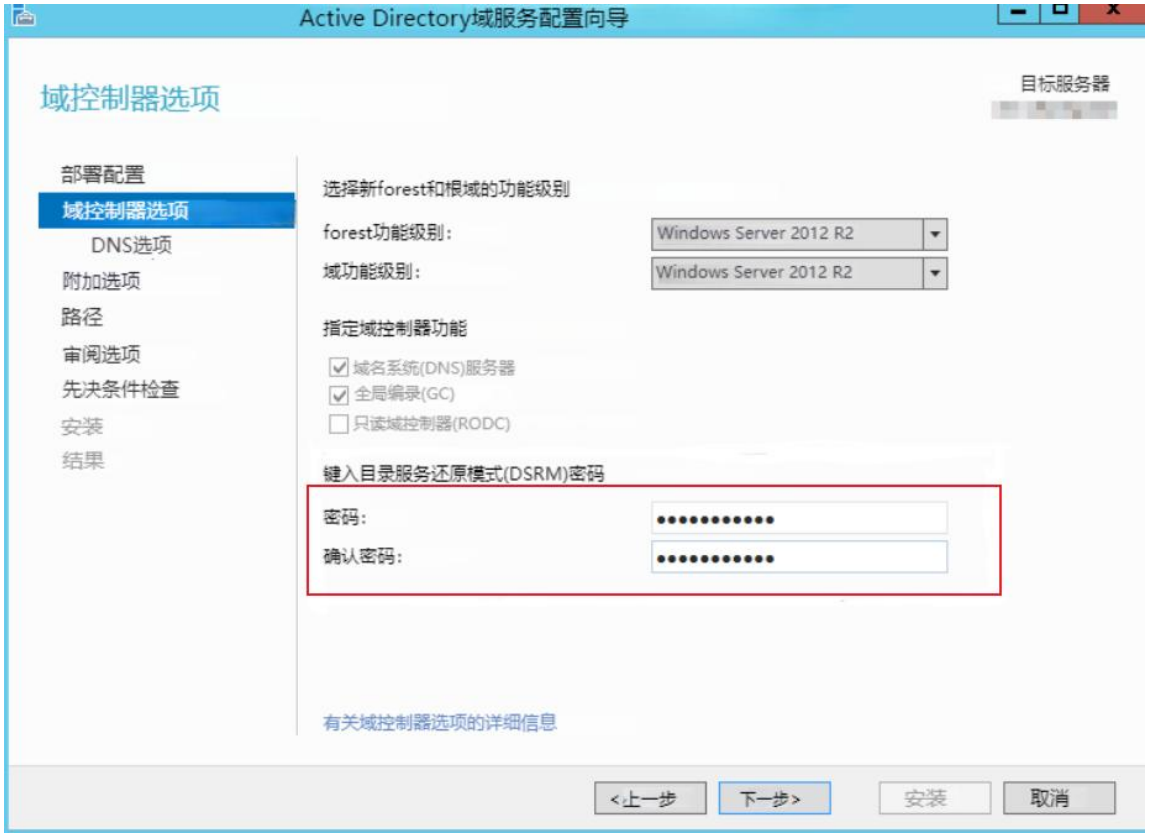


11. 因为没有已经存在的域环境，选择“添加一个新 forest”，创建一个新的 forest 并设置根域名。如“根域名”可以设置为“sp160.com.cn”。



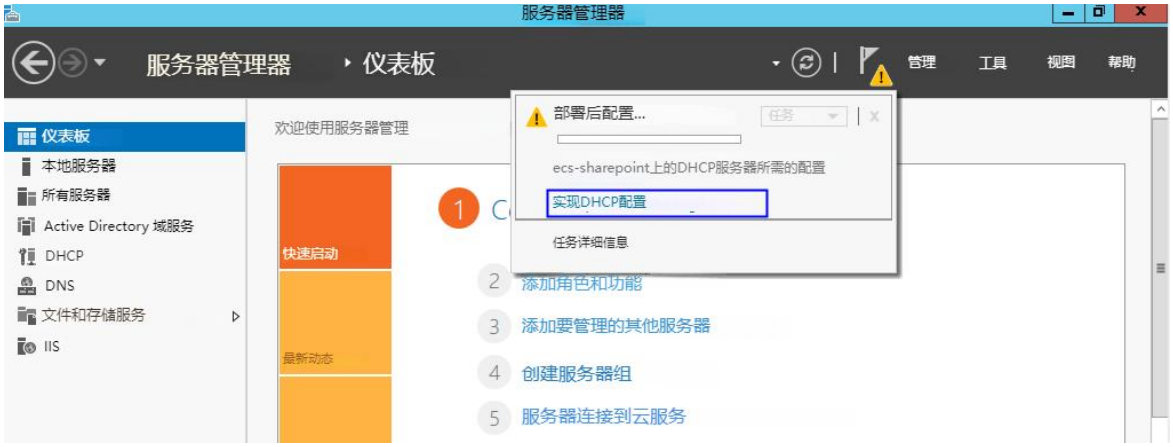
12. 单击“下一步”。

13. 设置密码，该密码用于备份还原域控。

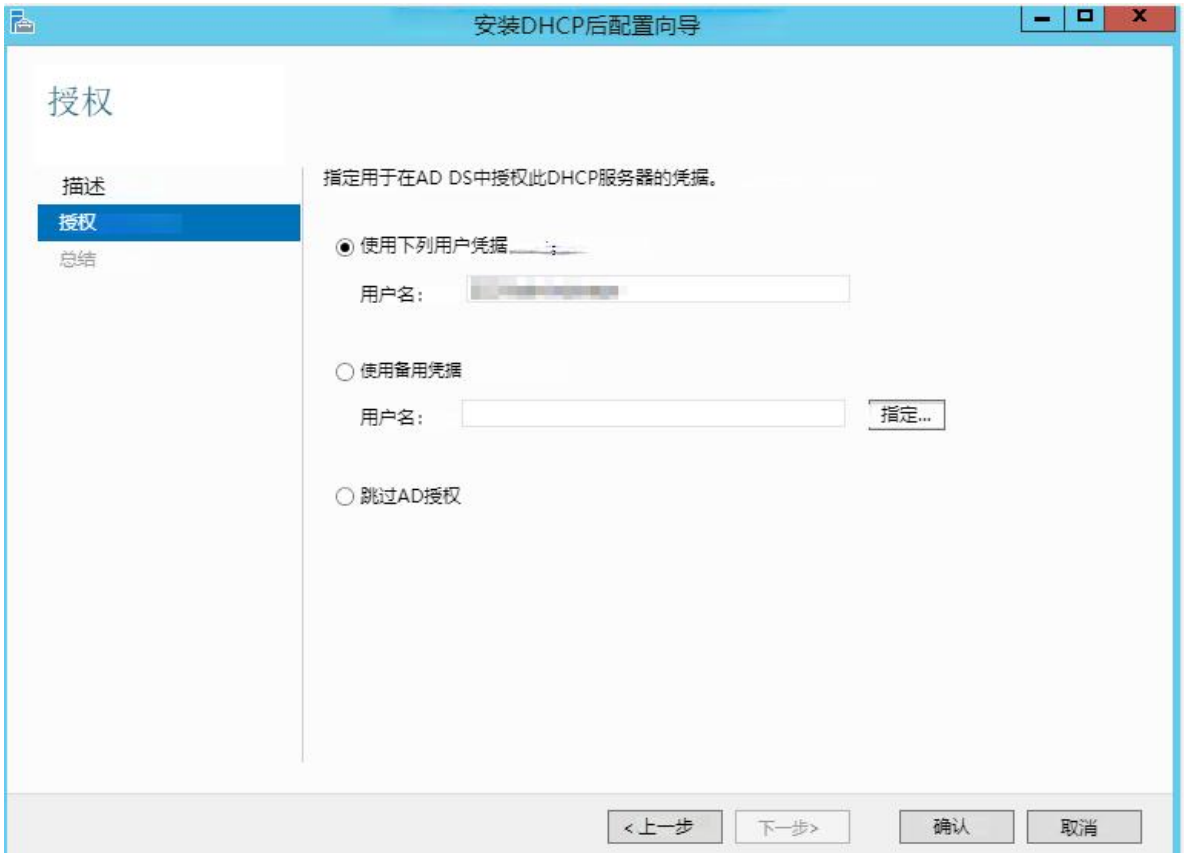


14. 单击“下一步”，直至安装完成。

15. 单击“实现 DHCP 配置”，开始配置 DHCP 功能。



16. 单击“下一步”，选择默认配置。



17. 单击“确认”。

18. 配置完成，单击“关闭”。

10.3 基于 Tomcat 构建 Java web 环境（CentOS 7.4）

应用场景

Tomcat 是一个被广泛使用的 Java Web 应用服务器。本文介绍了在天翼云弹性云主机上部署 Java Web 环境的操作步骤。首先需要下载部署 Java Web 环境所需的安装包，并将安装包上传至云主机，然后设置弹性云主机安全组规则，再安装并配置相关软件，完成开发环境的配置。

本文使用的云主机以天翼云 CentOS 7.4 64bit 操作系统云主机为例。

方案架构

图 1 基于 Tomcat 构建 Java web 环境示意图

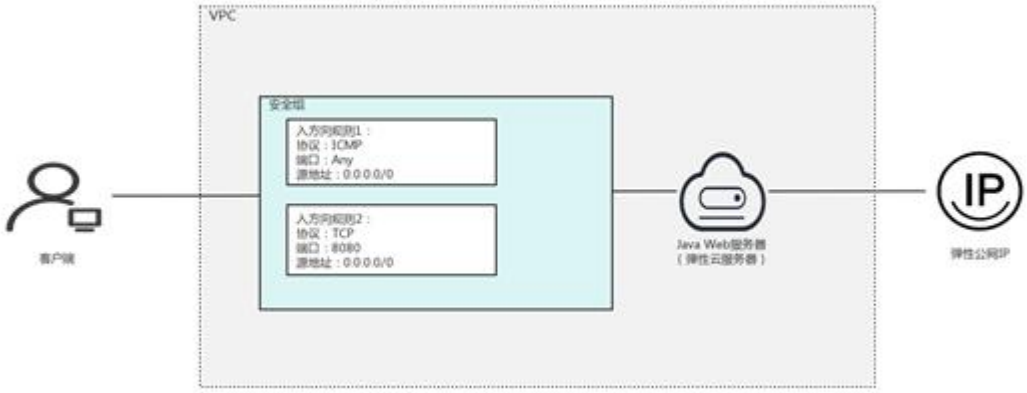


表 1 资源和成本规划

资源	资源说明	成本说明
虚拟私有云 VPC	VPC 网段：192.168.0.0/16	免费
虚拟私有云子网	可用区：可用区 1 子网网段：192.168.0.0/24	免费
安全组	入方向规则 1： 协议/应用：ICMP 端口：Any 源地址：0.0.0.0/0 入方向规则 2： 协议/应用： TCP 端口：8080 源地址：0.0.0.0/0	免费

资源	资源说明	成本说明
弹性云主机	计费模式：包年/包月 可用区：可用区 1 规格：s6.large.2 镜像：CentOS 7.4 64bit 系统盘：40G 弹性公网 IP：现在购买 公网带宽：按流量计费 带宽大小：5Mbit/s	ECS 涉及以下几项费用：云主机、云硬盘、弹性公网 IP。具体的计费方式及标准请参考计费说明。
jdk	Java 开发工具软件。 获取方式： http://www.oracle.com/technetwork/java/javase/downloads	
tomcat	是一款开源的 Web 应用服务器。 获取方式： http://tomcat.apache.org/download-80.cgi	
PuTTY	跨平台远程访问工具。用于在软件安装过程中在 Windows 系统 上访问云主机。 获取方式： https://www.chiark.greenend.org.uk/~sgtatham/putt	

资源	资源说明	成本说明
	y/latest.html	
WinSCP	跨平台文件传输工具。用于在 Windows 系统和 Linux 系统间传 输文件。 获取方式：http://winscp.net/	

说明

表 1 中为 jdk 和 tomcat 官方获取地址，您还可以参考其他开源镜像地址获取安装包。

实施步骤（手动）

安装前准备

- 创建弹性云主机，且弹性云主机已绑定弹性公网 IP。
- 为了更好的获取和更新系统和软件，建议您更新镜像源为天翼云镜像源，详细操作，请参见如何使用天翼云提供的内网 yum 源。
- 登录弹性云主机，执行如下命令，新建 jdk 目录和 tomcat 目录。

```
cd /home/mkdir webDemocd webDemo/mkdir jdkmkdir tomcat
```

- 您可以选择将安装包下载至本地后使用文件传输工具将安装包上传至云主机。或者选择使用 wget 命令直接下载安装包至云主机。
- 方法一：使用文件传输工具上传安装包至云主机。
  - 使用 WinSCP 工具上传 jdk 软件包至云主机 jdk 文件夹。

- 使用 WinSCP 工具上传 tomcat 软件包至云主机 tomcat 文件夹。

- 方法二：使用 wget 命令直接下载安装包至云主机。
- 执行如下命令，进入 jdk 目录。

```
cd /home/webDemo/jdk
```

- 执行如下命令，下载 jdk 软件包。软件包下载地址请参考表 1 查询 jdk 下载地址，或者使用其他开源镜像地址获取安装包。

```
wget jdk 软件包下载地址
```

- 例如：单击开源镜像地址，查询当前可用版本，以 v8.5.93 版本安装包为例，执行以下命令。

```
wget
https://dlcdn.apache.org/tomcat/tomcat-8/v8.5.93/bin/apache-tomcat-8.
5.93.tar.gz --no-check-certificate
```

- 执行如下命令，进入 tomcat 目录。

```
cd /home/webDemo/tomcat
```

- 执行如下命令，下载 tomcat 软件包。

请参考表 1 查询 tomcat 下载地址，或者使用其他开源镜像地址获取安装包。

```
wget

https://dlcdn.apache.org/tomcat/tomcat-x/vx.x.xx/bin/ap

ache-tomcat-x.x.xx.tar.gz
```

例如：单击开源镜像地址，查询当前可用版本，以 v8.5.93 版本安装包为例，执行以下命令。

```
wget
https://dlcdn.apache.org/tomcat/tomcat-8/v8.5.93/bin/apache-tomcat-8.
5.93.tar.gz --no-check-certificate
```

设置弹性云主机安全组规则

1.单击弹性云主机名称，查看弹性云主机详情，在弹性云主机详情页面，选择“安全组”。

2.在“安全组”界面，进入安全组详情界面。

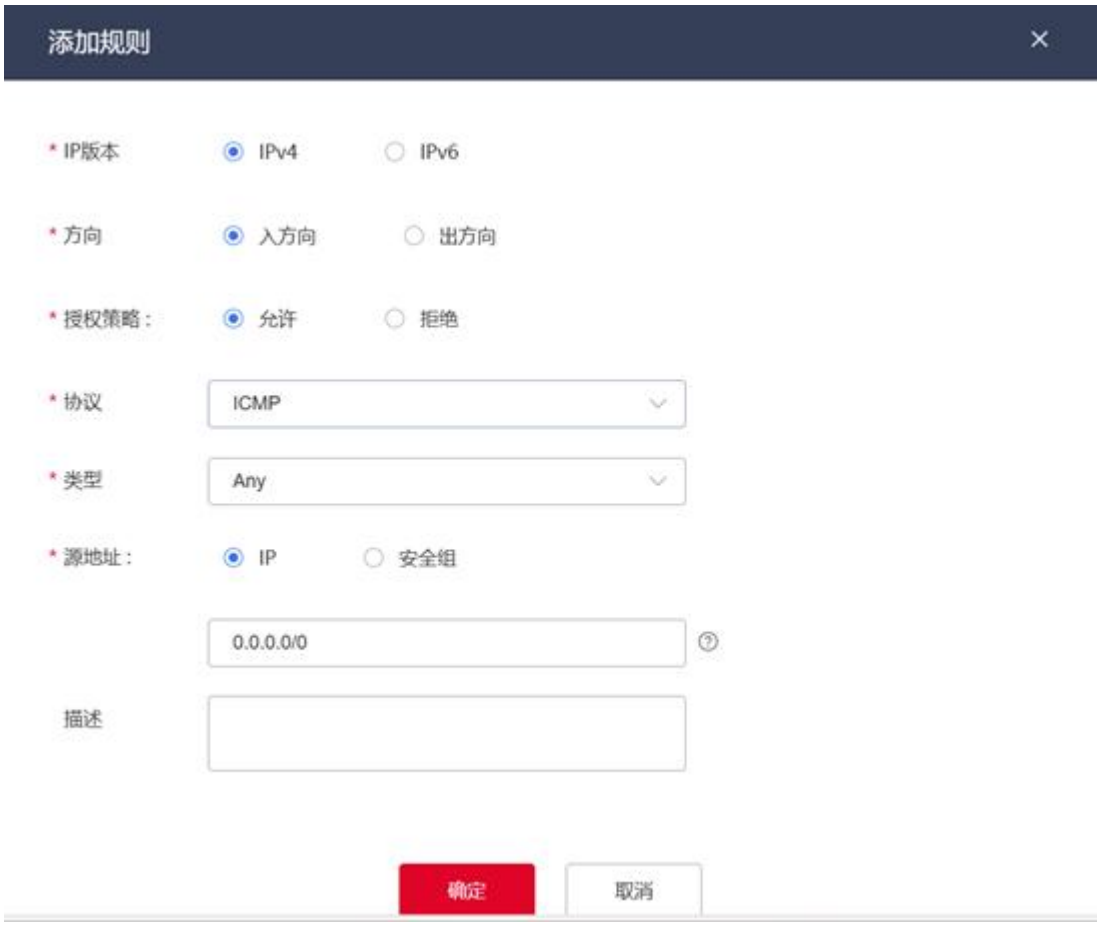
3.在安全组详情界面，单击“添加规则”，弹出添加规则窗口。

4.根据界面提示配置安全组规则。部署 JavaWeb 环境需为弹性云主机添加两个安全组规则。

a. 为云主机添加 ICMP 安全组规则。

如果云主机默认设置是禁止 ICMP 规则，当 ping 弹性服务器 IP 时会显示超时。因此首先为云主机添加 ICMP 规则。

图 3 添加 ICMP 规则

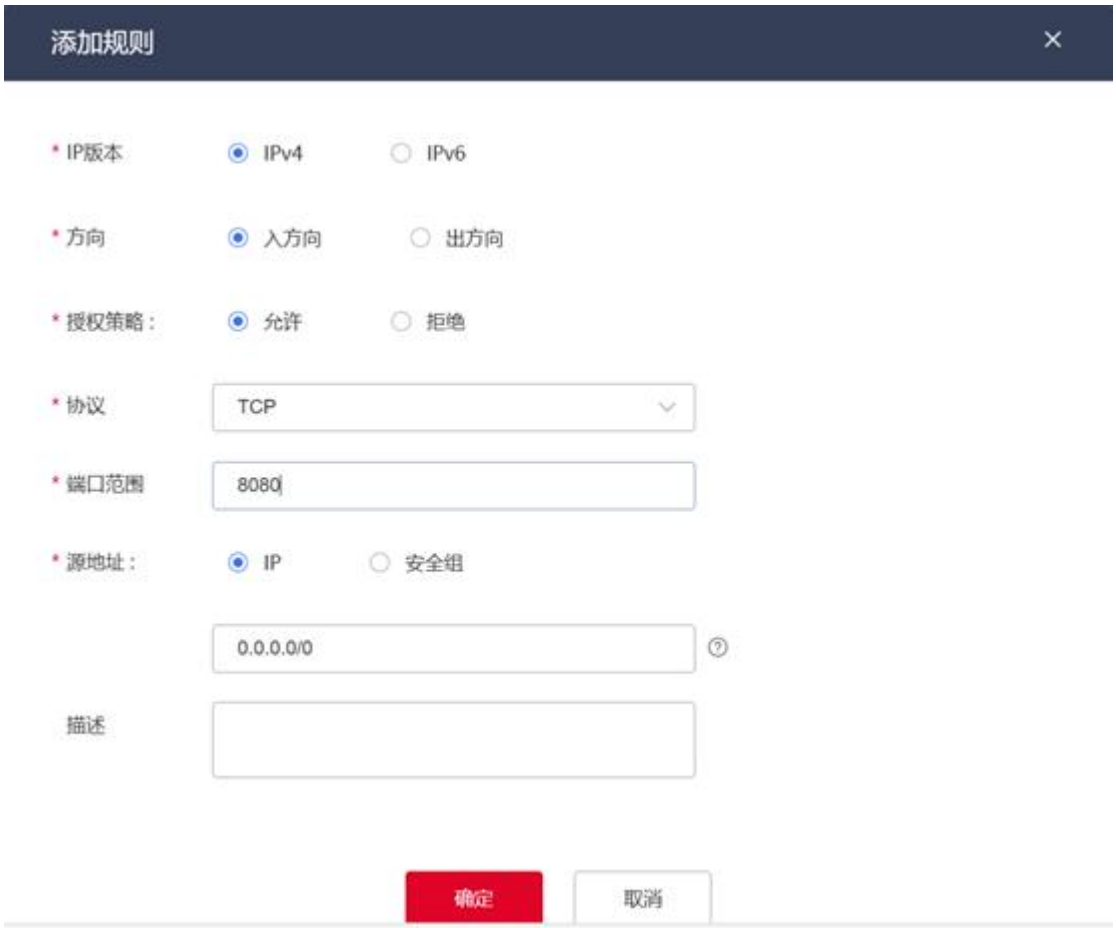


注意

天翼云云主机默认设置是放行 ICMP 规则。

b. 为云主机添加 web 项目分配端口的访问规则，以 8080 端口为例。

图 4 添加 8080 端口



安装 jdk

1.执行如下命令，进入 jdk 目录。

```
cd /home/webDemo/jdk
```

2.解压 jdk 安装包到 jdk 目录下。

```
tar -xvf jdk-17_linux-x64_bin.tar.gz -C /home/webDemo/jdk/
```

3.配置环境变量。

```
vim /etc/profile
```

4.在底部添加以下内容。

```
#set java environment

JAVA_HOME=/home/webDemo/jdk/jdk-17.0.x
```

```
JRE_HOME=$JAVA_HOME

PATH=JAVA_HOME/bin:JAVAHOME/bin:PATH

CLASSPATH=.:JAVA_HOME/lib/dt.jar:JAVAHOME/lib/dt.jar:JRE_HOME/lib/tools.jar

export JAVA_HOME JRE_HOME PATH CLASSPATH
```

说明

“jdk-17.0.x”表示 jdk 安装包的具体版本，实际值需要从步骤 2 的返回值中获取。例如：jdk-17.0.8。

5.执行以下命令保存并退出。

```
:wq
```

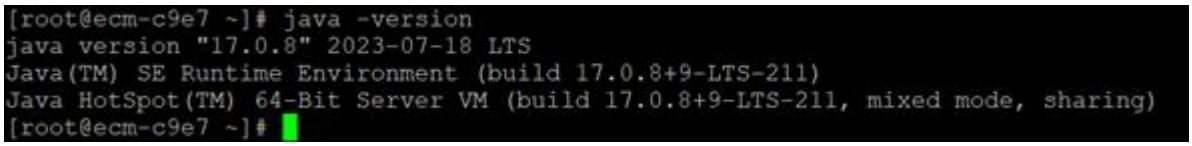
6.执行以下命令使/etc/profile 里的配置生效。

```
source /etc/profile
```

7.验证安装。

```
java -version
```

回显信息如下所示验证安装 jdk 成功。



安装 tomcat

1.执行如下命令，进入 tomcat 目录。

```
cd /home/webDemo/tomcat
```

2.解压 tomcat 安装包到 tomcat 目录下。

```
tar -xvf apache-tomcat-x.x.xx.tar.gz -C /home/webDemo/tomcat/
```

例如：以 v8.5.93 版本安装包为例，执行以下命令。

```
tar -xvf apache-tomcat-8.5.93.tar.gz -C /home/webDemo/tomcat/
```

3.进入 tomcat 的 bin 目录，执行以下命令安装 tomcat。

```
cd /home/webDemo/tomcat/apache-tomcat-x.x.xx/

cd bin/
```

例如：以 v8.5.93 版本安装包为例，执行以下命令。

```
cd /home/webDemo/tomcat/apache-tomcat-8.5.93/

cd bin/
```

4.执行如下命令编辑 setclasspath.sh 脚本。

```
vi setclasspath.sh
```

并在 setclasspath.sh 脚本底部添加以下内容。

请根据资源和成本规划中 jdk 的版本号替换如下内容中的 jdk 版本号。

```
export JAVA_HOME=/home/webDemo/jdk/jdk-17.0.8export
JRE_HOME=$JAVA_HOME
```

5.执行如下命令保存后退出。

```
:wq
```

6.可输入以下命令启动 tomcat。

```
./startup.sh
```

7.执行如下命令查看 tomcat 进程。

```
ps -ef | grep tomcat
```

若返回如下图所示，表示 tomcat 启动成功。

图 5 查看 tomcat 进程



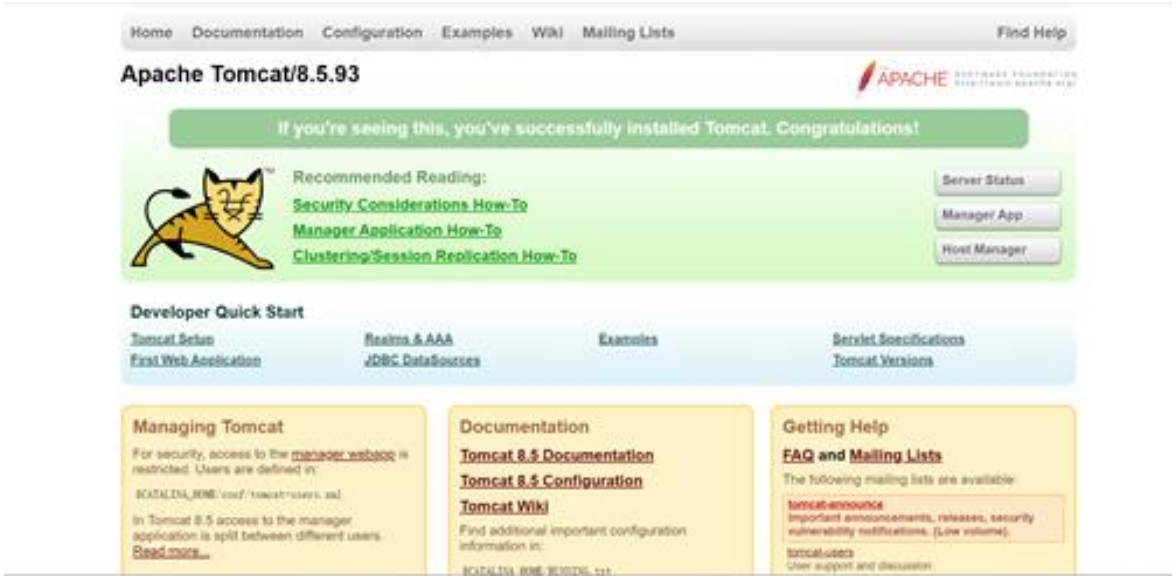
验证 Java Web 环境搭建完成

在浏览器输入以下内容。

http:// 云主机弹性公网 IP:8080

如果界面跳转至默认的 Tomcat 界面，证明 Java Web 环境搭建完成。我们就可以在公网访问云主机的 8080 端口了。

图 6 访问云主机的 8080 端口



10.4 基于弹性云主机部署 WordPress 的安全防护

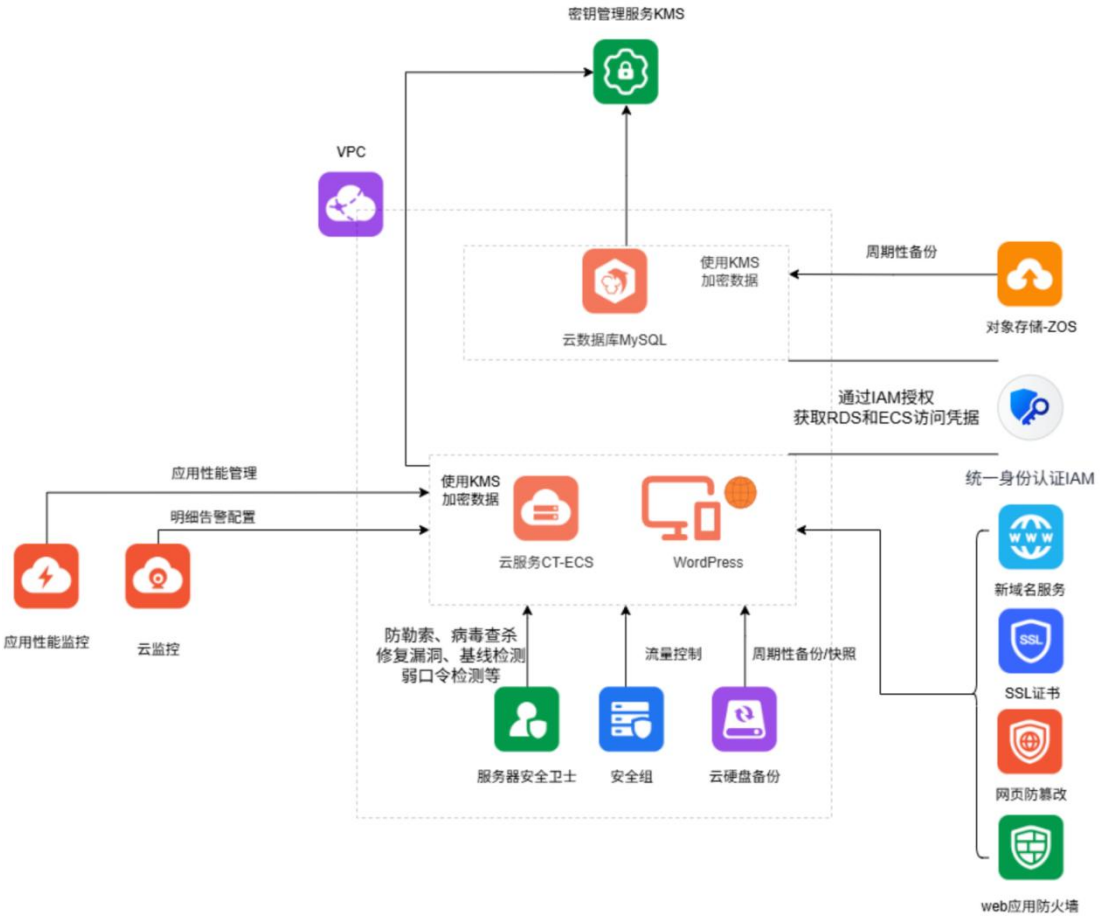
安全设计原则

- 规定时间内修复已知 OS 安全漏洞，满足安全补丁管理基线；同步配置弱口令检测、病毒查杀、重要文件防勒索及网页防篡改功能。
- 定期备份关键数据，防意外丢失，确保数据可恢复。
- 对云硬盘与云数据库存储加密，配置 SSL 加密数据连接。
- 集成精细化权限管理、多因子认证及动态访问控制，构建全链路安全防护。

- 通过逻辑隔离与最小权限访问，阻断威胁横向渗透路径。

部署架构

在云服务器 CT-ECS 中部署 WordPress 使用的安全架构如下。



在部署 WordPress 过程中需要创建 ECS 实例和 RDS MySQL 数据库，本实践针对这两者做的安全配置如下：

- 弹性云主机实例

1. 通过安全组控制 CT-ECS 实例的出入流量安全组是一种网络安全防护机制，用于防止未经授权的访问和保护计算机网络免受恶意攻击。它是一种虚拟防火墙，用于限制入向和出向网络流量。关于安全组的使用请参考：安全组概述。
2. 使用云硬盘快照策略对云硬盘进行周期性备份天翼云云硬盘快照是一种数据备份方式，云硬盘快照服务可以备份或者恢复整个云硬盘的数据，常用于数据备份、制作镜像、应用容灾等。关于云硬盘快照的使用请参考：云硬盘快照策略。

3. 使用 KMS 加密数据密钥管理服务（Key Management Service，KMS）是一站式密钥管理和数据加密服务平台，提供安全合规、可靠易用的资源托管及密码运算服务。同时与天翼云云硬盘、对象存储、弹性文件、关系型数据库 MYSQL 等云产品无缝集成，实现云上原生数据的加密保护。更多信息，请见云硬盘加密概述。
4. 使用 IAM 授权获取访问凭证统一身份认证（Identity and Access Management，简称 IAM）服务，是提供用户进行权限管理的基础服务，可以帮助您安全的控制云服务和资源的访问及操作权限。更多信息，请参见产品定义。
5. 使用服务器安全卫士进行防护服务器安全卫士（原生版）（CT-CSS，Cloud Security System）是一款全方位保障云上服务器安全的产品，能全面识别并管理服务器中的信息资产、实时监测服务器风险并阻止非法入侵行为，当发现服务器出现安全问题时，第一时间向您发出告警通知。主要包括资产清点、漏洞扫描、入侵检测、基线检查、弱口令检测、病毒查杀、文件防勒索等功能，帮助您构建服务器安全防护体系。更多信息，请参见产品简介。

- 关系型数据库 MySQL 版
- 使用 SSL 数据加密 SSL（Secure Socket Layer，安全套接层），位于可靠的面向连接的网络层协议和应用层协议之间的一种协议层。SSL 通过互相认证、使用数字签名确保完整性、使用加密确保私密性，以实现客户端和服务器之间的安全通讯。更多信息，请参考设置 SSL 数据加密。
- 使用 IAM 身份认证关系数据库 MySQL 版支持 CTIAM 身份认证和多种访问控制，多维度保障您云数据库的安全。更多信息，请参考主子账号和 IAM 权限管理。

资源规划说明

本实践将指导用户在 CTyunOS 2.0.1 操作系统上的云主机实例上部署 WordPress，均在华东 1 资源池操作

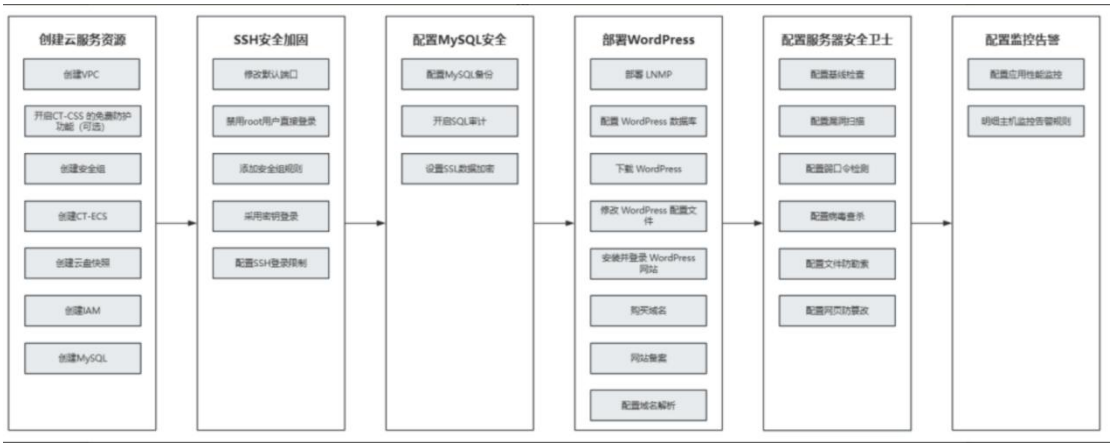
- 资源规划总体说明

资源类型	资源数量	说明
弹性云主机	1	● 名称：根据实际情况进行填写。 ● 镜像：根据实际情况进行选择，本示例为：CTyunOS 2.0.1。 ● 系统盘：通用型 SSD 盘 40GB。 ● 磁盘模式：VBD。 ● 数据盘：本示例未选购数据盘，请您根据实际业务需求选购数据盘。 ● 网络： 1. 虚拟私有云：选择您的虚拟私有云，本示例为 vpc-test。 2. 子网：选择子网，本示例为 subnet-test。 ● 安全组：请根据实际情况选择，安全组名称为 sgs-test ● 私有 IP 地址：192.168.1.100。
虚拟私有云	1	● VPC 名称：请根据实际情况填写，本示例为 vpc-test。 ● IPv4 网段：请根据实际情况填写，本示例为 192.168.0.0/16。 ● 子网名称：请根据实际情况填写，本示例为 subnet-test。 ● 子网 IPv4 网段：请根据实际情况填写，本示例为 192.168.1.0/24。
弹性 IP	1	● 计费模式：请根据情况选择计费模式，本示例为按需计费。 ● EIP 地址：EIP 地址系统随机分配，本示例为

资源类型	资源数量	说明
		121.X.X.2。
安全组	1	● 安全组名称：sgs-test。
关系型数据库 MySQL 版	1	● 创建内网数据库实例。 ● 实例名称：mysql-test。 ● 网络：选择已创建的虚拟私有云和子网，vpc-test 和 subnet-test。 ● 安全组：选择已创建的安全组，名称 sgs-test。 ● IPv4 地址：192.168.1.11。

操作步骤

流程图如下



步骤一：创建云服务资源

1. 创建一个 VPC 和 1 个子网具体操作步骤请参考：创建 VPC、子网搭建私有网络。
2. 开启服务器安全卫士（原生版）的免费防护功能（建议选择）具体操作步骤请参考：开启免费防护。
3. 创建安全组，建议创建最小权限的安全组规则具体操作步骤请参考：创建安全组。

方向	授权策略	类型	优先级	协议	端口范围	远端	描述
入方向	允许	IPv4	1	Tcp	22	0.0.0.0/0	放通安全组内 ECS 的 SSH(22)端口，用于远程登录 Linux ECS。
入方向	允许	IPv4	1	Tcp	81	0.0.0.0/0	临时使用，用于测试 WordPress。
入方向	允许	IPv4	1	Any	Any	sgs-test(sg-my7uktu281)	针对 IPv4，用于安全组内 ECS 之间网络互通。
出方向	允许	IPv4	1	Any	Any	0.0.0.0/0	针对 IPv4，用于安全组内 ECS 访问外部，允许流量从安全组内 ECS 流出。
出	允	IPv	1	An	An	::/0	针对 IPv6，

方向	授权策略	类型	优先级	协议	端口范围	远端	描述
方向	许	6		y	y		用于安全组内 ECS 访问外部，允许流量从安全组内 ECS 流出。

4. 创建弹性云主机具体操作步骤请参考：创建弹性云主机。
   - 网络：选择已创建的虚拟私有云和子网，vpc-test 和 subnet-test
   - 安全组：选择已创建的安全组，名称 sgs-test
   - 开启主机安全防护：开启（建议勾选）
   - 存储：磁盘加密 （建议勾选）
     1. 如果没有密钥则创建密码，参考创建密钥
   - 弹性 IP：自动分配，或使用已有的弹性 IP
5. 创建云盘快照策略具体步骤请参考：创建云硬盘快照。
6. 创建 IAM 相关授权，按照用户职责规划用户组，并将对应职责的权限授予用户组。具体步骤请参考：创建用户组和授权具体步骤请参考：创建 IAM 用户和登录。
7. 创建内网关系型数据库 MySQL 版具体步骤请参考：创建实例。
   - 版本：8.0
   - 实例规格：x86
   - 存储：
     1. 存储空间自动扩展 （可选）
     2. 磁盘加密 （可选）

- 网络：选择已创建的虚拟私有云和子网，vpc-test 和 subnet-test
- 安全组：选择已创建的安全组，名称 sgs-test
- IPv4 地址，手动指定：192.168.1.101
- 加入白名单：是
- mysql 设置 IAM，参考：主子账号和 IAM 权限管理

步骤二：SSH 安全加固

SSH 是远程登录 Linux 服务器的主要方式，但它也是黑客进行暴力破解和恶意攻击的主要入口之一。为了加强 Linux 云主机的 SSH 登录安全，请参考：Linux 服务器 SSH 登录的安全加固，进行安全加固。

步骤三：配置 MySQL 安全

- 配置 MySQL 备份，请参考：备份功能简介
- 配置 SQL 审计，请参考：开启 SQL 审计日志（开启后您可以选择性的是否使用 DMS 操作审计，请参考：DMS 操作审计），SQL 审计日志保留期最长为 7 天，如果您需要永久保存，请参考：云审计
- 配置 MySQL-SSL 数据加密，请参考：设置 SSL 数据加密

步骤四：在弹性云主机上部署 WordPress

说明：

- 不同的操作系统搭建 WordPress 的命令存在差异，请根据您的 CT-ECS 实例操作系统选择对应的操作，否则会安装失败。
- 本文档以 CTyunOS 2.0.1 64 位操作系统，MySQL 8.0，PHP 8.1，为例部署 WordPress 6.8。
- WordPress 对 PHP、MySQL 版本存在依赖关系，如果版本不匹配则会导致 WordPress 安装失败，具体依赖信息，请移步到 WordPress 官网 --> WordPress Compatibility 进行查看。

1.部署 LNMP

1.1、部署 nginx：

添加 nginx 官方源：

```
sudo tee /etc/yum.repos.d/nginx.repo <<-'EOF'
[nginx-stable]
name=nginx stable repo
baseurl=http://nginx.org/packages/centos/7/$basearch/
gpgcheck=1
enabled=1
gpgkey=https://nginx.org/keys/nginx_signing.key
module_hotfixes=true
EOF
```

安装 nginx：

```
yum -y install nginx
```

启动 nginx 服务，并设置为开机自启：

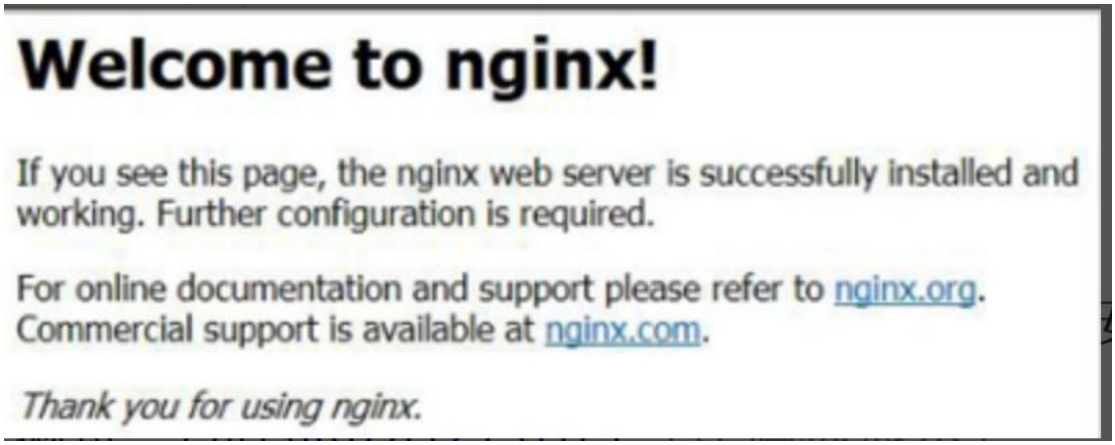
```
systemctl start nginx && systemctl enable nginx
```

使用浏览器访问“http://服务器 IP 地址”，显示如下页面，说明 nginx 安装成功，如未备案可以在系统内使用 curl 命令进行测试：

```
curl http://127.0.0.1
```

注意

web 服务默认通过 80 端口访问，但由于政策原因，该端口需要备案才能访问，默认无法访问，具体备案操作详见备案基础知识-网站备案-备案介绍 - 天翼云(ctyun.cn)。



- 如果您未备案，需要浏览器访问，可以先修改 nginx 端口进行测试：

```
vim /etc/nginx/conf.d/default.conf
```

修改 listen 字段，80 改为 81 重启 nginx 使配置生效：

```
systemctl restart nginx 或 nginx -s reload
```

- 如果您想要配置使用 https，您可以购买 SSL 证书，来搭配域名进行配置使用

1.2、部署 mysql 客户端

配置 IPv4 连接优先级（如果 vpc 没配 IPv6 忽略）：

```
echo "precedence ::ffff:0:0/96 100" > /etc/gai.conf
```

下载 MySQL YUM 仓库配置：

```
yum install -y
https://dev.mysql.com/get/mysql80-community-release-el7-7.noarch.rpm
```

更新仓库缓存：

```
yum clean all && yum makecache
```

安装 MySQL 客户端：

```
yum install -y --nogpgcheck mysql-community-client
```

测试连接，参考内网连接 MySQL：

```
mysql -h 192.168.1.11 -P 13049 -u root -p --ssl-ca=ca.pem
```

1.3、部署 PHP 8.1 版本

安装编译依赖：

```
yum install -y gcc make autoconf libxml2-devel openssl-devel curl-devel libpng-devel libjpeg-devel libzip-devel oniguruma-devel sqlite-devel bzip2-devel libxslt-devel
```

下载并编译 PHP 8.1：

```
wget https://www.php.net/distributions/php-8.1.0.tar.gz

tar -zxvf php-8.1.0.tar.gz

cd php-8.1.0

./configure \
  --prefix=/usr/local/php81 \
  --with-openssl \
  --with-config-file-path=/usr/local/php81/etc \
  --enable-fpm \
  --with-mysqli=mysqlnd \
  --with-pdo-mysql=mysqlnd \
  --enable-opcache \
  --enable-zip \
  --enable-calendar \
  --enable-exif \
  --enable-ftp \
  --with-zlib \
  --with-bz2 \
  --with-gettext \
  --with-zlib-dir \
```

```
  --with-freetype \

  --with-jpeg \

  --with-webp \

  --with-xsl \

  --enable-soap \

  --enable-sockets \

  --enable-shmop \

  --enable-sysvmsg \

  --enable-sysvsem \

  --enable-sysvshm
```

输出如下符合预期

```
+---------------------------------------------------------------------+
| License:                                                            |
| This software is subject to the PHP License, available in this      |
| distribution in the file LICENSE. By continuing this installation   |
| process, you are bound by the terms of this license agreement.      |
| If you do not agree with the terms of this license, you must abort  |
| the installation process at this point.                             |
+---------------------------------------------------------------------+
```

- 安装：

```
make -j$(nproc) && make install
```

- 配置环境变量：

```
echo 'export PATH=/usr/local/php81/bin:$PATH' >> ~/.bashrc && source ~/.bashrc
```

- 复制配置文件模板：

```
cp php.ini-production /usr/local/php81/etc/php.ini

cp sapi/fpm/php-fpm.conf /usr/local/php81/etc/php-fpm.conf
```

```
cp /usr/local/php81/etc/php-fpm.d/www.conf.default
/usr/local/php81/etc/php-fpm.d/www.conf
```

- 创建系统服务：

```
echo '[Unit]
Description=The PHP FastCGI Process Manager
After=network.target

[Service]
Type=simple
PIDFile=/usr/local/php81/var/run/php-fpm.pid
ExecStart=/usr/local/php81/sbin/php-fpm --nodaemonize --fpm-config
/usr/local/php81/etc/php-fpm.conf
ExecReload=/bin/kill -USR2 $MAINPID

[Install]
WantedBy=multi-user.target' >
/usr/lib/systemd/system/php-fpm81.service
```

重载服务：

```
systemctl daemon-reload && systemctl enable --now php-fpm81
```

重启服务：

```
systemctl restart php-fpm81.service
```

添加 PHP 转发规则，并重启 nginx 服务：

```
systemctl restart nginx
sudo tee /etc/nginx/conf.d/default.conf <<-'EOF'
```

```
server {
    listen       81; #使用 81 端口临时测试，80 端口需备案后使用
    server_name  localhost;
    root /usr/share/nginx/html;
    location / {
            index index.php index.html index.htm;
    }
    location ~ .php$ {
            root /usr/share/nginx/html;
            fastcgi_pass 127.0.0.1:9000;
            fastcgi_index index.php;
            fastcgi_param SCRIPT_FILENAME
$document_root$fastcgi_script_name;
            include fastcgi_params;
    }
    error_page   500 502 503 504  /50x.html;
    location = /50x.html {
        root   /usr/share/nginx/html;
    }
}
EOF
```

2.配置 WordPress 数据库

2.1、为 WordPress 网站创建一个名称为 wordpress 的数据库，可以参考图形化的方式：创建数据库或使用下述命令行的方式：

- 连接数据库：

```
mysql -h 192.168.1.11 -P 13049 -u root -p --ssl-ca=ca.pem
```

- 创建数据库：

```
create database wordpress;
```

- 创建新用户 user，新用户密码自定义：

```
create user 'user'@'%' identified by 'XXXXX';
```

- 赋予用户对 wordpress 数据库的全部权限：

```
grant all privileges on wordpress.* to 'user'@'%';
```

- 使上述配置生效：

```
flush privileges;
```

- 退出 MySQL：

```
exit;
```

- 新建 test.php 文件，用于测试数据库连接的代码，并修改下述配置信息为你实际的配置信息

```
sudo tee /usr/share/nginx/html/test.php <<-'EOF'
<?php
error_reporting(E_ALL);
ini_set('display_errors', 1);


$servername = "192.168.1.11";  // 修改此处，填写数据库的 IP 地址
$username = "user"; //刚刚创建的 user 用户
$password = "XXXXX"; //user 用户的密码
$dbname = "wordpress"; //指定访问的库 wordpress
```

```
$conn = new mysqli($servername, $username, $password, $dbname, 13049);
// 13049 为你数据库的端口
if ($conn->connect_error) {
    die("Connection failed: " . $conn->connect_error);  // 输出具体错误
}
echo "Success!";
$conn->close();
?>
EOF
```

- 在本地物理机上使用浏览器访问 http://CT-ECS 实例公网 IP/test.php，返回 success 说明 PHP 代理设置成功并成功连接 MySQL 数据库。



Success!

3. 下载 WordPress

a. 进入 nginx 根目录，并下载 WordPress:

```
cd /usr/share/nginx/html && wget
https://cn.wordpress.org/wordpress-6.8.1-zh_CN.zip
```

b. 解压 WordPress 压缩包：

```
unzip wordpress-6.8.1-zh_CN.zip
```

c. 进入 WordPress 安装目录：

```
cd /usr/share/nginx/html/wordpress
```

d. 复制 wp-config-sample.php 文件：

```
cp wp-config-sample.php wp-config.php
```

4.修改 WordPress 配置文件

a.编辑 wp-config.php 文件：

```
vim wp-config.php
```

b.按 i 键进入编辑模式。修改下述关于 MySQL 部分的配置信息，如果有问题可以从 wp-config-sample.php 文件进行恢复。

```
/ database_name_here 修改为实际的数据库名称，本文示例创建为：wordpress
define( 'DB_NAME', 'database_name_here' );

/ username_here 修改为实际创建的用户名，本文示例创建为：user
define( 'DB_USER', 'username_here' );

/ password_here 修改为实际创建用户名的密码
define( 'DB_PASSWORD', 'password_here' );

/ localhost 修改为数据库的主机加端口，格式：IP 地址 + 端口号，本文示例创建为：192.168.1.11:13049
define( 'DB_HOST', 'localhost' );

/ utf8 修改为 utf8mb4
define( 'DB_CHARSET', 'utf8' );

/ '' 添加 utf8mb4_0900_ai_ci，如果你使用的非云数据库，请修改为 utf8mb4_unicode_ci
```

```
define( 'DB_COLLATE', '' );


/可选，配置 SSL 加密连接，把 /path/ca.pem 替换为实际的 CA 证书路径
/可以参考 步骤三：配置 MySQL 安全，设置 SSL 数据加密，下载数据库的证书
define('MYSQL_CLIENT_FLAGS', MYSQLI_CLIENT_SSL);
define('MYSQL_SSL_CA', '/path/ca.pem');
```

c.按 Esc 键，输入:wq 后按 Enter 键，保存退出配置文件。

5.安装并登录 WordPress 网站

a.浏览器访问“http://CT-ECS 实例公网 IP/wordpress”进入安装向导。



b.配置站点标题、管理员账号、密码和邮箱，并单击“安装 WordPress”。

参数	说明
站点标题	WordPress 网站名称。
用户名	WordPress 管理员名称。
密码	可以使用默认的密码或者自定义密码。 不建议重复使用现有密码，并确保将密码保存在安全的

参数	说明
	位置。
您的电子邮箱地址	接收通知的电子邮件地址。

c.安装成功。



d.安装成功，点击登录，或浏览器访问“http://CT-ECS 实例公网 IP/wordpress/wp-login.php，进入 WordPress。

6.购买域名为了便于网站的访问和使用，您可以给网站设置一个域名，使用域名访问网站。您可以在天翼云域名快速注册中购买域名。为了您网站的安全性起见，建议您的域名可以搭配 CT-WAF 一起使用，您可以参考：web 应用防火墙（原生版）-产品定义。

7.如网站未进行备案，您可以参考：备案基础知识-网站备案-备案介绍 - 天翼云（ctyun.cn）。

8.配置域名解析。配置域名解析后才能使用域名访问网站，具体操作请创建域名解析。

步骤五：配置服务器安全卫士（原生版）

- 服务器安全卫士（原生版）防护实践请参考：主机安全防护最佳实践，您可以根据您开通的对应防护，选择性的配置对应的安全设置，不同的防护版本

有不同的防护功能，您可以参考：产品规格。

- 配置基线检查，请参考：本节介绍服务器安全卫士（原生版）基线检测操作指南。
- 配置漏洞扫描，请参考：本节为您介绍漏洞扫描原理、漏洞扫描版本规格、漏洞等级。
- 配置弱口令检测，请参考：本节介绍服务器安全卫士（原生版）弱口令检测操作指南。
- 配置病毒查杀，请参考：病毒查杀概述。
- 配置文件防勒索，请参考：文件防勒索概述。
- 配置网页防篡改，请参考，购买网页防篡改配额。

步骤六：配置监控告警

- 如果您是使用云服务器部署数据库，建议您可以使用应用性能监控-产品定义，为您的应用健康保驾护航。
- 默认云主机和数据库都接入云监控服务，建议您配置明细的告警规则，请参考创建告警规则。并创建对应的告警模板。

总结

安全防御从来不局限于单点产品安全能力。安全防御需构建多维协同防护体系：通过全链路数据加密、零信任架构、操作系统加固、应用安全沙箱等技术形成纵深防御，实现动态风险阻断与持续安全运营。

10.5 典型网站类场景

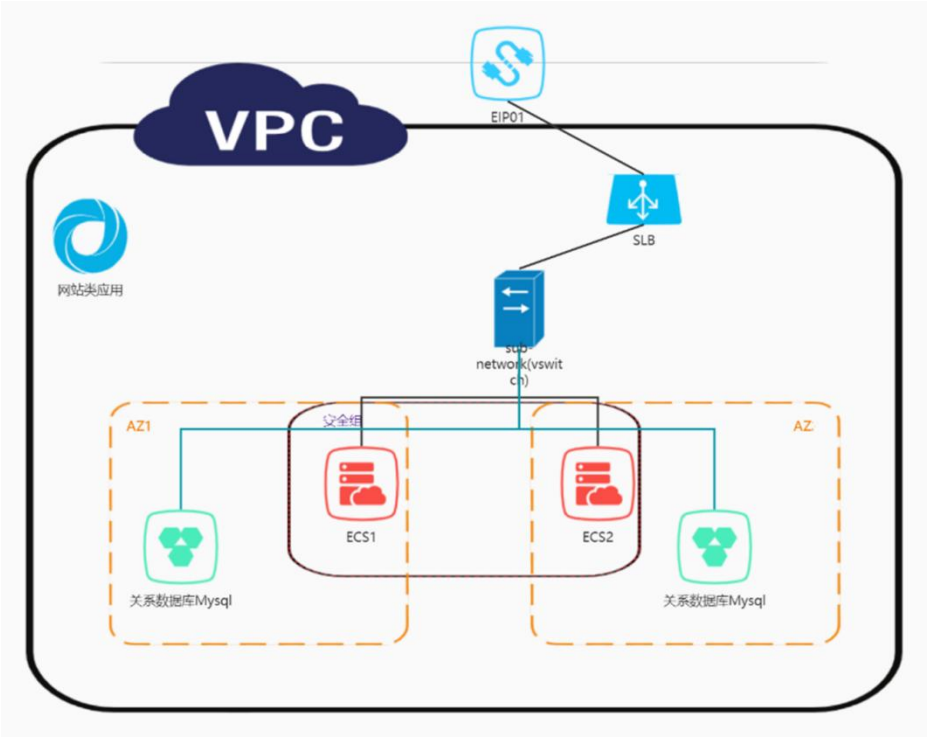
典型网站类场景

场景以 WordPress 网站高可用场景为例，在不同可用区部署前端云主机，并通过弹性负载均衡实例分摊业务流量，同时后端数据分离到关系数据库。

资源参数

产品	配置	可用区
虚拟私有云	VpcforWebsite	VPC 网段 192.168.0.0/16，子网 192.168.1.0/24
弹性云主机	2*s7.large.4（2 核 8G）；CentOS 7.6	分别位于可用区 1，可用区 2
关系型数据库 MySQL	主从 4 核 8G；MySQL 5.7；超高 IO	-
弹性负载均衡	SLBforWebsite	-
弹性 IP	5M 带宽	-

网站拓扑



操作步骤

1.进入控制中心，“网络”>“虚拟私有云”，点击创建虚拟私有云，VPC 网段设置为 192.168.0.0/16，子网网段设置为 192.168.0.0/24。

2.进入控制中心，“计算”>“弹性云主机”，点击创建云主机 ECS1，云主机规格选择 s7.large.4，操作系统设置为 CentOS 7.6，VPC 和子网选择步骤 1 所创建的 VPC 和子网。

3.进入控制中心，“网络”>“弹性 IP”，点击创建弹性 IP。

4.在云主机控制台或弹性 IP 控制台将弹性 IP 绑定至 ECS1。

5.使用控制台的远程登录功能，登录进入 ECS1，部署 WordPress。

1.

1）执行如下命令，安装 Apache 和 PHP 组件：

```
# 安装 Apache 和 PHP 组件

yum install -y httpd php php-mysql php-gd php-imap php-ldap php-odbc php-

pear php-xml php-xmlrpc# 启动 Apache 服务
```

```
service httpd start# 设置开机默认启动 Apache
sudo chkconfig httpd on
```

2）执行如下命令，创建 info.php 文件，验证 php 是否安装成功。

```
vim /var/www/html/info.php
```

3）执行如下命令，重启 httpd 服务。

```
service httpd restart
```

4)用可以访问公网的任何节点，访问 http://弹性 IP/info.php，可见运行成功，（请确保 PHP 所使用的端口，已在云主机所在安全组中放行）。

5）执行如下命令，卸载当前版本 PHP，安装高版本 PHP。

```
yum list installed | grep php | awk '{print $1}' | xargs yum remove -y
rpm -Uvh https://mirror.webtatic.com/yum/el7/epel-release.rpm
rpm -Uvh https://mirror.webtatic.com/yum/el7/webtatic-release.rpm
yum install -y php56w.x86_64 php56w-cli.x86_64 php56w-common.x86_64
php56w-gd.x86_64 php56w-imap.x86_64 php56w-ldap.x86_64 php56w-
mysql.x86_64 php56w-pdo.x86_64 php56w-odbc.x86_64 php56w-
process.x86_64 php56w-xml.x86_64 php56w-xmlrpc.x86_64
```

6）执行如下命令，重启 httpd 服务。

```
systemctl restart httpd
```

7）执行如下命令，下载并解压 WordPress 软件包。

```
mkdir -p /opt/WP
cd /opt/WP
wget https://cn.wordpress.org/latest-zh_CN.tar.gz
tar -xzvf latest-zh_CN.tar.gz
```

8）拷贝 wordpress 目录到/var/www/html/路径下，执行如下命令。

```
cd /var/www/html

cp -rf /opt/WP/wordpress/* /var/www/html/
```

9）进入控制中心，“数据库”>“关系数据库 MySQL 版”，创建 MySQL 数据库实例。

2.

10）ECS1 中安装 mysql 客户端，用于创建和访问 MySQL 数据库实例。

```
yum install mysql
```

11）连接 MySQL 数据库，创建数据库实例。

```
mysql -h<IP> -P<HOST> -u<user> -p<password>

create database <databasename> charset utf8
```

12）ECS1 中修改 wordpress 访问数据库配置。

```
cd /var/www/html/

vim wp-config.php
```

修改配置内容参考下面配置：

```
############// ** MySQL 设置 - 具体信息来自您正在使用的主机 ** ///** WordPress 数据库的名称 */

define('DB_NAME', 'db_wordpress');/** MySQL 数据库用户名 */

define('DB_USER', 'user_wordpress');/** MySQL 数据库密码 */

define('DB_PASSWORD', 'Hello1234');/** MySQL 主机 */

define('DB_HOST','<RDS 实例内网地址，不含端口号>');/** 创建数据表时默认的文字编码 */

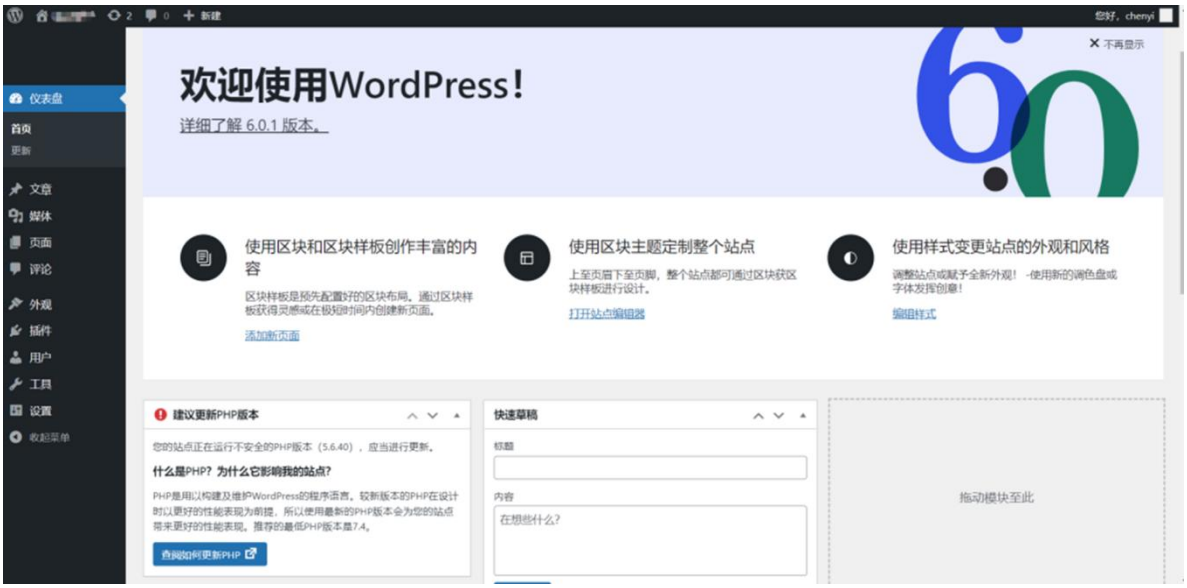
define('DB_CHARSET', 'utf8');/** 数据库整理类型。如不确定请勿更改 */
```

```
define('DB_COLLATE', '');

############
```

6.打开浏览器输入 http://弹性 IP:端口/wp-admin/install.php，访问 WordPress 站点。





参考步骤 5，申请 ECS2 并完成相应配置。配置 ECS2 的过程，可重复利用弹性 IP，挂载给 ECS2。

进入控制中心，“网络”>“弹性负载均衡”，创建弹性负载均衡，并绑定申请的弹性 IP。使用弹性负载均衡后，弹性 IP 将会从云主机上卸载下并关联到弹性负载均衡上。

执行以下命令，使用 ab 压测网站，并关闭其中 ECS，出现一定概率访问失败的现象。

```
ab. -n 10 -c 10 -t 30 -C 'cookies=xxx' -v 3 http://IP/wp-admin
```

10.6 镜像部署 windows 环境

简介

本节介绍如何使用天翼云镜像，通过重装系统部署 Windows 环境。如果您已经购买了弹性云主机，想切换成 Windows 系统，或者想使用镜像重新部署已经预装了其它软件的环境，可以参考本文档的介绍和操作指导。

重装系统的约束与限制可以参见重装操作系统。

操作步骤

1. 登录天翼云控制台。
2. 选择“计算>弹性云主机”。
3. 选中指定的云主机，将云主机进行关机操作。

如果云主机已经处于关机状态，则可以直接执行重装系统操作。重装系统后，系统盘数据将会丢失，如果想要保留某些数据，可以先进行备份。



4. 选择云主机列表操作栏中的“更多”选项，选择“一键重装”。



5. 如果需要批量云主机重装系统，勾选多个云主机实例，选择云主机列表上方的“更多”选项，选择“一键重装”。



6. 仔细阅读一键重装页面的提示信息，选择需要的 windows 镜像。镜像可以选择公共镜像、私有镜像、共享镜像、安全产品镜像。



7. 设置密码和用户数据。密码为必填项，用户数据可以暂不配置。点击“确定”按钮，即可进行重装系统的操作。

8. 可以在云主机列表看到重装云主机的状态。



9.重装完成后，云主机的状态会恢复成“关机”，此时云主机可以正常使用。

10.7 麒麟系统云主机配置图形化界面

操作场景

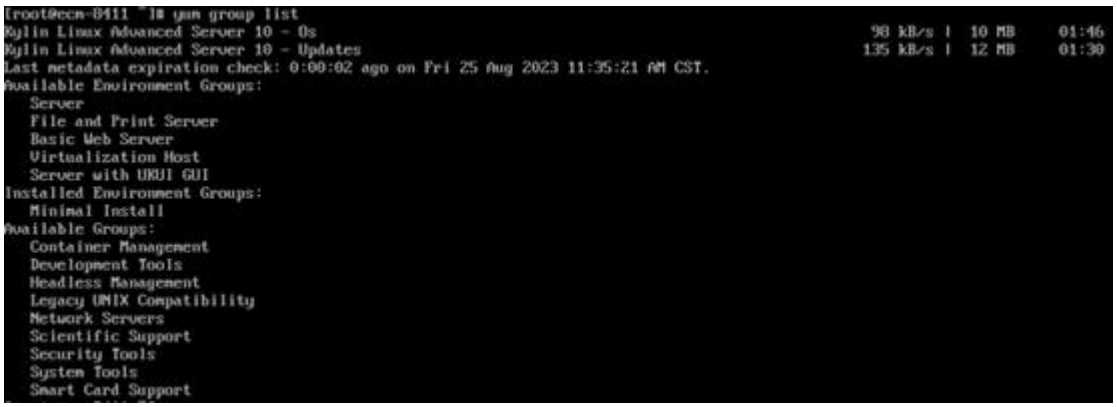
为了提供纯净的弹性云云主机系统给客户，麒麟系统云主机默认没有安装图形化界面，如果您需要图形化界面，请参见本节内容进行安装。

操作步骤

1.配置网络，确保 yum 源可以正常使用。银河麒麟云主机操作系统在有外网的环境下可以直接使用 yum 源，而在无外网的环境下需要挂载镜像作为本地源。

2.在字符界面执行以下命令，列出可安装的图形化包组。

```
#yum group list
```

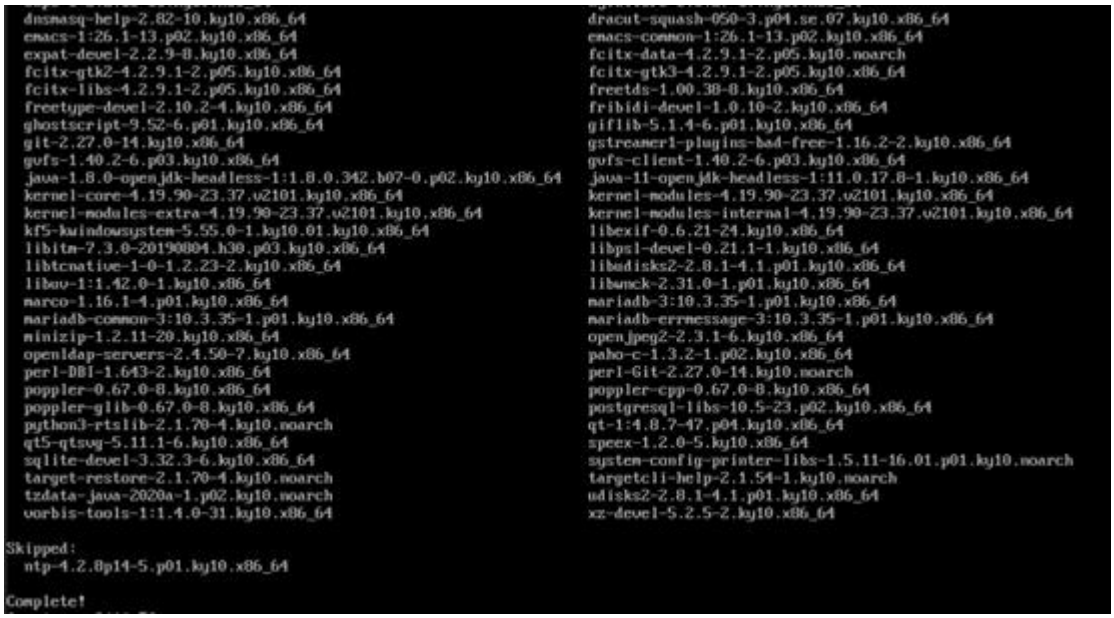


3.在字符界面执行以下命令，安装图形化桌面相关包。

```
#yum groupinstall -y "Server with UKUI GUI"
```

或者

```
#yum groupinstall -y "带 GUI 的云主机"
```



4. 在字符界面执行以下命令，设置系统默认的启动方式为 graphical.target，即图形化启动。

```
#systemctl set-default graphical.target
```



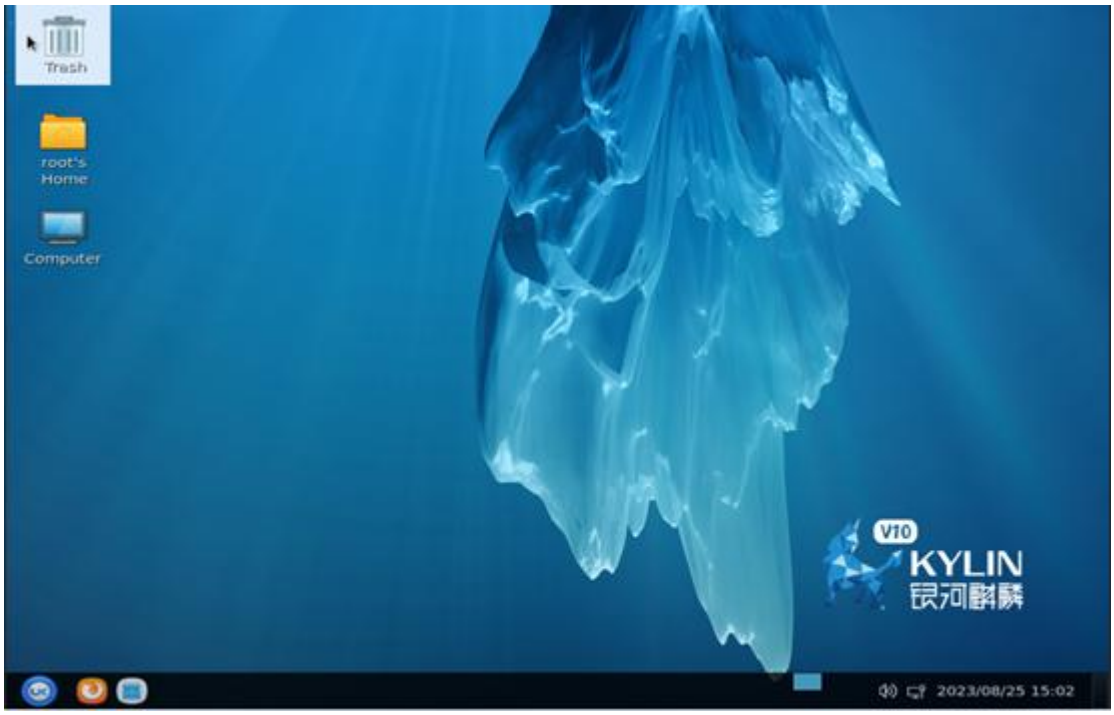
5. 在字符界面执行以下命令，重启系统使其生效。

```
#reboot
```

6. 重启系统后即可进入图形化界面。



7. 输入用户名与密码，进入系统。



10.8 下载并安装 SQL Server

SQL Server 各个版本的安装要求

SQL 版本	安装要求
SQL Server 2016，SQL Server 2017	SQL Server 2016 和 2017：硬件和软件要求
SQL Server 2019	SQL Server 2019：硬件和软件要求
SQL Server 2022	SQL Server 2022：硬件和软件要求

目前，Microsoft 官网仅提供 SQL Server 2016，SQL Server 2017，SQL Server 2019，SQL Server 2022 的下载。其余版本的 SQL Server 需要自行在网络上找资源下载。

SQL Server 2016 支持 Windows Server 2012，Windows Server 2016，Windows Server 2019 的操作系统，且不支持 Windows Server 2022 操作系统。更多的硬件和软件要求详情请查看上方表格链接。

SQL Server 2017 支持 Windows Server 2012，Windows Server 2016，Windows Server 2019 的操作系统，更多的硬件和软件要求详情请查看上方表格链接。

SQL Server 2019 支持 Windows Server 2016，Windows Server 2019 的操作系统，更多的硬件和软件要求详情请查看上方表格链接。

SQL Server 2022 支持 Windows Server 2016，Windows Server 2019 的操作系统，更多的硬件和软件要求详情请查看上方表格链接。

Windows Server 2008 的云主机请选用 SQL Server 2014 或更低的版本。

SQL Server 2016 安装步骤

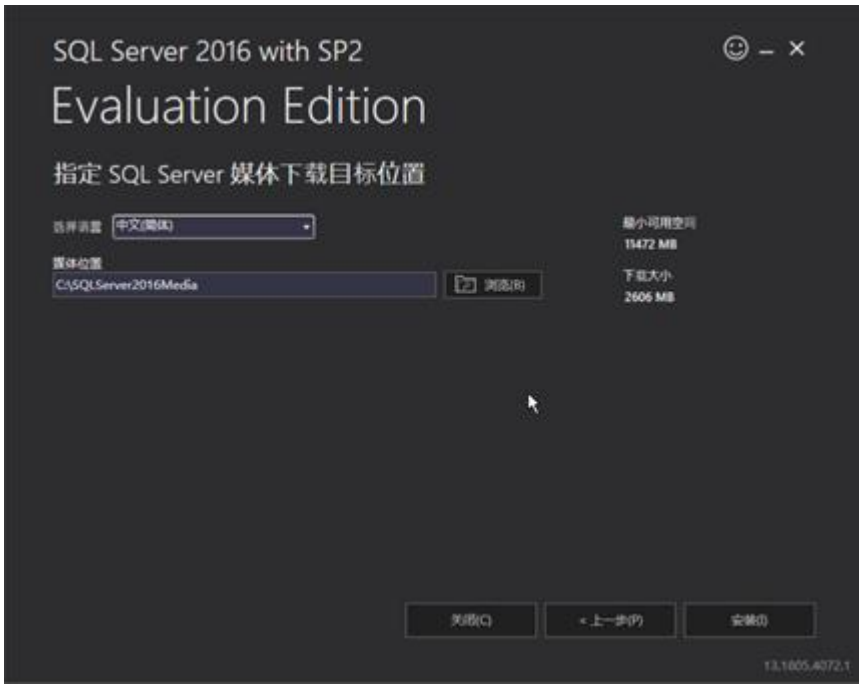
以下安装方法以 SQL Server 2016 安装为例，云主机操作系统为 Windows Server 2012 数据中心版 R2 64 位。

1. 点击下载好的“SQLServer2016-SSEI-Eval.exe”文件，进入 SQL 安装向导，进行安装操作。

2. 选择“自定义”安装，等待安装程序准备完毕。



3. 选择安装的语言，以及安装目录，点击“安装”。等待安装程序下载，下载完毕后自动进入 SQL Server 安装中心。

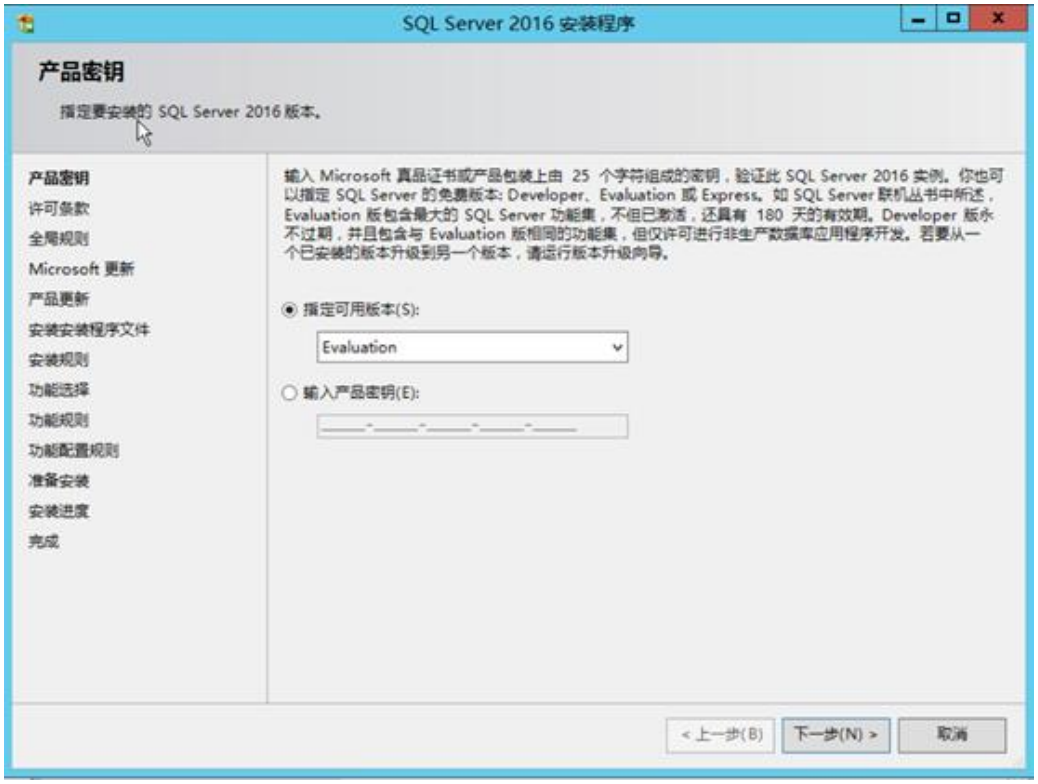




4. 点击左侧的“安装”导航栏，选择第一个选项，即“全新 SQL Server 独立安装或向现有安装添加功能”。



5. 选择“指定可用版本”，采用免费镜像模式进行安装。（若要正式搭建 Sharepoint，需要输入密钥安装 SQL Server 完整版本。）点击“下一步”。



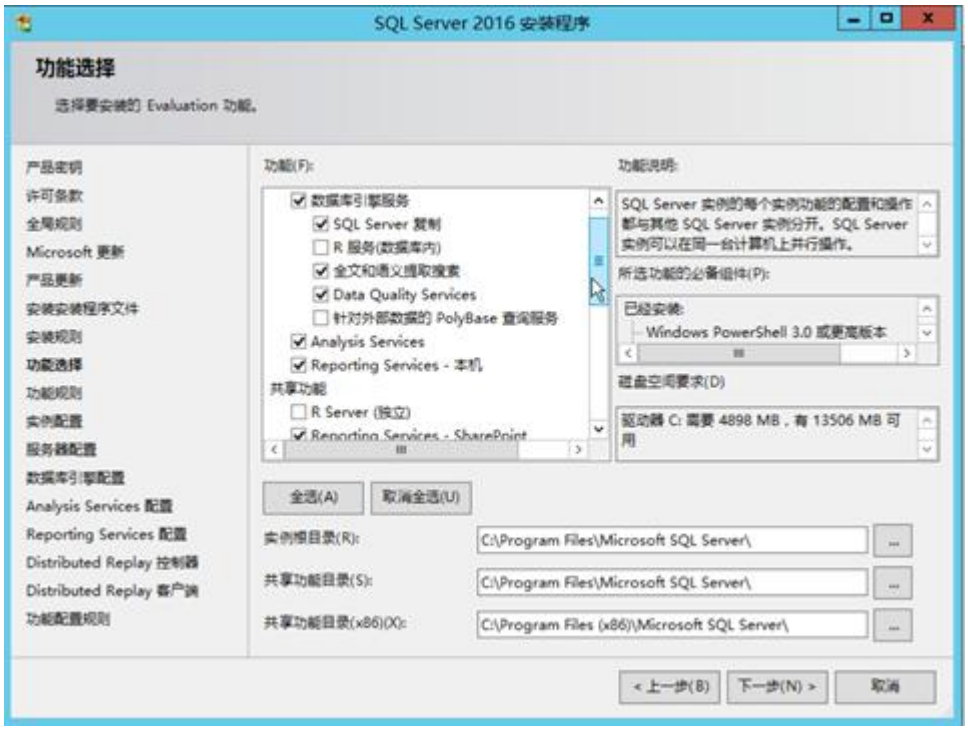
6. 勾选“我接受许可条款和隐私声明”。点击“下一步”。



7. 如果需要使用 Microsoft Update 更新则勾选。如果不需要则不用勾选。点击“下一步”。



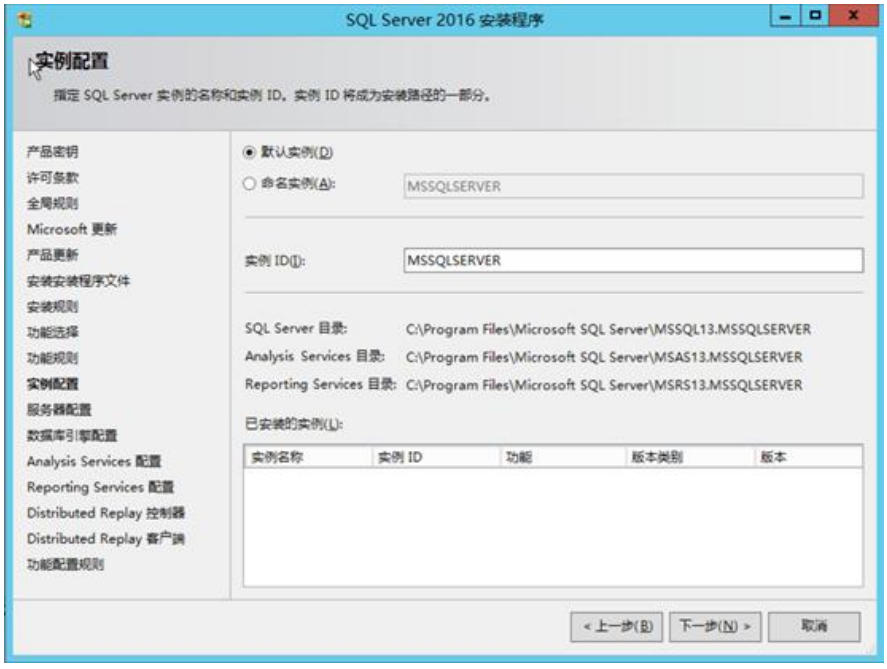
8.勾选“全选”，选择安装全部功能后，取消勾选“针对外部数据的 PolyBase 查询服务”、R 服务（数据库内）”以及“R Server（独立）”功能。该云主机为域控制器，PolyBase 无法在域控制器中使用，因此取消该选项。点击“下一步”。



9.勾选“默认实例”。使用默认值来配置实例，不需要任何修改。点击“下一步”。

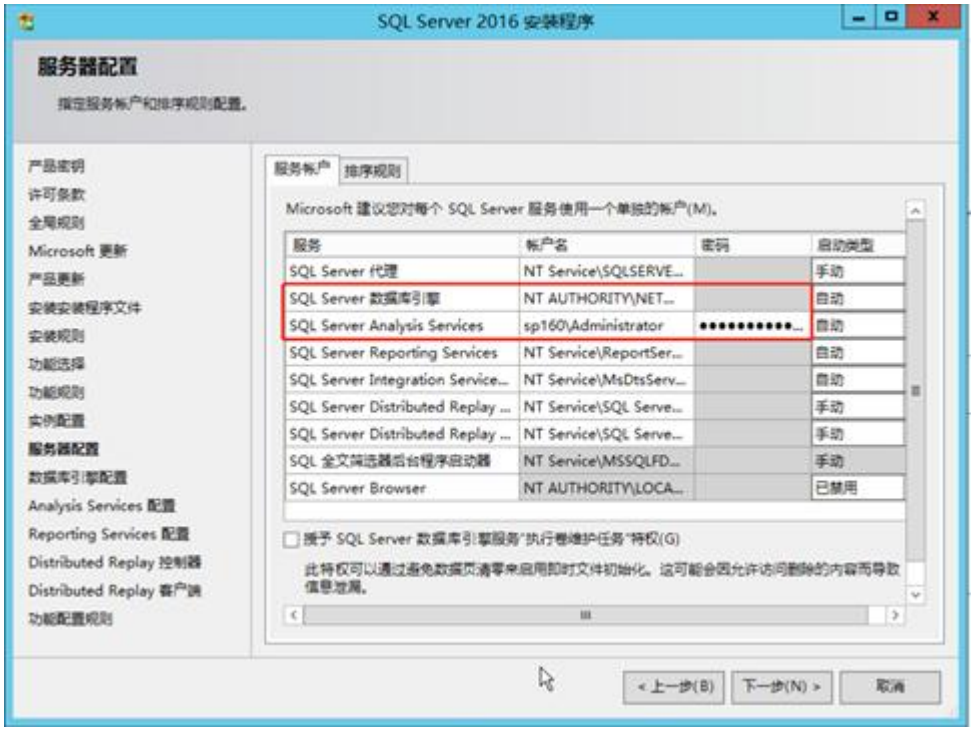
10.设置 SQL server Configuration。

修改“SQL Server 数据库引擎”的“账户名”为“NT AUTHORITY\NETWORK SERVICE”。

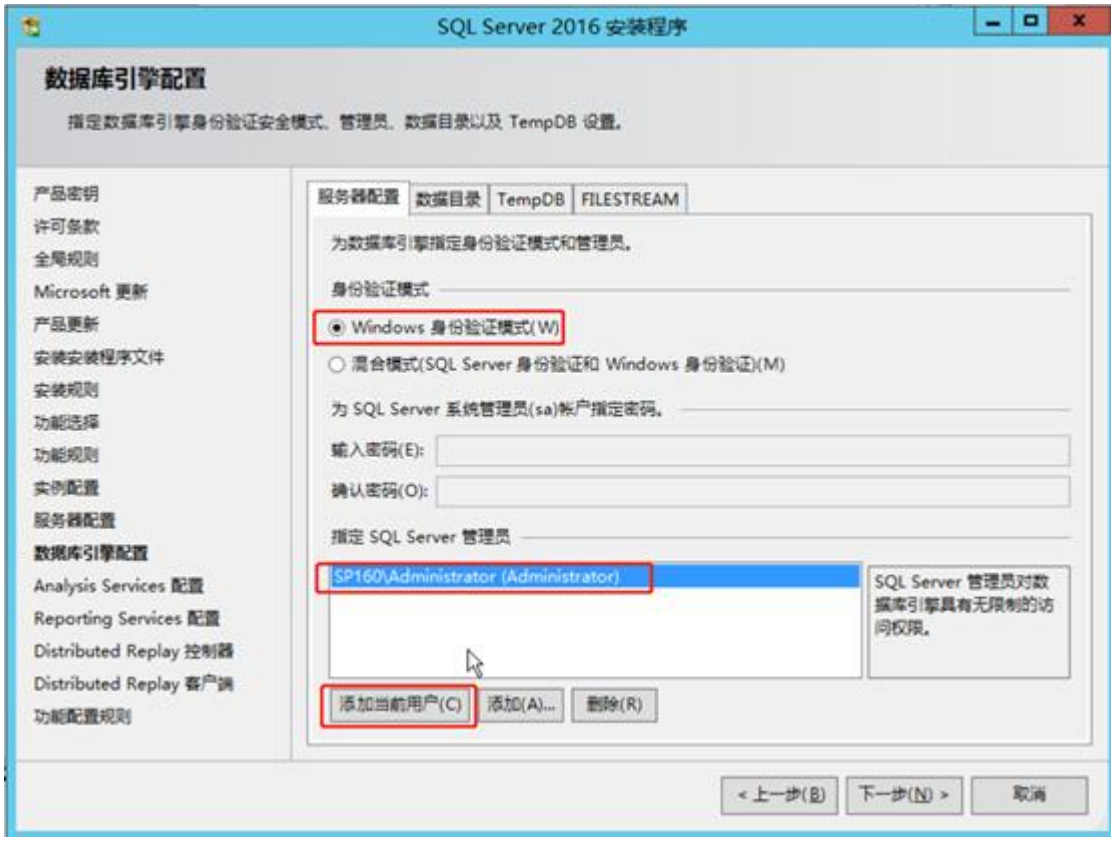


设置“SQL Server Analysis Services”为 Windows Server 添加 AD，DHCP，DNS，IIS 服务中设置的域账户和密码。详细步骤参见：为 Windows Server 添加 AD，DHCP，DNS，IIS 服务。

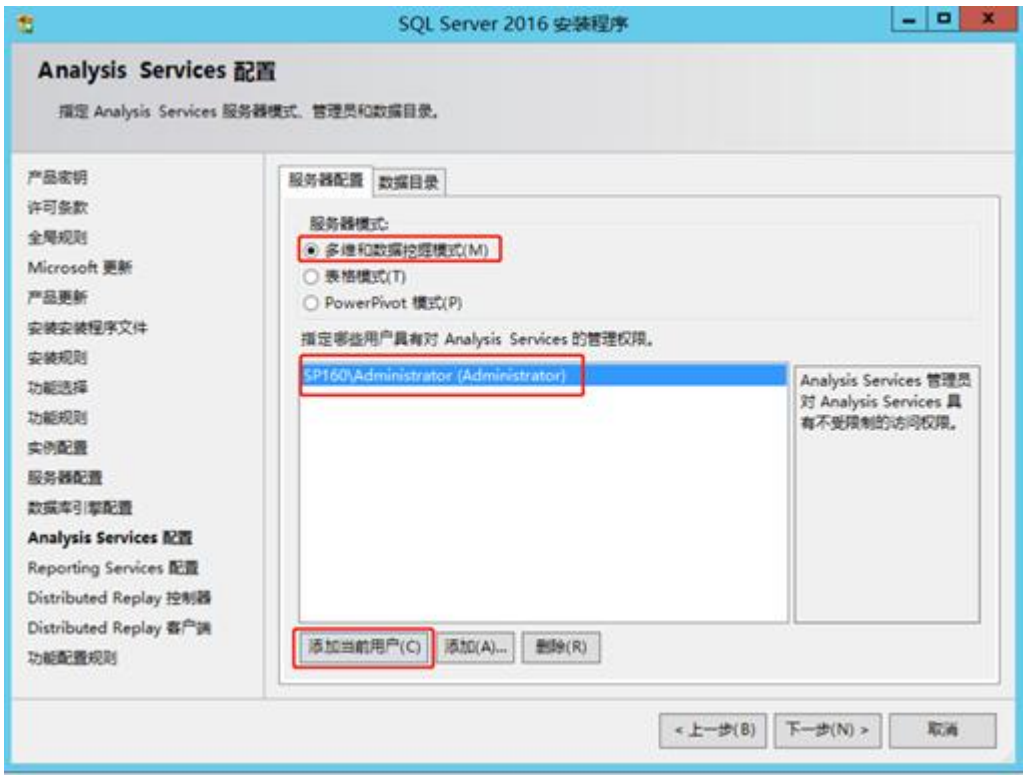
设置完毕之后点击“下一步”。



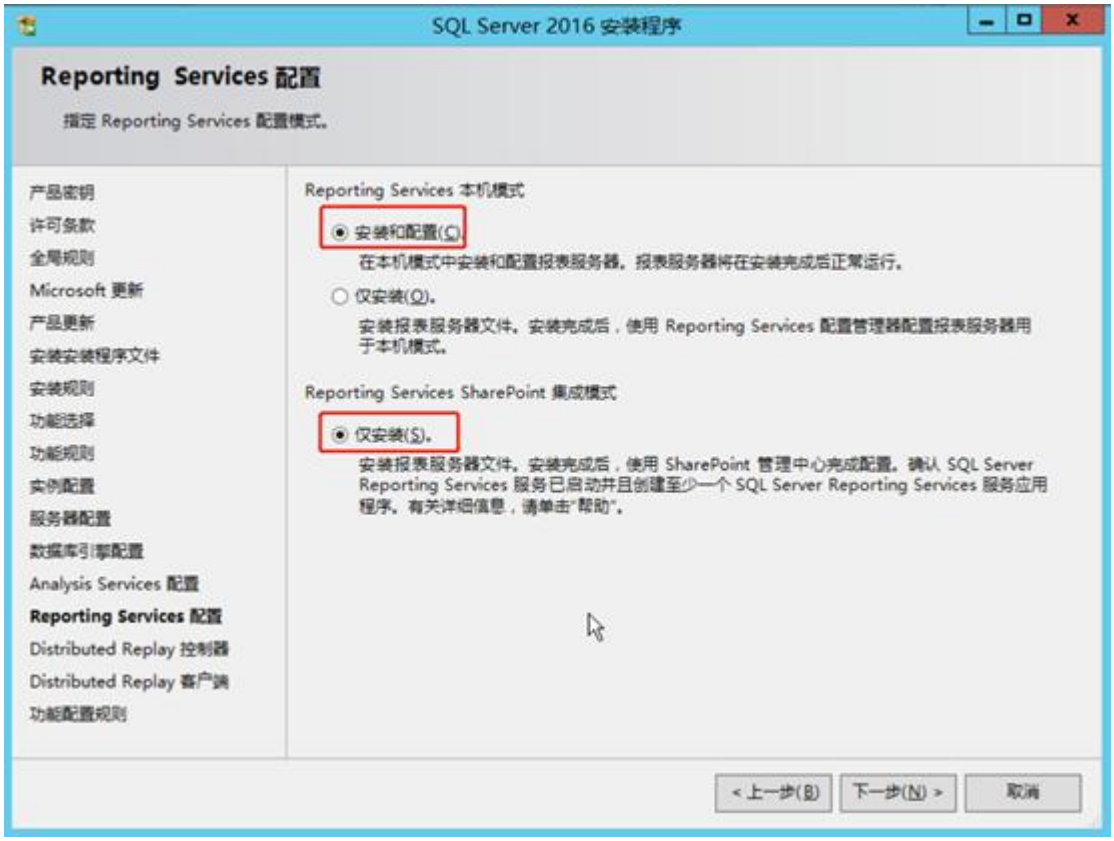
11. 勾选“Windows 身份验证模式”，点击“添加当前用户”，将当前账号作为 SQL Server 的管理员账号。点击“下一步”。



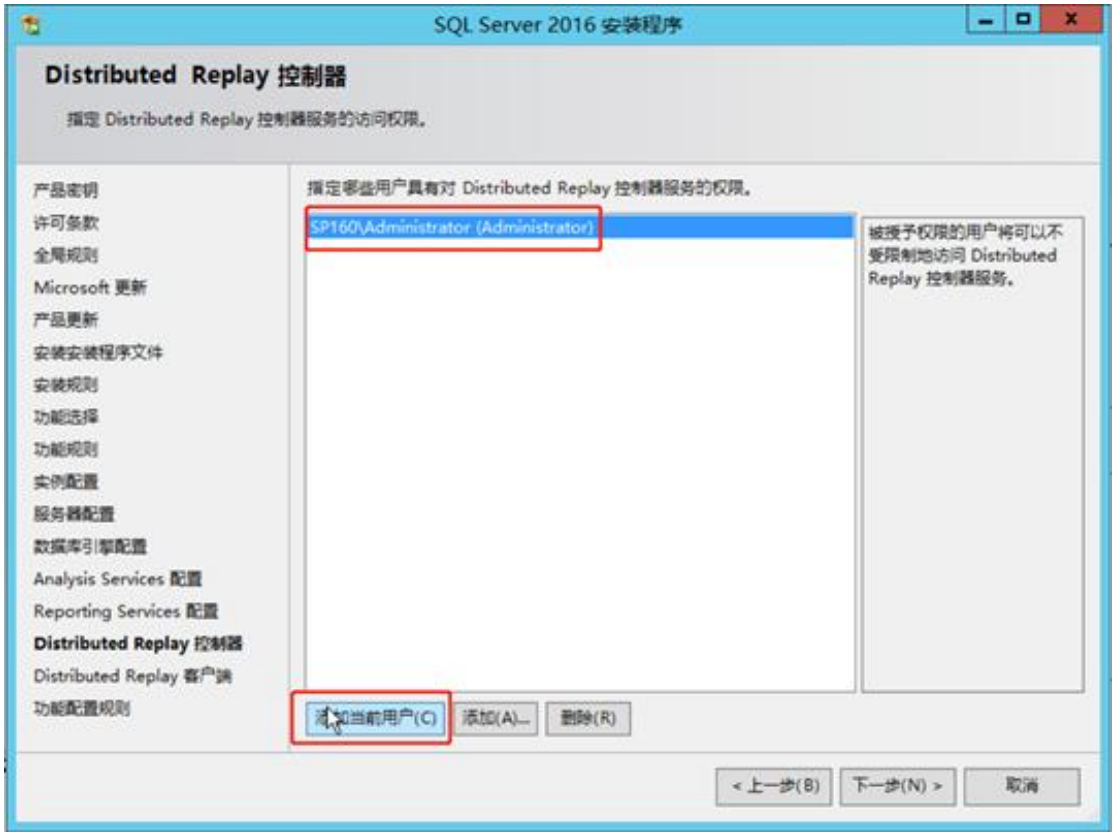
12. 勾选“多维和数据挖掘模式”。点击“添加当前用户”，为当前账号添加 Analysis Services 的管理员权限。点击“下一步”。



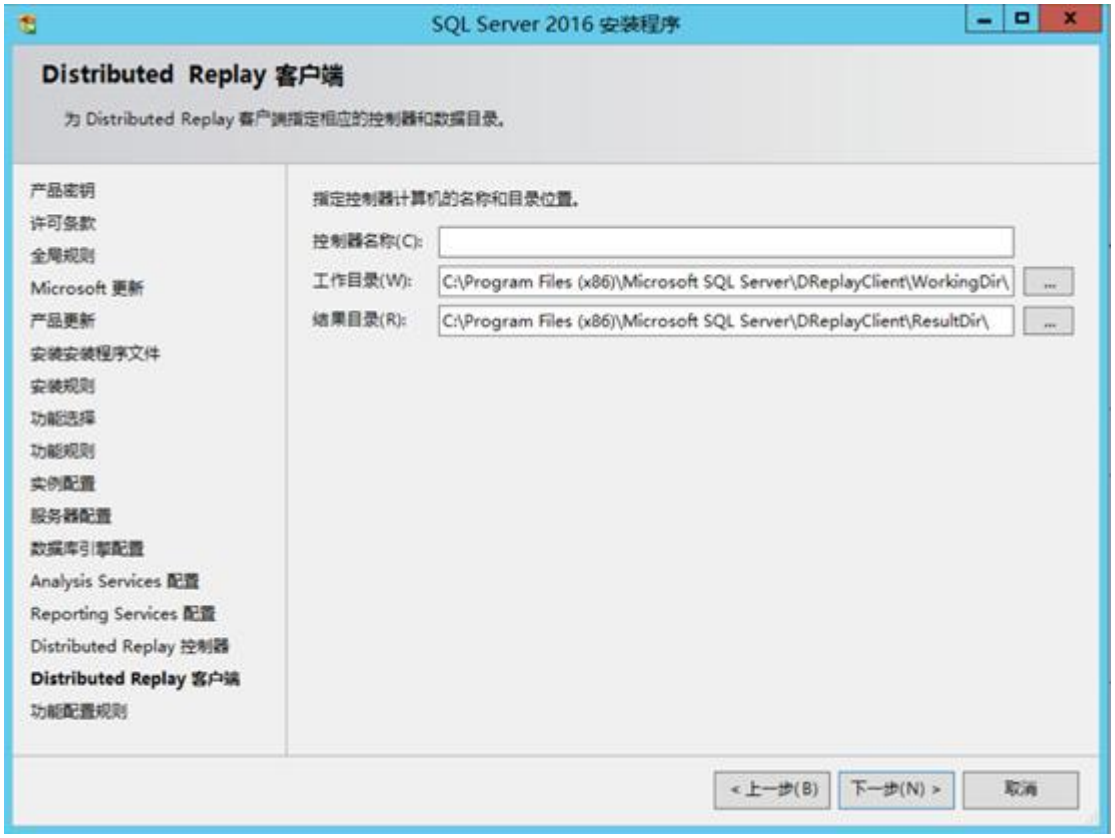
13. 设置 Reporting Services，在 Reporting Services 本机模式中，勾选“安装和配置”，在 Reporting Services SharePoint 集成模式下，勾选“仅安装”。单击“下一步”。



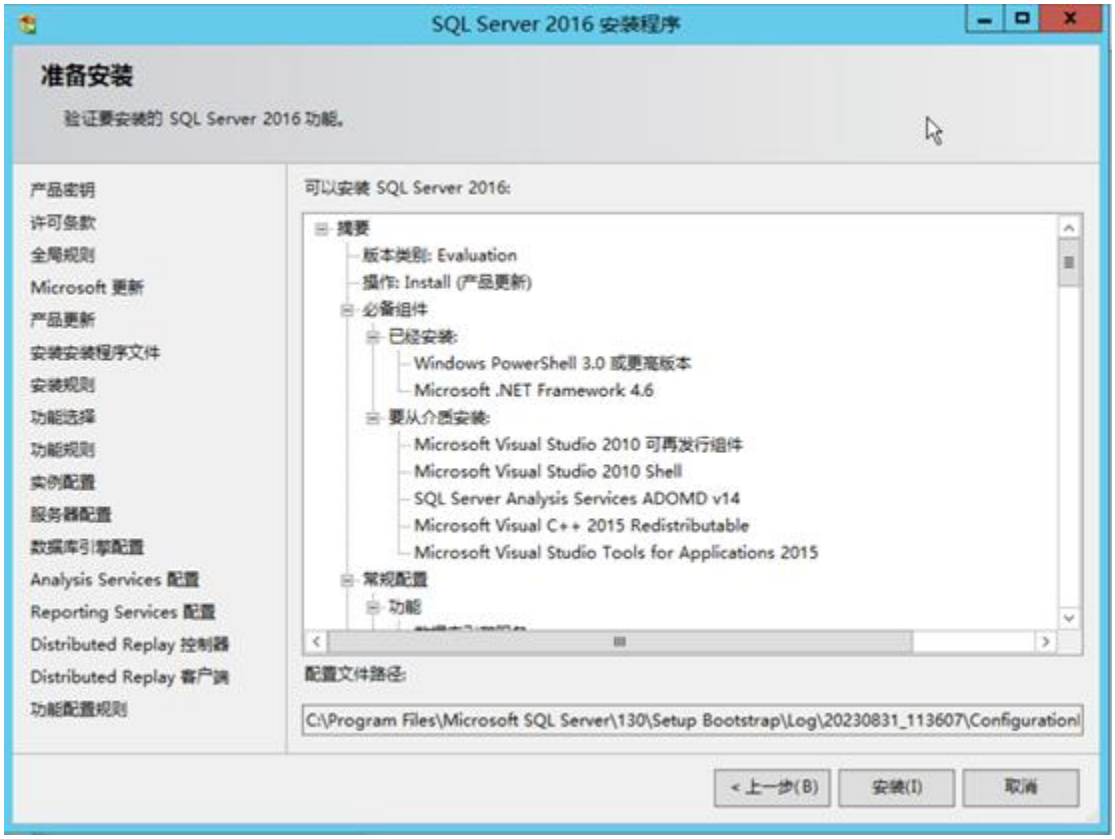
14. 点击“添加当前用户”，为当前账号添加 Distributed Replay 控制器的权限。点击“下一步”。



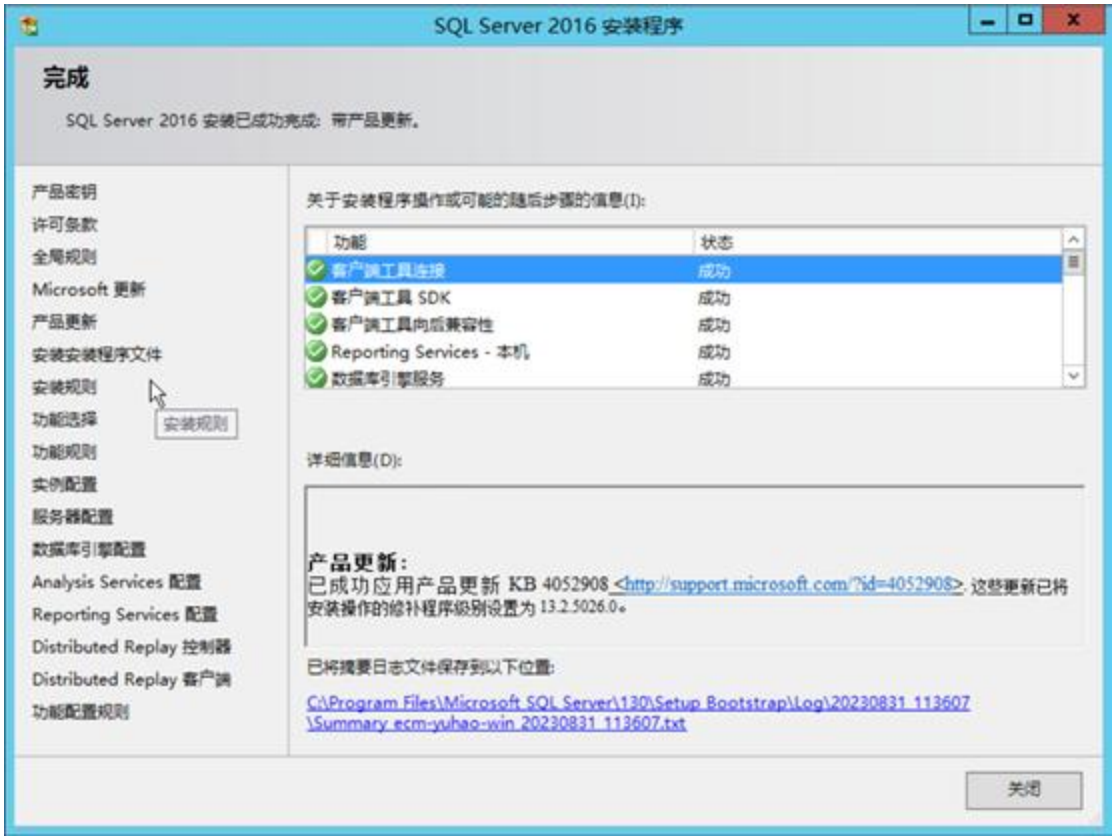
15. Distributed Replay 客户端设施使用默认值，不修改。点击“下一步”。



16. 确认 SQL Server 配置，点击“安装”开始安装。



17. 检查功能状态是否全部成功，如果全部成功，点击“关闭”，SQL Server 安装完成。



10.9Mysql 数据库自建，使用，问题排查最佳实践

创建 ECS 并自建 mysql 服务器

前提条件

1.已创建 ECS 虚机，请参考指定 ECS 规格创建实例创建 MySQL 客户端的弹性云服务器，如追求数据库 IO 读写性能，推荐使用本地盘云主机 Ip3。

2.已下载 MySQL 客户端安装包，MySQL 官网提供了针对不同操作系统的客户端安装包，单击此处下载 8.0 的最新版本，单击此处下载其他归档版本，用户需根据自己的操作系统类型选择符合要求的安装包。

mysql 部署

系统设置

确认系统及依赖库版本

本篇使用 OS 镜像版本为 CTyunOS 22.06.3

```
[root@localhost ~]# cat /etc/ctyunos-release

CTyunOS release 22.06.3



[root@localhost ~]# ldd --version

ldd (GNU libc) 2.28

Copyright (C) 2018 Free Software Foundation, Inc.

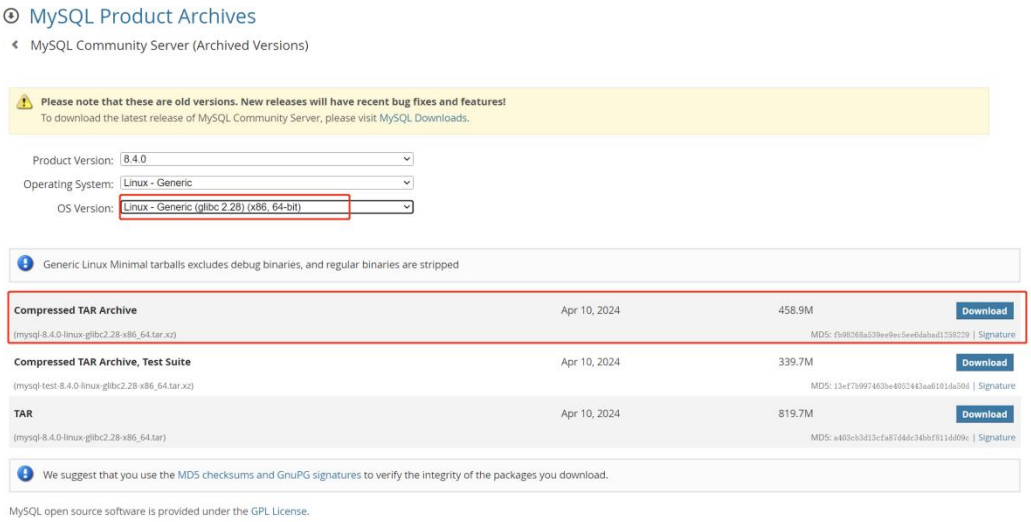
This is free software; see the source for copying conditions.  There is
NO

warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR
PURPOSE.

Written by Roland McGrath and Ulrich Drepper.
```

确认 mysql 版本

二进制搭建，mysql 对应的 OS Version 必须和操作系统依赖库版本保持一致，找到对应版本进行下载。



防火墙及 SELinux

```
[root@localhost ~]# systemctl disable firewalld --now #关闭并禁用防火墙


[root@localhost ~]# setenforce 0 #关闭 selinux

setenforce: SELinux is disabled


[root@localhost ~]# sed -i 's/^SELINUX=enforcing$/SELINUX=disabled/' /etc/selinux/config #配置持久化
```

创建用户及目录

```
[root@localhost ~]# groupadd mysql


[root@localhost ~]# useradd -r -g mysql -s /sbin/nologin mysql


[root@localhost ~]# mkdir -p /mysql/install/data
```

```
[root@localhost ~]# mkdir -p /mysql/install/tmp


[root@localhost ~]# mkdir -p /mysql/install/file


[root@localhost ~]# mkdir -p /mysql/install/log
```

创建 MySQL

上传并解压

```
[root@localhost install]# pwd
/mysql/install


[root@localhost install]# ls
data  file  log  mysql-8.4.0-linux-glibc2.28-x86_64.tar.xz  tmp


[root@localhost install]# tar -xf
mysql-8.4.0-linux-glibc2.28-x86_64.tar.xz


[root@localhost install]# ln -s mysql-8.4.0-linux-glibc2.28-x86_64
mysql  #创建软链接


[root@localhost install]# ll
total 459M
drwxr-xr-x 2 root root    6 Oct  8 17:04 data
drwxr-xr-x 2 root root    6 Oct  8 17:04 file
drwxr-xr-x 2 root root    6 Oct  8 17:04 log
```

```
lrwxrwxrwx 1 root root   34 Oct  8 17:08 mysql -&gt;
mysql-8.4.0-linux-glibc2.28-x86_64

drwxr-xr-x 9 root root  129 Oct  8 17:07
mysql-8.4.0-linux-glibc2.28-x86_64

-rw-r--r-- 1 root root 459M Oct  8 17:04
mysql-8.4.0-linux-glibc2.28-x86_64.tar.xz

drwxr-xr-x 2 root root    6 Oct  8 17:04 tmp
```

环境变量

```
[root@localhost install]# vi ~/.bash_profile


#PATH 增加路径 /mysql/install/mysql/bin

[root@localhost install]# cat ~/.bash_profile

# .bash_profile


# Get the aliases and functions

if [ -f ~/.bashrc ]; then

    . ~/.bashrc

fi


# User specific environment and startup programs


PATH=$PATH:$HOME/bin

PATH=$PATH:/mysql/install/mysql/bin:$HOME/bin

export PATH
```

```
[root@localhost install]# source ~/.bash_profile
```

初始化配置文件

```
#注意 bind_address 的 IP 地址和对应路径的修改

cat &lt;<eof> /mysql/install/my.cnf

[mysql]

no-beep

prompt="\u@mydb \R:\m:\s [\d]&gt; "

#no-auto-rehash

auto-rehash

default-character-set=utf8


[mysqld]

lower_case_table_names=1

server-id=3306

port=3306

user = mysql

bind_address= 192.168.2.134

basedir=/mysql/install/mysql

datadir=/mysql/install/data

#socket = /mysql/install/mysql.sock

socket = /tmp/mysql.sock

pid-file = /mysql/install/mysql.pid
```

```
character-set-server=utf8

autocommit = 0

#skip_name_resolve = 1

max_connections = 800

max_connect_errors = 1000

default-storage-engine=INNODB

transaction_isolation = READ-COMMITTED

explicit_defaults_for_timestamp = 1

sort_buffer_size = 32M

join_buffer_size = 128M

tmp_table_size = 72M

max_allowed_packet = 16M

#sql_mode =
"STRICT_TRANS_TABLES,NO_ENGINE_SUBSTITUTION,NO_ZERO_DATE,NO_ZERO_IN_D
ATE,ERROR_FOR_DIVISION_BY_ZERO,NO_AUTO_CREATE_USER"

interactive_timeout = 1800

wait_timeout = 1800

read_buffer_size = 16M

read_rnd_buffer_size = 32M


#query_cache_type = 1

#query_cache_size=1M

table_open_cache=2000

thread_cache_size=768
```

```
myisam_max_sort_file_size=10G

myisam_sort_buffer_size=135M

key_buffer_size=32M

read_buffer_size=8M

read_rnd_buffer_size=4M


back_log=1024

#flush_time=0

open_files_limit=65536

table_definition_cache=1400

#binlog_row_event_max_size=8K

#sync_master_info=10000

#sync_relay_log=10000

#sync_relay_log_info=10000


log-output=FILE

general_log = 0

general_log_file=/mysql/install/log/hiri-general.err

slow_query_log = ON

slow_query_log_file=/mysql/install/log/hiri-query.err

long_query_time=10

log-error=/mysql/install/log/hiri-error.err


log_queries_not_using_indexes = 1
```

```
log_slow_admin_statements = 1

log_slow_slave_statements = 1

log_throttle_queries_not_using_indexes = 10

#expire_logs_days = 90

#binlog_expire_logs_seconds=2592000

binlog_expire_logs_seconds=604800

min_examined_row_limit = 100


log_bin=/mysql/install/log/hiri-binlog

log_bin_index=/mysql/install/log/hiri-binlog.index

binlog_format='ROW'

binlog_rows_query_log_events=on


#master_info_repository = TABLE

#relay_log_info_repository = TABLE

#log_bin = bin.log

#sync_binlog = 1

#gtid_mode = on

#enforce_gtid_consistency = 1

#log_slave_updates

#binlog_format = row

#relay_log = relay.log

#relay_log_recovery = 1

#binlog_gtid_simple_recovery = 1
```

```
#slave_skip_errors = ddl_exist_errors


innodb_io_capacity = 4000

innodb_io_capacity_max = 8000

innodb_buffer_pool_size = 500M

innodb_buffer_pool_instances = 8

innodb_buffer_pool_load_at_startup = 1

innodb_buffer_pool_dump_at_shutdown = 1

innodb_lru_scan_depth = 2000

innodb_lock_wait_timeout = 5

#innodb_flush_method = O_DIRECT


innodb_log_file_size = 200M

innodb_log_files_in_group = 2

innodb_log_buffer_size = 16M


#innodb_undo_logs = 128

innodb_undo_tablespaces = 3

innodb_undo_log_truncate = 1

innodb_max_undo_log_size = 2G


innodb_flush_neighbors = 1

innodb_purge_threads = 4

#innodb_large_prefix = 1
```

```
innodb_thread_concurrency = 64

innodb_print_all_deadlocks = 1

innodb_strict_mode = 1

innodb_sort_buffer_size = 64M

innodb_flush_log_at_trx_commit=1

innodb_autoextend_increment=64

innodb_concurrency_tickets=5000

innodb_old_blocks_time=1000

innodb_open_files=65536

innodb_stats_on_metadata=0

innodb_file_per_table=1

innodb_checksum_algorithm=0

innodb_data_file_path=ibdata1:200M;ibdata2:200M;ibdata3:200M:autoexte
nd:max:5G

innodb_temp_data_file_path = ibtmp1:200M:autoextend:max:20G


innodb_buffer_pool_dump_pct = 40

innodb_page_cleaners = 4

innodb_purge_rseg_truncate_frequency = 128

binlog_gtid_simple_recovery=1

log_timestamps=system

#transaction_write_set_extraction=MURMUR32

#default_authentication_plugin=mysql_native_password  #mysql 8.4 版本
之后不允许这样配置了，使用 mysql_native_password=ON 可以正常使用
```

```
mysql_native_password=ON

#default_authentication_plugin=caching_sha2_password

EOF


[root@localhost install]# chown -R mysql.mysql /mysql #修改 mysql 目录的属组和属主


[root@localhost install]# ll

total 459M

drwxr-xr-x 2 mysql mysql    6 Oct  8 17:04 data

drwxr-xr-x 2 mysql mysql    6 Oct  8 17:04 file

drwxr-xr-x 2 mysql mysql    6 Oct  8 17:04 log

-rw-r--r-- 1 mysql mysql 3.5K Oct  8 17:13 my.cnf

lrwxrwxrwx 1 mysql mysql   34 Oct  8 17:08 mysql -&gt;
mysql-8.4.0-linux-glibc2.28-x86_64

drwxr-xr-x 9 mysql mysql  129 Oct  8 17:07
mysql-8.4.0-linux-glibc2.28-x86_64

-rw-r--r-- 1 mysql mysql 459M Oct  8 17:04
mysql-8.4.0-linux-glibc2.28-x86_64.tar.xz

drwxr-xr-x 2 mysql mysql    6 Oct  8 17:04 tmp
```

初始化数据库

```
[root@localhost install]# /mysql/install/mysql/bin/mysqld
--defaults-file=/mysql/install/my.cnf --initialize --user=mysql
--basedir=/mysql/install/mysql --datadir=/mysql/install/data
```

```
[root@localhost log]# grep 'temporary password'
/mysql/install/log/hiri-error.err
2024-10-08T17:34:43.831698+08:00 6 [Note] [MY-010454] [Server] A
temporary password is generated for root@localhost: ,H4xd+oyOM&gt;t
```

启库并修改密码

```
[root@localhost install]# /mysql/install/mysql/bin/mysqld_safe
--defaults-file=/mysql/install/my.cnf --datadir=/mysql/install/data
--pid-file=/mysql/install/mysql.pid &amp;
[1] 2152
[root@localhost install]# 2024-10-08T11:26:28.216260Z mysqld_safe
Logging to '/mysql/install/log/hiri-error.err'.
2024-10-08T11:26:28.470158Z mysqld_safe Starting mysqld daemon with
databases from /mysql/install/data


[root@localhost install]# mysqladmin -uroot -p',H4xd+oyOM&gt;t'
password 'redhat' -S /mysql/install/mysql.sock
mysqladmin: [Warning] Using a password on the command line interface can
be insecure.
Warning: Since password will be sent to server in plain text, use ssl
connection to ensure password safety.
```

配置远程登录

```
[root@localhost install]# ln -sf /mysql/install/mysql.sock
/tmp/mysql.sock


[root@localhost install]# mysql -uroot -predhat
```

```
mysql: [Warning] Using a password on the command line interface can be
insecure.
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 10
Server version: 8.4.0 MySQL Community Server - GPL

Copyright (c) 2000, 2024, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input
statement.

mysql&gt; show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
```

```
4 rows in set (0.02 sec)


mysql&gt; use mysql;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A


Database changed
mysql&gt; select Host,User from user;
+-----------+------------------+
| Host      | User             |
+-----------+------------------+
| localhost | mysql.infoschema |
| localhost | mysql.session    |
| localhost | mysql.sys        |
| localhost | root             |
+-----------+------------------+
4 rows in set (0.00 sec)


mysql&gt; create user 'root'@'%' identified by 'redhat'; #创建 root 用户并设置密码为 redhat
Query OK, 0 rows affected (0.09 sec)


mysql&gt; grant all privileges on *.* to 'root'@'%' with grant option; #赋予 root 用户所有权限
```

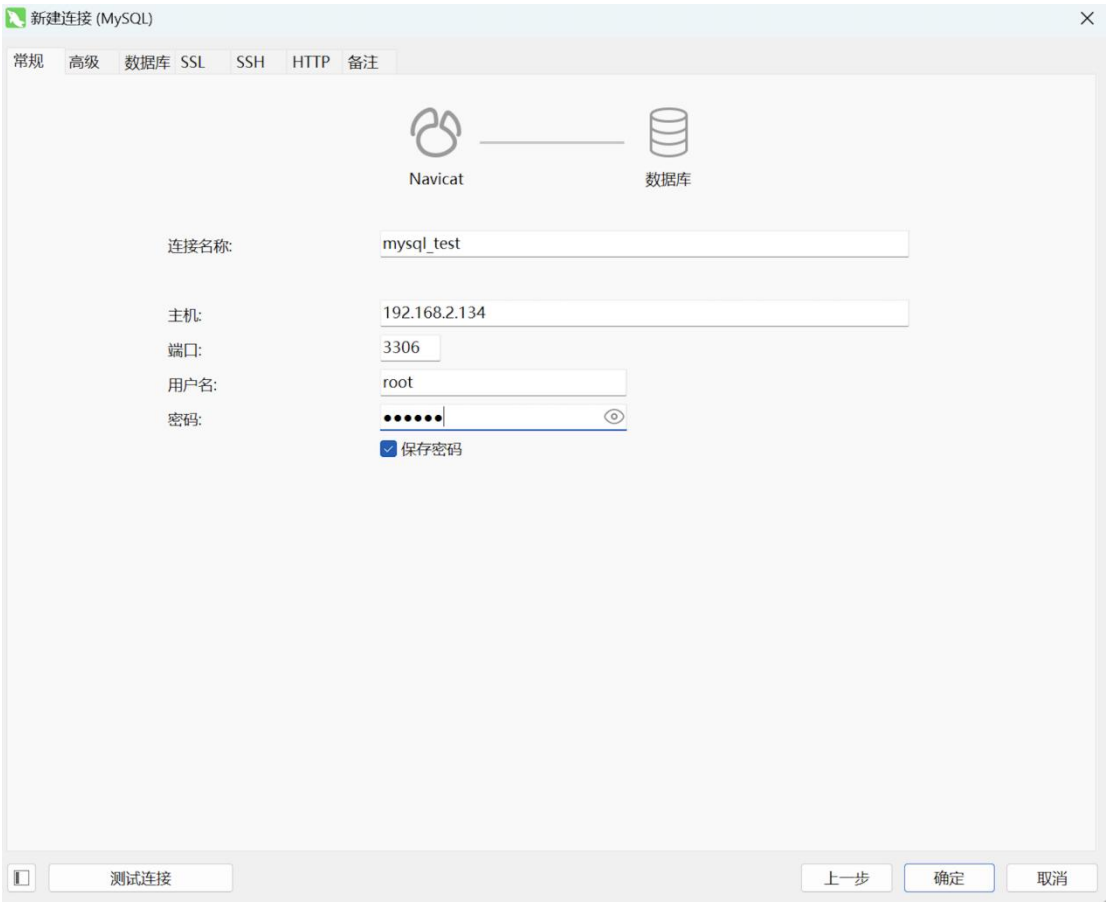
```
Query OK, 0 rows affected (0.03 sec)

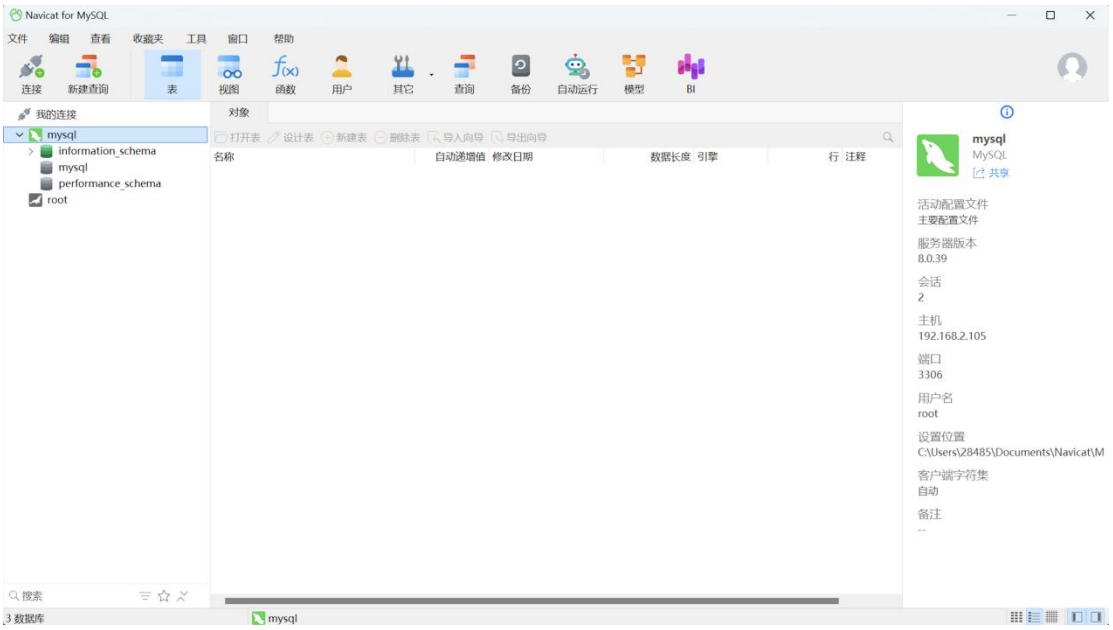

mysql&gt; flush privileges; #刷新权限使之生效
Query OK, 0 rows affected (0.01 sec)


mysql&gt; select Host,User from user;
+-----------+------------------+
| Host      | User             |
+-----------+------------------+
| %         | root             |
| localhost | mysql.infoschema |
| localhost | mysql.session    |
| localhost | mysql.sys        |
| localhost | root             |
+-----------+------------------+
5 rows in set (0.00 sec)


mysql&gt; exit
Bye
```

连接测试





配置启动服务

在/mysql/install/mysql/support-files 目录下，官方提供的有个脚本模板文件 mysql.server，可以拷贝一份起名叫 mysql 并修改脚本里面对应的路径和参数即可使用，这里已经改好的脚本内容如下：

天翼云
State Cloud

```
[root@localhost support-files]# vim mysql
#!/bin/sh
# Copyright Abandoned 1996 TCX DataKonsult AB &amp;amp; Monty Program KB &amp;amp;
Detron HB
# This file is public domain and comes with NO WARRANTY of any kind

# MySQL daemon start/stop script.

# Usually this is put in /etc/init.d (at least on machines SYSV R4 based
# systems) and linked to /etc/rc3.d/S99mysql and /etc/rc0.d/K01mysql.
# When this is done the mysql server will be started when the machine is
# started and shut down when the systems goes down.

# Comments to support chkconfig on RedHat Linux
# chkconfig: 2345 64 36
# description: A very fast and reliable SQL database engine.

# Comments to support LSB init script conventions
### BEGIN INIT INFO
# Provides: mysql
# Required-Start: $local_fs $network $remote_fs
# Should-Start: ypbind nscd ldap ntpd xntpd
# Required-Stop: $local_fs $network $remote_fs
# Default-Start:  2 3 4 5
```

```
# Default-Stop: 0 1 6
# Short-Description: start and stop MySQL
# Description: MySQL is a very fast and reliable SQL database engine.
### END INIT INFO

# If you install MySQL on some other places than /usr/local/mysql, then
you
# have to do one of the following things for this script to work:
#
# - Run this script from within the MySQL installation directory
# - Create a /etc/my.cnf file with the following information:
#   [mysqld]
#   basedir=<path-to-mysql-installation-directory>
# - Add the above to any other configuration file (for example ~/.my.ini)
#   and copy my_print_defaults to /usr/bin
# - Add the path to the mysql-installation-directory to the basedir
variable
#   below.
#
# If you want to affect other MySQL variables, you should make your changes
# in the /etc/my.cnf, ~/.my.cnf or other MySQL configuration files.

# If you change base dir, you must also change datadir. These may get
# overwritten by settings in the MySQL configuration files.
```

```
basedir=/mysql/install/mysql
datadir=/mysql/install/data

# Default value, in seconds, afterwhich the script should timeout waiting
# for server start.
# Value here is overriden by value in my.cnf.
# 0 means don't wait at all
# Negative numbers mean to wait indefinitely
service_startup_timeout=900

# Lock directory for RedHat / SuSE.
lockdir='/var/lock/subsys'
lock_file_path="$lockdir/mysql"

# The following variables are only set for letting mysql.server find
things.

# Set some defaults
mysqld_pid_file_path=/mysql/install/mysql.pid
if test -z "$basedir"
then
  basedir=/mysql/install/mysql
  bindir=/mysql/install/mysql/bin
```

```
  if test -z "$datadir"
  then
    datadir=/mysql/install/data
  fi
  sbindir=/mysql/install/mysql/bin
  libexecdir=/mysql/install/mysql/bin
else
  bindir="$basedir/bin"
  if test -z "$datadir"
  then
    datadir="$basedir/data"
  fi
  sbindir="$basedir/sbin"
  libexecdir="$basedir/libexec"
fi


# datadir_set is used to determine if datadir was set (and so should be
# *not* set inside of the --basedir= handler.)
datadir_set=


#
# Use LSB init script functions for printing messages, if possible
#
lsb_functions="/lib/lsb/init-functions"
```

```
if test -f $lsb_functions ; then
  . $lsb_functions
else
  log_success_msg()
  {
    echo " SUCCESS! $@"
  }
  log_failure_msg()
  {
    echo " ERROR! $@"
  }
fi

PATH="/sbin:/usr/sbin:/bin:/usr/bin:$basedir/bin"
export PATH

mode=$1    # start or stop

[ $# -ge 1 ] &amp;amp;&amp;amp; shift


other_args="$*"   # uncommon, but needed when called from an RPM upgrade
action
           # Expected: "--skip-networking --skip-grant-tables"
```

```
          # They are not checked here, intentionally, as it is the
resposibility
          # of the "spec" file author to give correct arguments only.

case `echo "testing\c"`,`echo -n testing` in
    *c*,-n*) echo_n=   echo_c=     ;;
    *c*,*)   echo_n=-n echo_c=     ;;
    *)       echo_n=   echo_c='\c' ;;
esac

parse_server_arguments() {
  for arg do
    case "$arg" in
      --basedir=*)  basedir=`echo "$arg" | sed -e 's/^[^=]*=//'`
                    bindir="$basedir/bin"
            if test -z "$datadir_set"; then
              datadir="$basedir/data"
            fi
            sbindir="$basedir/sbin"
            libexecdir="$basedir/libexec"
        ;;
      --datadir=*)  datadir=`echo "$arg" | sed -e 's/^[^=]*=//'`
            datadir_set=1
    ;;
```

```
      --pid-file=*) mysqld_pid_file_path=`echo "$arg" | sed -e
's/^[^=]*=//'` ;;
      --service-startup-timeout=*) service_startup_timeout=`echo
"$arg" | sed -e 's/^[^=]*=//'` ;;
    esac
  done
}

wait_for_pid () {
  verb="$1"           # created | removed
  pid="$2"            # process ID of the program operating on the
pid-file
  pid_file_path="$3" # path to the PID file.

  i=0
  avoid_race_condition="by checking again"

  while test $i -ne $service_startup_timeout ; do

    case "$verb" in
      'created')
        # wait for a PID-file to pop into existence.
        test -s "$pid_file_path" &amp;&amp; i='' &amp;&amp; break
        ;;
```

```
      'removed')
        # wait for this PID-file to disappear
        test ! -s "$pid_file_path" &amp;&amp; i='' &amp;&amp; break
        ;;
      *)
        echo "wait_for_pid () usage: wait_for_pid created|removed pid
pid_file_path"
        exit 1
        ;;
    esac

    # if server isn't running, then pid-file will never be updated
    if test -n "$pid"; then
      if kill -0 "$pid" 2&gt;/dev/null; then
        :  # the server still runs
      else
        # The server may have exited between the last pid-file check and
now.
        if test -n "$avoid_race_condition"; then
          avoid_race_condition=""
          continue  # Check again.
        fi

        # there's nothing that will affect the file.
```

天翼云
State Cloud

```
        log_failure_msg "The server quit without updating PID file
($pid_file_path)."
        return 1  # not waiting any more.
      fi
    fi

    echo $echo_n ".$echo_c"
    i=`expr $i + 1`
    sleep 1

  done

  if test -z "$i" ; then
    log_success_msg
    return 0
  else
    log_failure_msg
    return 1
  fi
}

# Get arguments from the my.cnf file,
# the only group, which is read from now on is [mysqld]
if test -x "$bindir/my_print_defaults";  then
```

```
  print_defaults="$bindir/my_print_defaults"
else
  # Try to find basedir in /etc/my.cnf
  conf=/mysql/install/my.cnf
  print_defaults=
  if test -r $conf
  then
    subpat='^[^=]*basedir[^=]*=\(.*\)$'
    dirs=`sed -e "/$subpat/!d" -e 's//\1/' $conf`
    for d in $dirs
    do
      d=`echo $d | sed -e 's/[     ]//g'`
      if test -x "$d/bin/my_print_defaults"
      then
        print_defaults="$d/bin/my_print_defaults"
        break
      fi
    done
  fi

  # Hope it's in the PATH ... but I doubt it
  test -z "$print_defaults" &amp;&amp;
print_defaults="my_print_defaults"
fi
```

```
#
# Read defaults file from 'basedir'.   If there is no defaults file there
# check if it's in the old (depricated) place (datadir) and read it from
there
#

extra_args=""
if test -r "/mysql/install/my.cnf"
then
  extra_args="-e /mysql/install/my.cnf"
fi

parse_server_arguments `$print_defaults $extra_args mysqld server
mysql_server mysql.server`

#
# Set pid file if not given
#
if test -z "$mysqld_pid_file_path"
then
  mysqld_pid_file_path=$datadir/`hostname`.pid
else
  case "$mysqld_pid_file_path" in
```

```
    /* ) ;;
    * )  mysqld_pid_file_path="$datadir/$mysqld_pid_file_path" ;;
  esac
fi


case "$mode" in
  'start')
    # Start daemon

    # Safeguard (relative paths, core dumps..)
    cd $basedir

    echo $echo_n "Starting MySQL"
    if test -x $bindir/mysqld_safe
    then
      # Give extra arguments to mysqld with the my.cnf file. This script
      # may be overwritten at next upgrade.
      $bindir/mysqld_safe --defaults-file=/mysql/install/my.cnf
--datadir="$datadir" --pid-file="$mysqld_pid_file_path" $other_args
&gt;/dev/null &amp;
      wait_for_pid created "$!" "$mysqld_pid_file_path";
return_value=$?

      # Make lock for RedHat / SuSE
```

```
      if test -w "$lockdir"
      then
        touch "$lock_file_path"
      fi

      exit $return_value
    else
      log_failure_msg "Couldn't find MySQL server
($bindir/mysqld_safe)"
    fi
    ;;

  'stop')
    # Stop daemon. We use a signal here to avoid having to know the
    # root password.

    if test -s "$mysqld_pid_file_path"
    then
      # signal mysqld_safe that it needs to stop
      touch "$mysqld_pid_file_path.shutdown"

      mysqld_pid=`cat "$mysqld_pid_file_path"`

      if (kill -0 $mysqld_pid 2&gt;/dev/null)
```

```
      then
        echo $echo_n "Shutting down MySQL"
        kill $mysqld_pid
        # mysqld should remove the pid file when it exits, so wait for
it.
        wait_for_pid removed "$mysqld_pid" "$mysqld_pid_file_path";
return_value=$?
      else
        log_failure_msg "MySQL server process #$mysqld_pid is not
running!"
        rm "$mysqld_pid_file_path"
      fi

      # Delete lock for RedHat / SuSE
      if test -f "$lock_file_path"
      then
        rm -f "$lock_file_path"
      fi
      exit $return_value
    else
      log_failure_msg "MySQL server PID file could not be found!"
    fi
    ;;
```

天翼云
State Cloud

```
  'restart')
    # Stop the service and regardless of whether it was
    # running or not, start it again.
    if $0 stop  $other_args; then
      $0 start $other_args
    else
      log_failure_msg "Failed to stop running server, so refusing to try
to start."
      exit 1
    fi
    ;;

  'reload'|'force-reload')
    if test -s "$mysqld_pid_file_path" ; then
      read mysqld_pid &lt;  "$mysqld_pid_file_path"
      kill -HUP $mysqld_pid &amp;&amp; log_success_msg "Reloading
service MySQL"
      touch "$mysqld_pid_file_path"
    else
      log_failure_msg "MySQL PID file could not be found!"
      exit 1
    fi
    ;;
  'status')
```

天翼云
State Cloud

```
    # First, check to see if pid file exists
    if test -s "$mysqld_pid_file_path" ; then
      read mysqld_pid &lt; "$mysqld_pid_file_path"
      if kill -0 $mysqld_pid 2&gt;/dev/null ; then
        log_success_msg "MySQL running ($mysqld_pid)"
        exit 0
      else
        log_failure_msg "MySQL is not running, but PID file exists"
        exit 1
      fi
    else
      # Try to find appropriate mysqld process
      mysqld_pid=`pidof $libexecdir/mysqld`

      # test if multiple pids exist
      pid_count=`echo $mysqld_pid | wc -w`
      if test $pid_count -gt 1 ; then
        log_failure_msg "Multiple MySQL running but PID file could not
be found ($mysqld_pid)"
        exit 5
      elif test -z $mysqld_pid ; then
        if test -f "$lock_file_path" ; then
          log_failure_msg "MySQL is not running, but lock file
($lock_file_path) exists"
```

```
          exit 2
        fi
        log_failure_msg "MySQL is not running"
        exit 3
      else
        log_failure_msg "MySQL is running but PID file could not be found"
        exit 4
      fi
    fi
    ;;
    *)
      # usage
      basename=`basename "$0"`
      echo "Usage: $basename
{start|stop|restart|reload|force-reload|status}  [ MySQL server
options ]"
      exit 1
    ;;
esac


exit 0


[root@localhost support-files]# pwd
/mysql/install/mysql/support-files
```

```
[root@localhost support-files]# chown mysql.mysql mysql

[root@localhost support-files]# chmod 755 mysql

[root@localhost support-files]# ll
total 32K
-rwxr-xr-x 1 mysql mysql  11K Oct  8 19:44 mysql
-rwxr-xr-x 1 mysql mysql 1.1K Apr 10 14:26 mysqld_multi.server
-rw-r--r-- 1 mysql mysql 2.1K Apr 10 15:58 mysql-log-rotate
-rwxr-xr-x 1 mysql mysql  11K Apr 10 15:58 mysql.server
[root@localhost support-files]# ./mysql status
 SUCCESS! MySQL running (3305)

[root@localhost support-files]# ./mysql stop
Shutting down MySQL.....2024-10-08T11:50:53.760574Z mysqld_safe mysqld
from pid file /mysql/install/mysql.pid ended
 SUCCESS!
[1]+  Done                    /mysql/install/mysql/bin/mysqld_safe
--defaults-file=/mysql/install/my.cnf --datadir=/mysql/install/data
--pid-file=/mysql/install/mysql.pid  (wd: /mysql/install)
(wd now: /mysql/install/mysql/support-files)

[root@localhost support-files]# ./mysql start
```

```
Starting MySQL.............. SUCCESS!


[root@localhost support-files]# ./mysql status

 SUCCESS! MySQL running (4653)
```

使用 systemd 启动服务器(可选)

```
[root@localhost support-files]# ./mysql stop

Shutting down MySQL.. SUCCESS!


[root@localhost support-files]#  cd /usr/lib/systemd/system


[root@localhost system]# touch mysqld.service


[root@localhost system]# chmod 644 mysqld.service


[root@localhost system]# vi mysqld.service

[root@localhost system]# cat mysqld.service

[Unit]

Description=MySQL Server

Documentation=man:mysqld(8)

Documentation=http://dev.mysql.com/doc/refman/en/using-systemd.html

After=network.target

After=syslog.target


[Install]
```

```
WantedBy=multi-user.target

[Service]
User=mysql
Group=mysql

# Have mysqld write its state to the systemd notify socket
Type=notify

# Disable service start and stop timeout logic of systemd for mysqld service.
TimeoutSec=0

# Start main service
ExecStart=/mysql/install/mysql/bin/mysqld --defaults-file=/mysql/install/my.cnf $MYSQLD_OPTS

# Use this to switch malloc implementation
EnvironmentFile=-/etc/sysconfig/mysql

# Sets open_files_limit
LimitNOFILE = 10000

Restart=on-failure
```

```
RestartPreventExitStatus=1

# Set environment variable MYSQLD_PARENT_PID. This is required for
restart.
Environment=MYSQLD_PARENT_PID=1

PrivateTmp=false

[root@localhost system]# systemctl enable mysqld.service
Created symlink
/etc/systemd/system/multi-user.target.wants/mysqld.service →
/usr/lib/systemd/system/mysqld.service.

[root@localhost system]# systemctl start mysqld.service

[root@localhost system]# systemctl status mysqld
● mysqld.service - MySQL Server
   Loaded: loaded (/usr/lib/systemd/system/mysqld.service; enabled;
vendor preset: disabled)
   Active: active (running) since Tue 2024-10-08 19:53:18 CST; 1s ago
     Docs: man:mysqld(8)
           http://dev.mysql.com/doc/refman/en/using-systemd.html
 Main PID: 4794 (mysqld)
```

```
   Status: "Server is operational"
    Tasks: 39
   Memory: 422.6M
   CGroup: /system.slice/mysqld.service
           └─4794 /mysql/install/mysql/bin/mysqld
--defaults-file=/mysql/install/my.cnf

Oct 08 19:53:10 localhost systemd[1]: Starting MySQL Server...
Oct 08 19:53:18 localhost systemd[1]: Started MySQL Server.

[root@localhost system]# systemctl stop mysqld
```

自此 mysql 服务器自建完成。

mysql 数据库基础使用介绍

mysql 常用命令

查看 mysql 中有哪些数据库：

```
mysql&gt; show databases;  #注意：以分号结尾，分号是英文的分号。
+--------------------+
| Database           |
+--------------------+
| information_schema |
| mysql              |
| performance_schema |
+--------------------+
3 rows in set (0.02 sec)
```

选择使用某个数据库:

```
mysql&gt; use mysql;

Reading table information for completion of table and column names

You can turn off this feature to get a quicker startup with -A



Database changed
```

查看某个数据库下有哪些表:

```
mysql&gt; show tables;

+-----------------------------------------------------------+

| Tables_in_mysql                                           |

+-----------------------------------------------------------+

| columns_priv                                              |

| component                                                 |

| db                                                        |

| default_roles                                             |

| engine_cost                                               |

| func                                                      |

| general_log                                               |

| global_grants                                             |

| gtid_executed                                             |

| help_category                                             |

| help_keyword                                              |

| help_relation                                             |

| help_topic                                                |
```

```
| innodb_index_stats                                 |
| innodb_table_stats                                 |
| ndb_binlog_index                                   |
| password_history                                   |
| plugin                                             |
| procs_priv                                         |
| proxies_priv                                       |
| replication_asynchronous_connection_failover       |
| replication_asynchronous_connection_failover_managed |
| replication_group_configuration_version            |
| replication_group_member_actions                   |
| role_edges                                         |
| server_cost                                        |
| servers                                            |
| slave_master_info                                  |
| slave_relay_log_info                               |
| slave_worker_info                                  |
| slow_log                                           |
| tables_priv                                        |
| time_zone                                          |
| time_zone_leap_second                              |
| time_zone_name                                     |
| time_zone_transition                               |
| time_zone_transition_type                          |
```

```
| user                                                   |
+--------------------------------------------------------+
```

创建数据库:

```
mysql&gt; create database example;
Query OK, 1 row affected (0.01 sec)

mysql&gt; show databases;
+--------------------+
| Database           |
+--------------------+
| example            |
| information_schema |
| mysql              |
| performance_schema |
+--------------------+
4 rows in set (0.00 sec)
```

查看 mysql 数据库的版本号：

```
mysql&gt; select version();
+-----------+
| version() |
+-----------+
| 8.4.0     |
+-----------+
1 row in set (0.00 sec)
```

查看当前使用的是哪个数据库:

```
mysql&gt; select database();
+------------+
| database() |
+------------+
| mysql      |
+------------+
1 row in set (0.01 sec)
```

导入数据库表:

```
mysql&gt; source D:\course\03-MySQL\document\test.sql  #注意：路径中不要有中文。
```

查看表的结构:

```
mysql&gt; desc <table_name>; #describe 缩写为：desc
```

简单查询 select

```
mysql&gt; select * from db;
+-----------+-------------------+--------------+------------+-------------+-------------+------------+-------------+-----------+------------+-----------------+------------+-----------------+-----------------+---------------+-------------+-----------------+
| Host      | Db                | User         | Select_priv | Insert_priv | Update_priv | Delete_priv | Create_priv | Drop_priv | Grant_priv | References_priv | Index_priv | Alter_priv | Create_tmp_table_priv | Lock_tables_priv | Create_view_priv |
```

```
Show_view_priv | Create_routine_priv | Alter_routine_priv | Execute_priv
| Event_priv | Trigger_priv |
+-----------+-------------------+--------------+------------+-----
--------+------------+------------+------------+----------+------
------+----------------+------------+------------+-----------------
-----+-----------------+-----------------+---------------+--------
------------+-------------------+-------------+-----------+------
--------+

| localhost | performance_schema | mysql.session | Y           | N
| N           | N           | N           | N         | N          | N
| N           | N           | N                    | N  | N
| N              | N                   | N                  | N
| N           | N           |

+-----------+-------------------+--------------+------------+-----
--------+------------+------------+------------+----------+------
------+----------------+------------+------------+-----------------
-----+-----------------+-----------------+---------------+--------
------------+-------------------+-------------+-----------+------
--------+

1 row in set (0.00 sec)
```

#这种方式的缺点：

1、效率低（他会先把*号转换为字段）

2、可读性差。

给查询的列起别名：使用 as 关键字起别名

```
mysql&gt; select host,user as username from db; #as 也可以省略

+-----------+--------------+
```

```
| host      | username      |
+-----------+---------------+
| localhost | mysql.session |
+-----------+---------------+
1 row in set (0.00 sec)
```

条件查询

查询语法格式：

select 字段 1,字段 2,字段 3.... from 表名 where 条件;

```
mysql&gt; select host,user from user;
+-----------+------------------+
| host      | user             |
+-----------+------------------+
| %         | root             |
| localhost | mysql.infoschema |
| localhost | mysql.session    |
| localhost | root             |
+-----------+------------------+
4 rows in set (0.00 sec)
mysql&gt; select host,user username  from user where
user='mysql.session';
+-----------+---------------+
| host      | username      |
+-----------+---------------+
| localhost | mysql.session |
```

```
+-----------+---------------+

1 row in set (0.00 sec
```

表的创建

```
#创建一个学生表

create table t_student(

      no int,

      name varchar(32),

      sex char(1),

      age int(3),

      email varchar(255)

   );
```

表的删除

```
 #删除表：

 drop table t_student;

 // 当这张表不存在的时候会报错

 // 如果这张表存在，删除

 drop table if exists t_student;
```

插入数据 insert

插入数据 insert （DML）

语法格式：insert into 表名(字段名 1,字段名 2,字段名 3...) values(值 1,值 2,值 3);

```
mysql&gt;  create table t_user(number bigint(32),username varchar(32),sex char(1) default '男');

mysql&gt; desc t_user;
```

天翼云
State Cloud

```
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key | Default | Extra |
+----------+-------------+------+-----+---------+-------+
| number   | bigint      | YES  |     | NULL    |       |
| username | varchar(32) | YES  |     | NULL    |       |
| sex      | char(1)     | YES  |     | 男      |       |
+----------+-------------+------+-----+---------+-------+
```

```
mysql&gt; insert into t_user (number,username,sex) values (1,'user1','男'), (2,'user2','女'), (3,'user3','男');
Query OK, 3 rows affected (0.00 sec)
Records: 3  Duplicates: 0  Warnings: 0
mysql&gt; select * from t_user;
+--------+----------+------+
| number | username | sex  |
+--------+----------+------+
|      1 | user1    | 男   |
|      2 | user2    | 女   |
|      3 | user3    | 男   |
+--------+----------+------+
3 rows in set (0.00 sec)
```

修改 update

语法格式：update 表名 set 字段名 1=值 1,字段名 2=值 2,字段名 3=值 3... where 条件;

注意：没有条件限制会导致所有数据全部更新！

```
update t_user set name = 'jack', birth = '2000-10-11' where id = 2;
+------+----------+------------+---------------------+
| id   | name     | birth      | create_time         |
+------+----------+------------+---------------------+
|    1 | zhangsan | 1990-10-01 | 2020-03-18 15:49:50 |
|    2 | jack     | 2000-10-11 | 2020-03-18 15:51:23 |
+------+----------+------------+---------------------+


update t_user set name = 'jack', birth = '2000-10-11', create_time = now()
where id = 2;
```

更新所有；

```
update t_user set name = 'abc';
```

删除数据 deleted

语法格式：delete from 表名 where 条件;

注意：没有条件，整张表的数据会全部删除！

```
delete from t_user where id = 2;


insert into t_user(id) values(2);


delete from t_user; // 删除所有
```

快速删除表中的数据；

```
//删除 dept_bak 表中的数据
delete from dept_bak;
```

```
mysql&gt; select * from dept_bak;

Empty set (0.00 sec)
```

删除表操作；

```
drop table 表名; // 删除表。
```

mysql 常见问题及解决思路

案例一：Too many connections（连接数过多，导致连接不上数据库，业务无法正常进行）

问题复现：

```
mysql> SHOW VARIABLES LIKE 'max_connections'; #将当前数据库允许最大连接数设置为 1

+-----------------+-------+

| Variable_name   | Value |

+-----------------+-------+

| max_connections | 800   |

+-----------------+-------+

1 row in set (0.03 sec)


mysql>  set global max_connections=1;

Query OK, 0 rows affected (0.00 sec)


mysql> SHOW VARIABLES LIKE 'max_connections';

+-----------------+-------+

| Variable_name   | Value |
```

```
+-----------------+-------+
| max_connections | 1     |
+-----------------+-------+
1 row in set (0.00 sec)


[root@mysql ~]# mysql -uroot -predhat
mysql: [Warning] Using a password on the command line interface can be insecure.
ERROR 1040 (HY000): Too many connections #报错，太多连接
```

解决思路：

首先先要考虑在我们 MySQL 数据库参数文件里面，对应的 max_connections 这个参数值是不是设置的太小了，导致客户端连接数超过了数据库所承受的最大值。该值默认大小是 151，我们可以根据实际情况进行调整。

对应解决办法：

```
set global max_connections=500
```

但这样调整会有隐患，因为我们无法确认数据库是否可以承担这么大的连接压力，就好比原来一个人只能吃一个馒头，但现在却非要让他吃 10 个，他肯定接受不了。反应到服务器上面，就有可能会出现宕机的可能。

所以这又反映出了，我们在新上线一个业务系统的时候，要做好压力测试。保证后期对数据库进行优化调整。

其次可以限制 InnoDB 的并发处理数量，如果 innodb_thread_concurrency = 0（这种代表不受限制） 可以先改成 16 或是 64 看服务器压力。

如果非常大，可以先改的小一点让服务器的压力下来之后，然后再慢慢增大，根据自己的业务而定。

案例二 主从复制报错类型

```
Last_SQL_Errno: 1062 （从库与主库数据冲突）

Last_Errno: 1062


Last_Error: Could not execute Write_rows event on table test.t;


Duplicate entry ‘4’ for key ‘PRIMARY’,


Error_code: 1062; handler error HA_ERR_FOUND_DUPP_KEY;


the event’s master log mysql-bin.000014, end_log_pos 1505
```

针对这个报错，我们首先要考虑是不是在从库中误操作导致的。结果发现，我们在从库中进行了一条针对有主键表的 SQL 语句的插入，导致主库再插入相同 sql 的时候，主从状态出现异常。发生主键冲突的报错。

解决方法：在确保主从数据一致性的前提下，可以在从库进行错误跳过。一般使用 percona-toolkit 中的 pt-slave-restart 进行。

在从库完成如下操作：

```
[root@zs bin]# ./pt-slave-restart -uroot -proot123
```

之后最好在从库中开启 read_only 参数，禁止在从库进行写入操作。

```
Last_IO_Errno: 1593（server-id 冲突）

Last_IO_Error:


Fatal error: The slave I/O thread stops because master and slave have equal MySQL server ids;
```

```
these ids must be different for replication to work


(or the  –replicate-same-server-id option must be used on slave but this


does not always make sense; please check the manual before using it)
```

这个报错出现之后，就能一目了然看到两台机器的 server-id 是一样的。

在搭建主从复制的过程中，我们要确保两台机器的 server-id 是唯一的。这里再强调一下 server-id 的命名规则（服务器 ip 地址的最后一位+本 MySQL 服务的端口号）。

解决方法：

在主从两台机器上设置不同的 server-id。

```
Last_SQL_Errno: 1032（从库少数据，主库更新的时候，从库报错）
Last_SQL_Error:


Could not execute Update_rows event on table test.t; Can’t find record


in ‘t’, Error_code: 1032; handler error HA_ERR_KEY_NOT_FOUND; the


event’s master log mysql-bin.000014, end_log_pos 1708
```

解决问题的办法：

根据报错信息，我们可以获取到报错日志和 position 号，然后就能找到主库执行的哪条 sql，导致的主从报错。

在主库执行：

```
/usr/local/mysql/bin/mysqlbinlog  -no-defaults -v -v  --
-output=decode-rows /data/mysql/mysql-bin.000014 |grep -A 10 1708 >
1.log


cat 1.log


#170720 14:20:15 server id 3 end_log_pos 1708 CRC32 0x97b6bdec
Update_rows: table id 113 flags: STMT_END_F


### UPDATE `test`.`t`


### WHERE


### @1=4 /* INT meta=0 nullable=0 is_null=0 */


### @2=' dd'  /* VARSTRING(60) meta=60 nullable=1 is_null=0 */


### SET


### @1=4 /* INT meta=0 nullable=0 is_null=0 */


### @2=' ddd'  /* VARSTRING(60) meta=60 nullable=1 is_null=0 */


# at 1708
```

```
#170720 14:20:15 server id 3 end_log_pos 1739 CRC32 0xecaf1922 Xid = 654

COMMIT/*!*/;

DELIMITER ;

# End of log file

ROLLBACK /* added by mysqlbinlog */;

/*!50003 SET COMPLETION_TYPE=@OLD_COMPLETION_TYPE*/;

/*!50530 SET @@SESSION.PSEUDO_SLAVE_MODE=0*/;
```

获取到 SQL 语句之后，就可以在从库反向执行 SQL 语句。把从库缺少的 SQL 语句补全，解决报错信息。

在从库依次执行：

```
mysql> insert into t (b) values ( ‘ddd’ );

Query OK, 1 row affected (0.01 sec)

mysql> stop slave;

Query OK, 0 rows affected (0.00 sec)
```

```
mysql> exit

Bye

[root@node4 bin]# ./pt-slave-restart -uroot -proot123
```

案例三 MySQL 安装过程中的报错

```
[root@zs data]# /usr/local/mysql/bin/mysqld_safe  -
defaults-file=/etc/my.cnf &[1] 3758

[root@zs data]# 170720 14:41:24 mysqld_safe Logging to
‘/data/mysql/error.log’.

170720 14:41:24 mysqld_safe Starting mysqld daemon with databases from
/data/mysql170720

14:41:25 mysqld_safe mysqld from pid file /data/mysql/node4.pid ended

170720 14:41:24 mysqld_safe Starting mysqld daemon with databases from
/data/mysql2017-07-20

14:41:25 0 [Warning] TIMESTAMP with implicit DEFAULT value is deprecated.

Please use  -explicit_defaults_for_timestamp server option
```

```
(see documentation for more details)./usr/local/mysql/bin/mysqld:

File ‘/data/mysql/mysql-bin.index’ not found (Errcode: 13 -
Permission denied)

2017-07-20 14:41:25 4388 [ERROR] Aborting
```

解决思路：

遇到这样的报错信息，我们要学会时时去关注错误日志 error log 里面的内容。看见了关键的报错点 Permission denied，证明当前 MySQL 数据库的数据目录没有权限。

解决方法：

```
[root@zs data]# chown mysql:mysql -R mysql

[root@zs data]# /usr/local/mysql/bin/mysqld_safe -
defaults-file=/etc/my.cnf &

[1] 4402

[root@zs data]# 170720 14:45:56 mysqld_safe Logging to
‘/data/mysql/error.log’.

170720 14:45:56 mysqld_safe Starting mysqld daemon with databases from
/data/mysql
```

如何避免这类问题，个人建议在安装 MySQL 初始化的时候，一定加上 –user=mysql，这样就可以避免权限问题。

```
./mysql_install_db –basedir=/usr/local/mysql/ –datadir=/data/mysql/ –defaults-file=/etc/my.cnf –user=mysql
```

案例四 数据库密码忘记的问题

```
[root@zs ~]# mysql -uroot -p

Enter password:

ERROR 1045 (28000): Access denied for user ‘root’@’localhost’ (using password: YES)

[root@zs ~]# mysql -uroot -p

Enter password:

ERROR 1045 (28000): Access denied for user ‘root’@’localhost’ (using password: YES)
```

解决思路：

目前是进入不了数据库的情况，所以我们要考虑是不是可以跳过权限。因为在数据库中，MySQL 数据库中 user 表记录着我们用户的信息。

解决方法：启动 MySQL 数据库的过程中，可以这样执行：

```
/usr/local/mysql/bin/mysqld_safe –defaults-file=/etc/my.cnf –skip-grant-tables &
```

这样启动，就可以不用输入密码，直接进入 MySQL 数据库了。然后在修改你自己想要改的 root 密码即可。

```
update mysql.user set password=password( ‘root123′ ) where user=’ root’;
```

案例五 truncate 删除数据，导致自动清空自增 ID，前端返回报错 not found

这个问题的出现，就要考虑下 truncate 和 delete 的区别了，看下实验演练：

首先先创建一张表：

```
CREATE TABLE `t` (

`a` int(11) NOT NULL AUTO_INCREMENT,

`b` varchar(20) DEFAULT NULL,

PRIMARY KEY (`a`),

KEY `b` (`b`)

) ENGINE=InnoDB AUTO_INCREMENT=300 DEFAULT CHARSET=utf8
```

插入三条数据：

```
mysql> insert into t (b) values ( ‘aa’ );

Query OK, 1 row affected (0.00 sec)

mysql> insert into t (b) values ( ‘bb’ );
```

```
Query OK, 1 row affected (0.00 sec)

mysql> insert into t (b) values ( ‘cc’ );

Query OK, 1 row affected (0.00 sec)

mysql> select * from t;

+—-+——+

| a | b |

+—-+——+

| 300 | aa |

| 301 | bb |

| 302 | cc |

+—-+——+

3 rows in set (0.00 sec)
```

先用 delete 进行删除全表信息，再插入新值。

结果发现 truncate 把自增初始值重置了，自增属性从 1 开始记录了。当前端用主键 id 进行查询时，就会报没有这条数据的错误。

个人建议不要使用 truncate 对表进行删除操作，虽然可以回收表空间，但是会涉及自增属性问题。这些坑，我们不要轻易钻进去。

案例六 can’t open file (errno:24)

有的时候，数据库跑得好好的，突然报不能打开数据库文件的错误了。
解决思路：首先我们要先查看数据库的 error log。然后判断是表损坏，还是权限问题。还有可能磁盘空间不足导致的不能正常访问表；操作系统的限制也要关注下；用 perror 工具查看具体错误！

```
linux:/usr/local/mysql/bin # ./perror 24

OS error code 24: Too many open files
```

超出最大打开文件数限制！ulimit -n 查看系统的最大打开文件数是 65535，不可能超出！那必然是数据库的最大打开文件数超出限制！

在 MySQL 里查看最大打开文件数限制命令：show variables like ‘open_files_limit’;

发现该数值过小，改为 2048，重启 MySQL，应用正常

处理方法：

```
repair table ;

chown mysql 权限
```

清理磁盘中的垃圾数据

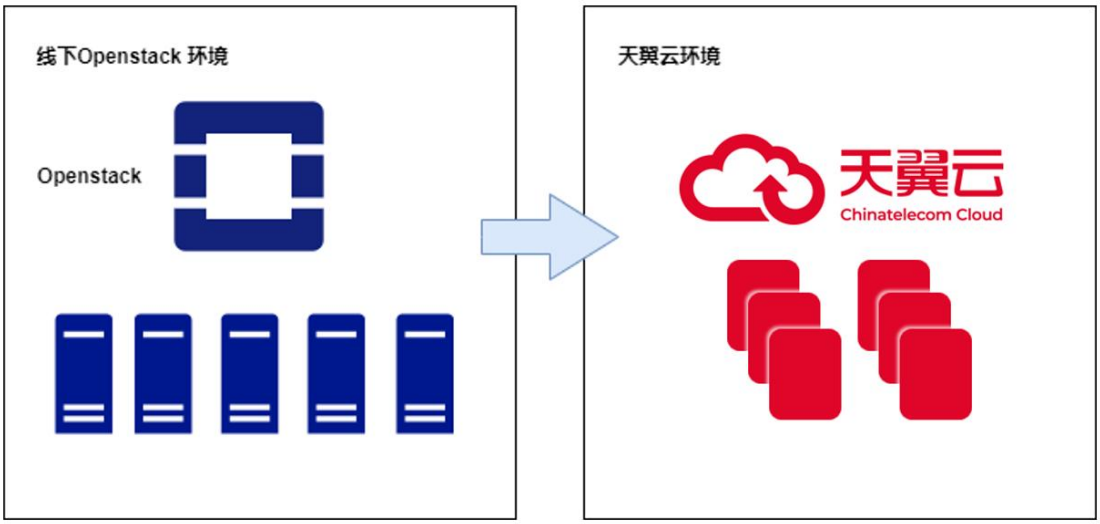
10.10 线下 OpenStack 环境导入公有云环境

操作场景

在线下环境中，很多用户使用 OpenStack 构建云环境。为了降低成本，部分用户选择将线下 OpenStack 环境迁移至云服务商。此次场景将模拟 OpenStack 线下环

境的云主机迁移到天翼云上，利用天翼云提供的服务提高资源运维管理能力，缩小成本。

架构迁移图示：



操作步骤

本文将介绍如何使用 VMWare 在本地搭建一套 OpenStack 环境，以模拟迁移和导入的过程。

模拟环境如下：

- OS：CentOS Linux release 7.9.2009
- VMWare Workstation:16.1.0
- 模拟流程包含如下步骤：
- 安装 OpenStack 环境
- 创建云主机
- 导出云主机
- 导入天翼云
- 创建云主机（天翼云环境）

安装 OpenStack 环境。

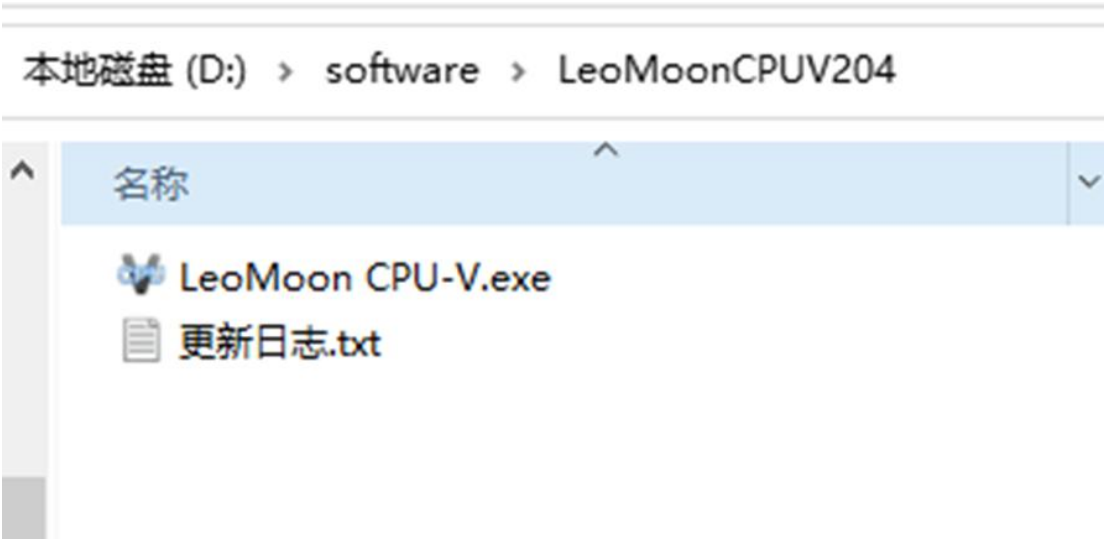
OpenStack 能够创建云主机，我们需要保障如下要求符合：

(1) CPU 支持虚拟化

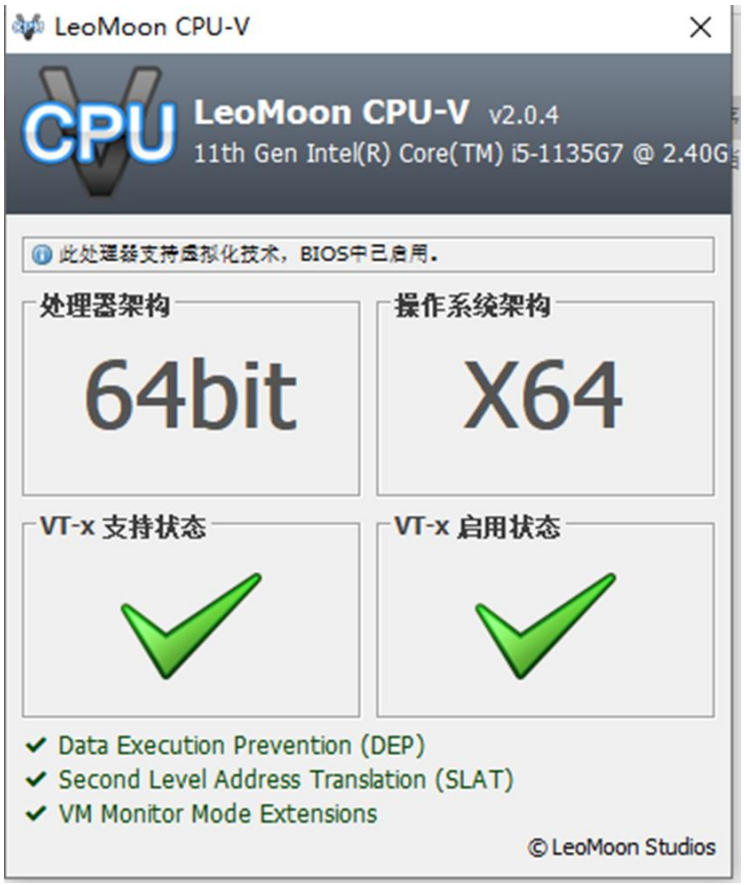
(2) 内存至少 8G 以上

步骤 1：检查 CPU 是否支持虚拟化

下载 LeoMoon CPU-V，并双击打开。



步骤 2：如果结果如下图所示，表明已经支持虚拟化。



步骤 3： 在虚拟机中安装 OpenStack（all-in-one）。

本次使用 packstack 自动安装 rocky 版本，准备工作如下：

1）执行以下命令，禁止使用 NetworkManager，配置静态 IP。

```
systemctl stop NetworkManager && systemctl disable NetworkManager

vim /etc/sysconfig/network-scripts/ifcfg-ens33
```

在配置文件中修改 IP 如下图后保存。

```
TYPE="Ethernet"
PROXY_METHOD="none"
BROWSER_ONLY="no"
BOOTPROTO="statics"
IPADDR=192.168.43.128
NETMASK=255.255.255.0
GATEWAY=192.168.43.2
DEFROUTE="yes"
IPV4_FAILURE_FATAL="no"
IPV6INIT="yes"
IPV6_AUTOCONF="yes"
IPV6_DEFROUTE="yes"
IPV6_FAILURE_FATAL="no"
IPV6_ADDR_GEN_MODE="stable-privacy"
NAME="ens33"
UUID="5a41ef4c-e3fd-4ed9-86cc-bc374c496ef0"
DEVICE="ens33"
ONBOOT="yes"
```

2）执行以下命令，修改主机名。

```
hostname set-hostname controller
```

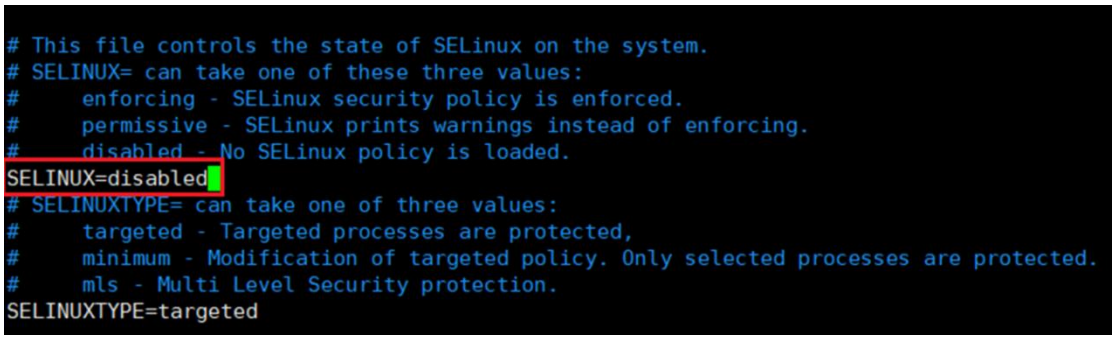
3）如下图所示，将主机名与 IP 映射写入 hosts。

```
127.0.0.1   localhost localhost.localdomain localhost4 localhost4.localdomain4
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.43.128 controller
```

4） 执行以下命令，禁用 firewalld。

```
hostname set-hostname controllersystemctl stop firewalld && systemctl
disable firewalldsetenforce 0
```

5）修改 selinux 配置文件，将 SELINUX 设置为 disabled。



6）执行以下命令，配置 yum 源。

```
[root@controller yum.repos.d(keystone_admin)]# cat CentOS-Base.repo

[base]

name=CentOS-$releasever - Base - mirrors.aliyun.com

failovermethod=priority

baseurl=http://mirrors.aliyun.com/centos/$releasever/os/$basearch/

        http://mirrors.aliyuncs.com/centos/$releasever/os/$basearch/


http://mirrors.cloud.aliyuncs.com/centos/$releasever/os/$basearch/

gpgcheck=1

gpgkey=http://mirrors.aliyun.com/centos/RPM-GPG-KEY-CentOS-7

 #released updates

[updates]

name=CentOS-$releasever - Updates - mirrors.aliyun.com

failovermethod=priority

baseurl=http://mirrors.aliyun.com/centos/$releasever/updates/$basearc
h/


http://mirrors.aliyuncs.com/centos/$releasever/updates/$basearch/
```

```
http://mirrors.cloud.aliyuncs.com/centos/$releasever/updates/$basearch/
gpgcheck=1
gpgkey=http://mirrors.aliyun.com/centos/RPM-GPG-KEY-CentOS-7
 #additional packages that may be useful
[extras]
name=CentOS-$releasever - Extras - mirrors.aliyun.com
failovermethod=priority
baseurl=http://mirrors.aliyun.com/centos/$releasever/extras/$basearch/

http://mirrors.aliyuncs.com/centos/$releasever/extras/$basearch/

http://mirrors.cloud.aliyuncs.com/centos/$releasever/extras/$basearch/
gpgcheck=1
gpgkey=http://mirrors.aliyun.com/centos/RPM-GPG-KEY-CentOS-7
 #additional packages that extend functionality of existing packages
[centosplus]
name=CentOS-$releasever - Plus - mirrors.aliyun.com
failovermethod=priority
baseurl=http://mirrors.aliyun.com/centos/$releasever/centosplus/$basearch/
```

```
http://mirrors.aliyuncs.com/centos/$releasever/centosplus/$basearch/

http://mirrors.cloud.aliyuncs.com/centos/$releasever/centosplus/$basearch/
gpgcheck=1
enabled=0
gpgkey=http://mirrors.aliyun.com/centos/RPM-GPG-KEY-CentOS-7
 #contrib - packages by Centos Users
[contrib]
name=CentOS-$releasever - Contrib - mirrors.aliyun.com
[root@controller yum.repos.d(keystone_admin)]# cat epel.repo[epel]name=Extra Packages for Enterprise Linux 7 - $basearch# It is much more secure to use the metalink, but if you wish to use a local mirror# place its address here.#baseurl=http://download.example/pub/epel/7/$basearchmetalink=https://mirrors.fedoraproject.org/metalink?repo=epel-7&arch=$basearch&infra=$infra&content=$contentdirfailovermethod=priorityenabled=1gpgcheck=1gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-7
[epel-debuginfo]name=Extra Packages for Enterprise Linux 7 - $basearch - Debug# It is much more secure to use the metalink, but if you wish to use a local mirror# place its address here.#baseurl=http://download.example/pub/epel/7/$basearch/debugmetalink=https://mirrors.fedoraproject.org/metalink?repo=epel-debug-7&arch=$basearch&infra=$infra&content=$contentdirfailovermethod=priorityenabled=0gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-7gpgcheck=1
```

```
[epel-source]name=Extra Packages for Enterprise Linux 7 - $basearch - Source# It is much more secure to use the metalink, but if you wish to use a local mirror# place it's address here.#baseurl=http://download.example/pub/epel/7/source/tree/metalink=https://mirrors.fedoraproject.org/metalink?repo=epel-source-7&arch=$basearch&infra=$infra&content=$contentdirfailovermethod=priorityenabled=0gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-7gpgcheck=1

[root@controller yum.repos.d(keystone_admin)]# cat CentOS-OpenStack-rocky.repo[centotack-rocky]name=openstack-rockybaseurl=https://mirrors.aliyun.com/centos/7/cloud/x86_64/openstack-rocky/enabled=1gpgcheck=0[qume-kvm]name=qemu-kvmbaseurl=https://mirrors.aliyun.com/centos/7/virt/x86_64/kvm-common/enabled=1gpgcheck=0
```

7） 执行以下命令，安装 packstack。

```
#安装 leatherman

yum -y install leatherman#安装 packstack

yum -y install openstack-utils openstack-packstack
```

8） 执行以下命令，生成应答文件并编辑配置文件。

```
packstack --gen-answer-file=/root/openstack.ini

vim openstack.ini#除修改以下所列项外，其他项不保持变（密码字段，可根据需要自行修改，这里全部配置为了 123456）
CONFIG_DEFAULT_PASSWORD=123456CONFIG_AODH_INSTALL=nCONFIG_MARIADB_USER=rootCONFIG_MARIADB_PW=123456CONFIG_KEYSTONE_DB_PW=123456CONFIG_KEYSTONE_ADMIN_EMAIL=root@localhostCONFIG_KEYSTONE_ADMIN_USERNAME=adminCONFIG_KEYSTONE_ADMIN_PW=123456CONFIG_NEUTRON_ML2_TYPE_DRIVERS=vxlan,flatCONFIG_NEUTRON_ML2_TENANT_NETWORK_TYPES=vxlanCONFIG_NEUTRON_ML2_MECHANISM_DRIVERS=openvswitchCONFIG_NEUTRON_L2_AGENT=openvswitchCONFIG_N
```

```
EUTRON_OVS_BRIDGE_MAPPINGS=extnet:br-exCONFIG_NEUTRON_OVS_BRIDGE_IFACES=br-ex:eth0CONFIG_NEUTRON_OVS_EXTERNAL_PHYSNET=extnetCONFIG_PROVISION_DEMO=n
```

9）执行以下命令，通过应答文件进行安装。

```
packstack --answer-file=/root/openstack.ini
```

10）打开浏览器，输入 http://IP，本示例为 http://192.168.43.128/，登录 OpenStack 管理页面。



11）执行以下命令，导入镜像。

```
openstack image create "centos"  --file centos.img  --disk-format qcow2 --container-format bare  --public
```

12）执行以下命令，创建网络。

```
openstack create net vnet

openstack network create --provider-network-type=flat --provider-physical-network=extnet --external --share ext_net
```

13）执行以下命令，创建子网。

```
openstack subnet create subnet1 --network int --subnet-range
172.16.1.0/24 --gateway 172.16.1.1openstack subnet create ex --network
ext_net --gateway 192.168.101.254 --allocation-pool
start=192.168.101.200,end=192.168.101.205 --subnet-range
192.168.101.0/24
```

14）执行以下命令，创建路由。

```
openstack router create routeopenstack router add subnet route ex

openstack router add subnet route subnet1
```

15）执行以下命令，创建规格。

```
openstack flavor create --vcpus 1 --ram 512 --disk 10 test1
```

16）执行以下命令，创建密钥。

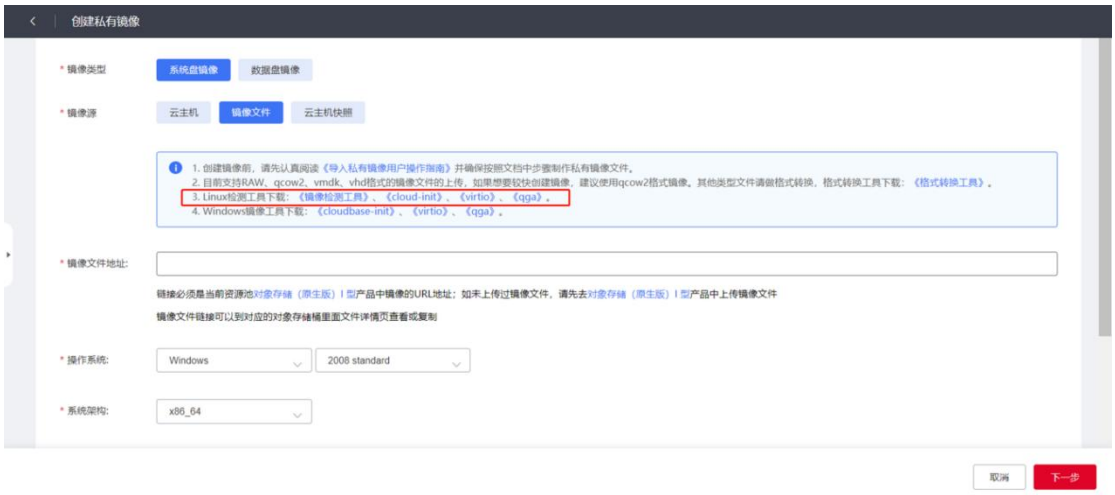
```
openstack keypair create  cy
```

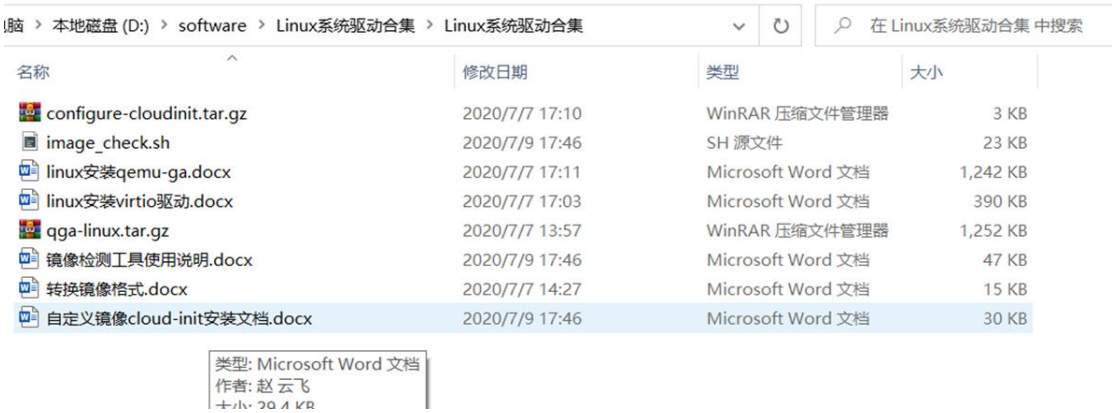
17）执行以下命令，创建虚拟机。

```
openstack server create --flavor test1 --image centos --nic
net-id=cc9d7710-d82e-4109-9412-56ae02a18d7d --security-group default
--key-name cy test
```

18）执行以下命令，连接虚拟机并上传 qemu-ga、cloudinit 以及 virtio 组件。

下载路径在“镜像服务”>“私有镜像”>“创建私有镜像”页面中连接下载。





根据说明文档安装后关闭虚拟机。

19）执行以下命令，将云主机转为 image。

```
openstack server image create 1b1d3411-b755-4980-9da6-db01d8331600
--name cytestimg
```

20）执行以下命令，将 image 保存到本地（此时的镜像为 raw 格式）。

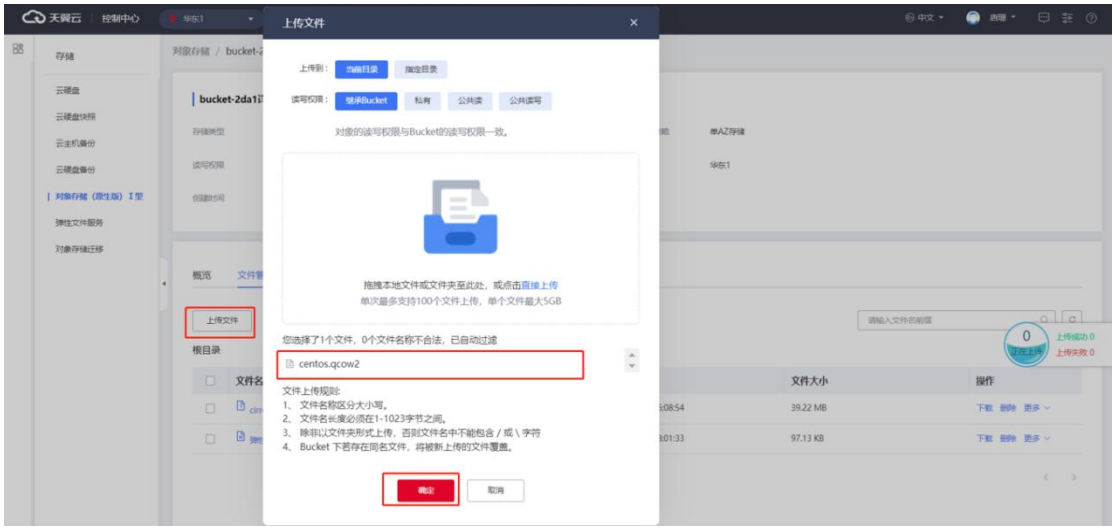
```
openstack server image create 1b1d3411-b755-4980-9da6-db01d8331600
--name cytestimg
```

21）执行以下命令，将镜像转换为 qcow2 格式（防止镜像过大，上传到桶失败）。

```
qemu-img convert -f raw -o qcow2 cytestimg centos.qcow2
```

22）将镜像上传到天翼云。

进入对象存储原生版，创建桶后，上传镜像到桶中（镜像较大建议通过 oss api 上传）。



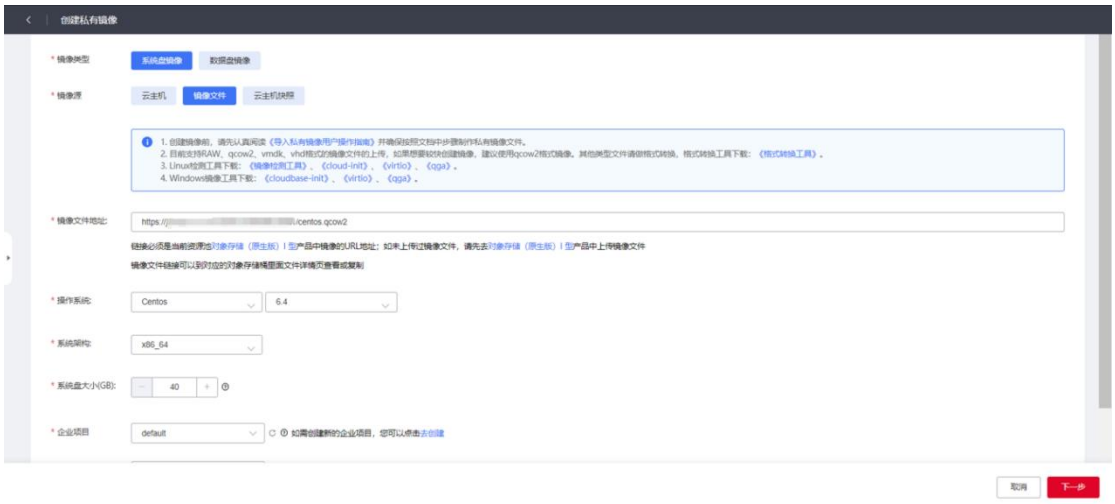


23）创建私有镜像。

通过“镜像服务”>“私有镜像”>“创建私有镜像”，选择镜像文件，填写对象的地址后，转为私有镜像

其中地址可以在对象存储控制台，“更多”>“复制 URL”将地址复制过来。





24）创建成功后，“弹性云主机”>“创建云主机”，镜像处选择私有镜像。创建完成后可进入云主机查看数据完整性。

10.11 创建双栈云主机

1.目前双栈云主机未面向线上用户开放，线下客户请向云公司受理岗申请双栈公测权限，注明账号名称，账号归属（哪个分公司、哪个客户），公测的资源池。

2.双栈是针对子网开启的，需在 VPC 内创建一个新的双栈子网，或对原有子网开启 IPv6。申请完双栈公测权限后，创建子网时可看到选项。



3.云主机必须创建在已开启 IPv6 的子网中，云主机目前只有在新创建的时刻支持开启双栈，存量主机暂无法增加 IPv6 地址。

4.新建一台云主机，选择支持双栈的规格，并非所有主机规格都支持双栈，具体见后续列表，但也存在双栈 POD 中主机资源售罄而无法开出，如果 1-4 步都正常走完，选择支持的主机类型还是看不到 IPv6 选项，就是双栈资源售罄了。

5.下图是新建云主机时，选择了合适的规格和子网，可以看到红框，选择自动分配即可获得 IPv6 地址。右边的蓝框是共享带宽，如果配置，则该 IPv6 地址将加入共享带宽，可访问 IPv6 公网。

6.弹性 IP 仅用于访问 IPv4 公网，如果不访问 IPv4 公网，无需使用。



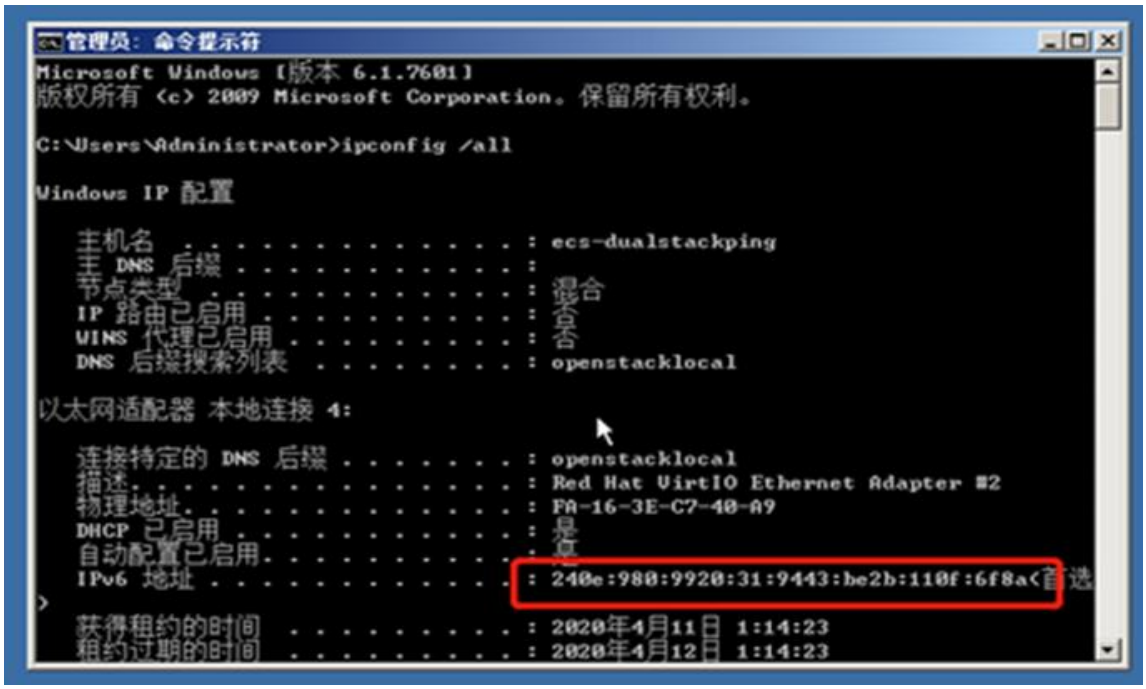
7.牢记：弹性 IP 与双栈功能没有关系，不要把弹性 IP 的 NAT64 转换功能与双栈混为一谈。双栈云主机是通过云主机上绑定的双栈网卡获得内网 IPv4 地址、内外合一的 IPv6 地址。如果访问 IPv4 公网，就申请弹性 IP；如果访问 IPv6 公网，就把双栈网卡的 IPv6 地址加入共享带宽。

8.申请双栈主机成功后，主机将获得 IPv6 地址，此地址是全球唯一的，内网公网都相同，但此时只支持内网访问。

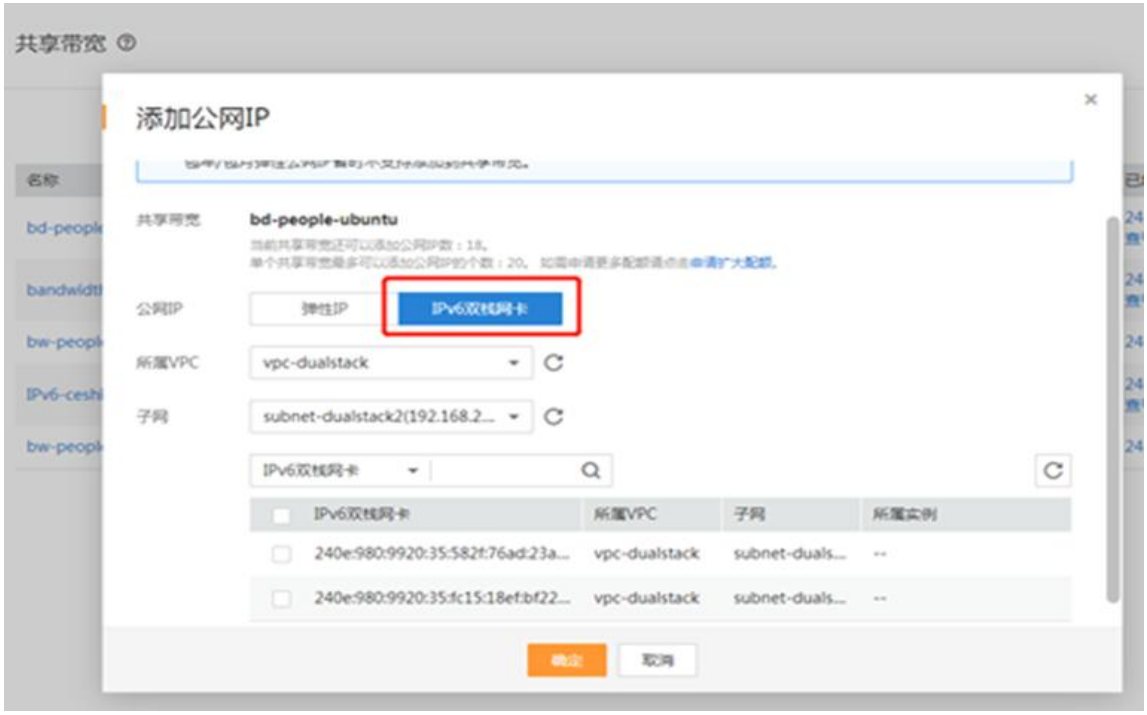


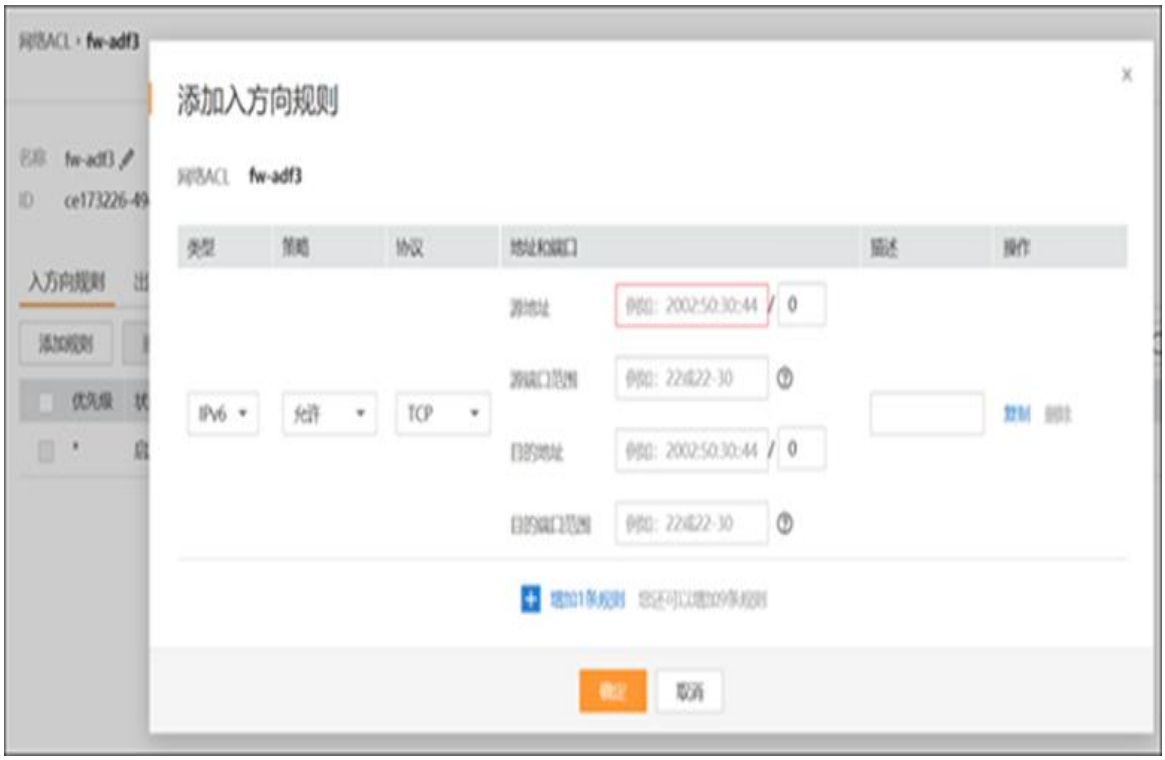
9.以上都操作完成后，可登录主机尝试使用 ipconfig /all 或 ifconfig 来查看主机是否获取了该 IPv6 地址。如果没有需要手动开

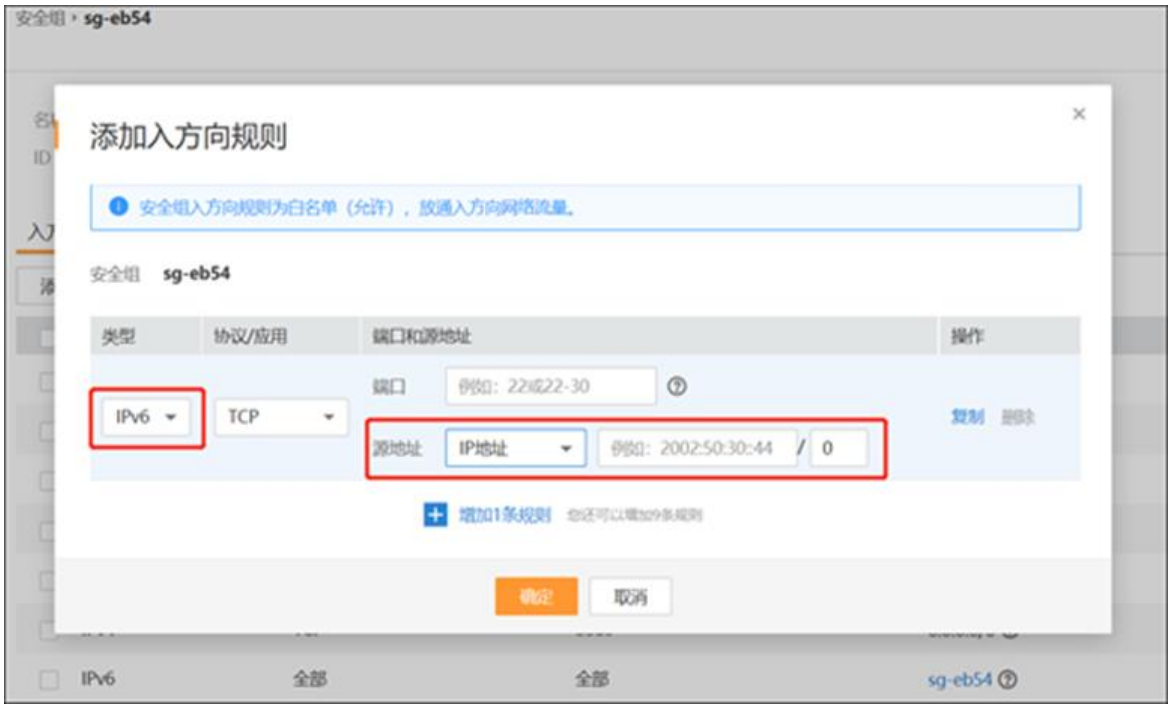
启 IPv6 协议。



10.如果要访问 IPv6 公网，将 IPv6 地址加入共享带宽，并查看网络 ACL 和安全组的出入方向是否已放通所需 IPv6 协议。 ::/0 表示全部 IPv6 地址，0.0.0.0/0 表示全部 IPv4 地址。




网络ACL › fw-adf3
添加入方向规则
网络ACL fw-adf3
类型
策略
协议
地址和端口
描述
操作
源地址
例如: 2002:50:30:44
/ 0
源端口范围
例如: 22或22-30
IPv6
允许
TCP
目的地址
例如: 2002:50:30:44
/ 0
目的端口范围
例如: 22或22-30
增加1条规则
您还可以增加99条规则
确定
取消


安全组 › sg-eb54
添加入方向规则
安全组入方向规则为白名单（允许），放通入方向网络流量。
安全组 sg-eb54
类型
协议/应用
端口和源地址
操作
IPv6
TCP
端口
例如: 22或22-30
源地址
IP地址
例如: 2002:50:30:44
/ 0
复制
删除
增加1条规则
您还可以增加99条规则
确定
取消
IPv6
全部
全部
sg-eb54

11. 尝试访问 http://www.edu.cn，正常显示网页则成功。如果失败，可将 IPv6 DNS 设置为 240c::6666 再尝试。



10.12 弹性云主机配置

景介绍

为了帮助用户最大程度地安全、可靠地使用弹性云主机，通常可以从以下几个方面进行弹性云主机配置的最佳实践：

- 考虑业务负载需求
- 网络与安全
- 数据备份与恢复策略
- 系统监控与告警

考虑业务负载需求

在选择弹性云主机时，需要根据应用程序的需求选择适当的实例类型和规格，考虑性能需求、应用需求以及成本效益等因素。

性能需求

性能需求是选择弹性云主机配置的首要考虑因素。性能需求包括计算性能、存储性能和网络性能。

- 计算性能越强，服务器能够同时处理的请求就越多。如果应用需要大量计算资源，比如进行大规模数据分析、科学计算等，就需要选择配置较高的 CPU 和内存。
- 存储性能是指服务器对数据的读写速度。对于需要频繁读写大量数据的应用，如数据库、视频处理等，需要选择配置较高的存储设备。
- 网络性能是指服务器的网络带宽和延迟。对于需要大量数据传输的应用，如视频直播、游戏等，需要选择具有较高网络带宽的配置。此外，延迟也是一个重要指标，特别是对于对实时性要求较高的应用，如在线交易、远程会议等。

应用需求

不同的应用有不同的特点和需求，因此需要选择适合应用的配置。

- 操作系统：弹性云主机支持多种操作系统，如 Windows、Linux 等。选择操作系统需要根据应用的实际需求来决定，比如是否支持特定软件、是否需要特定的开发环境等。
- 数据库：如果应用需要使用数据库，需要考虑数据库的类型和规模。对于大型数据库，需要选择配置较高的服务器，以保证数据库的性能和稳定性。
- 应用软件：如果应用需要运行特定的软件，比如 Web 服务器、应用服务器、大数据平台等，需要选择配置适合运行这些软件的服务器。

成本效益

在云计算领域，用户只需按需购买和使用资源，不需要投资大额资金购买硬件设备，因此成本效益非常重要。

- 价格：弹性云主机的价格与配置相关，一般来说，配置越高，价格越高。在选择配置时，需要根据应用的需求和预算来综合考虑。
- 弹性伸缩：弹性云主机搭配弹性伸缩使用，可以根据应用的需求进行资源灵活调整。在选择配置时，可以考虑未来应用的发展和变化。可以搭配负载均衡实现业务负载的均衡分布。

选择弹性云主机配置需要综合考虑性能需求、应用需求和成本效益等因素，选择最适合业务负载需求的云主机。

网络与安全

弹性云主机通常需要与网络结合使用，因此需要重点考虑到网络设计安全性，同时需要配置多项安全措施保障云主机的安全。

网络逻辑隔离

虚拟私有云提供了一个隔离的虚拟网络环境，可以在云平台上创建自己的私有网络。将弹性云主机部署在虚拟私有云中，可以更好地控制网络访问和保护数据的安全性。主要通过下列方法进行网络安全设置：

- 使用安全组：安全组是一种虚拟防火墙，用于控制弹性云主机的网络流向。配置适当的安全组规则，限制入站和出站流量，只允许必要的端口和协议访问。配置详细步骤参见弹性云主机-安全组配置示例。
- 使用网络 ACL：网络 ACL 是一种在子网级别控制流量的安全策略。通过配置网络 ACL 规则，实现对子网内所有 ECS 实例的访问控制，防止未经授权的访问。配置详细步骤参见虚拟私有云-创建 ACL。

安全登录

为了确保弹性云主机创建成功且状态正常，建议您第一次通过 VNC 方式远程登录弹性云主机，若登录成功则表明弹性云主机资源状态正常。

尽量选用密钥方式登录 Linux 弹性云主机。同时可以采取相应措施对 Linux 服务器 SSH 登录进行安全加固，详细操作，请参见 Linux 服务器 SSH 登录的安全加固。

使用密码登录的实例需要定期修改密码。

详细操作，请参见：

弹性云主机-登录 Linux 弹性云主机

弹性云主机-登录 Windows 弹性云主机

安全产品配置

安全防护：为了实现对弹性云主机的高阶安全防护，建议您根据防护需求开通服务器安全卫士服务，服务器安全卫士通过对主机信息和行为进行持续监控和分析，快速精准地发现安全威胁和入侵事件，保证云主机安全。具体操作参见服务器安全卫士-快速入门。

网络防护：为了防止弹性云主机绑定的弹性公网 IP 被攻击，建议您配合使用 Anti-DDoS 服务，具体操作参见 Anti-DDoS 流量清洗-开启 Anti-DDoS 防护。

数据备份与恢复策略

数据备份与恢复

为了在发生病毒入侵、人为误删除、软硬件故障等事件时，能够快速恢复数据，建议您定期进行数据备份，可以使用天翼云云备份服务，云备份提供简单易用的备份能力，当发生异常时，可将数据恢复到任意备份点，减轻经济损失。具体操作参见云备份-快速入门。

高可用容灾业务部署

为了实现较高的业务容灾能力，建议您将云主机资源部署在同一区域的不同可用区内。具体详情参见地域和可用区。

系统监控与告警

监控管理

为了使用户能够及时获取弹性云主机的各项性能指标及运行状态，建议您使用云监控服务，对实例进行实时监控，实例异常时进行告警和通知。详细内容参见云监控服务。

日志管理

为了提高弹性云主机和应用程序的可用性和性能，建议您通过日志审计服务收集日志数据，通过对海量日志数据的分析和处理，实现实时决策分析、设备运维管理、用户业务趋势分析等能力。

10.13 弹性云主机之间迁移最佳实践

介绍说明

当需要把不同天翼云账号下的主机资源整合到一个天翼云账号下，或者将同账户不同区域的 CT-ECS 资源整合到同一个区域下时，可以采用主机迁移服务实现 CT-ECS 实例间的快速迁移。主机迁移服务支持同一区域或者不同区域的源端到目的端的迁移，下面为您介绍天翼云 CT-ECS 实例间的实现逻辑和操作步骤。

实现逻辑

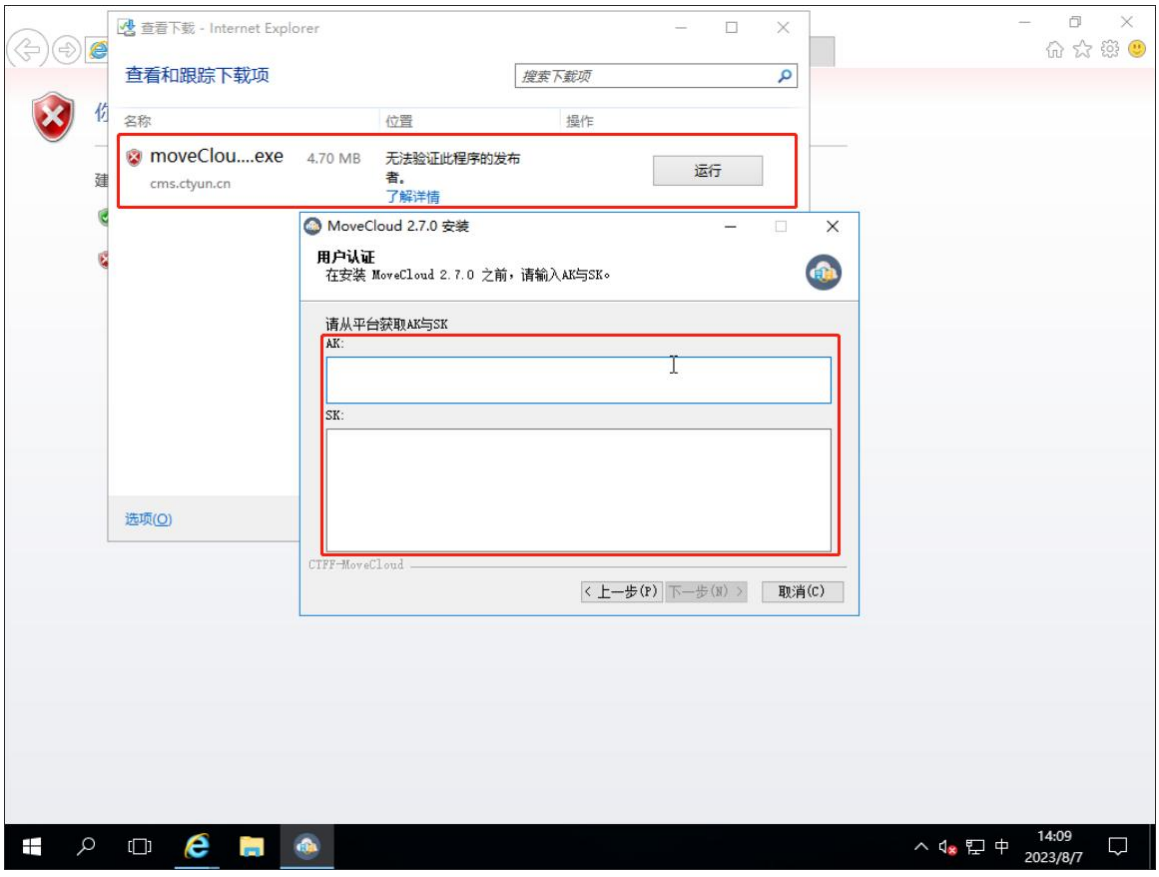
CMS 基础实现逻辑是在目标端部署 PE 轻量系统，在 CMS 控制台处进行任务配置与下发，通过协同 CMS-Agent 与 CMS-PE 实现迁移；

仅需通过确定 PE 轻量系统开通的账户与区域，即可限定迁移目标端的账户与区域。需要注意迁移过程中需要保证账户余额充足与目标区域已加载迁移专用镜像，未加载镜像区域不支持 CMS 迁移服务。

操作步骤

跨账号 CT-ECS 迁移（账户 A 迁移至 B 账户）

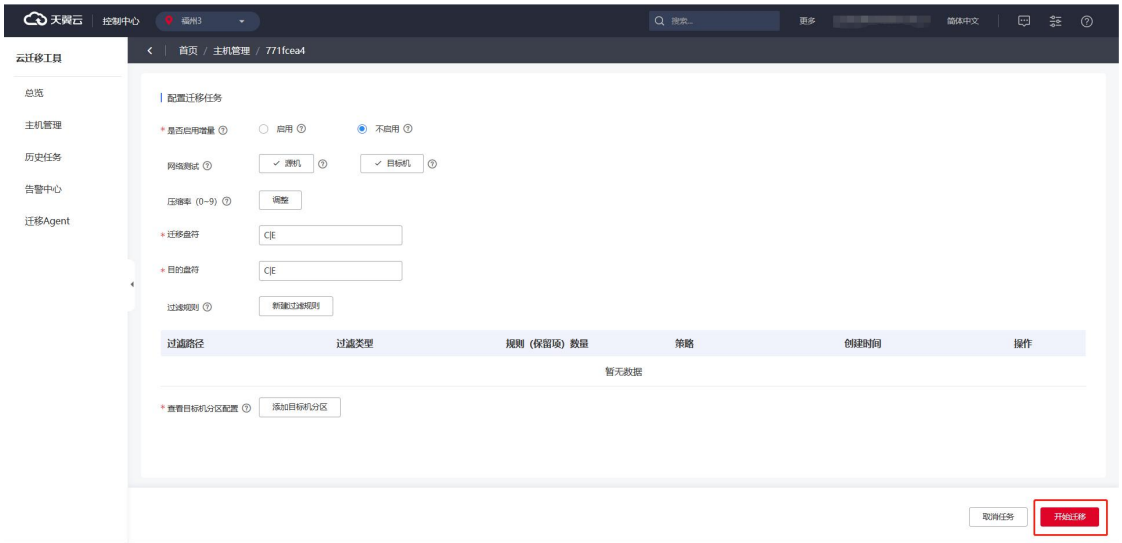
1. 在 A 账户下的 ECS 云主机下，使用 B 账户的云迁移服务 AK/SK 安装 CMS-Agent。具体步骤请参见在源端安装 Agent。



2. A 账户主机即会被上报至 B 账户迁移平台，接下来可以进行迁移目标机设置。具体步骤请参见创建目标机，检测目标机和绑定目标机。

3.绑定成功后可以配置并启动迁移任务。具体步骤请参见迁移任务配置与开始全量迁移。

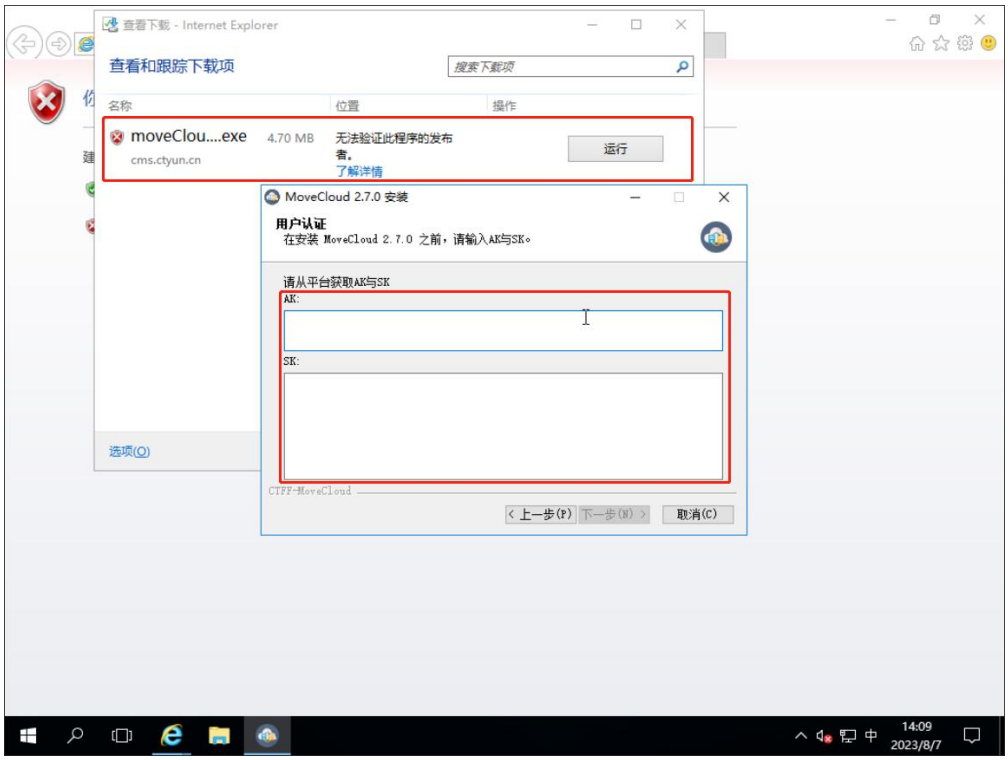
4.迁移任务配置完成后，单击“开始迁移”正常执行后续迁移操作，即可实现 ECS 云主机跨账户迁移。



同账户下 CT-ECS 区域间迁移（C 区域迁移至 D 区域）

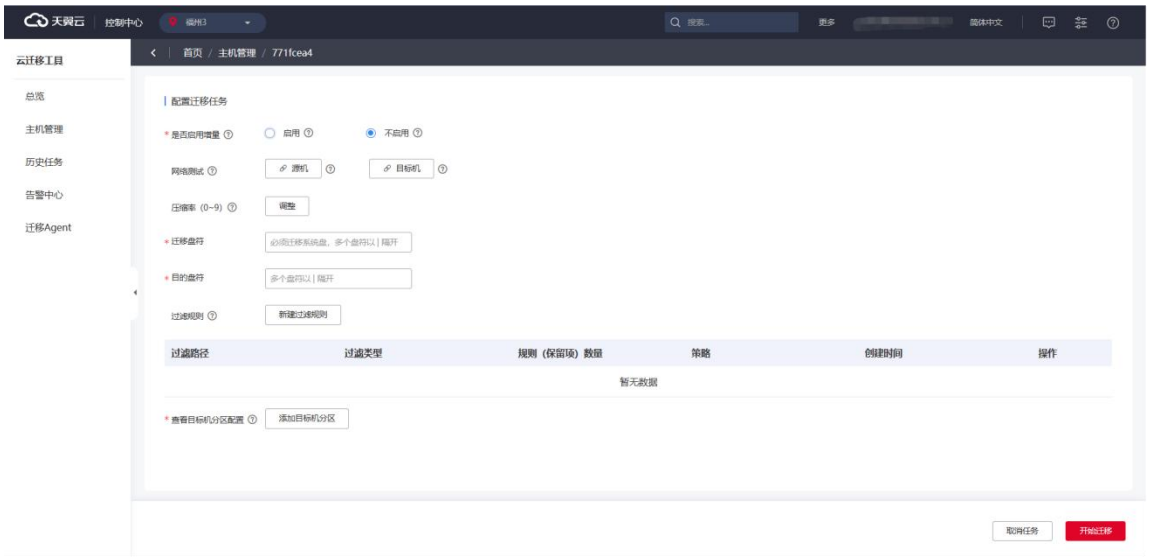
操作前提：D 区域需支持云迁移服务，并且在 D 区域使用云迁移服务。

1.在 C 区域云主机安装 CMS-Agent，将迁移源上报；具体步骤请参见在源端安装迁移 Agent。

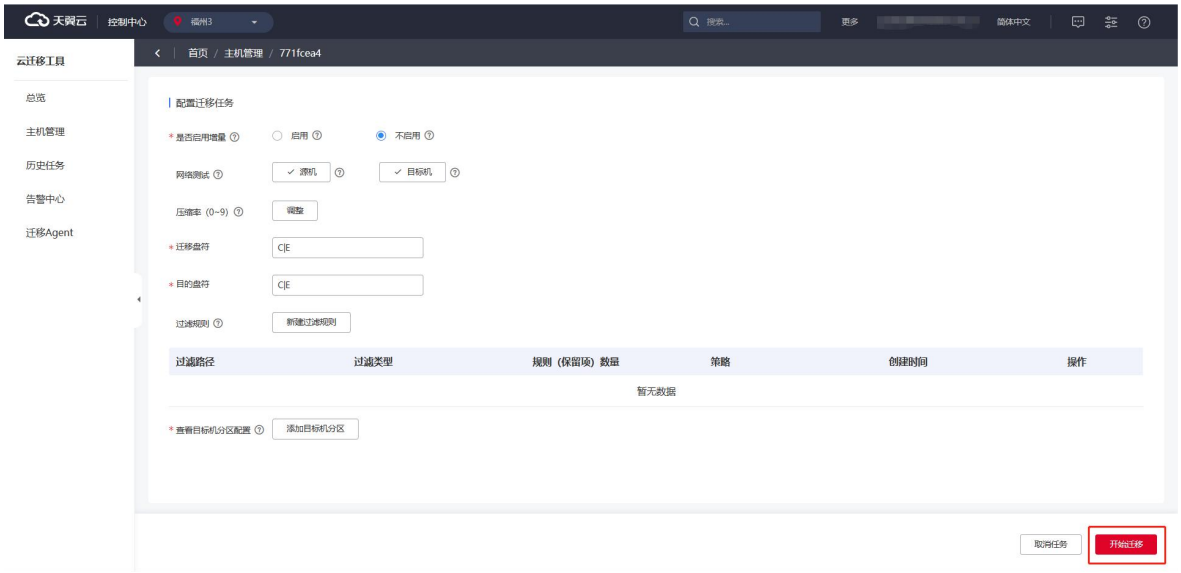


2. 在 D 区域创建目标机，具体步骤请参见创建目标机，检测目标机和绑定目标机。

3. 绑定成功后可以配置并启动迁移任务。具体步骤请参见迁移任务配置与开始全量迁移。



4. 迁移任务配置完成后，单击“开始迁移”正常执行后续迁移操作，即可实现同账户下 CT-ECS 云主机跨区域迁移。



10.14Linux 服务器 SSH 登录的安全加固

SSH 是远程登录 Linux 服务器的主要方式，但它也是黑客进行暴力破解和恶意攻击的主要入口之一。为了加强 Linux 云主机的 SSH 登录安全，您可以采用以下方式进行安全加固：

- 修改默认端口；
- 禁止 root 用户直接登录；
- 添加安全组规则；
- 采用密钥登录；
- 配置 SSH 登录限制。

我们以 centos 为例为大家介绍如何进行安全加固。

修改默认端口

1.远程登录弹性云主机。

2.打开 sshd 配置文件 /etc/ssh/sshd_config，可以使用编辑器打开该文件，例如使用 vi 命令：

```
vi /etc/ssh/sshd_config
```

3.找到以下代码行：

```
#Port 22
```

将该行代码的注释符号 # 去掉，并将 22 修改为想要设置的端口号，例如 2222：

```
Port 2222
```

4. 按 ESC 键，输入:wq 保存修改并退出编辑器。

5. 重启 sshd 服务，使配置生效。可以使用以下命令重启 sshd 服务：

```
systemctl restart sshd
```

6. 确认修改是否生效。可以使用以下命令检查 sshd 服务是否在新端口上监听：

```
netstat -tnl | grep 2222
```

如果输出类似于以下内容，则表示 sshd 服务已经在新端口 2222 上监听：

```
tcp        0      0 0.0.0.0:2222            0.0.0.0:*
LISTEN
```

7. 完成以上步骤后，就成功将 CentOS 系统的 SSH 默认端口号修改为了 2222。注意，在修改 SSH 默认端口号之后，需要使用新的端口号来进行 SSH 连接，例如：

```
ssh username@hostname -p 2222
```

其中，username 和 hostname 分别是 SSH 登录的用户名和远程主机名或 IP 地址。

禁用 root 用户直接登录

Linux 的默认管理员名即是 root，只需要知道 root 密码即可直接登录 SSH。禁止 root 从 SSH 直接登录可以提高服务器安全性。经过以下操作后即可实现。

1. 新建用户

```
useradd test
```

2. 配置密码

#使用 passwd 命令即可给相应帐户设置或修改密码。

passwd test#根据图示，设置或修改密码需要填写两次，第二次为效验密码，输入完毕后请回车确认。

```
Changing password for user test.

New password:

Retype new password:

passwd: all authentication tokens updated successfully.
```

3.配置不允许 root 用户直接登录，修改相关文件

```
vi /etc/ssh/sshd_config

查找 #PermitRootLogin yes”  默认为 132 行

将“#”去掉，末尾“Yes”改为“No”

并:wq 保存
```

4.重启 SSHD 服务

```
systemctl restart sshd
```

5.测试连接，可以看到 直接使用 root 连接服务器 ssh 会直接拒绝：



添加安全组规则

安全组中的入方向规则默认开启了 22 端口，当云服务器的 SSH 登录端口修改为 2222 时，需要为安全组新加一条规则。

登录管理控制台。

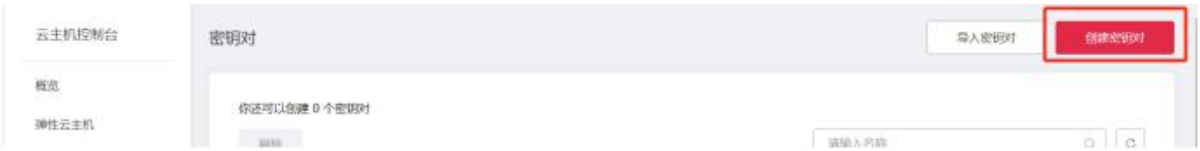
1. 选择“计算 > 弹性云主机”，进入云主机控制台。
2. 单击云服务器名称进入详情页面。
3. 选择“安全组”页签，单击展开安全组规则详情，单击列表右上角的“更改安全组规则”。
4. 添加一条入方向规则。



采用密钥登录

密钥对登录比起密码登录更加安全，因此在创建弹性云主机时您可以选择密钥对登录的方式。

1. 新建密钥对，或者选择已有的密钥对，并下载至本地。



2. 创建云主机时，登录方式选择密钥对。



3. 使用 xshell 登录弹性云主机

通过弹性公网 IP，执行以下命令，SSH 远程连接弹性云主机。

```
ssh 用户名@弹性公网 IP
```

选择“Public Key”，并单击“用户密钥(K)”栏的“浏览”。

在“用户密钥”窗口中，单击“导入”。

选择本地保存的密钥文件，并单击“打开”。

单击“确定”，登录弹性云主机。

配置 SSH 登录限制

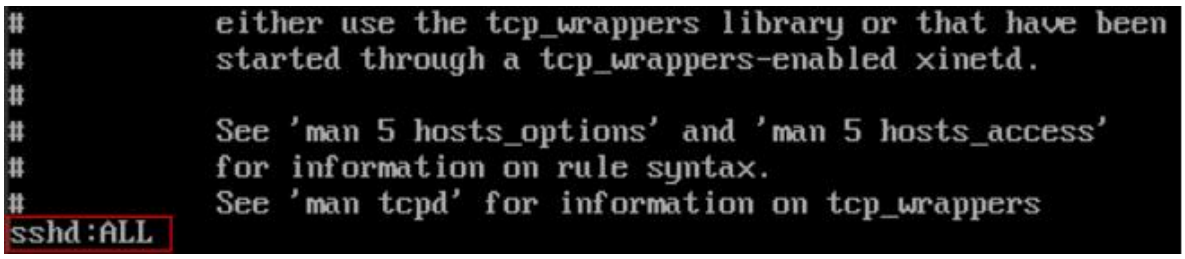
“/etc/hosts.allow”和“/etc/hosts.deny”是控制远程访问的文件，通过配置该文件可以允许或者拒绝某个 IP 或者 IP 段的客户访问 Linux 云服务器的某项服务。

比如 SSH 服务，通常只对管理员开放，那我们就可以禁用不必要的 IP，而只开放管理员可能使用到的 IP 段。

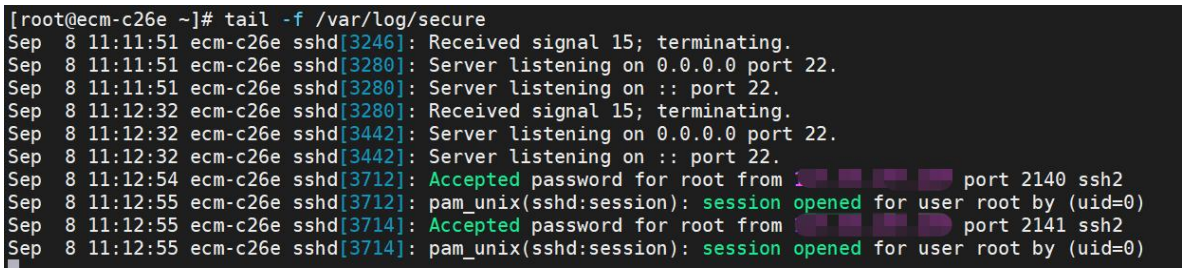
因为云主机需要在不同地点登录，建议编辑“/etc/hosts.allow”允许所有 IP 地址登录，这样不会影响正常使用。

```
vim /etc/hosts.allow
```

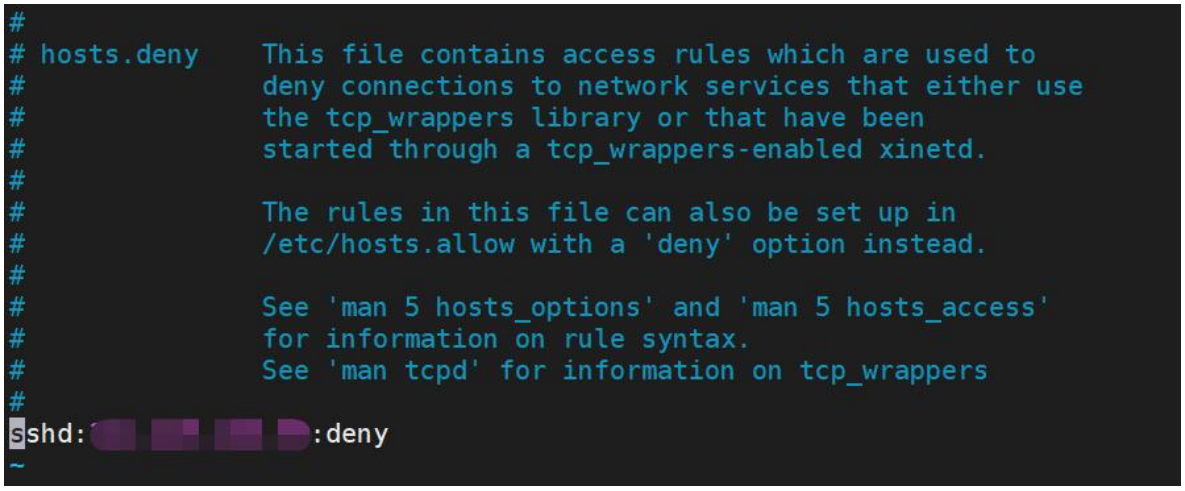
在最后一行增加“sshd:ALL”。



也可以通过一些方法识别云主机的安全风险，使用 tail -f /var/log/secure 打开 SSH 日志，查看疑似恶意登录的 IP。



然后在“/etc/hosts.deny”中将这些地址禁止。



10.15 使用 VNC Viewer 连接 Linux 云主机

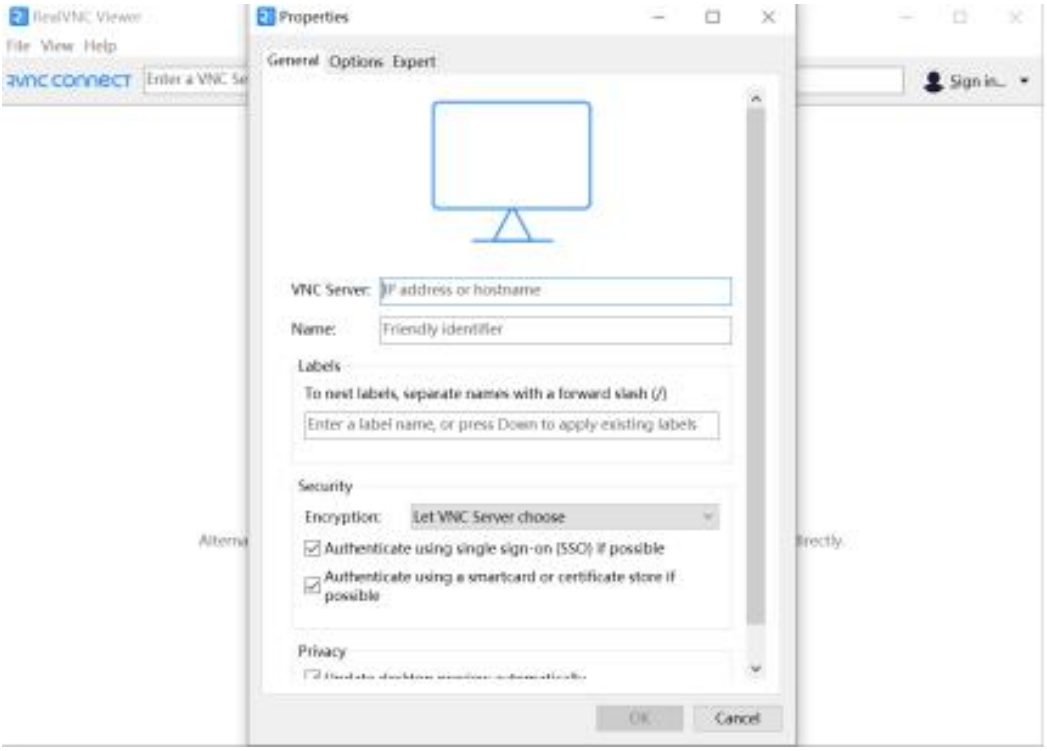
背景介绍

Linux 云主机一般采用 SSH 连接方式，使用密钥对进行安全地无密码访问。但是 SSH 连接一般都是字符界面，有时我们需要使用图形界面进行一些复杂操作。

VNC Viewer 是一款免费,并且开源的远程控制软件，可以实现批量管理，远程监控，批量打开链接，操作一台等于同时操作 N 台，使用起来十分便捷。

准备工作

1.已创建一台 CentOS7.6 操作系统的弹性云主机，并且为云主机绑定弹性公网 IP，确保可以连接互联网，具体操作参见弹性云主机-创建弹性云主机。
2.已在本地 PC 安装 VNC Viewer 客户端。



VNC Viewer 下载地址：
https://www.realvnc.com/en/connect/download/viewer/

给 CentOS 7.6 云主机安装 gnome 桌面环境

1.使用 SSH 工具或控制台远程登录云主机。帐号：root；密码：您在创建云主机时设置的密码。
2.执行如下命令，安装图形化界面。

```
yum groupinstall "GNOME Desktop" "Graphical Administration Tools" -y
```

3.执行以下命令，设置默认启动图形化界面。

```
ln -sf /lib/systemd/system/runlevel5.target
/etc/systemd/system/default.targ
```

4.执行以下命令，重启云主机。

```
reboot
```

5.通过天翼云控制台远程登录方式登录云主机，详情请参见登录 Linux 弹性云主机。

登录云主机后查看可视化界面即表示搭建成功，根据界面提示进行配置进入桌面后，可按需进行相关操作。如图所示：

安装 VNC Server 实现本地远程图形化访问

1.执行如下命令，安装 vncserver。

```
yum install -y tigervnc-server
```

2.配置 vncserver 执行 vncsever 命令配置 vnc 密码（初次启动的时候需要输入两次密码）。

```
vncserver

You will require a password to access your desktop.Password:（输入并记住该密码）

Verify:（重复输入并记住该密码）
```

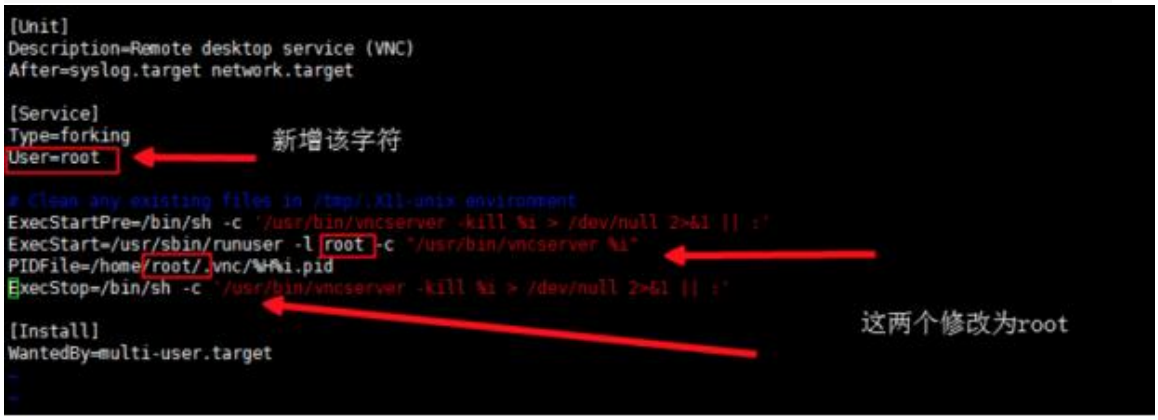


执行以下命令，生成并编辑配置文件。

```
cp -p /lib/systemd/system/vncserver@.service
/lib/systemd/system/vncserver@:1.service
```

执行如下命令修改配置：

```
vim /lib/systemd/system/vncserver@:1.service
```



按 Esc 键退出编辑模式，再输入:wq 保存并退出。

3.运行以下命令重启 systemd。

```
systemctl daemon-reload
```

4.配置为开机自启动。

```
systemctl enable vncserver@:1.service
```

5.运行以下命令启动 VNC Server。

```
systemctl start vncserver@:1.service
```

6.运行命令 ps -ef | grep vnc 确认服务是否已经启动。如果返回以下类似结果，说明服务已经启动。

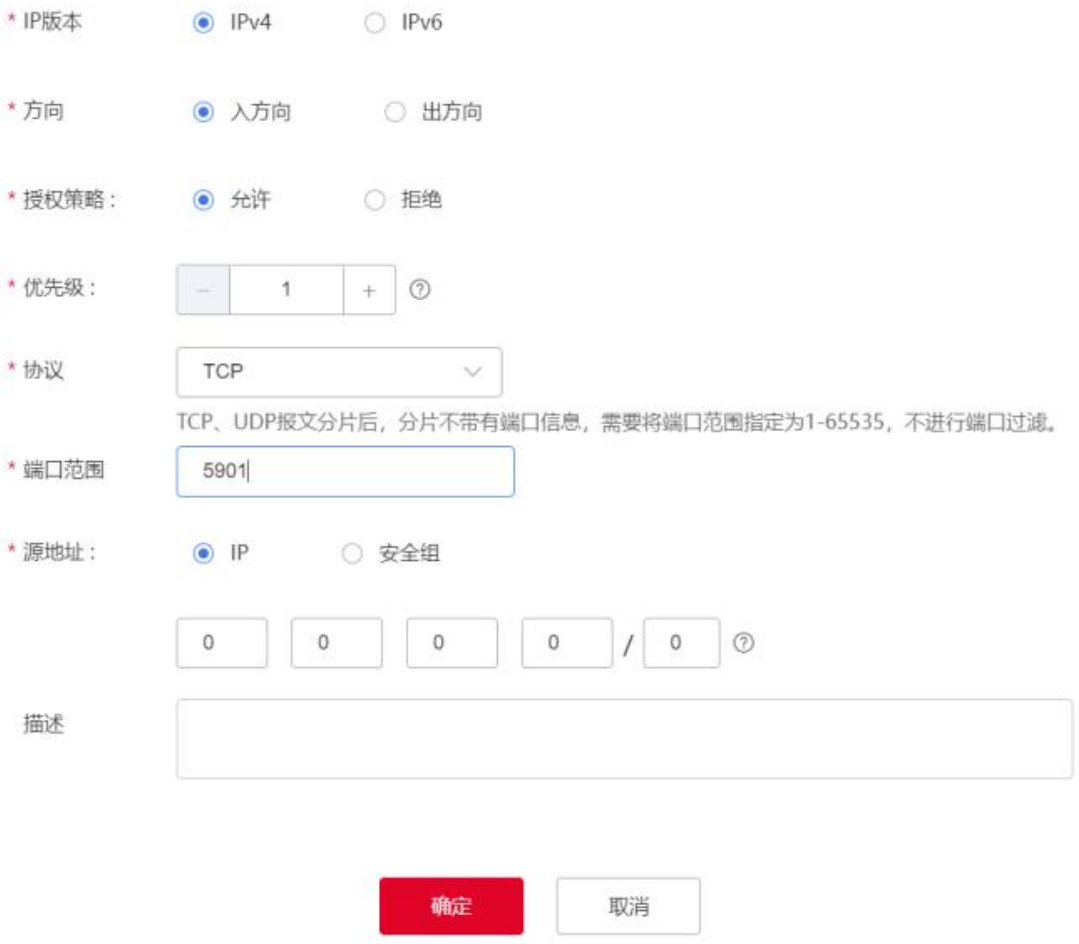


更改安全组规则

1.登录弹性云主机管理控制台。

2.单击弹性云主机名称，进入详情页面。

3.在“安全组”页签，单击“更改安全组规则”，放行 5901 端口。

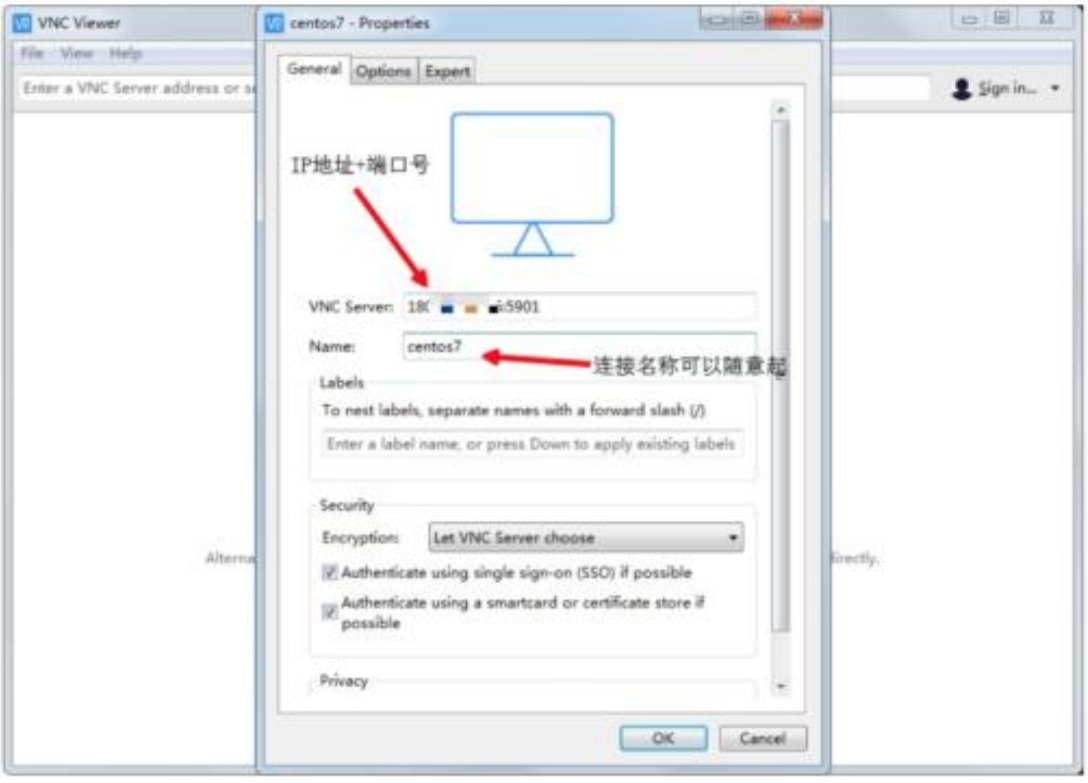


使用 viewer 客户端连接软件连接该(CentOS7.6)实例

1.使用准备工作中提供的 viewer 客户端连接软件连接该实例。点击 file——-New connection，创建新连接。



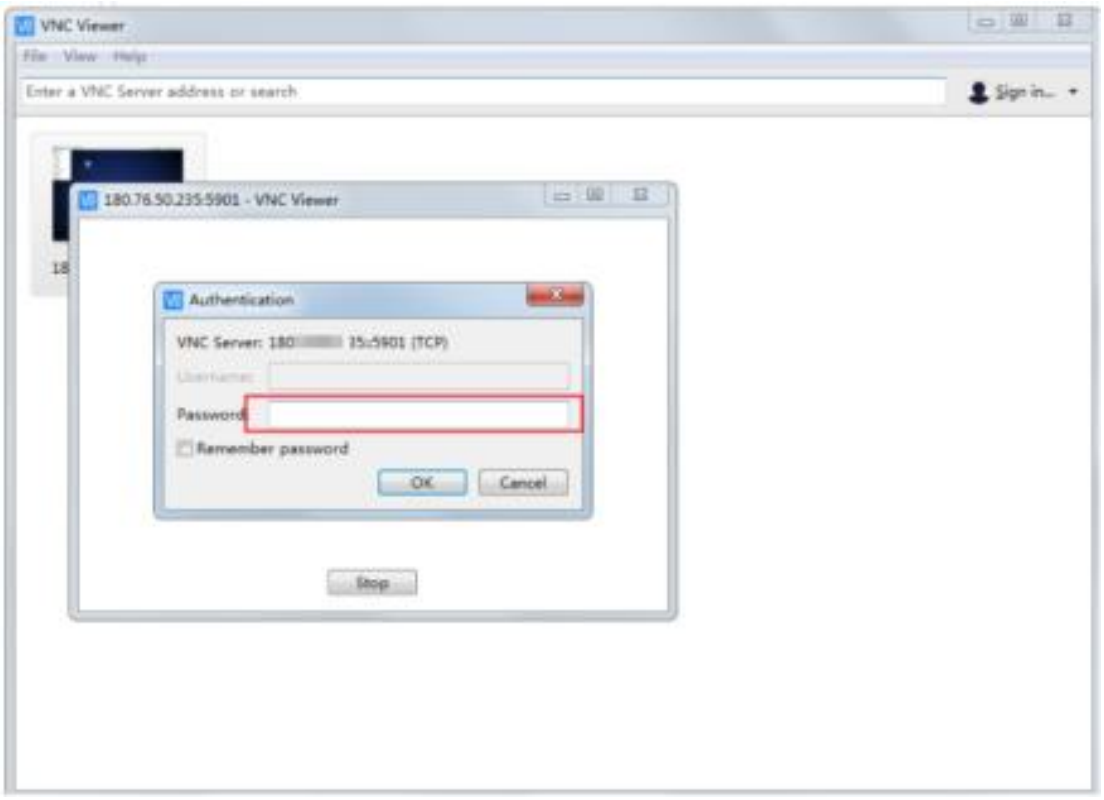
2.输入对应的信息。



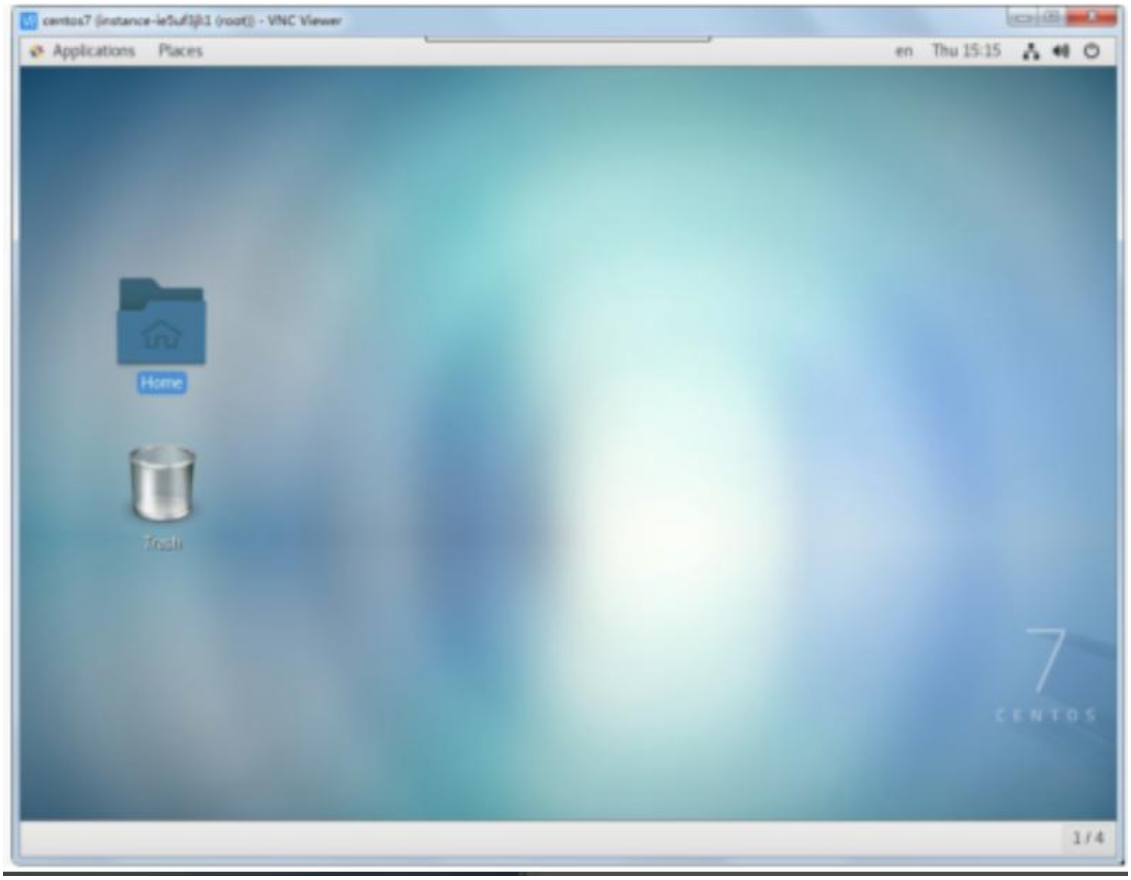
3. 双击之前创建的连接，点击进行连接。



4. 此处输入安装 VNC Server 时设置的密码。



5.连接成功。



10.16Linux 系统进入单用户模式

背景介绍

用户模式介绍

Linux 操作系统有两种重要的用户模式，一种是多用户模式，一种是单用户模式，具体介绍如下：

多用户模式（multi-user mode）：多用户模式是 Linux 系统默认的运行模式。在多用户模式下，系统可以同时支持多个用户登录，并运行各种服务和进程。多用户模式提供了完整的功能和服务，包括网络服务、图形界面等。

单用户模式（single-user mode）：单用户模式是一种特殊的启动模式，提供了一个最小化但强大的环境，允许管理员以超级用户权限进行操作。进入单

用户模式后，系统只会加载最基本的服务和功能，只有一个命令行界面，不会启动图形界面和网络服务。

单用户模式应用场景

单用户模式通常应用于以下场景：

- 系统故障排除和修复：当系统遇到启动问题、文件系统错误、网络配置问题或其他故障时。管理员可以以超级用户权限登录，并执行诊断、修复和恢复任务，如修复文件系统、检查硬件、还原配置文件等。
- 系统备份和还原：单用户模式提供了对系统磁盘的完全访问权限。这使得管理员可以在单用户模式下执行备份和还原操作，包括创建和还原系统快照、复制重要数据、修复损坏的文件系统等。
- 系统配置和维护：单用户模式提供了一个干净且最小化的系统环境，使管理员能够进行系统配置和维护操作，而不会受到其他服务或用户的干扰。这包括更新软件包、重建启动引导程序、更改网络配置等。
- 安全审计：单用户模式可以用于进行安全审计。管理员可以检查系统日志、分析安全事件、查找潜在的漏洞，并采取必要的措施来加固系统和提高安全性。

- 前提条件

- 要进入 Linux 系统的单用户模式，通常需要满足以下前提条件：

- 已经注册并登录天翼云账号，未完成的可参见注册天翼云账号。
- 已经在天翼云上购买了弹性云主机，且购买过程中镜像选择为 Linux 操作系统，例如 CentOS，Ubuntu 等。
- 进入单用户模式后，您将作为超级用户（root）登录。因此，您需要知道 root 用户的密码或者您具有其他获得 root 权限的方法。

- 约束与限制

- 当进入 Linux 操作系统的单用户模式时，存在一些约束与限制，包括：提前备份数据：在进入单用户模式之前，务必提前备份重要数据。单用户模式是一个强大的权限环境，错误的操作可能导致数据丢失或系统不稳定。
- 只有一个命令行界面：单用户模式不会加载图形界面，只提供基本的命令行界面。这意味着用户只能使用命令行工具进行操作，无法使用图形化界面。
- 确保系统文件可读写：单用户模式需要能够读写系统文件，因此需要确保您的文件系统没有损坏并且没有只读的挂载。
- 无网络连接：默认情况下，单用户模式不会启动网络服务，因此无法进行网络连接。这意味着不能访问网络资源，包括互联网、局域网和远程连接。

操作步骤

不同的 Linux 发行版使用不同的引导加载程序，如 GRUB（Grand Unified Bootloader）或 LILO（Linux Loader）。本文将通过系统引导器（GRUB）进入单用户模式。

CentOS 操作系统进入单用户模式

本示例将会对一台操作系统为 CentOS8.0 64 位镜像的弹性云主机实例进行单用户模式进入操作。

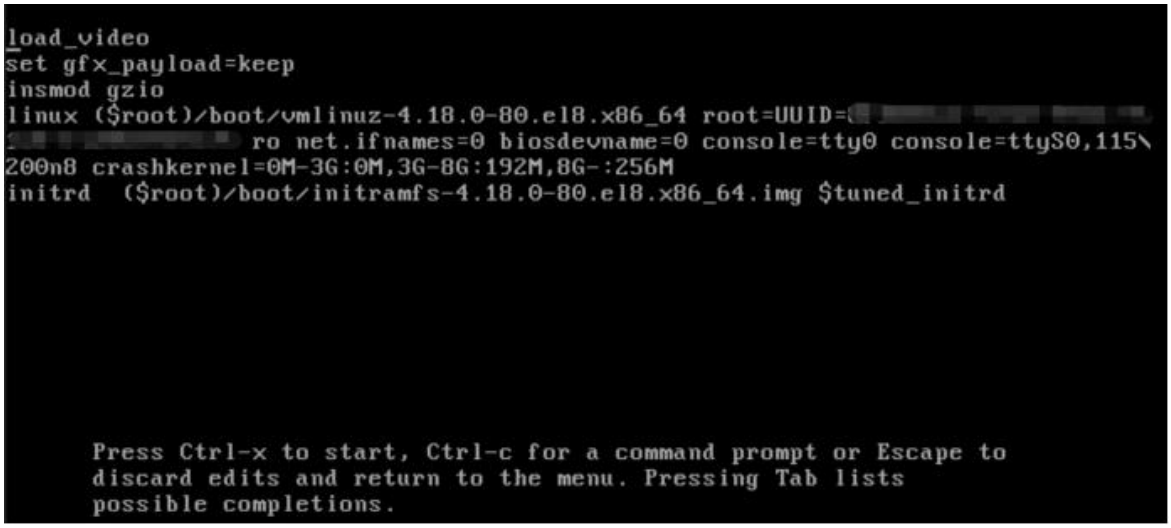
1. 登录控制中心。
2. 选择区域 华东-华东 1。
3. 单击“计算>弹性云主机”，进入弹性云主机页面。
4. 单击待操作的弹性云主机行的“操作>远程登录”按钮，远程连接弹性云主机实例。



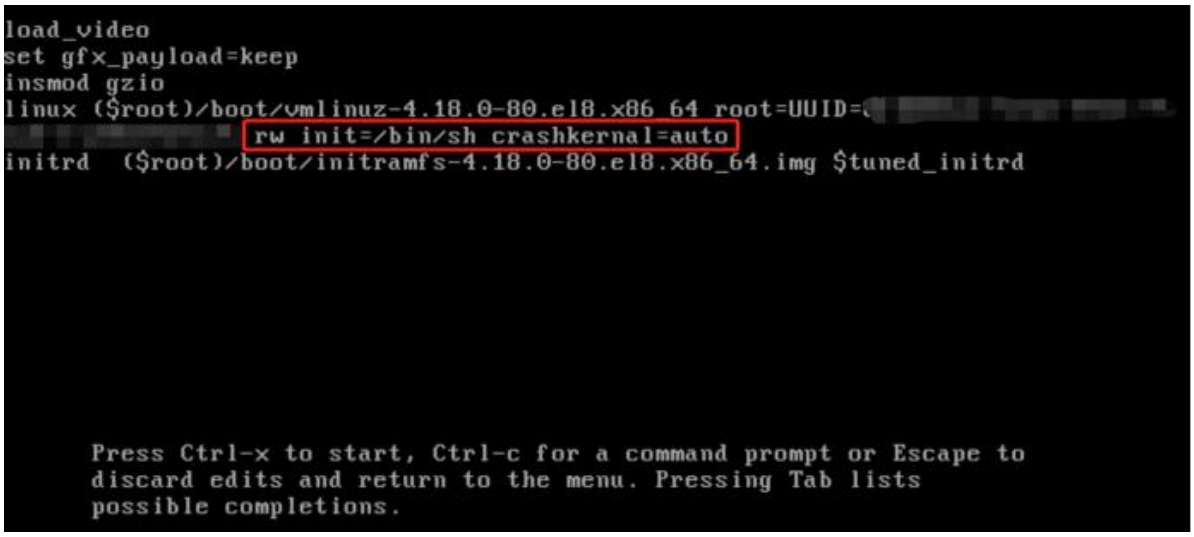
5.输入用户名 root，密码为购买弹性云主机时用户自定义的密码，登录成功之后如图：



6.输入重启命令 reboot，当重启过程中出现让您选择启动系统界面时按下键盘 e 键（为防止页面自动跳转，需要您及时关注重启页面），跳转至启动项配置界面。下图为启动项配置界面：



7.使用键盘的方向键，移动光标到 linux 开头的那一行命令，将本行中 ro 至本行末尾的内容删除，并替换为 rw init=/bin/sh crashkernel=auto，具体修改信息可见下图：



8.截至目前，进入单用户模式的配置已经基本完成，现在需要用户按下键盘的 ctrl+x 组合键或按 F10 键，系统会直接进入单用户模式。输入 passwd 命令重置系统密码，密码输入完成还需要进行密码的二次输入进行确认。示例如图所示。



至此，一台操作系统为 CentOS8.0 64 位的弹性云主机实例就进入到了单用户模式。

Ubuntu 操作系统进入单用户模式

本示例将会对一台操作系统为 Ubuntu18.04 64 位镜像的弹性云主机实例进行单用户模式进入操作。

1.登录控制中心。

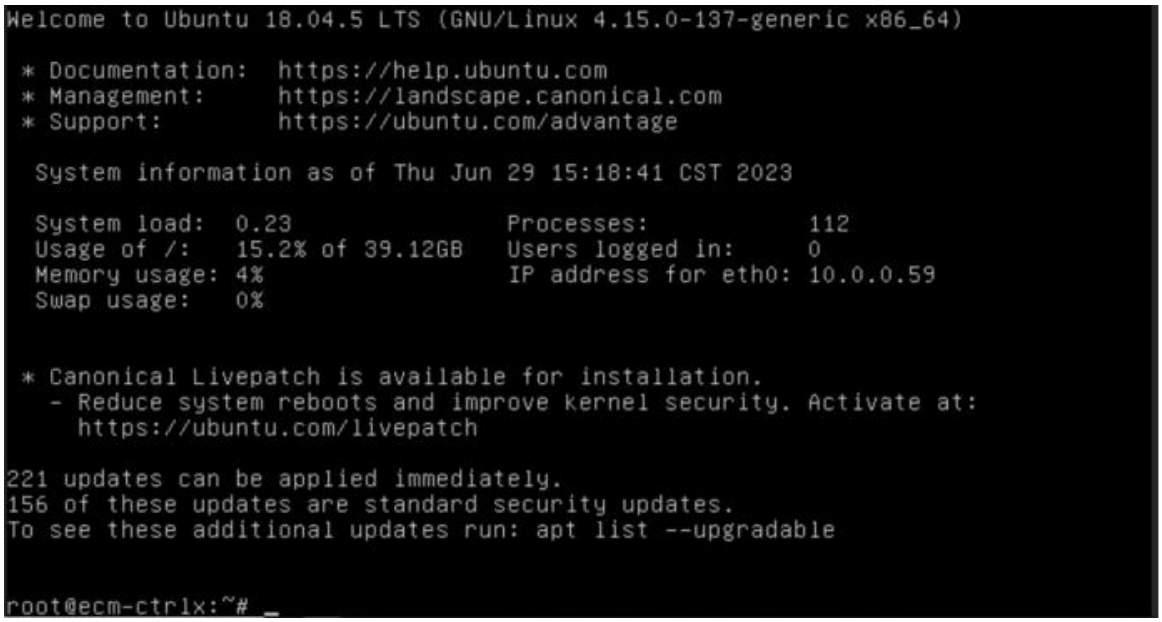
2.选择区域 华东-华东 1。

3.单击“计算>弹性云主机”，进入弹性云主机页面。

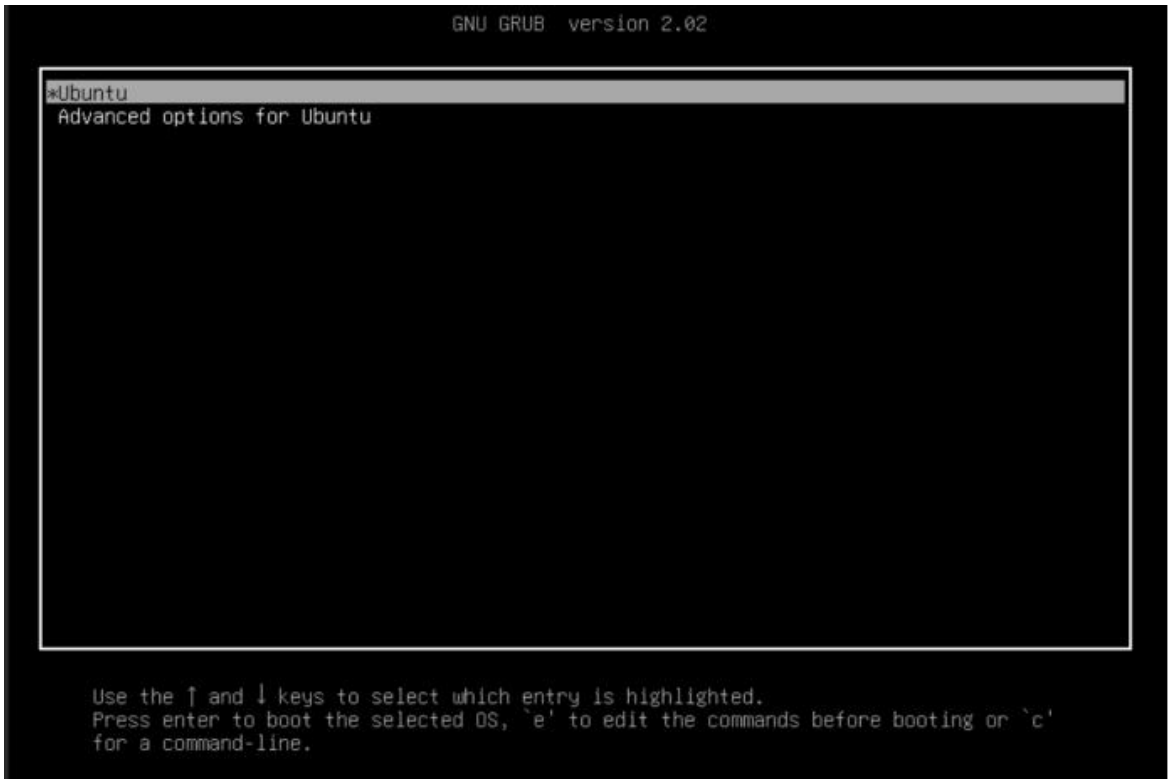
4.单击待操作的弹性云主机行的“操作>远程登录”按钮，远程连接弹性云主机实例。



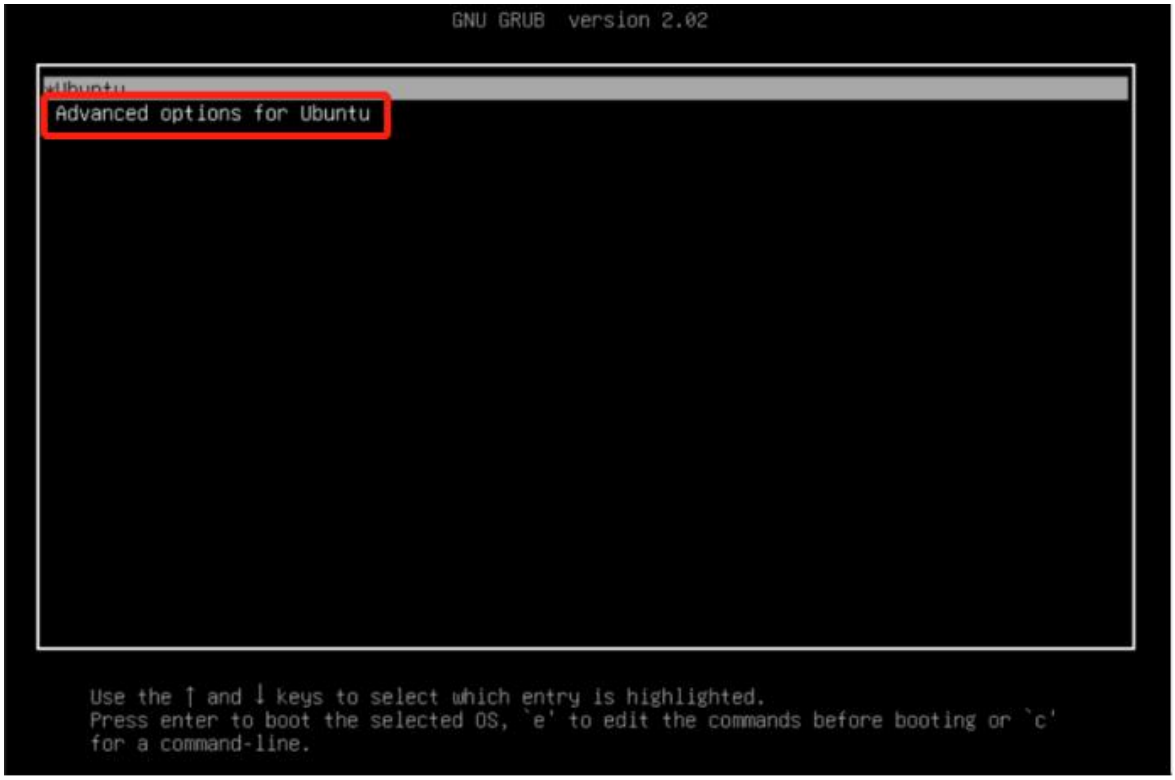
5.输入用户名 root，密码为购买弹性云主机时用户自定义的密码，登录成功之后如图：



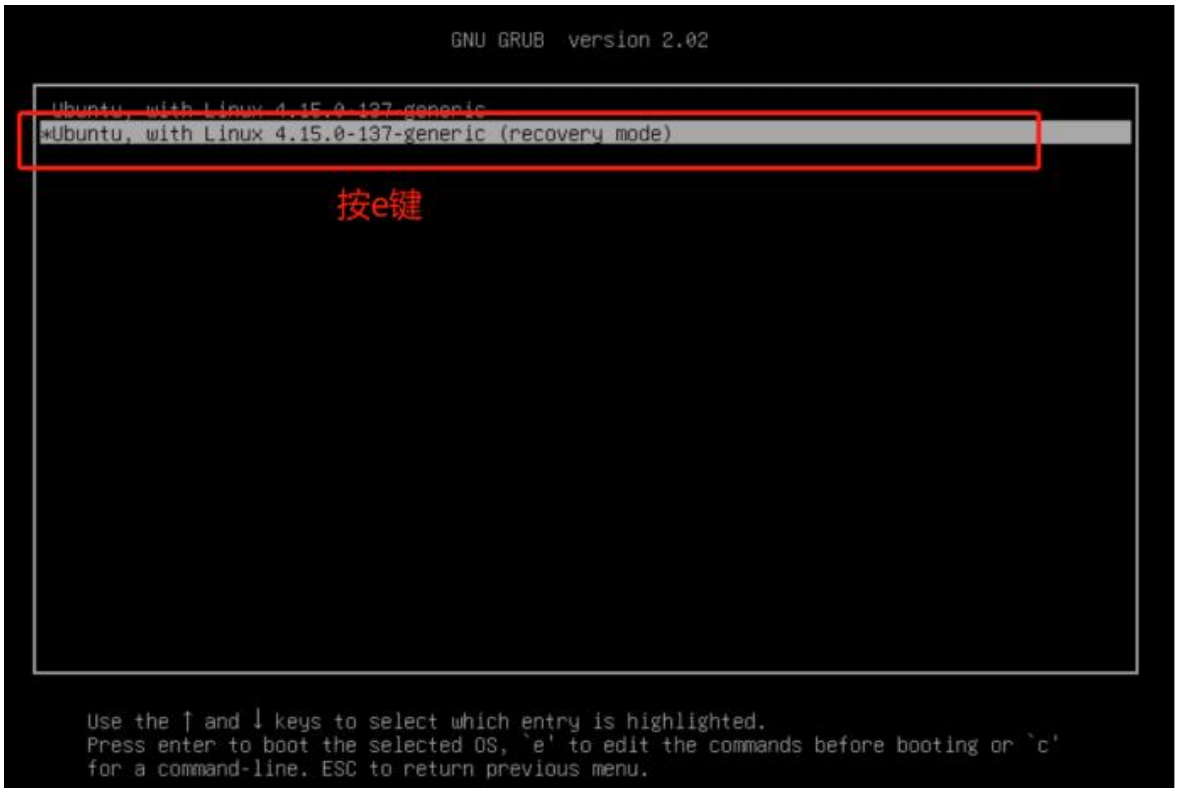
6.输入重启命令 reboot，并在重启过程中按 shift 键，进入 GRUB 界面，GRUB 界面示例如下。



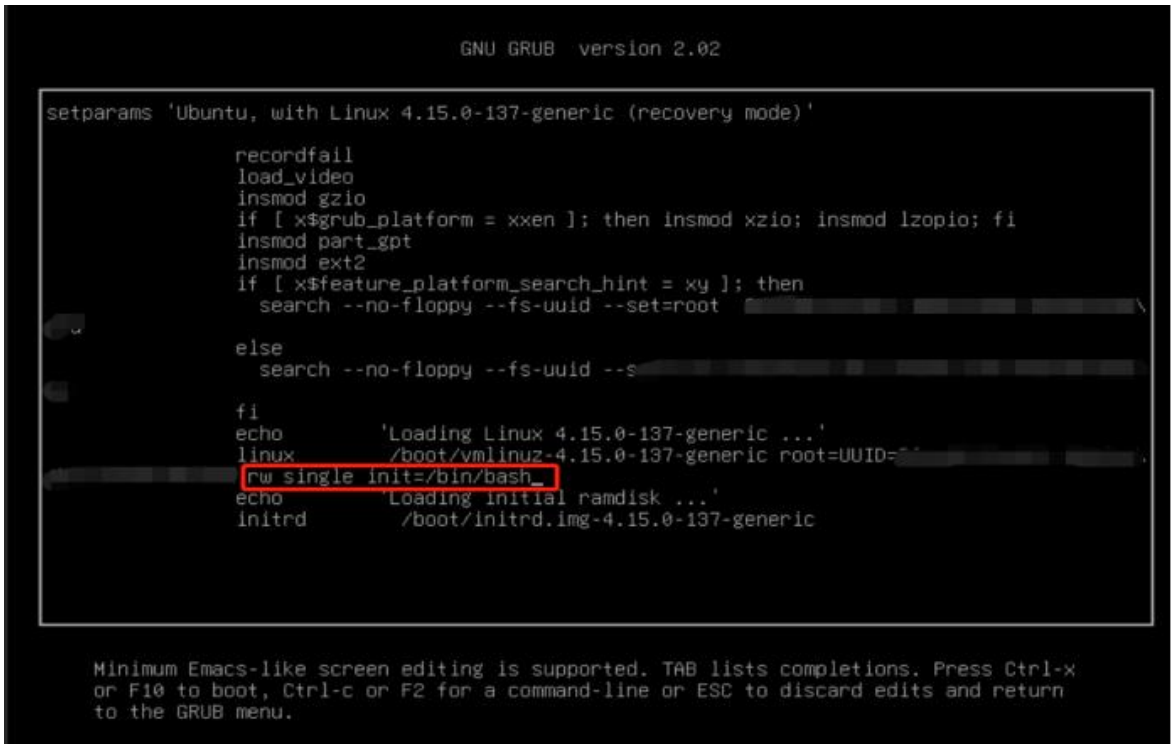
7.使用键盘方向键，向下移动选择 GRUB 页面第二行的高级选项，并按下键盘的 enter 键。



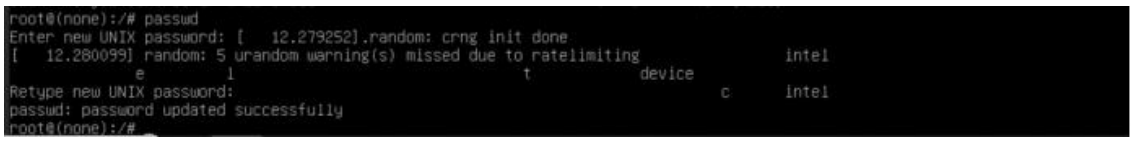
8.接下来选择新页面中第二行的恢复模式（recovery mode），并按 e 键编辑启动项。



9.进入编辑页面，使用键盘的方向键，移动光标向下至 linux 命令开头的一行，并在本行中将 ro 至末尾的内容删除并替换为 rw single init=/bin/bash。具体的修改信息如下图：



10.截至目前，进入单用户模式的配置已经基本完成，现在需要用户按下键盘的 ctrl+x 组合键或按 F10 键，系统会直接进入单用户模式。输入 passwd 命令重置系统密码，密码输入完成还需要进行密码的二次输入进行确认。示例如图所示。



至此，一台操作系统为 Ubuntu18.04 64 位的弹性云主机实例就进入到了单用户模式。

10.17 修改云主机默认远程端口

弹性云主机实例的默认远程端口是 22。这个端口是 SSH（Secure Shell）协议的默认端口，用于远程登录和管理云主机实例。通过 SSH 协议，您可以通过终端或 SSH 客户端与云主机进行安全的远程连接，并执行各种管理任务和操作。

修改 Windows 系统实例默认远程端口

本章以 Windows Server 2012 R2 为例介绍如何修改 Windows 系统实例默认远程端口。

1. 登录控制中心。

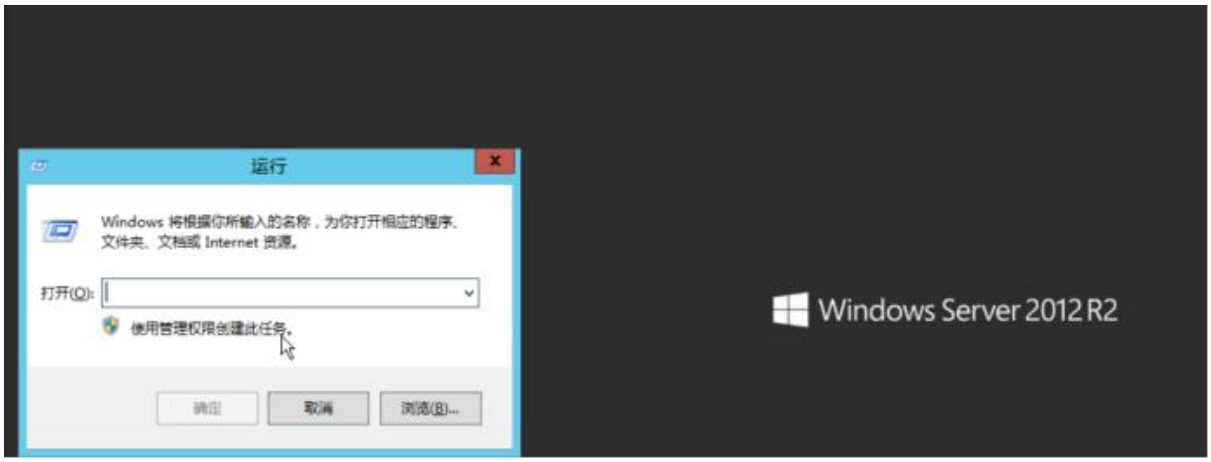
2. 选择区域 华东-华东 1。

3. 单击“计算>弹性云主机”，进入弹性云主机页面。

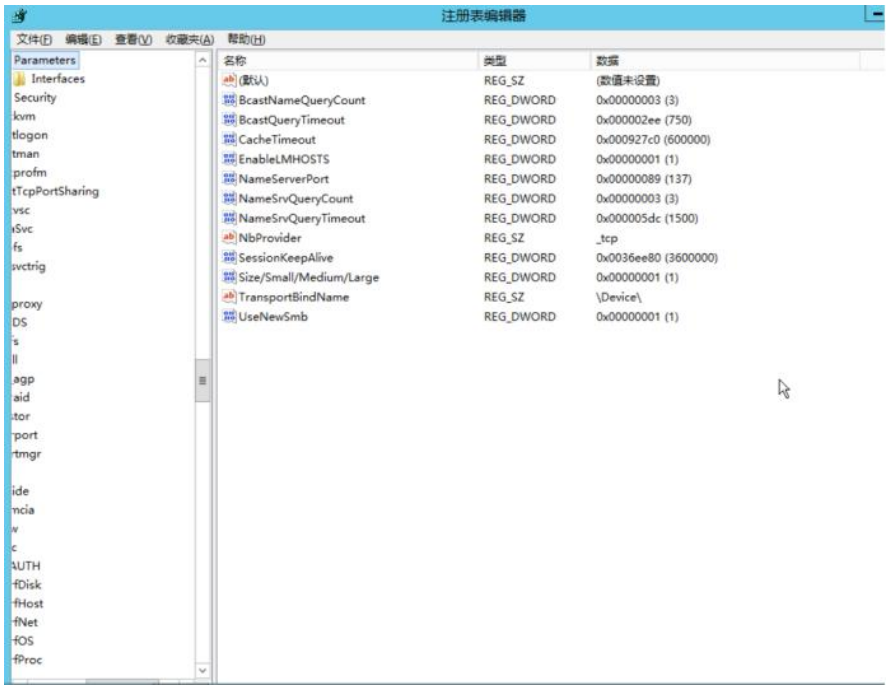
4. 参考登录 Windows 弹性云主机，远程登录待修改端口的 Windows 弹性云主机实例。

5. 修改注册表子项 PortNumber 的值。

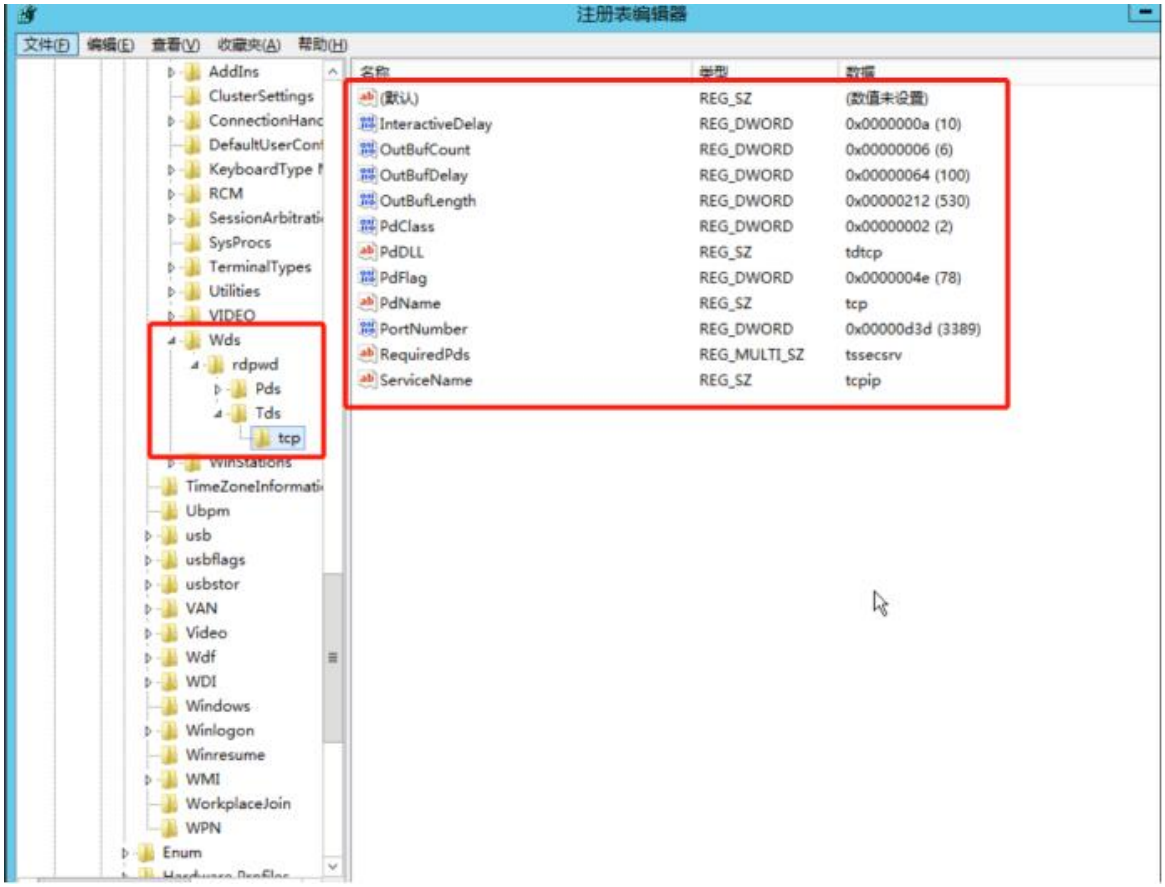
1）右键单击 Windows 徽标键，选择“运行”，启动运行窗口。



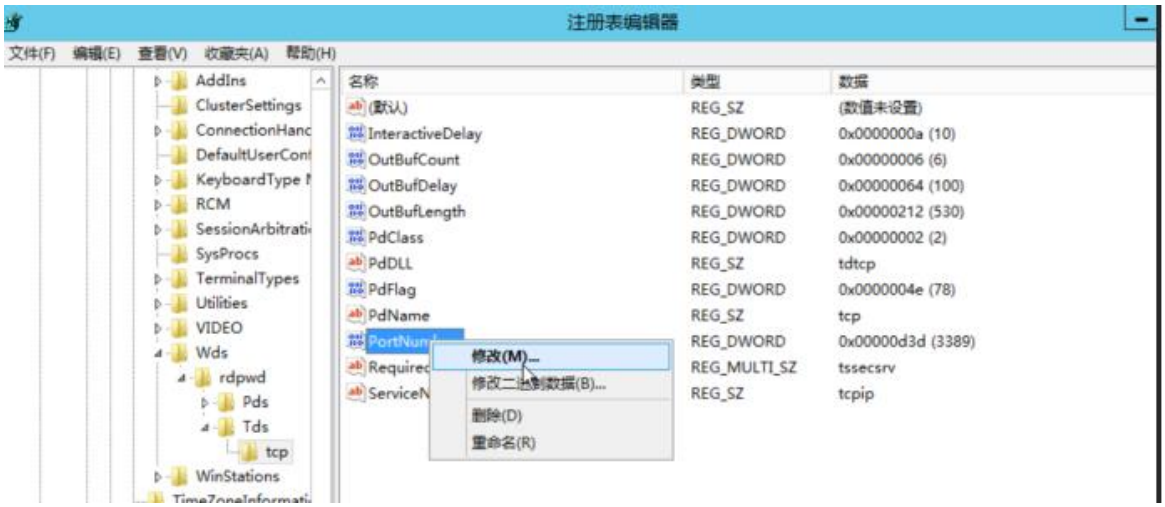
2）在运行窗口的文本框内输入 regedit.exe 后按回车键，打开“注册表编辑器”。



3）在左侧导航树中逐层选择 HKEY_LOCAL_MACHINE > System > CurrentControlSet > Control > Terminal Server > Wds > rdpwd > Tds > tcp，如下图所示：



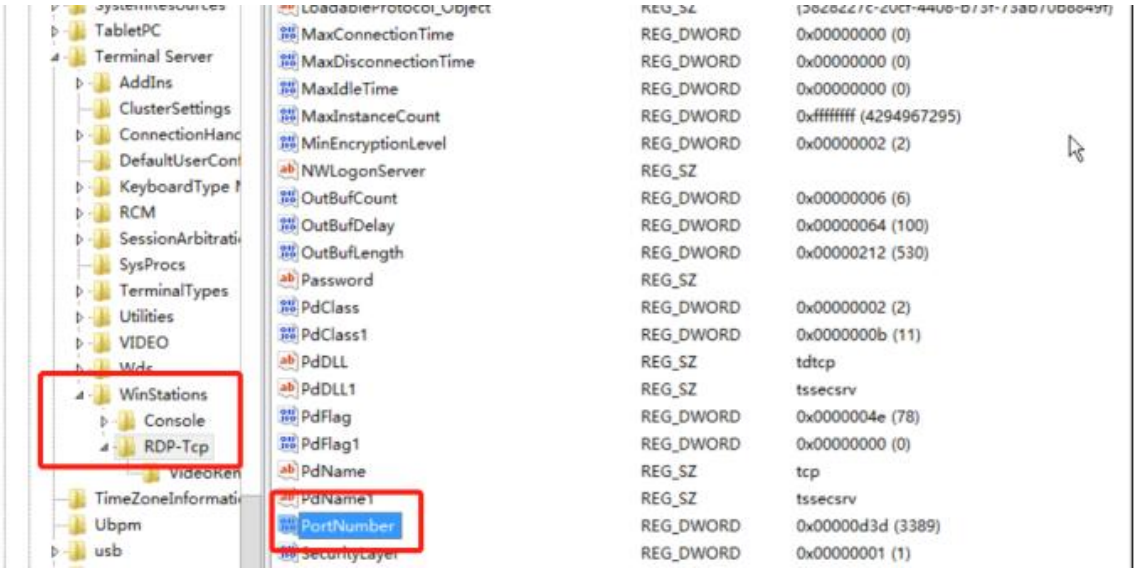
4）在列表中找到注册表子项 PortNumber 并右键单击，选择“修改”，进入修改窗口。



5）在弹出的修改窗口中，在数值数据的文本框中输入新的远程端口号，以 3398 为例。在基数框中勾选十进制，然后单击“确定”。



6）在左侧导航树中逐层选择 HKEY_LOCAL_MACHINE > System > CurrentControlSet > Control > Terminal Server > WinStations > RDP-Tcp。



7）在右侧列表中找到注册表子项 PortNumber 并右键单击，选择“修改”（可以使用键盘方向键向下寻找）。

8）在弹出的对话框中，在数值数据的文本框中输入新的远程端口号，在本示例中即 3398。在基数框中勾选十进制，然后单击“确定”。



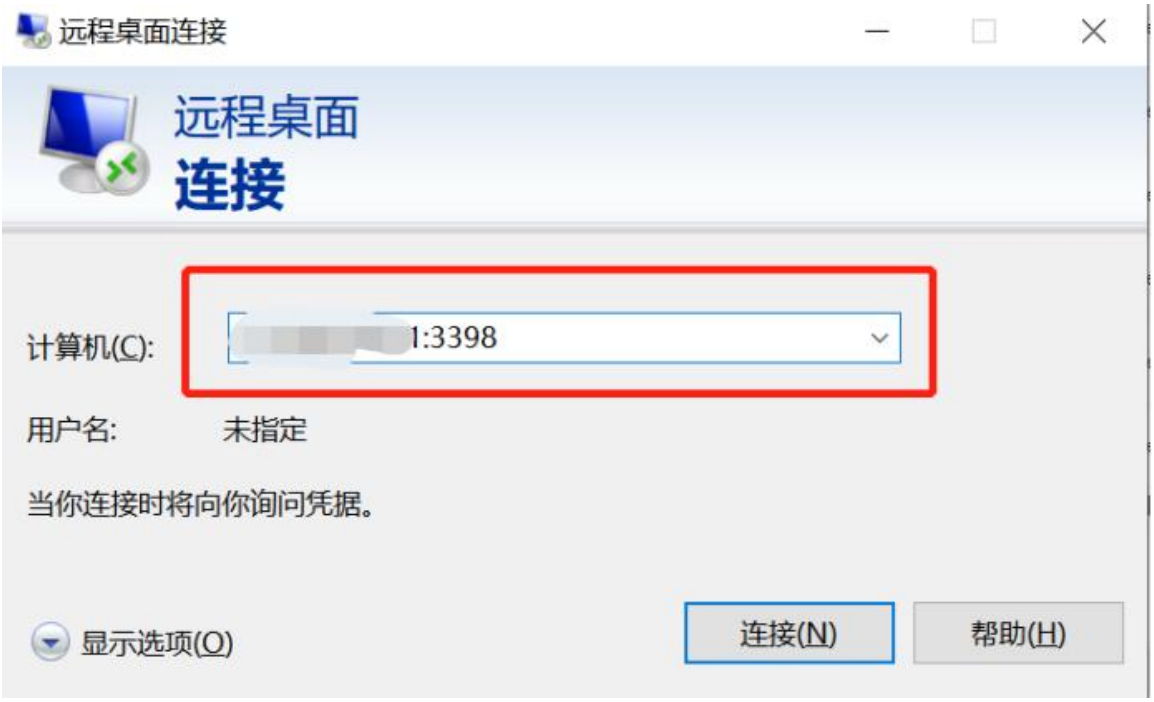
6. 在弹性云主机管理控制台中将此台云主机进行重启，具体如下图：



7. 为该实例添加安全组规则，允许新配置的 3398 远程端口进行连接。

具体操作，请参见配置安全组规则。

8. 最后使用远程桌面连接功能远程访问实例，在远程地址后面添加新端口号 3398 即可连接实例。



修改 Linux 系统实例默认远程端口

本章以 CentOS 8.0 64 位为例介绍如何修改 Linux 系统实例默认远程端口。

1. 登录控制中心。

2. 选择区域 华东-华东 1。

3. 单击“计算>弹性云主机”，进入弹性云主机页面。

4. 单击待修改的弹性云主机行的“操作>远程登录”按钮，远程连接弹性云主机实例。



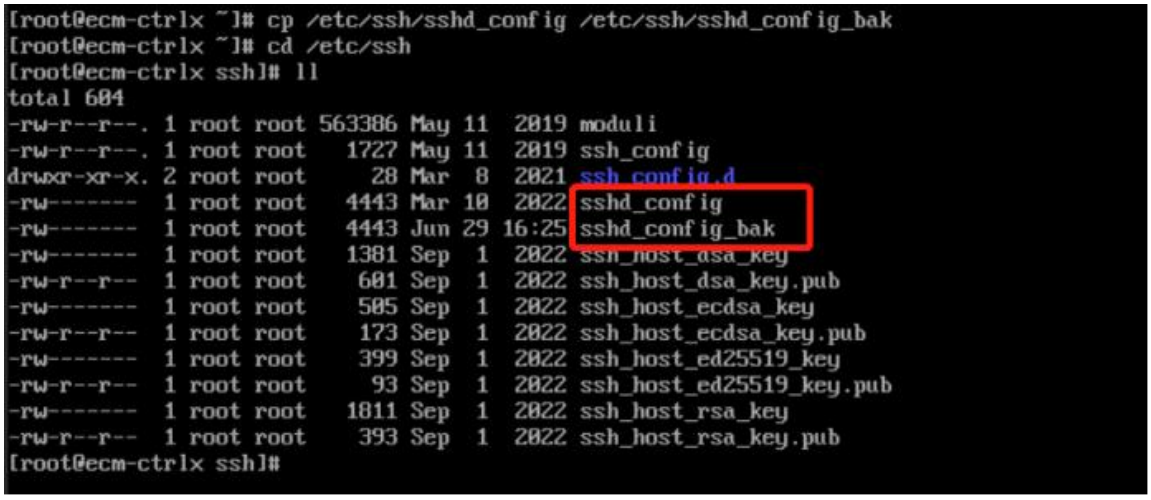
5. 输入用户名 root，密码为购买弹性云主机时用户自定义的密码，登录成功之后如图：



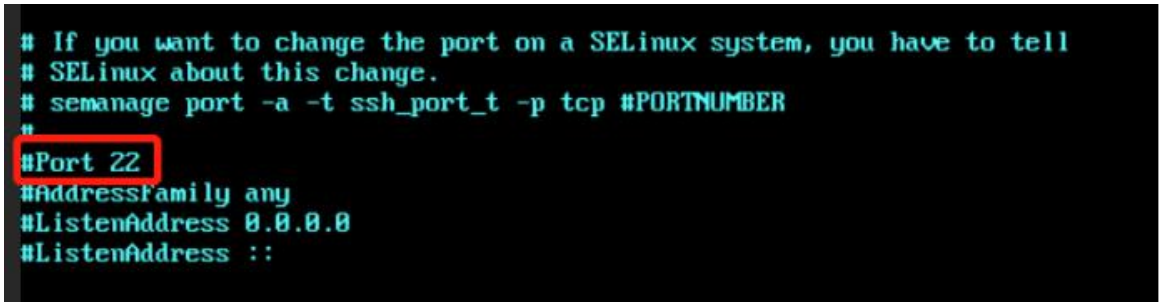
6. 因 sshd_config 是 Linux 中重要的配置文件，为了避免误操作所带来的故障，在运行前首先对 sshd_config 进行备份，请运行以下命令：

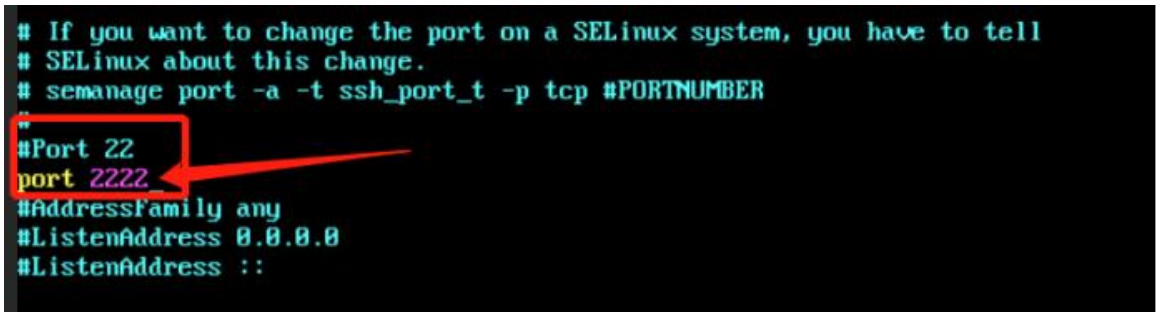
```
cp /etc/ssh/sshd_config /etc/ssh/sshd_config_bak
```

进入到/etc/ssh 路径下输入 ll 命令查看，具体信息如下图：



7. 修改 sshd 服务的端口号，因为已经在/etc/ssh 路径下，因此直接运行 vim sshd_config 编辑 sshd_config 配置文件，在键盘上按 i 键，进入编辑状态，添加新的远程服务端口，本节以 2222 端口为例。在 Port 22 下输入 Port 2222。





8. 在键盘上按 Esc 键，输入:wq 后保存并退出编辑模式。

9. 运行以下命令重启 sshd 服务。

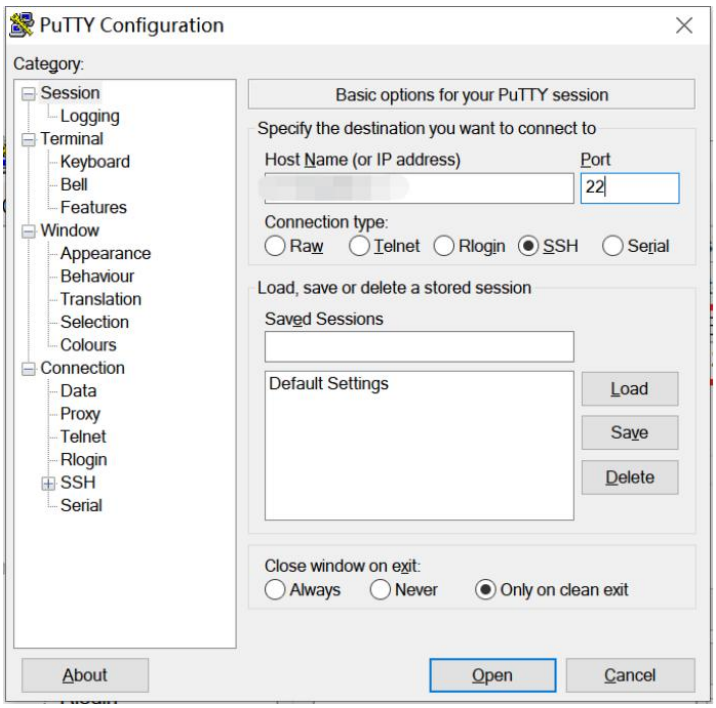
```
systemctl restart sshd
```

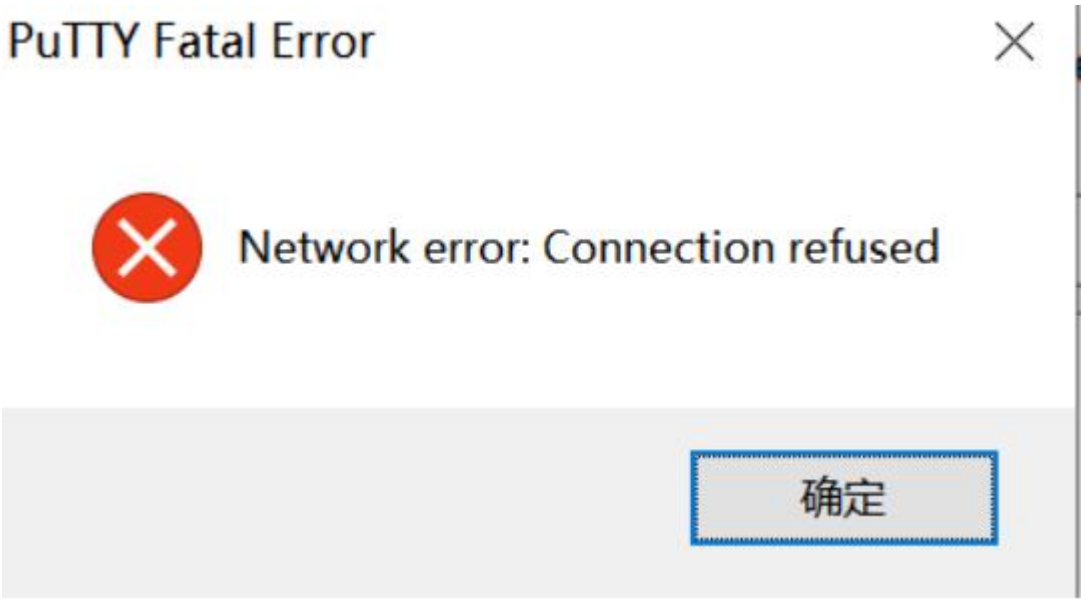


至此，此弹性云主机的远程登录默认端口已经从 22 改为了 2222。

功能验证：

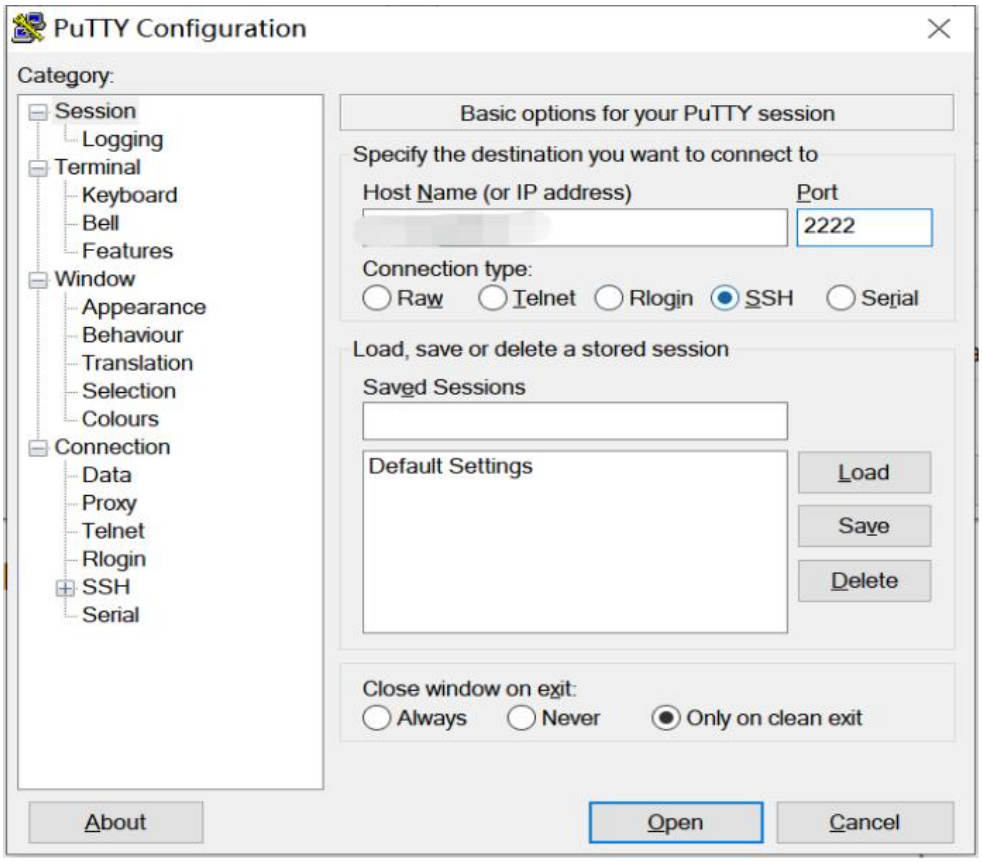
1. 打开远程连接工具 putty，输入此弹性云主机实例的弹性公网 IP，先默认使用 22 端口登录，可以看到连接被拒绝，网络已中断，说明目前 22 端口已经无法进行远程登录。

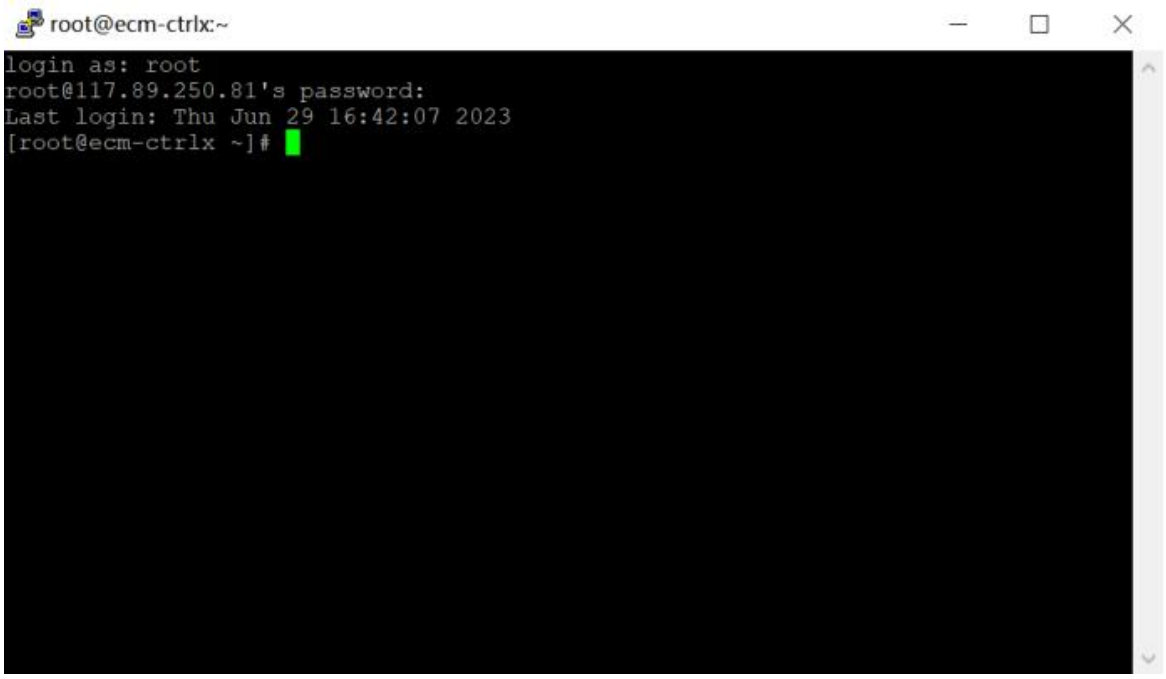




2.配置实例的安全组，放行 TCP 协议 2222 端口，具体操作，请参见配置安全组规则

3.配置完成之后，使用 SSH 工具连接新端口 2222，在 port 下输 2222，可看到能够成功连接。





10.18 设置 Windows 操作系统首选语言

前提条件

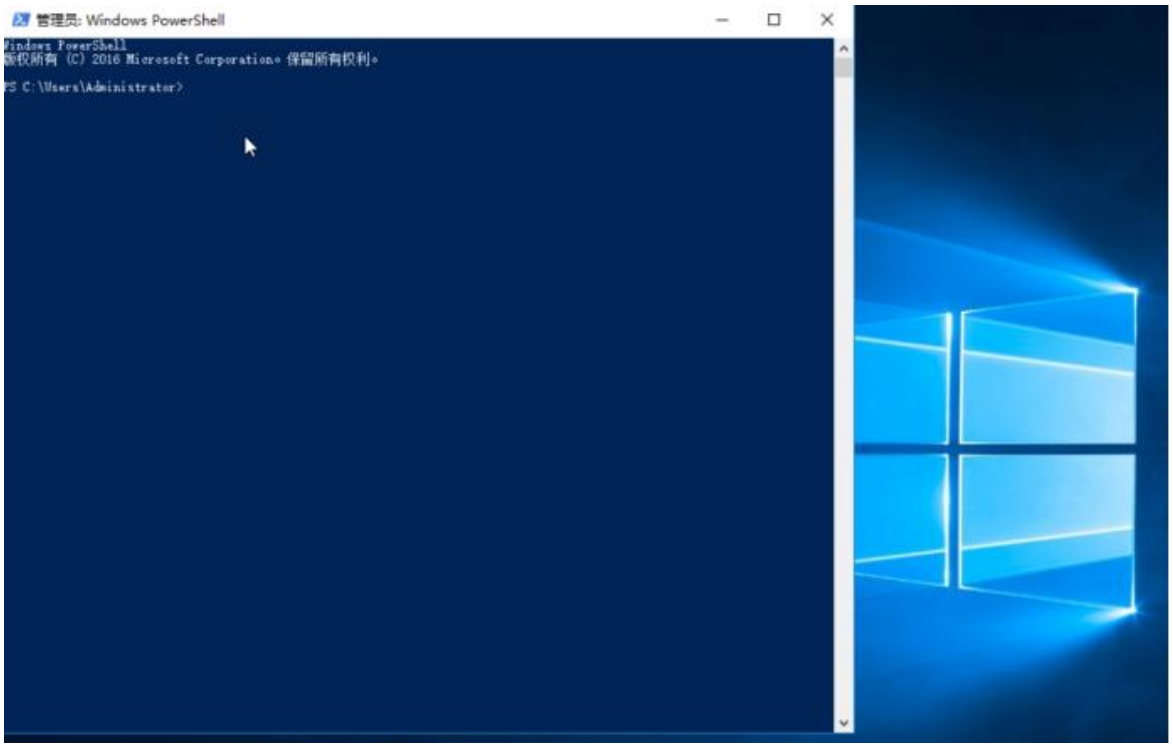
- 已经注册并登录天翼云账号，未完成的可参见注册天翼云账号。
- 已经在天翼云上购买了弹性云主机，且购买过程中镜像选择为 Windows 2016 数据中心版操作系统。

背景信息

天翼云云服务器 ECS 目前仅提供中文版的 Windows Server 公共镜像。如果您因工作需求而需要使用其他的语言版本，例如英文，那么您可以参考本章节为弹性云主机实例设置新的首选语言。

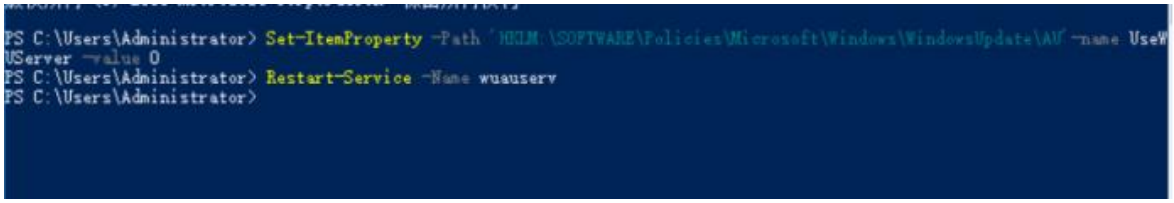
操作步骤

1. 登录控制中心。
2. 选择区域 华东-华东 1。
3. 单击“计算>弹性云主机”，进入弹性云主机页面。
4. 参考登录 Windows 弹性云主机，远程登录待设置操作系统首选语言的 Windows 弹性云主机实例。
5. 打开 PowerShell 模块。

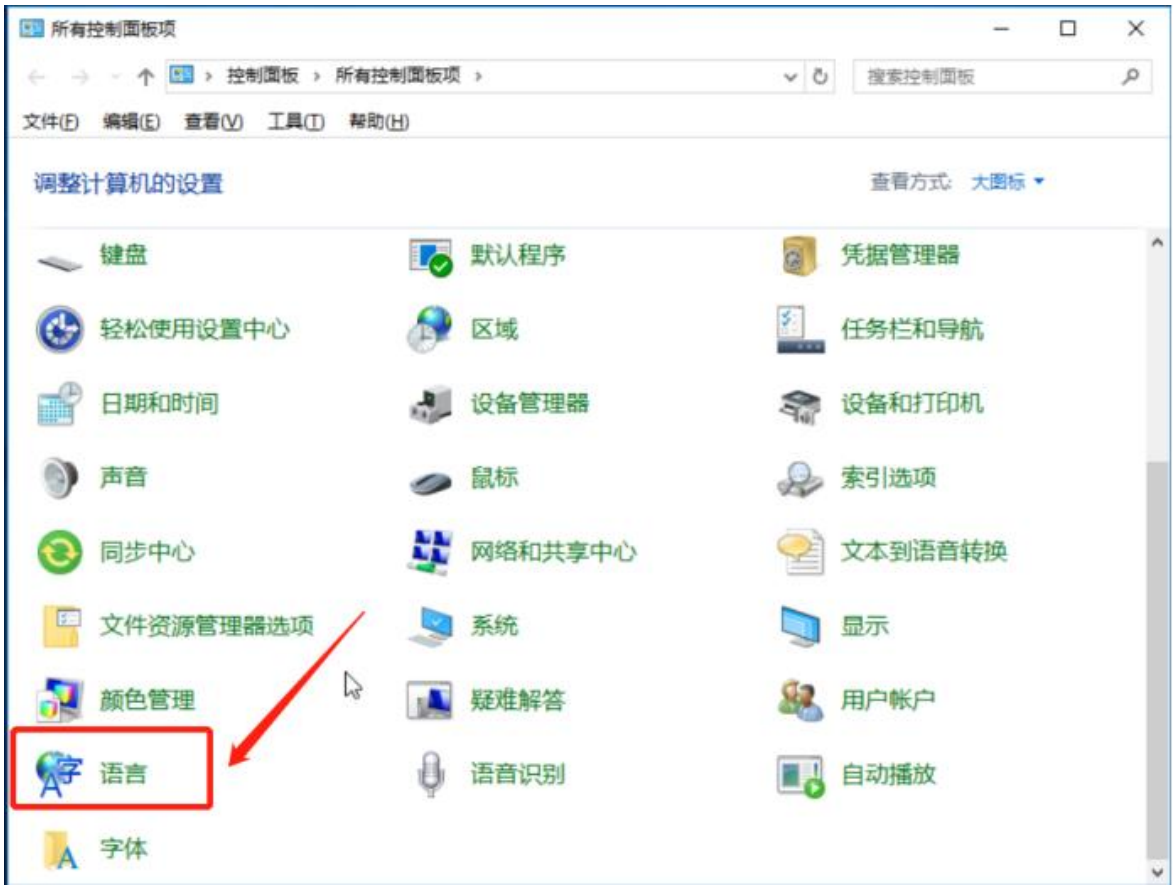


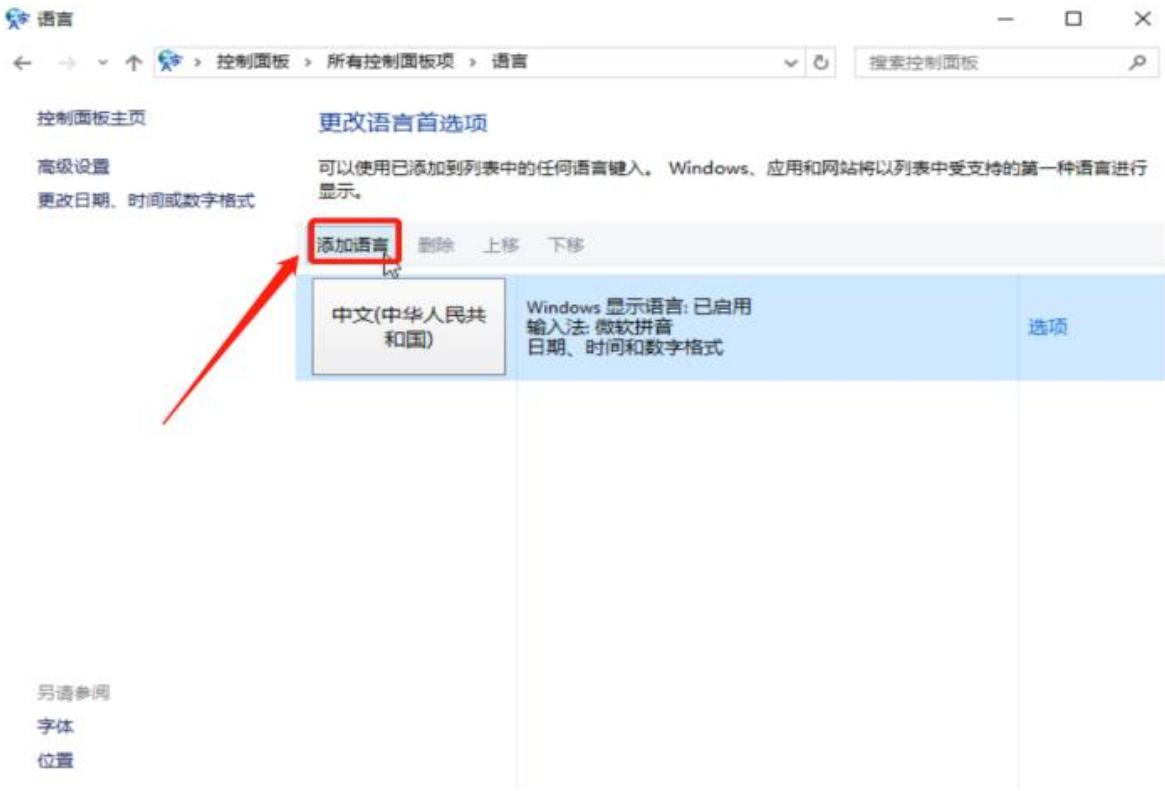
6.在操作之前首先需要临时禁用掉 WSUS（Windows Server Update Services）更新源，请您运行以下命令：

```
Set-ItemProperty -Path
'HKLM:\SOFTWARE\Policies\Microsoft\Windows\WindowsUpdate\AU' -Name
UseWUServer -Value ORestart-Service -Name wuauserv
```



7.找到控制面板，单击“语言” >“添加语言”。

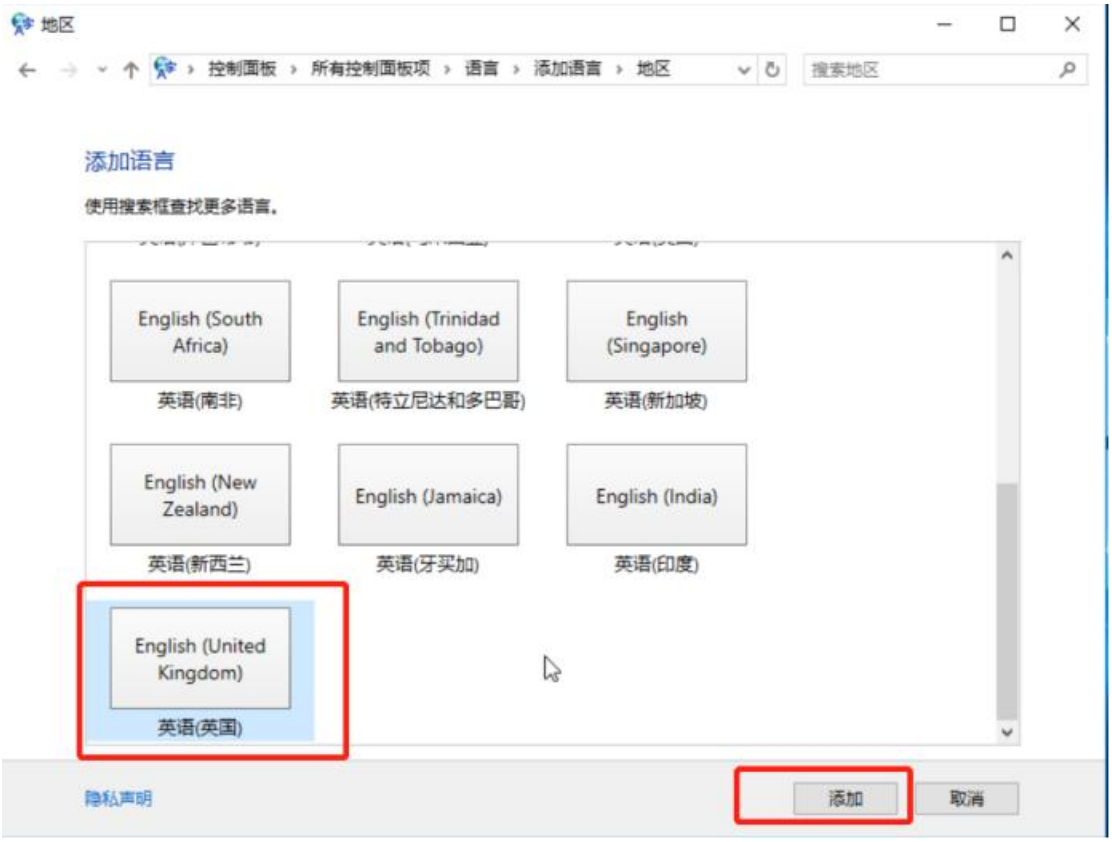

所有控制面板项
控制面板 › 所有控制面板项 ›
搜索控制面板
文件(F) 编辑(E) 查看(V) 工具(T) 帮助(H)
调整计算机的设置
查看方式: 大图标
键盘
默认程序
凭据管理器
轻松使用设置中心
区域
任务栏和导航
日期和时间
设备管理器
设备和打印机
声音
鼠标
索引选项
同步中心
网络和共享中心
文本到语音转换
文件资源管理器选项
系统
显示
颜色管理
疑难解答
用户帐户
语言
语音识别
自动播放
字体


语言
控制面板 › 所有控制面板项 › 语言
搜索控制面板
控制面板主页
高级设置
更改日期、时间或数字格式
更改语言首选项
可以使用已添加到列表中的任何语言键入。 Windows、应用和网站将以列表中受支持的第一种语言进行显示。
添加语言
删除
上移
下移
中文(中华人民共和国)
Windows 显示语言: 已启用
输入法: 微软拼音
日期、时间和数字格式
选项
另请参阅
字体
位置

8.在“添加语言”对话框中，选择一种语言，例如English(英语），单击下方的“打开”。



9.进入跳转窗口后，选择具体的英语区域，选择English（United kingdom），点击“添加”。



10. 添加之后，此语言会更新到语言列表。

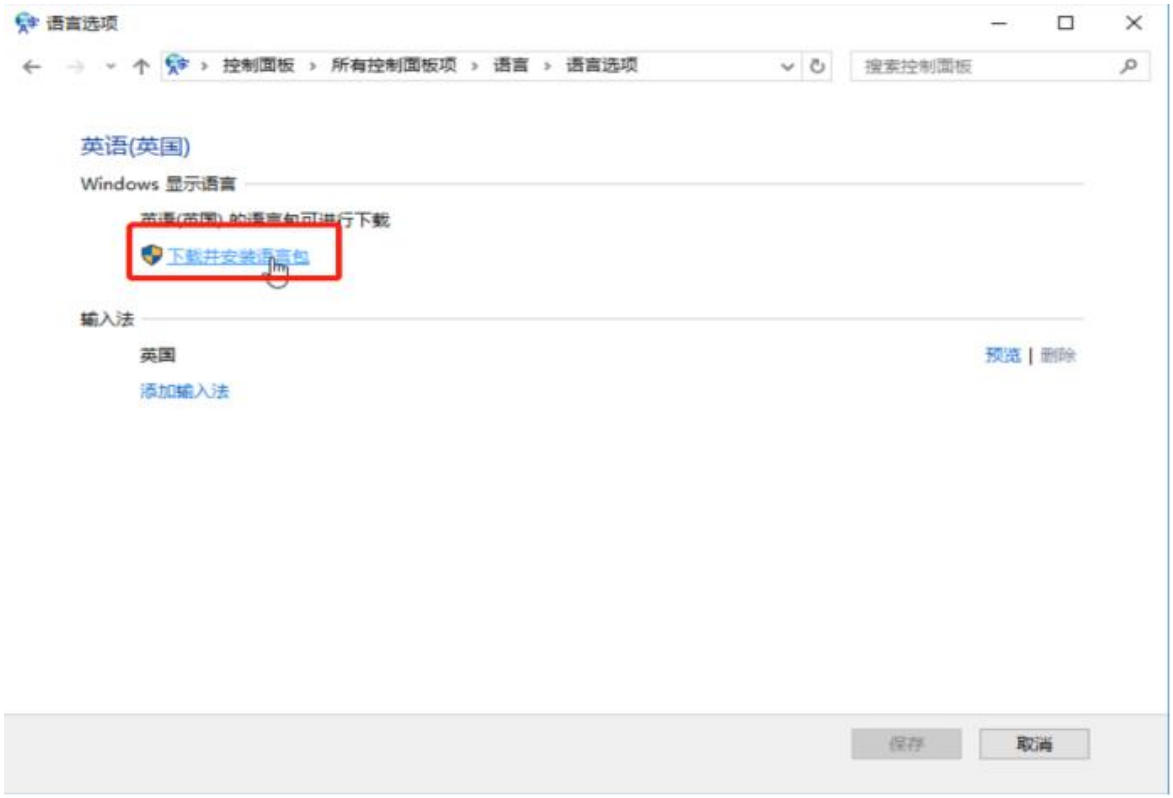


11.选中新增的 English，单击“上移”更改语言优先级，下图为已经上移之后的状态：

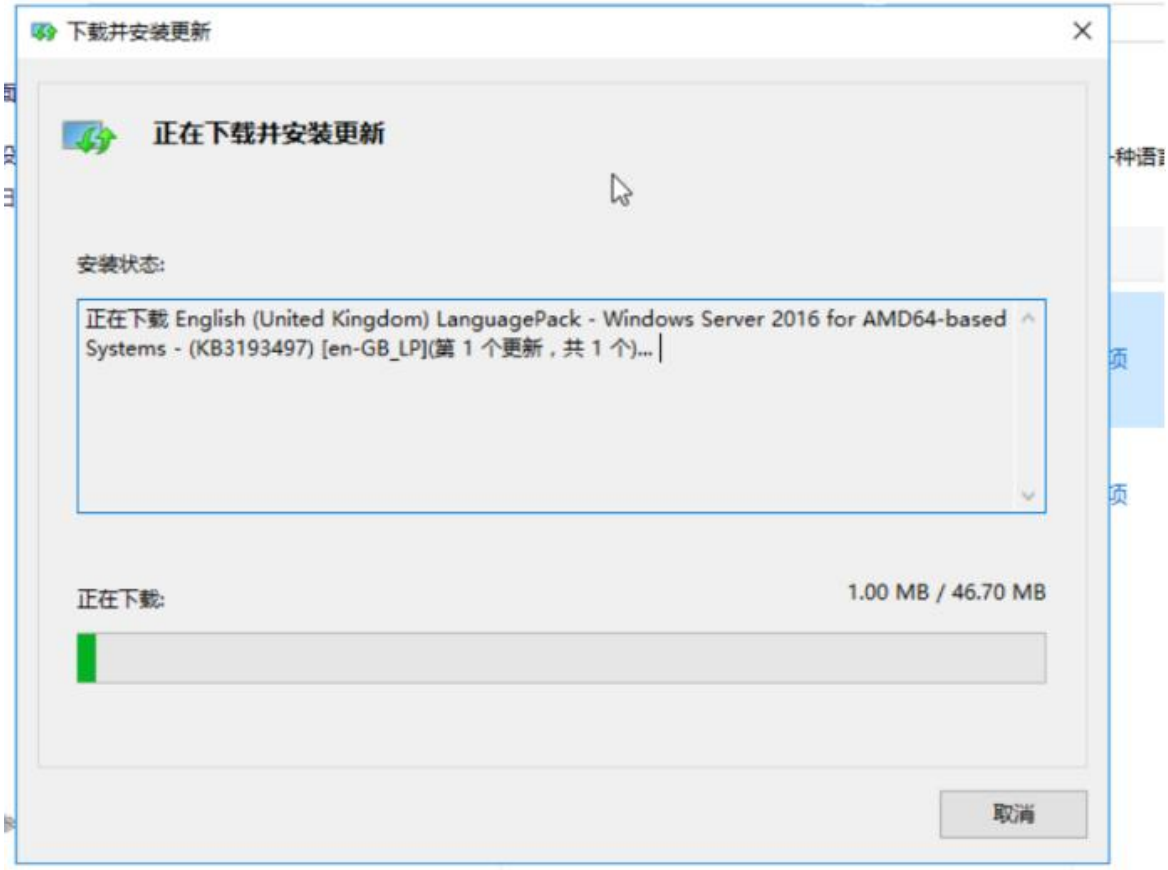


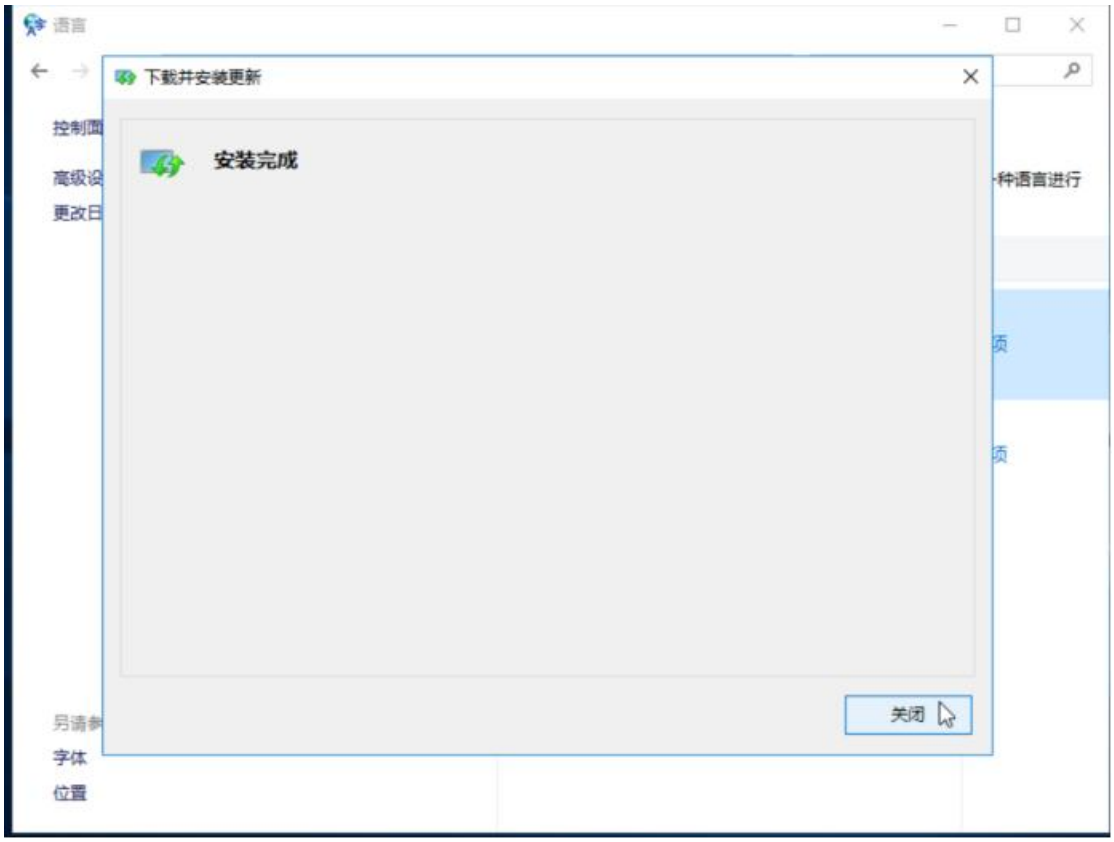
12. 单击所选语言右侧的“选项”按键，进入语言选项页面，单击“下载并安装语言包”，若没有此链接，请再等待三分钟进行更新。





13. 用户此时可以等待下载并安装完成。


下载并安装更新
正在下载并安装更新
安装状态:
正在下载 English (United Kingdom) LanguagePack - Windows Server 2016 for AMD64-based Systems - (KB3193497) [en-GB_LP](第 1 个更新，共 1 个)...
正在下载:
1.00 MB / 46.70 MB
取消


语言
下载并安装更新
安装完成
关闭
字体
位置

14. 在弹性云主机管理控制台中将此台云主机进行重启，具体如下图：

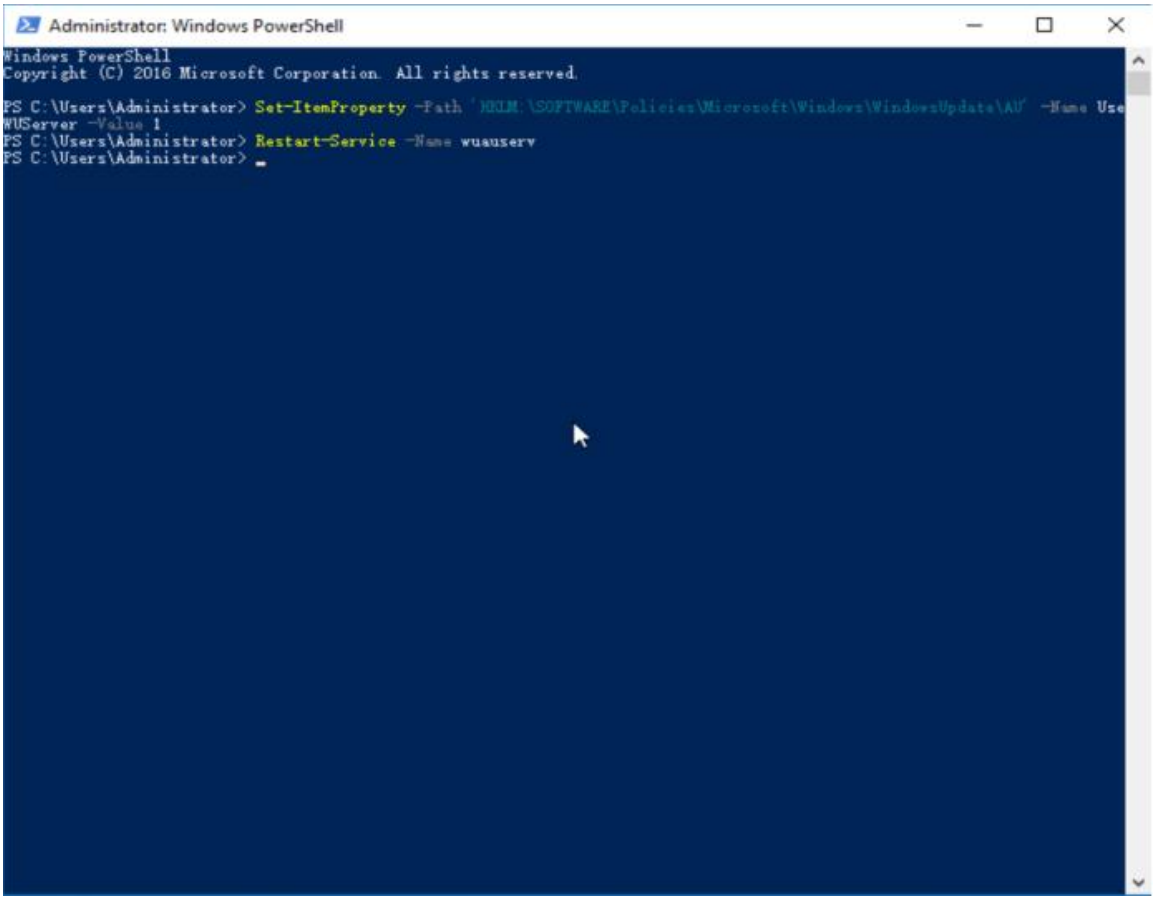


15. 再次连接 Windows 实例。显示语言更改为英语。



16. 打开 Windows PowerShell，运行以下命令重新启用 WSUS。

```
Set-ItemProperty -Path
'HKLM:\SOFTWARE\Policies\Microsoft\Windows\WindowsUpdate\AU' -Name
UseWUServer -Value 1Restart-Service -Name wuauserv
```



10.19ECS 安全组实践（入方向规则）

弹性云主机是天翼云提供的云服务器。安全组（Security Group）是用于设置云服务器实例的网络访问控制的虚拟防火墙，一般是作为流量访问白名单存在，以下是一些 ECS 安全组的最佳实践：

开放原则

- 默认拒绝所有流量：新建的安全组规则默认情况下拒绝所有入站和出站流量，这是一种良好的安全实践。
- 仅开放必要的端口和协议，避免使用 0.0.0.0/0 规则：根据应用需求，只开放必要的端口和协议，例如 HTTP（80 端口）、HTTPS（443 端口）等。避免开放不必要的端口，以减少攻击面。

分组原则

- 根据应用需求进行分组：根据应用程序或服务的需求，可以将安全组规则进行分组。例如，将 Web 服务器相关的规则放在一个 Web 层安全组中，数据库服务器相关的规则放在另一个 Database 层安全组中。这样可以更好地管理和组织安全组规则，暴露不同的出入规则和权限。
- 避免过度复杂化：尽量避免创建过多的安全组，以免管理和维护变得复杂。根据实际情况合理划分安全组，保持简洁和易于管理。

授权原则

- 基于最小权限原则：为安全组授权时，应遵循最小权限原则，仅授予实例所需的访问权限。只开放必要的端口和 IP 地址范围，避免授权过度，减少潜在的安全风险。
- 限制访问来源：根据实际需求，限制访问来源的 IP 地址或 IP 地址段。仅允许特定的 IP 地址或 IP 地址段访问云主机，以增加安全性。

安全原则

- 定期审查和更新安全组规则：定期审查安全组规则，确保只有必要的端口和 IP 地址被允许访问。如果某些规则不再需要或存在安全风险，及时进行更新和删除。
- 启用日志记录和监控：开启安全组的日志记录功能，对网络流量进行审计和监控。及时发现潜在的安全问题，并采取相应的措施。
- 使用网络安全组合 ACL：结合使用网络安全组合 ACL（Access Control List），可以在更细粒度的网络层面控制流量，提供额外的安全层。

变更安全组规则

用户可以参照配置安全组规则来进行安全组规则的配置，如果要变更安全组规则，需要注意变更安全组规则可能会影响用户实例间的网络通信，通常我们会选择放行必要的实例，再执行安全组策略收紧变更来保证必要的网络通信。

- 将需要互通访问的实例加入到一个新建的安全组，再执行变更操作。
- 如果授权类型为安全组访问，则将需要互通访问的对端实例所绑定的安全组 ID 添加为授权对象。
- 如果授权类型为地址段访问，则将需要互通访问的对端实例内网 IP 添加为授权对象。

10.20 搭建微信公众号后台

准备事项

- 申请微信公众号。

微信公众号申请链接：https://mp.weixin.qq.com/

- 购买天翼云弹性云主机服务。

如果没有天翼云账号，需先注册天翼云账号并完成实名认证。

本实践中，使用公共镜像 CentOS 7.4。

- 购买弹性 IP。

建议同时购买弹性 IP，后面需要在微信公众号上配置公网 IP 的地址。

操作步骤

基础软件安装

本实践中使用 Python+Web.py 组合完成微信公众号开发，需要安装或升级 Python、pip、Web.py 框架、WinSCP 软件版本。

1.升级默认 Python 版本；

CentOS 7.4 自带 Python 版本比较老，建议升级到 Python3。

1）查看 Python 版本，使用如下命令：

```
python -version
```

```
[root@ecs- ~]# python --version
Python 2.7.5
[root@ecs- ~]#
```

2）下载 Python 安装包，这里以 Python 3.6.0 版本为例，使用命令：

```
wget https://www.python.org/ftp/python/3.6.0/Python-3.6.0a1.tar.xz
```

```
[root@ecs- ~]# wget https://www.python.org/ftp/python/3.6.0/Python-3.6.0a1.tar.xz
--2020-12-28 09:25:56--  https://www.python.org/ftp/python/3.6.0/Python-3.6.0a1.tar.xz
Resolving www.python.org (www.python.org)...  2a04:4e42:1a::223
Connecting to www.python.org (www.python.org)|  |:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 15328032 (15M) [application/octet-stream]
Saving to: 'Python-3.6.0a1.tar.xz'

100%[=======================================================================================

2020-12-28 09:25:57 (12.7 MB/s) - 'Python-3.6.0a1.tar.xz' saved [15328032/15328032]

[root@ecs- ~]#
```

3）解压安装包，使用如下命令：

```
tar xvf Python-3.6.0a1.tar.xz
```

```
configure: creating ./config.status
config.status: creating Makefile.pre
config.status: creating Modules/Setup.config
config.status: creating Misc/python.pc
config.status: creating Misc/python-config.sh
config.status: creating Modules/ld_so_aix
config.status: creating pyconfig.h
creating Modules/Setup
creating Modules/Setup.local
creating Makefile
```

如果出现“configure: error: no acceptable C compiler found in $PATH”异常提示，是因为未安装合适的编译器。

解决方法：

执行以下命令，安装/升级 gcc 及其他依赖的包。

```
sudo yum install gcc-c++
```

并在随后提示安装包是否 OK 时，输入 y 并回车。出现如下图提示，说明依赖的包安装成功。

```
Installed:
  gcc-c++.x86_64 0:4.8.5-44.el7

Dependency Installed:
  libstdc++-devel.x86_64 0:4.8.5-44.el7

Dependency Updated:
  cpp.x86_64 0:4.8.5-44.el7          gcc.x86_64 0:4.8.5-44.el7
```

在编译器安装完成后，重新执行 ./configure 命令。

4）执行命令：

```
make && make install
```

执行成功。但提示 pip 错误，原因是我的系统中少了 openssl-devel 包，可以先忽略。

```
rm -f /usr/local/bin/pyvenv
(cd /usr/local/bin; ln -s pyvenv-3.6 pyvenv)
if test "x" != "x" ; then \
        rm -f /usr/local/bin/python3-32; \
        (cd /usr/local/bin; ln -s python3.6-32 python3-32) \
fi
rm -f /usr/local/share/man/man1/python3.1
(cd /usr/local/share/man/man1; ln -s python3.6.1 python3.1)
if test "xupgrade" != "xno"  ; then \
        case upgrade in \
                upgrade) ensurepip="--upgrade" ;; \
                install|*) ensurepip="" ;; \
        esac; \
         ./python -E -m ensurepip \
                $ensurepip --root=/ ; \
fi
Ignoring ensurepip failure: pip 8.1.1 requires SSL/TLS
```

5）查看 Python3 版本，使用命令：

python3 -version

```
[root@ecs-      Python-3.6.0a1]# python3 --version
Python 3.6.0a1
```

6）执行命令：

```
python3
```

出现如下提示，则说明 Python3 安装成功。

```
[root@ecs-      Python-3.6.0a1]# python3
Python 3.6.0a1 (default, Dec 18 2020, 15:45:57)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-44)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

2.升级默认 pip 版本；

pip 是通用的 Python 包管理工具。提供了对 Python 包的查找、下载、安装、卸载功能。Python3 安装成功后自带 pip3，但版本比较老，建议升级到 pip 最新版本。同时前面安装 python3 提示“Ignoring ensurepip failure: pip 8.1.1 requires SSL/TLS”错误，导致 pip 未成功安装，所以需要重新安装 pip。

1）安装 openssl-devel 包，使用命令：

```
yum install openssl-devel -y
```

```
Installed:
  openssl-devel.x86_64 1:1.0.2k-21.el7_9

Dependency Installed:
  keyutils-libs-devel.x86_64 0:1.5.8-3.el7     krb5-devel.x86_64 0:1.15.1-50.el7
  libselinux-devel.x86_64 0:2.5-15.el7         libsepol-devel.x86_64 0:2.5-10.el7
  zlib-devel.x86_64 0:1.2.7-18.el7

Dependency Updated:
  openssl.x86_64 1:1.0.2k-21.el7_9

Complete!
```

2）执行命令：

```
make && make install
```

出现如下提示说明 pip 安装成功。

```
Collecting setuptools
Collecting pip
Installing collected packages: setuptools, pip
Successfully installed pip-8.1.1 setuptools-20.10.1
```

3）升级 pip3，使用命令：

```
pip3 install --upgrade pip
```

出现如下提示说明升级 pip 到最新版本了。

```
[root@ecs-      Python-3.6.0a1]# pip3 install --upgrade pip
Collecting pip
  Downloading https://files.pythonhosted.org/packages/54/eb/4a36
    100% |████████████████████████████████| 1.5MB 32kB/s
Installing collected packages: pip
  Found existing installation: pip 8.1.1
    Uninstalling pip-8.1.1:
      Successfully uninstalled pip-8.1.1
Successfully installed pip-20.3.3
```

3. 安装 Web.py 框架；

Web.py 官方教程地址：http://webpy.org/，使用如下命令安装 web.py：

```
pip3 install web.py==0.40.dev0
```

```
[root@ecs-c438 Python-3.6.0a1]# pip3 install web.py==0.40.dev0
Collecting web.py==0.40.dev0
  Downloading web.py-0.40.dev0.tar.gz (116 kB)
     |████████████████████████████████| 116 kB 76 kB/s
Using legacy 'setup.py install' for web.py, since package 'wheel'
Installing collected packages: web.py
    Running setup.py install for web.py ... done
Successfully installed web.py-0.40.dev0
```

4.安装 WinSCP；

通常情况下，我们在本地 Windows 操作系统上编辑代码，完成后再上传至 ECS 上（CentOS Linux 系统）。WinSCP 是一个 Windows 环境下使用的 SSH 的开源图形化 SFTP 客户端，同时支持 SCP 协议。它的主要功能是在本地与远程计算机间安全地复制文件，并且可以直接编辑文件。

WinSCP 安装链接：https://winscp.net/eng/docs/lang:chs

上传代码

1.新建 main.py 文件，复制如下代码：

```
# -*- coding: utf-8 -*-# filename: main.pyimport webfrom handle import Handle

urls = (

    '/wx', 'Handle',

)if __name__ == '__main__':

    app = web.application(urls, globals())

    app.run()
```

2.新建 handle.py 文件，复制如下代码：

```
# -*- coding: utf-8 -*-# filename: handle.pyimport hashlibimport webimport receiveimport timeimport osclass Handle(object):

    def __init__(self):
```

```
        self.app_root = os.path.dirname(__file__)
        self.templates_root = os.path.join(self.app_root, 'templates')
        self.render = web.template.render(self.templates_root)
    def GET(self):
        try:
            data = web.input()
            if len(data) == 0:
                return "hello, this is handle view"
            signature = data.signature
            timestamp = data.timestamp
            nonce = data.nonce
            echostr = data.echostr
            token = "此处内容与公众号基本配置里 Token 字段取值保持一致"
            list = [token, timestamp, nonce]
            list.sort()
            s = list[0] + list[1] + list[2]
            hashcode = hashlib.sha1(s.encode('utf-8')).hexdigest()
            print( "handle/GET func: hashcode, signature: ", hashcode,
signature)
            if hashcode == signature:
                return echostr
            else:
```

```
                return echostr
        except (Exception) as Argument:
            return Argument
    def POST(self):
        try:
            webData = web.data()
            print("Handle Post webdata is:\n", webData)
            #打印消息体日志
            recMsg = receive.parse_xml(webData)
            if isinstance(recMsg, receive.Msg) and recMsg.MsgType ==
'text':
                toUser = recMsg.FromUserName
                fromUser = recMsg.ToUserName
                content = "欢迎" + str(recMsg.Content)
                print('Reply message info:\n')
                print('toUser =', toUser)
                print('fromUser = ', fromUser)
                print('content = ', content)
                return self.render.reply_text(toUser, fromUser,
int(time.time()), content)
            else:
                print("不支持的消息类型：",recMsg.MsgType)
                return "success"
```

```
        except (Exception) as Argment:
            return Argment
```

3.新建 receive.py 文件，复制如下代码：

```
# -*- coding: utf-8 -*-# filename: receive.pyimport
xml.etree.ElementTree as ETdef parse_xml(web_data):
    if len(web_data) == 0:
        return None
    xmlData = ET.fromstring(web_data)
    msg_type = xmlData.find('MsgType').text
    if msg_type == 'text':
        return TextMsg(xmlData)
    elif msg_type == 'image':
        return ImageMsg(xmlData)
    elif msg_type == 'location':
        return LocationMsg(xmlData)
    elif msg_type == 'event':
        return EventMsg(xmlData)class Event(object):
    def __init__(self, xmlData):
        self.ToUserName = xmlData.find('ToUserName').text
        self.FromUserName = xmlData.find('FromUserName').text
        self.CreateTime = xmlData.find('CreateTime').text
        self.MsgType = xmlData.find('MsgType').text
```

```
        self.Eventkey = xmlData.find('EventKey').textclass
Msg(object):
    def __init__(self, xmlData):
        self.ToUserName = xmlData.find('ToUserName').text
        self.FromUserName = xmlData.find('FromUserName').text
        self.CreateTime = xmlData.find('CreateTime').text
        self.MsgType = xmlData.find('MsgType').text
        self.MsgId = xmlData.find('MsgId').textclass TextMsg(Msg):
    def __init__(self, xmlData):
        Msg.__init__(self, xmlData)
        self.Content = xmlData.find('Content').textclass
ImageMsg(Msg):
    def __init__(self, xmlData):
        Msg.__init__(self, xmlData)
        self.PicUrl = xmlData.find('PicUrl').text
        self.MediaId = xmlData.find('MediaId').textclass
LocationMsg(Msg):
    def __init__(self, xmlData):
        Msg.__init__(self, xmlData)
        self.Location_X = xmlData.find('Location_X').text
        self.Location_Y = xmlData.find('Location_Y').textclass
EventMsg(Msg):
    def __init__(self, xmlData):
```

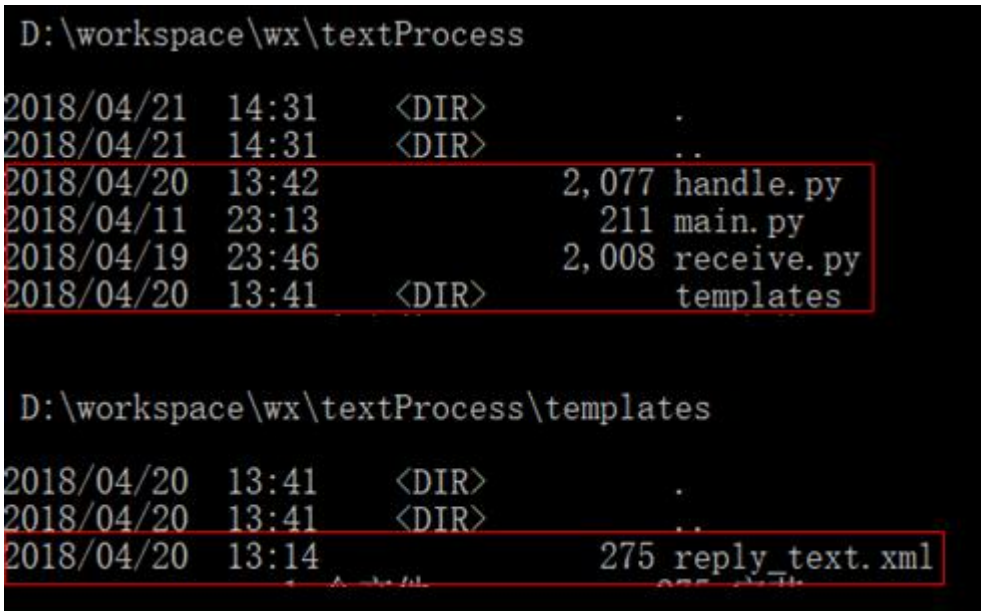
```
        Event.__init__(self, xmlData)

        self.Event = xmlData.find('Event').text
```

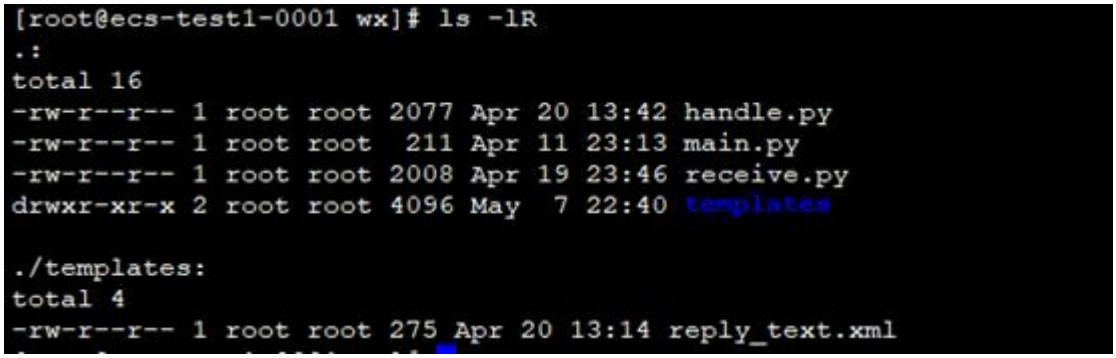
4.新建 templates 文件夹，在文件夹下新建 reply_text.xml 文件，复制如下代码：

```
$def with
(toUser,fromUser,createTime,content)<xml><ToUserName><![CDATA[$toUser]
]></ToUserName><FromUserName><![CDATA[$fromUser]]></FromUserName><Cre
ateTime>$createTime</CreateTime><MsgType><![CDATA[text]]></MsgType><C
ontent><![CDATA[$content]]></Content></xml>
```

5.最终本地代码文件形成如下：



6.通过 WinSCP 工具将上述文件与目录上传至 ECS 指定目录下：



启动服务

使用如下命令启动服务：

```
python3 main.py 80
```

```
[root@ecs-test1-0001 wx]# python3 main.py 80
http://0.0.0.0:80/
```

启用开发者模式

1.登录微信公众平台，选择“开发 > 基本配置”，单击“修改配置”。
2.填写配置信息，单击“提交”。

- URL：https://ECS 的弹性公网 IP/wx，不用添加 80 端口。
- Token：需要与 handle.py 中对应 token 取值完全一致。
- EncodingAESKey：随机生成。
- 消息加解密方式：此为示例，选择简单的“明文模式”。

3.验证 token 成功，单击“启用”。

说明：如果 token 验证失败，请检查 Token 配置与 handle.py 中 GET 消息处理代码是否一致。

验证

使用微信关注公众号，任意发送一条文本消息，看是否能够收到回复。如能收到回复则表明系统处理正常。

10.21 网络性能测试的最佳实践

iperf 可用于测试两个计算机之间的网络带宽和延迟。它是一个开源项目，提供了用于 Windows、Linux 和其他操作系统的版本。通过在两台计算机之间运行 iPerf 服务器和客户端，您可以测量它们之间的网络性能。iPerf 客户端发送数据流到服务器，并且服务器测量接收到的数据量和速度，支持在 TCP 和 UDP 协议上运行。

本文将详细介绍如何使用 iperf3 工具，测试弹性云主机间网络性能。内容主要包括“测试前准备”、“TCP 带宽测试”、“UDP PPS 测试”。

安装使用

```
[root@ss-daliu1 ~]# sudo yum install iperf3

[root@ss-daliu2 ~]# sudo yum install iperf3
```

参数说明

iperf 常用参数中文说明：

（1）-s, -server：iperf 服务器模式，默认启动的监听端口为 5201，eg：iperf -s

（2）-c, -client host：iperf 客户端模式，host 是 server 端地址，eg：iperf -c 222.35.11.23

（3）-i，-interval：指定每次报告之间的时间间隔，单位为秒，eg：iperf3 -c 192.168.12.168 -i 2

（4）-p，-port：指定服务器端监听的端口或客户端所连接的端口，默认是 5001 端口。

（5）-u，-udp：表示采用 UDP 协议发送报文，不带该参数表示采用 TCP 协议。

（6）-l，-len：设置读写缓冲区的长度，(default 128 KB for TCP, dynamic or 1 for UDP)。通常测试 PPS 的时候该值为 16，测试 BPS 时该值为 1400。

（7）-b，-bandwidth [K|M|G]：指定 UDP 模式使用的带宽，单位 bits/sec，默认值是 1 Mb/s。

（8）-t，-time：指定数据传输的总时间，即在指定的时间内，重复发送指定长度的数据包。默认 10 秒。

（9）-A：CPU 亲和性，可以将具体的 iperf3 进程绑定对应编号的逻辑 CPU，避免 iperf 进程在不同的 CPU 间调度。

通用参数（Server 端和 Client 端共用）：

(1）-f，-format [k|m|g|K|M|G]：指定带宽输出单位，“[k|m|g|K|M|G]”分别表示以 Kb，Mb，Gb，KBytes，MBytes,GBytes 显示输出结果，默认 Mb，eg：iperf3 -c 192.168.12.168 -f M

（2）-p，-port：指定服务器端监听的端口或客户端所连接的端口，默认是 5001

端口。

（3）-i，－interval：指定每次报告之间的时间间隔，单位为秒，eg：iperf3 -c 192.168.12.168 -i 2

（4）-F：指定文件作为数据流进行带宽测试。例如：iperf3 -c 192.168.12.168 -F web-ixdba.tar.gz

Server 端专用参数：

（1）-s,－server：iperf 服务器模式，默认启动的监听端口为 5201，eg：iperf -s

（2）-c,－client host：如果 iperf 运行在服务器模式，并且用-c 参数指定一个主机，那么 iperf 将只接受指定主机的连接。此参数不能工作于 UDP 模式。

（3）-D：Unix 平台下将 Iperf 作为后台守护进程运行。在 Win32 平台下，Iperf 将作为服务运行。

（4）-R：卸载 Iperf 服务(仅用于 Windows)。

（5）-o：重定向输出到指定文件(仅用于 Windows)。

（6）-P,－parallel：服务器关闭之前保持的连接数。默认是 0，这意味着永远接受连接。

Client 端专用参数：

（1）-c,－client host：iperf 客户端模式，host 是 server 端地址，eg：iperf -c 222.35.11.23

（2）-u，－udp：表示采用 UDP 协议发送报文，不带该参数表示采用 TCP 协议。

（3）-b，－bandwidth [K|M|G]：指定 UDP 模式使用的带宽，单位 bits/sec，默认值是 1 Mbit/sec。

（4）-t，－time：指定数据传输的总时间，即在指定的时间内，重复发送指定长度的数据包。默认 10 秒。

（5）-l，－len：设置读写缓冲区的长度，(default 128 KB for TCP, dynamic or 1 for UDP)。通常测试 PPS 的时候该值为 16，测试 BPS 时该值为 1400。

（6）-n，－num [K|M|G]：指定传输数据包的字节数，例如：iperf3 -c 192.168.12.168 －n 100M

（7）-P,－parallel：指定客户端与服务端之间使用的线程数。默认是 1 个线程。

需要客户端与服务器端同时使用此参数。

（8）-w，－window：指定套接字缓冲区大小，在 TCP 方式下，此设置为 TCP 窗口的大小。在 UDP 方式下，此设置为接受 UDP 数据包的缓冲区大小，用来限制可以接收数据包的最大值

（9）-B，－bind：用来绑定一个主机地址或接口，这个参数仅用于具有多个网络接口的主机。在 UDP 模式下，此参数用于绑定和加入一个多播组。

（10）-M，－mss：设置 TCP 最大信息段的值

（11）-N，－nodelay：设置 TCP 无延时

（12）-V：绑定一个 IPv6 地址。

（13）-d，－dualtest：运行双测试模式。将使服务器端反向连接到客户端，使用-L 参数中指定的端口（或默认使用客户端连接到服务器端的端口）。使用参数-r 以运行交互模式。

（14）-L,－listenport：指定服务端反向连接到客户端时使用的端口。默认使用客户端连接至服务端的端口。

（15）-r，－tradeoff：往复测试模式。当客户端到服务器端的测试结束时，服务器端反向连接至客户端。当客户端连接终止时，反向连接随即开始。如果需要同时进行双向测试，请尝试-d 参数。

一对一场景

测试准备

类型	数量	镜像	规格
被测机（被测机：被压力测试网络性能的弹性云服务器，可作为 iperf3 测试中的 client 端（发送端））	1台	CTyunOS 23.01（推荐）	4C8G(推荐)
辅助云服务器（作为 iperf3 测试中的 server 端（接收端））	1台	CTyunOS 23.01（推荐）	4C8G(推荐)

准备两台处于不同计算节点的弹性云主机，网络可达，分别安装 iperf，添加对

应的安全组规则、防护墙规则放行。

测试 TCP 吞吐量

1. TCP 发送带宽。

```
# Server 端开启 iperf 的服务器模式，指定 TCP 端口：10000，安全组规则入方向放行 TCP 10000 端口
[root@ss-daliu2 ~]# iperf3 -s -i 1 -p 10000
-----------------------------------------------------------Server listening on 10000 (test #1)
# Client 端启动 iperf 的客户端模式，连接服务端。ip 为 server 端地址
[root@ss-daliu1 ~]# iperf3 -c 192.168.0.15 -i 1 -t 60 -p 10000Connecting to host 192.168.0.15, port 10000
[  5] local 192.168.0.8 port 37720 connected to 192.168.0.15 port 10000
[ ID] Interval           Transfer     Bitrate         Retr  Cwnd
[  5]   0.00-1.00   sec   332 MBytes  2.78 Gbits/sec  4394    167 KBytes...
...
...
[  5]  52.00-53.00  sec   180 MBytes  1.51 Gbits/sec  6546    130 KBytes
[  5]  53.00-54.00  sec   166 MBytes  1.39 Gbits/sec  5237    180 KBytes
[  5]  54.00-55.00  sec   166 MBytes  1.40 Gbits/sec  5505   19.8 KBytes
[  5]  55.00-56.00  sec   207 MBytes  1.74 Gbits/sec  7651   12.7 KBytes
```

```
[  5]  56.00-57.00  sec   167 MBytes  1.40 Gbits/sec  5822    314
KBytes
[  5]  57.00-58.00  sec   168 MBytes  1.41 Gbits/sec  5952   58.0
KBytes
[  5]  58.00-59.00  sec   178 MBytes  1.49 Gbits/sec  6940    372
KBytes
[  5]  59.00-60.00  sec   205 MBytes  1.72 Gbits/sec  6756   2.83
KBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bitrate         Retr
[  5]   0.00-60.00  sec  10.7 GBytes  1.53 Gbits/sec  343419
sender
[  5]   0.00-60.00  sec  10.7 GBytes  1.53 Gbits/sec
receiver
```

结果说明：

Interval 表示时间间隔。

Transfer 表示时间间隔里面转输的数据量。

Bitrate 是时间间隔里的传输速率。

2. TCP 接收带宽。

```
# Server 端开启 iperf 的服务器模式，指定 TCP 端口：10000，安全组规则入方向放行 TCP 10000 端口[root@ss-daliu2 ~]# iperf3 -s -i 1 -p 10000
-----------------------------------------------------------
Server listening on 10000 (test #1)


# Client 端启动 iperf 的客户端模式，连接服务端。ip 为 server 端地址
[root@ss-daliu1 ~]# iperf3 -c 192.168.0.15 -i 1 -t 60 -p 10000 -R
```

```
Connecting to host 192.168.0.15, port 10000

Reverse mode, remote host 192.168.0.15 is sending[  5] local 192.168.0.8
port 38316 connected to 192.168.0.15 port 10000[ ID] Interval
Transfer     Bitrate[  5]   0.00-1.00   sec   336 MBytes  2.81
Gbits/sec[  5]   1.00-2.00   sec   172 MBytes  1.45 Gbits/sec

...

...

...[  5]  59.00-60.00  sec   179 MBytes  1.50 Gbits/sec

- - - - - - - - - - - - - - - - - - - - - - - - -[ ID] Interval
Transfer     Bitrate         Retr[  5]   0.00-60.00  sec  10.7 GBytes
1.53 Gbits/sec  360910             sender[  5]   0.00-60.00  sec
10.7 GBytes  1.53 Gbits/sec                  receiver
```

结果说明：

Interval 表示时间间隔。

Transfer 表示时间间隔里面转输的数据量。

Bitrate 是时间间隔里的传输速率。

3. TCP 上下行带宽测试（双向传输）。

```
# Server 端开启 iperf 的服务器模式，指定 TCP 端口：10000，安全组规则入方向放行 TCP 10000 端口

[root@ss-daliu2 ~]# iperf3 -s -i 1 -p 10000

-----------------------------------------------------------

Server listening on 10000 (test #1)# Client 端启动 iperf 的客户端模式，连接服务端。ip 为 server 端地址

[root@ss-daliu1 ~]# iperf3 -c 192.168.0.15 -i 1 -d -t 60 -p 10000

...

Thread FD 5 stopped
```

```
Sender threads stopped
tcpi_snd_cwnd 335 tcpi_snd_mss 1448 tcpi_rtt 1054
send_results
{
        "cpu_util_total":       4.7314293702543742,
        "cpu_util_user":        0.0694572749595791,
        "cpu_util_system":      4.6619720952947947,
        "sender_has_retransmits":       1,
        "congestion_used":      "cubic",
        "streams":      [{
                        "id":   1,
                        "bytes":        11446517760,
                        "retransmits":  354575,
                        "jitter":       0,
                        "errors":       0,
                        "omitted_errors":       0,
                        "packets":      0,
                        "omitted_packets":      0,
                        "start_time":   0,
                        "end_time":     60.00129
                }]
}
get_results
{
```

```
        "cpu_util_total":       3.3835424728514676,
        "cpu_util_user":        0.14802798323701713,
        "cpu_util_system":      3.2355161562396457,
        "sender_has_retransmits":       -1,
        "congestion_used":      "cubic",
        "streams":      [{
                        "id":   1,
                        "bytes":        11442323456,
                        "retransmits":  -1,
                        "jitter":       0,
                        "errors":       0,
                        "omitted_errors":       0,
                        "packets":      0,
                        "omitted_packets":      0,
                        "start_time":   0,
                        "end_time":     60.00149
                }]
}
interval_len 1.000265 bytes_transferred 176291840
interval forces keep
[  5]  59.00-60.00  sec   168 MBytes  1.41 Gbits/sec  5895    474
KBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bitrate         Retr
```

```
[  5]   0.00-60.00  sec  10.7 GBytes  1.53 Gbits/sec  354575
sender
[  5]   0.00-60.00  sec  10.7 GBytes  1.53 Gbits/sec
receiver
Receiver threads stopped


iperf Done.
# 测试 4 线程 TCP 吞吐量,如果没有指定发送方式，iperf 客户端只会使用单线程。
$ iperf3 -c 172.16.0.9 -i 1 -P 4 -t 60 -p 10000
```

结果说明：

Interval 表示时间间隔。

Transfer 表示时间间隔里面转输的数据量。

Bitrate 是时间间隔里的传输速率。

测试 UDP 吞吐量

带宽测试通常采用 UDP 模式，因为能测出极限带宽、时延抖动、丢包率。在进行测试时，首先以链路理论带宽作为数据发送速率进行测试，例如，从客户端到服务器之间的链路的理论带宽为 100Mbps，先用-b 100M 进行测试，然后根据测试结果（包括实际带宽，时延抖动和丢包率），再以实际带宽作为数据发送速率进行测试，会发现时延抖动和丢包率比第一次好很多，重复测试几次，就能得出稳定的实际带宽。

```
# Server 端开启 iperf 的服务器模式，指定 UDP 端口：
[root@ss-daliu2 ~]#  iperf3 -s -i 1 -p 10001
-----------------------------------------------------------
Server listening on 10001
-----------------------------------------------------------
```

```
# Client 端启动 iperf 的客户端模式，连接服务端
[root@ss-daliu1 ~]# iperf3 -u -c 192.168.0.15 -b 100m -t 60 -p 10001
Connecting to host 172.16.0.9, port 10001
[  4] local 172.16.0.2 port 46011 connected to 172.16.0.9 port 10001
[ ID] Interval           Transfer     Bandwidth       Total Datagrams
[  4]   0.00-1.00   sec  10.8 MBytes  90.3 Mbits/sec  7799
...
[  4]  58.00-59.00  sec  11.9 MBytes   100 Mbits/sec  8640
[  4]  59.00-60.00  sec  11.9 MBytes  99.9 Mbits/sec  8625
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bitrate         Jitter
Lost/Total Datagrams
[  4]   0.00-60.00  sec   714 MBytes  99.8 Mbits/sec  0.001 ms
133914/517118 (26%)
[  4] Sent 517118 datagrams


iperf Done.
# 测试多线程 UDP 吞吐量
$ iperf3 -u -c 192.168.0.15 -b 5m -P 4 -t 60 -p 10001
# 进行上下行带宽测试（双向传输）
$ iperf3 -u -c 192.168.0.15 -b 100M -d -t 60 -p 10001
```

结果说明：

jitter 为抖动，在连续传输中的平滑平均值差。

Lost 为丢包数量。

Total Datagrams 为包数量。

iperf 测试 udp 丢包严重：可以尝试增大收发包缓冲区。

```
echo 16777216 > /proc/sys/net/core/rmem_max

echo 16777216 > /proc/sys/net/core/rmem_default

echo 16777216 > /proc/sys/net/core/wmem_max

echo 16777216 > /proc/sys/net/core/wmem_default
```

多流打流测试方案（针对大规格云主机）

针对大规格云主机，单进程打流可能无法发挥出云主机的性能，可以参考下述多流打流测试方案。

虚机网络性能说明

虚机从同一个队列发出的报文，在底层只能被一个 cpu 转发。即虚机一个队列的发包性能 <= 虚拟交换机单核转发性能。

性能测试步骤

1. 确定两个用于性能测试的计算节点：A，B.
2. 在每个计算节点上分别创建 2 个虚机，虚机规格尽量大，虚机内核版本 4.19 及以上(建议使用 CTyunOS 23.01)。
3. 4 个测试虚机内安装 iperf3 包

```
yum install iperf3
```

4. B 计算节点上的两个虚机(B1，B2)分别起 iperf 服务端：

```
for i in {0..63}; do iperf3 -s -D -p $((5200 + $i)) ; done
```

5. A 计算节点上两个虚机(A1，A2)起 iperf 客户端(A1 给 B1 发包，A2 给 B2 发包)：

   创建发包脚本：

```
#!/usr/bin/env bash

flow_per_thread=2

max_thread=64

run_time=60

udp_opts="-u"# 默认用全力(100G)发送 udp 报文，但可能存在丢包.
```

```
udp_band_Mb=100000

tcp_opts=""

use_tcp=1usage() {

    echo '

usage:

    $0 -s <server-ip> [--pps] [-n <nprocess>] [-t <time>] [--tcp]


-s          Iperf server ip address.

--pps       Send small pkts for pps benchmark. By default send large pkts
for bandwidth benckmark.

-n          Set max process number. If not set, tx queue number is used.

-t          time in seconds to transmit for, default 60 seconds.

--udp       Send udp instead of tcp(default). Udp is good for some
special scenarios.

-b          Bandwidth to send for udp mode. i.e. 1M, 2G,

'

}while [ $# -gt 0 ]; do

    case $1 in

        -s | --server)

            server="$2"

            shift

            ;;

        -n | --nprocess)

            max_thread="$2"
```

```
        shift
        ;;
    -t | --time)
        run_time="$2"
        shift
        ;;
    --pps)
        # 144 bytes
        tcp_opts="$tcp_opts -M 90"
        # 18(eth) + 28(ip) + 18(udp) = 64
        udp_opts="$udp_opts -l 18"
        ;;
    --udp)
        use_tcp=0
        ;;
    -b)
        udp_band="$2"
        _unit=$(echo $udp_band | grep -oiP '\d+\K[mg]')
        _num=$(echo $udp_band | grep -oP '\d+')
        if [ -z "$_unit" ] || [ -z "$_num" ];then
            echo "bad bandwidth"
            exit 1
        fi
        if [ ${_unit,,} == 'g' ];then
```

```
                _num=$((_num*1024))
            fi
            udp_band_Mb=$_num

            ;;
        -h | --help)
            usage
            exit 0
            ;;
        *)
            usage
            exit 1
        ;;
    esac
    shiftdone

interface=$(ip route get $server |grep -oP "(?<=dev )\S+")
n_txq=$(ls /sys/class/net/$interface/queues/tx-*/xps_cpus | wc -l)
band_per_flow=$((udp_band_Mb/flow_per_thread/max_thread))
i=0for fxps in `ls /sys/class/net/$interface/queues/tx-*/xps_cpus |
sort`do
    aff=$(cat $fxps)
    if ((i >= max_thread));then
        break
```

```
fi
no_zero_aff=$(echo "$aff" | tr -d '0xX')
if [ -z "$no_zero_aff" ];then
    _aff=$((1 << $i))
    aff=$(printf "%x" $_aff)
    echo $aff > $fxps
fi

# Affinity on only one core
suffix=""
for (( j=${#aff}-1; j>=0; j-- )); do
    if [ "${aff:$j:1}" == "," ];then
        continue
    fi
    if [ "${aff:$j:1}" != "0" ];then
        aff_one="${aff:$j:1}${suffix}"
        break
    fi
    suffix+="0"
done
if [ -z "$aff_one" ];then
    echo "Failed to find txq affinity core"
    exit 1
fi
```

```
    aff_one=`echo $aff_one | tr '13579bdf' '1' | tr '26ae' '2' | tr '4c'
'4'`


    if ((use_tcp > 0)) ;then
        taskset $aff_one iperf3 -c $server $tcp_opts -P 2 -p $((5200+$i))
-t $run_time &
    else
        taskset $aff_one iperf3 -c $server $udp_opts -b ${udp_band_Mb}M
-P 2 -p $((5200+$i))  -t $run_time &
    fi
    ((i++))donewait
```

测试 pps(144B/64B 小包)

```
bash iperf-client.sh -s <B1-ip> --pps -t 300
```

测试带宽(1500B 大包)

```
bash iperf-client.sh -s <B1-ip> -t 300
```

6. 打流的过程中虚拟交换机会根据流量做 balance，所以我们流量测试应持续 5 分钟。等待底层 balance 到最佳状态，以得到较好的性能结果。

10.22 网站性能测试的最佳实践

天翼云通用型 S6、计算增强型 C6 等新一代云主机上提供超高网络性能，您可通过本实践提供的 netperf 和 DPDK 两种网络性能测试方法，进行云主机高吞吐网络性能测试。

推荐选择 netperf 方法进行测试，netperf 为通常使用的测试方法，可满足大多数测试场景。但当云主机配置较高（pps 超过 1000 万且带宽大于 50Gbps）时，netperf 包含云主机机内核协议栈的完整处理路径对网络性能损耗较大，而

DPDK 可屏蔽虚拟机内核协议栈的差异，获取虚拟机网卡的网络性能，此时可选择 DPDK 方法进行测试。

netperf 测试

工具介绍

Netperf

HP 开发的网络性能测量工具，主要测试 TCP 及 UDP 吞吐量性能。测试结果主要反应系统向其他系统发送数据的速度，以及其他系统接收数据的速度。

- SAR

用于监控网络流量，运行示例如下：

```
sar -n DEV 102:41:03 PM     IFACE   rxpck/s   txpck/s    rxkB/s
txkB/s   rxcmp/s   txcmp/s  rxmcst/s02:41:04 PM      eth0 1626689.00
8.00  68308.62      1.65      0.00      0.00  0.0002:41:04 PM
lo      0.00      0.00      0.00      0.00      0.00      0.00
0.0002:41:04 PM     IFACE   rxpck/s   txpck/s    rxkB/s    txkB/s
rxcmp/s   txcmp/s  rxmcst/s02:41:05 PM      eth0 1599900.00      1.00
67183.30      0.10      0.00      0.00  0.0002:41:05 PM        lo
0.00      0.00      0.00      0.00      0.00      0.00      0.00
```

sar -n DEV 1：该命令每秒钟报告一次网络设备的性能统计信息

字段解释

字段	单位	说明

字段	单位	说明
rxpck/s	pps	每秒收包量，即接收 pps
txpck/s	pps	每秒发包量，即发送 pps
rxkB/s	kB/s	接收带宽
txkB/s	kB/s	发送带宽

测试场景

测试场景	客户端运行命令	SAR 监控指标
UDP 64	netperf -t UDP_STREAM -H-l 10000 -- -m 64 -R 1 &	PPS
TCP 1500	netperf -t TCP_STREAM -H-l 10000 -- -m 1500 -R 1 &	带宽

测试场景	客户端运行命令	SAR 监控指标
TCP RR	netperf -t TCP_RR -H-l 10000 -- -r 32,128 -R 1 &	PPS

命令说明：

- netperf: 这是 Netperf 工具的命令，用于执行网络性能测试。
- -t UDP_STREAM/TCP_STREAM: 指定测试的类型为 UDP/TCP 流量测试。
- -H: 指定远程主机的地址。在这里，-H 后面应该跟上远程主机的地址或主机名。
- -l ：指定测试的时间限制，后接时间，单位为秒。
- -m ：设置消息大小，单位为字节。
- -R ：指定测试运行的数量 。
- &: 将测试命令放入后台运行。

性能指标

指标	说明
64 字节 UDP 收发 PPS(包/秒)	表示通过 UDP 进行批量数据传输时的数据传输吞吐量， 能反映网络极限转发能力（可能会存在丢包）。

指标	说明
1500 字节 TCP 收发带宽（Mbits/秒）	表示通过 TCP 进行批量数据传输时的数据传输吞吐量， 能反映网络极限带宽能力（可能会存在丢包）。
TCP-RR（次/秒）	表示在 TCP 长链接中反复进行 Request/Response 操作的交易吞吐量， 能反映 TCP 不丢包网络转发能力。

操作步骤

准备测试环境

1.准备 3 台测试机器，请参见创建弹性云主机购买测试机器。本文测试机器使用 CentOS 8.2 操作系统。

2.依次登录测试机器，并执行以下命令安装 netperf 工具。

```
yum install -y sysstat wget tar automake make gccwget -O netperf-2.7.0.tar.gz -c https://codeload.github.com/HewlettPackard/netperf/tar.gz/netperf-2.7.0tar zxf netperf-2.7.0.tar.gzcd netperf-netperf-2.7.0./autogen.sh && ./configure && make && make install
```

测试发包性能

1.分别在机器中执行命令“pkill netserver && pkill netperf”，停止残余的 netperf 和 netserver 进程。

2.将其中的机器 a 作为客户端，机器 b 和机器 c 作为服务端。在服务端中执行命令“netserver”，运行 netserver。

- 若返回结果如下图所示，则说明仍存在其他 netserver 进程。请执行步骤 1 中的命令，停止该进程。

```
[root@VM-2-8-centos ~]# netserver
Unable to start netserver with  'IN(6)ADDR_ANY' port '12865' and family AF_UNSPEC
[root@VM-2-8-centos ~]#
```

- 若返回结果如下图所示，则说明已成功运行 netserver，请继续下一步操作。

```
[root@VM-2-8-centos ~]# netserver
Starting netserver with host 'IN(6)ADDR_ANY' port '12865' and family AF_UNSPEC
[root@VM-2-8-centos ~]#
```

3.在客户端中执行 测试场景 中提供的命令，不断增减 netperf 进程，直到客户端发包性能不再增加。

（1）执行以下命令，启动 10 个 netperf 客户端进程

```
bash netperf 10 <serverip1> <serverip2> <serverip3> <serverip4> <serverip5> <serverip6> <serverip7> <serverip8> <serverip9> <serverip10>
```

（2）同时向 b、c 两台或者多台客户端发送数据，代码如下

```
#!/bin/bash
count=$1shiftwhile [ $# != 0 ];dofor ((i=1;i<=count;i++))do
    server_ip=$1
    [ -z $server_ip ] && exit 0# 下方命令可以替换为测试场景表格中的命令# -H 后填写服务器 IP 地址;# -l 后为测试时间，为了防止 netperf 提前结束，因此时间设为 10000;
    netperf -t UDP_STREAM -H $server_ip -l 10000 -- -m 64 -R 1 &
    echo "$server_ip run $i times"doneshiftdone
```

4.在客户端执行命令“sar -n DEV 1”，观察客户端发包性能变化，取最大值。

根据所得结果，参考性能指标进行分析，即可测出云主机高吞吐网络性能。

测试收包性能

1.分别在机器中执行命令“pkill netserver && pkill netperf”，停止残余的 netperf 和 netserver 进程。
2.将其中的机器 a 作为服务端，机器 b 和机器 c 作为客户端。在服务端中执行命令“netserver”，运行 netserver。

若返回结果如下图所示，则说明仍存在其他 netserver 进程。请执行 步骤 1 中的命令，停止该进程。

```
[root@VM-2-8-centos ~]# netserver
Unable to start netserver with  'IN(6)ADDR_ANY' port '12865' and family AF_UNSPEC
[root@VM-2-8-centos ~]#
```

若返回结果如下图所示，则说明已成功运行 netserver，请继续下一步操作。

```
[root@VM-2-8-centos ~]# netserver
Starting netserver with host 'IN(6)ADDR_ANY' port '12865' and family AF_UNSPEC
[root@VM-2-8-centos ~]#
```

3.在客户端中执行 测试场景 中提供的命令，不断增减 netperf 进程，直到客户端发包性能不再增加。

在服务端执行命令“sar -n DEV 1”，观察服务端收包性能变化，取最大值。

4.根据所得结果，参考 性能指标 进行分析，即可测出云主机高吞吐网络性能。

附录

测试辅助脚本：

执行该脚本，可快速发起多个 netperf 进程。

```
#!/bin/bash

count=$1for ((i=1;i<=count;i++))doecho "Instance:$i-------"# 下方命令可以替换为测试场景表格中的命令# -H 后填写服务器 IP 地址;# -l 后为测试时间，为了防止 netperf 提前结束，因此时间设为 10000;

netperf -t UDP_STREAM -H <server ip> -l 10000 -- -m 64 -R 1 &done
```

DPDK 测试

编译安装 DPDK

1.准备 2 台测试机器，请参见 自定义配置 Linux 云主机 购买测试机器。本文测试机器使用 CentOS 8.2 操作系统。

2.依次登录测试机器，并执行以下命令下载 DPDK 工具。

```
yum install -y sysstat wget tar automake make gccwget
http://git.dpdk.org/dpdk/snapshot/dpdk-17.11.tar.gztar -xf
dpdk-17.11.tar.gzmv dpdk-17.11 dpdk
```

3.修改 txonly 引擎，使每个 DPDK 发包 CPU 上的 UDP 流量的端口变动产生多条流。

1）执行以下命令，修改 dpdk/app/test-pmd/txonly.c 文件。

```
vim dpdk/app/test-pmd/txonly.c
```

按 i 进入编辑模式，修改以下内容：

- 找到 #include "testpmd.h"，另起一行输入如下内容：

```
RTE_DEFINE_PER_LCORE(struct udp_hdr, lcore_udp_hdr);
```

修改完成后，如下图所示：

```
#include "testpmd.h"
RTE_DEFINE_PER_LCORE(struct udp_hdr, lcore_udp_hdr);

#define UDP_SRC_PORT 1024
#define UDP_DST_PORT 1024
```

- 找到 ol_flags |= PKT_TX_MACSEC;，另起一行输入如下内容：

```
 /* dummy test udp port */

static uint16_t test_port = 0;

 test_port++;
```

```
memcpy(&RTE_PER_LCORE(lcore_udp_hdr), &pkt_udp_hdr,
sizeof(pkt_udp_hdr));

 RTE_PER_LCORE(lcore_udp_hdr).src_port =
rte_cpu_to_be_16(rte_lcore_id() * 199 + test_port %
16);RTE_PER_LCORE(lcore_udp_hdr).dst_port =
rte_cpu_to_be_16(rte_lcore_id() * 1999 + test_port % 16);
```

修改完成后，如下图所示：

```
if (txp->tx_ol_flags & TESTPMD_TX_OFFLOAD_INSERT_QINQ)
        ol_flags |= PKT_TX_QINQ_PKT;
if (txp->tx_ol_flags & TESTPMD_TX_OFFLOAD_MACSEC)
        ol_flags |= PKT_TX_MACSEC;

/* dummy test udp port */
static uint16_t test_port = 0;
test_port++;
memcpy(&RTE_PER_LCORE(lcore_udp_hdr), &pkt_udp_hdr, sizeof(pkt_udp_hdr));
RTE_PER_LCORE(lcore_udp_hdr).src_port = rte_cpu_to_be_16(rte_lcore_id() * 199 + test_port % 16);
RTE_PER_LCORE(lcore_udp_hdr).dst_port = rte_cpu_to_be_16(rte_lcore_id() * 1999 + test_port % 16);

for (nb_pkt = 0; nb_pkt < nb_pkt_per_burst; nb_pkt++) {
        pkt = rte_mbuf_raw_alloc(mbp);
        if (pkt == NULL) {
        nomore_mbuf:
                if (nb_pkt == 0)
                        return;
                break;
        }
```

- 找到 copy_buf_to_pkt(&pkt_udp_hdr, sizeof(pkt_udp_hdr), pkt,，将其替换为如下内容：

```
copy_buf_to_pkt(&RTE_PER_LCORE(lcore_udp_hdr),
sizeof(RTE_PER_LCORE(lcore_udp_hdr)), pkt,
```

修改完成后，如下图所示：

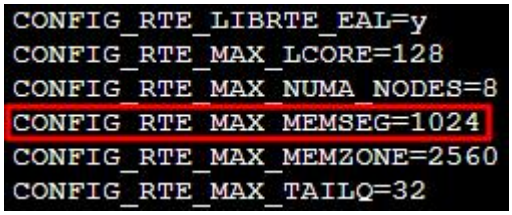
```
copy_buf_to_pkt(&eth_hdr, sizeof(eth_hdr), pkt, 0);
copy_buf_to_pkt(&pkt_ip_hdr, sizeof(pkt_ip_hdr), pkt,
                sizeof(struct ether_hdr));
copy_buf_to_pkt(&RTE_PER_LCORE(lcore_udp_hdr), sizeof(RTE_PER_LCORE(lcore_udp_hdr)), pkt,
                sizeof(struct ether_hdr) +
                sizeof(struct ipv4_hdr));
```

按 Esc 输入 :wq 保存修改并退出。

2）执行以下命令，修改 dpdk/config/common_base 文件。

```
vim dpdk/config/common_base
```

按 i 进入编辑模式，找到 CONFIG_RTE_MAX_MEMSEG=256，将其修改为 1024。修改完成后如下图所示：



按 Esc 输入 :wq 保存修改并退出。

4.执行以下命令，将 dpdk/app/test-pmd/txonly.c 的 IP 地址修改为测试机器所用 IP。

```
vim dpdk/app/test-pmd/txonly.c
```

按 i 进入编辑模式，找到如下内容：

```
#define IP_SRC_ADDR (198U << 24) | (18 << 16) | (0 << 8) | 1;

#define IP_DST_ADDR (198U << 24) | (18 << 16) | (0 << 8) | 2;
```

将数字 198、18、0、1 替换为机器 IP，SRC_ADDR 为发送端 IP，DST_ADDR 为接收端 IP。

5.对应机器操作系统，执行以下命令，安装 numa 库。

```
yum install numactl-devel
```

6.在 dpdk/ 目录下执行以下命令，关闭 KNI。

```
sed -i  "s/\(^CONFIG_.KNI.\)=y/\1=n/g" ./config/*
```

7.若您的操作系统内核版本较高（例如 5.3），则请执行以下命令，屏蔽差异。

```
sed -i "s/\(^WERROR_FLAGS += -Wundef -Wwrite-strings$\)/\1
-Wno-address-of-packed-member/g" ./mk/toolchain/gcc/rte.vars.mksed -i
"s/fall back/falls through
-/g" ./lib/librte_eal/linuxapp/igb_uio/igb_uio.c
```

8.执行以下命令，编译 DPDK。

```
make defconfigmake -j
```

配置大页内存

执行以下命令，配置大页内存。

```
let hugepagesall=$(cat /proc/meminfo | grep MemTotal | awk '{print $2}')/1024/2048#配置时不超过理论可配大页个数

echo $hugepagesall  > /sys/kernel/mm/hugepages/hugepages-2048kB/nr_hugepages
```

装载内核模块及绑定接口

1.切换登录方式为使用 VNC。由于将网卡驱动绑定至 igb_uio 用户态驱动后，该网卡将无法通过 ssh 或 IP 访问，仅支持通过 VNC 或 console 方式访问。
2.依次执行以下命令，装载 UIO 模块及绑定 virito 接口。

```
ifconfig eth0 0

ifconfig eth0 down

modprobe uio

insmod /root/dpdk/build/kmod/igb_uio.ko

cd /root/dpdk/usertools/

python3 dpdk-devbind.py --bind=igb_uio 00:05.0
```

完成测试后，可通过请执行以下命令，恢复网卡变更。

```
cd /root/dpdk/usertools/

python3 dpdk-devbind.py --bind=virtio-pci 00:05.0

ifconfig eth0 up
```

测试带宽及吞吐量

1.执行以下命令，发送端采用 TX only 模式启动 testpmd，接收端启用 rxonly 模式。

- 发送端：

```
/root/dpdk/build/app/testpmd -- --txd=128 --rxd=128 --txq=16 --rxq=16 --nb-cores=1 --forward-mode=txonly --txpkts=1430 --stats-period=1
```

- 接收端：

```
/root/dpdk/build/app/testpmd -- --txd=128 --rxd=128 --txq=48 --rxq=48 --nb-cores=16 --forward-mode=rxonly --stats-period=1
```

2.执行以下命令，测试 pps（UDP 64B 小包）。

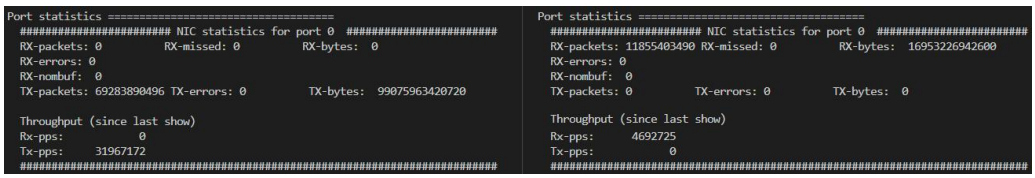
- 发送端：

```
/root/dpdk/build/app/testpmd -- --txd=128 --rxd=128 --txq=16 --rxq=16 --nb-cores=3 --forward-mode=txonly --txpkts=64 --stats-period=1
```

- 接收端：

```
/root/dpdk/build/app/testpmd -- --txd=128 --rxd=128 --txq=48 --rxq=48 --nb-cores=16 --forward-mode=rxonly --stats-period=1
```

得出如下图所示测试结果：



网络带宽计算

可根据接收端 PPS 和测试包长来计算当前网络的接收带宽，公式如下：

PPS × packet length × 8bit/B × 10^-9 = 带宽

结合测试得出数据，可得当前带宽为：

4692725pps × 1430B × 8bit/B × 10^-9 ≈ 53Gbps

10.23 网站搭建指导

网站搭建一般有以下两种方式：

对比项	镜像部署	手动搭建
搭建方式	选择天翼云市场镜像直接安装部署	以手动的方式安装所需软件，可定制化
特点	配套的软件版本相对比较固定	配套版本可以灵活选择
所需时间	比较短，一键部署	比较长，需自行安装相关软件
难易程度	相对比较简单	需要对软件配套版本及安装方法有一定的了解

如果您或您的团队，在建站时，有个性化的部署需求，并且有专业的维护人员，可以采用手动搭建方式。

如果您或您的团队，在建站时，对于操作系统、数据库、中间件等各类软件的选择和配置不是很了解，并且缺少专业维护人员，建议使用镜像部署，后期有镜像服务商的专业售后工程师团队支持。

10.24 手动部署 Java Web

Tomcat 是一个被广泛使用的 Java Web 应用云主机。本文介绍了在天翼云弹性云主机上部署 Java Web 环境的操作步骤。首先需要下载部署 Java Web 环境所需的安装包，并将安装包上传至云主机，然后设置弹性云主机安全组规则，再安装并配置相关软件，完成开发环境的配置。

适用对象：本文档适用于使用天翼云弹性云主机部署 Java Web 环境的用户。

相关软件及工具

软件包名称	获取方式
jdk	http://www.oracle.com/technetwork/java/javase/downloads
tomcat	http://tomcat.apache.org/download-80.cgi

说明 ： 上表中为 jdk 和 tomcat 软件包官方获取地址，您还可以参考其他开源镜像地址获取安装包。

工具名称	说明	获取方式
PuTTY	跨平台远程访问工具。用于在软件安装过程中在	https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html

工具名称	说明	获取方式
	Windows 系统上访问云主机。	
WinSCP	跨平台文件传输工具。用于在 Windows 系统和 Linux 系统间传输文件。	http://winscp.net/

说明 ： 上表中为 PuTTY 和 WinSCP 工具包官方获取地址，您还可以参考其他开源镜像地址获取安装包。

必备事项

1.创建弹性云主机，且弹性云主机已绑定弹性 IP。

2.登录弹性云主机，执行如下命令，新建 jdk 目录。

```
cd /home/mkdir webDemocd webDemo/mkdir jdk
```

3.登录弹性云主机，执行如下命令，新建 tomcat 目录。

```
cd webDemo/mkdir tomcat
```

4.您可以选择将安装包下载至本地后使用文件传输工具将安装包上传至云主机。或者选择使用 wget 命令直接下载安装包至云主机。

- 方法一

1）使用文件传输工具上传安装包至云主机。

2）使用 WinSCP 工具上传 jdk 软件包至云主机 jdk 文件夹。

3）使用 WinSCP 工具上传 tomcat 软件包至云主机 tomcat 文件夹。

- 方法二

1）使用 wget 命令直接下载安装包至云主机。

I.执行如下命令，进入 jdk 目录。

```
cd /home/webDemo/jdk
```

II.执行如下命令，下载 jdk 软件包。

```
wget
https://download.oracle.com/java/17/latest/jdk-17_linux-x64_bin.tar.g
z
```

III.执行如下命令，进入 tomcat 目录。

```
cd /home/webDemo/tomcat
```

IV.执行如下命令，下载 tomcat 软件包。

```
wget
https://mirrors.tuna.tsinghua.edu.cn/apache/tomcat/tomcat-8/v8.5.88/b
in/apache-tomcat-8.5.88.tar.gz --no-check-certificate
```

说明

本文使用的云主机以天翼云 CentOS 7.3 64bit 操作系统云主机为例。

- *JDK 软件包以 jdk-17_linux-x64_bin.tar.gz 安装包为例。

- Tomcat 以 apache-tomcat-8.5.78.tar.gz 安装包为例。
- 如果当前操作步骤中下载链接过期，您可以参考表 1 查询其他版本 jdk 和 tomca 下载地址，或者使用其他开源镜像地址获取安装包。

设置弹性云主机安全组规则

安全组规则设置步骤

1.单击弹性云主机名称，查看弹性云主机详情，在弹性云主机详情页面，选择“安全组”。

2.在“安全组”界面，单击“更改安全组规则”，进入安全组详情界面。

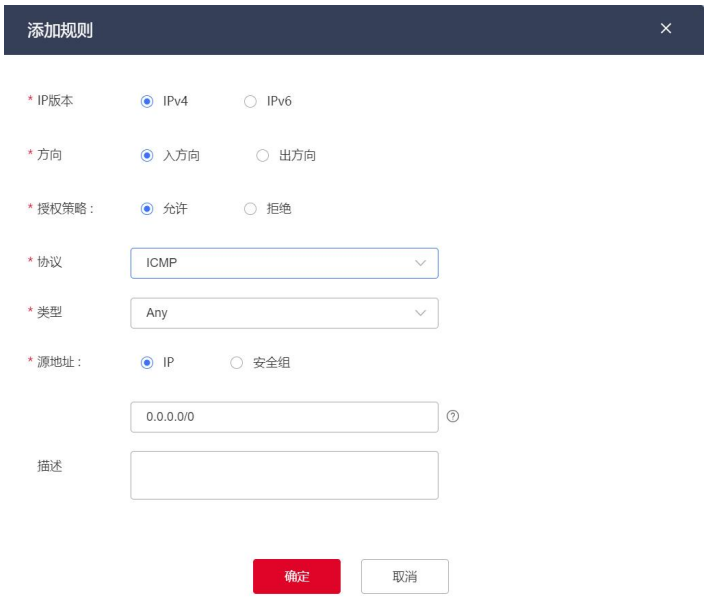
3.在安全组详情界面，单击“添加规则”，弹出添加规则窗口。

4.根据界面提示配置安全组规则。

添加 ICMP 安全组规则

如果云主机默认设置是禁止 ICMP 规则，当 ping 弹性云主机 IP 时会显示超时。因此首先为云主机添加 ICMP 规则。

为云主机添加 web 项目分配端口的访问规则，以 8080 端口为例:



说明

- 入方向：指从外部访问安全组规则下的弹性云主机。

- 出方向：指安全组规则下的弹性云主机访问安全组外的实例。
- 端口范围：安全组规则的端口范围，取值范围为：1～65535,Any 代表所有端口。其中 0.0.0.0/0 表示所有 IPv4 地址，::/0 表示所有 IPv6 地址。

WEB 服务端口默认使用得端口为：80（http）、443（https），用户也可选择将 Web 云主机配置为在非标准端口上运行，如 8080，此时需要配置对应的安全组端口。

安装 JDK

- 解压 jdk 安装包到 jdk 目录下。

```
tar -xvf jdk-8u231-linux-x64.tar.gz -C /home/webDemo/jdk/
```

- 配置环境变量。

```
vi /etc/profile
```

- 在底部添加以下内容。

```
#set java      environmentexport
JAVA_HOME=/home/webDemo/jdk/jdk1.8.0_231export
JRE_HOME=/home/webDemo/jdk/jdk1.8.0_231/jreexport
CLASSPATH=.:JAVA_HOME/lib/dt.jar:JAVA_HOME/lib/dt.jar:JRE_HOME/lib/tools.jar  export PATH=JAVA_HOME/bin:JAVA_HOME/bin: $PATH
```

- 执行以下命令保存并退出。

```
:wq
```

- 执行以下命令使/etc/profile 里的配置生效。

```
source /etc/profile
```

- 验证安装。

```
java -version
```

回显信息如下所示验证安装 jdk 成功。

```
[root@ecs-c525-web ~]# java -versionjava version "1.8.0_231"Java(TM) SE Runtime Environment (build 1.8.0_231-b11)Java HotSpot(TM) 64-Bit Server VM (build 25.231-b11, mixed mode
```

安装 tomcat

- 解压 tomcat 安装包到 tomcat 目录下。

```
tar -xvf apache-tomcat-8.5.47.tar.gz -C /home/webDemo/tomcat/
```

- 进入 tomcat 的 bin 目录，执行以下命令安装 tomcat。

```
cd /home/webDemo/tomcat/apache-tomcat-8.5.47/cd bin/
```

- 执行如下命令编辑 setclasspath.sh 脚本。

```
vi setclasspath.sh
```

并在 setclasspath.sh 脚本添加以下内容。

```
export JAVA_HOME=/home/webDemo/jdk/jdk-17.0.3export JRE_HOME=$JAVA_HOME
```

- 保存后退出，可输入以下命令启动 tomcat。

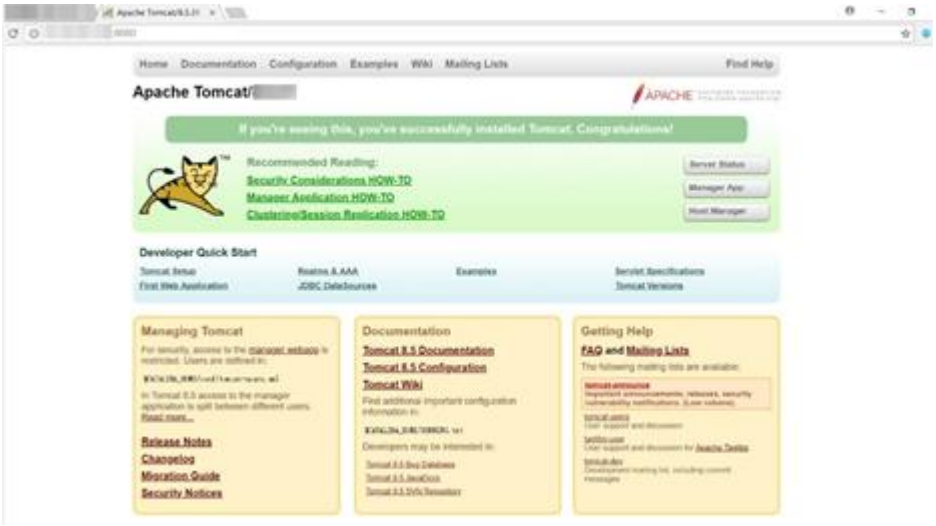
```
./startup.sh
```

验证 Java Web 环境搭建完成

- 在浏览器输入以下内容。

http://云主机弹性 IP:8080,如果界面跳转至默认的 Tomcat 界面,证明 Java Web 环境搭建完成。我们就可以在公网访问云主机的 8080 端口了。

访问云主机的 8080 端口如图所示：



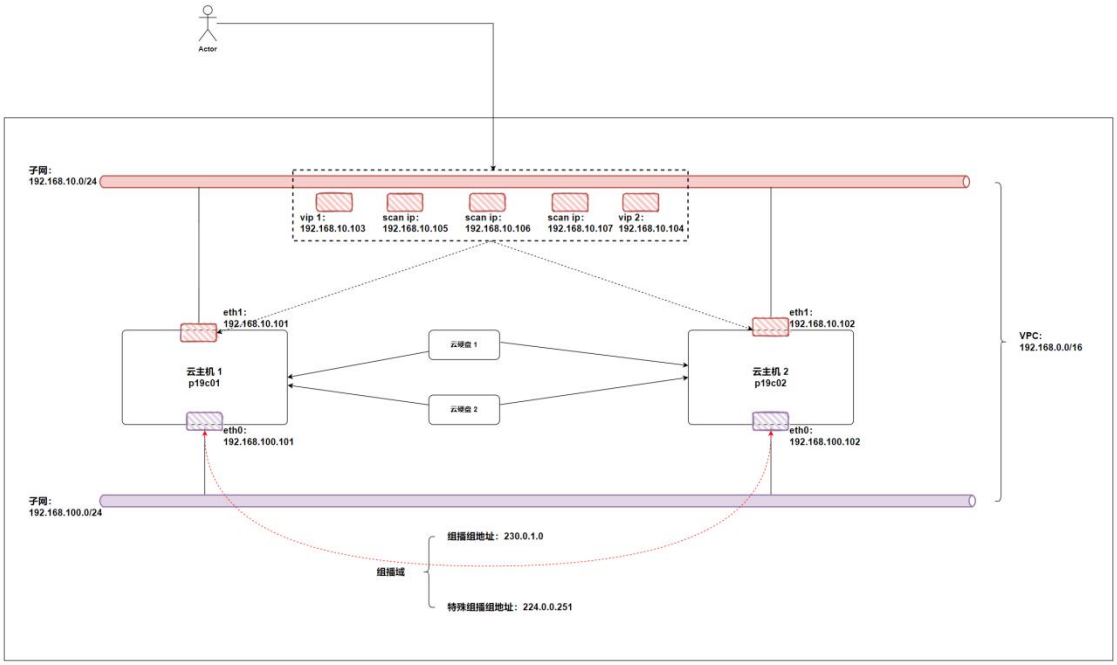
注意

天翼云云主机 80/8080/443/8443 端口需备案完成后才可开通使用，除上述四个端口，云主机其余端口均为开通状态，直接调用即可。但域名+IP+端口号如能通过互联网访问，则属于未备案，只做内部访问无妨。

80/8080/443/8443 端口会在备案成功后一个工作日内开通，开通后会电话联系网站负责人。

10.25 基于天翼云 ECS 搭建 Oracle RAC 集群

资源规划



系统规划

序号	节点主机名称	操作系统	虚机规格
1	p19c01	CentOS 7.9 64 位	s7.2xlarge.2
2	p19c02	CentOS 7.9 64 位	s7.2xlarge.2

网络规划

序号	节点主机名称	Private IP	Public IP	Virtual IP	Scan IP

序号	节点主机名称	Private IP	Public IP	Virtual IP	Scan IP
1	p19c01	192.168.100.101	192.168.10.101	192.168.10.103	192.168.10.105 192.168.10.106 192.168.10.107
2	p19c02	192.168.100.102	192.168.10.102	192.168.10.104	192.168.10.105 192.168.10.106 192.168.10.107

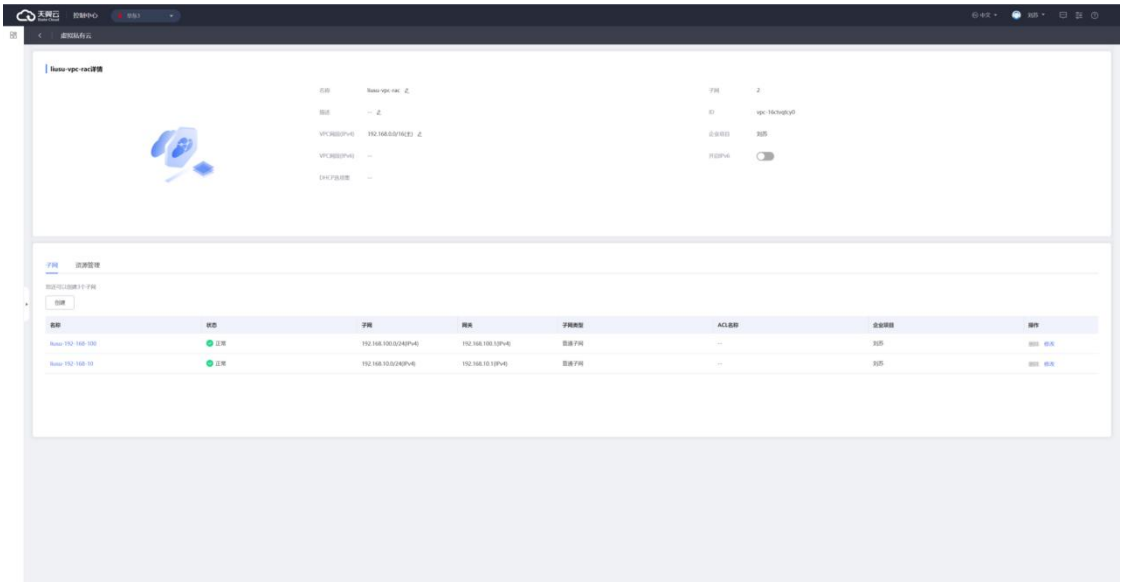
存储规划

序号	角色	冗余模式	大小
1	裁决盘 OCR	EXTERNAL	100G
2	数据盘 DATA	EXTERNAL	200G

资源创建

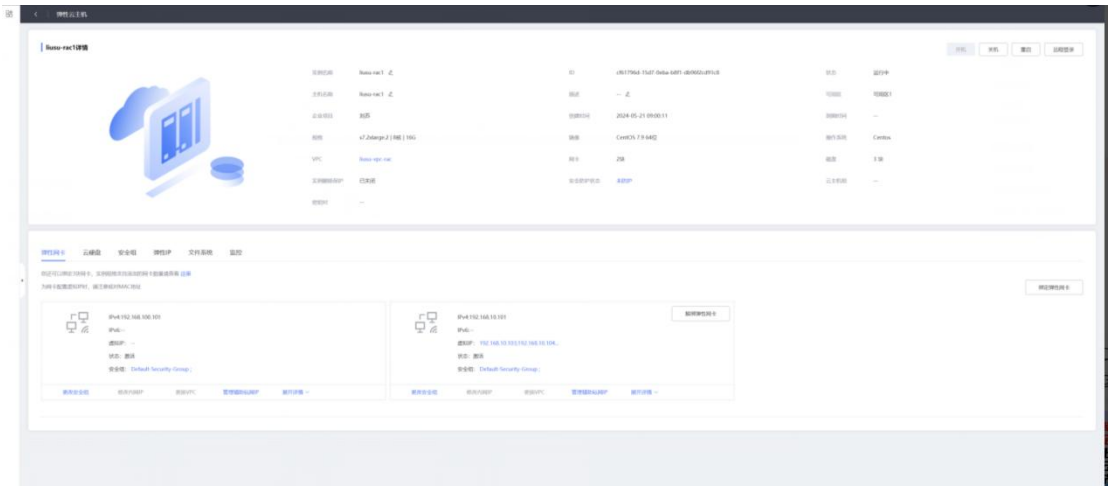
1. 创建 VPC 网络。

如下图所示为创建的 VPC 网络并创建了 2 个子网：



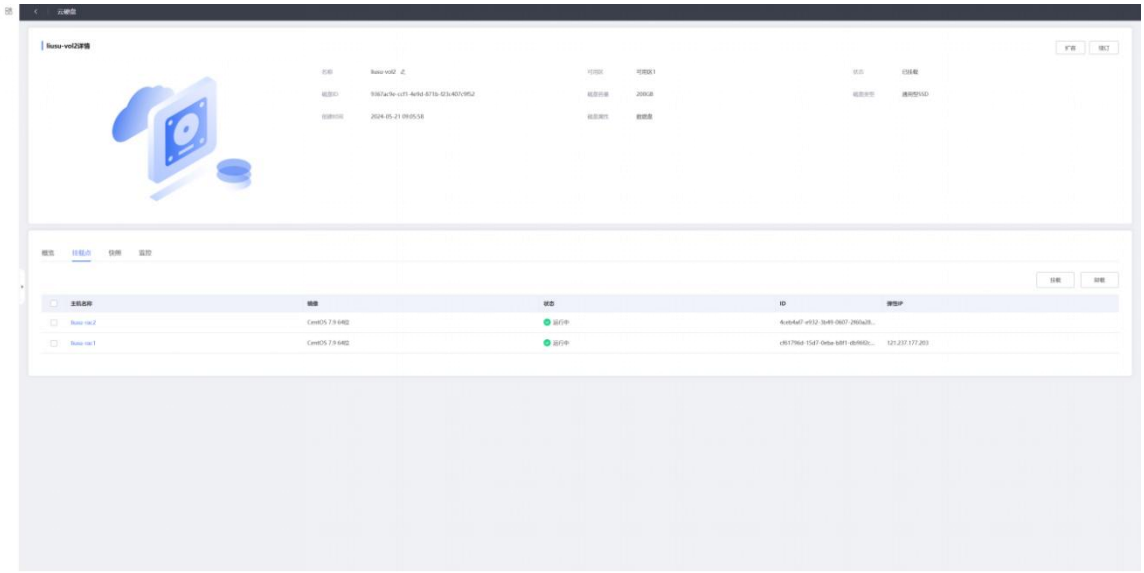
2. 创建云主机。

如下图所示为创建的云主机，共创建两台同规格的云主机：



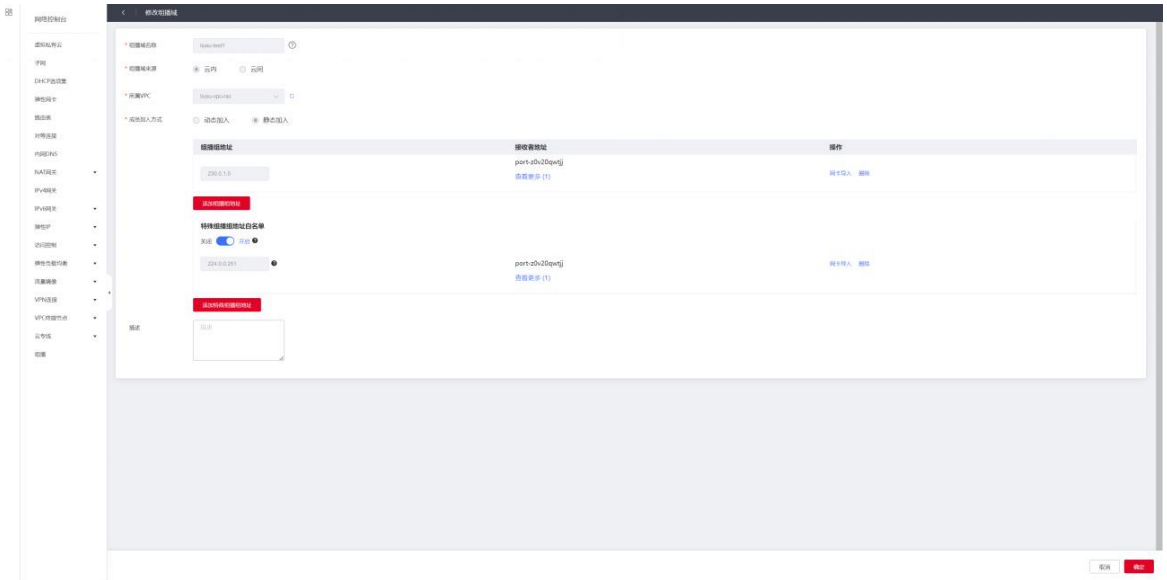
3. 创建共享盘。

创建 2 个共享盘云硬盘并挂载至 2 个云主机上：



4.创建组播。

创建一个组播，所属 VPC 选择上述步骤创建的 VPC，并设置为静态加入方式，配置组播组地址：



节点配置

1.Host 文件配置。

设置两个节点的主机名：

```
hostnamectl set-hostname p19c01
hostnamectl set-hostname p19c02
```

配置两个节点的/etc/hosts/文件：

192.168.10.101 p19c01

192.168.10.102 p19c02

192.168.100.101 p19c01-priv

192.168.100.102 p19c02-priv

192.168.10.103 p19c01-vip

192.168.10.104 p19c02-vip

192.168.10.105 p19c-scan

192.168.10.106 p19c-scan

192.168.10.107 p19c-scan

注意：两个节点均需执行。

2.存储配置。

1）SCSI 磁盘模式场景

配置 udev 规则，步骤如下所示：

```
[root@p19c01 soft]# cat asm.sh #!/bin/bashfor i in a b ;do

echo
"KERNEL==\"sd*\",ENV{DEVTYPE}==\"disk\",SUBSYSTEM==\"block\",PROGRAM=
=\"/usr/lib/udev/scsi_id -g -u -d
\$devnode\",RESULT==\"`/usr/lib/udev/scsi_id -g -u /dev/sd$i`\",
RUN+=\"/bin/sh -c 'mknod /dev/asmdisk$i b  \$major \$minor; chown
grid:asmadmin /dev/asmdisk$i; chmod 0660 /dev/asmdisk$i'\""

done

[root@p19c01 soft]# cat  /etc/udev/rules.d/99-oracle-asmdevices.rules

KERNEL=="sd*",ENV{DEVTYPE}=="disk",SUBSYSTEM=="block",PROGRAM=="/usr/
lib/udev/scsi_id -g -u -d
$devnode",RESULT=="360000000000000000e000000002f0001", RUN+="/bin/sh
```

```
-c 'mknod /dev/asmdiska b  $major $minor; chown grid:asmadmin
/dev/asmdiska; chmod 0660 /dev/asmdiska'"

KERNEL=="sd*",ENV{DEVTYPE}=="disk",SUBSYSTEM=="block",PROGRAM=="/usr/
lib/udev/scsi_id -g -u -d
$devnode",RESULT=="360000000000000000e000000002b0001", RUN+="/bin/sh
-c 'mknod /dev/asmdiskb b  $major $minor; chown grid:asmadmin
/dev/asmdiskb; chmod 0660 /dev/asmdiskb'"

[root@p19c01 soft]# udevadm control --reload

[root@p19c01 soft]# udevadm trigger

[root@p19c01 soft]#

[root@p19c01 soft]# ls -l /dev/ |grep asm

brw-rw----  1 root root      8,    0 May 24 17:08 asmdiska

brw-rw----  1 root root      8,   16 May 24 17:08 asmdiskb

[root@p19c01 soft]#

[root@p19c02 ~]# sh asm.sh

KERNEL=="sd*",ENV{DEVTYPE}=="disk",SUBSYSTEM=="block",PROGRAM=="/usr/
lib/udev/scsi_id -g -u -d
$devnode",RESULT=="360000000000000000e000000002f0001", RUN+="/bin/sh
-c 'mknod /dev/asmdiska b  $major $minor; chown grid:asmadmin
/dev/asmdiska; chmod 0660 /dev/asmdiska'"

KERNEL=="sd*",ENV{DEVTYPE}=="disk",SUBSYSTEM=="block",PROGRAM=="/usr/
lib/udev/scsi_id -g -u -d
$devnode",RESULT=="360000000000000000e000000002b0001", RUN+="/bin/sh
-c 'mknod /dev/asmdiskb b  $major $minor; chown grid:asmadmin
/dev/asmdiskb; chmod 0660 /dev/asmdiskb'"
```

```
[root@p19c02 ~]#

[root@p19c02 ~]#

[root@p19c02 ~]# vim /etc/udev/rules.d/99-oracle-asmdevices.rules

[root@p19c02 ~]#

[root@p19c02 ~]# udevadm control --reload

[root@p19c02 ~]# udevadm trigger

[root@p19c02 ~]#

[root@p19c02 ~]#

[root@p19c02 ~]# ls -l /dev |grep asm

brw-rw----  1 root root      8,   0 May 24 17:09 asmdiska

brw-rw----  1 root root      8,  16 May 24 17:09 asmdiskb
```

1.

注意：两个节点均需执行。

2.

2）VBD 磁盘模式场景。

3.

```
[root@p19c01 ~]# lsblk

NAME   MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT

sr0     11:0    1  380K  0 rom

vda    253:0    0  100G  0 disk

└─vda1 253:1    0  100G  0 part /

vdb    253:16   0  100G  0 disk

vdc    253:32   0  200G  0 disk
```

```
[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]# cat /sys/class/block/vdb/serial
965fcbe9-4f9c-4aa0-9[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]# cat /sys/class/block/vdc/serial
9367ac9e-ccf1-4e9d-8[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]#

[root@p19c01 ~]# vim
/etc/udev/rules.d/75-persistent-net-generator.rules
90-configure-multi-queue.rules     99-gen-net-conf.rules

[root@p19c01 ~]# vim /etc/udev/rules.d/99-oracle-asmdevices.rules

[root@p19c01 ~]#

[root@p19c01 ~]# cat /etc/udev/rules.d/99-oracle-asmdevices.rules

KERNEL=="vd*",ATTR{serial}=="965fcbe9-4f9c-4aa0-9",SYMLINK+="asm_ocr",
OWNER="grid",GROUP="asmadmin",MODE="0660"

KERNEL=="vd*",ATTR{serial}=="9367ac9e-ccf1-4e9d-8",SYMLINK+="asm_data
",OWNER="grid",GROUP="asmadmin",MODE="0660"

[root@p19c01 ~]#

[root@p19c01 ~]#
```

```
[root@p19c01 ~]#
[root@p19c01 ~]# udevadm control --reload
[root@p19c01 ~]#
[root@p19c01 ~]# udevadm trigger
[root@p19c01 ~]#
[root@p19c01 ~]#
[root@p19c01 ~]# ls -l /dev/ |grep asm
lrwxrwxrwx  1 root root           3 May 27 15:30 asm_data -> vdc
lrwxrwxrwx  1 root root           3 May 27 15:30 asm_ocr -> vdb
[root@p19c02 ~]# lsblk
NAME   MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sr0     11:0    1  380K  0 rom
vda    253:0    0  100G  0 disk
└─vda1 253:1    0  100G  0 part /
vdb    253:16   0  100G  0 disk
vdc    253:32   0  200G  0 disk
[root@p19c02 ~]#
[root@p19c02 ~]#
[root@p19c02 ~]# vim /etc/udev/rules.d/99-oracle-asmdevices.rules
[root@p19c02 ~]#
[root@p19c02 ~]#
[root@p19c02 ~]# udevadm control --reload
```

```
[root@p19c02 ~]#

[root@p19c02 ~]# udevadm trigger

[root@p19c02 ~]#

[root@p19c02 ~]#

[root@p19c02 ~]# ls -l /dev/ |grep asm

lrwxrwxrwx  1 root root           3 May 27 15:31 asm_data -> vdc

lrwxrwxrwx  1 root root           3 May 27 15:31 asm_ocr -> vdb

[root@p19c02 ~]#

[root@p19c02 ~]#

[root@p19c02 ~]#

[root@p19c02 ~]# cat /etc/udev/rules.d/99-oracle-asmdevices.rules

KERNEL=="vd*",ATTR{serial}=="965fcbe9-4f9c-4aa0-9",SYMLINK+="asm_ocr",
OWNER="grid",GROUP="asmadmin",MODE="0660"

KERNEL=="vd*",ATTR{serial}=="9367ac9e-ccf1-4e9d-8",SYMLINK+="asm_data
",OWNER="grid",GROUP="asmadmin",MODE="0660"

[root@p19c02 ~]#
```

注意：两个节点均需执行。

avahi-daemon 配置。

```
yum install -y avahi*

systemctl disable avahi-daemon.service
```

注意：两个节点均需执行。

4.系统参数配置。

配置参数文件如下所示：

```
memTotal=$(grep MemTotal /proc/meminfo | awk '{print
$2}')totalMemory=$((memTotal / 2048))shmall=$((memTotal / 4))if
[ $shmall -lt 2097152 ]; then
        shmall=2097152
fishmmax=$((memTotal * 1024 - 1))if [ "$shmmax" -lt 4294967295 ]; then
        shmmax=4294967295
fi
cat <<EOF>>/etc/sysctl.conf
fs.aio-max-nr = 1048576
fs.file-max = 6815744
kernel.shmall = $shmall
kernel.shmmax = $shmmax
kernel.shmmni = 4096
kernel.sem = 250 32000 100 128
net.ipv4.ip_local_port_range = 9000 65500
net.core.rmem_default = 262144
net.core.rmem_max = 4194304
net.core.wmem_default = 262144
net.core.wmem_max = 1048576
net.ipv4.conf.eth0.rp_filter = 1
net.ipv4.conf.eth1.rp_filter = 2
EOF
```

生效系统参数：

```
sysctl -p
```

5.系统资源限制配置。

配置 limits.conf：

```
cat <<EOF>>/etc/security/limits.conf

oracle soft nofile 1024

oracle hard nofile 65536

oracle soft stack 10240

oracle hard stack 32768

oracle soft nproc 2047

oracle hard nproc 16384

oracle hard memlock 134217728

oracle soft memlock 134217728



grid soft nofile 1024

grid hard nofile 65536

grid soft stack 10240

grid hard stack 32768

grid soft nproc 2047

grid hard nproc 16384

EOF
```

配置 pam.d/login：

```
cat <<EOF>>/etc/pam.d/loginsession required pam_limits.so session
required /lib64/security/pam_limits.so
```

```
EOF
```

6.用户及组、目录创建。

用户组创建：

```
/usr/sbin/groupadd -g 54321 oinstall/usr/sbin/groupadd -g 54322
dba/usr/sbin/groupadd -g 54323 oper/usr/sbin/groupadd -g 54324
backupdba/usr/sbin/groupadd -g 54325 dgdba/usr/sbin/groupadd -g 54326
kmdba/usr/sbin/groupadd -g 54327 asmdba/usr/sbin/groupadd -g 54328
asmoper/usr/sbin/groupadd -g 54329 asmadmin/usr/sbin/groupadd -g 54330
racdba
```

用户 grid/oracle 创建：

```
/usr/sbin/useradd -u 11012 -g oinstall -G
asmadmin,asmdba,asmoper,dba,racdba,oper grid/usr/sbin/useradd -u 54321
-g oinstall -G asmdba,dba,backupdba,dgdba,kmdba,racdba,oper oracleecho
"oracle" |passwd oracle --stdinecho "oracle" |passwd grid --stdin

id grid

id oracle

udevadm control --reload-rules

udevadm trigger

ll /dev |grep asm
```

创建软件目录：

```
mkdir -p /u01/app/19.3.0/gridmkdir -p /u01/app/gridmkdir -p
/u01/app/oracle/product/19.3.0/dbmkdir -p /u01/app/oraInventorymkdir
-p /backupmkdir -p /home/oracle/scriptschown -R oracle:oinstall
/backupchown -R oracle:oinstall /home/oracle/scriptschown -R
grid:oinstall /u01chown -R grid:oinstall /u01/app/gridchown -R
```

```
grid:oinstall /u01/app/19.3.0/gridchown -R grid:oinstall
/u01/app/oraInventorychown -R oracle:oinstall /u01/app/oraclechmod -R
775 /u01
```

注意：两个节点均需执行。

7.环境变量配置。

用户 grid 环境变量：

```
cat
<<EOF>>/home/grid/.bash_profile################OracleBegin###########
###############

umask 022export TMP=/tmpexport TMPDIR=\$TMPexport
NLS_LANG=AMERICAN_AMERICA.AL32UTF8export
ORACLE_BASE=/u01/app/gridexport
ORACLE_HOME=/u01/app/19.3.0/gridexport ORACLE_TERM=xtermexport
TNS_ADMIN=\$ORACLE_HOME/network/adminexport
LD_LIBRARY_PATH=\$ORACLE_HOME/lib:/lib:/usr/libexport
ORACLE_SID=+ASM1export PATH=/usr/sbin:\$PATHexport
PATH=\$ORACLE_HOME/bin:\$ORACLE_HOME/OPatch:\$PATH

alias sas='sqlplus / as sysasm'export
PS1="[\`whoami\`@\`hostname\`:"'\$PWD]\$ '

EOF

cat
<<EOF>>/home/grid/.bash_profile################OracleBegin###########
###############

umask 022export TMP=/tmpexport TMPDIR=\$TMPexport
NLS_LANG=AMERICAN_AMERICA.AL32UTF8export
ORACLE_BASE=/u01/app/gridexport
ORACLE_HOME=/u01/app/19.3.0/gridexport ORACLE_TERM=xtermexport
```

```
TNS_ADMIN=\$ORACLE_HOME/network/adminexport
LD_LIBRARY_PATH=\$ORACLE_HOME/lib:/lib:/usr/libexport
ORACLE_SID=+ASM2export PATH=/usr/sbin:\$PATHexport
PATH=\$ORACLE_HOME/bin:\$ORACLE_HOME/OPatch:\$PATH

alias sas='sqlplus / as sysasm'export
PS1="[\`whoami\`@\`hostname\`:"'\$PWD]\$ '

EOF
```

用户 oracle 环境变量：

```
cat <<EOF>>/home/oracle/.bash_profile

##################OracleBegin##########################

umask 022export TMP=/tmpexport TMPDIR=\$TMPexport
NLS_LANG=AMERICAN_AMERICA.AL32UTF8export
ORACLE_BASE=/u01/app/oracleexport
ORACLE_HOME=\$ORACLE_BASE/product/19.3.0/dbexport
ORACLE_HOSTNAME=p19c01export ORACLE_TERM=xtermexport
TNS_ADMIN=\$ORACLE_HOME/network/adminexport
LD_LIBRARY_PATH=\$ORACLE_HOME/lib:/lib:/usr/libexport
ORACLE_SID=p19c01export PATH=/usr/sbin:\$PATHexport
PATH=\$ORACLE_HOME/bin:\$ORACLE_HOME/OPatch:\$PATH

alias sas='sqlplus / as sysdba'export
PS1="[\`whoami\`@\`hostname\`:"'\$PWD]\$ '

##################OracleEnd##########################

EOF

cat <<EOF>>/home/oracle/.bash_profile

##################OracleBegin##########################
```

```
umask 022export TMP=/tmpexport TMPDIR=\$TMPexport
NLS_LANG=AMERICAN_AMERICA.AL32UTF8export
ORACLE_BASE=/u01/app/oracleexport
ORACLE_HOME=\$ORACLE_BASE/product/19.3.0/dbexport
ORACLE_HOSTNAME=p19c02export ORACLE_TERM=xtermexport
TNS_ADMIN=\$ORACLE_HOME/network/adminexport
LD_LIBRARY_PATH=\$ORACLE_HOME/lib:/lib:/usr/libexport
ORACLE_SID=p19c02export PATH=/usr/sbin:\$PATHexport
PATH=\$ORACLE_HOME/bin:\$ORACLE_HOME/OPatch:\$PATH

alias sas='sqlplus / as sysdba'export
PS1="[\`whoami\`@\`hostname\`:"'\$PWD]\$ '

#################OracleEnd############################

EOF
```

注意：两个节点均需执行。

8.SWAP 配置。

```
dd if=/dev/zero of=/swap bs=1M count=16390

mkswap /swap

swapon /swap

echo "/swap swap swap default 0 0" >> /etc/fstab
```

注意：两个节点均需执行。

9.安装依赖包。

```
yum groupinstall -y "Server with GUI"



yum install -y bc \

binutils \
```

```
compat-libcap1 \
compat-libstdc++-33 \
gcc \
gcc-c++ \
elfutils-libelf \
elfutils-libelf-devel \
glibc \
glibc-devel \
ksh \
libaio \
libaio-devel \
libgcc \
libstdc++ \
libstdc++-devel \
libxcb \
libX11 \
libXau \
libXi \
libXtst \
libXrender \
libXrender-devel \
make \
```

```
net-tools \

nfs-utils \

smartmontools \

sysstat \

e2fsprogs \

e2fsprogs-libs \

fontconfig-devel \

expect \

unzip \

openssh-clients \

readline* \

tigervnc* \

psmisc
```

注意：两个节点均需执行。

10 上传并解压安装包。

注意：只在第一个节点执行。

```
mkdir /soft
```

上传安装包到该目录下：

```
[root@p19c01 soft]# ls -lh

total 5.6G

-rw-r--r-- 1 root root 2.9G May 21 10:44 LINUX.X64_193000_db_home.zip

-rw-r--r-- 1 root root 2.7G May 21 09:56 LINUX.X64_193000_grid_home.zip
```

解压 grid 安装包：

```
[root@p19c01 ~]# cd /soft
```

```
[root@p19c01 soft]# ls -l
total 5809468
-rw-r--r-- 1 root root 3059705302 May 21 10:44
LINUX.X64_193000_db_home.zip
-rw-r--r-- 1 root root 2889184573 May 21 09:56
LINUX.X64_193000_grid_home.zip
[root@p19c01 soft]#
[root@p19c01 soft]# chown -R grid:oinstall /soft
[root@p19c01 soft]#
[root@p19c01 soft]# su - grid -c "unzip -q
/soft/LINUX.X64_193000_grid_home.zip -d /u01/app/19.3.0/grid/"
```

解压 oracle 安装包：

```
[root@p19c01 soft]# ls -lh
total 5.6G
-rw-r--r-- 1 grid oinstall 2.9G May 21 10:44 LINUX.X64_193000_db_home.zip
-rw-r--r-- 1 grid oinstall 2.7G May 21 09:56
LINUX.X64_193000_grid_home.zip
[root@p19c01 soft]#
[root@p19c01 soft]# chown -R oracle:oinstall /soft
[root@p19c01 soft]#
[root@p19c01 soft]# su - oracle -c "unzip -q
/soft/LINUX.X64_193000_db_home.zip -d
/u01/app/oracle/product/19.3.0/db/"
```

安装 cvuqdisk（两个节点都需执行）：

```
[root@p19c01 soft]# cd /u01/app/19.3.0/grid/cv/rpm/

[root@p19c01 rpm]# ls -l

total 12

-rw-r--r-- 1 grid oinstall 11412 Mar 13  2019 cvuqdisk-1.0.10-1.rpm

[root@p19c01 rpm]# rpm -ivh cvuqdisk-1.0.10-1.rpm

Preparing...
########################################## [100%]

Using default group oinstall to install package

Updating / installing...

   1:cvuqdisk-1.0.10-1
########################################## [100%]

[root@p19c01 rpm]# scp cvuqdisk-1.0.10-1.rpm root@p19c02:/root

[root@p19c01 rpm]#

[root@p19c01 rpm]# ssh p19c02

root@p19c02's password:

Last failed login: Thu May 23 14:22:55 CST 2024 from 183.81.169.238 on
ssh:notty

There were 22 failed login attempts since the last successful login.

Last login: Thu May 23 09:12:17 2024 from 101.227.46.166

[root@p19c02 ~]# ls -lh

total 12K

-rw-r--r-- 1 root root 12K May 23 15:05 cvuqdisk-1.0.10-1.rpm
```

```
[root@p19c02 ~]# rpm -ivh cvuqdisk-1.0.10-1.rpm

Preparing...
################################## [100%]

Using default group oinstall to install package

Updating / installing...

   1:cvuqdisk-1.0.10-1
################################## [100%]
```

安装 Grid 软件

配置 grid 用户 vnc 图形界面：

```
[root@p19c01 ~]# chown -R grid:oinstall /soft

[root@p19c01 ~]#

[root@p19c01 ~]# su - grid

Last login: Thu May 23 14:05:58 CST 2024 on pts/0

[grid@p19c01:/home/grid]$

[grid@p19c01:/home/grid]$ vncserver


You will require a password to access your desktops.


Password:

Verify:

Would you like to enter a view-only password (y/n)? n

A view-only password is not used
```

```
xauth:  file /home/grid/.Xauthority does not exist

New 'p19c01:1 (grid)' desktop is p19c01:1


Creating default startup script /home/grid/.vnc/xstartup

Creating default config /home/grid/.vnc/config

Starting applications specified in /home/grid/.vnc/xstartup

Log file is /home/grid/.vnc/p19c01:1.log


[grid@p19c01:/home/grid]$

[grid@p19c01:/home/grid]$

[grid@p19c01:/home/grid]$ ps aux |grep vnc

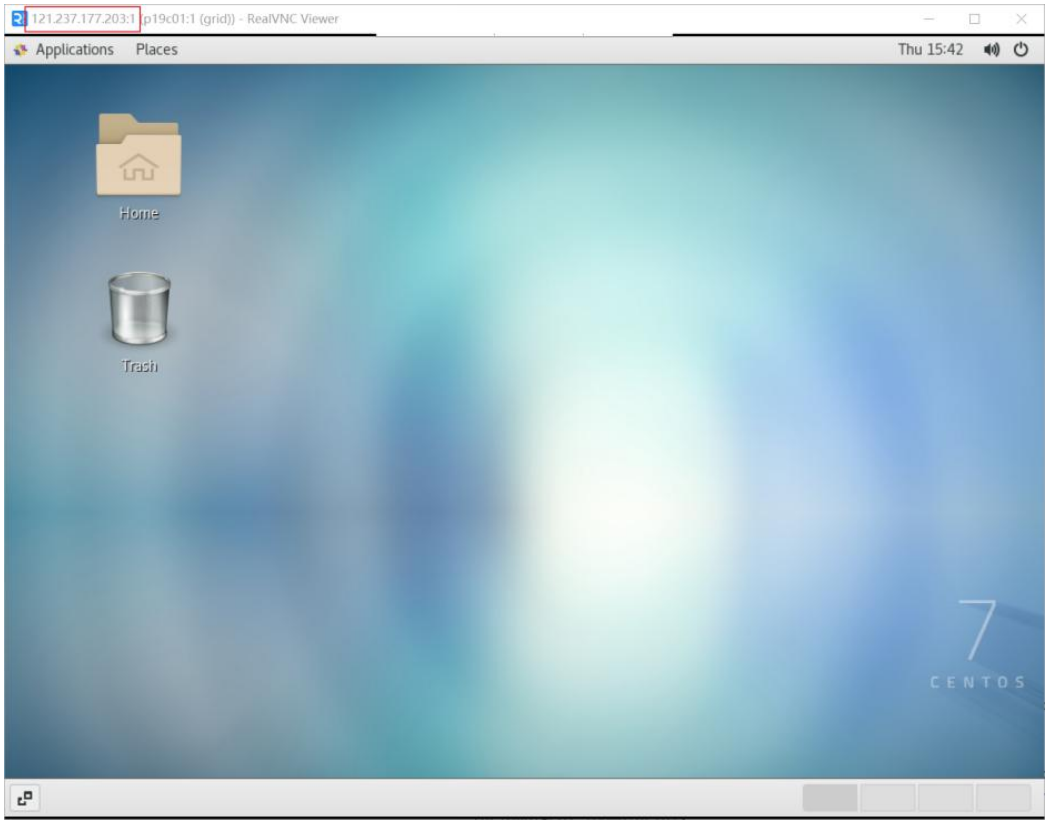
grid      1090  1.4  0.2 286708 46392 pts/0    Sl   15:39   0:00
/bin/Xvnc :1 -auth /home/grid/.Xauthority -desktop p19c01:1 (grid) -fp
catalogue:/etc/X11/fontpath.d -geometry 1024x768 -httpd
/usr/share/vnc/classes -pn -rfbauth /home/grid/.vnc/passwd -rfbport
5901 -rfbwait 30000

grid      1103  0.0  0.0 113288  1200 pts/0    S    15:39   0:00
/bin/sh /home/grid/.vnc/xstartup

grid      1898  0.0  0.0 112820   980 pts/0    S+   15:39   0:00 grep
--color=auto vnc

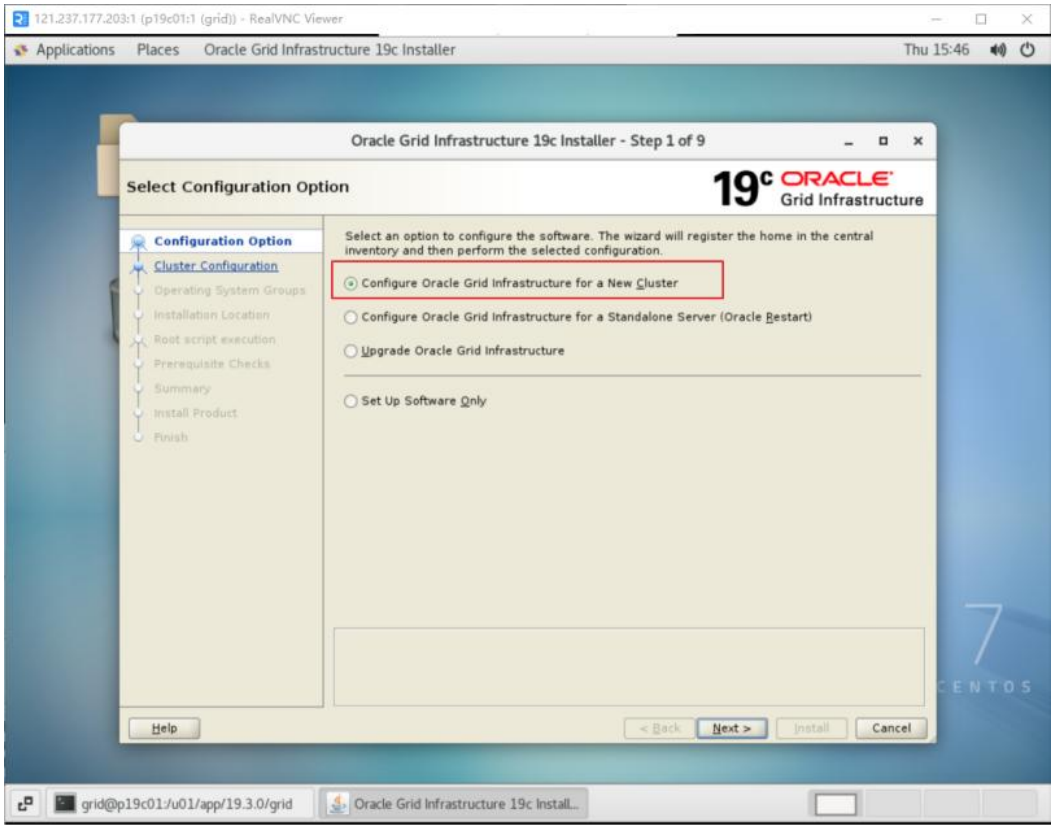
[grid@p19c01:/home/grid]$
```

在 vnc 客户端通过 VNC Viewer 来访问，输入地址为节点 1 的弹性 EIP 加上一个端口 1：

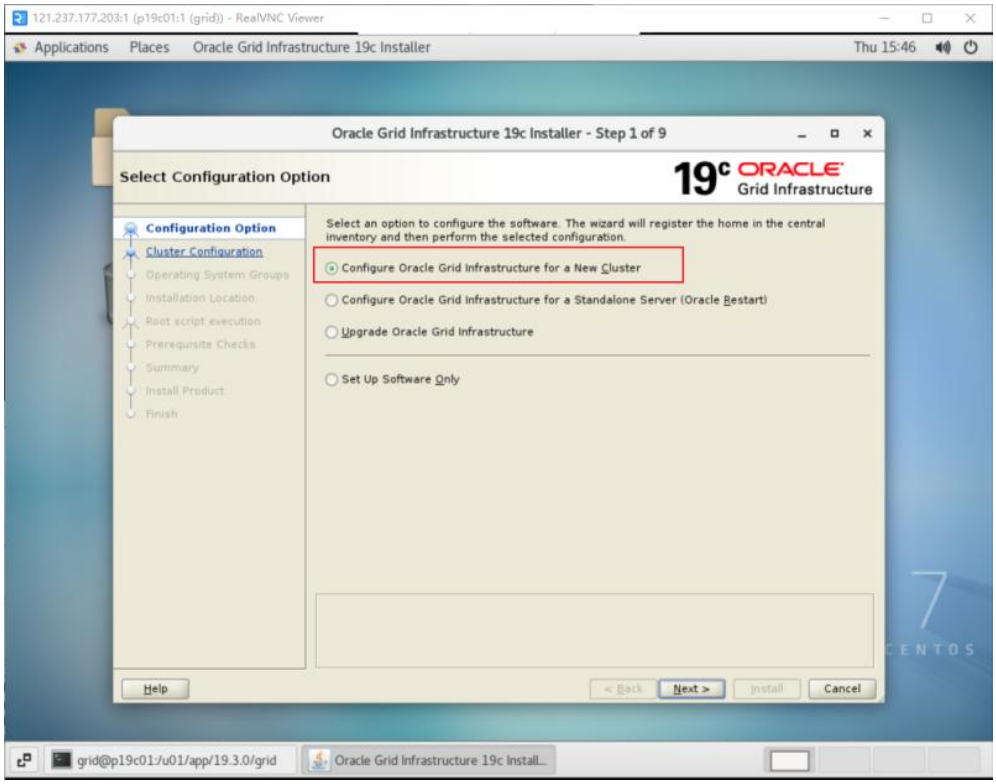


如下图所示，开始执行安装：

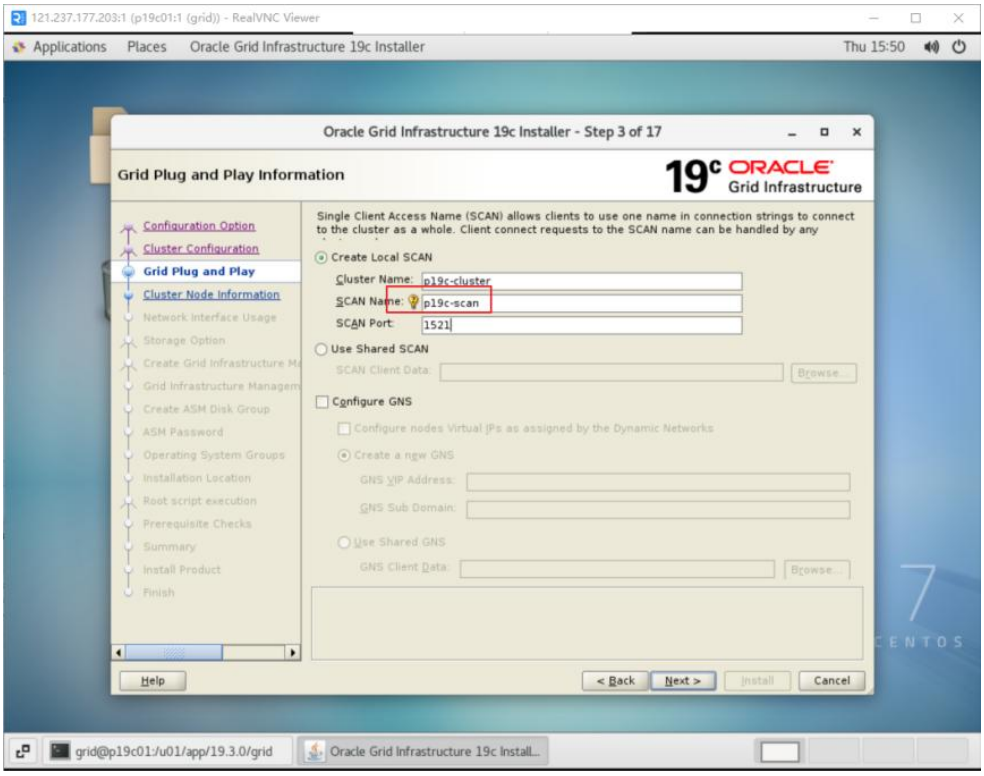
选择集群安装：

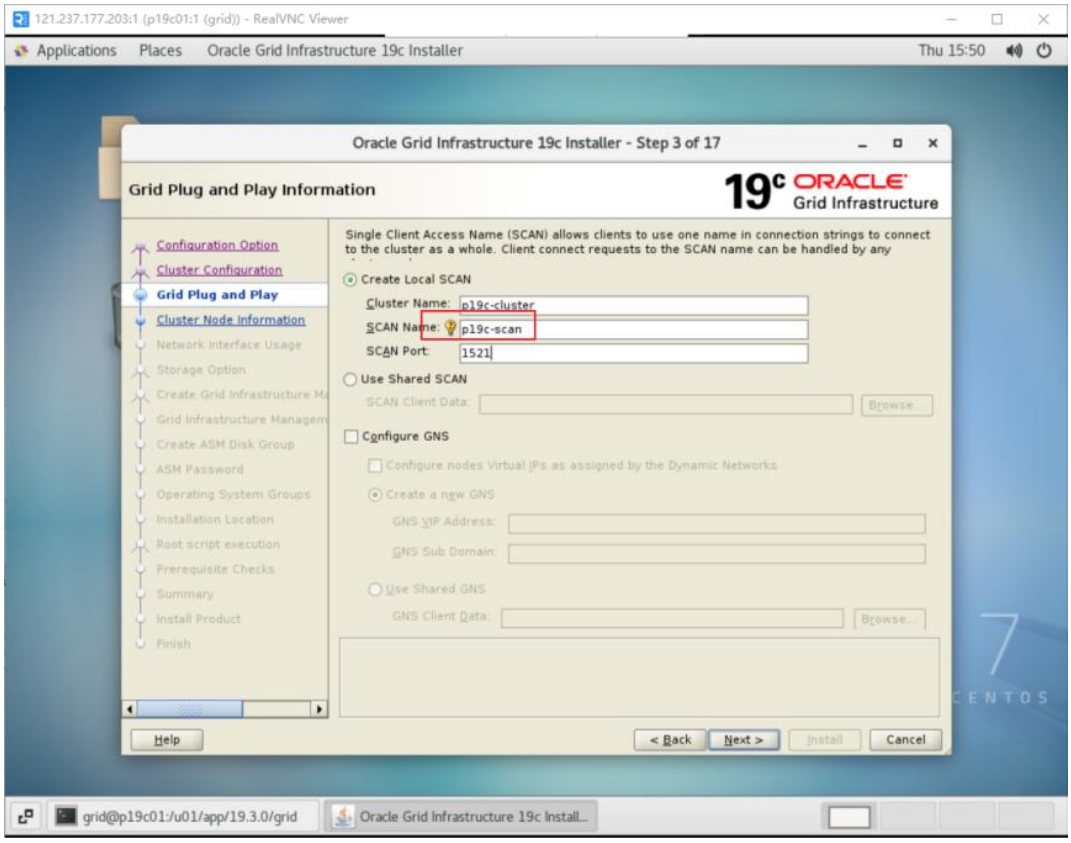


选择 Standalone 模式：

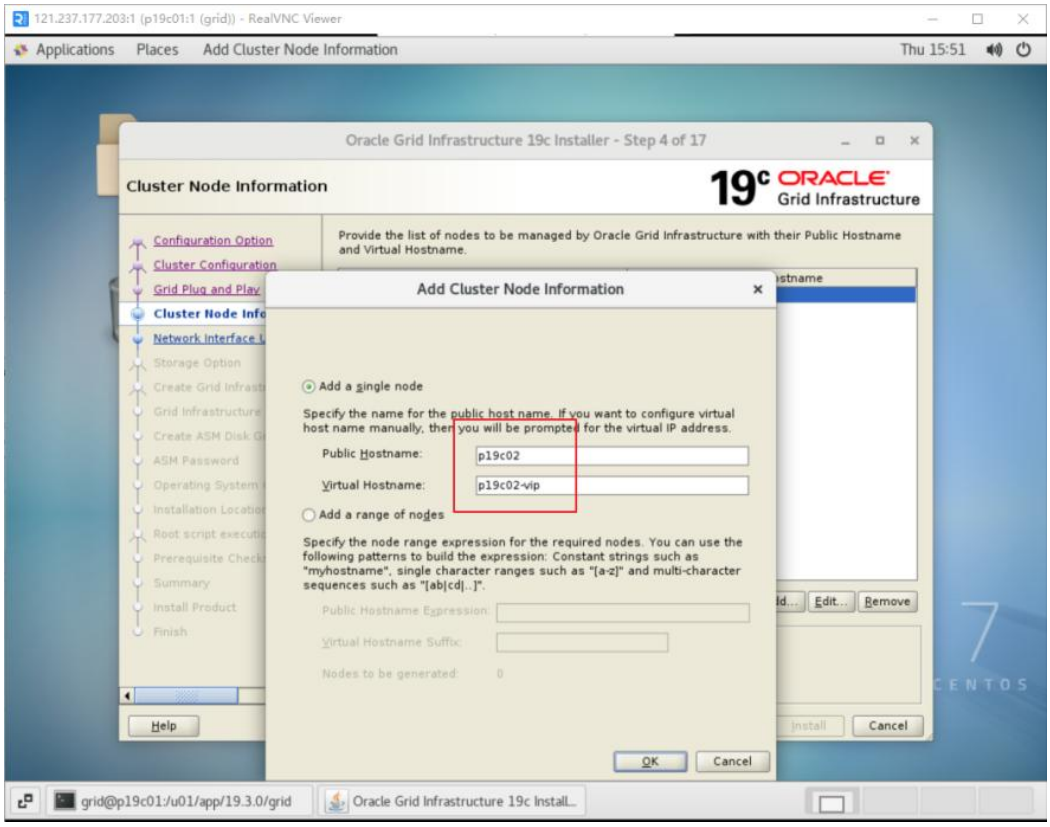


修改 scan 名称，与 hosts 文件配置 scan 名称保持一致：





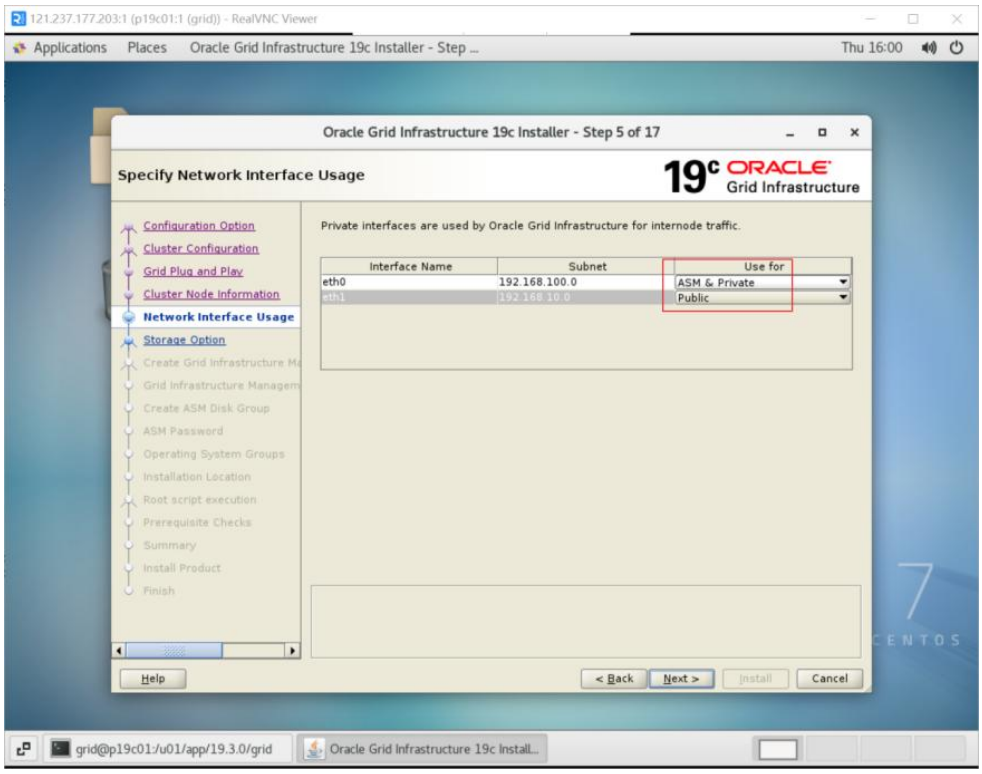
添加节点 2 信息，进行互信：



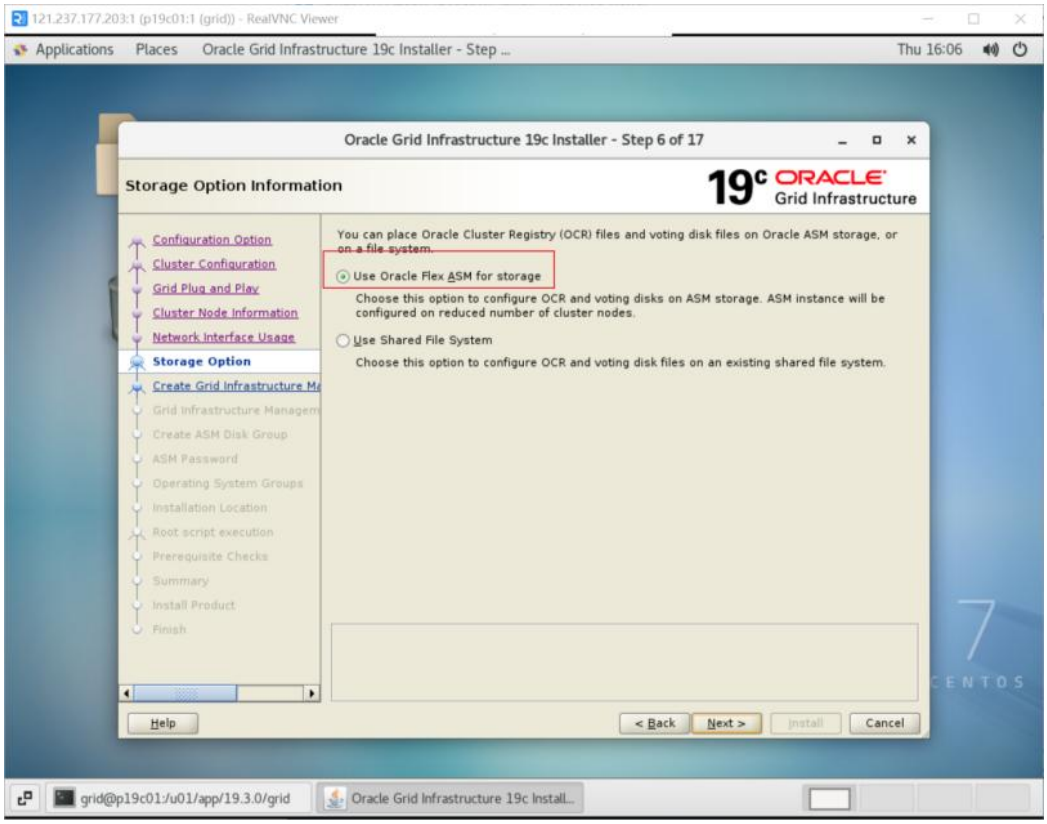
输入 grid 用户密码，创建用户时两节点必须保持一致。

先执行 Setup，再执行 Test：

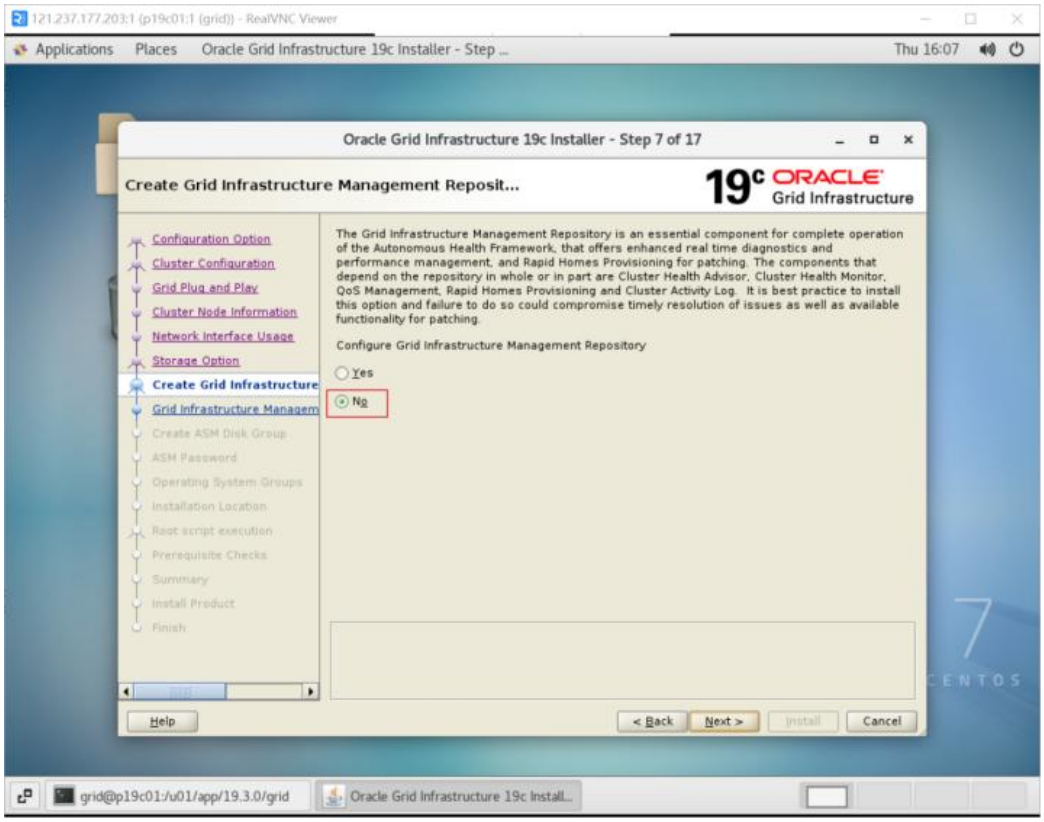
在网卡接口用途确认界面，修改私网网络用途为 ASM&Private：



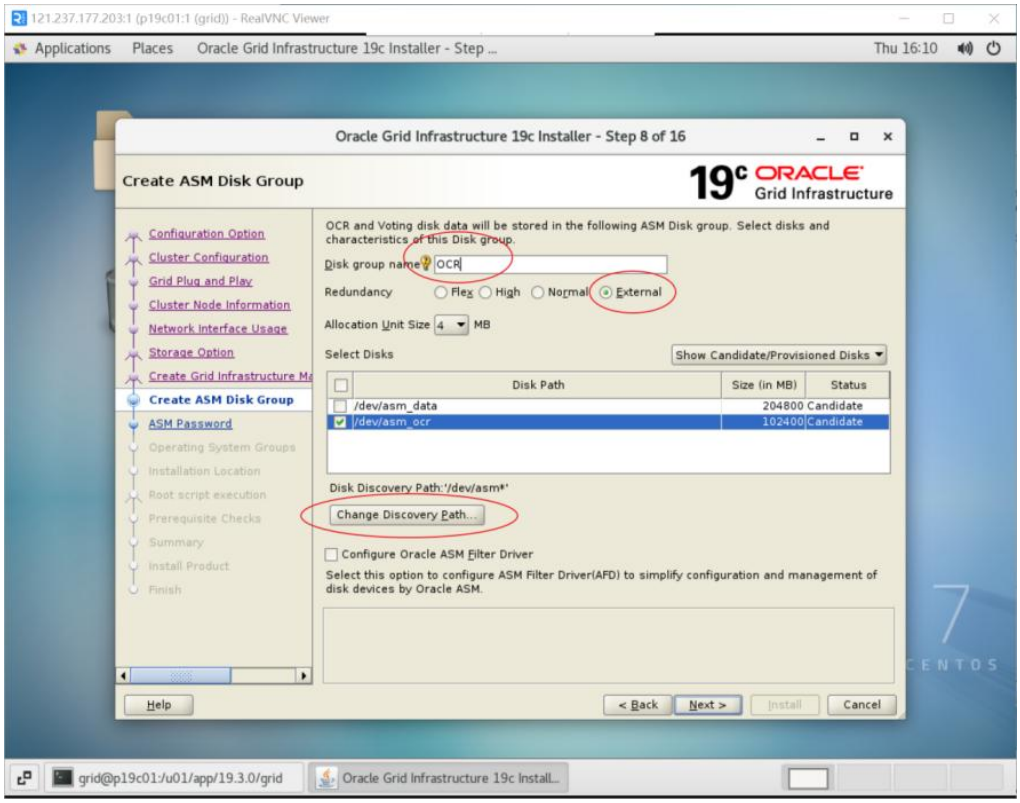
选择存储类型：



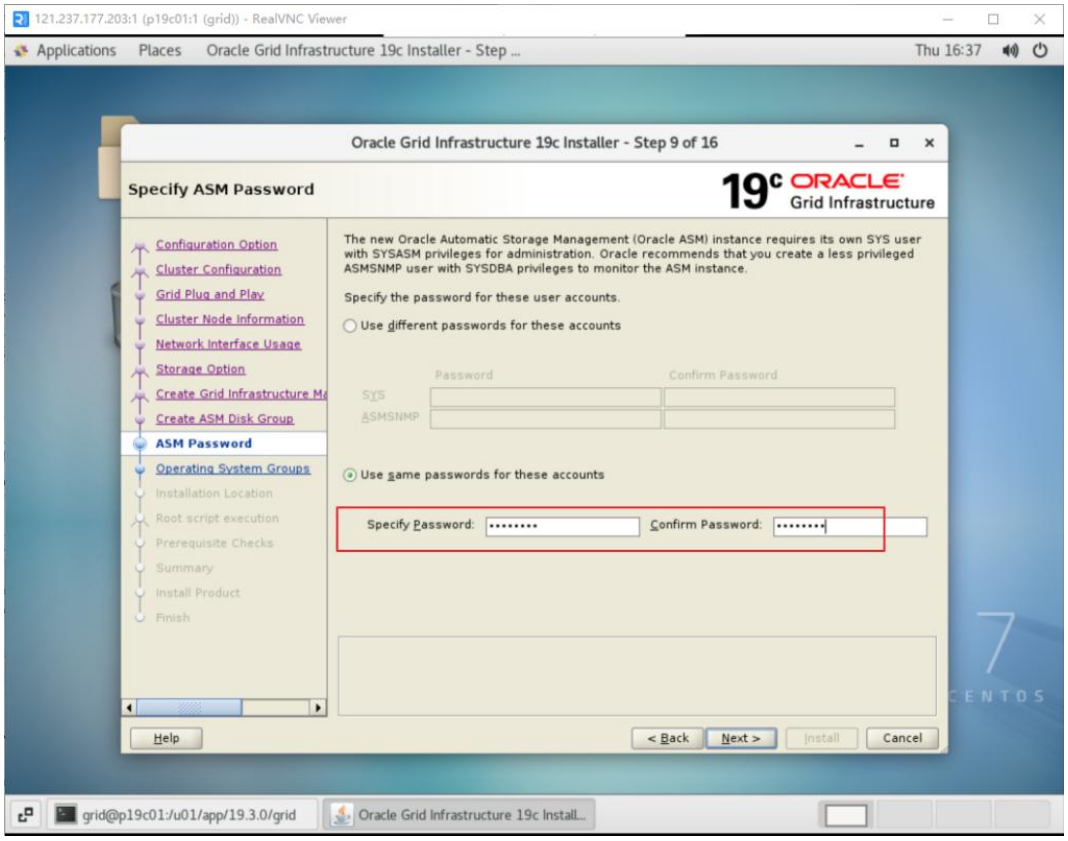
步骤 GRIM 采用默认不安装的选择：



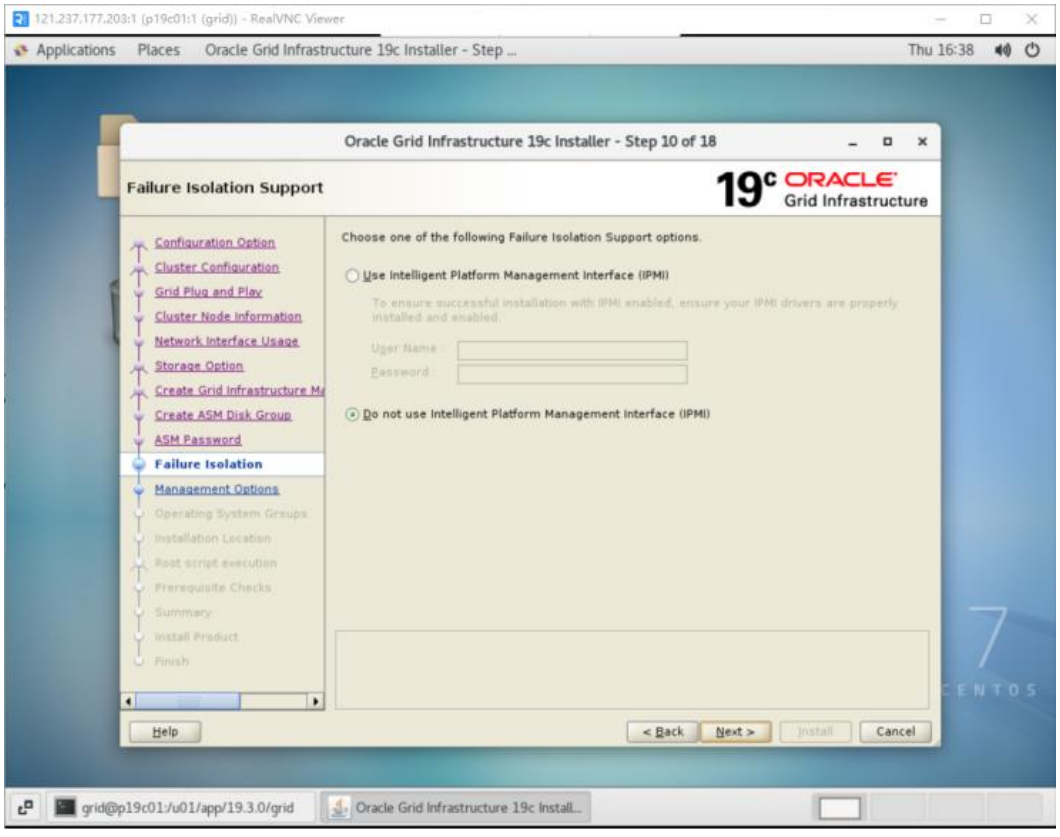
安装填写 OCR 盘，一块盘冗余 External，目录选择 udev 绑定的路径：



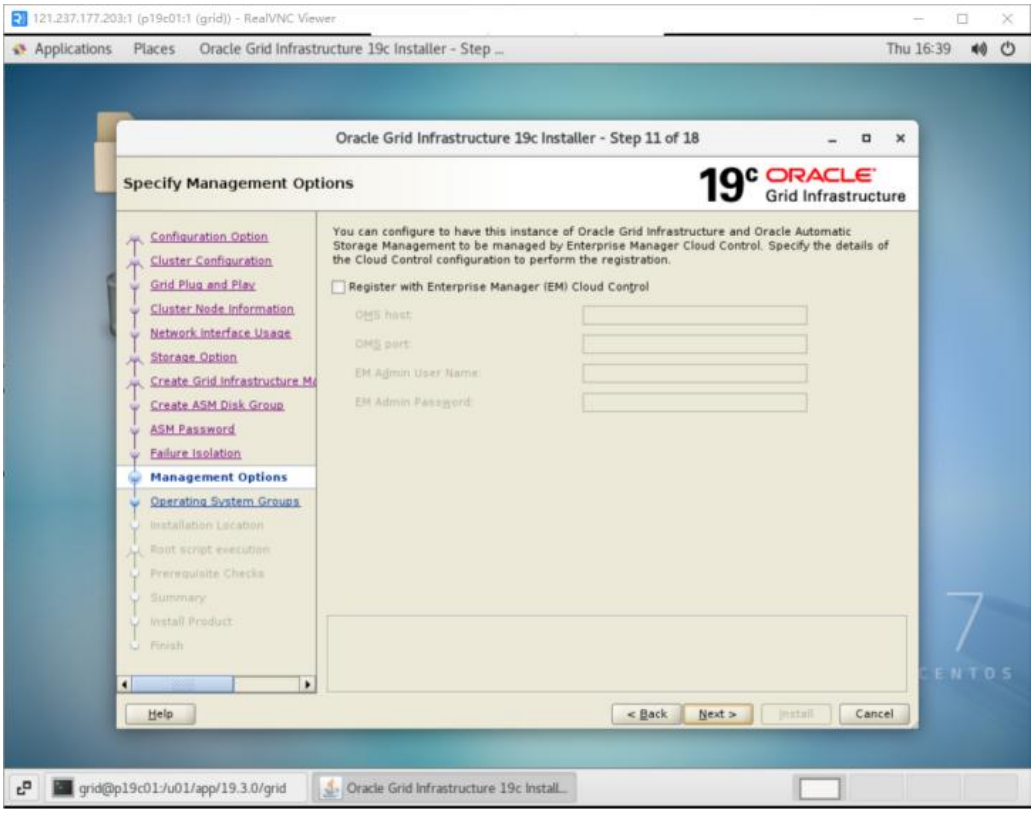
填写 ASM 密码：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications Places Oracle Grid Infrastructure 19c Installer - Step ... Thu 16:37
Oracle Grid Infrastructure 19c Installer - Step 9 of 16
Specify ASM Password
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure Ma
Create ASM Disk Group
ASM Password
Operating System Groups
Installation Location
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
The new Oracle Automatic Storage Management (Oracle ASM) instance requires its own SYS user with SYSASM privileges for administration. Oracle recommends that you create a less privileged ASMSNMP user with SYSDBA privileges to monitor the ASM instance.
Specify the password for these user accounts.
Use different passwords for these accounts
Password
Confirm Password
SYS
ASMSNMP
Use same passwords for these accounts
Specify Password:
Confirm Password:
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

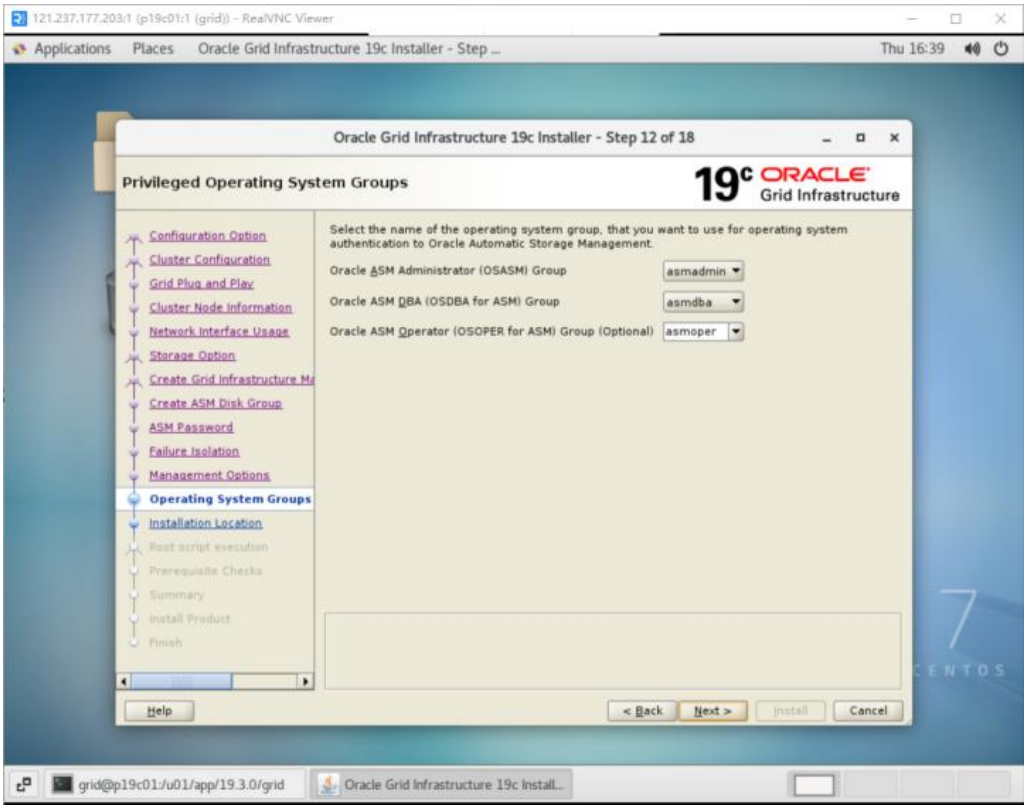
默认：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications Places Oracle Grid Infrastructure 19c Installer - Step ... Thu 16:38
Oracle Grid Infrastructure 19c Installer - Step 10 of 18
Failure Isolation Support
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure Ma
Create ASM Disk Group
ASM Password
Failure Isolation
Management Options
Operating System Groups
Installation Location
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
Choose one of the following Failure Isolation Support options.
Use Intelligent Platform Management Interface (IPMI)
To ensure successful installation with IPMI enabled, ensure your IPMI drivers are properly installed and enabled.
User Name
Password
Do not use Intelligent Platform Management Interface (IPMI)
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

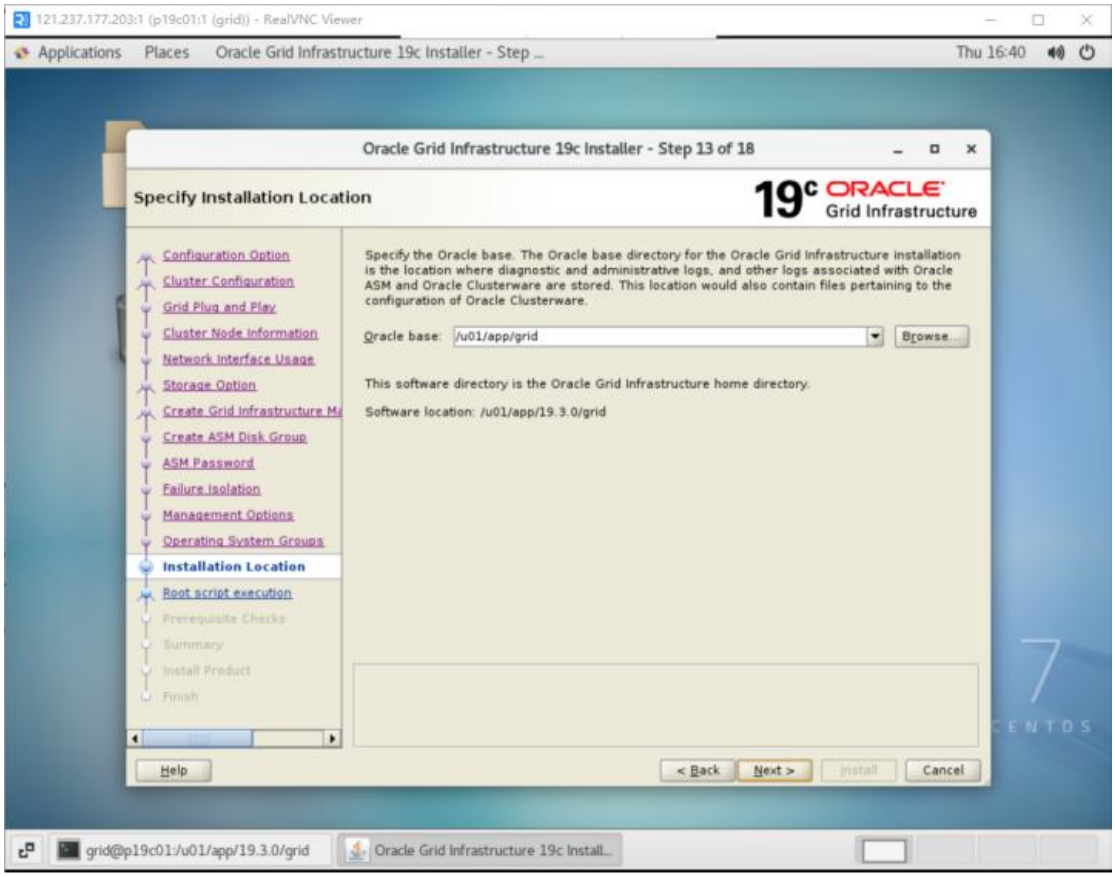
默认：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications
Places
Oracle Grid Infrastructure 19c Installer - Step ...
Thu 16:39
Oracle Grid Infrastructure 19c Installer - Step 11 of 18
Specify Management Options
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure Ma
Create ASM Disk Group
ASM Password
Failure Isolation
Management Options
Operating System Groups
Installation Location
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
You can configure to have this instance of Oracle Grid Infrastructure and Oracle Automatic Storage Management to be managed by Enterprise Manager Cloud Control. Specify the details of the Cloud Control configuration to perform the registration.
Register with Enterprise Manager (EM) Cloud Control
OMS host:
OMS port:
EM Admin User Name:
EM Admin Password:
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

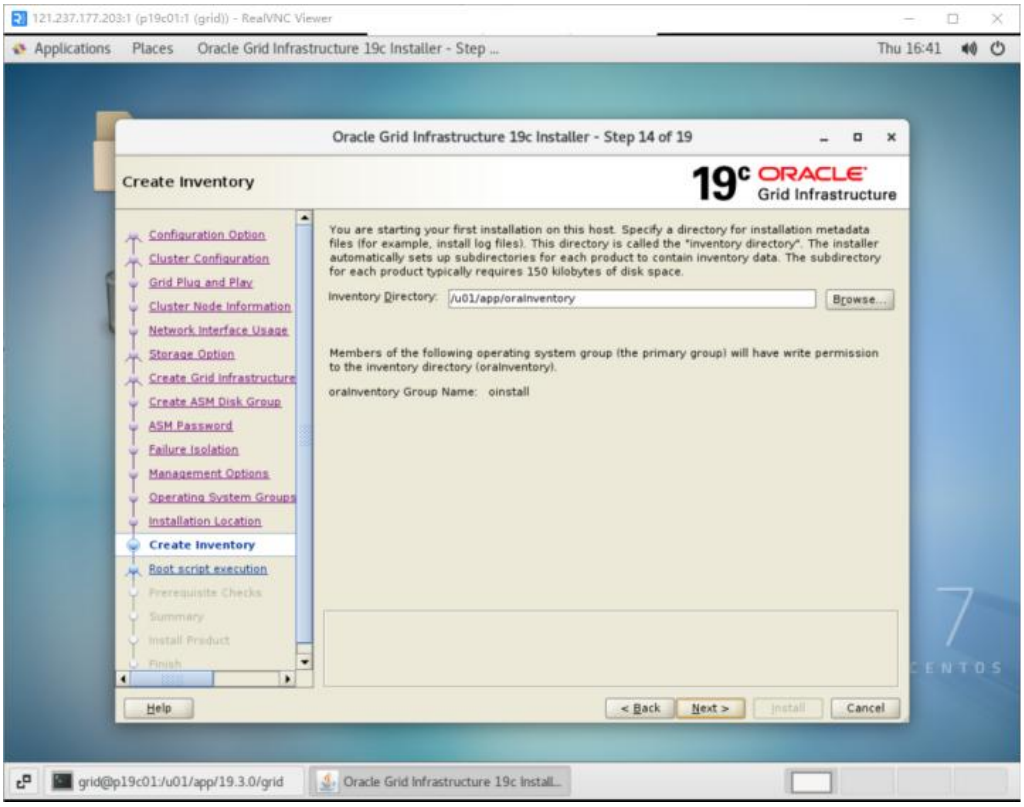
默认：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications
Places
Oracle Grid Infrastructure 19c Installer - Step ...
Thu 16:39
Oracle Grid Infrastructure 19c Installer - Step 12 of 18
Privileged Operating System Groups
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure Ma
Create ASM Disk Group
ASM Password
Failure Isolation
Management Options
Operating System Groups
Installation Location
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
Select the name of the operating system group, that you want to use for operating system authentication to Oracle Automatic Storage Management.
Oracle ASM Administrator (OSASM) Group
asmadmin
Oracle ASM DBA (OSDBA for ASM) Group
asmdba
Oracle ASM Operator (OSOPER for ASM) Group (Optional)
asmoper
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

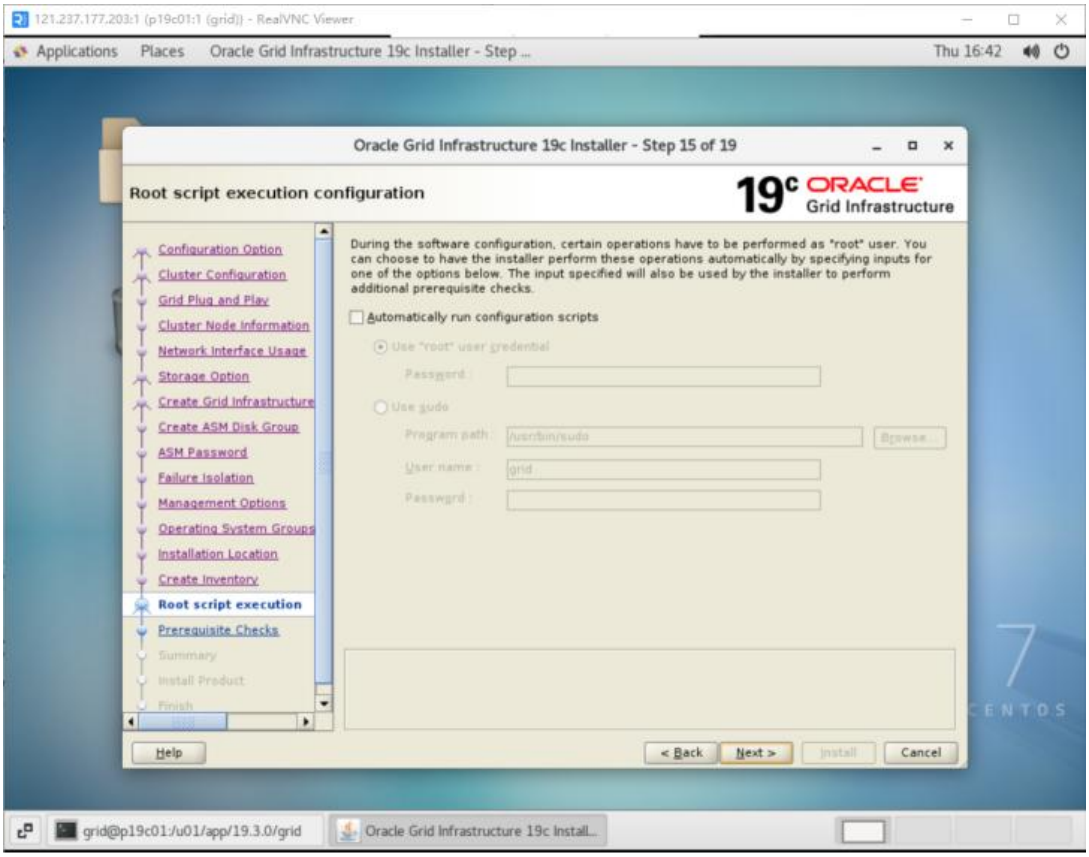
默认：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications
Places
Oracle Grid Infrastructure 19c Installer - Step ...
Thu 16:40
Oracle Grid Infrastructure 19c Installer - Step 13 of 18
Specify Installation Location
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure Ma
Create ASM Disk Group
ASM Password
Failure Isolation
Management Options
Operating System Groups
Installation Location
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
Specify the Oracle base. The Oracle base directory for the Oracle Grid Infrastructure installation is the location where diagnostic and administrative logs, and other logs associated with Oracle ASM and Oracle Clusterware are stored. This location would also contain files pertaining to the configuration of Oracle Clusterware.
Oracle base: /u01/app/grid
Browse...
This software directory is the Oracle Grid Infrastructure home directory.
Software location: /u01/app/19.3.0/grid
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

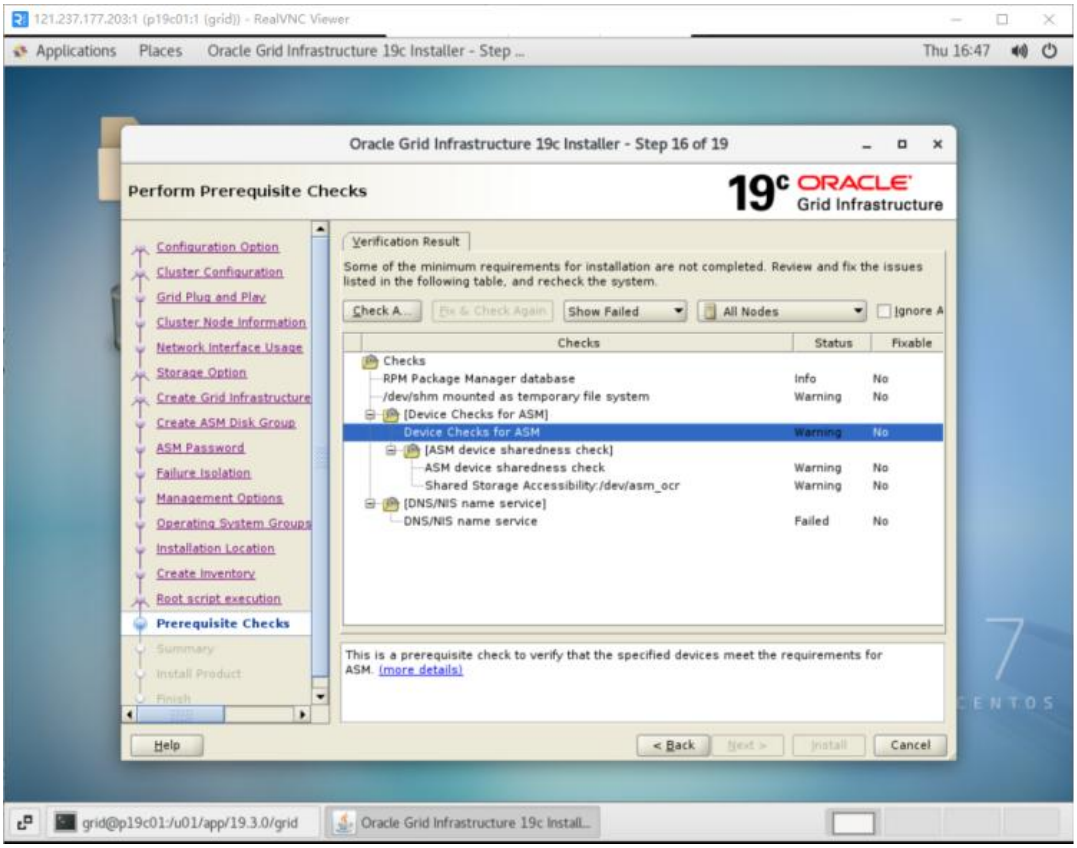
默认：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications Places Oracle Grid Infrastructure 19c Installer - Step ... Thu 16:41
Oracle Grid Infrastructure 19c Installer - Step 14 of 19
Create Inventory
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure
Create ASM Disk Group
ASM Password
Failure Isolation
Management Options
Operating System Groups
Installation Location
Create Inventory
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
You are starting your first installation on this host. Specify a directory for installation metadata files (for example, install log files). This directory is called the "inventory directory". The installer automatically sets up subdirectories for each product to contain inventory data. The subdirectory for each product typically requires 150 kilobytes of disk space.
Inventory Directory: /u01/app/oraInventory
Browse...
Members of the following operating system group (the primary group) will have write permission to the inventory directory (oraInventory).
oraInventory Group Name: oinstall
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

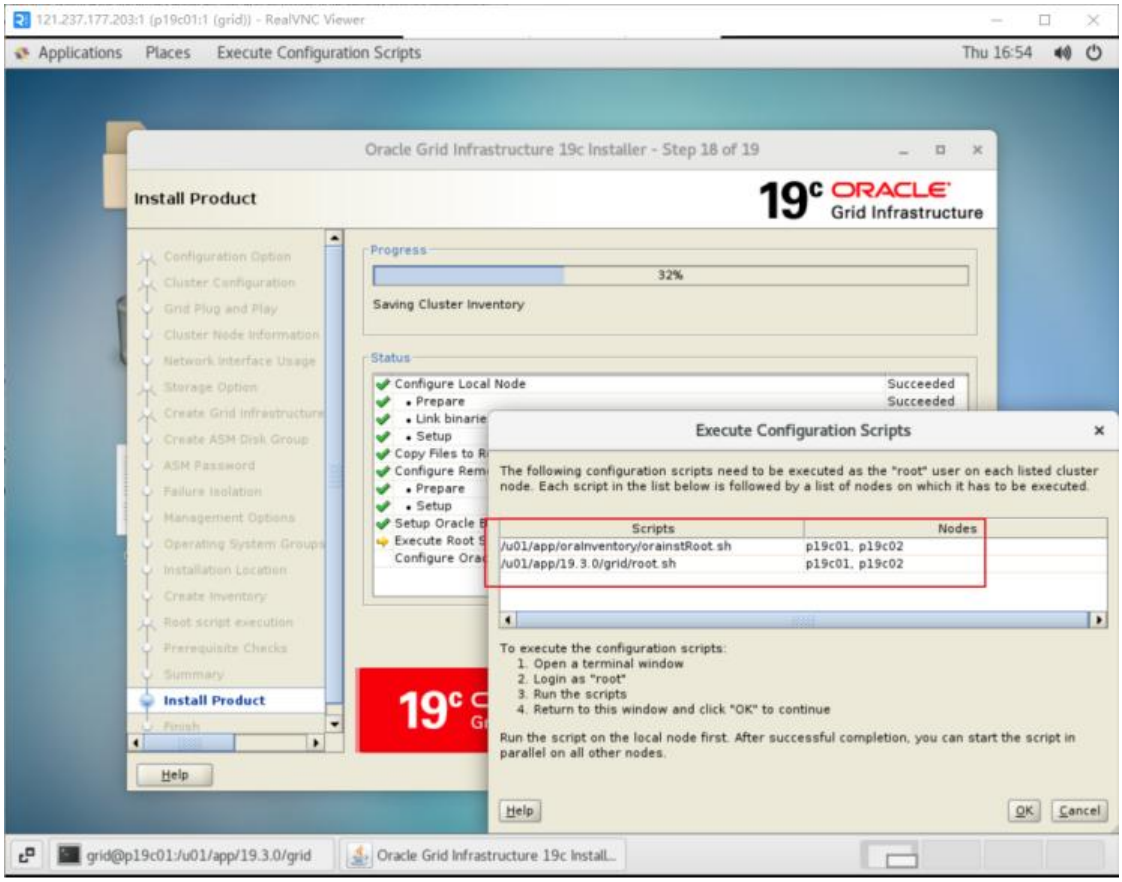
默认：


121.237.177.203:1 (p19c01:1 (grid)) - RealVNC Viewer
Applications Places Oracle Grid Infrastructure 19c Installer - Step ... Thu 16:42
Oracle Grid Infrastructure 19c Installer - Step 15 of 19
Root script execution configuration
19c ORACLE Grid Infrastructure
Configuration Option
Cluster Configuration
Grid Plug and Play
Cluster Node Information
Network Interface Usage
Storage Option
Create Grid Infrastructure
Create ASM Disk Group
ASM Password
Failure Isolation
Management Options
Operating System Groups
Installation Location
Create Inventory
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
During the software configuration, certain operations have to be performed as "root" user. You can choose to have the installer perform these operations automatically by specifying inputs for one of the options below. The input specified will also be used by the installer to perform additional prerequisite checks.
Automatically run configuration scripts
Use "root" user credential
Password:
Use sudo
Program path: /usr/bin/sudo
Browse...
User name: grid
Password:
Help
< Back
Next >
Install
Cancel
grid@p19c01:/u01/app/19.3.0/grid
Oracle Grid Infrastructure 19c Install...

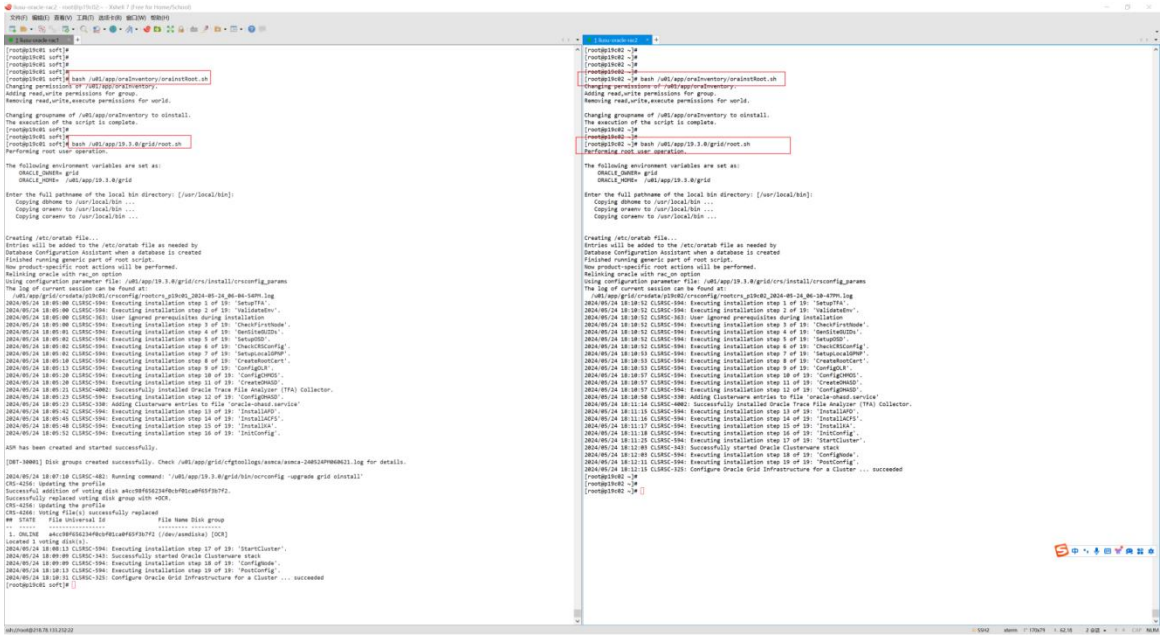
安装预检查结果：



选择忽略按钮 Ignore All，开始进行安装：

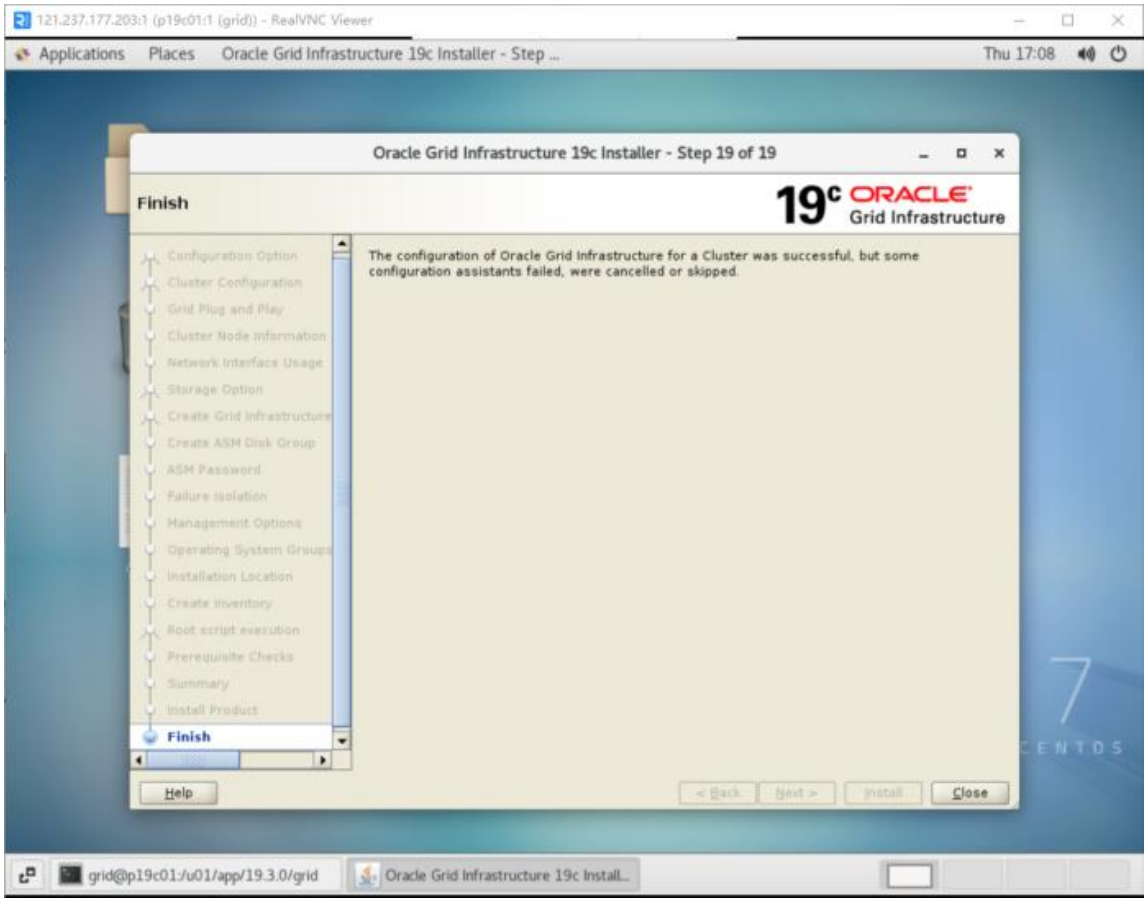


在两个节点上按顺序分别执行上述脚本：



两个节点都执行完成后，点击 OK，继续安装。

安装完成：



这个时候查看服务状态发现只有一个节点是在线的：

```
[grid@p19c01:/home/grid]$ crsctl stat res
-t--------------------------------------------------------------------------
-------------

Name           Target  State        Server                   State
details

--------------------------------------------------------------------------
-----------Local Resources
--------------------------------------------------------------------------
-----------

ora.LISTENER.lsnr                 ONLINE  ONLINE       p19c01
STABLE                ONLINE  ONLINE       p19c02
STABLE
```

```
ora.chad                ONLINE  ONLINE       p19c01
STABLE                  ONLINE  ONLINE       p19c02
STABLE

ora.net1.network                ONLINE  ONLINE       p19c01
STABLE                  ONLINE  ONLINE       p19c02
STABLE

ora.ons                 ONLINE  ONLINE       p19c01
STABLE                  ONLINE  ONLINE       p19c02
STABLE

--------------------------------------------------------------------------------
----------Cluster Resources
--------------------------------------------------------------------------------
----------

ora.ASMNET1LSNR_ASM.lsnr(ora.asmgroup)
      1         ONLINE  ONLINE       p19c01                   STABLE
      2         ONLINE  ONLINE       p19c02                   STABLE
      3         OFFLINE OFFLINE                               STABLE
ora.LISTENER_SCAN1.lsnr
      1         ONLINE  ONLINE       p19c02                   STABLE
ora.LISTENER_SCAN2.lsnr
      1         ONLINE  ONLINE       p19c01                   STABLE
ora.LISTENER_SCAN3.lsnr       1         ONLINE  ONLINE       p19c01
STABLE

ora.OCR.dg(ora.asmgroup)       1         ONLINE  ONLINE       p19c01
STABLE      2         OFFLINE OFFLINE
```

```
STABLE      3        OFFLINE OFFLINE
STABLE

ora.asm(ora.asmgroup)      1        ONLINE  ONLINE       p19c01
Started,STABLE      2        ONLINE  OFFLINE
Instance Shutdown,ST
ABLE      3        OFFLINE OFFLINE
STABLE

ora.asmnet1.asmnetwork(ora.asmgroup)      1        ONLINE  ONLINE
p19c01                   STABLE      2        ONLINE  ONLINE
p19c02                   STABLE      3        OFFLINE OFFLINE
STABLE

ora.cvu      1        ONLINE  ONLINE       p19c01
STABLE

ora.p19c01.vip      1        ONLINE  ONLINE       p19c01
STABLE

ora.p19c02.vip      1        ONLINE  ONLINE       p19c02
STABLE

ora.qosmserver      1        ONLINE  ONLINE       p19c01
STABLE

ora.scan1.vip      1        ONLINE  ONLINE       p19c02
STABLE

ora.scan2.vip      1        ONLINE  ONLINE       p19c01
STABLE

ora.scan3.vip      1        ONLINE  ONLINE       p19c01
STABLE
```

```
--------------------------------------------------------------------------
-----------
```

这是因为 2 个云主机节点的 169.254 网段不通，即 Oracle RAC 的 HAIP 不通。这里采用将 HAIP 禁用掉的方案来解决，分别在两个节点执行如下操作：

```
/u01/app/19.3.0/grid/bin/crsctl stop crs

/u01/app/19.3.0/grid/bin/crsctl start crs -excl -nocrs

/u01/app/19.3.0/grid/bin/crsctl modify res
ora.cluster_interconnect.haip -attr "ENABLED=0" -init

/u01/app/19.3.0/grid/bin/crsctl modify res ora.asm -attr
"START_DEPENDENCIES='hard(ora.cssd,ora.ctssd)pullup(ora.cssd,ora.ctssd)weak(ora.drivers.acfs)',STOP_DEPENDENCIES='hard(intermediate:ora.cssd)'" -init

/u01/app/19.3.0/grid/bin/crsctl stop crs
```

最后再重新两个节点的 crs，然后查看集群状态就正常了：

```
[grid@p19c01:/home/grid]$ crsctl stat res
-t--------------------------------------------------------------------------
-------------
Name           Target  State        Server                   State
details
--------------------------------------------------------------------------
-----------Local Resources
--------------------------------------------------------------------------
-----------
```

```
ora.LISTENER.lsnr                 ONLINE  ONLINE       p19c01
STABLE                 ONLINE  ONLINE       p19c02
STABLE

ora.chad                   ONLINE  ONLINE       p19c01
STABLE                 ONLINE  ONLINE       p19c02
STABLE

ora.net1.network                  ONLINE  ONLINE       p19c01
STABLE                 ONLINE  ONLINE       p19c02
STABLE

ora.ons                    ONLINE  ONLINE       p19c01
STABLE                 ONLINE  ONLINE       p19c02
STABLE

--------------------------------------------------------------------------------
-----------Cluster Resources
--------------------------------------------------------------------------------
-----------

ora.ASMNET1LSNR_ASM.lsnr(ora.asmgroup)
      1        ONLINE  ONLINE       p19c01                   STABLE
      2        ONLINE  ONLINE       p19c02                   STABLE
      3        ONLINE  OFFLINE                               STABLE
ora.LISTENER_SCAN1.lsnr
      1        ONLINE  ONLINE       p19c02                   STABLE
ora.LISTENER_SCAN2.lsnr
      1        ONLINE  ONLINE       p19c01                   STABLE
```

```
ora.LISTENER_SCAN3.lsnr        1         ONLINE  ONLINE        p19c01
STABLE

ora.OCR.dg(ora.asmgroup)       1         ONLINE  ONLINE        p19c01
STABLE       2         ONLINE  ONLINE        p19c02
STABLE       3         OFFLINE OFFLINE
STABLE

ora.asm(ora.asmgroup)          1         ONLINE  ONLINE        p19c01
Started,STABLE       2         ONLINE  ONLINE        p19c02
Started,STABLE       3         OFFLINE OFFLINE
STABLE

ora.asmnet1.asmnetwork(ora.asmgroup)       1         ONLINE  ONLINE
p19c01                   STABLE       2         ONLINE  ONLINE
p19c02                   STABLE       3         OFFLINE OFFLINE
STABLE

ora.cvu        1         ONLINE  ONLINE        p19c01
STABLE

ora.p19c01.vip       1         ONLINE  ONLINE        p19c01
STABLE

ora.p19c02.vip       1         ONLINE  ONLINE        p19c02
STABLE

ora.qosmserver       1         ONLINE  ONLINE        p19c01
STABLE

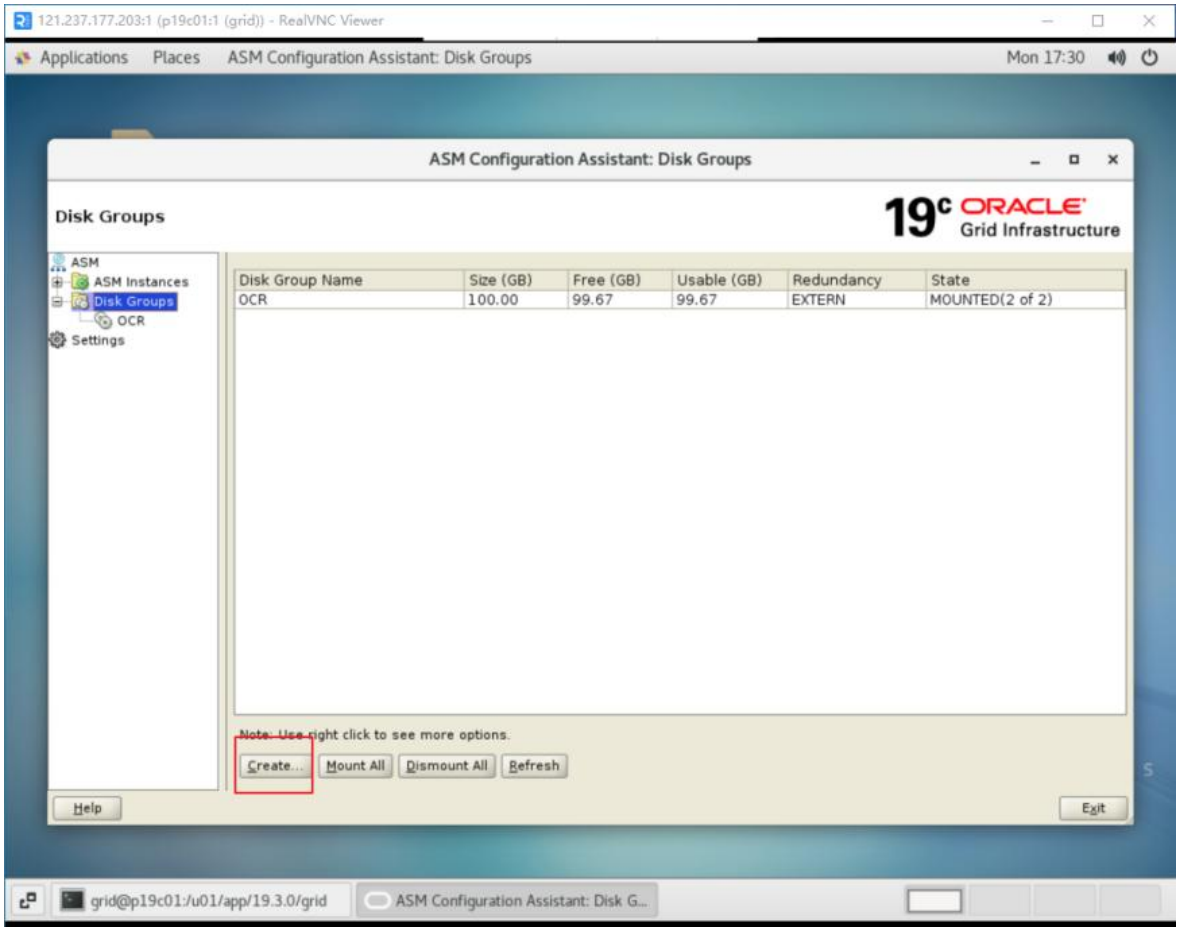
ora.scan1.vip        1         ONLINE  ONLINE        p19c02
STABLE

ora.scan2.vip        1         ONLINE  ONLINE        p19c01
STABLE
```

```
ora.scan3.vip       1         ONLINE  ONLINE       p19c01
STABLE
--------------------------------------------------------------------------------
```

添加 ASM DATA 数据盘：

执行 asmca 弹出磁盘组管理界面：



查看添加成功：

```
[grid@p19c01:/home/grid]$ asmcmd lsdg
State    Type    Rebal  Sector  Logical_Sector  Block       AU
Total_MB  Free_MB  Req_mir_free_MB  Usable_file_MB  Offline_disks
Voting_files  Name
```

```
MOUNTED  EXTERN  N         512             512    4096  4194304
204800   204668                0          204668              0
N  DATA/
MOUNTED  EXTERN  N         512             512    4096  4194304
102400   102064                0          102064              0
Y  OCR/
```

安装 Oracle 软件

配置 oracle 用户 vnc 图形界面：

```
[root@p19c01 ~]# chown -R oracle:oinstall /soft
[root@p19c01 ~]# su - oracle
Last login: Mon May 27 15:54:16 CST 2024 on pts/0
[oracle@p19c01:/home/oracle]$
[oracle@p19c01:/home/oracle]$ vncserver

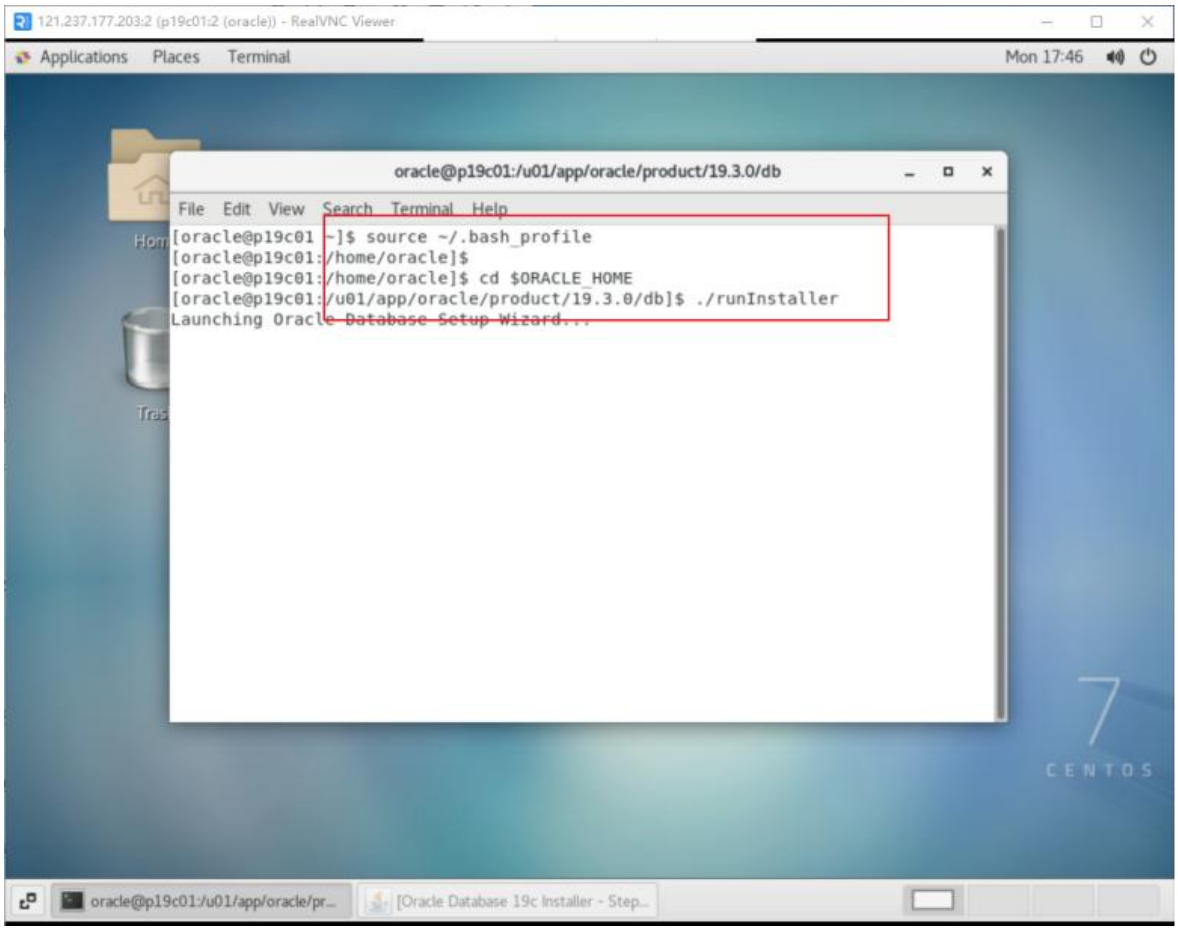
You will require a password to access your desktops.

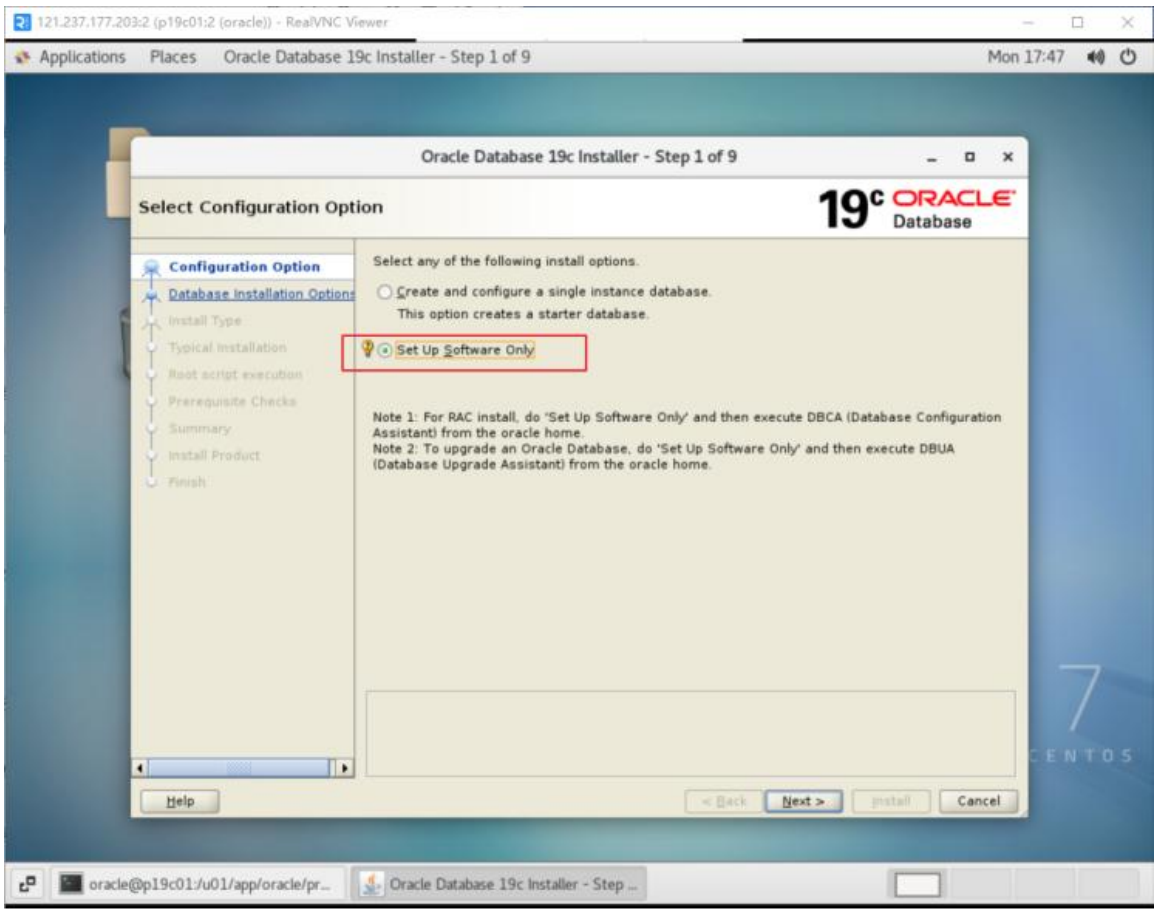
Password:
Verify:
Would you like to enter a view-only password (y/n)? n
A view-only password is not used
xauth:  file /home/oracle/.Xauthority does not exist
New 'p19c01:2 (oracle)' desktop is p19c01:2
```

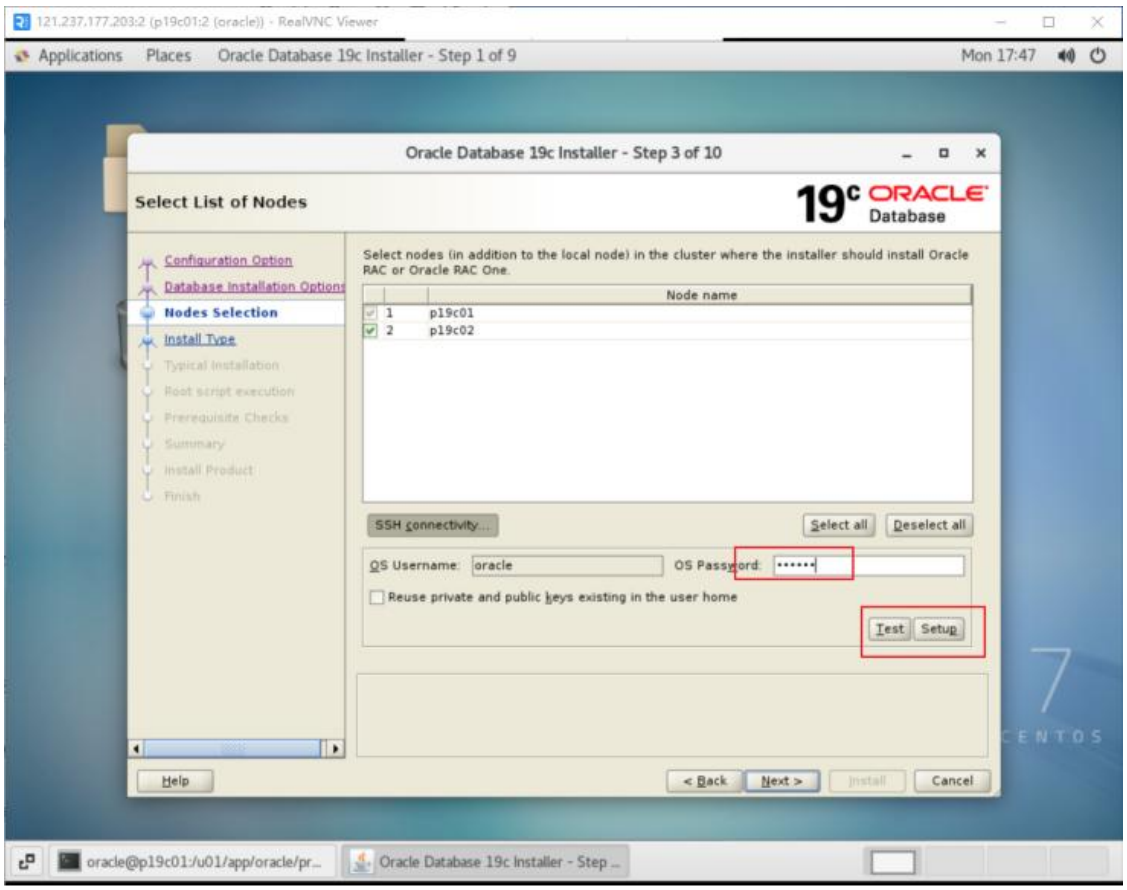
```
Creating default startup script /home/oracle/.vnc/xstartup
Creating default config /home/oracle/.vnc/config
Starting applications specified in /home/oracle/.vnc/xstartup
Log file is /home/oracle/.vnc/p19c01:2.log

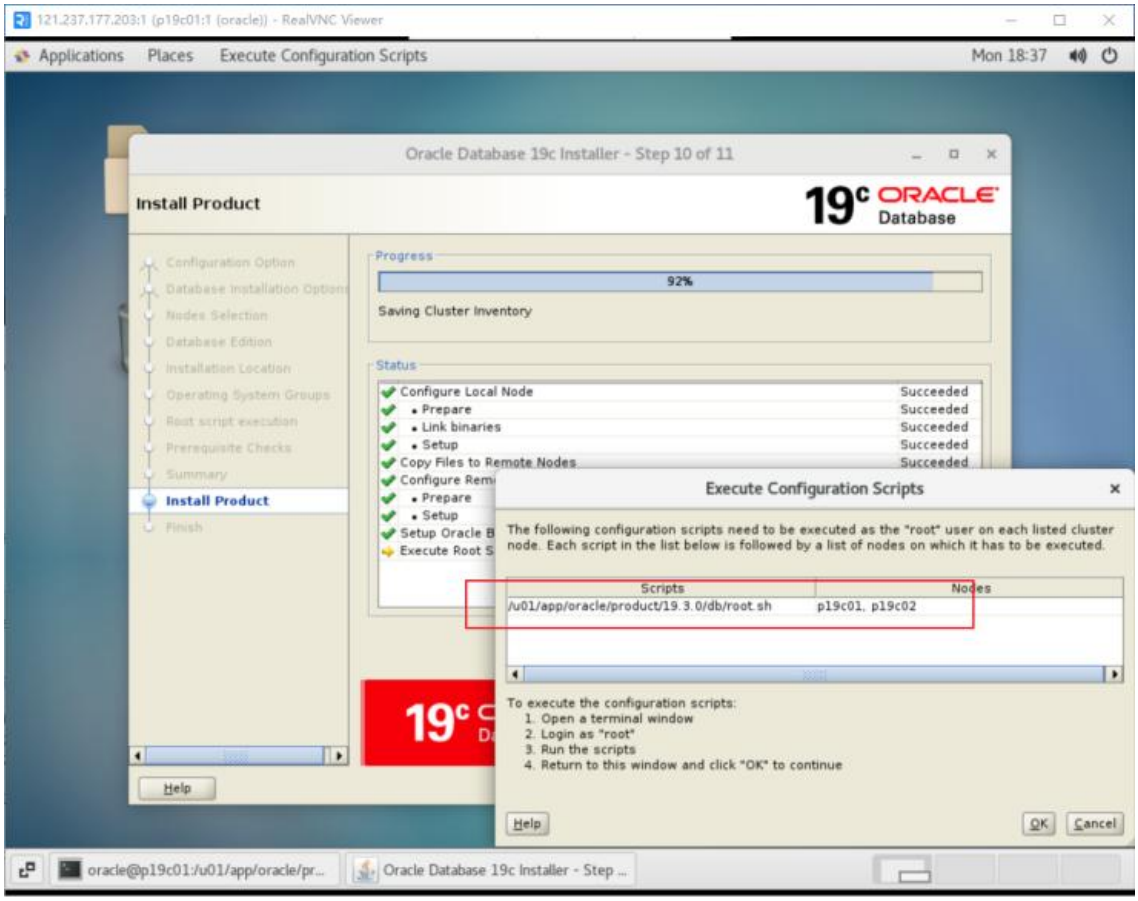
[oracle@p19c01:/home/oracle]$
[oracle@p19c01:/home/oracle]$ ps aux |grep vnc
oracle    1495  1.8  0.2 286620 46376 pts/0    Sl   17:33   0:00
/bin/Xvnc :2 -auth /home/oracle/.Xauthority -desktop p19c01:2 (oracle)
-fp catalogue:/etc/X11/fontpath.d -geometry 1024x768 -httpd
/usr/share/vnc/classes -pn -rfbauth /home/oracle/.vnc/passwd -rfbport
5902 -rfbwait 30000
oracle    1523  0.0  0.0 113288  1200 pts/0    S    17:33   0:00
/bin/sh /home/oracle/.vnc/xstartup
oracle    2262  0.0  0.0 112820   984 pts/0    S+   17:34   0:00 grep
--color=auto vnc
grid     20009  0.0  0.3 305292 54360 pts/0    Sl   15:56   0:04
/bin/Xvnc :1 -auth /home/grid/.Xauthority -desktop p19c01:1 (grid) -fp
catalogue:/etc/X11/fontpath.d -geometry 1024x768 -httpd
/usr/share/vnc/classes -pn -rfbauth /home/grid/.vnc/passwd -rfbport
5901 -rfbwait 30000
grid     20022  0.0  0.0 113288  1200 pts/0    S    15:56   0:00
/bin/sh /home/grid/.vnc/xstartup
[oracle@p19c01:/home/oracle]$
```

在 vnc 客户端输入弹性 IP 加上端口 2，接入图形界面。




121.237.177.203:2 (p19c01:2 (oracle)) - RealVNC Viewer
Applications Places Oracle Database 19c Installer - Step 1 of 9 Mon 17:47
Oracle Database 19c Installer - Step 1 of 9
Select Configuration Option
19c ORACLE Database
Configuration Option
Database Installation Options
Install Type
Typical Installation
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
Select any of the following install options.
Create and configure a single instance database.
This option creates a starter database.
Set Up Software Only
Note 1: For RAC install, do 'Set Up Software Only' and then execute DBCA (Database Configuration Assistant) from the oracle home.
Note 2: To upgrade an Oracle Database, do 'Set Up Software Only' and then execute DBUA (Database Upgrade Assistant) from the oracle home.
Help
< Back
Next >
Install
Cancel
oracle@p19c01:/u01/app/oracle/pr...
Oracle Database 19c Installer - Step ...

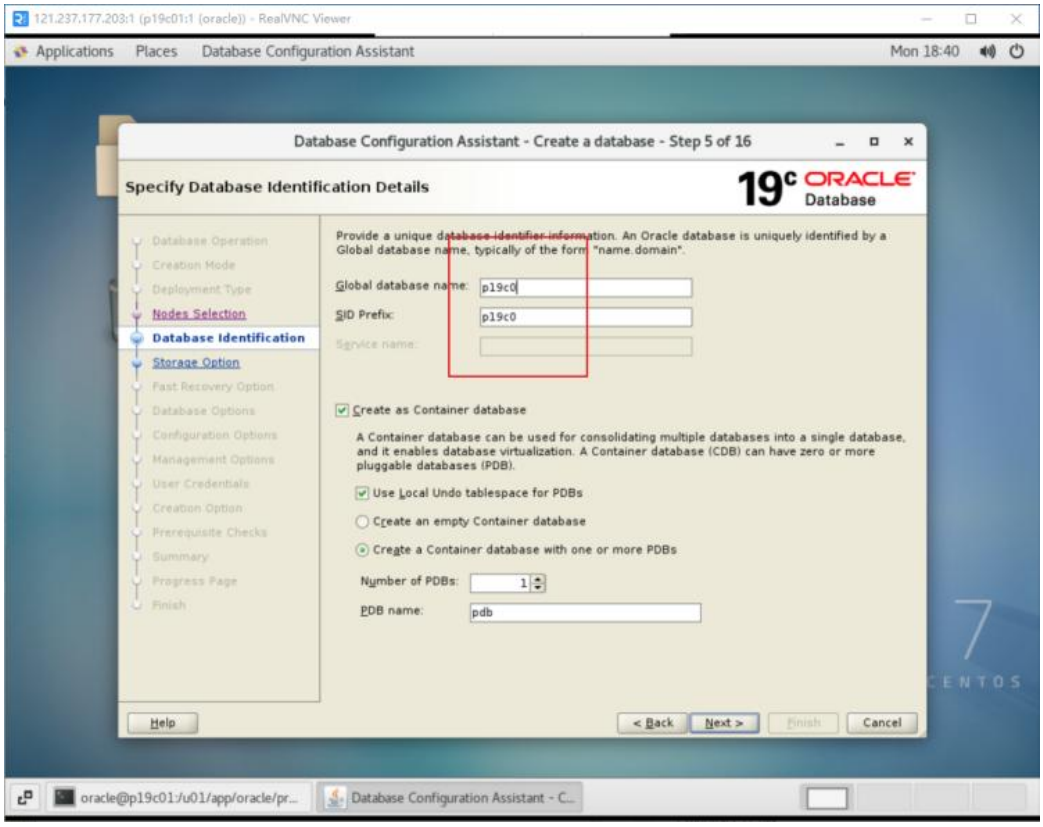

121.237.177.203:2 (p19c01:2 (oracle)) - RealVNC Viewer
Applications Places Oracle Database 19c Installer - Step 1 of 9 Mon 17:47
Oracle Database 19c Installer - Step 3 of 10
Select List of Nodes
19c ORACLE Database
Configuration Option
Database Installation Options
Nodes Selection
Install Type
Typical Installation
Root script execution
Prerequisite Checks
Summary
Install Product
Finish
Select nodes (in addition to the local node) in the cluster where the installer should install Oracle RAC or Oracle RAC One.
Node name
1 p19c01
2 p19c02
SSH connectivity...
Select all
Deselect all
OS Username: oracle
OS Password:
Reuse private and public keys existing in the user home
Test
Setup
Help
< Back
Next >
Install
Cancel
oracle@p19c01:/u01/app/oracle/pr...
Oracle Database 19c Installer - Step ...



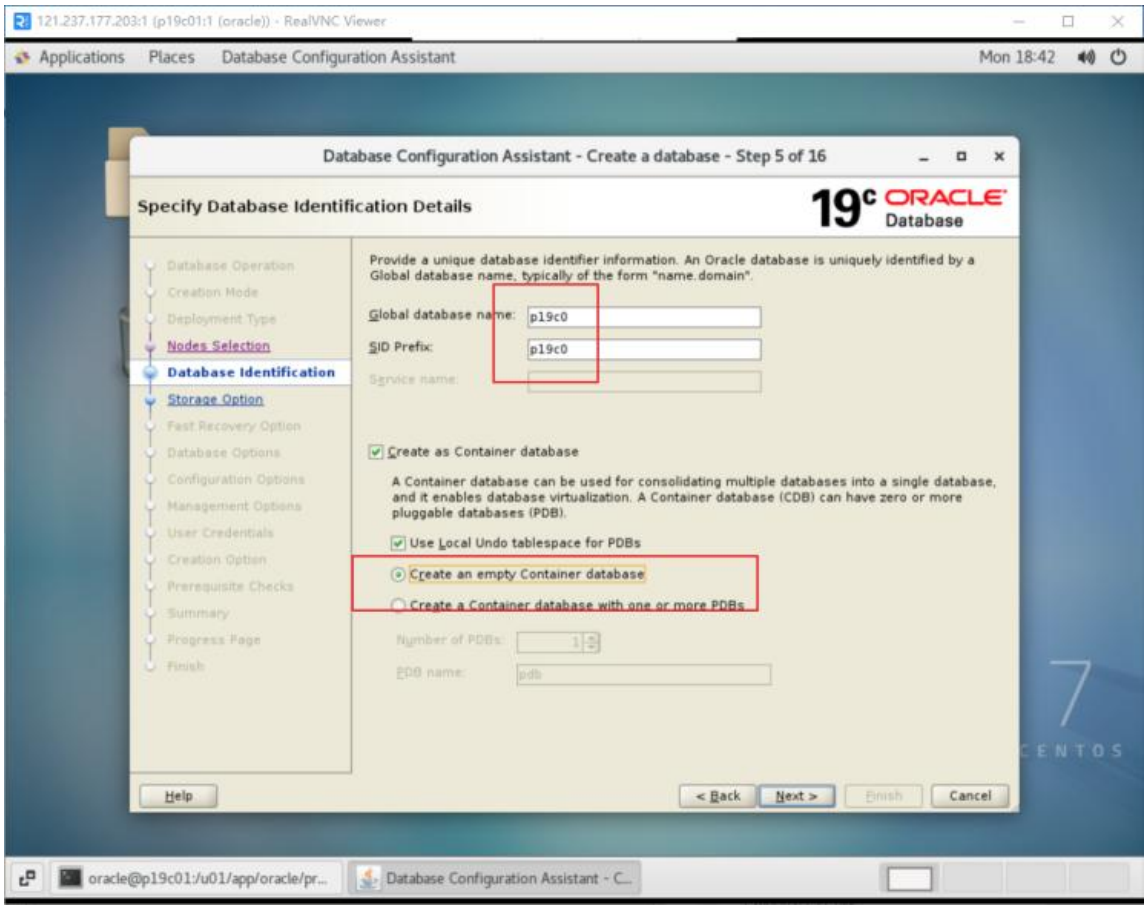
分别在两个节点上执行配置脚本，完成数据库的安装。

创建数据库实例

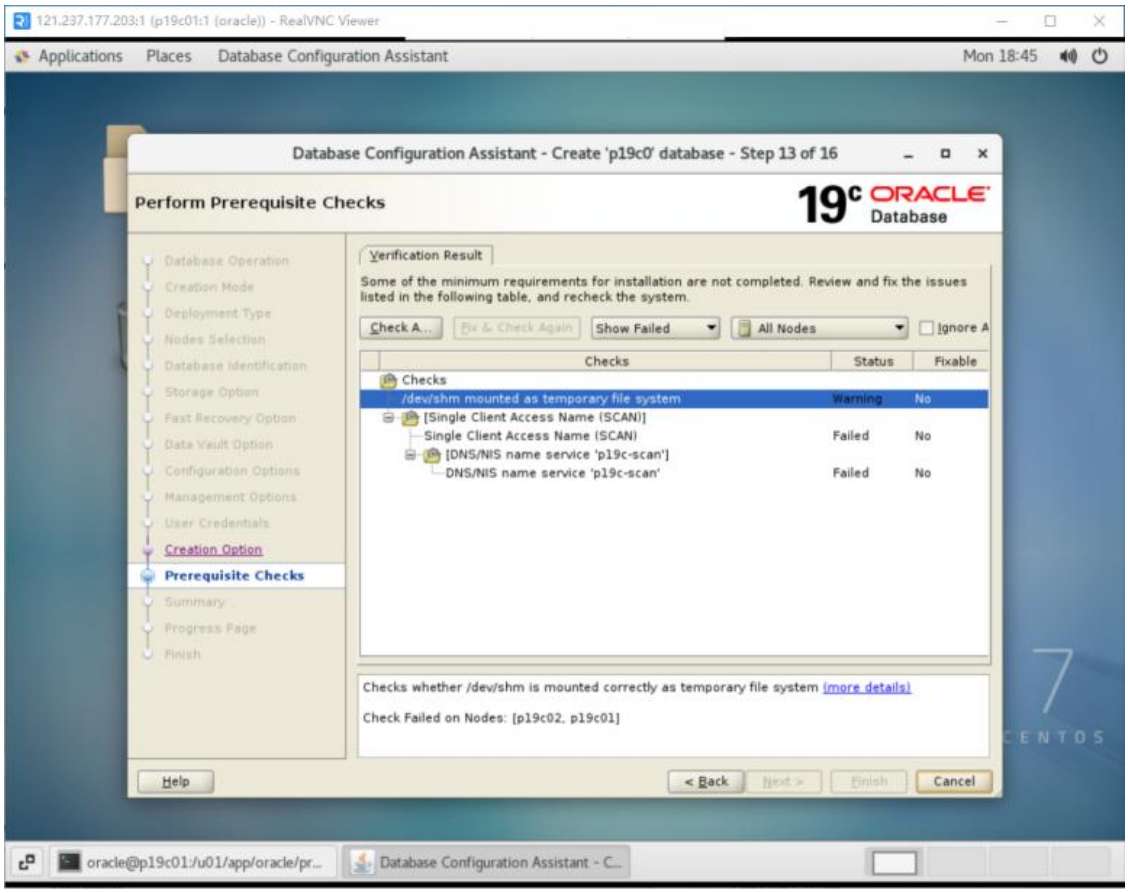
执行 dbca 命令弹出创建数据库界面：


121.237.177.203:1 (p19c01:1 (oracle)) - RealVNC Viewer
Applications Places Database Configuration Assistant
Mon 18:40
Database Configuration Assistant - Create a database - Step 5 of 16
Specify Database Identification Details
19c ORACLE Database
Database Operation
Creation Mode
Deployment Type
Nodes Selection
Database Identification
Storage Option
Fast Recovery Option
Database Options
Configuration Options
Management Options
User Credentials
Creation Option
Prerequisite Checks
Summary
Progress Page
Finish
Provide a unique database identifier information. An Oracle database is uniquely identified by a Global database name, typically of the form "name.domain".
Global database name: p19c0
SID Prefix: p19c0
Service name:
Create as Container database
A Container database can be used for consolidating multiple databases into a single database, and it enables database virtualization. A Container database (CDB) can have zero or more pluggable databases (PDB).
Use Local Undo tablespace for PDBs
Create an empty Container database
Create a Container database with one or more PDBs
Number of PDBs: 1
PDB name: pdb
Help
< Back
Next >
Finish
Cancel
oracle@p19c01:/u01/app/oracle/pr...
Database Configuration Assistant - C...

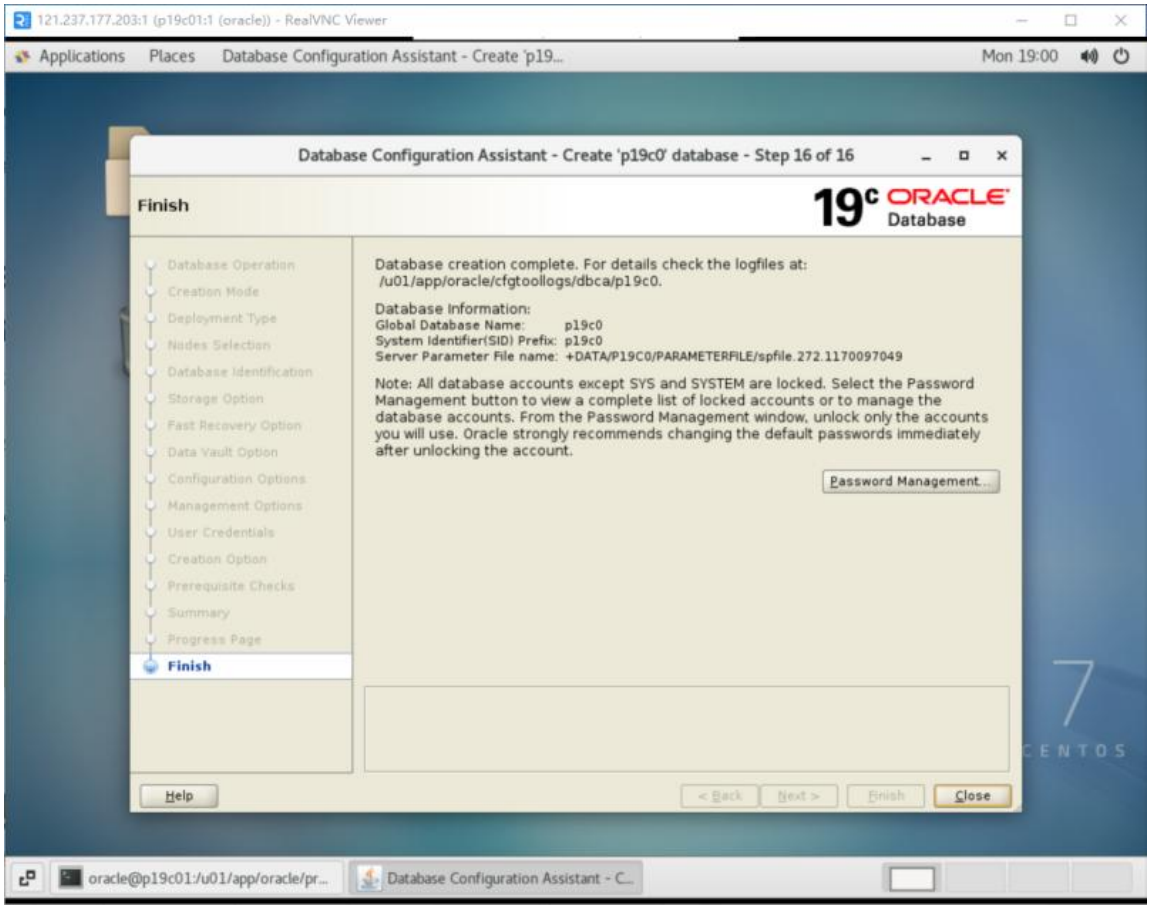
填写实例名称：


121.237.177.203:1 (p19c01:1 (oracle)) - RealVNC Viewer
Applications Places Database Configuration Assistant
Mon 18:42
Database Configuration Assistant - Create a database - Step 5 of 16
Specify Database Identification Details
19c ORACLE Database
Database Operation
Creation Mode
Deployment Type
Nodes Selection
Database Identification
Storage Option
Fast Recovery Option
Database Options
Configuration Options
Management Options
User Credentials
Creation Option
Prerequisite Checks
Summary
Progress Page
Finish
Provide a unique database identifier information. An Oracle database is uniquely identified by a Global database name, typically of the form "name.domain".
Global database name: p19c0
SID Prefix: p19c0
Service name:
Create as Container database
A Container database can be used for consolidating multiple databases into a single database, and it enables database virtualization. A Container database (CDB) can have zero or more pluggable databases (PDB).
Use Local Undo tablespace for PDBs
Create an empty Container database
Create a Container database with one or more PDBs
Number of PDBs: 1
PDB name: pdb
Help
< Back
Next >
Finish
Cancel
oracle@p19c01:/u01/app/oracle/pr...
Database Configuration Assistant - C...

安装预检查，忽略：



完成数据库的创建：



10.26 手动搭建 LNMP 环境

操作流程

- 安装 nginx。
- 安装 MySQL。
- 安装 PHP。
- 浏览器访问测试。

前提条件

- 弹性云主机已绑定弹性 IP。

- 弹性云主机所在安全组添加了如下表所示的安全组规则，具体步骤参见为安全组添加安全组规则。

方向	协议/应用	端口/范围	源地址
入方向	HTTP(80)	80	0.0.0.0/0

操作步骤

安装 nginx。

1.登录弹性云主机。

2.执行以下命令，下载对应当前系统版本的 nginx 包。

```
wget http://nginx.org/packages/centos/7/noarch/RPMS/nginx-release-centos-7-0.el7.ngx.noarch.rpm
```

3.执行以下命令，建立 nginx 的 yum 仓库。

```
rpm -ivh nginx-release-centos-7-0.el7.ngx.noarch.rpm
```

4.执行以下命令，安装 nginx。

```
yum -y install nginx
```

5.执行以下命令，启动 nginx 并设置开机启动。

```
systemctl start nginxsystemctl enable nginx
```

6.使用浏览器访问 “http://服务器 IP 地址”，显示如下页面，说明 nginx 安装成功。



安装 MySQL

1.依次执行以下命令，安装 MySQL。

```
rpm -Uvh
http://dev.mysql.com/get/mysql57-community-release-el7-8.noarch.rpm

yum -y install mysql-community-server
```

2.依次执行以下命令，启动 MySQL 服务并设置开机自启动。

```
systemctl start mysqldsystemctl enable mysqld
```

3.执行以下命令，获取安装 MySQL 时自动设置的 root 用户密码。

```
grep 'temporary password' /var/log/mysqld.log
```

回显如下类似信息。

```
2018-08-29T07:27:37.541944Z 1 [Note] A temporary password is generated
for root@localhost: 2YY?3uHUA?Ys
```

4.执行以下命令，并按照回显提示信息进行操作，加固 MySQL。

```
mysql_secure_installation

Securing the MySQL server deployment.

Enter password for user root:     #输入上一步骤中获取的安装 MySQL 时自动
设置的 root 用户密码
```

```
The existing password for the user account root has expired. Please set a new password.

New password:  #设置新的 root 用户密码

Re-enter new password:   #再次输入密码

The 'validate_password' plugin is installed on the server.

The subsequent steps will run with the existing configuration of the plugin.

Using existing password for root.

Estimated strength of the password: 100

Change the password for root ? ((Press y|Y for Yes, any other key for No) : N   #是否更改 root 用户密码，输入 N

... skipping.

By default, a MySQL installation has an anonymous user,

allowing anyone to log into MySQL without having to havea user account created for them. This is intended only for

testing, and to make the installation go a bit smoother.

You should remove them before moving into a production

environment.

Remove anonymous users? (Press y|Y for Yes, any other key for No) : Y   #是否删除匿名用户，输入 Y

Success.

Normally, root should only be allowed to connect from 'localhost'. This ensures that someone cannot guess at the root password from the network.
```

```
Disallow root login remotely? (Press y|Y for Yes, any other key for No) :
Y    #禁止 root 远程登录，输入 Y

Success.

By default, MySQL comes with a database named 'test' that anyone can access.
This is also intended only for testing, and should be removed before moving
into a production environment.

Remove test database and access to it? (Press y|Y for Yes, any other key
for No) : Y    #是否删除 test 库和对它的访问权限，输入 Y

-Dropping test database...

Success.

-Removing privileges on test database...

Success.

Reloading the privilege tables will ensure that all changes

made so far will take effect immediately.

Reload privilege tables now? (Press y|Y for Yes, any other key for No) :
Y    #是否重新加载授权表，输入 Y

Success.

All done!
```

安装 PHP

1.依次执行以下命令，安装 PHP 7 和一些所需的 PHP 扩展。

```
rpm -Uvh https://mirror.webtatic.com/yum/el7/epel-release.rpmrpm -Uvh
https://mirror.webtatic.com/

yum/el7/webtatic-release.rpm
```

```
yum -y install php70w-tidy php70w-common php70w-devel php70w-pdo
php70w-mysql php70w-gd php70w-ldap php70w-mbstring php70w-mcrypt
php70w-fpm
```

2.执行以下命令，验证 PHP 的安装版本。

```
php -v
```

回显如下类似信息：

```
PHP 7.0.31 (cli) (built: Jul 20 2018 08:55:22) ( NTS )Copyright (c)
1997-2017 The PHP GroupZend Engine v3.0.0, Copyright (c) 1998-2017 Zend
Technologies
```

3.执行以下命令，启动 PHP 服务并设置开机自启动。

```
systemctl start php-fpmsystemctl enable php-fpm
```

4.修改 nginx 配置文件以支持 PHP。

执行以下命令打开配置文件“default.conf”。

```
vim /etc/nginx/conf.d/default.conf
```

按 i 键进入编辑模式。

修改打开的“default.conf”文件。

在所支持的主页面格式中添加 php 格式的主页，如下所示：

```
location / {
root   /usr/share/nginx/html;index index.php          index.html
index.htm;
}
```

取消如下内容的注释，并设置字体加粗部分为 nginx 的默认路径，如下所示：

```
location ~ \.php$ {
```

```
    root           html;

    fastcgi_pass   127.0.0.1:9000;

    fastcgi_index  index.php;

    fastcgi_param  SCRIPT_FILENAME
/usr/share/nginx/html$fastcgi_script_name;

    include        fastcgi_params;

        }
```

按 Esc 键退出编辑模式，并输入 :wq 保存后退出。

执行以下命令，重新载入 nginx 的配置文件。

```
service nginx reload
```

浏览器访问测试

1./usr/share/nginx/html/目录下创建“info.php”的测试页面。

1）执行以下命令创建并打开“info.php”的测试文件。

```
vim /usr/share/nginx/html/info.php
```

2）按 i 键进入编辑模式。

3）修改打开的“info.php”文件，将如下内容写入文件。

4）按 Esc 键退出编辑模式，并输入:wq 保存后退出。

2.使用浏览器访问 http://服务器 IP 地址/info.php，显示如下页面，说明搭建环境成功。



10.27 搭建 FTP

Windows 云主机

添加 IIS 以及 FTP 服务角色

1.登录弹性云主机。

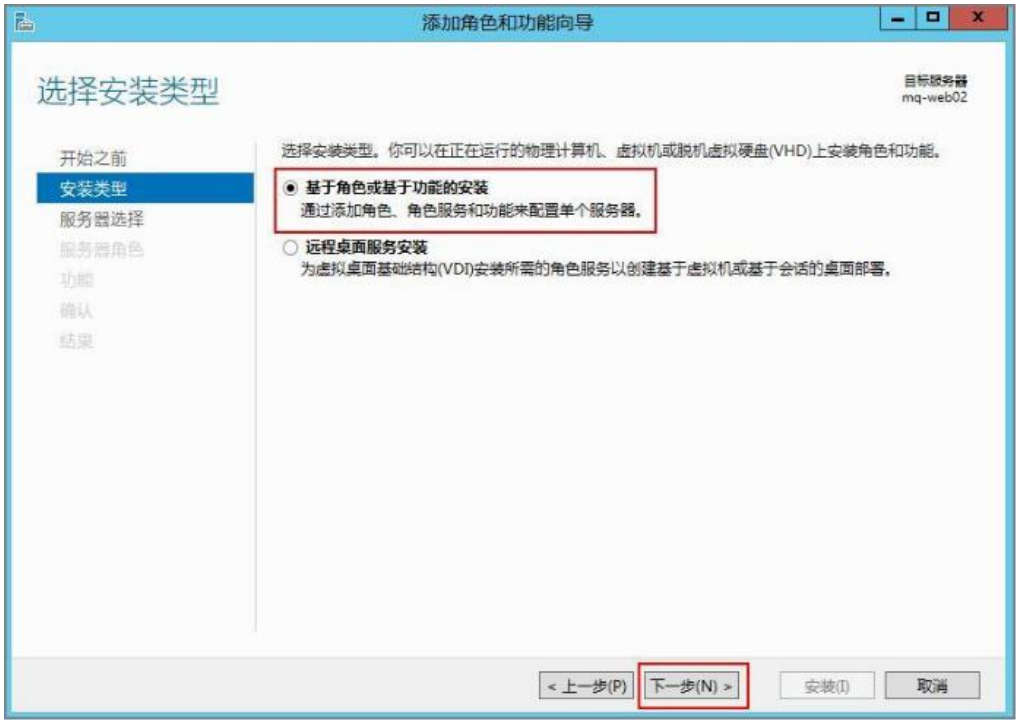
2.选择“开始 > 服务器管理器”。

3.单击“添加角色和功能”。

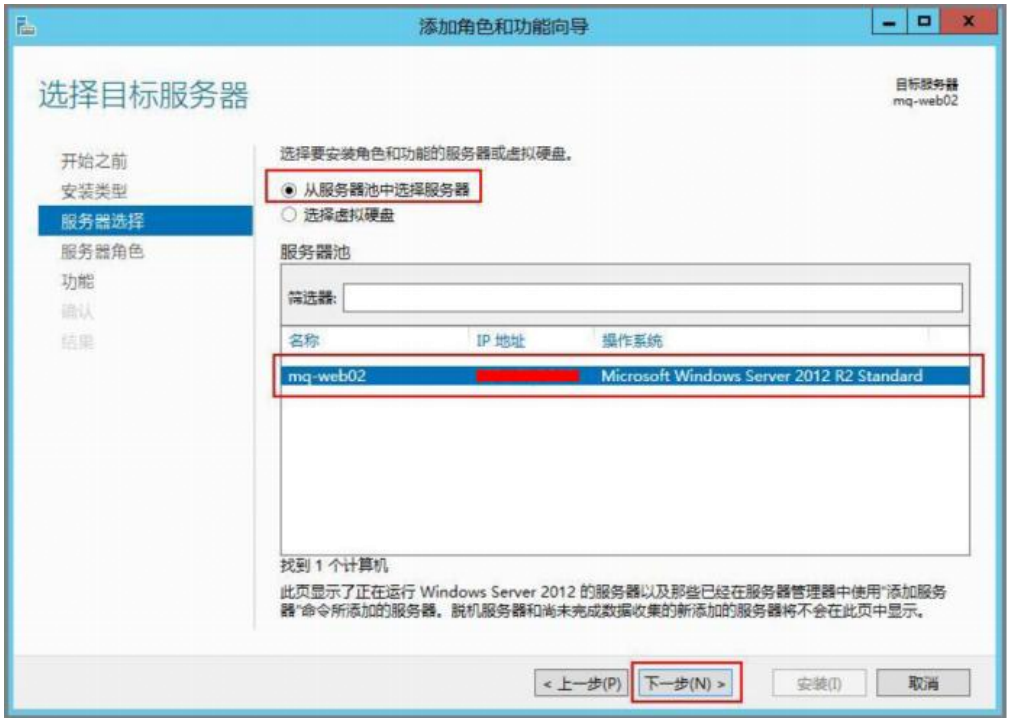


4.在弹出的“开始之前”对话框中，单击“下一步”。

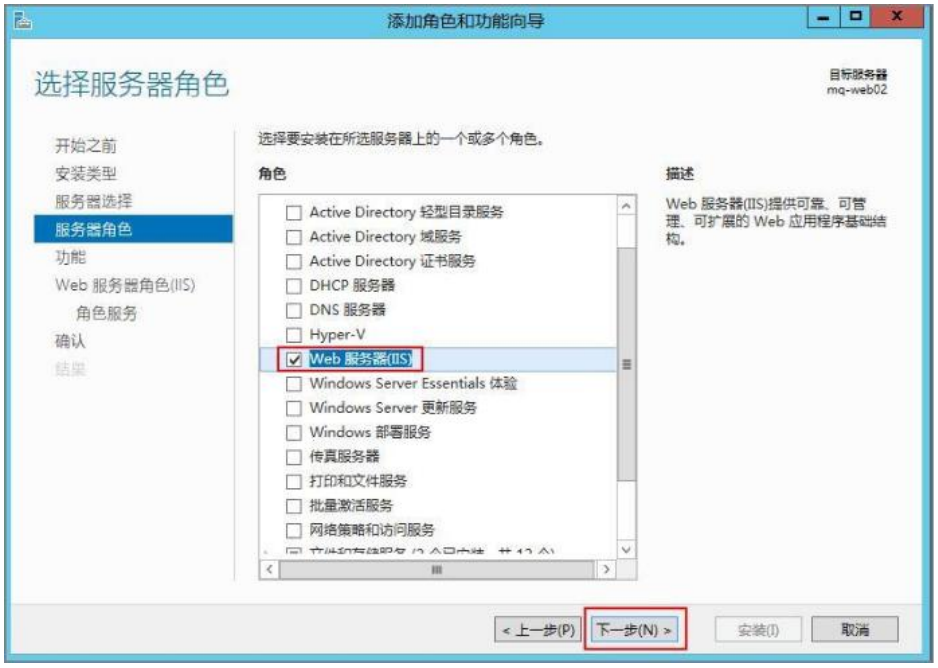
5.选择“基于角色或基于功能的安装”，单击“下一步”。



6.选择需要部署 FTP 的服务器，单击“下一步”。



7.选择“ Web 服务器（IIS）”，并在弹出的对话框中单击“添加功能”，然后单击“下一步”。





8. 连续单击“下一步”，到“角色服务”页面。

9. 选择“FTP 服务器” 以及“IIS 管理控制台”，单击“下一步”。

10. 单击“安装”，开始部署服务角色。

11. 安装完成后，单击“关闭”。

创建 FTP 用户名及密码

创建 Windows 用户名和密码，用于 FTP 使用。如果您希望匿名用户可以访问，可以不创建 FTP 用户和密码。

1. 在“服务器管理器”中，选择“仪表板 > 工具 > 计算机管理”。



2. 选择“系统工具 > 本地用户和组 > 用户”，在右侧空白处右击，并选择“新用户”。

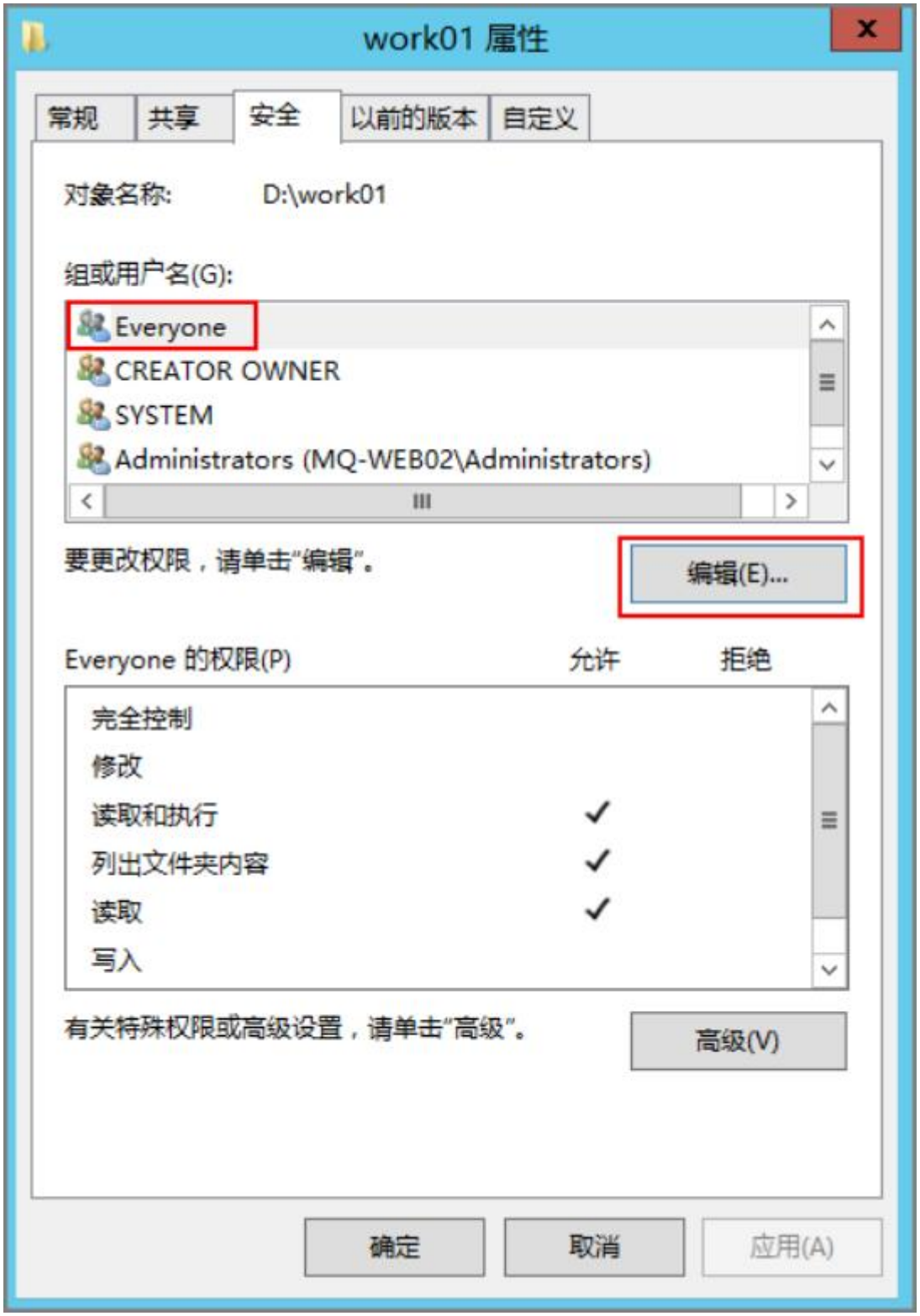


3.设置“用户名”和“密码”，此处用户名以“ftpadmin”为例。

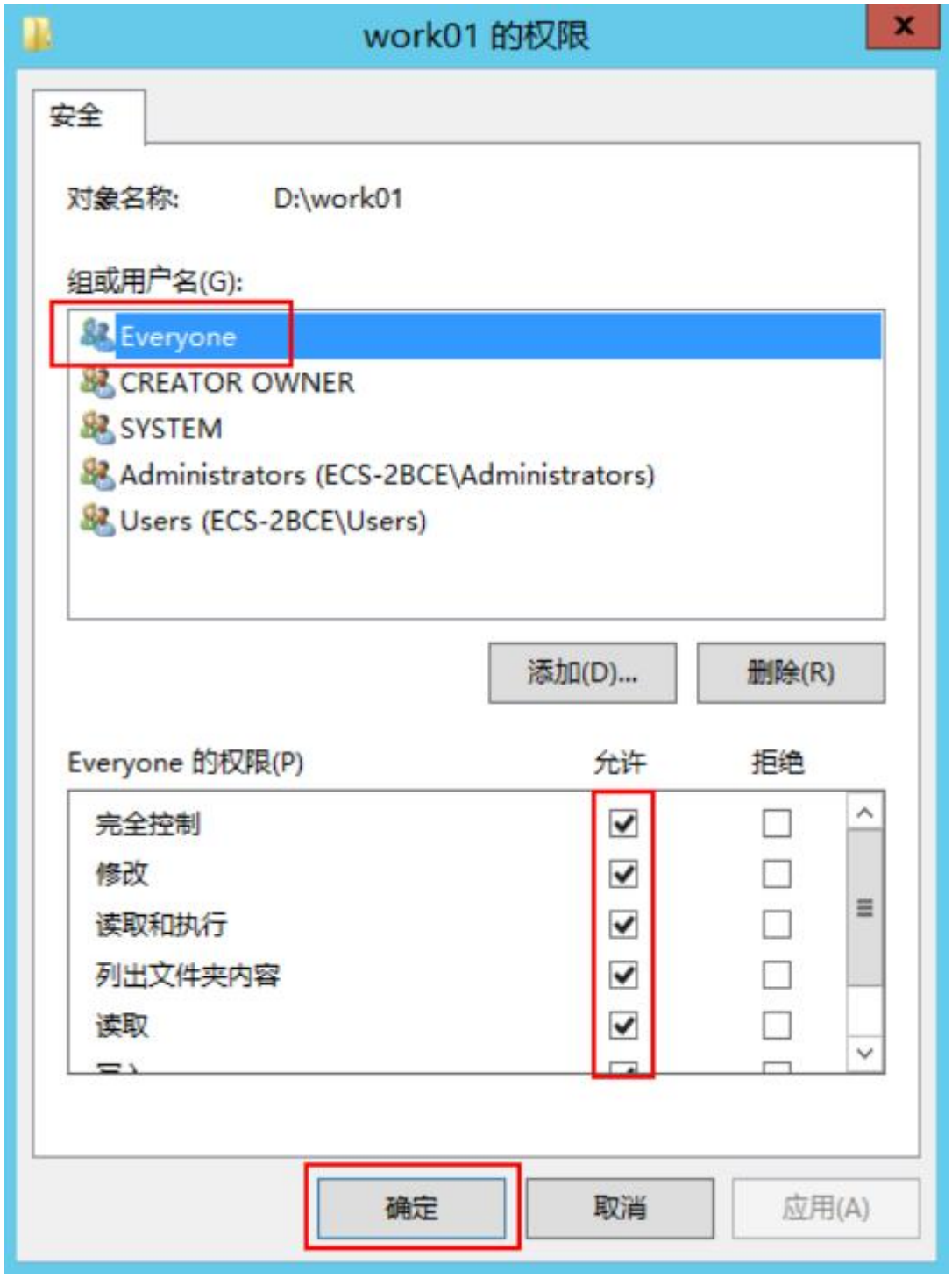
设置共享文件的权限

需要在 FTP 站点为共享给用户的文件夹设置访问及修改等权限。

1.在服务器上创建一个供 FTP 使用的文件夹，选择文件夹，并单击右键选择“属性”。此处以“work01”文件夹为例。
2.在“安全”页签，选择 “Everyone”，单击“编辑”。如果没有“Everyone”用户可以直接选择，需要先进行添加。



3. 选择“Everyone”，然后根据需要，选择“Everyone”的权限，并单击“确定”。此处以允许所有权限为例。

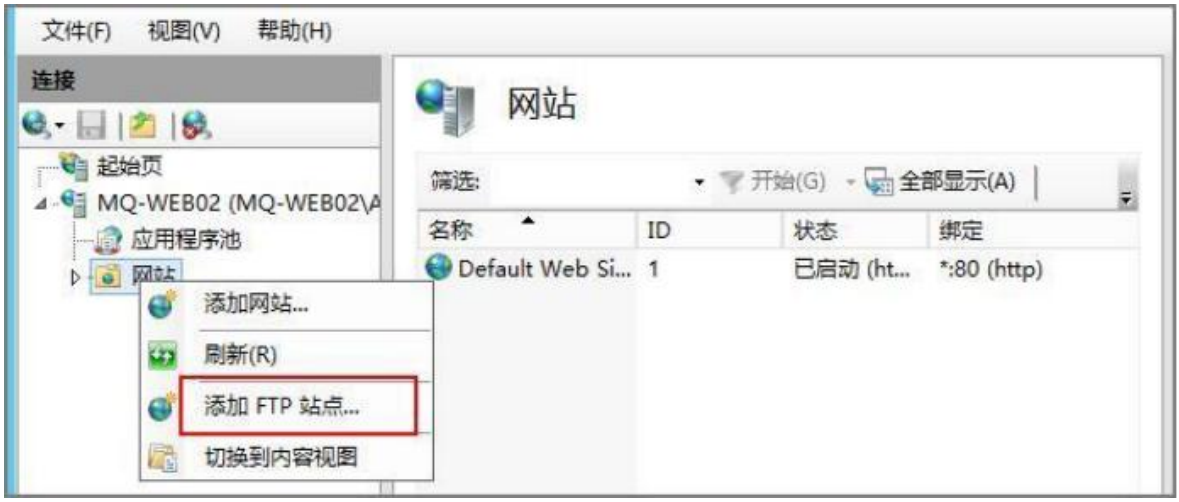


添加及设置 FTP 站点

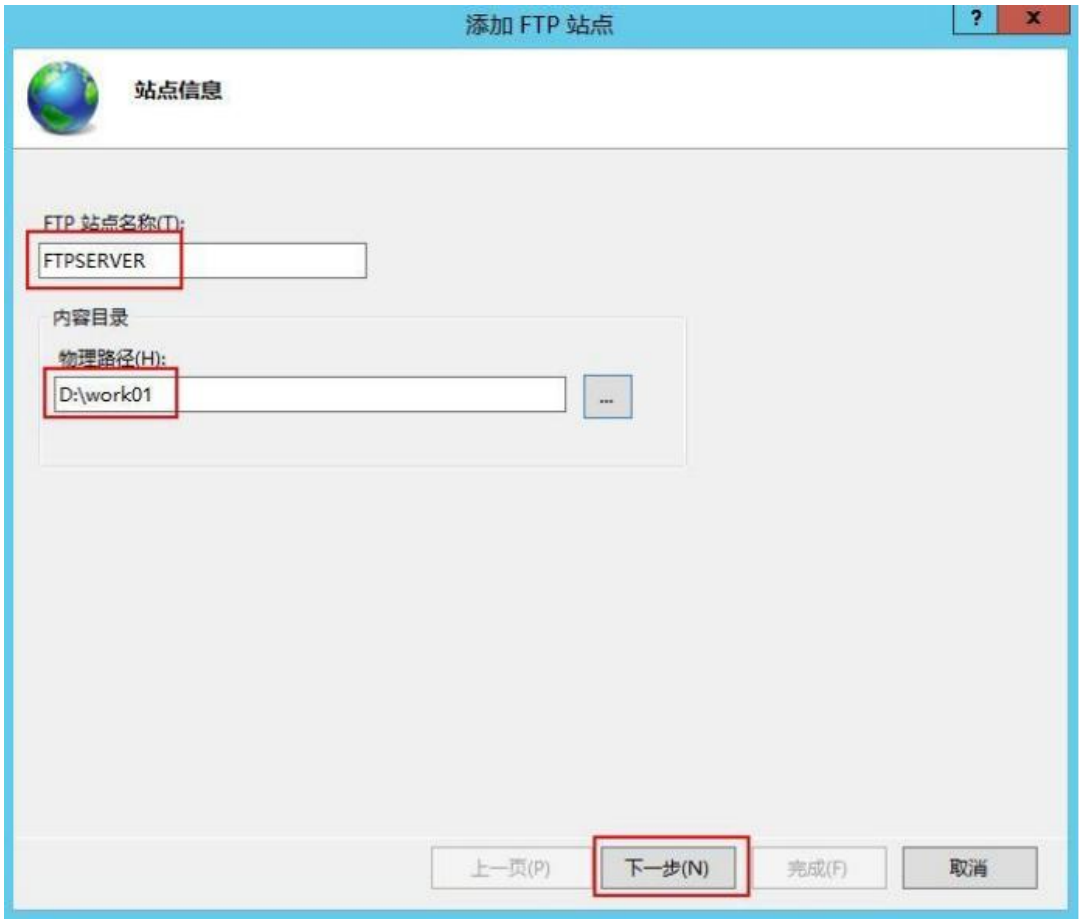
1. 在“服务器管理器”中，选择“仪表板 > 工具 > Internet Information Services (IIS)管理器 ”。



2. 选择“网站”并单击右键，然后选择“添加 FTP 站点”。



3. 在弹出的窗口中，填写 FTP 站点名称及共享文件夹的物理路径，然后单击“下一步”。此处站点名称以“FTPSERVER”为例。

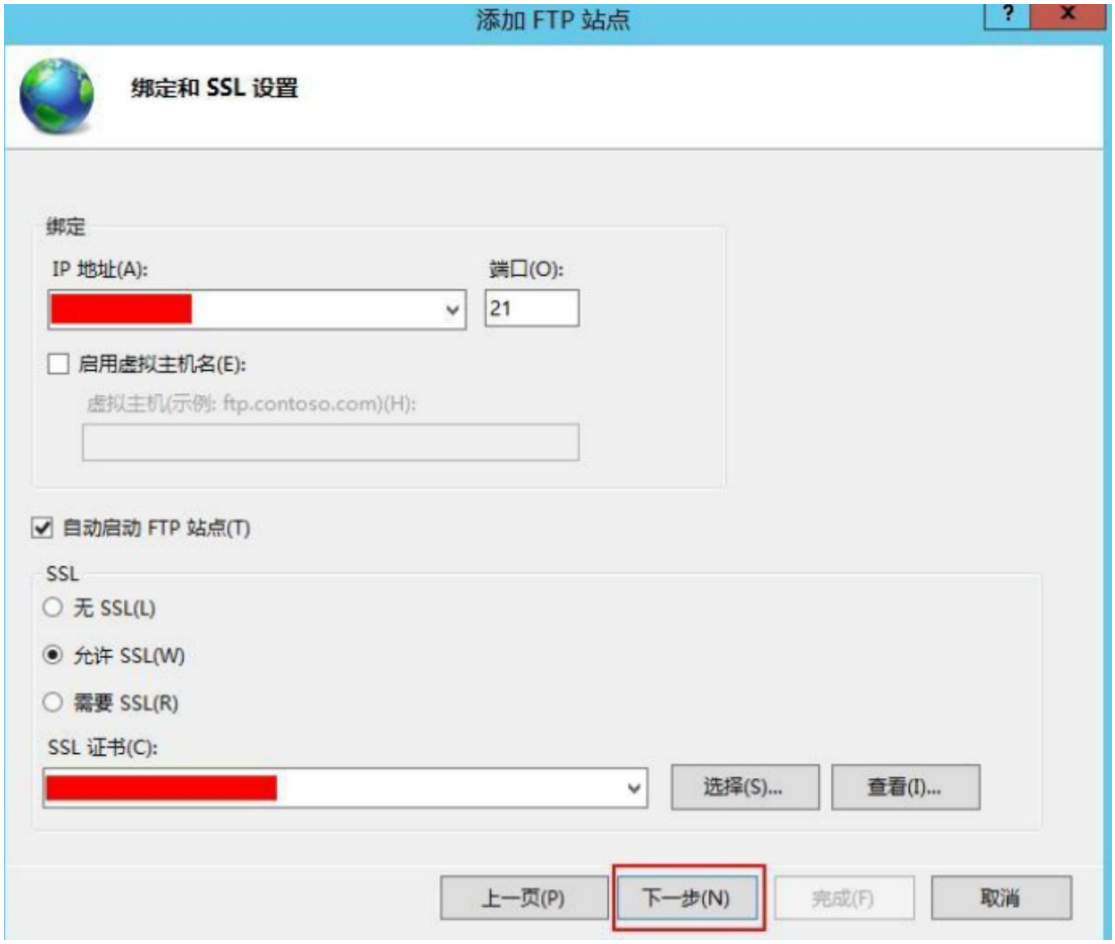


4. 输入该弹性云主机的公网 IP 地址以及端口号，并设置 SSL，单击“下一步”。端口号默认为 21，也可自行设置 。

5. SSL 根据需要进行设置。

- 无：不需要 SSL 加密。
- 允许：允许 FTP 服务器与客户端的非 SSL 和 SSL 连接。
- 需要：需要对 FTP 服务器和客户端之间的通信进行 SSL 加密。

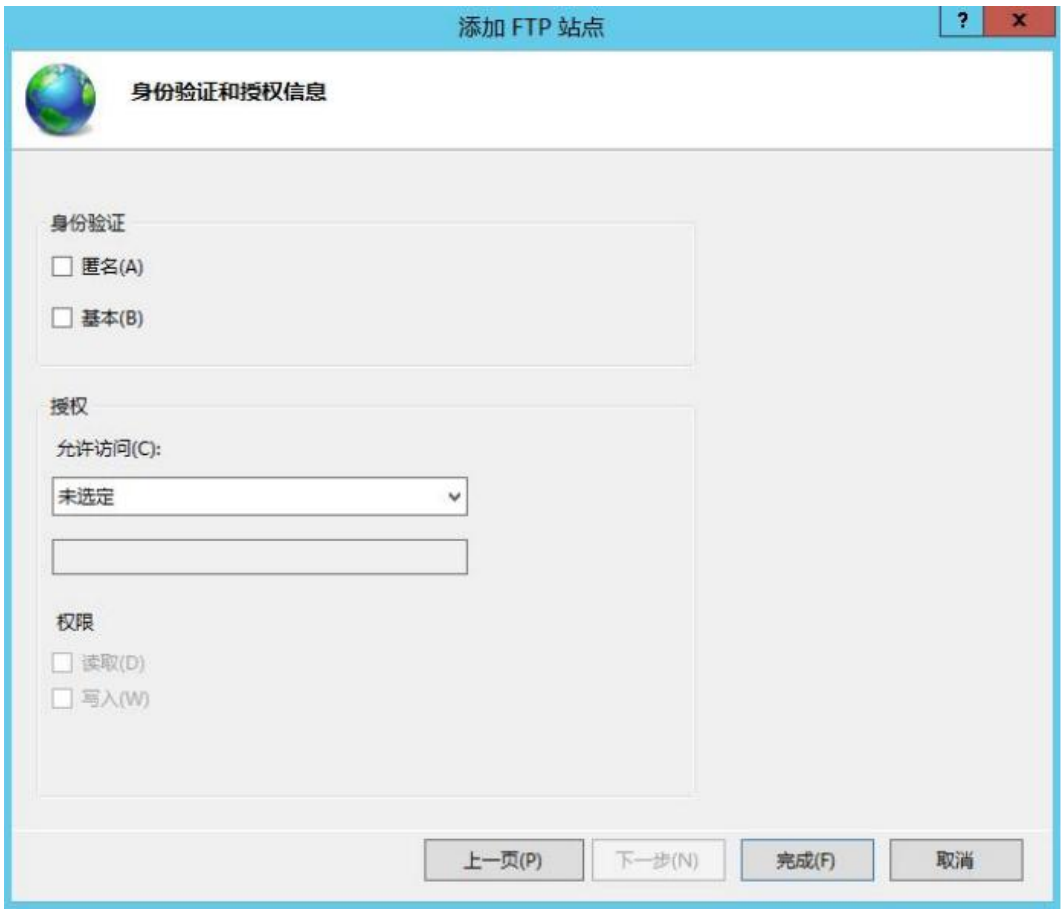
说明 ：当 SSL 选择“允许”和“需要”时，需要选择 SSL 证书。可以选择已有的 SSL 证书，也可以制作一个 SSL 证书，具体制作证书的方法请参见制作服务器证书。



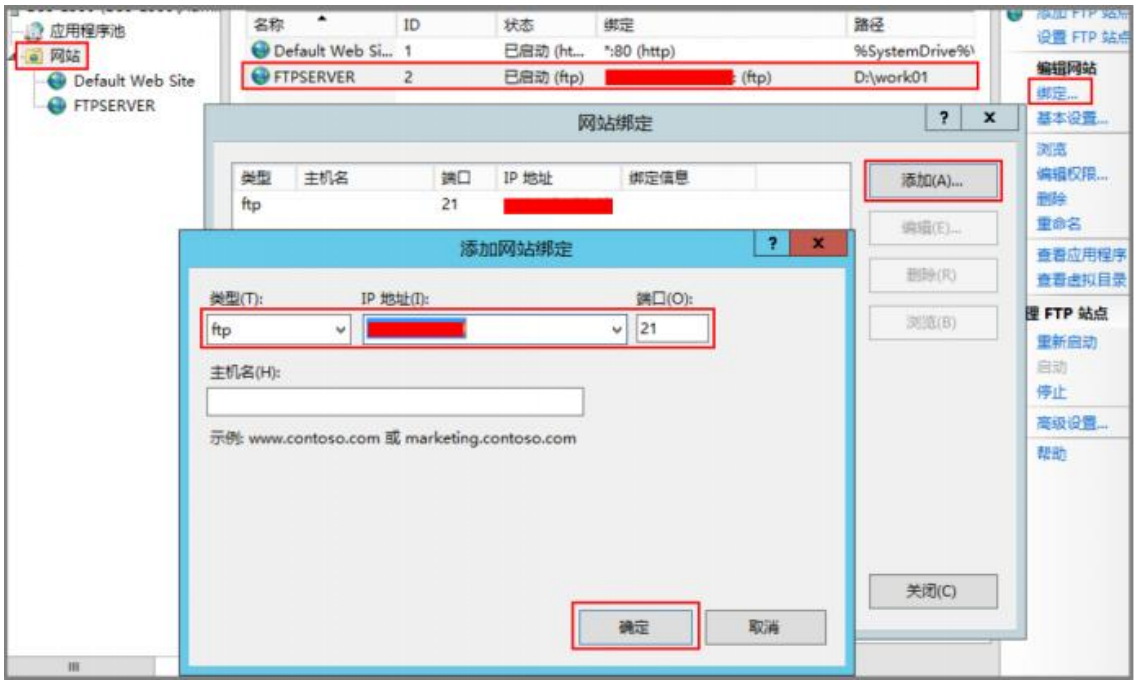
6. 设置身份认证和授权信息，并单击“完成”。

- 身份认证
  - 匿名 ： 允许任何仅提供用户名 “anonymous” 或 “ftp” 的用户访问内容。
  - 基本 ： 需要用户提供有效用户名和密码才能访问内容。但是基本身份验证通过网络传输密码时不加密，因此建议在确认客户端和 FTP 服务器之间的网络连接安全时使用此身份验证方法。
- 授权
  - 允许访问
  - 权限：选择经过授权的用户的“读取”和“写入”权限。
- 所有用户 ： 所有用户均可访问相应内容。
- 匿名用户 ： 匿名用户可访问相应内容。
- 指定角色或用户组：仅指定的角色或用户组的成员才能访问相应内容。如果选择此项，需要在下面输入框中输入指定的角色或用户组。

- 指定用户：仅指定的用户才能访问相应内容。如果选择此项，需要在下面输入框中输入指定的用户。



7. 绑定弹性云主机的私网 IP。选择“网站”，选中创建的 FTP 站点，单击“绑定”；在弹出的“网站绑定”窗口单击“添加”，然后在弹出的窗口中添加弹性云主机的私网 IP 地址，并单击“确定”。



可选）配置 FTP 防火墙支持

如果需要使用 FTP 服务器的被动模式，则需要配置 FTP 防火墙支持。

如果天翼云上的服务器需要通过公网 IP 地址访问天翼云上的实例搭建的 FTP 服务器时，需要配置 FTP 服务器的被动模式。

1. 双击“FTP 防火墙支持”，打开 FTP 防火墙支持的配置界面。



2.配置相关参数，并单击“应用”。

数据通道端口范围：指定用于被动连接的端口范围。可指定的有效端口范围为1025-65535。请根据实际需求进行设置。

防火墙的外部 IP 地址：输入该弹性云主机的公网 IP 地址。



3.重启云主机使防火墙配置生效。

设置安全组及防火墙

搭建好 FTP 站点后，需要在弹性云主机安全组的入方向添加一条放行 FTP 端口的规则，具体步骤参见虚拟私有云-安全组-添加安全组规则。

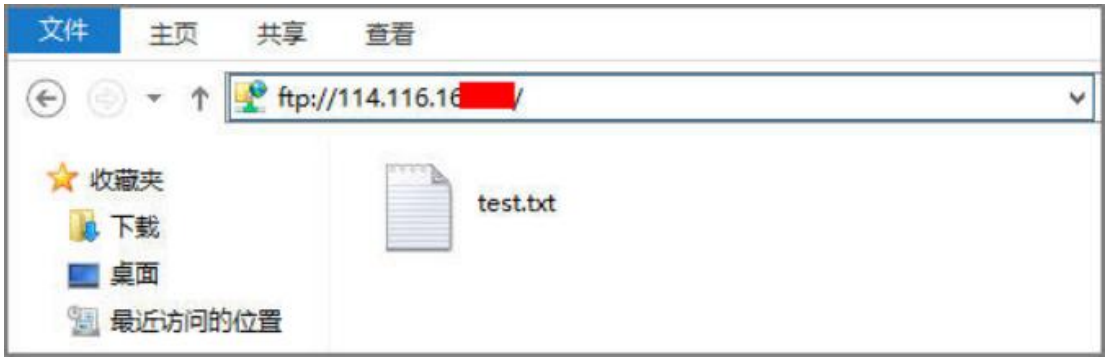
如果配置了“FTP 防火墙支持”，需要在安全组中同时放行 FTP 站点使用的端口和 FTP 防火墙使用的数据通道端口。

服务器防火墙默认放行 TCP 的 21 端口用于 FTP 服务。如果选用其他端口，需要在防火墙中添加一条放行此端口的入站规则。

客户端测试

打开客户端的计算机，在路径栏输入“ftp://FTP 服务器 IP 地址:FTP 端口”(如果不填端口则默认访问 21 端口）。弹出输入用户名和密码的对话框表示配置成功，正确的输入用户名和密码后，即可对 FTP 文件进行相应权限的操作。

说明： 如果没有配置“FTP 防火墙支持”，客户端使用此方法访问 FTP 站点时，需要对 IE 浏览器进行设置，才能打开 FTP 的文件夹。打开 IE 浏览器，选择“设置 > Internet 选项 > 高级”。勾选“启用 FTP 文件夹视图”，取消勾选“使用被动 FTP”。

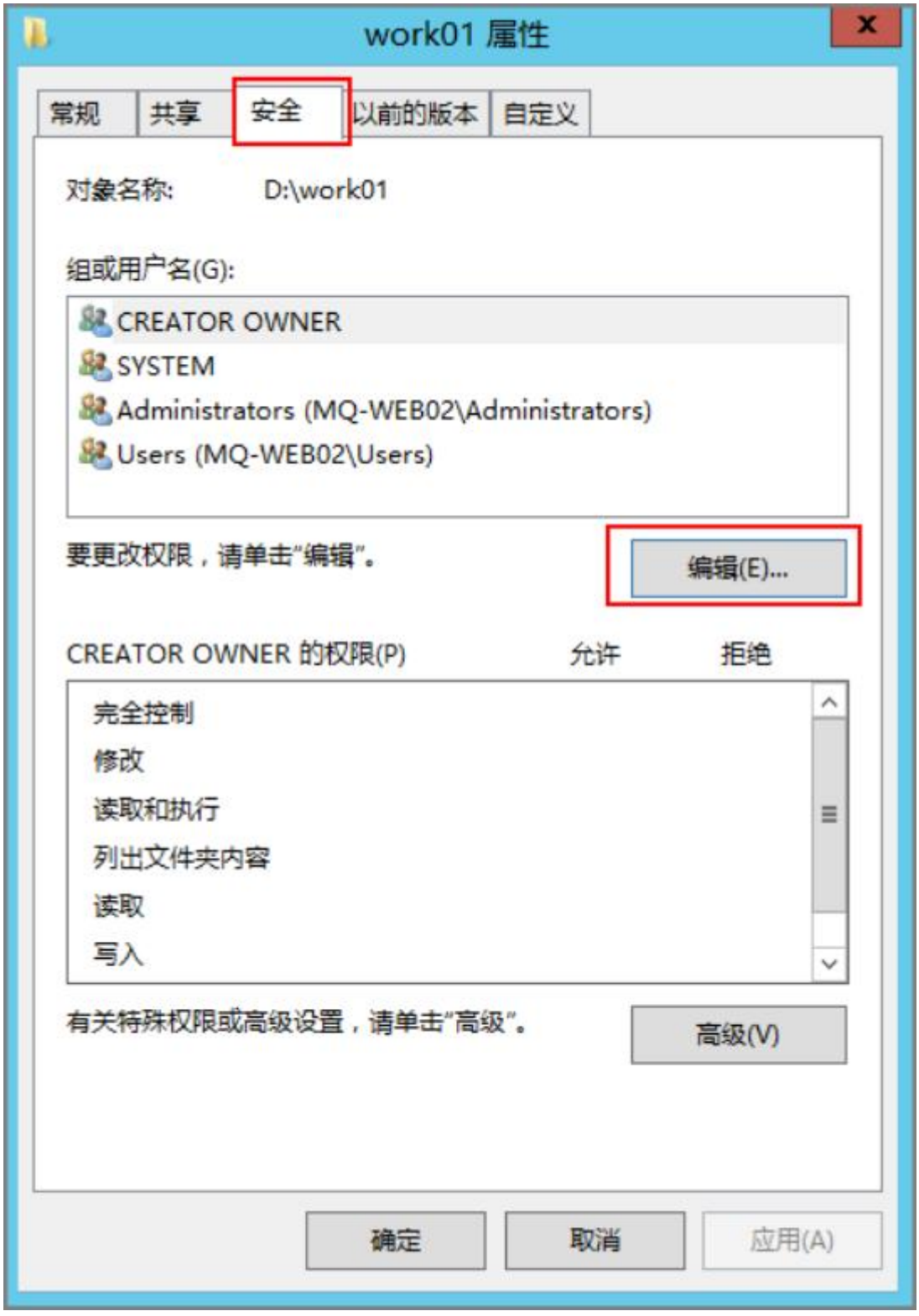


FAQ

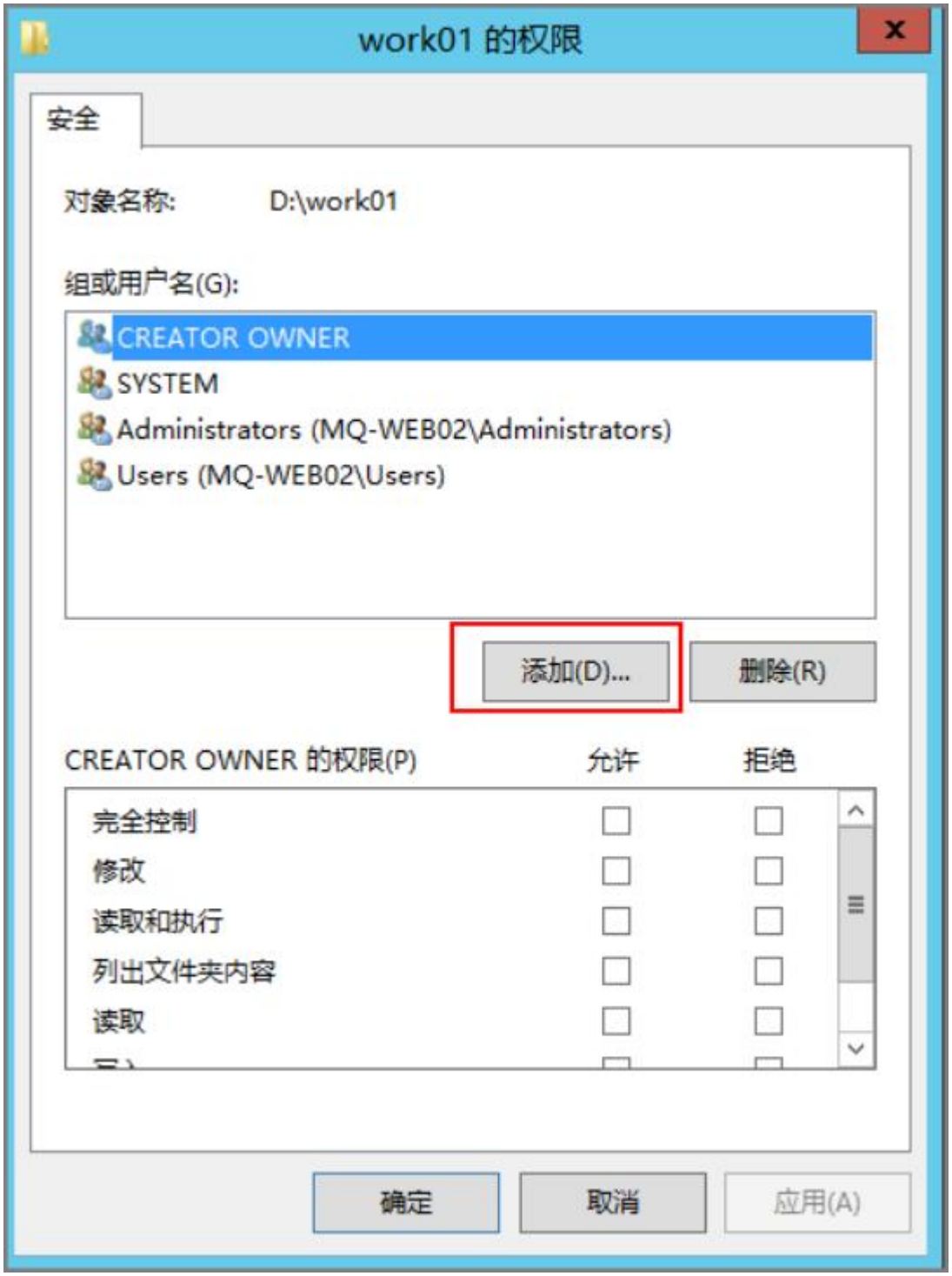
使用 Windows 系统搭建 FTP 站点的更多信息请参见微软官方文档。

如果设置文件夹的属性时，没有“Everyone”用户可直接选择，可按照如下步骤添加“Everyone”用户。

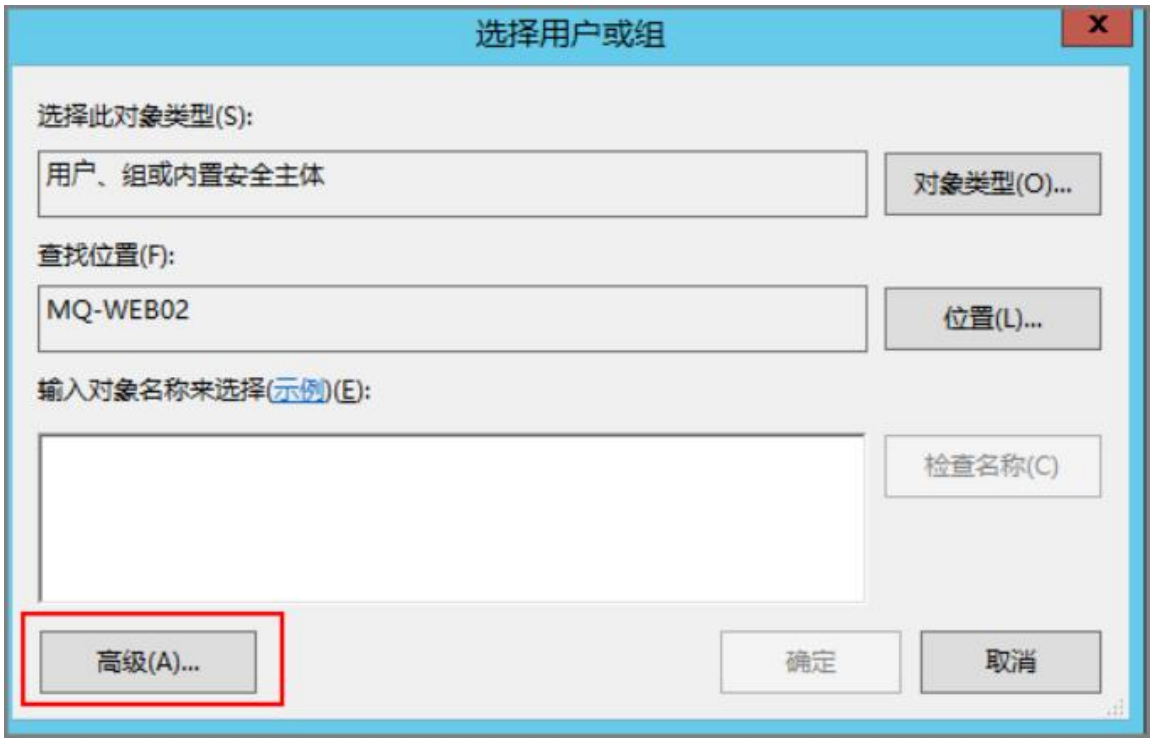
1.在“安全”页签，单击“编辑”。



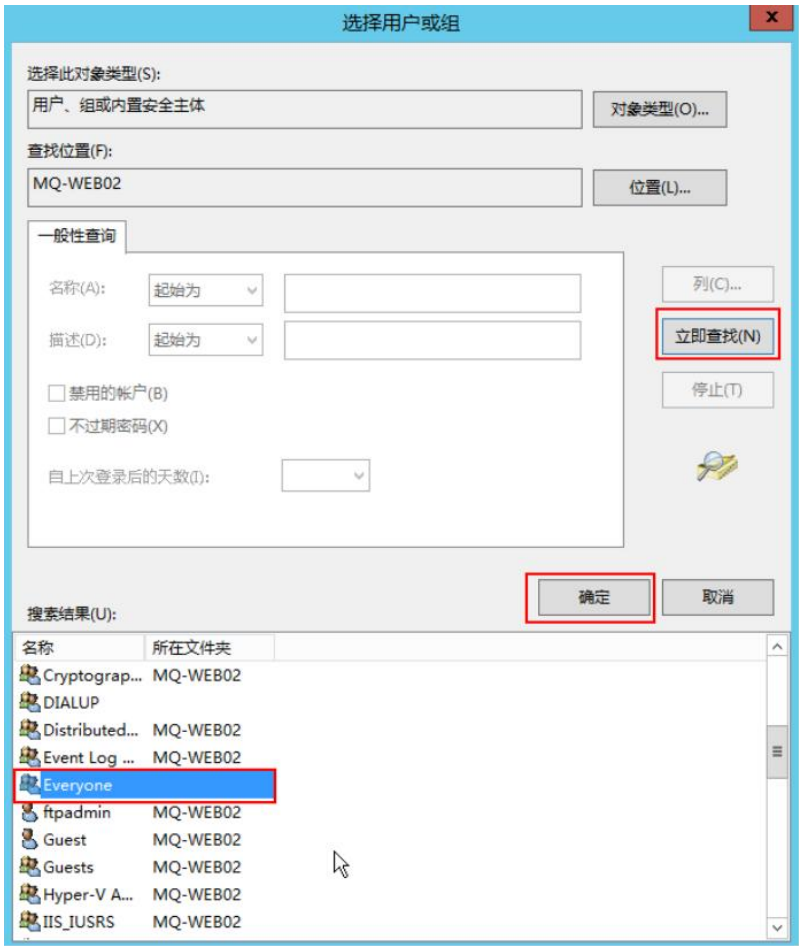
2. 在新弹出的窗口中，单击“添加”。



3. 在新弹出的窗口中，单击“高级”。



4.在新弹出的窗口中，单击“立即查找(N)”，然后在搜索结果中选择“Everyone”并单击“确定”。



5. 单击“确定”，返回权限窗口。

6. 单击“确定”，完成添加。

7. 制作服务器证书。

1）在“服务器管理器”中，选择“仪表板 > 工具 > Internet Information Services (IIS)管理器 ”。



2）在左侧列表单击服务器，然后在服务器主页“IIS”区域，双击“服务器证书”，进入“服务器证书”页面。



3）单击“创建自签名证书”。

4）输入证书的名称，并选择证书存储类型，单击“确定”。

5）创建成功后会在服务器证书界面显示已经创建的证书。

Linux 云主机

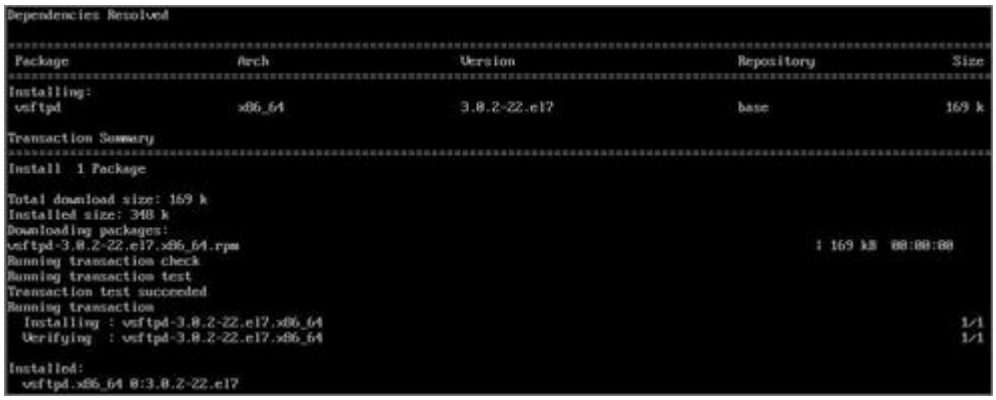
安装 vsftpd

1.登录弹性云主机。

2.执行以下命令安装 vsftpd。

```
yum install -y vsftpd
```

回显如下类似信息时，表示软件安装成功。



3.执行以下命令设置 FTP 服务开机自启动。

```
systemctl enable vsftpd.service
```

4.执行以下命令启动 FTP 服务。

```
systemctl start vsftpd.service
```

5.执行以下命令查看 FTP 服务端口。

```
netstat -antup | grep ftp
```

回显如下类似信息。



配置 vsftpd

vsftpd 安装后默认开启了匿名 FTP 的功能，使用匿名 FTP，用户无需输入用户名密码即可登录 FTP 服务器，但没有权限修改或上传文件。用户如果试图使用 Linux 操作系统中的账号登录服务器，将会被 vsftpd 拒绝，但可以在 vsftpd 里配置用户账号和密码登录。以下操作以在 vsftpd 里配置用户账号和密码登录 FTP 服务器为例。

1.执行以下命令创建“ftpadmin”用户。此处“ftpadmin”用户为示例，可根据实际情况创建相应的用户。

```
useradd ftpadmin
```

2.执行以下命令并按照提示设置“ftpadmin”用户密码。

```
passwd ftpadmin
```

3.执行以下命令创建供 FTP 使用的文件目录，此处以“/var/ftp/work01”为例。

```
mkdir /var/ftp/work01
```

4.执行以下命令将创建的文件目录所有者改为用于登录 FTP 的本地用户。

```
chown -R ftpadmin:ftpadmin /var/ftp/work01
```

5.修改“vsftpd.conf”配置文件。

1）执行以下命令打开配置文件“vsftpd.conf”。

```
vi /etc/vsftpd/vsftpd.conf
```

2.按 i 键进入编辑模式。

修改打开的“vsftpd.conf”文件。可根据实际需求选择将 FTP 配置为主动模式或者被动模式。如果天翼云上的服务器需要通过公网 IP 地址访问天翼云上的实例搭建的 FTP 服务器时，需要将 FTP 服务器配置为被动模式。

主动模式需要配置的参数如下：

#设置以下参数，不允许匿名登录 FTP 服务器，允许本地用户登录 FTP 服务器，并指定 FTP 本地用户使用的文件目录。

anonymous_enable=NO #不允许匿名登录 FTP 服务器

local_enable=YES #允许本地用户登录 FTP 服务器

local_root=/var/ftp/work01 #FTP 本地用户使用的文件目录

#设置以下参数，限制用户只能访问自身的主目录。

chroot_local_user=YES #所有用户都被限制在其主目录

chroot_list_enable=YES #启用例外用户名单

chroot_list_file=/etc/vsftpd/chroot_list #例外用户名单

allow_writeable_chroot=YES

被动模式除了需要配置主动模式所需的所有参数外，还需要配置的参数如下：

#设置以下参数，配置 FTP 支持被动模式。并指定 FTP 服务器的公网 IP 地址，以及可供访问的端口范围，端口范围请根据实际环境进行设置。

listen=YES

listen_ipv6=NO

pasv_address=xx.xx.xx.xx #FTP 服务器的公网 IP 地址

pasv_min_port=3000 #被动模式下的最小端口

pasv_max_port=3100 #被动模式下的最大端口

III.按 Esc 键退出编辑模式，并输入:wq 保存后退出

IV.在“/etc/vsftpd/”目录下创建“chroot_list”文件。

touch chroot_list

“chroot_list”文件是限制在主目录下的例外用户名单。如果需要设置某个用户不受只可以访问其主目录的限制，可将对应的用户名写入该文件。如果没有例外也必须要有“chroot_list”文件，内容可为空。

6.执行以下命令重启 vsftpd 服务使配置生效。

```
service vsftpd restart
```

设置安全组

搭建好 FTP 站点后，需要在弹性云主机安全组的入方向添加一条放行 FTP 端口的规则，具体步骤参见虚拟私有云-安全组-添加安全组规则。

根据 FTP 的不同模式需要放通的端口如下：

- FTP 为主动模式时：端口 21。

- FTP 为被动模式时：端口 21，以及配置文件“/etc/vsftpd/vsftpd.conf”中参数“pasv_min_port”和“pasv_max_port”之间的所有端口。

客户端测试

打开客户端的计算机，在路径栏输入“ftp://FTP 服务器 IP 地址:FTP 端口”(如果不填端口则默认访问 21 端口）。弹出输入用户名和密码的对话框表示配置成功，正确的输入用户名和密码后，即可对 FTP 文件进行相应权限的操作。

说明：　如果 FTP 服务器配置为主动模式，客户端使用此方法访问 FTP 站点时，需要对 IE 浏览器进行设置，才能打开 FTP 的文件夹。打开 IE 浏览器，选择“工具 > Internet 选项 > 高级”。勾选“启用 FTP 文件夹视图”，取消勾选“使用被动 FTP”。

使用浏览器访问 FTP 服务器出错时，建议您清除浏览器缓存后再尝试。

10.28 快速构建 FTP 站点（Windows）

Windows 实例搭建 FTP 站点具体操作步骤

1.添加 IIS 的角色和功能。
2.创建 FTP 服务用户名及密码。
3.设置共享文件权限。
4.配置 FTP 站点。
5.配置 FTP 防火墙。
6.配置安全组规则及防火墙策略。
7.客户端进行连接测试。

示例环境

实例类型：s3.large.2 | 2 核 | 4G 通用型云主机

所在区域：山西

系统盘：40GB

操作系统：Windows 2012 Standard R2 64 位中文版

公网弹性 IP 带宽：1Mbps

操作步骤

1.添加 IIS 的角色和功能。

a）登录弹性云主机。

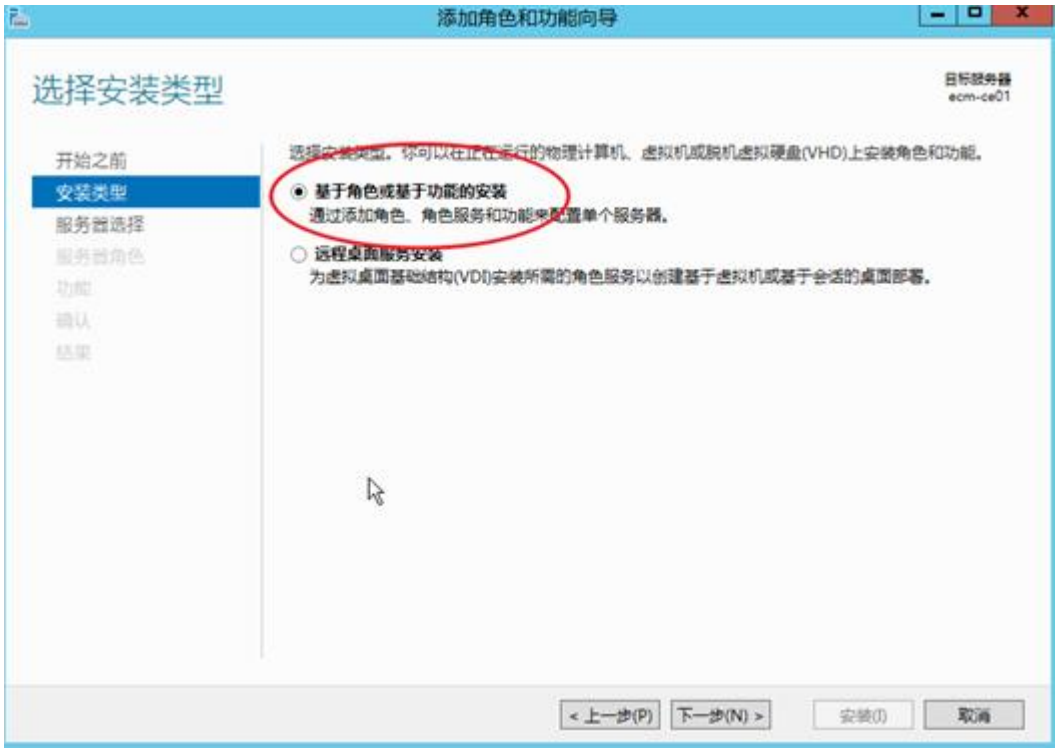
b）选择“开始 > 服务器管理器”。

c）单击“添加角色和功能”。

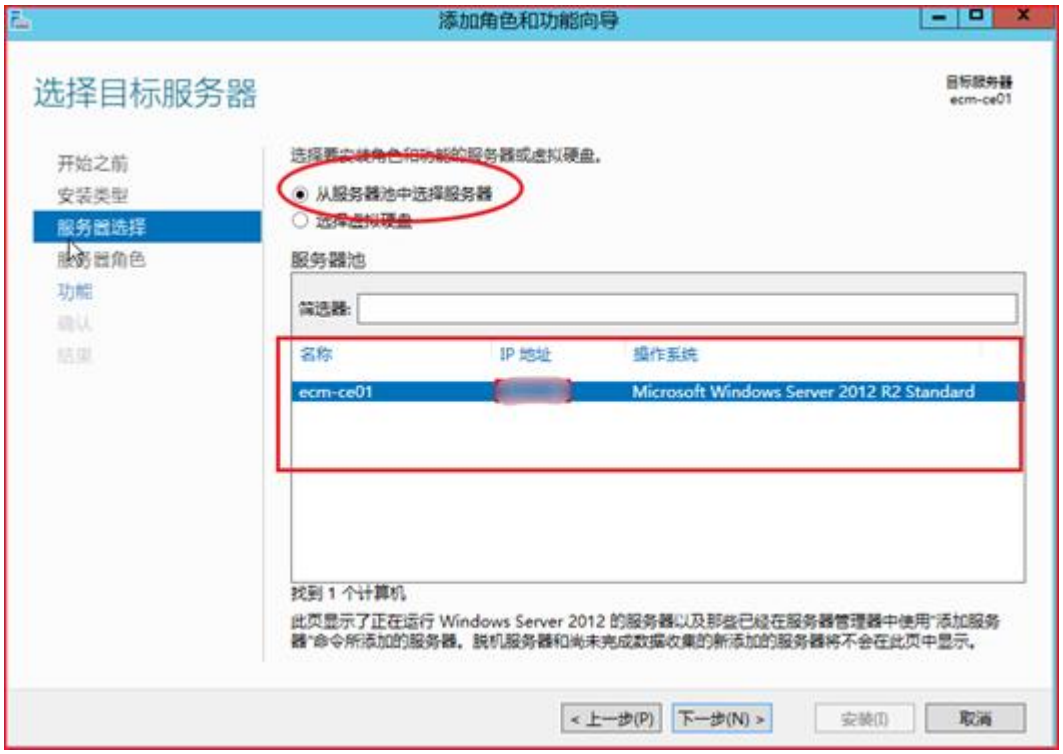


d）在弹出的“开始之前”对话框中，单击“下一步”。

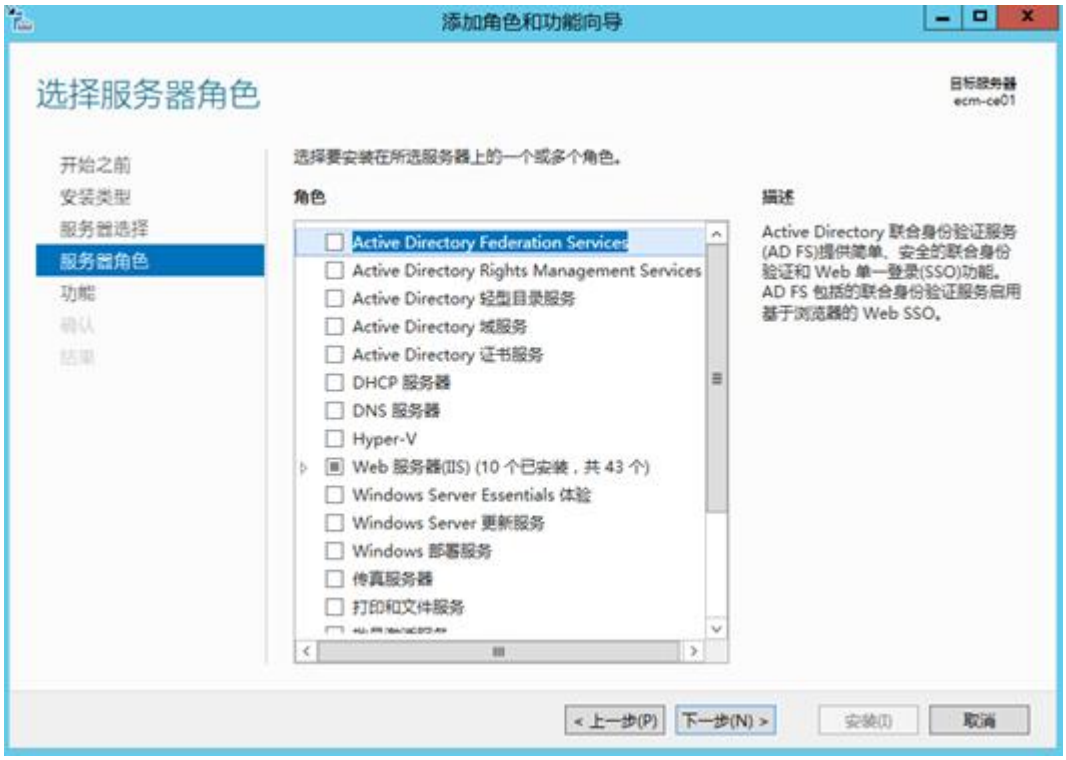
e）选择“基于角色或基于功能的安装”，单击“下一步”。



f）选择需要部署 FTP 的服务器，单击“下一步”。



g）选择“Web 服务器（IIS）”，并在弹出的对话框中单击“添加功能”，然后单击“下一步”。



h）连续单击“下一步”，到“角色服务”页面。

i）选择“FTP 服务器”以及“IIS 管理控制台”，单击“下一步”。



j）单击“安装”，开始部署服务角色。

k）安装完成后，单击“关闭”。

a 添加 IIS 的角色和功能

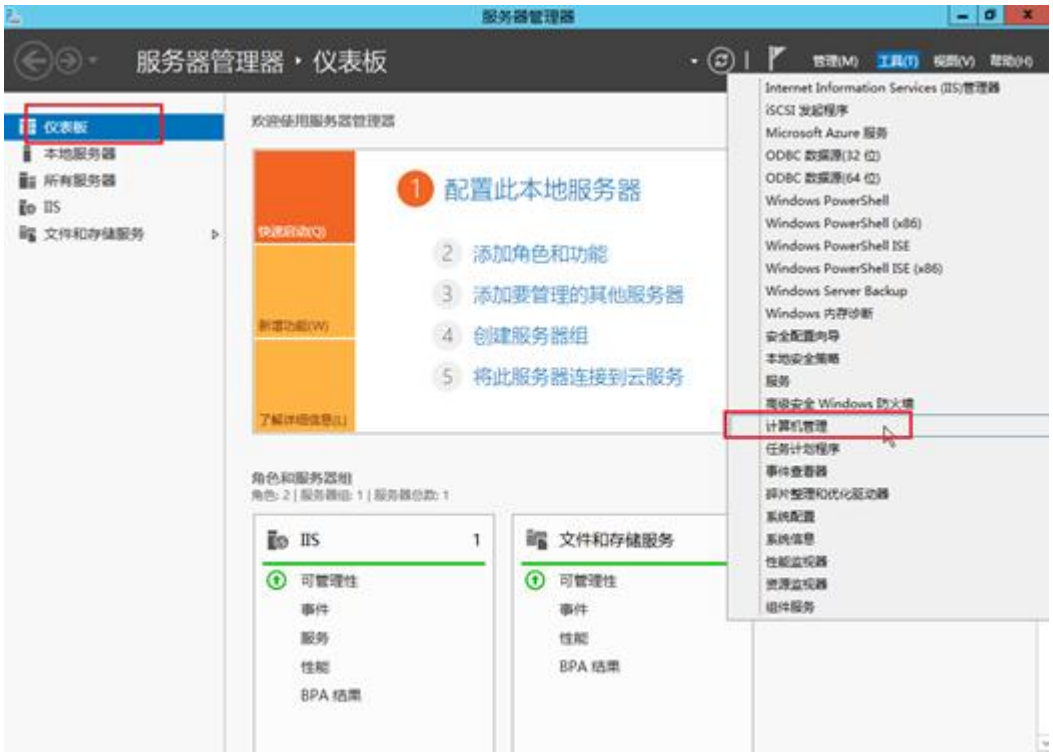
b 添加 IIS 的角色和功能

c 添加 IIS 的角色和功能

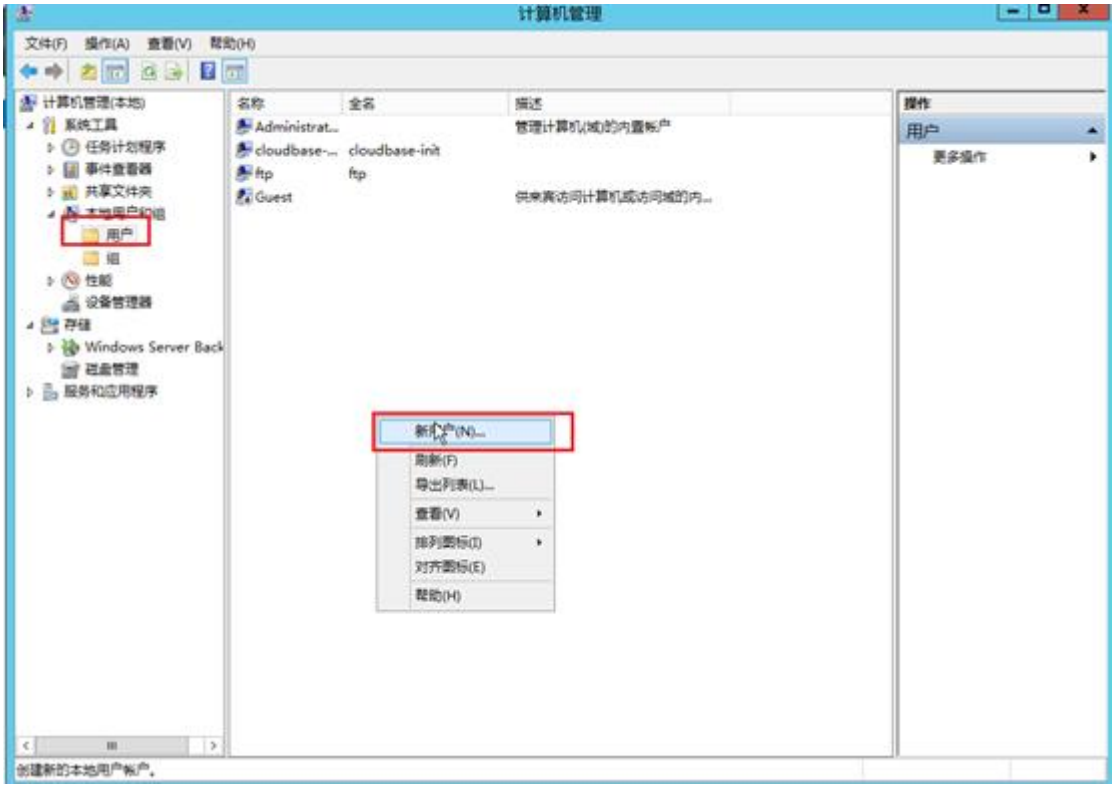
d 添加 IIS 的角色和功能

2. 创建 FTP 服务用户名及密码。

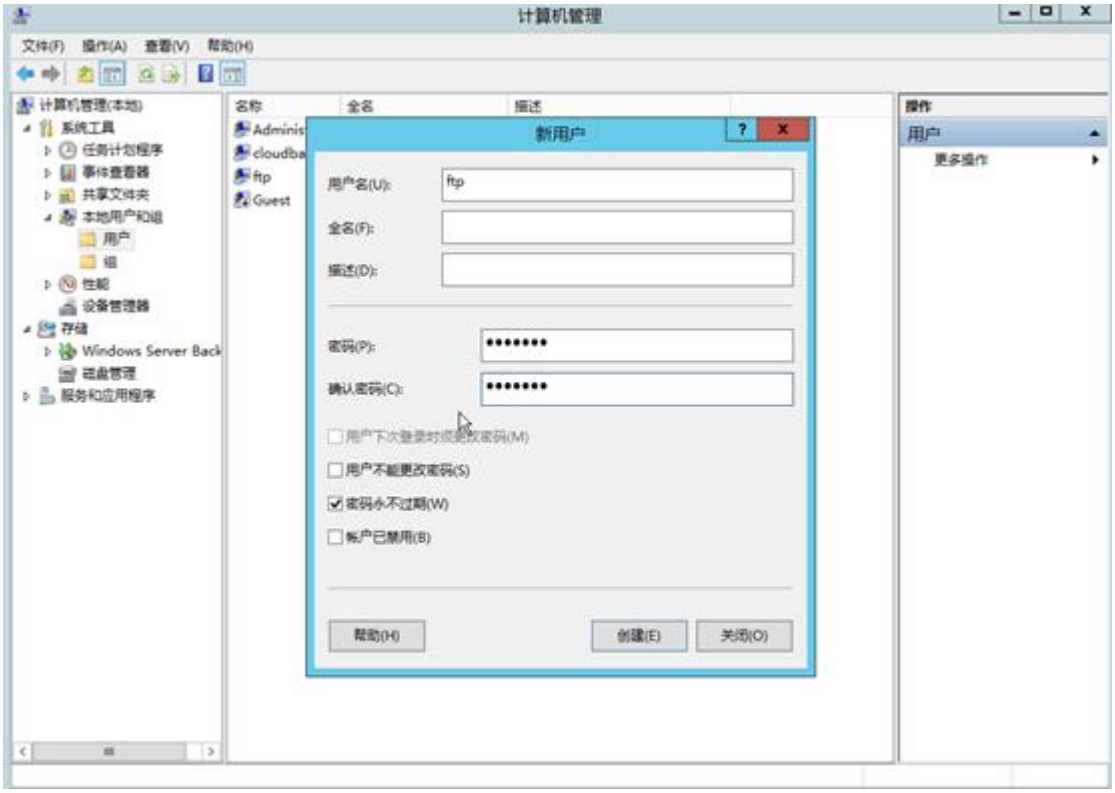
a）在“服务器管理器”中，选择“仪表板 > 工具 > 计算机管理”。



b）选择“系统工具 > 本地用户和组 >用户”，在右侧空白处右击，并选择“新用户”。



c）设置“用户名”和“密码”，此处用户名以“ftp”为例。



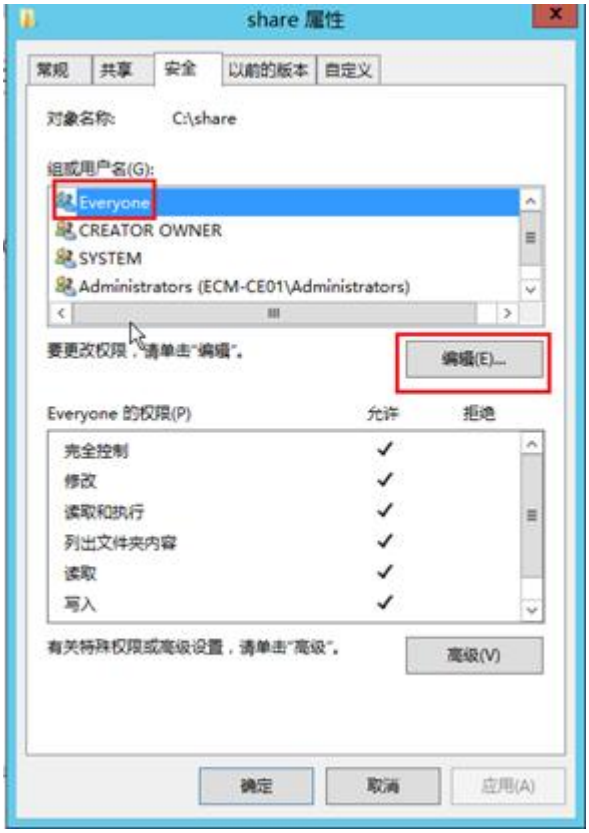
3. 设置共享文件权限。

需要在 FTP 站点为共享给用户的文件夹设置访问及修改等权限。

a）在服务器上创建一个供 FTP 使用的文件夹，选择文件夹，并单击右键选择“属性”。[]()[]()此处以“share”文件夹为例。

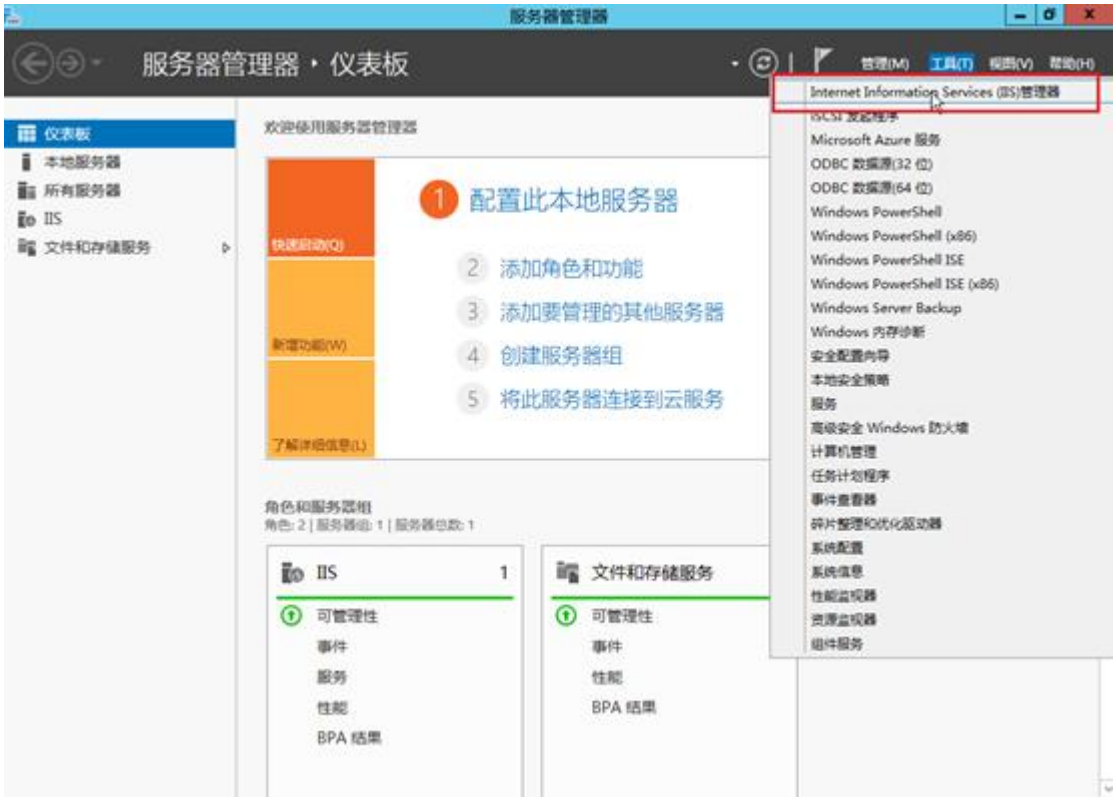
b）在“安全”页签，选择 “Everyone”，单击“编辑”。如果没有“Everyone”用户可以直接选择，需要先进行添加。

c）选择“Everyone”，然后根据需要，选择“Everyone”的权限，并单击“确定”。此处以允许所有权限为例。

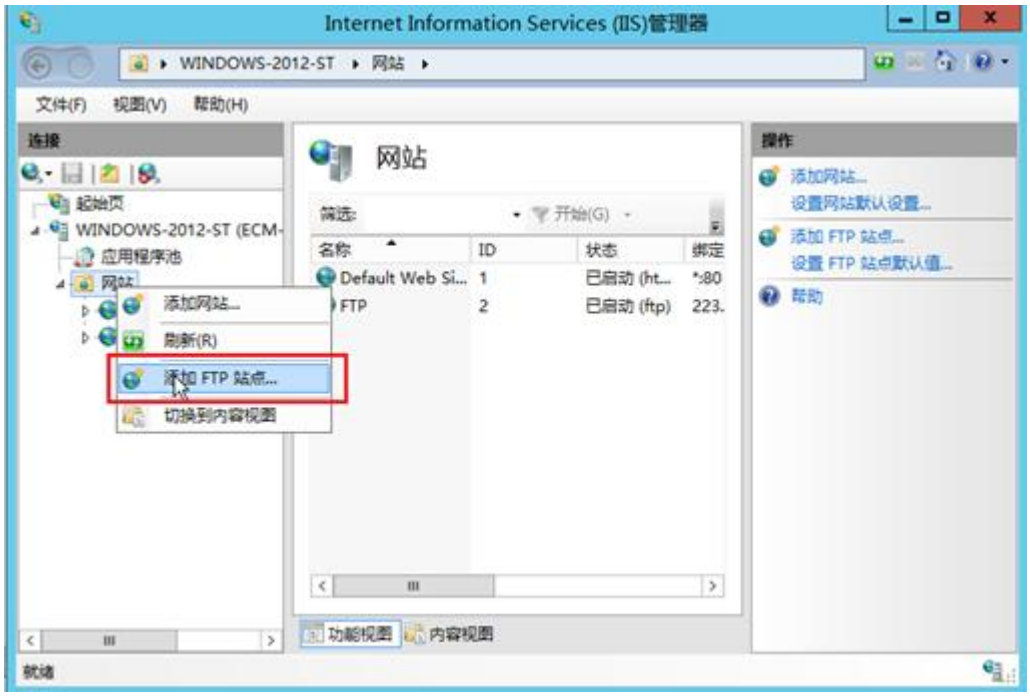


4.配置 FTP 站点。

a）在“服务器管理器”中，选择“仪表板 > 工具 >Internet Information Services (IIS)管理器 ”。



b）选择“网站”并单击右键，然后选择“添加 FTP 站点”。

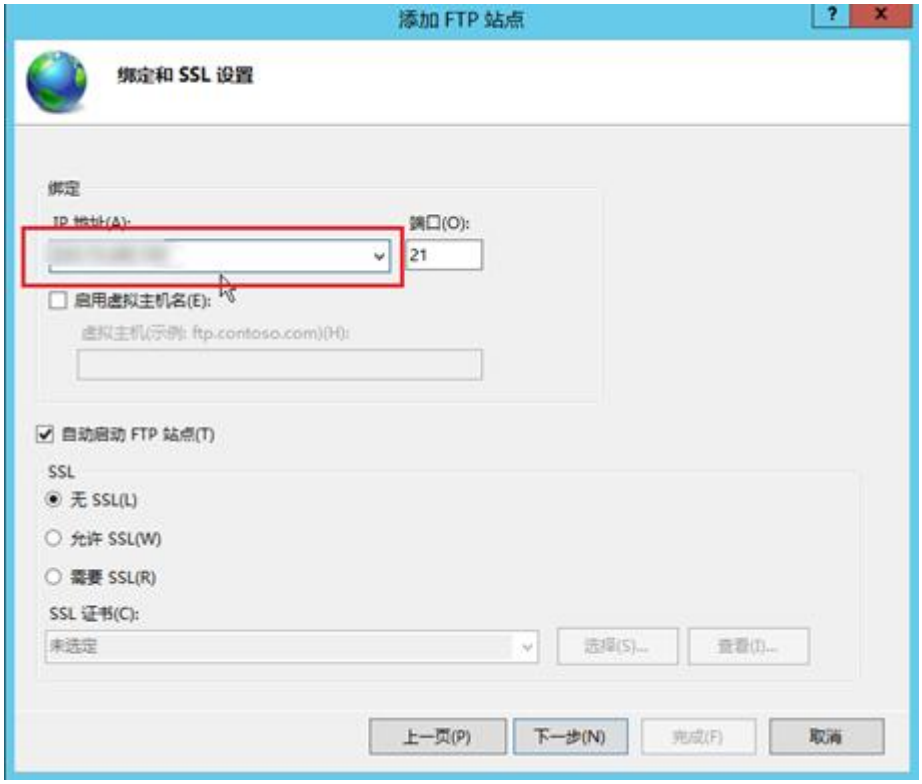


c）在弹出的窗口中，填写 FTP 站点名称及共享文件夹的物理路径，然后单击“下一步”。此处站点名称以“ftp”为例。



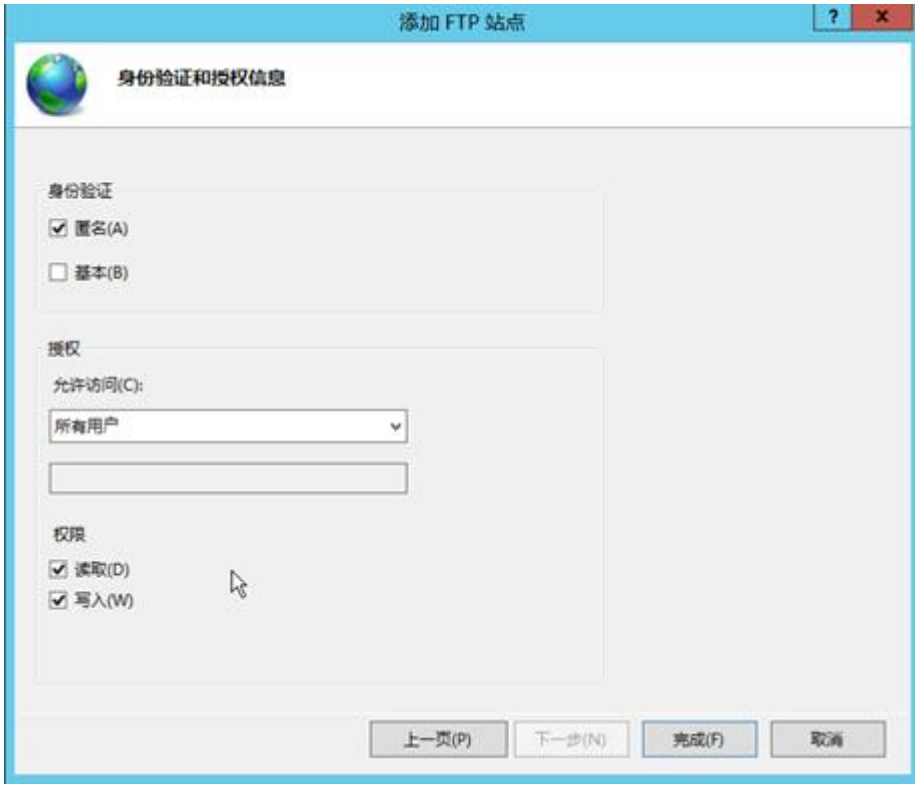
d）输入该弹性云主机的公网 IP 地址以及端口号，并设置 SSL，单击“下一步”。

- 端口号默认为 21，也可自行设置。
- SSL 根据需要进行设置。
  - 无： 不需要 SSL 加密。
  - 允许：允许 FTP 服务器与客户端的非 SSL 和 SSL 连接。
  - 需要：需要对 FTP 服务器和客户端之间的通信进行 SSL 加密。



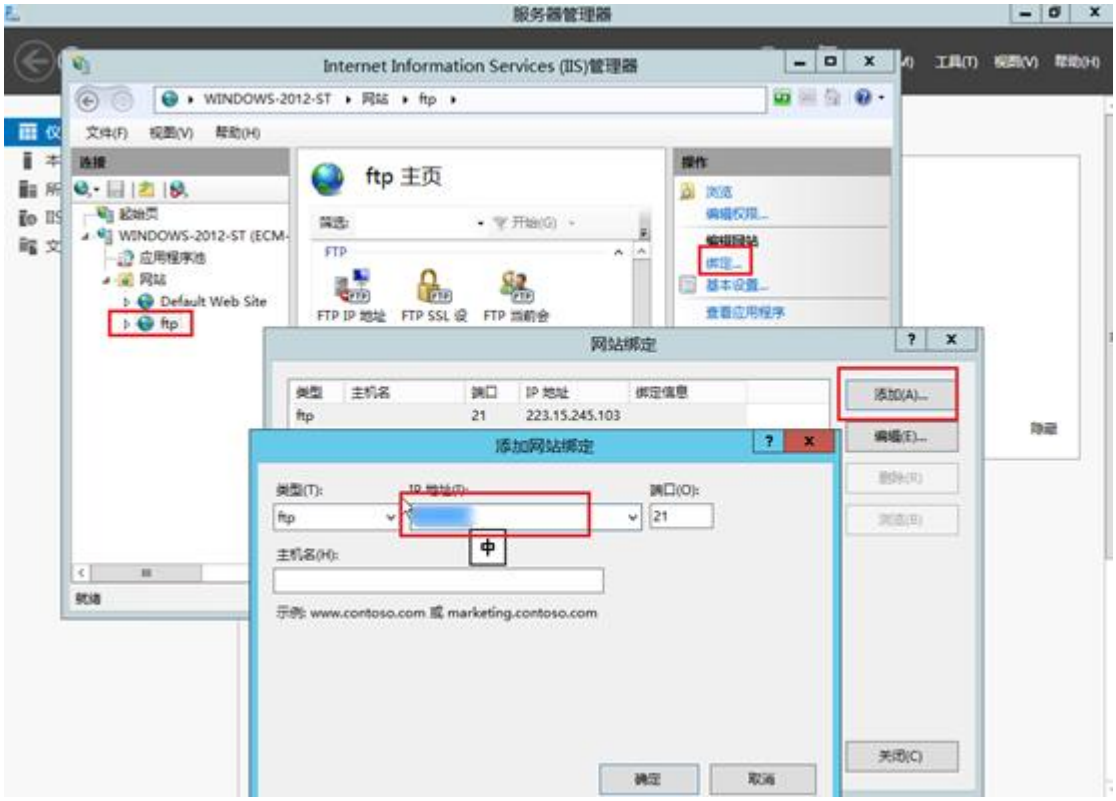
e）设置身份认证和授权信息，并单击“完成”。

- 身份认证
  - 匿名：允许任何仅提供用户名 “anonymous” 或 “ftp” 的用户访问内容。
  - 基本：需要用户提供有效用户名和密码才能访问内容。但是基本身份验证通过网络传输密码时不加密，因此建议在确认客户端和 FTP 服务器之间的网络连接安全时使用此身份验证方法。
- 授权
  - 允许访问
    - 所有用户：所有用户均可访问相应内容。
    - 匿名用户：匿名用户可访问相应内容。
    - 指定角色或用户组：仅指定的角色或用户组的成员才能访问相应内容。如果选择此项，需要在下面输入框中输入指定的角色或用户组。
    - 指定用户：仅指定的用户才能访问相应内容。如果选择此项，需要在下面输入框中输入指定的用户。
      - 权限：选择经过授权的用户的“读取”和“写入”权限



f）绑定弹性云主机的公网 IP。

选择“网站”，选中创建的 FTP 站点，单击“绑定”；在弹出的“网站绑定”窗口单击“添加”，然后在弹出的窗口中添加弹性云主机的私网 IP 地址，并单击“确定。



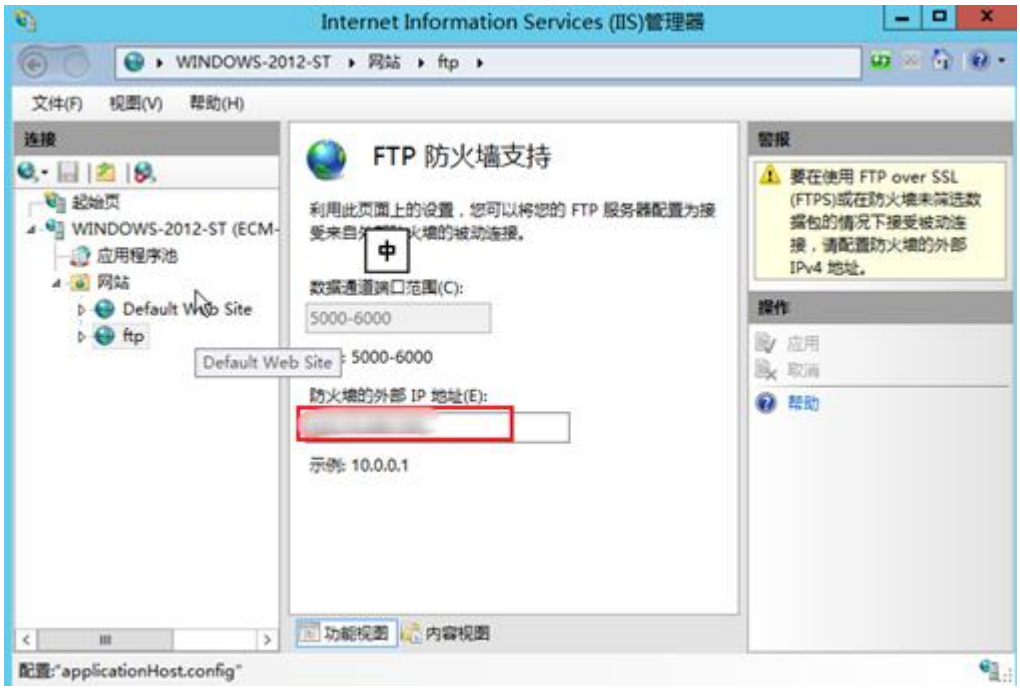
配置 FTP 防火墙支持。

- 如果需要使用 FTP 服务器的被动模式，则需要配置 FTP 防火墙支持。
- 如果天翼云上的服务器需要通过公网 IP 地址访问天翼云上的实例搭建的 FTP 服务器时，需要配置 FTP 服务器的被动模式。

  - 双击“FTP 防火墙支持”，打开 FTP 防火墙支持的配置界面。



- 配置相关参数，数据通道端口范围：指定用于被动连接的端口范围。可指定的有效端口范围为 1025-65535，请根据实际需求进行设置，此处配置 5000-6000，防火墙的外部 IP 地址：输入该弹性云主机的公网 IP 地址，并单击“应用”。



。

- 重启云服务器使防火墙配置生效。

6. 配置安全组规则及防火墙策略。

搭建好 FTP 站点后，需要在弹性云主机安全组的入方向添加一条放行 FTP 端口的规则，如果配置了“FTP 防火墙支持”，需要在安全组中同时放行 FTP 站点使用的端口和 FTP 防火墙使用的数据通道端口。

服务器防火墙默认放行 TCP 的 21 端口用于 FTP 服务。如果选用其他端口，需要在防火墙中添加一条放行此端口的入站规则。

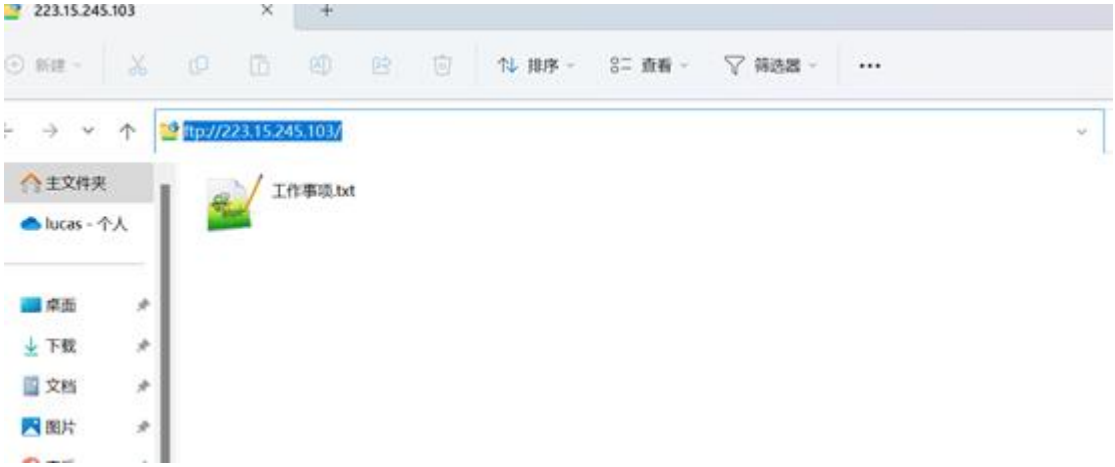
FTP 模式	方向	协议	端口	源地址
主动模式	入方向	TCP	20 端口和 21 端口	0.0.0.0/0

FTP 模式	方向	协议	端口	源地址
被动模式	入方向	TCP	21 端口和 1024~65535 间的端口（根据上文配置，此处 5000-6000）	0.0.0.0/0

7.客户端进行连接测试。

打开客户端的计算机，在路径栏输入“ftp://FTP 服务器 IP 地址 :FTP 端口 ”（如果不填端口则默认访问 21 端口）。弹出输入用户名和密码的对话框表示配置成功，正确的输入用户名和密码后，即可对 FTP 文件进行相应权限的操作，share 文件夹下创建工作事项.txt 文件，在本地访问共享文件夹可以查看并下载。





10.29 快速构建 FTP 站点（Linux）

Linux 实例手动搭建 FTP 站点具体操作步骤

1. 安装 vsftpd
2. 配置 vsftpd
3. 设置安全组
4. 客户端测试

示例环境

实例类型：s3.large.2 | 2 核 | 4G 通用型云主机

所在区域：山西

系统盘：40GB

操作系统：CentOS7.2 64 位

公网弹性 IP 带宽：1Mbps

操作步骤

1. 安装 vsftpd

a）登录弹性云主机实例。

b)执行以下命令安装 vsftpd， yum install -y vsftpd ， 回显如下类似信息时，表示软件安装成功。



c)执行以下命令设置 FTP 服务开机自启动。

```
systemctl enable vsftpd.service
```

d) 执行以下命令启动 FTP 服务。

```
systemctl start vsftpd.service
```

e)执行以下命令查看 FTP 服务端口。

```
netstat -antup | grep ftp
```

回显如下类似信息。

```
[root@ecm-c9e7 ~]# netstat -antup | grep ftp
tcp6       0      0 :::21                   :::*                    LISTEN      3503/vsftpd
```

2.配置 vsftpd

vsftpd 安装后默认开启了匿名 FTP 的功能，使用匿名 FTP，用户无需输入用户名密码即可登录 FTP 服务器,但没有权限修改或上传文件。用户如果试图使用 Linux 操作系统中的账号登录服务器，将会被 vsftpd 拒绝，但可以在 vsftpd 里配置用户账号和密码登录。以下操作以在 vsftpd 里配置用户账号和密码登录 FTP 服务器为例。

a)执行以下命令创建“ftptest”用户，此处“ftptest”用户为示例，可根据实际情况创建相应的用户。

```
useradd ftptest
```

b) 执行以下命令并按照提示设置“ftptest”用户密码 passwd ftptest，输入密码即可。

```
[root@ecm-c9e7 ~]# useradd ftptest
[root@ecm-c9e7 ~]# passwd ftptest
Changing password for user ftptest.
New password:
Retype new password:
passwd: all authentication tokens updated successfully.
```

c)执行以下命令创建供 FTP 使用的文件目录，此处以“/var/ftp/test”为例。

mkdir /var/ftp/test

d)执行以下命令将创建的文件目录所有者改为用于登录 FTP 的本地用户。

chown -R ftptest:ftptest /var/ftp/test

e)修改“vsftpd.conf”配置文件。

- 执行以下命令打开配置文件。

  “vsftpd.conf”， vi /etc/vsftpd/vsftpd.conf

- 按 i 键进入编辑模式。
- 修改打开的“vsftpd.conf”文件。

anonymous_enable=YES #开启匿名用户访问。默认已开启。

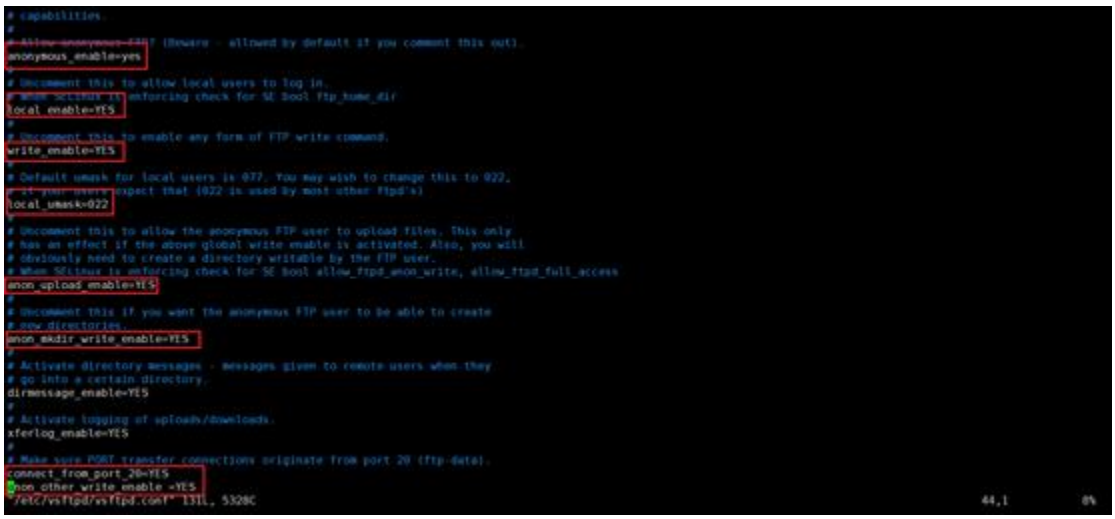
write_enable=YES #开放服务器的写权限（若要上传，必须开启）。默认已开启。

anon_umask=022 #设置匿名用户所上传数据的权限掩码（反掩码）。默认已开启。

anon_upload_enable=YES #允许匿名用户上传文件。默认已注释，需取消注释。

anon_mkdir_write_enable=YES #允许匿名用户创建（上传）目录。默认已注释，需取消注释。

anon_other_write_enable =YES #允许删除、重命名、覆盖等操作。需添加。



f)按 Esc 键退出编辑模式，并输入:wq 保存后退出。

g)执行以下命令重启 vsftpd 服务使配置生效。

```
systemctl restart vsftpd.service
```

h)关闭防火墙和增强型安全功能。

```
systemctl stop firewalld
setenforce 0
```

3.设置安全组

搭建好 FTP 站点后，需要配置弹性云主机的安全组规则，入方向添加一条放行 FTP 端口的规则，设置安全组规则如下表。

FTP 模式	方向	协议	端口	源地址
主动模式	入方向	TCP	20 端口和 21 端口	0.0.0.0/0
被动模式	入方向	TCP	21 端口和配置文件“/etc/vsftpd/vsftpd.conf”中参数“pasv_min_port”和“pasv_max_port”之间的所有端口	0.0.0.0/0

4.客户端测试

a)在 Windows 系统打开开始菜单，输入 cmd 命令打开命令提示符。

b)建立 ftp 连接：ftp XX.XX.XX.XX （虚拟机公网 ip 地址）。

c)输入配置好的用户名和密码，反馈如下表示访问成功。

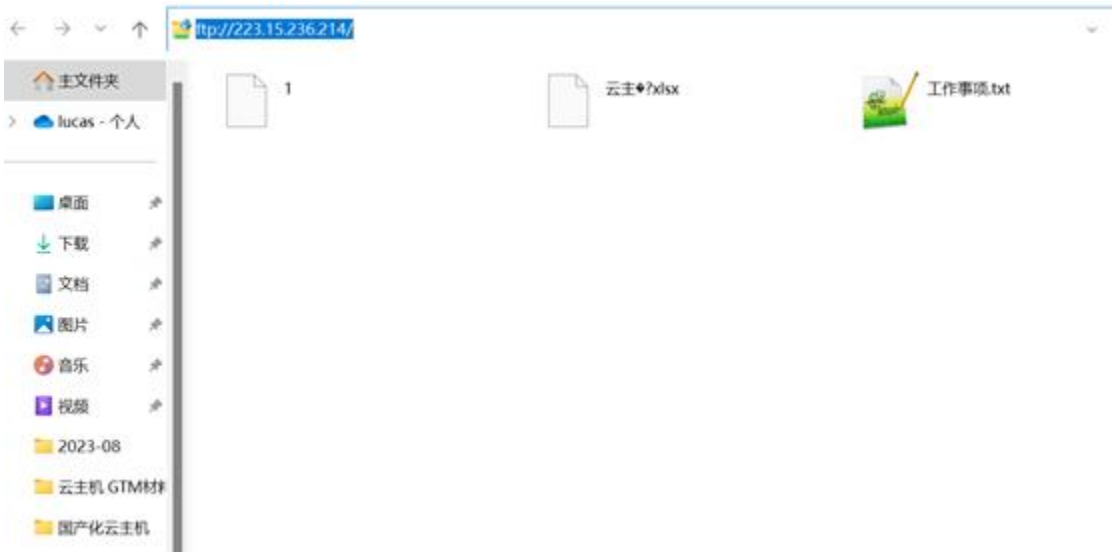


d)ftp> pwd #匿名访问 ftp 的根目录为 Linux 系统的/var/ftp/目录。

e)ftp> ls #查看当前目录。

```
ftp> pwd
257 "/home/ftptest"
ftp> ls
200 PORT command successful. Consider using PASV.
150 Here comes the directory listing.
226 Directory send OK.
ftp>
```

f)打开客户端的计算机，在路径栏输入“ftp://FTP 服务器 IP 地址:FTP 端口”(如果不填端口则默认访问 21 端口)。弹出输入用户名和密码的对话框表示配置成功，正确的输入用户名和密码后，即可对 FTP 文件进行相应权限的操作。



在 linux 云主机上文件目录下一致。

```
[root@ecm-c9e7 ftpadmin]# ls
1  云主???xlsx  工作事项.txt
[root@ecm-c9e7 ftpadmin]#
```

10.30 手工部署 RabbitMQ（CentOS 7.4）

本文介绍了如何在天翼云上使用弹性云主机的 Linux 实例部署 RabbitMQ。RabbitMQ 是采用 Erlang 语言实现 AMQP（Advanced Message Queuing Protocol，高级消息队列协议）的消息中间件，它最初起源于金融系统，用于在分布式系统中存储转发消息。RabbitMQ 凭借其高可靠、易扩展、高可用及丰富的功能特性成为目前非常热门的一款消息中间件。

前提条件

弹性云主机所在安全组添加了如下表所示的安全组规则，具体步骤参见为安全组添加安全组规则。

方向	类型	协议	端口/范围	源地址
入方向	IPv4	TCP	5672	0.0.0.0/0
入方向	IPv4	TCP	15672	0.0.0.0/0

操作步骤

1.安装相关依赖包和 perl。

1）登录弹性云主机。

2）执行以下命令，安装相关依赖包。

```
yum -y install make gcc gcc-c++ m4 ncurses-devel openssl-devel
unixODBC-devel
```

3）执行如下命令，安装 perl。

```
yum install perl
```

2.安装 erlang。

1）执行如下命令，下载 erlang 安装包。

```
wget http://erlang.org/download/otp_src_19.3.tar.gz
```

2）执行如下命令，解压缩安装包。

```
tar xzf otp_src_19.3.tar.gz
```

解压后生成一个“otp_src_19.3”文件夹。

3）执行如下命令，创建文件夹“erlang”。

```
mkdir /usr/local/erlang
```

4）执行如下命令，进入解压后生成的文件夹“otp_src_19.3”。

```
cd otp_src_19.3
```

5）执行如下命令，检查系统是否符合安装要求。

```
./configure --prefix=/usr/local/erlang --without-javac
```

6）执行如下命令，编译并安装 Erlang。

```
make && make install
```

7）执行如下命令，配置 erlang 环境变量。

- 执行如下命令打开配置文件"profile"。

```
vi /etc/profile
```

- 按 i 键进入编辑模式。
- 修改打开的“profile”文件，将如下内容添加到文件末尾。

```
export PATH=$PATH:/usr/local/erlang/bin
```

- 按 Esc 键退出编辑模式，并输入:wq 保存后退出。
- 执行如下命令，使环境变量生效。

```
source /etc/profile
```

- 执行如下命令，检查安装结果。

```
erl -version
```

回显类似如下信息，说明 erlang 安装成功。

```
[root@ecs-rabbitmq ~]# erl -version

Erlang (ASYNC_THREADS,HIPE) (BEAM) emulator version 8.3
```

3.安装 RabbitMQ

1）执行如下命令，进入用户主目录。

```
cd
```

2）执行如下命令，下载 RabbitMQ 安装包。

```
wget
https://www.rabbitmq.com/releases/rabbitmq-server/v3.6.9/rabbitmq-se
rver-generic-unix-3.6.9.tar.xz
```

3）执行如下命令，解压 RabbitMQ 安装包。

```
tar xvJf rabbitmq-server-generic-unix-3.6.9.tar.xz
```

4）执行如下命令，将解压目录移至“/usr/local/rabbitmq”。

```
mv rabbitmq_server-3.6.9 /usr/local/rabbitmq
```

5）执行如下命令，配置 RabbitMQ 环境变量。

6）执行如下命令打开配置文件"profile"。

```
vi /etc/profile
```

- 按 i 键进入编辑模式。
- 修改打开的“profile”文件，将如下内容添加到文件末尾。

```
export PATH=$PATH:/usr/local/rabbitmq/sbin
```

- 按 Esc 键退出编辑模式，并输入:wq 保存后退出。
- 执行如下命令，使环境变量生效。

```
source /etc/profile
```

4.执行如下命令，启用 RabbitMQ 的 web 管理界面。

```
rabbitmq-plugins enable rabbitmq_management
```

回显类似如下信息:

```
[root@ecs-rabbitmq ~]# rabbitmq-plugins enable rabbitmq_management

The following plugins have been enabled:

amqp_client

cowlib

cowboy

rabbitmq_web_dispatch

rabbitmq_management_agent

rabbitmq_management

Applying plugin configuration to rabbit@ecs-rabbitmq... started 6
plugins.
```

5.执行如下命令，创建一个新用户。

```
rabbitmqctl add_user 用户名 密码
```

命令示例：

```
rabbitmqctl add_user root 123456
```

6.执行如下命令，设置用户为管理员。

```
rabbitmqctl set_user_tags 用户名 administrator
```

命令示例：

```
rabbitmqctl set_user_tags root administrator
```

7.执行如下命令，赋予用户所有权限。

```
rabbitmqctl set_permissions -p / 用户名 '.' '.' '.*'
```

命令示例：

```
rabbitmqctl set_permissions -p / root '.' '.' '.*'
```

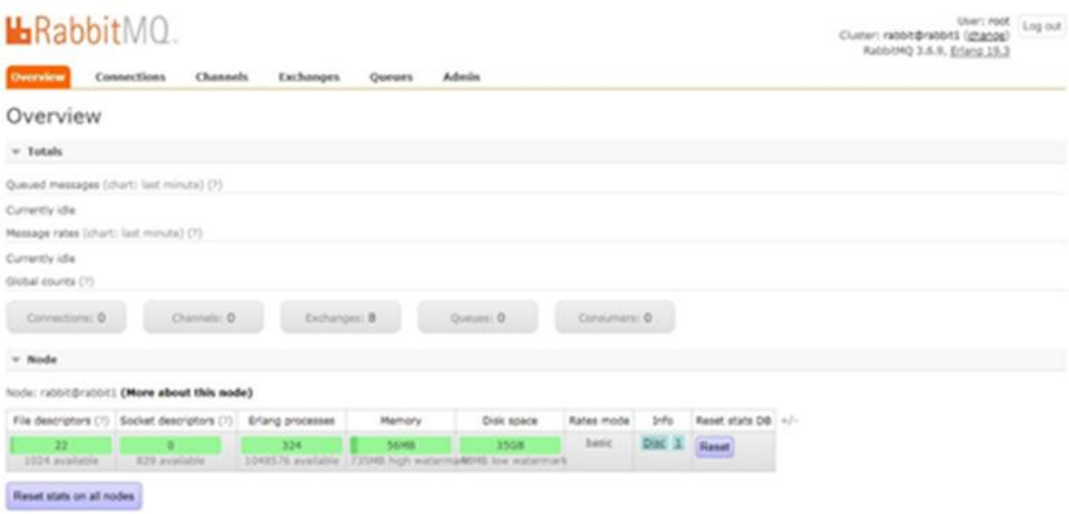
8.执行如下命令，在后台启动 RabbitMQ。

```
rabbitmq-server -detached
```

9.使用浏览器访问 “http://弹性 IP:15672”，显示如下页面，说明 RabbitMQ 安装成功。



10.输入步骤 5 创建的用户名和密码后点击“Login”，进入 RabbitMQ 管理界面。



10.32 在天翼云上配置 NTP 服务器

该 NTP 服务器仅限于在天翼云控制台上购买的弹性云主机安装使用。该功能使用的前提是该节点具备内网 DNS 功能。

如需在购买的弹性云主机上安装 NTP 服务器，您可以选择使用天翼云控制台自己提供的 NTP 服务器，也可以选择其他 NTP 服务器。安装 NTP 服务器的操作相同，本节以天翼云控制台提供的 NTP 服务器为例，介绍在弹性云主机上安装 NTP 服务器的操作。

背景信息

使用天翼云提供的 NTP 服务器时，需和天翼云 DNS 服务器配套使用。天翼云 NTP 服务器配置请参考：内网 yum 源及 NTP 配置。

天翼云参考以下时钟服务器域名：

cn.pool.ntp.org

asia.pool.ntp.org

pool.ntp.org

hk.pool.ntp.org

Linux 操作系统

1.检查弹性云主机的 DNS 服务器域名是否正确。

1）登录 Linux 弹性云主机。

2）执行以下命令，打开 resolv.conf 文件。

```
vi /etc/resolv.conf
```

3）查看文件中 nameserver 的值是否和提供的 DNS 服务器域名相同。

- 是，执行 3。
- 否，执行 2。

2.（可选）给弹性云主机配置 DNS 服务器。

1）登录 Linux 弹性云主机。

2）执行以下命令，编辑 resolv.conf 文件。

```
vi /etc/resolv.conf
```

3）添加如下语句，配置 DNS 服务器。

nameserver DNS 服务器的 IP 地址（具体见内网 DNS 控制台提示）。

3.给弹性云主机配置 NTP 服务器。

1）登录 Linux 弹性云主机。

2）执行以下命令，编辑 ntp.conf 文件。

```
vi /etc/ntp.conf
```

3）添加以下语句，配置 NTP 服务器。

server NTP 服务器域名

示例：

server ntp.ctyun.cn

4）执行以下命令，系统重新启动时启动服务。

- 对于 SUSE 操作系统：

```
service ntp restart
```

- 对于 Euler、CentOS 操作系统：

```
systemctl restart ntpd
```

请根据弹性云主机实际使用的操作系统，以及实际 ntp 服务名，选择具体的命令进行重启。

5）执行以下命令，检查与上层 NTP 服务器是否时钟同步成功。

```
ntpq -p
```

如果在您所配置的时钟服务器所在行的前面出现”*“， 则表示系统已成功同步到您当前配置的时钟服务器。

首次进行 NTP 时钟同步需要一定时间，请等待几分钟再进行检查。

6）请根据弹性云主机实际使用的操作系统，以及实际 ntp 服务名，选择具体的命令设置 NTP 服务的开机自启动。

- 如果是 SUSE 操作系统：

```
chkconfig ntp on
```

- 如果是 Euler、CentOS 操作系统：

```
chkconfig ntpd on
```

Windows 操作系统

以 Windows 7 操作系统为例：

1. 以用户名 Administrator，登录 Windows 弹性云主机。

2. 打开本地连接。

1）在任务栏的右下角，单击网络连接的图标。

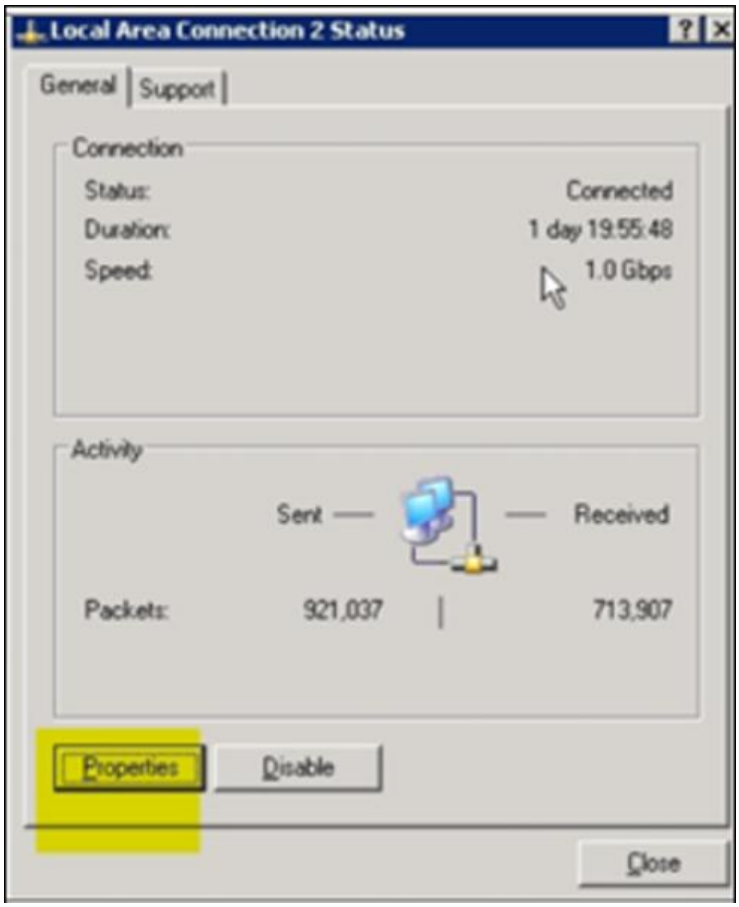
2）单击“打开网络和共享中心”。

3）在左侧导航栏，单击“更改适配器设置”。

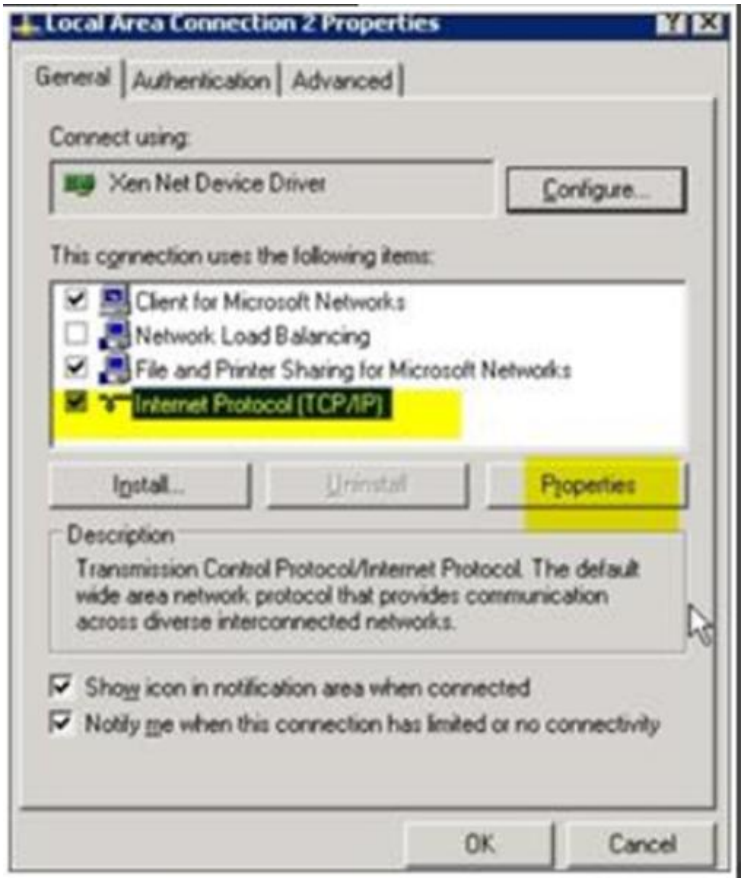
3.给弹性云主机配置 DNS 服务器。

1）双击“本地连接”。

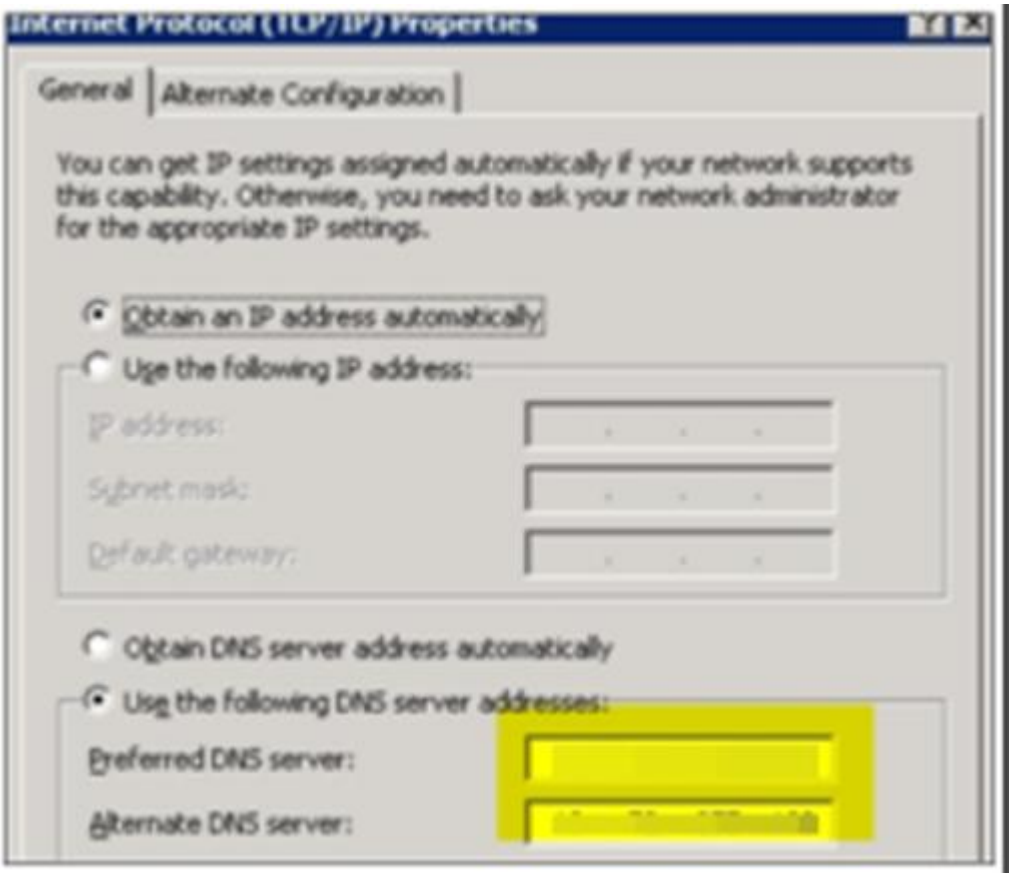
2）单击左下角的“属性”，如下图所示。



3）选择“Internet 协议版本（TCP/IP）”，并单击“属性”，如下图所示。



4）选择“使用下面的 DNS 服务器域名”，并根据界面提示填写 DNS 服务器的 IP 地址，如下图所示。



4.给弹性云主机配置 NTP 服务器。

1）使用快捷键“Win+R”，打开“运行”窗口。

2）在“打开”栏，输入“regedit”，并单击“确定”。打开注册表编辑器。

3）修改如下注册表信息：

- 打开文件“HKEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ Services \ W32Time \ TimeProviders \ NtpClient”，修改“Enabled”的“数值数据”为“1”，表示该设备为 NTP 服务器。
- 打开文件“HKEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ Services \ W32Time \ TimeProviders \ NtpServer”，修改“Enabled”的“数值数据”为“0”，关闭 NTP 服务器功能。
- 打开文件“KEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ Services \ W32Time \ Parameters”,设置“NtpServer”的“数值数据”为“ntp.ctyun.cn”,设置“TYPE”的“数值数据”为“NTP”。

- 打开文件“HKEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ Services \ W32Time \ TimeProviders \ NtpClient”，设置“SpecialPollInterval”的“数值数据”为“60”，“基数”为“十进制”，表示时钟同步周期为60s。
- 打开文件“HKEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ Services \ W32Time \ config”，设置“MaxPosPhaseCorrection”和“MaxNegPhaseCorrection”的“数值数据”为“ffffffff”，“基数”为“十六进制”。

4）使用快捷键“Win+R”，打开“运行”窗口。

5）在“打开”栏，输入“services.msc”，并单击“确定”。打开“服务”窗口。

6）查找名称为“Windows Time”的服务，并设置“启动类型”为“自动”，从NTP服务中同步时间。

7）使用快捷键“Win+R”，打开“运行”窗口。

8）依次执行以下命令，重启Windows Time服务。

```
net stop w32timenet start w32time
```

9）手动修改客户端时间，使得客户端时间与服务器时间不同。一分钟后，检查客户端时间与服务器时间是否一致，如果一致，则表示时间同步配置成功。

后续处理

弹性云主机重启后，DNS配置会被重置为VPC子网的DNS服务器地址。因此，建议在重启弹性云主机前，查询该弹性云主机的VPC子网DNS配置是否与目标DNS配置一致，如果不一致，请先修改弹性云主机VPC子网的DNS服务器地址。

10.31 跨账号同区域迁移云主机

操作场景

本文通过采用镜像迁移方式实现跨账号同区域迁移云主机，以实现跨账号同区域使用镜像快速创建相同配置的弹性云主机。

方案介绍

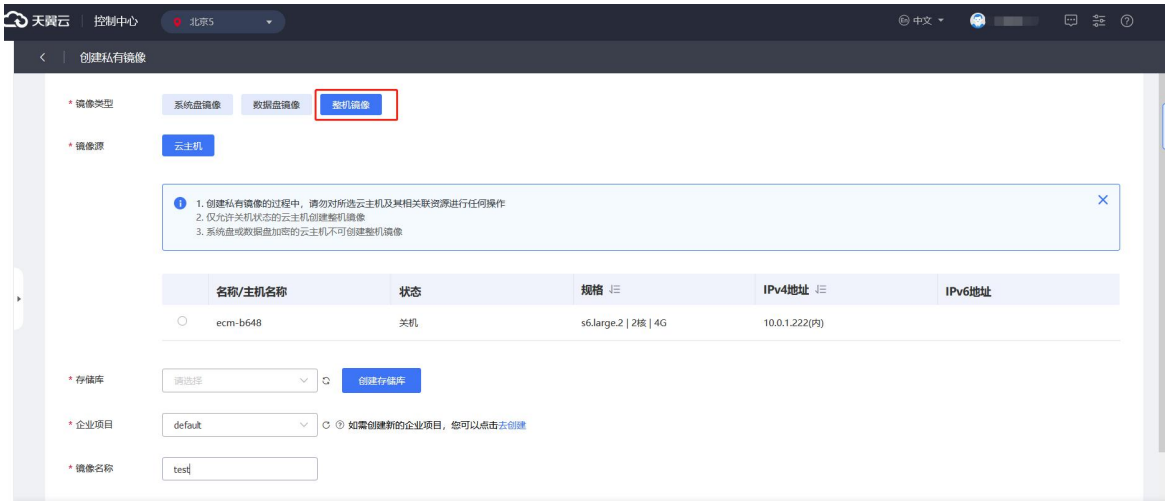
镜像迁移方式实现跨账号同区域迁移云主机的方案为：账号 A 将区域 A 的云主机做成私有镜像，再共享给账号 B；账号 B 接受账号 A 的共享镜像后，使用该镜像创建新的云主机。

解决方案

以下以“北京-北京 5”区域的云主机迁移到另一个账号为实例进行详细介绍。

步骤 1：创建私有镜像

1. 登陆控制中心。
2. 选择区域“北京-北京 5”。
3. 选择“产品服务列表>计算>弹性云主机”，进入云主机列表页面。
4. 选择需要迁移的云主机，单击右侧操作列的“更多>制作镜像”，进入创建私有镜像页面。



5. 填写页面参数，这里镜像类型选择“整机镜像”，镜像源选择“云主机”，选择要迁移的云主机，选择存储库和填写镜像名称。单击“下一步”。

确认无误后，阅读并勾选协议，单击“确定下单”。

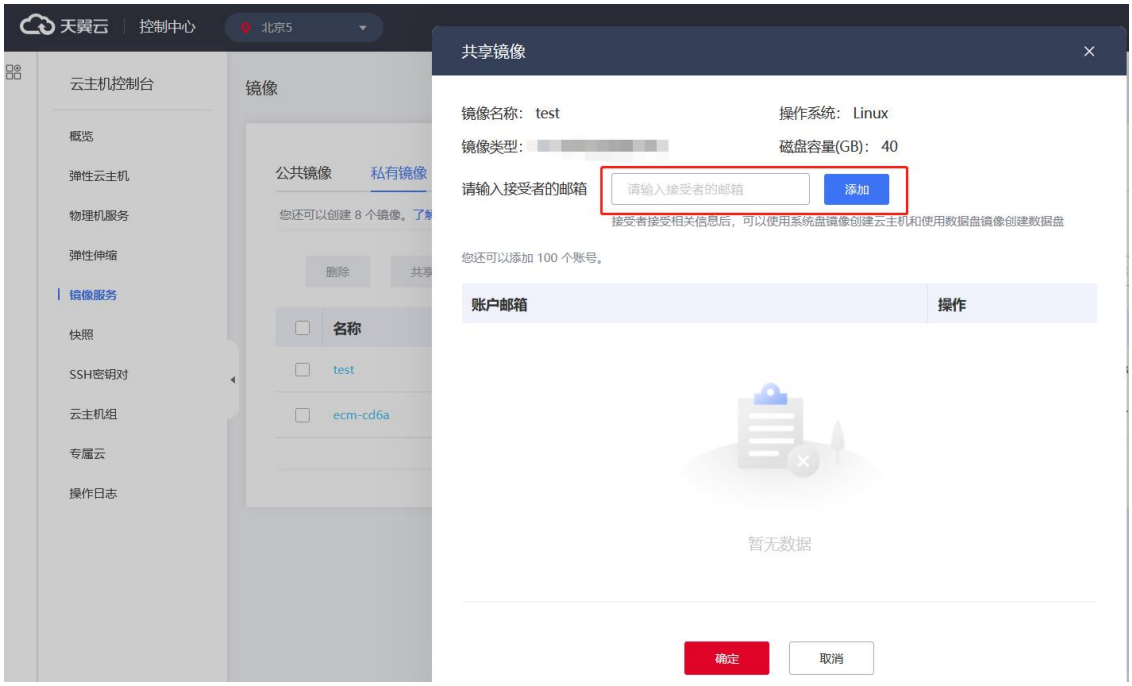
5. 返回私有镜像列表，在列表中查看新创建的私有镜像。

步骤 2：共享私有镜像

1. 选择需要迁移的云主机，单击右侧操作列的“更多>共享”，进入共享镜像页面。

2. 填写共享镜像页面信息。

3. 选择需要共享的镜像和需要共享的账户邮箱，点击确认。



步骤 3：接受共享私人镜像

1. 登陆需要将云主机迁移至的账号。

2. 登陆控制中心。

3. 选择区域 北京-北京 5。

4. 选择“产品服务列表>镜像服务”，进入镜像服务页面，单击“共享镜像”页签。

5. 选择需要迁移的共享镜像申请，单击接受共享，成功接受共享私人镜像。

步骤 4：使用共享镜像创建云主机

6.在共享镜像界面右侧操作栏，单击“申请云主机”，进入云主机创建界面。



7.填写信息，完成弹性云主机的创建，可参考“创建弹性云主机”。

10.32 手工搭建 Ghost 博客（Ubuntu 20.04）

Ghost 是基于 Node.js 开发的一款免费的非常出色的开源博客平台，具有即时预览、极致简约、支持多用户等特性，本文将指导用户在 Ubuntu20.04 操作系统上的云主机实例上部署 Ghost 博客。

前置条件

使用本文的操作步骤，弹性云服务器实例必须满足以下前置：

- 实例需要分配公网 IP 或绑定弹性公网 IP。
- 操作系统：Ubuntu 16.04、Ubuntu 18.04、Ubuntu 20.04、Ubuntu 22.04。
- 实例所在的安全组入方向安全规则放行 22、80、4423、2368 端口。

创建新用户

由于 Ghost 官方不推荐使用 root 用户直接进行操作。因此需要重新创建新的用户，并且为其配置权限。

1.执行以下命令，创建新用户。

1.

以创建 test 用户为例子。

```
adduser test
```

具体配置如下：

```
root@ecm-2d20-Ghost:~# adduser test

Adding user `test' ...
```

```
Adding new group `test' (1000) ...
Adding new user `test' (1000) with group `test' ...
Creating home directory `/home/test' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for test
Enter the new value, or press ENTER for the default
        Full Name []:
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
```

2.执行以下命令，将新创建的用户添加到组。

```
usermod -a -G sudo test
```

3.执行以下命令，切换到 test 用户。

```
su - test
```

安装 gcc 和 g++

1.执行以下命令安装常用开发编译工具。

```
sudo apt-get install build-essential
```

2.执行如下命令，安装 gcc。

```
apt-get install gcc
```

3.执行如下命令查看 gcc 版本。

```
gcc --version
```

回显信息：

```
root@ecm-da5b:~# gcc --versiongcc (Ubuntu 9.3.0-17ubuntu1~20.04)
9.3.0Copyright (C) 2019 Free Software Foundation, Inc.
```

4.执行以下命令安装 g++。

```
sudo apt-get install g++
```

5.使用以下命令查看 g++版本。

```
g++ --version
```

回显信息：

```
root@ecm-da5b:~# g++ --versiong++ (Ubuntu 9.3.0-17ubuntu1~20.04)
9.3.0Copyright (C) 2019 Free Software Foundation, Inc.
```

安装 Node.js

1.在安装 Ghost 前，需要先安装 Ghost 所需要的 Node.js。

执行以下命令，进行 Node.js 安装源的下载。

```
curl -sL https://deb.nodesource.com/setup_12.x | sudo -E bash -
```

2.执行以下命令，进行 Node.js 的安装。

```
sudo apt-get install -y nodejs
```

3.在 Node.js 安装完成后，执行以下命令查看 Node.js 版本和 npm 版本。

```
node -v

npm  - v
```

回显信息：

```
test@ecm-2d20-Ghost:~$ node -vv16.20.2test@ecm-2d20-Ghost:~$ npm -v8.19.4
```

安装 Nginx

部署 Ghost 博客，需要安装 Nginx 作为 HTTP 服务器。

1.执行以下命令，进行 Nginx 安装。

```
sudo apt-get update

sudo apt-get install nginx
```

2.安装完成后，执行以下命令查看 Nginx 版本。

```
nginx  -v
```

回显如下：

```
test@ecm-2d20-Ghost:~$ nginx -vnginx version: nginx/1.18.0 (Ubuntu)
```

3.验证 Nginx 是否正常启动。

浏览器输入公网 IP 地址或者域名进行访问 Nginx,出现如下页面，则说明 Nginx 安装成功。



4.Nginx 配置信息修改（可选）。

输入以下命令修改配置文件。

```
vim  /etc/nginx/nginx.conf
```

输入以下内容：

```
server {
    listen 8001; //监听端口号 本文选择监听端口为8001，避免出现80端口冲突问题
    server_name 171.111.xxx.xxx; #这里写你的域名或者ip地址
    location / {
        proxy_set_header   X-Real-IP $remote_addr;
        proxy_set_header   Host      $http_host;
        proxy_pass         http://127.0.0.1:2368;
    }
}
```

安装MySQL

1.执行以下命令，安装MySQL。

```
sudo apt-get install -y mysql-server
```

2.输入以下命令查看MySQL版本。

```
mysql  -V
```

回显如下：

```
root@ecm-da5b:~# mysql -Vmysql  Ver 8.0.34-0ubuntu0.20.04.1 for Linux on x86_64 ((Ubuntu))
```

3.打开MySQL控制台

通常，在未执行任何配置的情况下在 Ubuntu 上全新安装 MySQL 后，访问服务器的用户将使用身份验证套接字（auth_socket）插件进行身份验证。

auth_socket 的使用会阻碍服务器使用密码对用户进行身份验证。它不仅会引发

安全问题，而且还会使用户无法使用外部程序访问数据库。因此我们需要将身份验证方法从 auth_socket 更改为使用 mysql_native_password(密码验证)。输入以下命令，打开 MySQL 控制台。

```
sudo mysql
```

4.设置 root 用户密码。

输入以下命令将 MySQL 身份验证方式从 auth_socket 更改为使用。

mysql_native_password（密码验证）

```
ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY '输入 root 账号的密码'
```

5.执行以下命令，创建 Ghost 使用的数据库。

本文以 ghost 为例。

```
create database ghost
```

6.输入以下命令查看数据库创建情况。

```
show databases
```

回显信息：

```
mysql> show databases;
+--------------------+| Database           |
+--------------------+
| ghost              |
| information_schema |
| mysql              |
| performance_schema || sys                |
+--------------------+
 rows in set (0.00 sec)
```

5.执行以下命令,退出 MySQL。

```
exit
```

避免数据库可能存在的中文乱码问题，可执行以下命令设置 MySQL 编码。

```
sudo vi /etc/my.cnf
```

6.输入以下内容，之后保存并退出。

```
[client]default-character-set=utf8  [mysql]default-character-set=utf8
[mysqld]character-set-server=utf8  collation-server=utf8_general_ci
```

7.执行以下命令,重启 MySQL 生效

```
sudo /usr/sbin/service mysql restart
```

安装 Ghost

1.输入以下命令创建 Ghost-CLI 命令工具，快速配置 Ghost。

```
sudo npm install ghost-cli@latest -g
```

2.依次执行以下命令，创建 Ghost 安装目录并进授权。

```
sudo mkdir -p /var/www/ghost

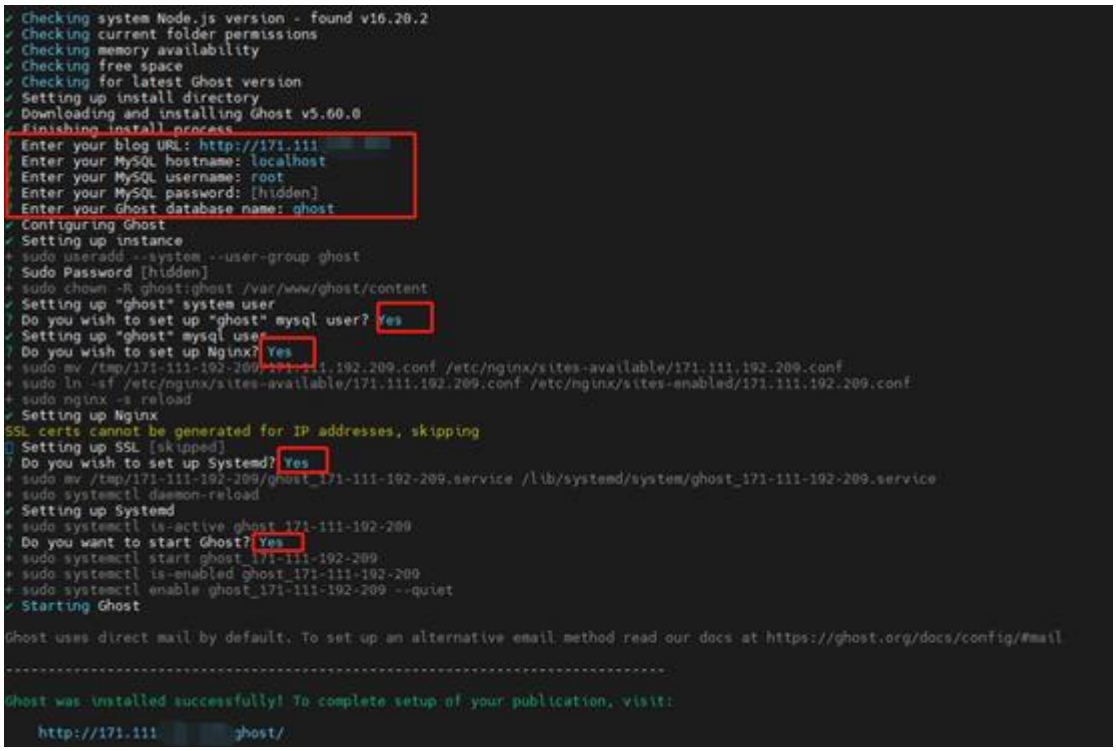
sudo chown test:test /var/www/ghost //配置权限

sudo chmod 775 /var/www/ghost
```

3.执行以下命令，进入创建的文件夹，安装 Ghost。

```
cd /var/www/ghost

ghost install
```



一些配置参数如下：

Enter your blog URL：请输入已解析的域名或者公网 IP，请输入 http://(您的域名或者公网 IP)。

Enter your MySQL hostname：请输入数据库连接地址，请输入 localhost 后按 Enter。

Enter your MySQL username：请输入数据库用户名，请输入 root 后按 Enter。

Enter your MySQL password：请输入数据库密码，请输入在设置 root 账户中已设置的密码后按 Enter。

Enter your database name：请输入 Ghost 使用的数据库的名称，请输入在创建数据库中已创建的 ghost 后按 Enter。

之后会有一些参数的配置，可按照需要进行设置，本文均输入 yes。

4.测试访问 Ghost 博客。

访问 Ghost 前台首页。

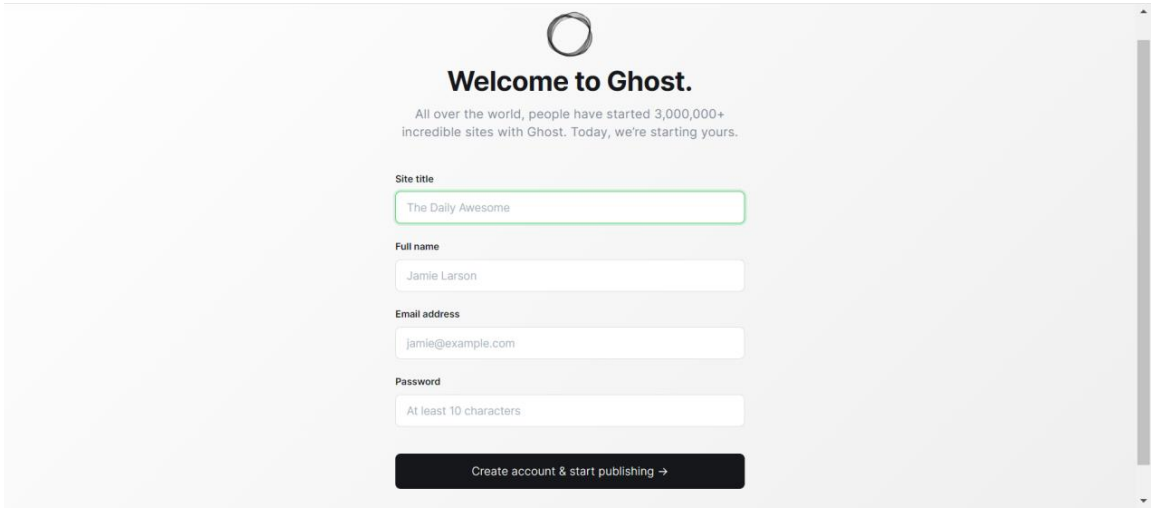
在任何具有访问公网能力的主机中，打开浏览器输入云主机实例的公网 IP 地址或者域名，即可访问 Ghost 前台首页。

例如：http://公网 IP 地址或者域名



如果想要进入管理员后台的话只需要在域名或者公网 IP 地址的后面加上/ghost 即可进入，第一次进入需要进行账号注册，然后便能进入后台。

例如：https://公网 IP 或者域名/ghost



10.33 手工安装宝塔面板（CentOS7.2）

应用场景

本文主要介绍如何使用 Linux 操作系统云主机手工安装宝塔面板。宝塔面板是一款服务器管理软件，支持 Windows 和 Linux 系统，可以通过 Web 端轻松管理服务器，提升运维效率。例如：创建管理网站、FTP、数据库，拥有可视化文件管理器、可视化软件管理器，以及可视化 CPU、内存、流量监控图表、计划任务等功能。同时，宝塔面板还提供了简单直观的用户界面，使得初学者也能够轻松

地管理自己的服务器。本文以 CentOS 7.2 64 位操作系统为例，介绍安装宝塔 Linux 面板过程。

前提条件

云主机资源配置和操作系统要求：

- 云主机规格最低 1 核 1G（软件要求内存最低 512MB，推荐 768MB 以上，纯面板约占系统 60MB 内存）。
- 云硬盘：任意空间大小（软件要求 300M 以上可用硬盘空间，纯面板约占 20M 磁盘空间）。
- 架构：支持 x86_64 架构（ARM 架构不完整兼容，面板环境安装慢，部分软件可能无法安装，请谨慎尝试）。
- 宝塔 Linux6.0 版本基于 Centos7 开发，建议使用 Centos 7.x 系统。
- 云主机全新且环境干净，未安装过 Apache、Nginx、php、MySQL。

安装宝塔面板操作流程

Linux 实例手工安装宝塔面板的具体操作步骤如下：

1.安装宝塔面板。
2.修改云主机所在安全组规则。
3.登录宝塔面板。

实施步骤安装宝塔面板。
1）登录云主机。
2）执行以下命令，下载并安装宝塔面板。

```
yum install -y wget && wget -O install.sh
http://download.bt.cn/install/install_6.0.sh && sh install.sh
```

待命令执行完毕。在执行过程中，会提示是否将磁盘挂载到/www，如下图所示：

```
Do you want to install Bt-Panel to the /www directory now?(y/n): y
```

输入小写 y 字母确认挂载。

安装完成后，会提示安装成功，并显示面板 URL、端口号、“username”、“password”等参数值。



说明

请妥善记录回显信息中 URL、端口号、“username”、“password”等参数值，于后续步骤中使用。

2. 修改云主机所在安全组规则。

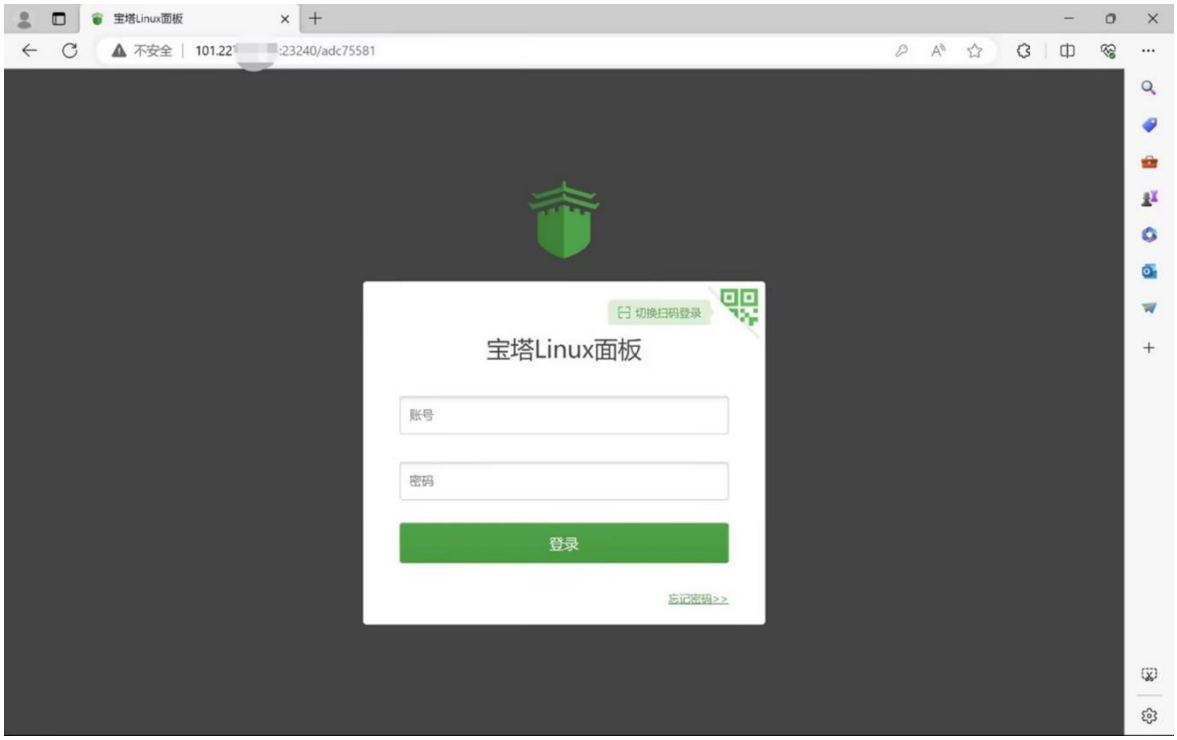
为云主机所在安全组添加安全组规则，规则如下表所示，具体步骤参见添加安全组规则。

方向	协议	端口/范围	源地址
入方向	TCP	回显信息中的端口号（本示例中为23240）	0.0.0.0/0

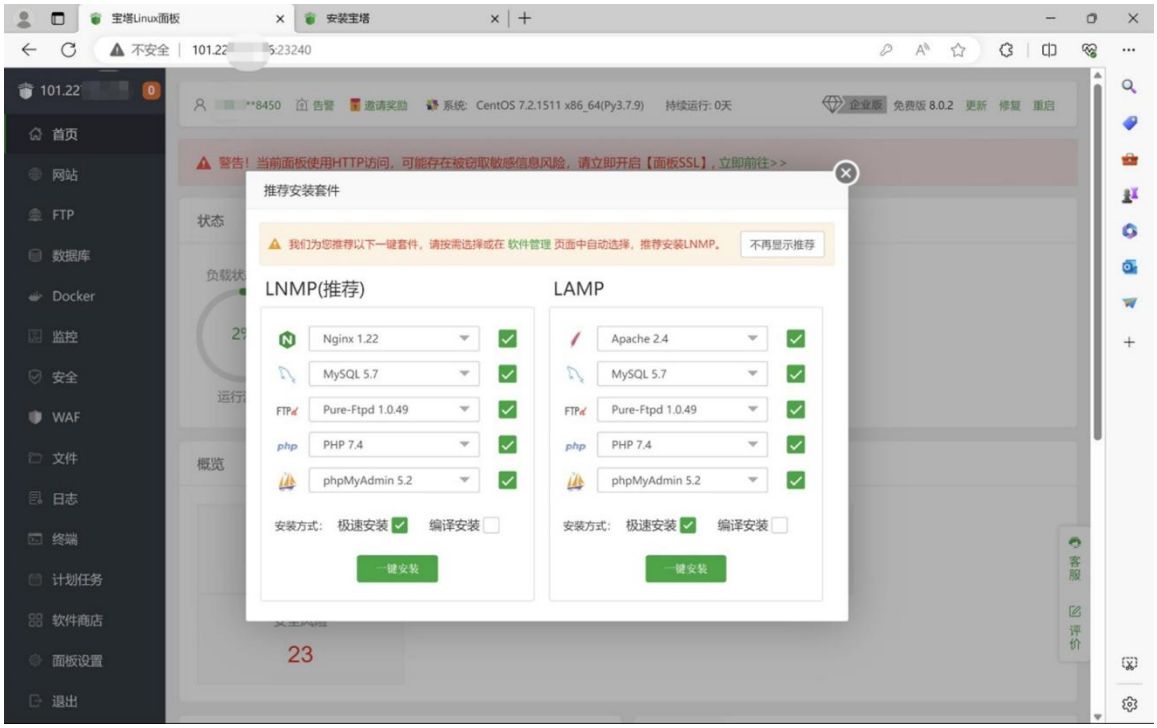
3. 登录宝塔面板。

1）在浏览器地址栏中输入回显信息中的 URL 地址，例如：

http://101.227.xxx.xx:23240/adc75581。成功访问，显示如下图：



2）输入安装步骤中记录的账户和密码，勾选宝塔用户协议，绑定宝塔账号后，进入宝塔管理界面。



3）根据业务需要可以通过面板安装相关的套件并部署网站。

10.34 快速部署 SD-WAN vCPE 最佳实践

镜像描述

预装天翼云 SD-WAN 智能网关软件版 vCPE，支持客户快速部署入云服务。推荐配置：CPU 核数≥2， 内存≥4G、挂载弹性 IP。仅对 ECS 的主网卡，并且选择使用自动分配 IPv4 地址，挂载弹性 IP 场景使用。

配置要求

本镜像推荐配置：vCPU≥2 核，内存≥4G，系统盘≥40G。

免责说明

本公司发布的镜像作为 SDWAN 虚拟智能网关设备提供入云服务，仅供用户参考和使用。本公司不对因不当使用该镜像而导致的任何损害承担法律责任，用户应当确保其使用行为符合相关法律法规，并遵守相应协议。

实践指南

智能网关软件版（vCPE）是 SD-WAN 智能网关的软件镜像版，使用 vCPE 无需购买物理设备，vCPE 作为镜像部署在云主机中，使云主机作为一个虚拟智能网关设备为用户提供更加灵活、便捷的入云服务。开通并部署 vCPE 后，用户可以通过 SD-WAN 智能网关控制台管理 vCPE，并对其进行网络配置，将用户的网络接入云资源池。本文将为您介绍如何在云主机中部署 vCPE 镜像。

步骤一：开通 SD-WAN 智能网关 vCPE 实例

1. 登录天翼云，点击页面右上方“控制中心”，进入控制台。
2. 单击管理控制台左上角的 选择区域。
3. 依次选择“广域云网（CloudWAN）”，单击“天翼云 SD-WAN”，进入天翼云 SD-WAN 总览页面。



4. 单击“智能网关”，在智能网关列表页面，单击“创建智能网关”按钮，进入创建智能网关页面。
5. 根据页面提示配置参数，关键参数信息如下：

参数	描述	样例
实例名称	长度为 1-64 字符，以大小写字母或中文开头，可包含数字、”.”、”_”、”-”	Test-01
组网方案	可选高性能核心网、互联网直连	高性能核心网
区域	所属位置信息	中国内地
基础带宽	带宽大小	5Mbps
是否购	可选是、否	是

参数	描述	样例
买设备		
网关形态	可选择硬件版、软件版	软件版
设备类型	vCPE	vCPE
使用方式	可选择单机、双机备份	单机
购买数量	购买智能网关的个数	1
购买时长	按需购买	1 个月
企业项目	企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理	default

6. 提交订单，确认订单配置信息，勾选协议，确认下单并支付。

步骤二：开通云服务器

使用镜像市场-"标准 VCPE 镜像-CTyunOS2.0.1-X86 64 位"镜像开通云主机。云主机配置要求：配置要求 2c4g 以上，推荐计算型云主机，网卡使用自动分配内网 IPv4 地址，并有用弹性 IP，带宽大小根据需要设置。

1. 登录天翼云，点击页面右上方“控制中心”，进入控制台。
2. 单击管理控制台左上角的 选择区域。
3. 点击计算-弹性云主机进入创建云主机页面。
4. 单击右上角创建云主机进入云主机购买页面。

5. 进行基础配置。

a. 根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、“实例名称”、“主机名称”等。

b. 选择规格。此处选择"CPU 架构"为"X86 计算"、分类可以选 "通用型"或"计算型"，推荐选用计算型。规格为 2c4G 以上。

c. 选择镜像。“镜像类型”选择“云镜像市场”，在云镜像市场中选择预置了 vCPE 的"标准 VCPE 镜像-CTyunOS2.0.1-X86 64 位"镜像。

d. 设置云盘类型和大小。此处选择"系统盘"为"通用型 SSD"，40GB。

6. 网络配置

设置网络，此处"网卡"选择“单网卡”，设置“自动分配内网地址”；"安全组"根据客户需求自行配置； "弹性 IP"用于连通 SDWAN POP 接入点，带宽根据客户网络需求进行配置。

7. 高级配置

设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。用户数据需配置为 vCPE 实例的 SN 号。vCPE SN 号信息可通过“天翼云 SD-WAN 控制台->目标智能网关详情页->基本信息模块”查看，操作步骤可参考：查看基本信息。

用户数据填写格式如下：

其中，CTCEXXXXXXXXXSN 根据客户订购的 vCPE 实例 SN 号进行替换。

```
#!/bin/bash

set_sn  CTCEXXXXXXXXXSN
```

8. 确认配置信息并支付。

步骤三：激活 vCPE 实例

1. 登录控制中心。

2. 在控制中心页面左上角点击 ，选择区域。

3. 依次选择“广域云网（CloudWAN）”，单击“天翼云 SD-WAN”，进入天翼云 SD-WAN 总览页面。

4. 选择“智能网关”，单击目标智能网关 vCPE 实例“操作”列的“更多-激活”。

5. 根据页面提示，配置私网网络地址。

6. 单击“确定”，开始激活操作。
7. 激活成功后，业务状态变为“已激活”。

步骤四：绑定 SD-WAN 网络实例

1. 登录天翼云 SD-WAN 控制台。
2. 选择“智能网关”，单击目标智能网关实例“操作”列的“绑定网络实例”。
3. 根据页面提示，选择要加入的天翼云 SD-WAN 实例。
4. 单击“确定”，完成智能网关实例绑定天翼云 SD-WAN 实例。

步骤五：创建互联

1. 登录天翼云 SD-WAN 控制台。
2. 选择“天翼云 SD-WAN”，单击上述已经绑定的目标 SD-WAN 实例名称。
3. 进入“设备互联”页签，单击“添加设备互联”。
4. 根据页面提示，选择要建立互联关系的两个智能网关实例。
5. 单击“确定”，完成在两个智能网关设备之间建联。

互联链路建立成功后，您可通过对端智能网关实例的下挂设备访问 vCPE 所在 VPC 下的云主机，测试是否连通。

10.35 通过配置监控 Agent 实现指定进程监控

目前，标准监控 Agent 可提供云主机内运行的全部进程数、全部进程的 CPU 使用率等指标的监控服务。若您希望对特定关键字的进程进行监控和统计，可通过修改监控插件配置的方式实现云主机内指定进程数量、CPU 使用率、内存使用率、打开文件数指标的统计。

前提条件

请确保您的云主机已安装最新监控 Agent，具体监控 Agent 支持的地域、说明和安装方法请参考安装监控 Agent。

背景信息

监控 Agent 实现指定进程监控的限制如下：

1. 目前只有安装的为 Linux 操作系统的支持通过自定义配置实现指定进程监控。
2. 一台云主机最大支持同时监控 10 个指定进程。
3. 云主机监控 Agent 配置更新后，约 3-5 分钟后可在控制台查看指定进程的相关指标。

操作步骤

1. 修改监控 Agent 配置文件：

1）以 root 用户登录需要配置的云主机。

2）找到配置文件：

```
vim /etc/telegraf/telegraf.conf
```

3）将 exe="telegraf"修改为指定的进程名，如需监控 java 进程，则修改如下：

```
[[inputs.procstat]]
  exe="java"
  fieldpass = [
    "pid_count",
    "cpu_usage",
    "memory_usage",
    "num_fds"
  ]
```

4）若需要监控多个进程，如需要同时监控 java 和 int 进程，则修改如下：

```
[[inputs.procstat]]
  exe="java"
```

```
  fieldpass = [
    "pid_count",
    "cpu_usage",
    "memory_usage",
    "num_fds"
  ]
[[inputs.procstat]]
  exe="int"
  fieldpass = [
    "pid_count",
    "cpu_usage",
    "memory_usage",
    "num_fds"
  ]
```

5）修改完保存配置文件，并重启监控 Agent。

```
systemctl restart telegraf
```

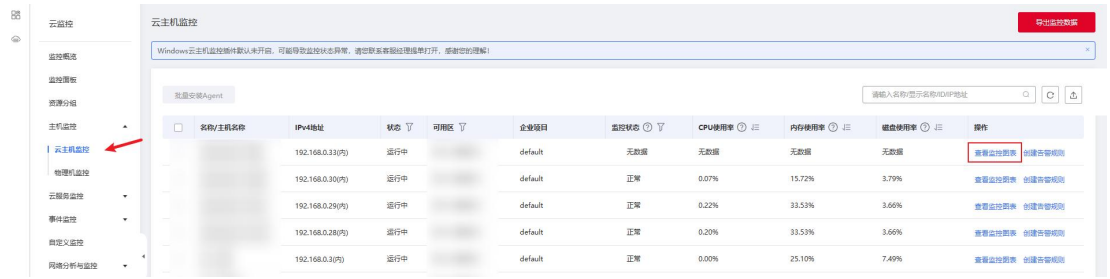
2. 创建指定进程监控任务

1）登录控制中心。

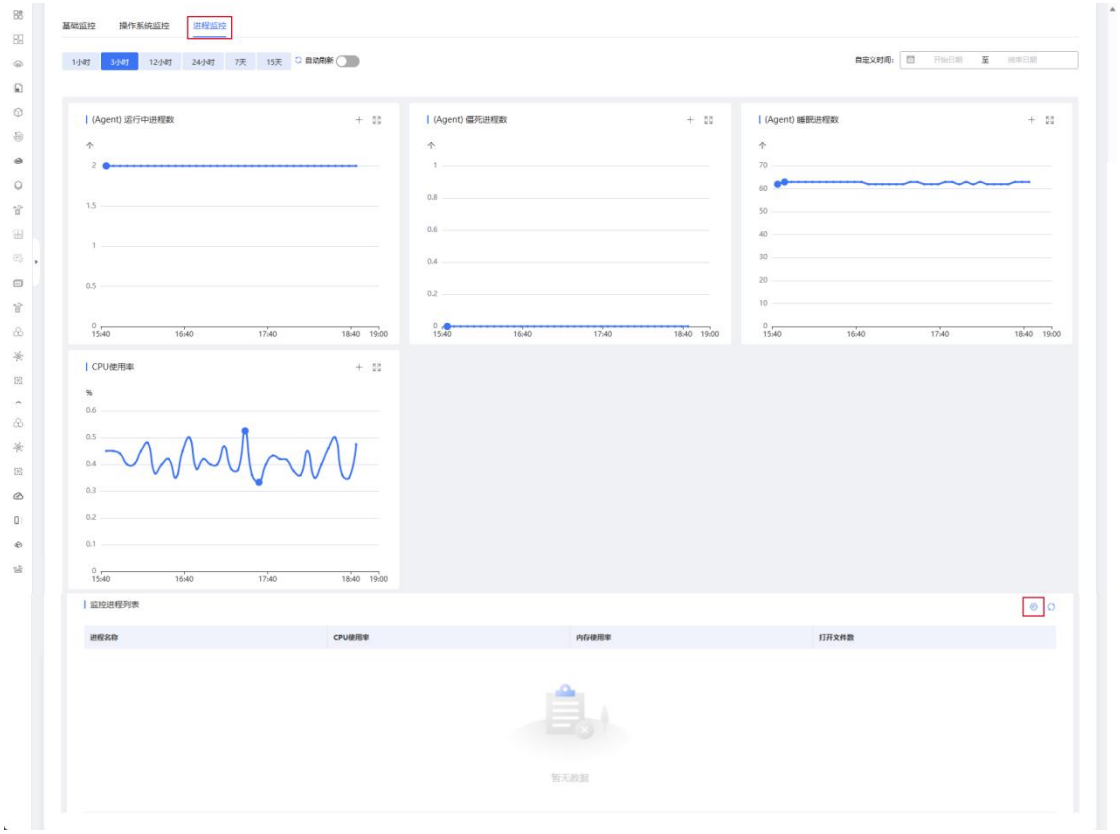
2）选择即将创建云主机所在的地域。

3）单击“管理与部署>云监控”进入云监控控制台。

4）在左侧导航栏选择“主机监控>云主机监控”，选择需要配置指定进程监控的云主机，查看监控图表。



5）在“进程监控”tab 页，点击“设置”按钮。



6）在配置监控进程中，可数据指定进程的关键字。

注意

注意：指定进程关键字需与步骤 1 中配置的关键字一致，才可以正确采集数据。



说明

进程名称关键字仅支持精确匹配。

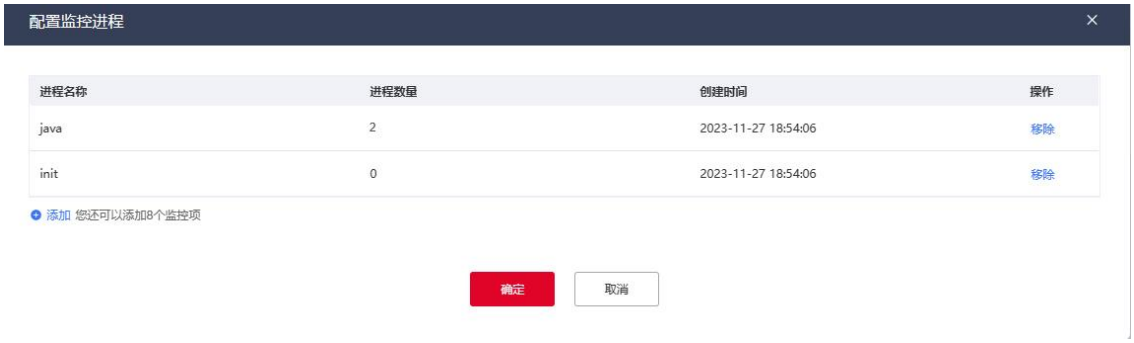
一台云主机最大可配置 10 个进程关键字进行采集。

3. 查看指定进程监控数据

1）在“进程监控”tab 页，可在监控进程列表中查看指定进程的 CPU 使用率、内存使用率、打开文件数的监控数据。



2）在“进程监控”tab 页，点击“设置”按钮，可在弹窗中查看指定进程的数量。



10.36 部署 Palworld 幻兽帕鲁服务器最佳实践

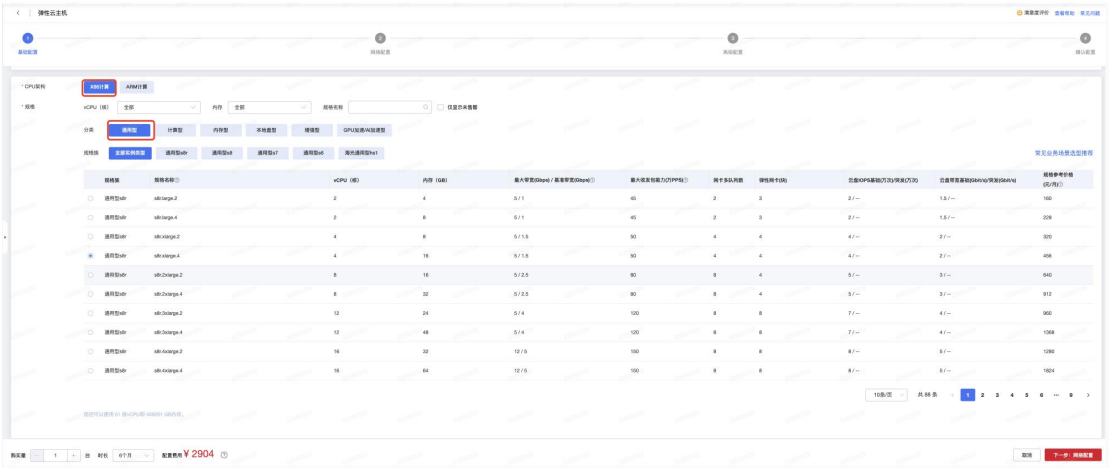
10.36.1 部署 Palworld 幻兽帕鲁服务器最佳实践（Ubuntu）

1.登录云主机控制台，选择创建云主机的资源池，点击“创建云主机”按钮。



2.基础配置。

CPU 架构选择“X86 计算”，规格-分类选择“通用型”。为确保游戏顺畅运行，建议您选择 4C16G 及以上的规格。



镜像类型选择“应用镜像”，镜像下拉菜单中选择“ubuntu”-“幻兽帕鲁（Palworld）-Ubuntu”。



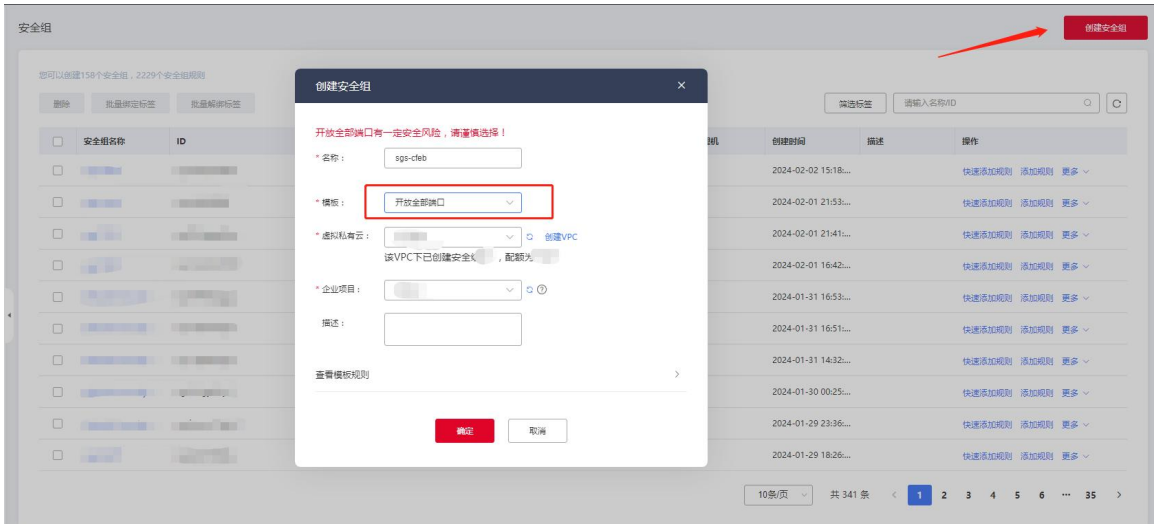
点击“下一步：网络配置”。

3.网络配置。

点击“创建安全组”按钮，跳转至网络控制台-安全组页面。

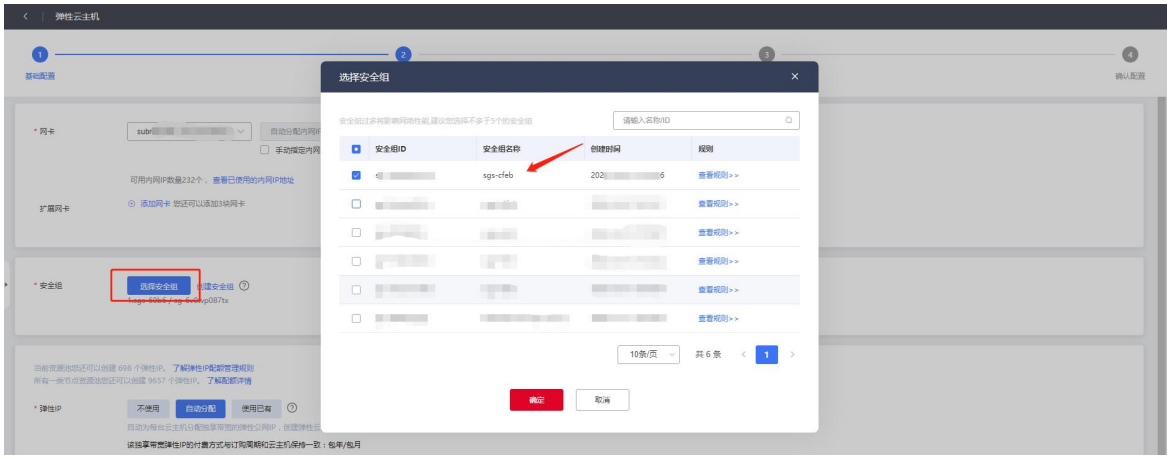


点击“创建安全组”按钮，在弹窗中等待模板下拉菜单选择“开放全部端口”。



点击“确定”，完成安全组创建。

返回云主机订购页面，点击“选择安全组”按钮，在安全组列表中勾选刚才创建的安全组。



点击“确定”，完成安全组选择。

弹性 IP 选择“自动分配”，根据需要选择带宽大小。为确保游戏顺畅运行，建议带宽选择 10M 以上。



点击“下一步：高级配置”。

4.根据提示完成高级配置，点击“下一步：确认配置”。

确认配置无误后，点击“立即购买”。

5.完成支付后，返回云主机控制台，确认云主机状态为“运行中”，则可以进行后续操作。



6.添加安全组规则

在云主机列表中点击云主机名称，进入云主机详情页。点击安全组页签，点击“添加规则”按钮，新增 IP 放行。



在添加规则弹窗中：

1）协议选择“UDP”。

2）端口范围填写“8211”。

端口也可自定义进行修改，修改后可提高服务器的安全性：

登录云主机后，首先使用密码登录 root 用户，执行以下命令进入编辑配置界面。

```
vim /etc/rc.local
```

按 i 进入编辑状态，将第二行改为：

```
su - steam -c "cd
/home/steam/.steam/SteamApps/common/PalServer;steamcmd +login
anonymous +app_update 2394010 validate +quit;nohup ./PalServer.sh
-port=xxx &"
```

其中 xxx 为 1-65535 中的任意希望启用的端口，修改完成后按 esc 退出编辑，再输入:wq 进行保存并退出。

3）地址处将“1.1.1.1”更改为自己的 IP 地址。



点击“确定”，完成安全组规则修改。

7.进入游戏。

先在云主机控制台-云主机列表中找到云主机的公网 IP 地址。



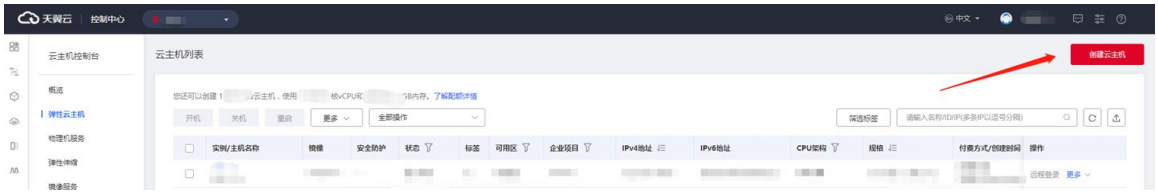
启动游戏，在游戏界面选择“加入多人游戏（专用服务器）”。在地址输入框输入“公网 IP:8211”，点击“联系”即可进入帕鲁世界。





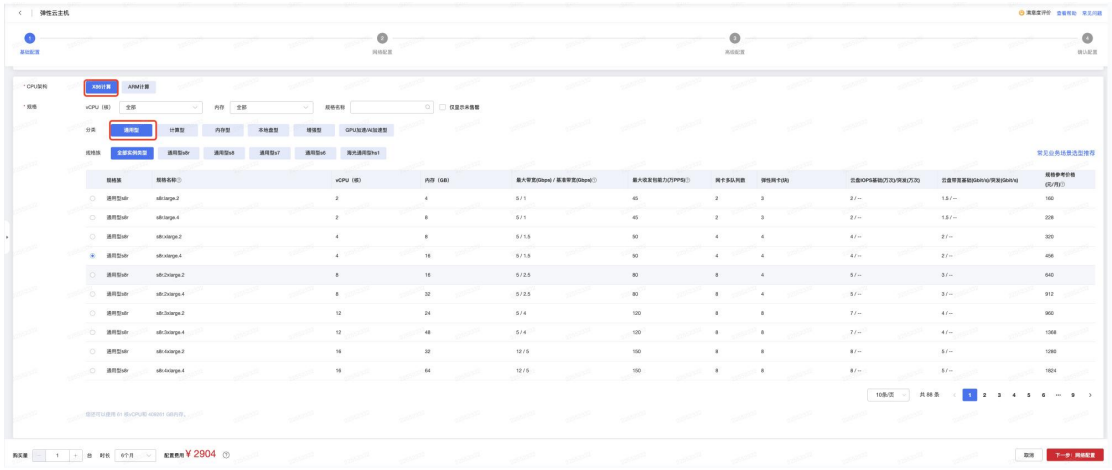
10.36.2 部署 Palworld 幻兽帕鲁服务器最佳实践（Windows）

1.登录云主机控制台，选择创建云主机的资源池，点击“创建云主机”按钮。



2.基础配置。

CPU 架构选择“X86 计算”，规格-分类选择“通用型”。为确保游戏顺畅运行，建议您选择 4C16G 及以上的规格。



镜像类型选择“应用镜像”，镜像下拉菜单中选择“windows”-“幻兽帕鲁（Palworld）”。



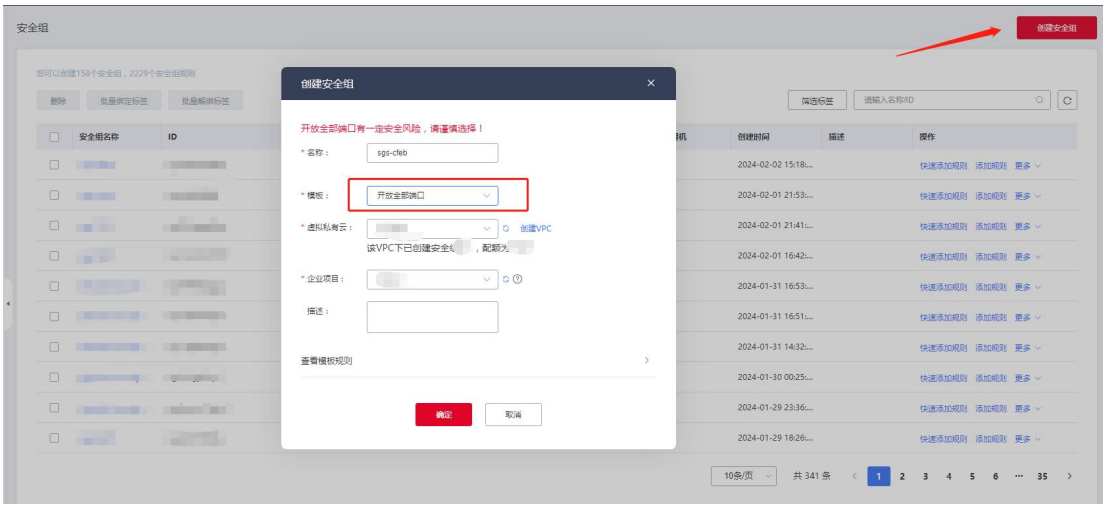
点击“下一步：网络配置”。

3.网络配置。

点击“创建安全组”按钮，跳转至网络控制台-安全组页面。

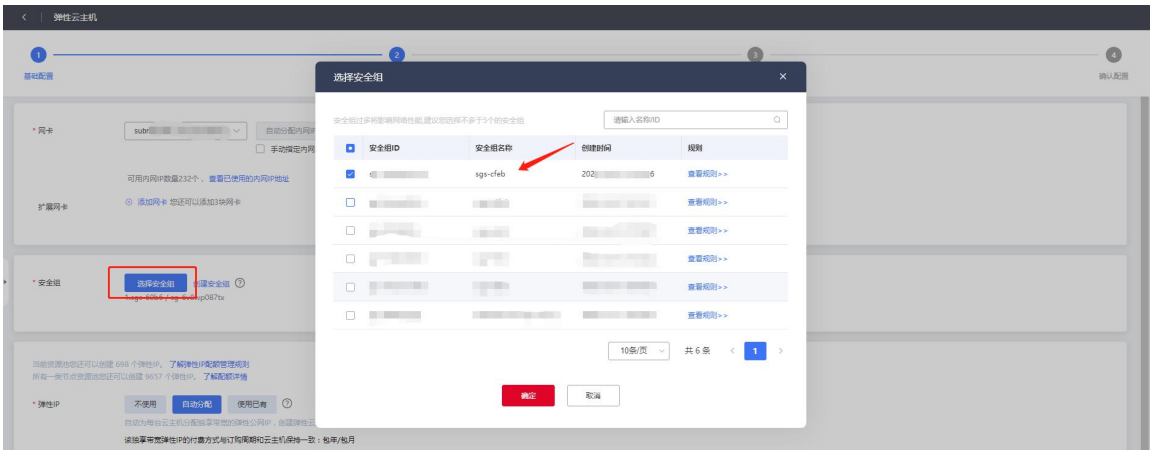


点击“创建安全组”按钮，在弹窗中等待模板下拉菜单选择“开放全部端口”。



点击“确定”，完成安全组创建。

返回云主机订购页面，点击“选择安全组”按钮，在安全组列表中勾选刚才创建的安全组。



点击“确定”，完成安全组选择。

弹性 IP 选择“自动分配”，根据需要选择带宽大小。为确保游戏顺畅运行，建议带宽选择 10M 以上。



点击“下一步：高级配置”。

4. 根据提示完成高级配置，点击“下一步：确认配置”。

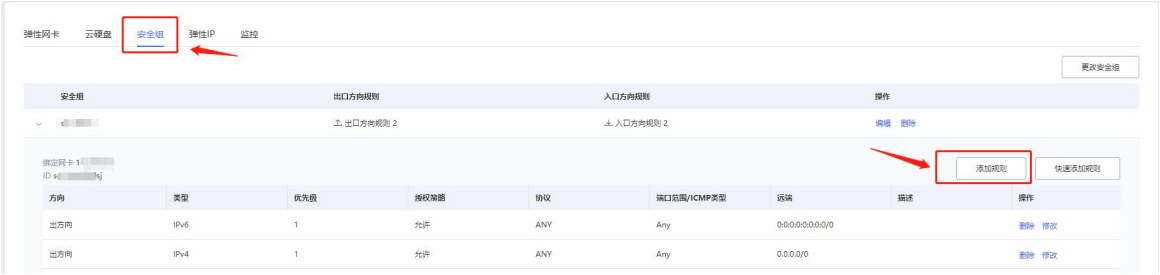
确认配置无误后，点击“立即购买”。

5. 完成支付后，返回云主机控制台，确认云主机状态为“运行中”，则可以进行后续操作。



6. 添加安全组规则。

在云主机列表中点击云主机名称，进入云主机详情页。点击安全组页签，点击“添加规则”按钮，新增 IP 放行。



在添加规则弹窗中：

1）协议选择“UDP”。

2）端口范围填写“8211”。

端口也可自定义进行修改，修改后可提高服务器的安全性：Windows 镜像云主机登录后，将 C:\steamcmd 路径下 start.bat 文件中修改最后一行，改为：

```
start "" "C:\PalServer\PalServer.exe" -port=xxx
```

其中 xxx 为 1-65535 中的任意希望启用的端口，并重新启动服务器即可生效。

3）地址处将“1.1.1.1”更改为自己的 IP 地址。



点击“确定”，完成安全组规则修改。

7. 进入游戏。

先在云主机控制台-云主机列表中找到云主机的公网 IP 地址。



启动游戏，在游戏界面选择“加入多人游戏（专用服务器）”。在地址输入框输入“公网 IP:8211”，点击“联系”即可进入帕鲁世界。





10.36.3Palworld 幻兽帕鲁世界参数修改最佳实践（Ubuntu）

操作场景

本文以 Ubuntu 操作系统为例，为您介绍通过 Palworld 专有镜像一键部署 Palworld 服务器后对游戏世界参数进行修改的具体操作。

操作步骤

1.登陆 root 用户。

在云主机控制台点击云主机列表中“远程登录”按钮，登录云主机。

输入 root 后回车，再输入创建时使用的密码，即可进入 root 用户。

2.执行命令　su - steam。

3.执行命令　cd /home/steam/.steam/SteamApps/common/PalServer，进入该目录下。

4.执行命令　cp DefaultPalWorldSettings.ini Pal/Saved/Config/LinuxServer/PalWorldSettings.ini。

5.使用 vim 编辑器对配置进行编辑。

```
vim Pal/Saved/Config/LinuxServer/PalWorldSettings.ini
```

进入后按 i 进入编辑状态，修改完成后按一下 esc 退出编辑，再输入:wq 进行保存并退出。

5. 输入　reboot 重启服务器，配置即可生效。

6.

10.36.4Palworld 幻兽帕鲁世界参数修改最佳实践（Windows）

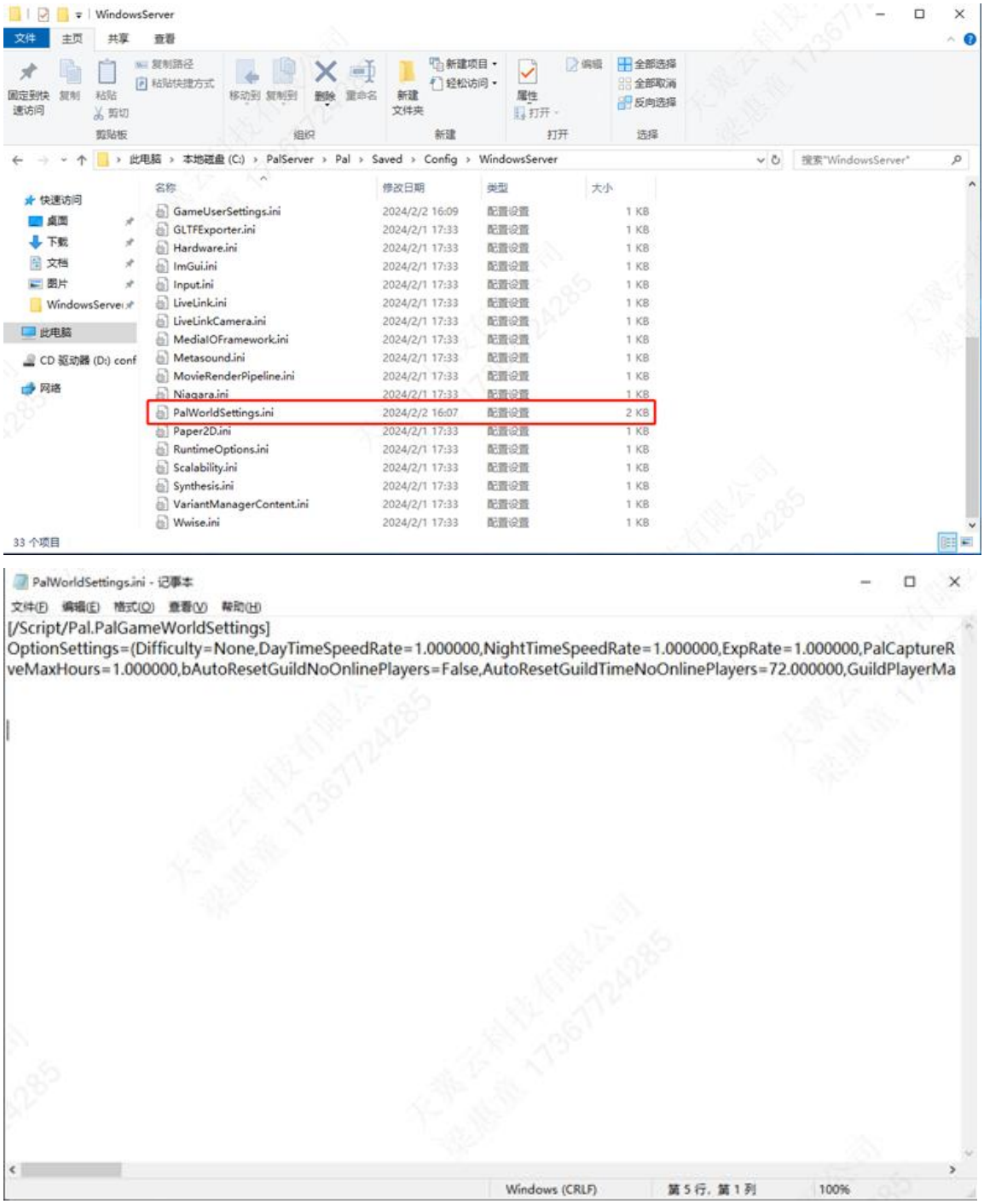
操作场景

本文以 Windows 操作系统为例，为您介绍通过 Palworld 专有镜像一键部署 Palworld 服务器后对游戏世界参数进行修改的具体操作。

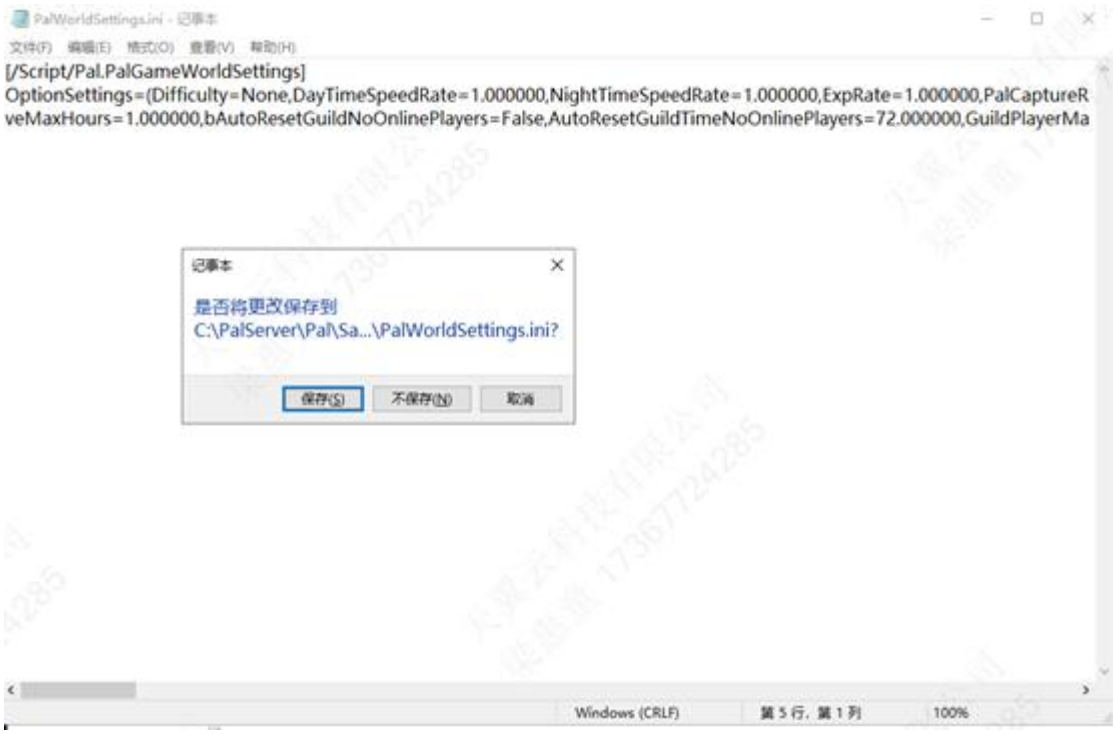
操作步骤

1.进入 C:\PalServer\Pal\Saved\Config\WindowsServer 路径。

2.使用记事本打开 PalWorldSettings.ini 文件。



3. 根据参数对应表修改自己喜欢的世界参数，修改完成后点击“保存”。



4.重启云主机，配置的参数即可生效。

参数对照表

Difficulty	英文	中文
DayTimeSpeedRate	Day time speed	白天时间流逝速度
NightTimeSpeedRate	Night time speed	夜晚时间流逝速度
ExpRate	EXP rate	经验获取倍率
PalCaptureRate	Pal capture rate	帕鲁捕获率

Difficulty	英文	中文
PalSpawnNumRate	Pal appearance rate	帕鲁刷新倍率
PalDamageRateAttack	Damage from pals multipiler	帕鲁攻击力倍率
PalDamageRateDefense	Damage to pals multipiler	帕鲁防御力倍率
PlayerDamageRateAttack	Damage from player multipiler	玩家攻击力倍率
PlayerDamageRateDefense	Damage to player multipiler	玩家防御力倍率
PlayerStomachDecreaceRate	Player hunger depletion rate	玩家饥饿消耗率
PlayerStaminaDecreaceRate	Player stamina reduction rate	玩家体力减少率
PlayerAutoHPRegeneRate	Player auto HP regeneration rate	玩家自动 HP 回复率
PlayerAutoHpRegeneRateInSleep	Player sleep HP regeneration rate	玩家睡眠 HP 回复率

Difficulty	英文	中文
PalStomachDecreaceRate	Pal hunger depletion rate	帕鲁饥饿消耗率
PalStaminaDecreaceRate	Pal stamina reduction rate	帕鲁体力减少率
PalAutoHPRegeneRate	Pal auto HP regeneration rate	帕鲁自动 HP 回复率
PalAutoHpRegeneRateInSleep	Pal sleep health regeneration rate (in Palbox)	帕鲁睡眠 HP 回复率（帕鲁终端中）
BuildObjectDamageRate	Damage to structure multipiler	建筑物血量
BuildObjectDeteriorationDamageRate	Structure determination rate	建筑物风化速率
CollectionDropRate	Getherable items multipiler	采集物掉落率
CollectionObjectHpRate	Getherable objects HP	采集物回复率

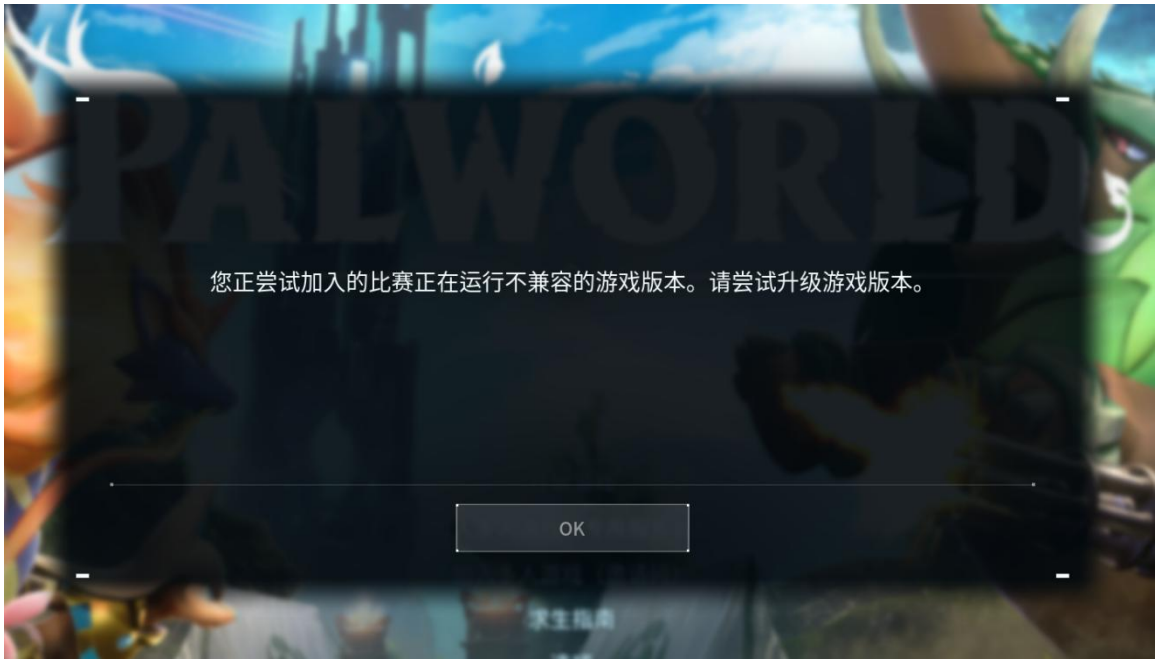
Difficulty	英文	中文
	multipiler	
CollectionObjectRespawnSpeedRate	Getherable objects respawn interval	采集物刷新率
EnemyDropItemRate	Dropped Items Multipiler	击败敌人物品掉落率
DeathPenalty	Death penalty None : No lost, Item : Lost item without equipment, ItemAndEquipment : Lost item and equipment, All : Lost All item, equipment, pal(in inventory)	死亡惩罚 None ：没有丢失；Item ：丢失物品，不丢失帕鲁；ItemAndEquipment ：丢失的物品和装备；All ：丢失所有物品、装备、帕鲁（背包中）
GuildPlayerMaxNum	Max player of Guild	公会最大玩家数
PalEggDefaultHatchingTime	Time(h) to incubate massive	孵化巨大蛋的时间(h)

Difficulty	英文	中文
	egg	
BaseCampWorkerMaxNum	Max worker in camp	每个据点工作帕鲁数量上限
WorkSpeedRate	Rate of working speed	工作效率
ServerPlayerMaxNum	Maximum number of people who can join the server	服务器最多可加入人数
ServerName	Server name	服务器名称
ServerDescription	Server description	服务器描述
AdminPassword	AdminPassword	管理员密码
ServerPassword	Set the server password.	设置服务器密码(仅作为社群服务器时有效)
PublicPort	Public port number	公共端口号

Difficulty	英文	中文
PublicIP	Public IP	公共 IP
RCONEnabled	Enable RCON	启用 RCON
RCONPort	Port number for RCON	RCON 的端口号

10.36.5Palworld 幻兽帕鲁更新游戏最佳实践

如果您进入游戏时遇到如图报错，请参考本文对游戏进行更新。



Ubuntu

启动云主机时可以触发后台的自动检查游戏版本，如需更新游戏，在云主机控制台点击“重启”即可。



Windows

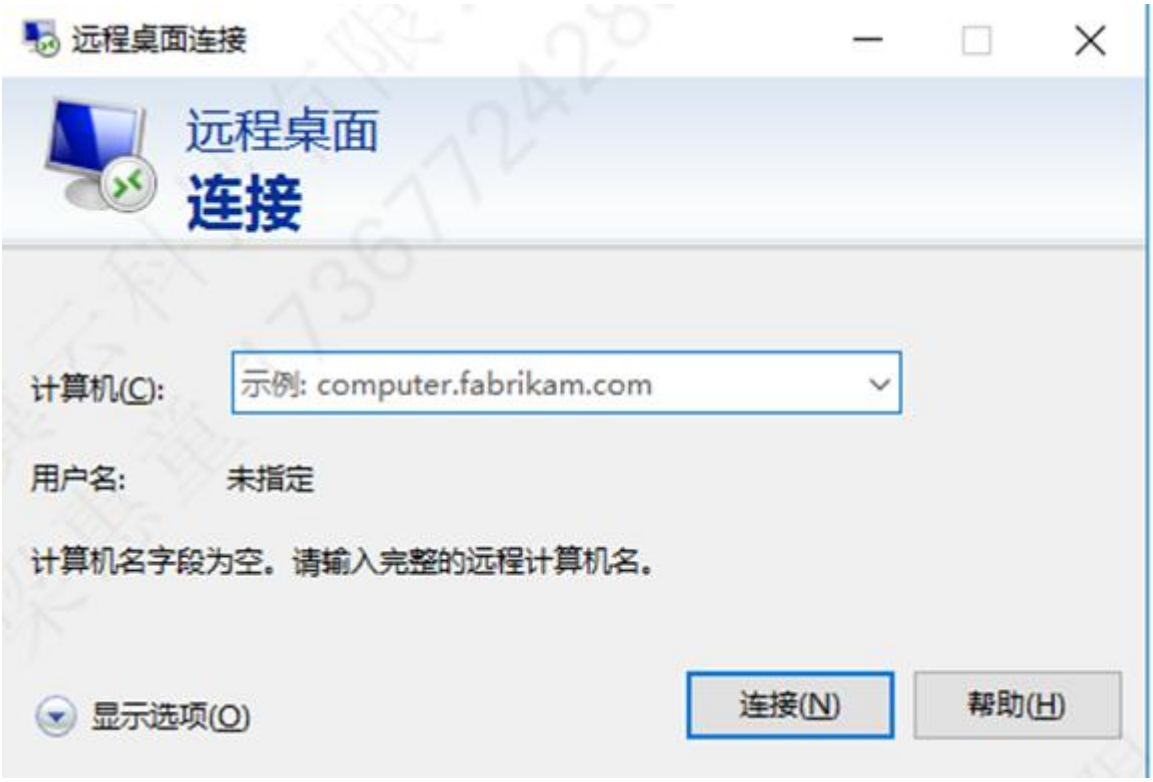
自动更新

启动云主机时可以触发后台的自动检查游戏版本，如需更新游戏，在云主机控制台点击“重启”即可。



手动更新

1. 在本地 Windows 电脑搜索远程桌面连接，并输入云主机的公网 IP 地址。



2.使用设置的密码进行登录，用户为 Administrator，密码输入设置的云主机密码并连接。

3.打开任务管理器，找到 Pal 进程，右键点击并选择结束任务。

9.35.6 已有天翼云 Ubuntu 云主机，搭建幻兽帕鲁服务器

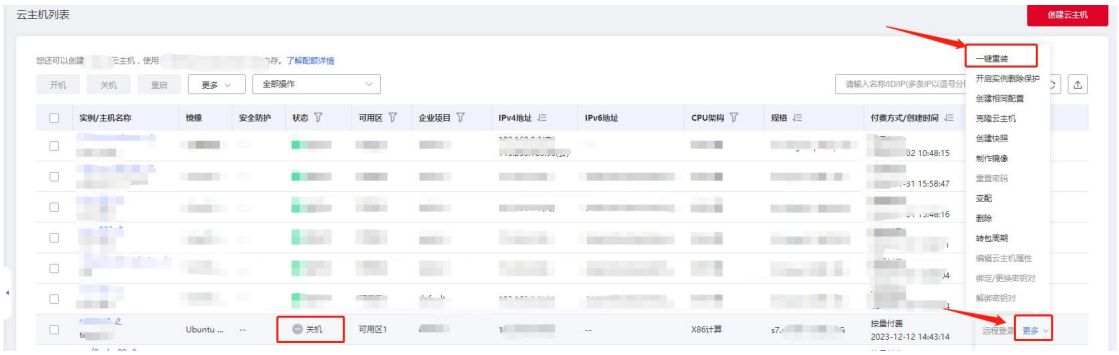
操作前提

- 如果您已购买的云主机位于【青岛 20、西南 1】资源池（其余资源池镜像正在加载中），且满足以下条件，请参考本文部署幻兽帕鲁服务器。
  - 操作系统：Ubuntu 20.04
  - 配置：4 核 16G 及以上
  - 已开通公网 IP
  - 云主机未到期

操作步骤

1. 登录云主机控制台，前往您已购买的云主机所在的资源池。确认云主机状态为“已关机”，则可以进行重装操作系统操作。

点击“更多”按钮，点击下拉菜单中“一键重装”按钮。



2. 在重装操作系统弹窗中，镜像类型选择“应用镜像”，并在重装选择下拉菜单中选择“ubuntu”、“幻兽帕鲁（Palworld）”。



设置并确认密码，点击“确定”按钮，完成重装操作系统操作。

3. 返回云主机控制台，选中已完成重装操作系统的云主机，点击“开机”按钮。确认云主机状态为“运行中”，则可以进行后续操作。

4. 添加安全组规则。

在云主机列表中点击云主机名称，进入云主机详情页。点击安全组页签，点击“添加规则”按钮，新增 IP 放行。



在添加规则弹窗中：

1）协议选择“UDP”。

2）端口范围填写“8211”。

端口也可自定义进行修改，修改后可提高服务器的安全性：

登录云主机后，首先使用密码登录 root 用户，执行以下命令进入编辑配置界面。

```
vim /etc/rc.local
```

按 i 进入编辑状态，将第二行改为：

```
su - steam -c "cd
/home/steam/.steam/SteamApps/common/PalServer;steamcmd +login
anonymous +app_update 2394010 validate +quit;nohup ./PalServer.sh
-port=xxx &"
```

其中 xxx 为 1-65535 中的任意希望启用的端口，修改完成后按 esc 退出编辑，再输入:wq 进行保存并退出。

3）地址处将“1.1.1.1”更改为自己的 IP 地址。



点击“确定”，完成安全组规则修改。

4. 登录游戏。

在云主机控制台找到云主机的公网 IP 地址。



启动游戏，在游戏界面选择“加入多人游戏（专用服务器）”。在地址输入框输入该地址，点击“联系”即可进入帕鲁世界（“:8211”不用进行修改）。





10.36.6 已有天翼云 Ubuntu 云主机，搭建幻兽帕鲁服务器

操作前提

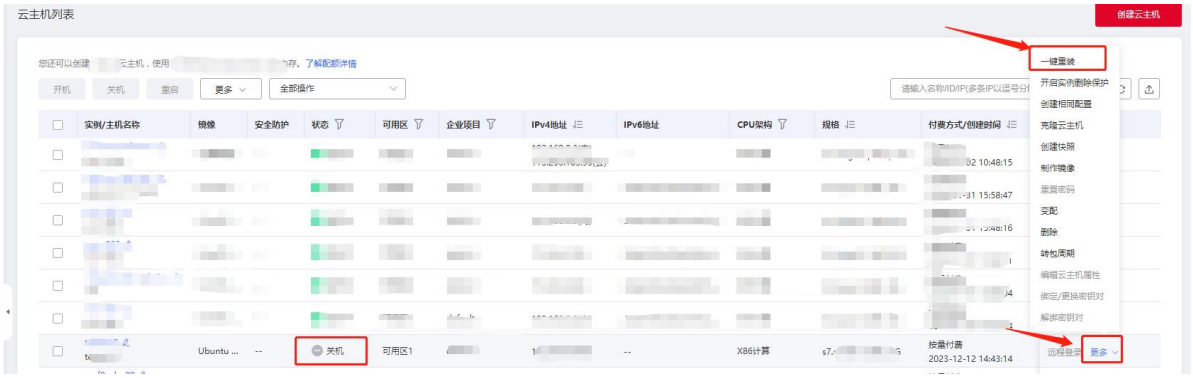
- 如果您已购买的云主机位于【青岛 20、西南 1】资源池（其余资源池镜像正在加载中），且满足以下条件，请参考本文部署幻兽帕鲁服务器。

- 操作系统：Ubuntu 20.04
- 配置：4 核 16G 及以上
- 已开通公网 IP
- 云主机未到期

操作步骤

1.登录云主机控制台，前往您已购买的云主机所在的资源池。确认云主机状态为“已关机”，则可以进行重装操作系统操作。

点击“更多”按钮，点击下拉菜单中“一键重装”按钮。



2.在重装操作系统弹窗中，镜像类型选择“应用镜像”，并在重装选择下拉菜单中选择“ubuntu”、“幻兽帕鲁（Palworld）”。



设置并确认密码，点击“确定”按钮，完成重装操作系统操作。

3.返回云主机控制台，选中已完成重装操作系统的云主机，点击“开机”按钮。确认云主机状态为“运行中”，则可以进行后续操作。

4.添加安全组规则。

在云主机列表中点击云主机名称，进入云主机详情页。点击安全组页签，点击“添加规则”按钮，新增 IP 放行。



在添加规则弹窗中：

1）协议选择“UDP”。

2）端口范围填写“8211”。

端口也可自定义进行修改，修改后可提高服务器的安全性：

登录云主机后，首先使用密码登录 root 用户，执行以下命令进入编辑配置界面。

```
vim /etc/rc.local
```

按 i 进入编辑状态，将第二行改为：

```
su - steam -c "cd /home/steam/.steam/SteamApps/common/PalServer;steamcmd +login anonymous +app_update 2394010 validate +quit;nohup ./PalServer.sh -port=xxx &"
```

其中 xxx 为 1-65535 中的任意希望启用的端口，修改完成后按 esc 退出编辑，再输入:wq 进行保存并退出。

3）地址处将“1.1.1.1”更改为自己的 IP 地址。



点击“确定”，完成安全组规则修改。

5. 登录游戏。

在云主机控制台找到云主机的公网 IP 地址。



启动游戏，在游戏界面选择“加入多人游戏（专用服务器）”。在地址输入框输入该地址，点击“联系”即可进入帕鲁世界（“:8211”不用进行修改）。





10.36.7 已有天翼云 Windows 云主机，搭建幻兽帕鲁服务器

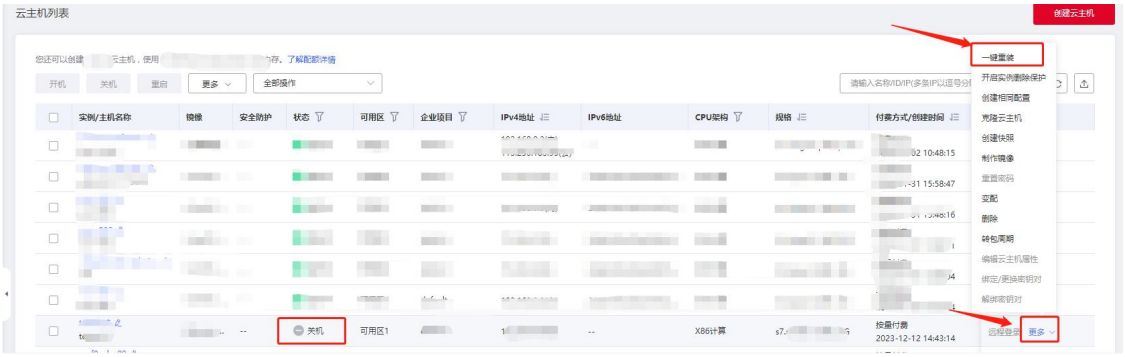
操作前提

- 如果您已购买的云主机位于【青岛 20、西南 1】资源池（其余资源池镜像正在加载中），且满足以下条件，请参考本文部署幻兽帕鲁服务器。
  - 操作系统：Windows
  - 配置：4 核 16G 及以上
  - 已开通公网 IP
  - 云主机未到期

操作步骤

1. 登录云主机控制台，前往您已购买的云主机所在的资源池。确认云主机状态为“已关机”，则可以进行重装操作系统操作。

点击“更多”按钮，点击下拉菜单中“一键重装”按钮。



2. 在重装操作系统弹窗中，镜像类型选择“应用镜像”，并在重装选择下拉菜单中选择“windows”、“幻兽帕鲁（Palworld）”。



设置并确认密码，点击“确定”按钮，完成重装操作系统操作。

3. 返回云主机控制台，选中已完成重装操作系统的云主机，点击“开机”按钮。确认云主机状态为“运行中”，则可以进行后续操作。

4. 添加安全组规则。

在云主机列表中点击云主机名称，进入云主机详情页。点击安全组页签，点击“添加规则”按钮，新增 IP 放行。



在添加规则弹窗中：

1）协议选择“UDP”。

2）端口范围填写“8211”。

端口也可自定义进行修改，修改后可提高服务器的安全性：Windows 镜像云主机登录后，将 C:\steamcmd 路径下 start.bat 文件中修改最后一行，改为：

```
start "" "C:\PalServer\PalServer.exe" -port=xxx
```

其中 xxx 为 1-65535 中的任意希望启用的端口，并重新启动服务器即可生效。

3）地址处将“1.1.1.1”更改为自己的 IP 地址。



点击“确定”，完成安全组规则修改。

5. 登录游戏。

在云主机控制台找到云主机的公网 IP 地址。



启动游戏，在游戏界面选择“加入多人游戏（专用服务器）”。在地址输入框输入该地址，点击“联系”即可进入帕鲁世界（“:8211”不用进行修改）。





10.37 大模型学习机最佳实践

10.37.1 一键部署与登录 Llama 3/Llama 2 大模型学习机

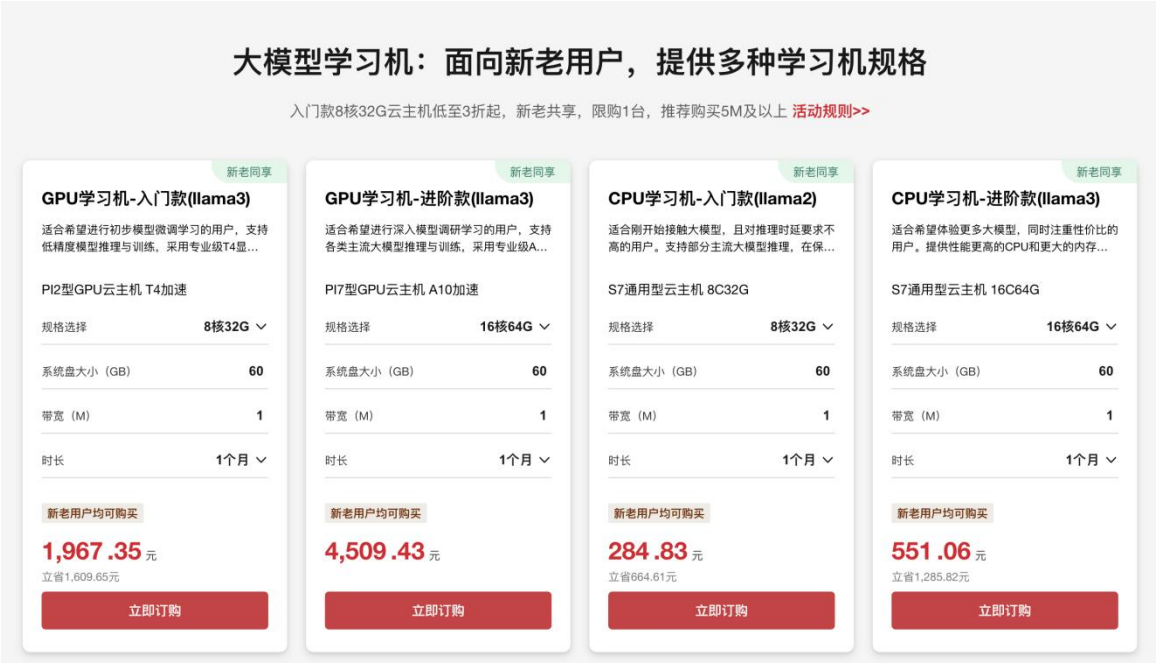
按需购买学习机

通过大模型学习机专属活动优惠页一键购买开通大模型学习机。





点击“新老同享优惠”，选择您要购买的学习机规格，点击“立即订购”。



您可以根据您的实际需求，调整学习机的系统盘大小、带宽大小以及购买时长。



再次点击“立即订购”，支付成功后返回云主机控制台，选择订购云主机所在的资源池。云主机状态变更为“运行中”后即为开通成功，可以进行后续操作。



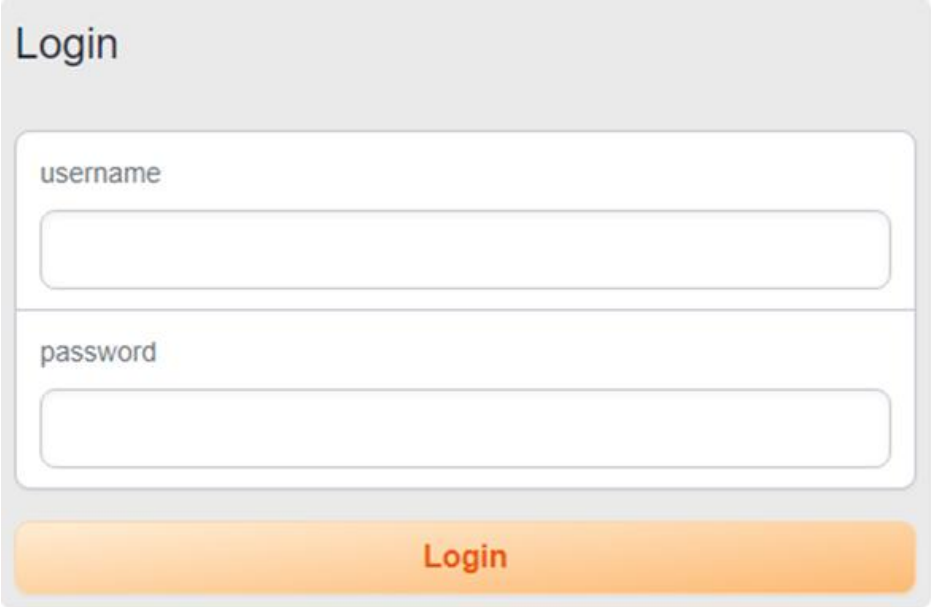
登录大模型

登录前准备：在开通云主机列表的 ipv4 地址处，复制列表中的公共弹性 IP。



（1）登录文本对话大模型（LLaMA2/LLaMA3）

将刚刚复制到的弹性 IP 替换到http://弹性 IP:7861/ 中并在浏览器中输入跳转至登录页面。



您可以任选以下两个账号其一进行登录：

注意

建议您登录后尽快修改密码，修改密码方法在文档“修改大模型学习机登录密码”可查看。

账号一：

username：user

password：ICFLvkJ69I

账户二：

username：user2

password：C7O0hgW5fxkN

（2）登录 AI 图片生成大模型（StableDiffusion）

将刚刚复制到的弹性 IP 替换到http://弹性 IP:7860/ 中并在浏览器中输入跳转至于登录页面。

您可以任选以下两个账号其一进行登录：

注意
建议您登录后尽快修改密码，修改密码方法在文档“修改大模型学习机登录密码”可查看。

账户一：

username：user

password：ICFLvkJ69I

账户二：

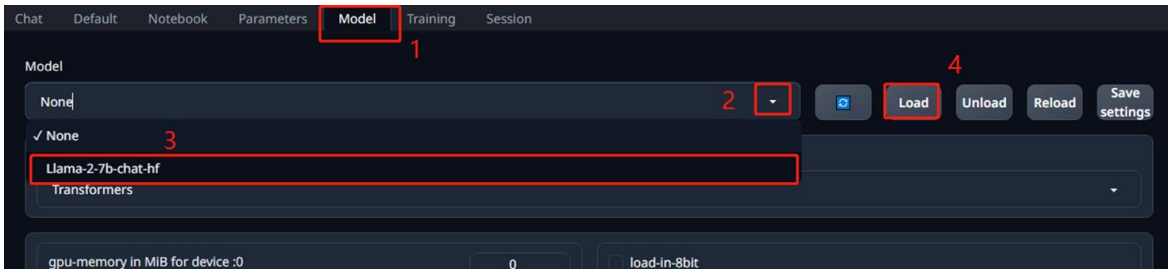
username：user2

password：C7O0hgW5fxkN

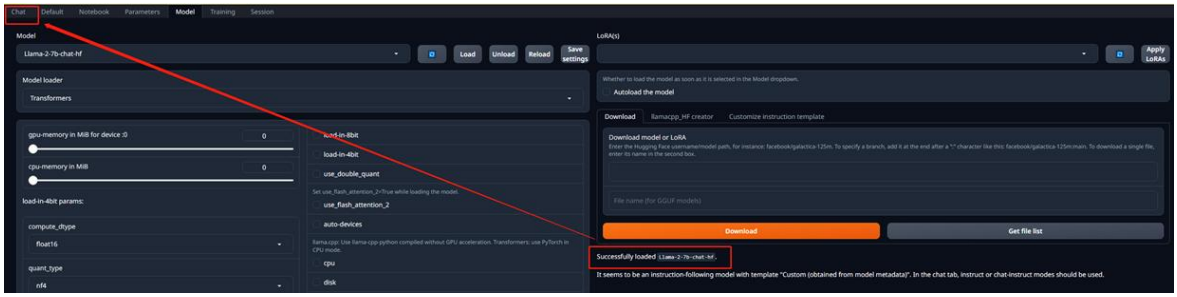
使用大模型进行推理计算

（1）文本对话大模型（LLaMA2/LLaMA3）

首先需要在 Model 标签页点击 LLaMA2/LLaMA3 大模型，然后点击 Load 按钮进行加载。



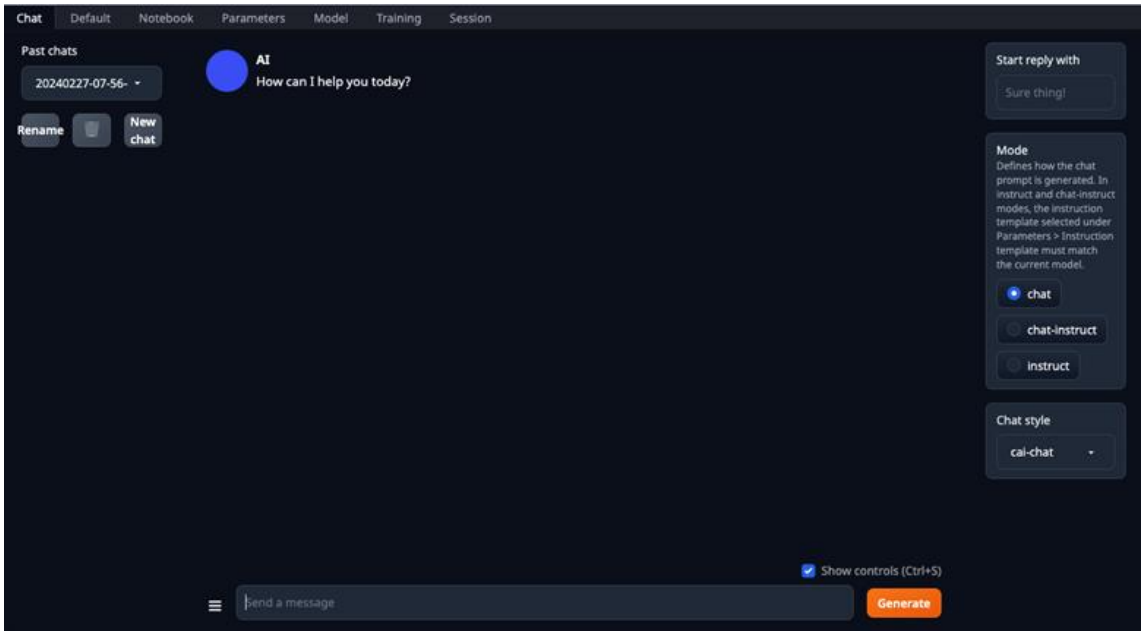
稍等一会，提示“Successfully loaded”，即可点击左上角的“chat”按钮，体验使用文本对话大模型（LLaMA2/LLaMA3）啦！



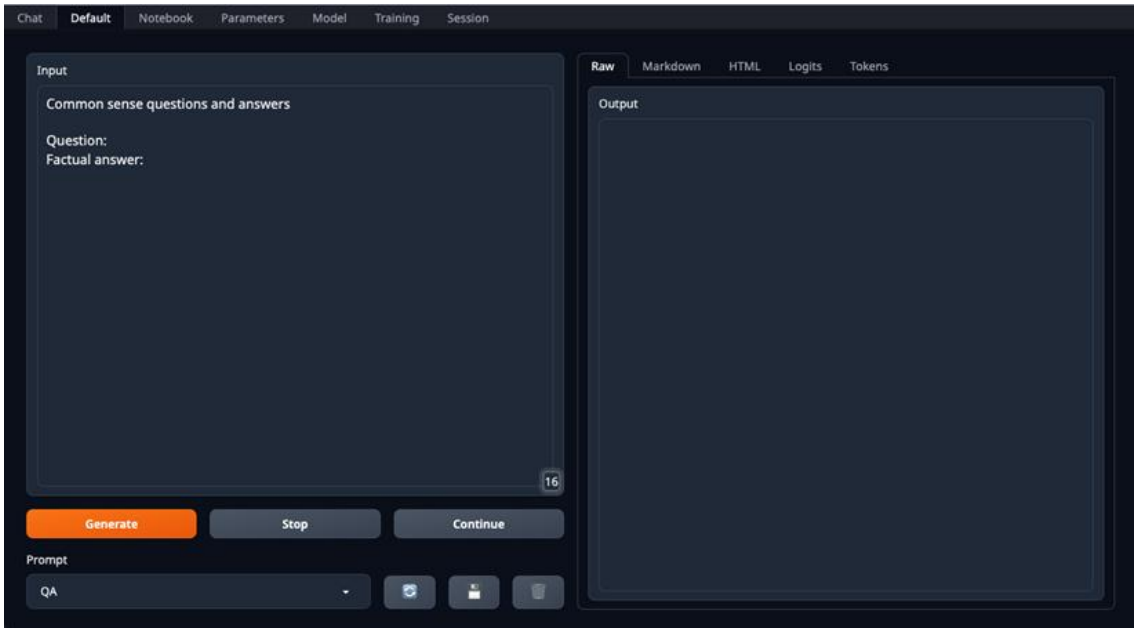
注意

在加载 Llama 2 或者 LLaMA3 模型时， Pi2 和 s7.2xlarge.4 规格使用默认配置会因为显存/内存不足加载失败， 需要通过勾选 load-in-8bit 选项降低模型精度或关闭 StableDiffusion 服务后再加载 Llama 2/LLaMA3 模型. 参考文档: 大模型学习机服务启停。

使用文本多轮对话页面:

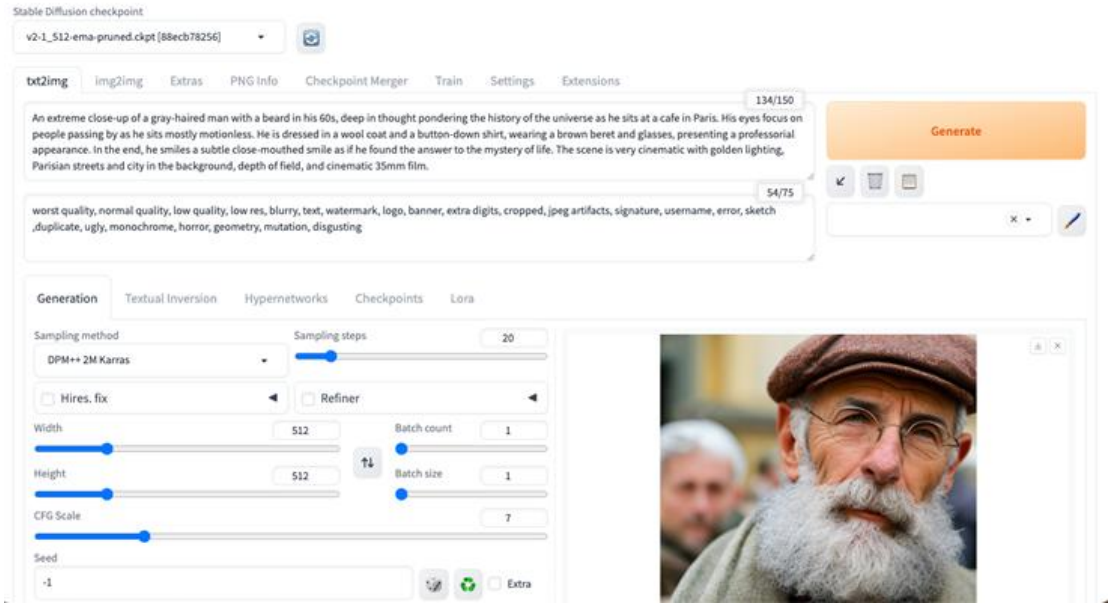


文章生成页面:



（2）AI 图片生成大模型（stable diffusion）：

AI 图片生成大模型登录后可以直接使用，以下为图片生成页面。



10.37.2 部署与登录 Llama 3/Llama 2 大模型学习机

购买云主机

登录云主机控制台，选择创建云主机资源池，点击“创建云主机”按钮。



您可在以下资源池进行不同规格的开通：

注意

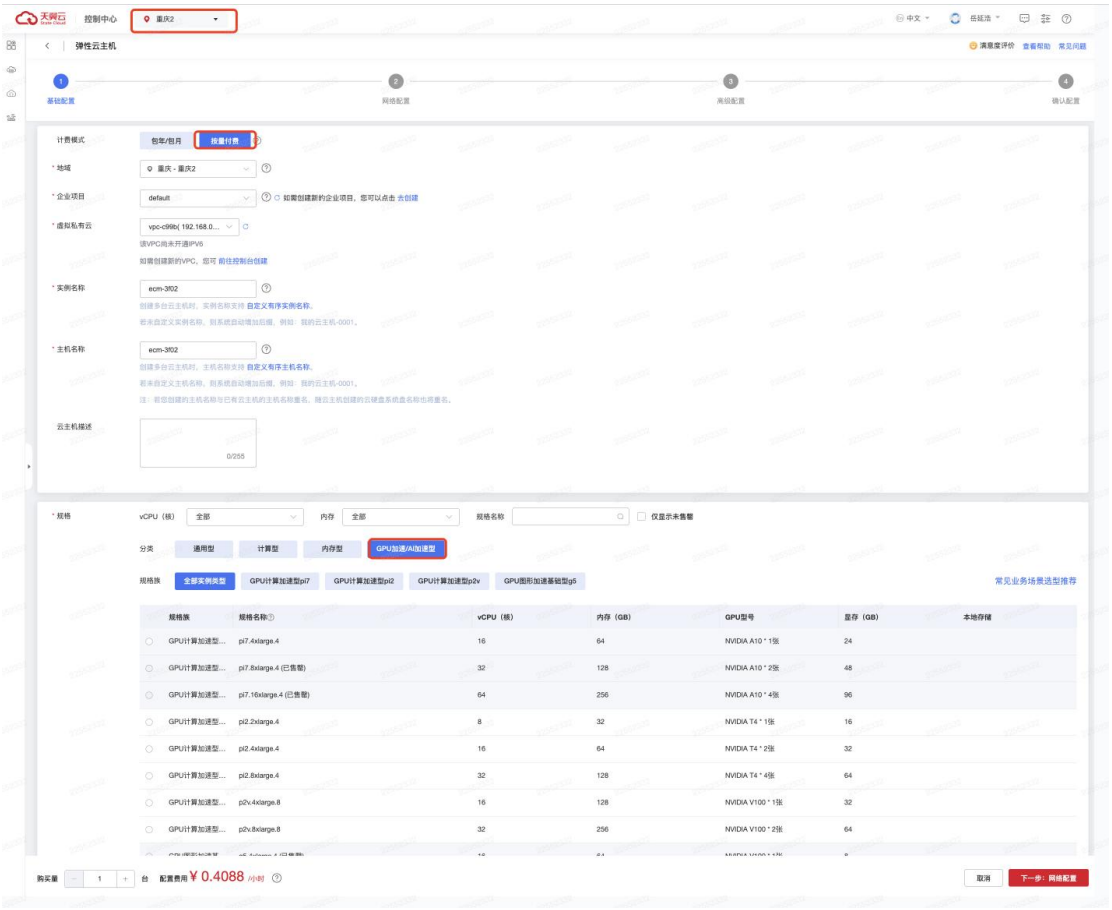
在加载 Llama 2 或者 LLaMA3 模型时， Pi2 和 s7.2xlarge.4 规格使用默认配置会因为显存/内存不足加载失败， 需要通过勾选 load-in-8bit 选项降低模型精度或关闭 StableDiffusion 服务后再加载 Llama 2/LLaMA3 模型. 参考文档:大模型学习机服务启停。

云服务器规格族	云服务器规格	支持大模型镜像资源池	镜像
GPU 计算加速型	PI7	重庆-重庆 2、宁夏-中卫 5、辽宁-辽阳 1	LLaMA2-StableDffusion-WebUI-GPU LLaMA3-StableDffusion-WebUI-GPU
GPU 计算加速型	PI2	重庆-重庆 2、上海-上海 7	LLaMA2-StableDffusion-WebUI-GPU LLaMA3-StableDffusion-WebUI-GPU
通用型	S7(64G 以上规格)	西南 1-az1/az2	LLaMA2-StableDffusion-WebUI-CPU LLaMA3-StableDffusion-WebUI-CPU

基础配置

根据您的需要选择计费方式：包年/包月或按量付费。

规格分类处选择“GPU 加速/AI 加速型”，在下方的规格列表的 pi2 和 pi7 中选择需要的规格。



选择镜像“ubuntu-LLaMA2-StableDiffusion-WebUI-GPU（预装大模型环境）(60GB)”或“ubuntu-LLaMA3-StableDiffusion-WebUI-GPU（预装大模型环境）(60GB)，点击“下一步：网络配置”。



网络配置

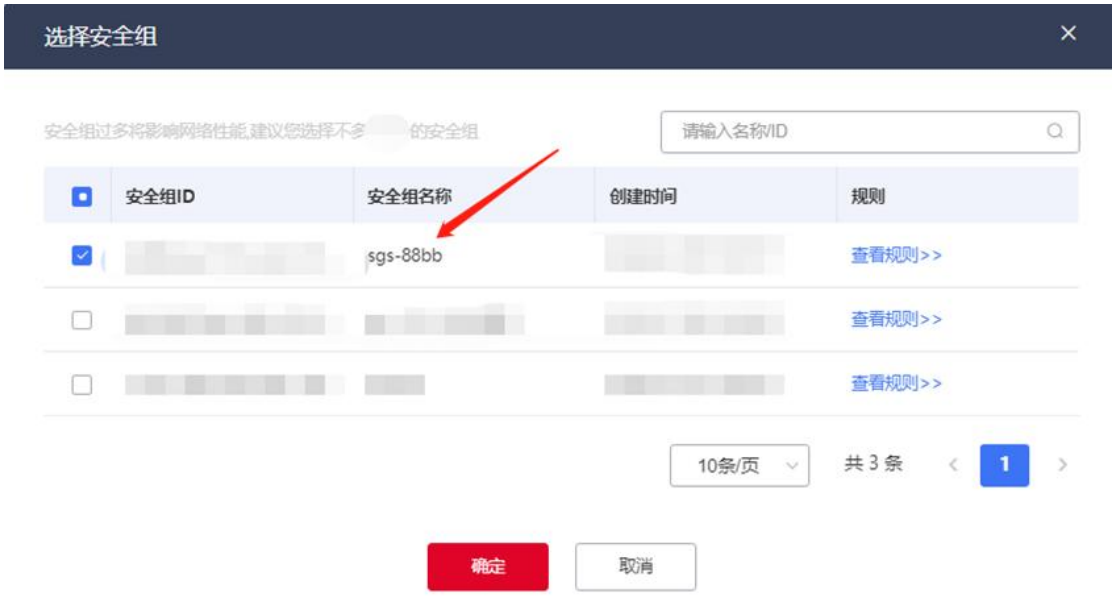
1. 点击“创建安全组”按钮，跳转至网络控制台-安全组页面。



2. 点击“创建安全组”按钮，在弹窗中等待模板下拉菜单选择“开放全部端口”。



3. 点击“确定”，完成安全组创建。
4. 返回云主机订购页面，点击“选择安全组”按钮，在安全组列表中勾选刚才创建的安全组。



5. 点击“确定”，完成安全组选择。
6. 弹性 IP 选择“自动分配”，根据需要选择带宽大小。为确保大模型学习机顺畅运行，如有其他数据下载需求，建议 10M 以上，否则建议 5M 及以下。



7. 点击“下一步：高级配置”。
8. 根据提示完成高级配置，点击“下一步：确认配置”。确认配置无误后，点击“立即购买”
9. 支付成功后返回云主机控制台，选择订购云主机所在的地域。云主机状态变更为“运行中”后即为开通成功，可以进行后续操作。

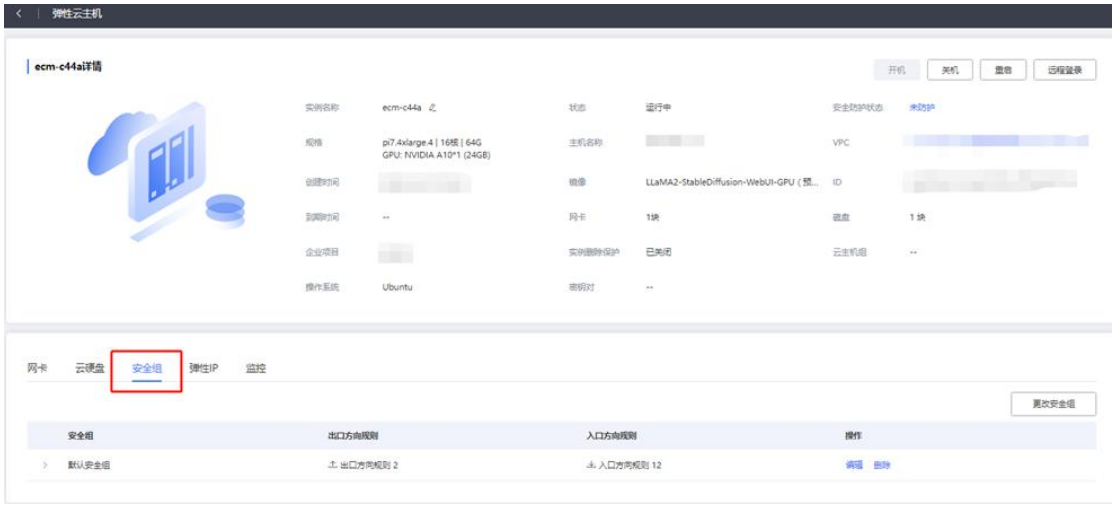


添加安全组规则

（1）点击云主机名称进入云主机详情页。



（2）选择云主机详情页下方的安全组页签。

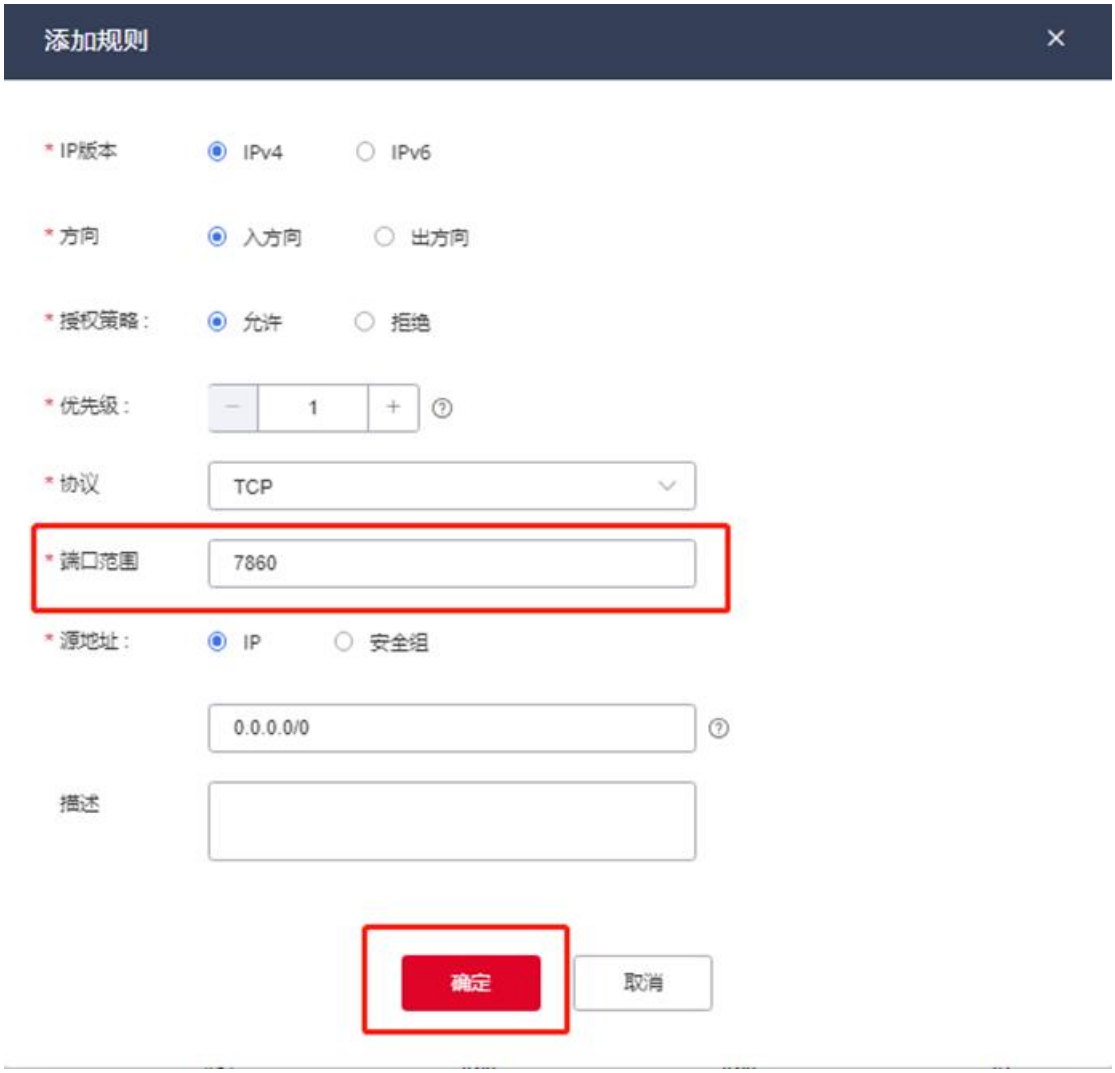


（3）点击默认安全组展开详情按钮。



（4）点击“添加规则”，在端口范围处填写 7860，点击确定。

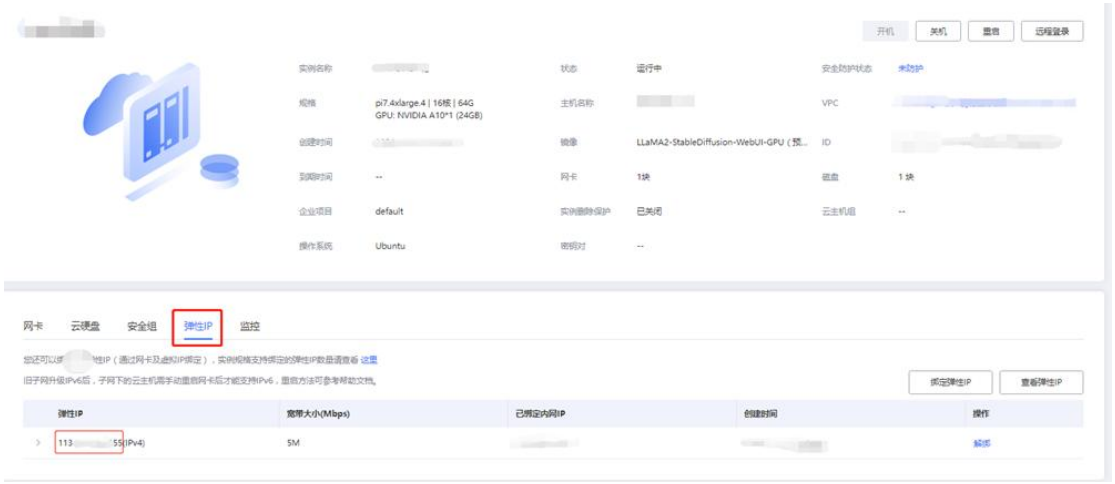




（5）再次点击“添加规则”，在端口范围处填写 7861，点击确定。

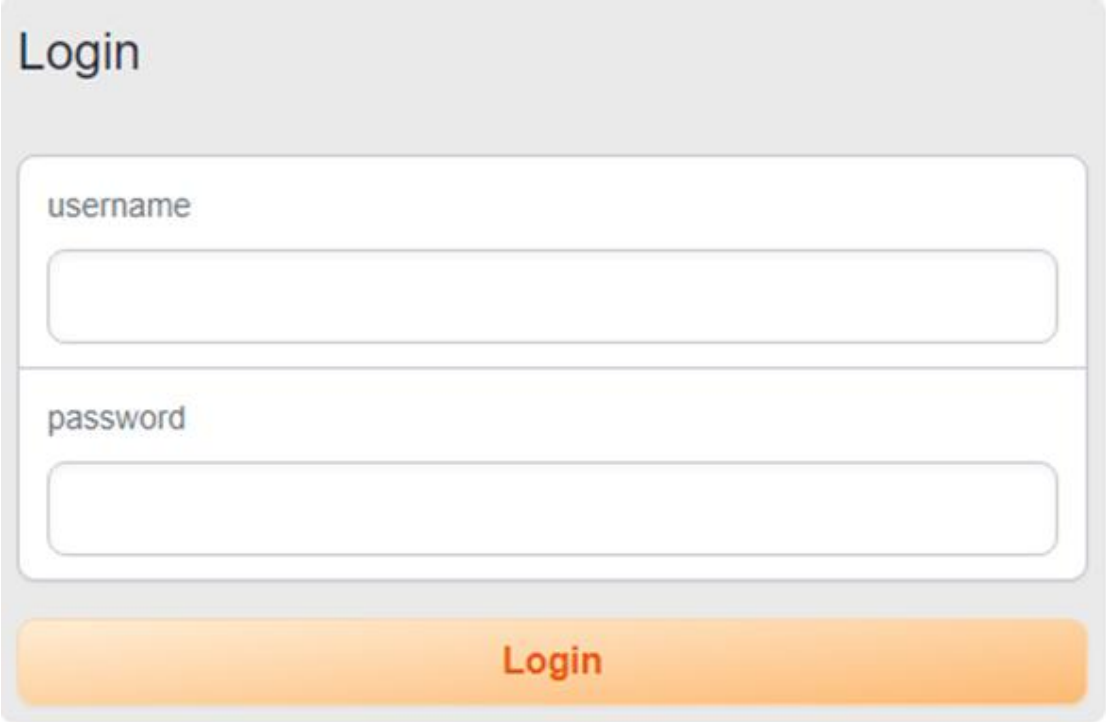
登录大模型

登录前准备：选择云主机详情的“弹性 IP”页签，复制列表中的弹性 IP。



（1）登录文本对话大模型（LLaMA2/LLaMA3）

将刚刚复制到的弹性 IP 替换到http://弹性 IP:7861/ 中并在浏览器中输入跳转至登录页面。



您可以任选以下两个账号其一进行登录：

注意

建议您登录后尽快修改密码，修改密码方法在文档“修改大模型学习机登录密码”可查看。

账号一：

username：user

password：ICFLvkJ69I

账户二：

username：user2

password：C7O0hgW5fxkN

（2）登录 AI 图片生成大模型（StableDiffusion）

将刚刚复制到的弹性 IP 替换到http://弹性 IP:7860/ 中并在浏览器中输入跳转至于登录页面。

您可以任选以下两个账号其一进行登录：

注意

建议您登录后尽快修改密码，修改密码方法在文档“修改大模型学习机登录密码”可查看。

账户一：

username：user

password：ICFLvkJ69I

账户二：

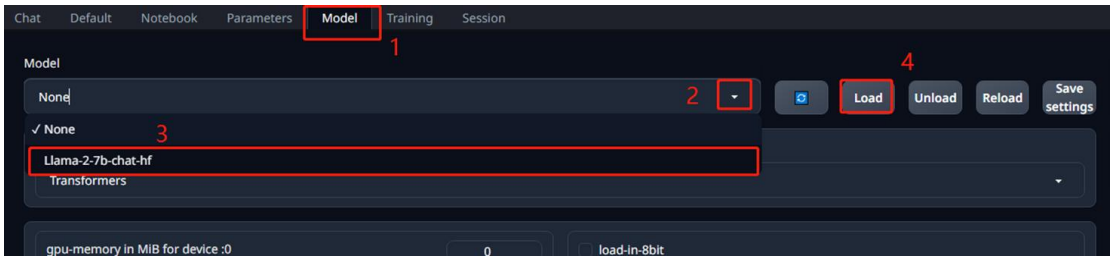
username：user2

password：C7O0hgW5fxkN

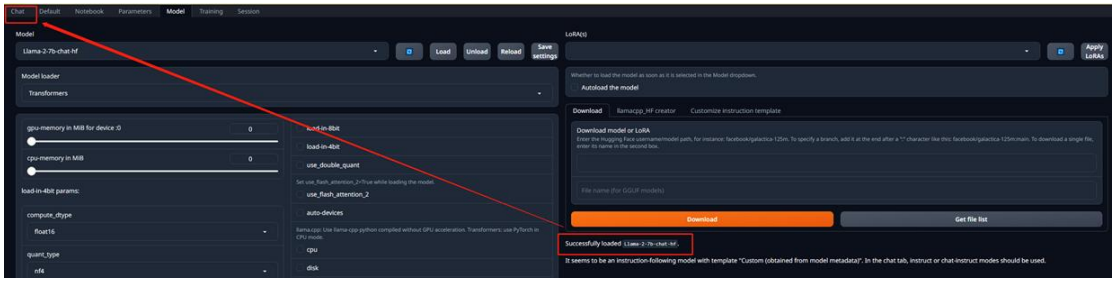
使用大模型进行推理计算

（1）文本对话大模型（LLaMA2/LLaMA3）

首先需要在 Model 标签页点击 LLaMA2/LLaMA3 大模型，然后点击 Load 按钮进行加载。



稍等一会，提示“Successfully loaded”，即可点击左上角的“chat”按钮，体验使用文本对话大模型（LLaMA2/LLaMA3）啦！

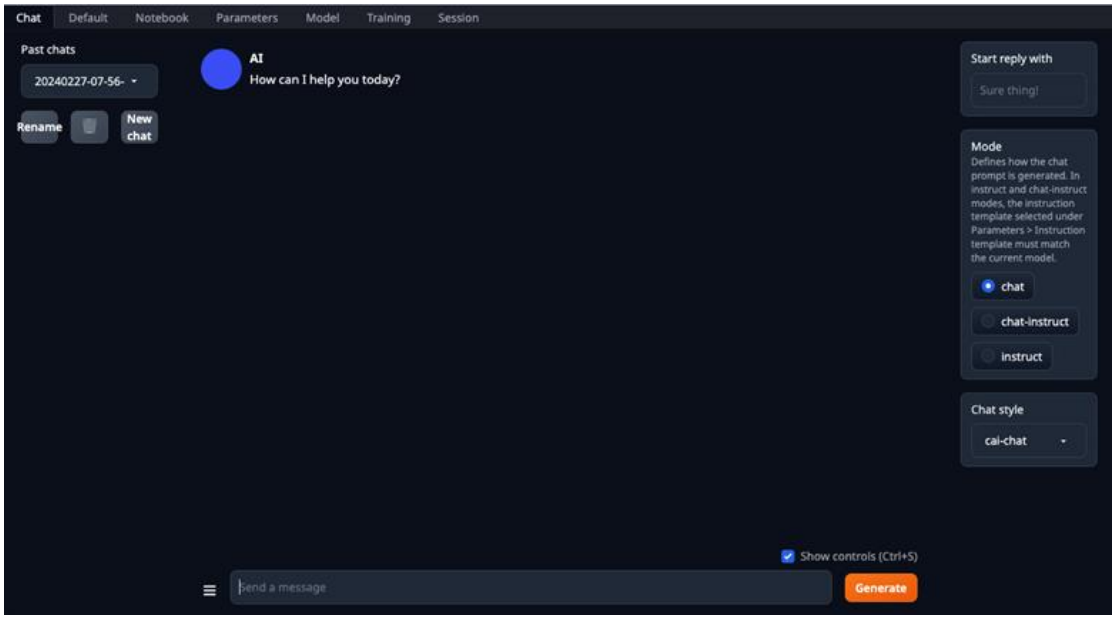


注意

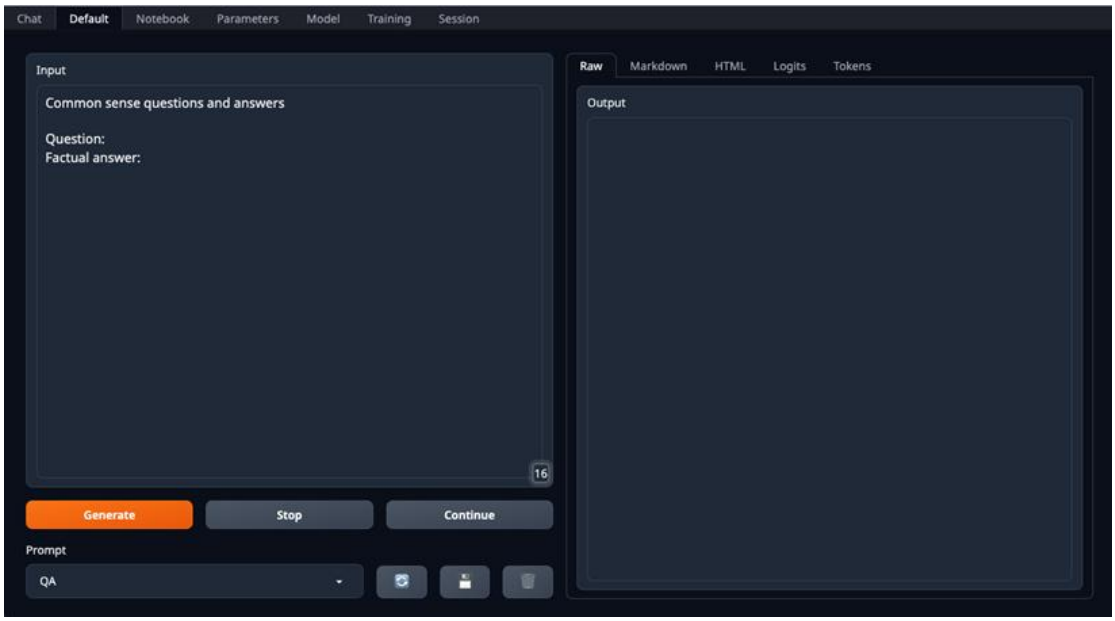
在加载 Llama 2 或者 LLaMA3 模型时， Pi2 和 s7.2xlarge.4 规格使用默认配置会因为显存/内存不足加载失败， 需要通过勾选 load-in-8bit 选项降低模型

精度或关闭 StableDiffusion 服务后再加载 Llama 2/LLaMA3 模型. 参考文档: 大模型学习机服务启停。

使用文本多轮对话页面:

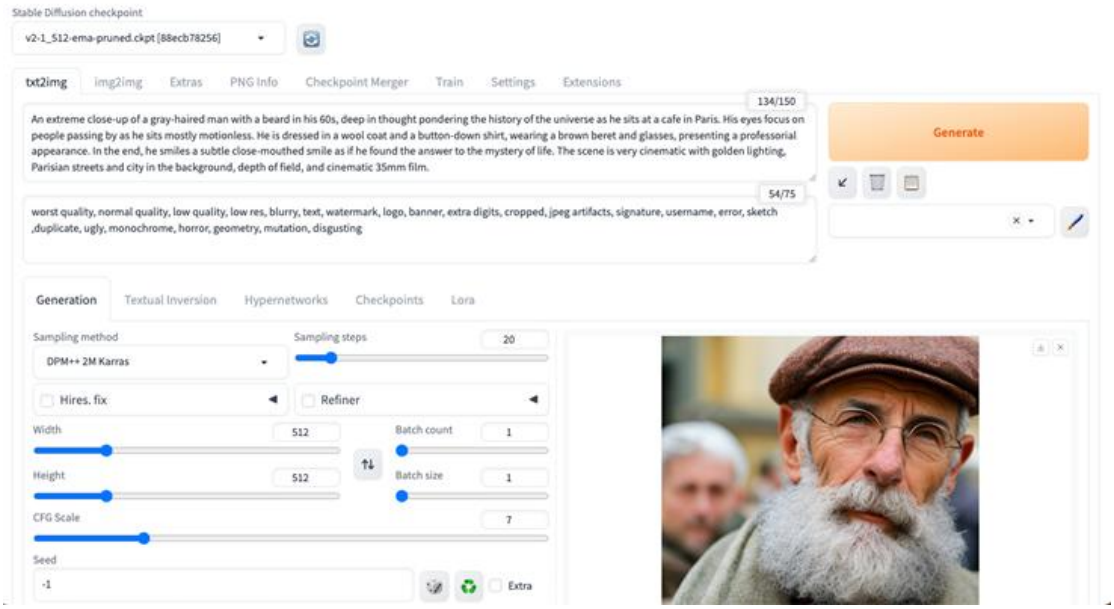


文章生成页面:



（2）AI 图片生成大模型（stable diffusion）：

AI 图片生成大模型登录后可以直接使用，以下为图片生成页面。



10.37.3 已有天翼云云主机，部署与登录 Llama 3/Llama 2 大模型学习机

操作前提

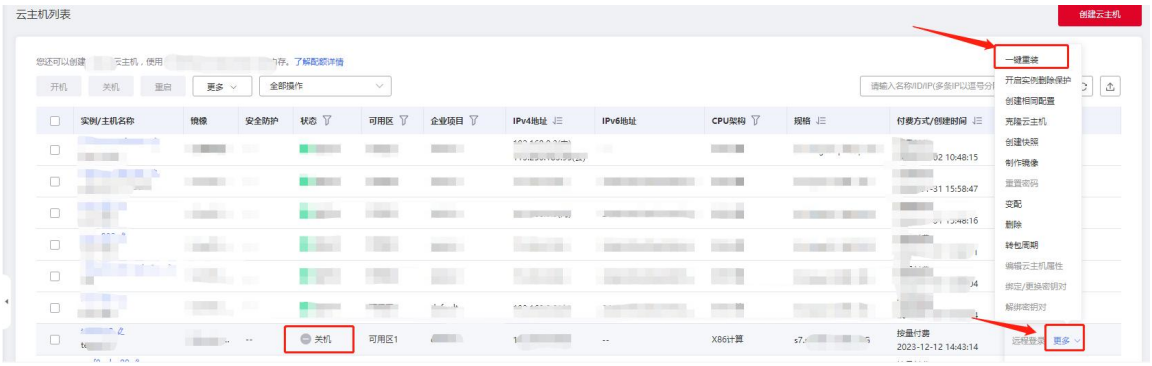
- 如果您尚未购买天翼云云主机，推荐您使用更简单快捷的活动页一键部署方案。
- 如果您已购买的云主机位于【重庆-重庆 2、宁夏-中卫 5】资源池（其余资源池镜像正在加载中），且满足以下条件，请参考本文部署大模型学习机。
- 配置：内存 32G， 显存 16G 以上
- 已开通公网 IP
- 云主机未到期

操作步骤

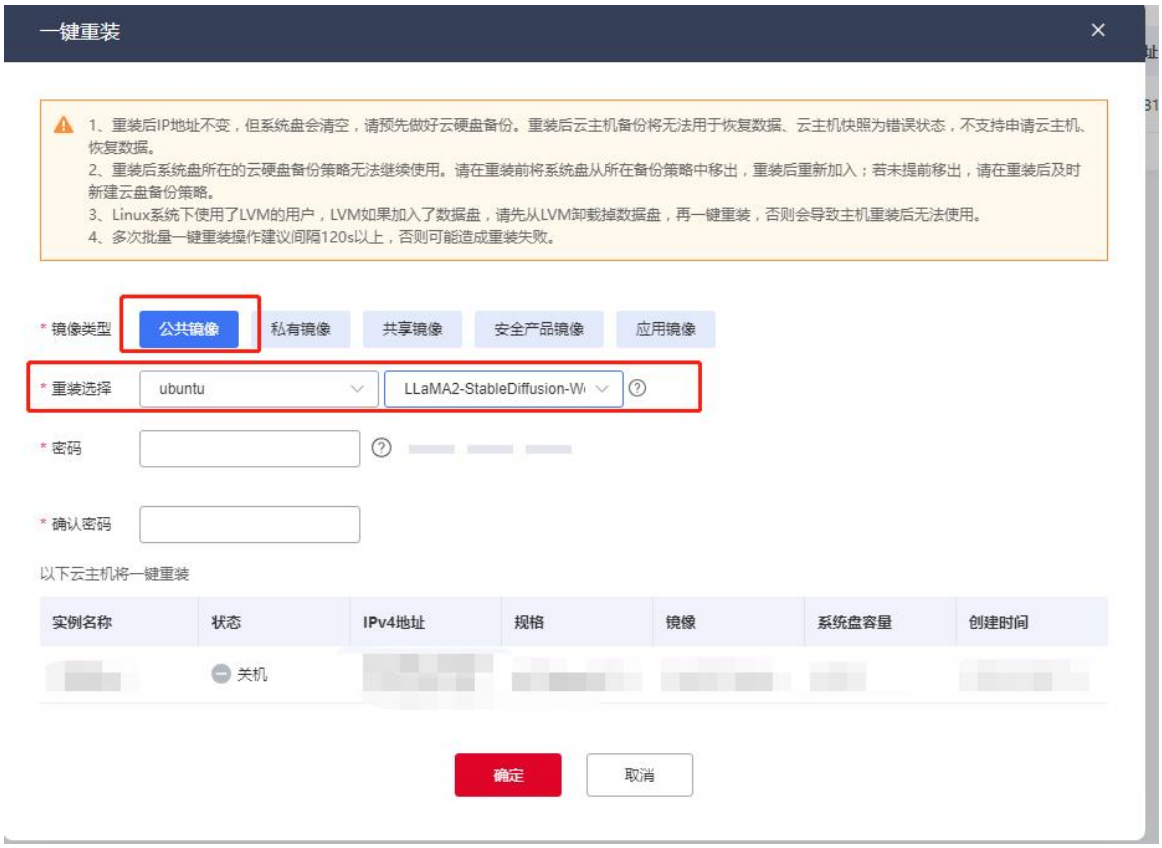
重装操作系统

1. 登录云主机控制台，前往您已购买的云主机所在的资源池。确认云主机状态为“已关机”，则可以进行重装操作系统操作。

点击“更多”按钮，点击下拉菜单中“一键重装”按钮。



2. 在重装操作系统弹窗中，镜像类型选择“公共镜像”，并在重装选择下拉菜单中选择“ubuntu-LLaMA2-StableDiffusion-WebUI-GPU(预装大模型环境)(60GB)”或 ubuntu-LLaMA3-StableDiffusion-WebUI-GPU（预装大模型环境）(60GB)。



设置并确认密码，点击“确定”按钮，完成重装操作系统操作。

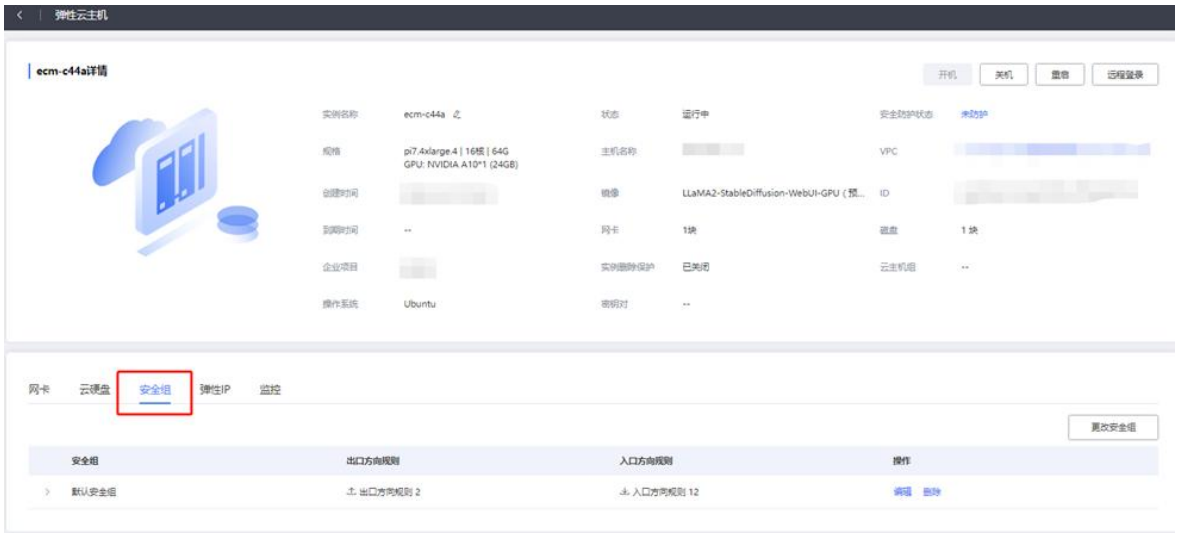
3. 返回云主机控制台，选中已完成重装操作系统的云主机，点击“开机”按钮。确认云主机状态为“运行中”，则可以进行后续操作。

添加安全组规则

（1）点击云主机名称进入云主机详情页。



（2）选择云主机详情页下方的安全组页签。

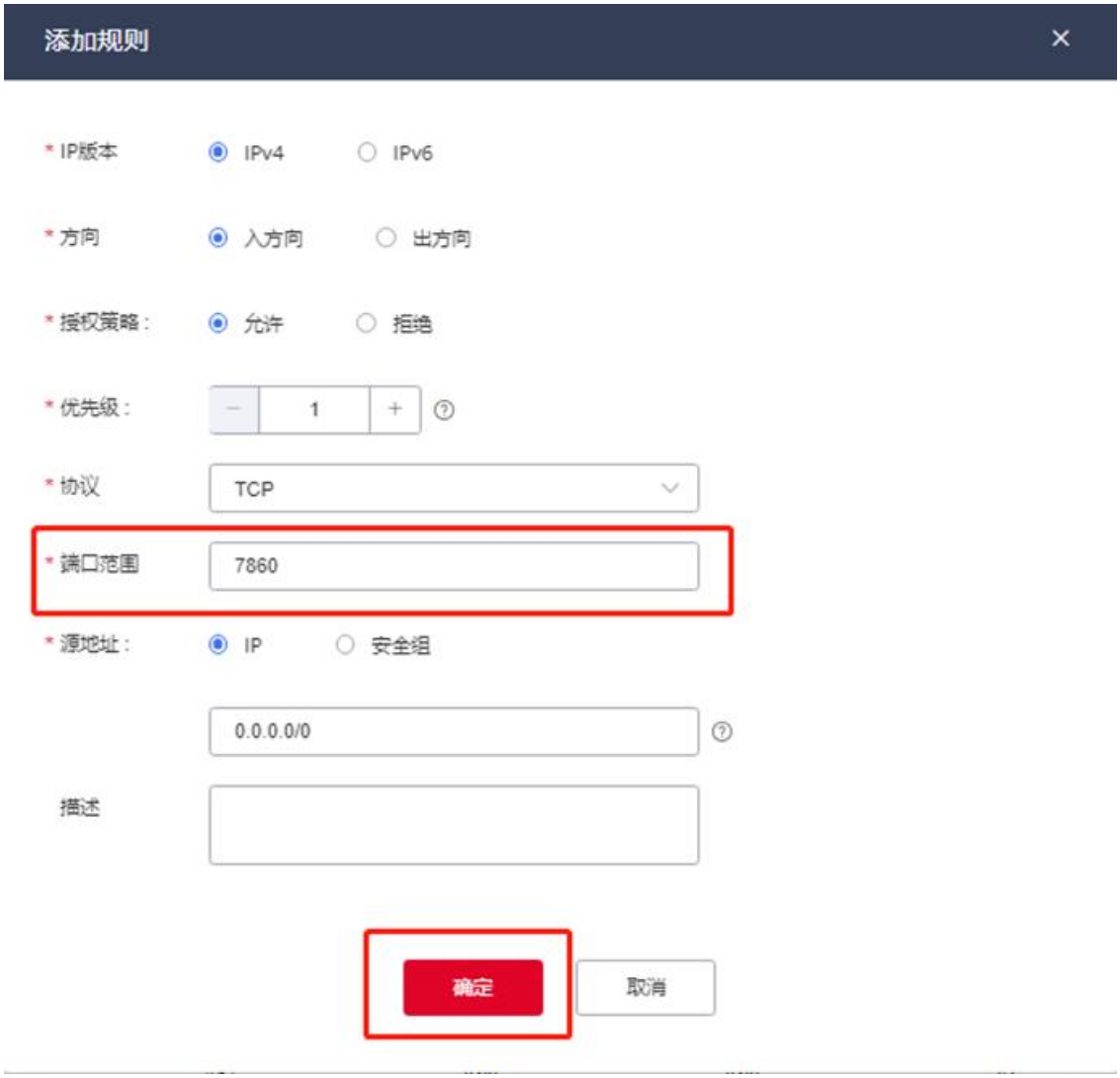


（3）点击默认安全组展开详情按钮。

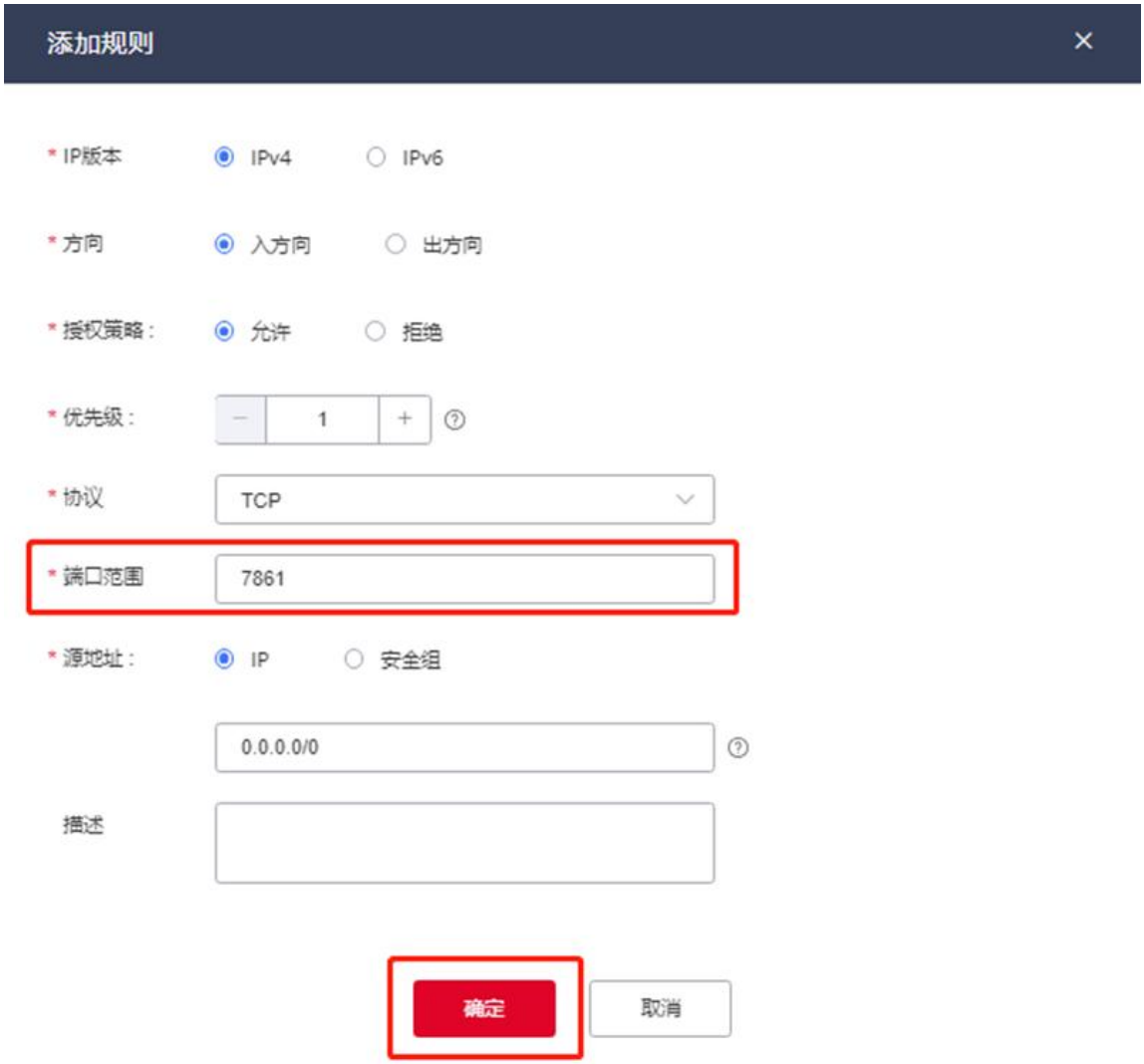


（4）点击“添加规则”，在端口范围处填写 7860，点击确定。



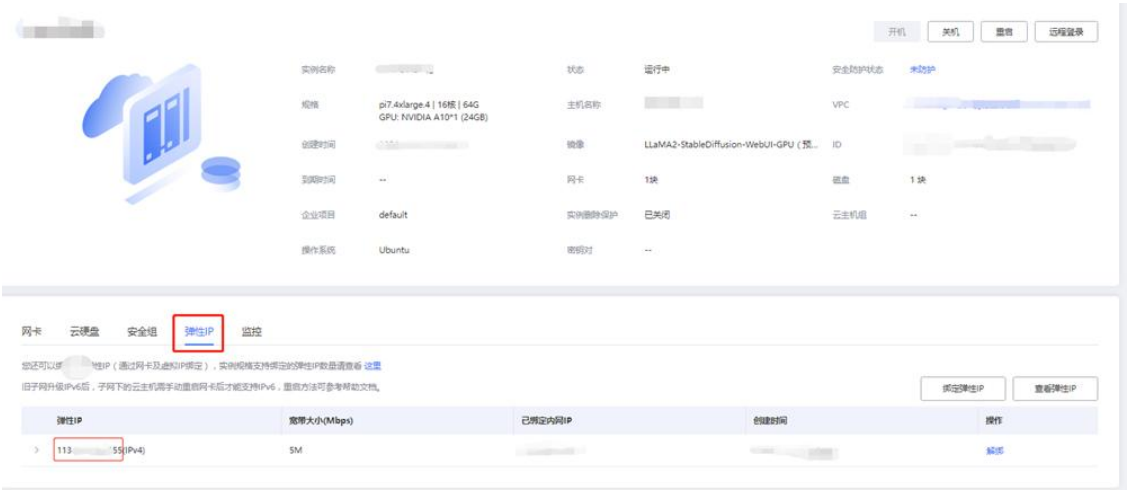


（5）再次点击“添加规则”，在端口范围处填写 7861，点击确定。



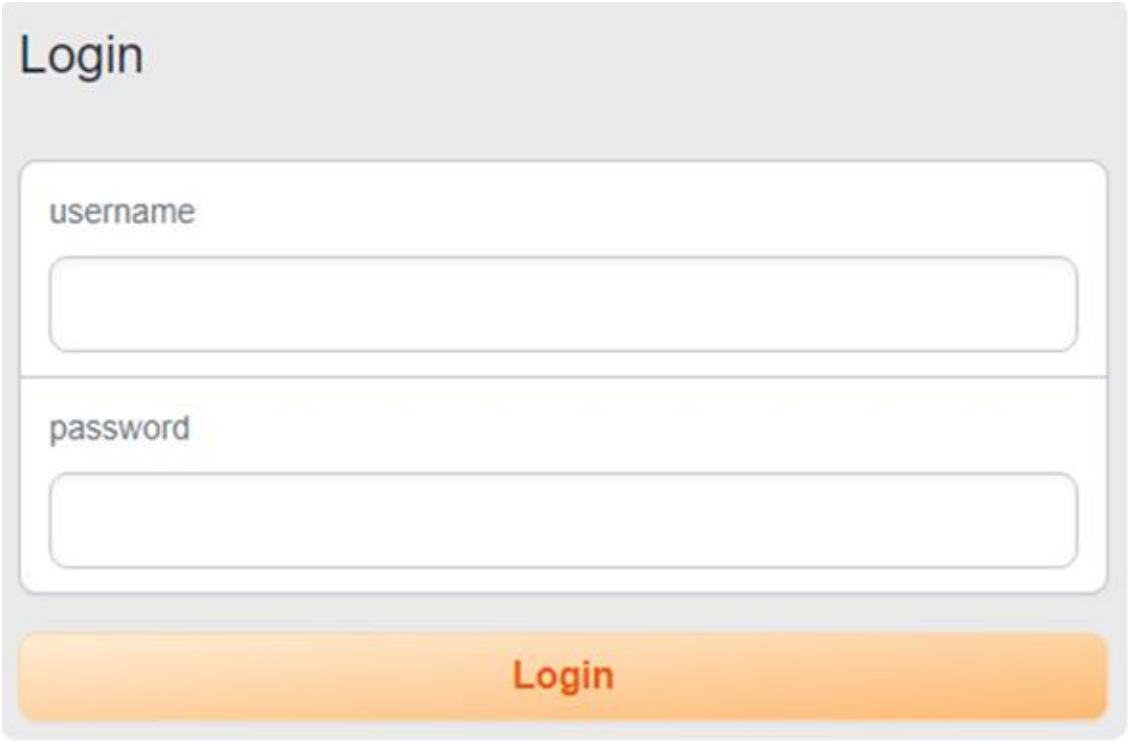
登录大模型

登录前准备：选择云主机详情的“弹性 IP”页签，复制列表中的弹性 IP。



（1）登录文本对话大模型（LLaMA2/LLaMA3）

将刚刚复制到的弹性 IP 替换到 http://弹性 IP:7861/ 中并在浏览器中输入跳转至登录页面。



您可以任选以下两个账号其一进行登录：

注意

建议您登录后尽快修改密码，修改密码方法在文档“修改大模型学习机登录密码”可查看。

账号一：

username：user

password：ICFLvkJ69I

账户二：

username：user2

password：C7O0hgW5fxkN

（2）登录 AI 图片生成大模型（StableDiffusion）

将刚刚复制到的弹性 IP 替换到http://弹性 IP:7860/ 中并在浏览器中输入跳转至于登录页面。

您可以任选以下两个账号其一进行登录：

注意
建议您登录后尽快修改密码，修改密码方法在文档“修改大模型学习机登录密码”可查看。

账户一：

username：user

password：ICFLvkJ69I

账户二：

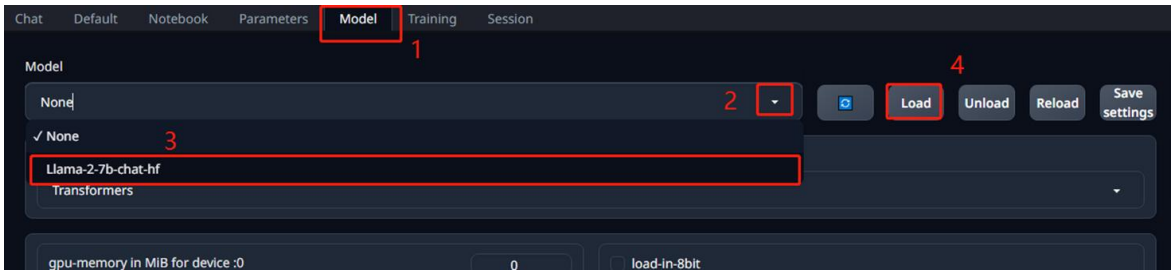
username：user2

password：C7O0hgW5fxkN

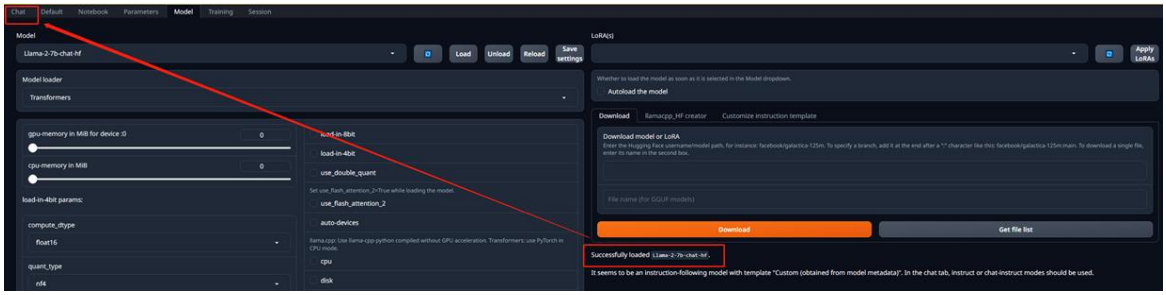
使用大模型进行推理计算

（1）文本对话大模型（LLaMA2/LLaMA3）

首先需要在 Model 标签页点击 LLaMA2/LLaMA3 大模型，然后点击 Load 按钮进行加载。



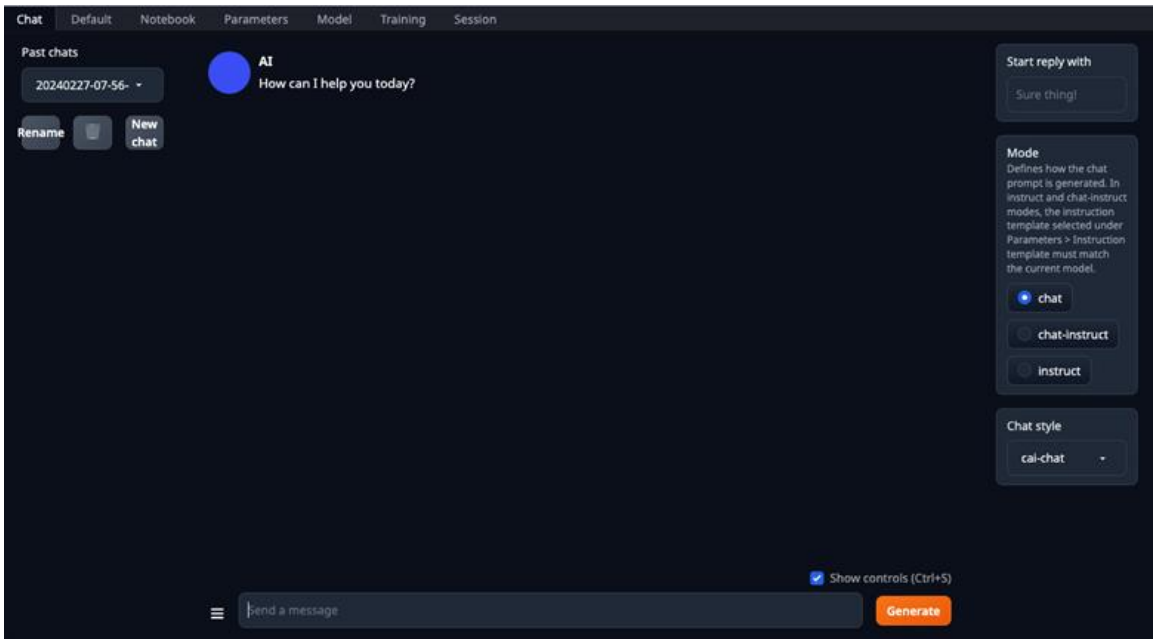
稍等一会，提示“Successfully loaded”，即可点击左上角的“chat”按钮，体验使用文本对话大模型（LLaMA2/LLaMA3）啦！



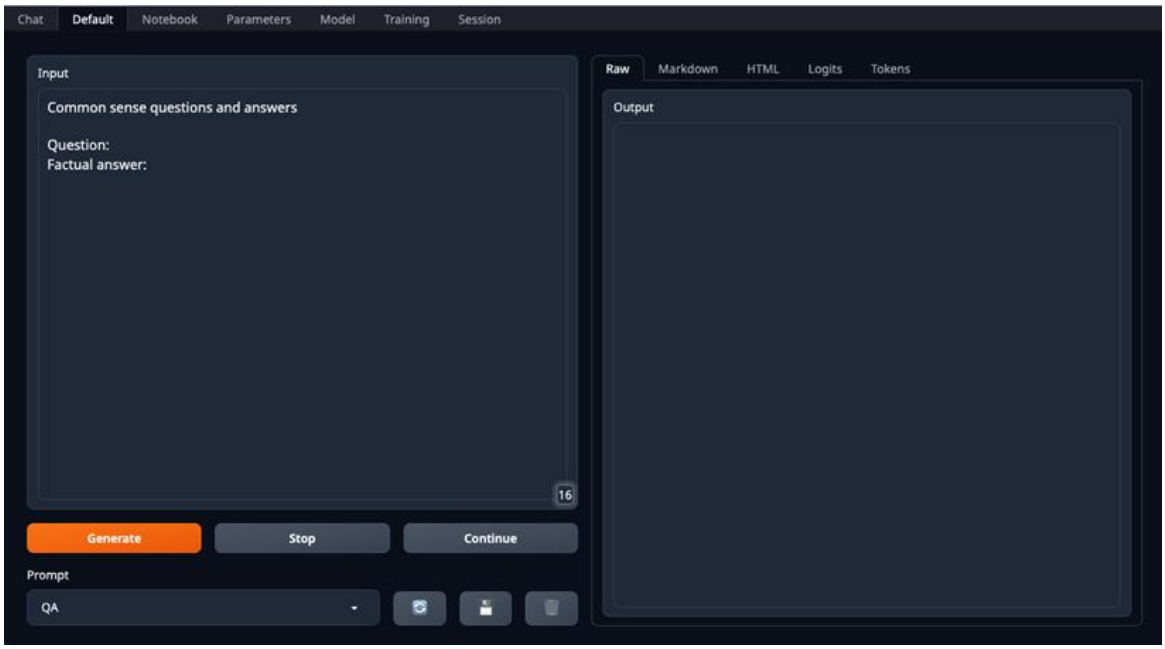
注意

在加载时参数默认以 fp16 精度进行加载，Pi2 和 s7.2xlarge.4 规格需要勾选 load in 8 bit，否则会因显存/内存不足而加载失败

使用文本多轮对话页面:

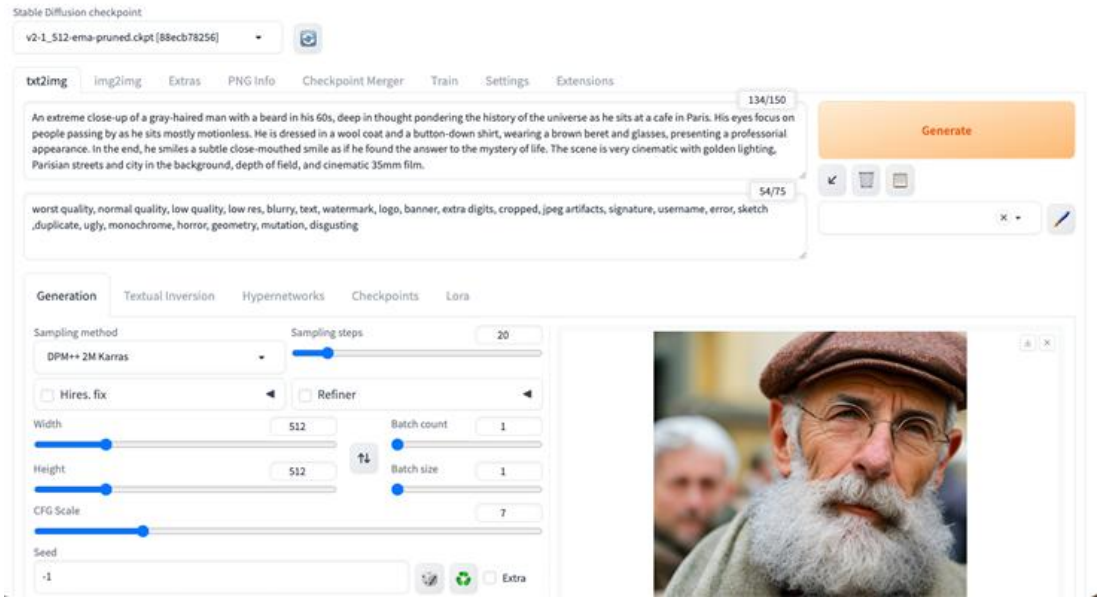


文章生成页面:



（2）AI 图片生成大模型（stable diffusion）：

AI 图片生成大模型登录后可以直接使用，以下为图片生成页面。



10.37.4 修改大模型学习机登录密码

本文向您介绍如何修改大模型学习机的登录密码，您也可通过视频教程（访问码：hbw8）进行大模型学习机登录密码的修改，以下为修改大模型学习机的登录密码的文字教程。

1.在云主机控制台，选择您购买云主机的地域，点击云主机的操作-更多-重置密码。

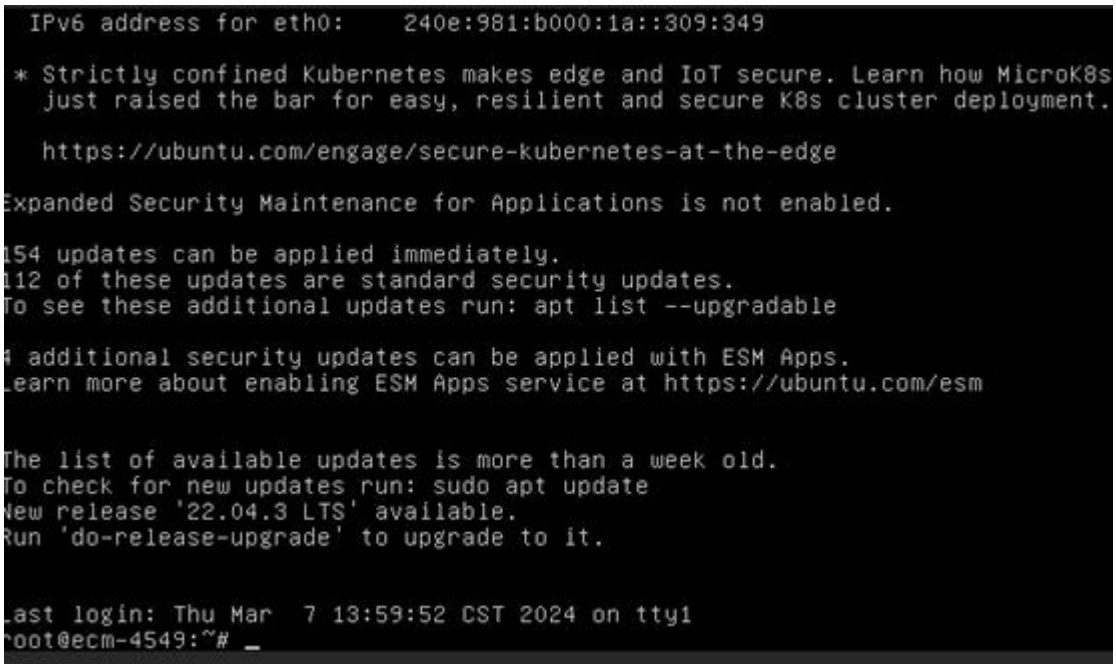




2. 点击“远程登录”按钮。



3. 进入虚机内部，输入 root 回车，在 password 处输入刚刚修改的密码（密码不可见，请输入完直接回车）。



4. 击右上角的“粘贴输入”，在输入框中粘贴以下指令，点击“确定发送”，点击虚机屏幕后回车发送指令。

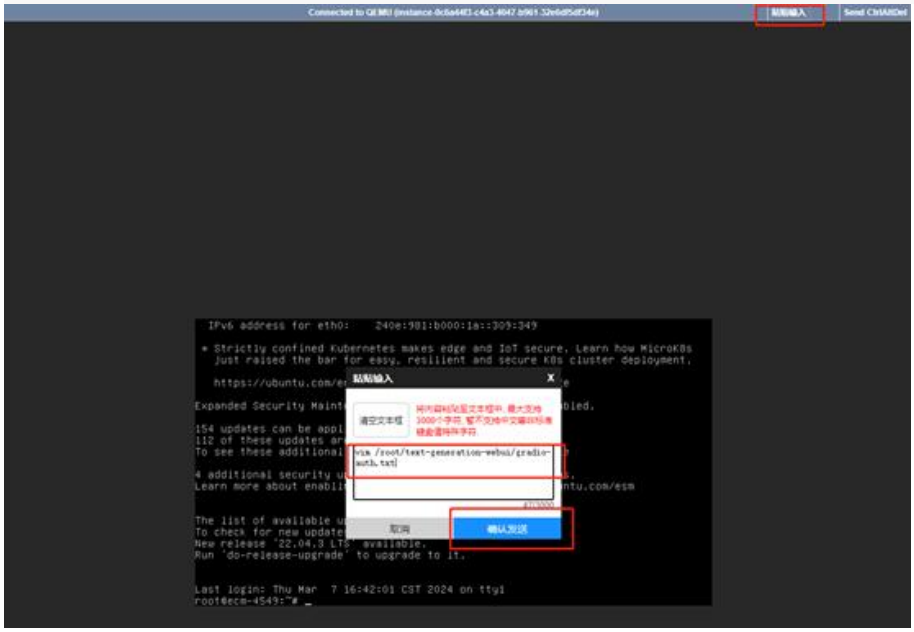
修改文本大模型密码指令：

```
vim /root/text-generation-webui/gradio-auth.txt
```

或

修改图像生成大模型密码指令：

```
vim /root/stable-diffusion-webui/gradio-auth.txt
```



进入修改对应大模型登录密码界面。



5. 改 user 和 user 后的密码后，按 Esc 退出编辑模式。再次点击右上角的“粘贴输入”，在输入框中粘贴 :wq 指令，点击“确定发送”，点击虚机屏幕后回车发送指令。返回以下内容表示修改成功。

```
~
<-generation-webui/gradio-auth.txt" 2L, 32C written
root@ecm-4549:~# _
```

6. 再次点击右上角的“粘贴输入”，在输入框中依次粘贴输入以下命令，“确认发送”后点击虚机屏幕，回车。使修改的大模型密码登录密码生效。

```
cd /root

sh restart_llama_webui.sh

sh restart_sd_webui.sh
```

```
<-generation-webui/gradio-auth.txt" 2L, 32C written
root@ecm-4549:~# cd /root
root@ecm-4549:~# sh restart_llama_webui.sh
container_status:running
llama is running, restart it
sh restart_sd_webui.shllama
root@ecm-4549:~# sh restart_sd_webui.sh_
```

此时密码修改成功，您可以通过使用修改后的密码登录大模型。

10.37.5 大模型学习机服务启停最佳实践

前提条件

学习机内通过 docker 容器的方式部署了 text-generation-webui 与 stable-diffusion-webui 两套框架的运行环境. 可以通过以下命令进行查看:

```
docker ps -a
```

```
root@ecm-4549:~# docker ps -a
CONTAINER ID   IMAGE                    COMMAND       CREATED          STATUS          PORTS     NAMES
dfc4c8ab835a   stable_diffusion:gpu     "/bin/bash"   16 minutes ago   Up 4 minutes              stable_diffusion
61de5996b808   llama:gpu                "/bin/bash"   16 minutes ago   Up 4 minutes              llama
```

网页服务的进程在容器内启动，框架代码与模型放置在本机的/root 目录下，通过挂载的方式传入到容器中。

您无需进入容器执行操作，只需要直接操作/root/text-generation-webui 和/root/stable-diffusion-webui 下的文件即可将改动同步到容器内部。

停止服务

直接停止 docker 容器即可停止对应服务，可选择执行以下命令：

```
docker stop llama

docker stop stable_diffusion
```



重启服务

通过执行预置的重启脚本启动容器与服务，可选择执行以下命令：

```
sh /root/restart_llama_webui.sh

sh /root/restart_sd_webui.sh
```



也可以只重启容器，不启动服务，随后手动进入容器启动服务。

此方式可以用来查看框架的具体执行日志，以便于追查问题等。

```
sh /root/docker_run_llama.sh

sh /root/docker_run_sd.sh
```



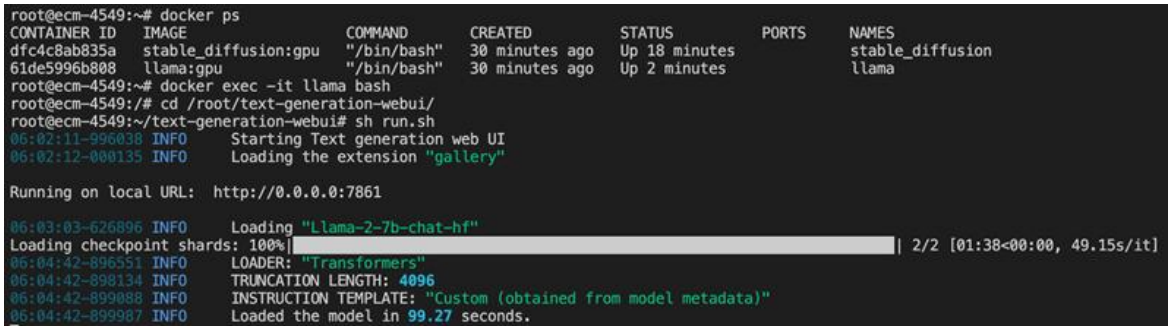
随后手动进入容器，启动服务，以 llama 为例：

```
docker exec -it llama bash

cd /root/text-generation-webui

sh run.sh
```

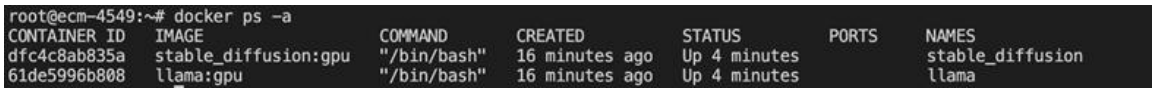
此时会在命令行界面打印日志. 如果关闭命令行界面或者通过 ctrl + c 终止命令，则会关闭服务。



前提条件

学习机内通过 docker 容器的方式部署了 text-generation-webui 与 stable-diffusion-webui 两套框架的运行环境. 可以通过以下命令进行查看:

```
docker ps -a
```



网页服务的进程在容器内启动，框架代码与模型放置在本机的/root 目录下，通过挂载的方式传入到容器中。

您无需进入容器执行操作，只需要直接操作/root/text-generation-webui 和 /root/stable-diffusion-webui 下的文件即可将改动同步到容器内部。

停止服务

直接停止 docker 容器即可停止对应服务，可选择执行以下命令：

```
docker stop llama

docker stop stable_diffusion
```



重启服务

通过执行预置的重启脚本启动容器与服务，可选择执行以下命令：

```
sh /root/restart_llama_webui.sh

sh /root/restart_sd_webui.sh
```



也可以只重启容器，不启动服务，随后手动进入容器启动服务。

此方式可以用来查看框架的具体执行日志，以便于追查问题等。

```
sh /root/docker_run_llama.sh

sh /root/docker_run_sd.sh
```



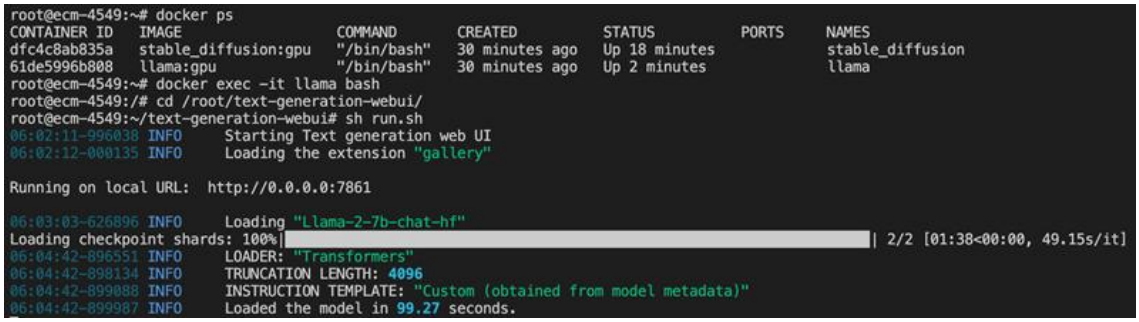
随后手动进入容器，启动服务，以 llama 为例：

```
docker exec -it llama bash

cd /root/text-generation-webui

sh run.sh
```

此时会在命令行界面打印日志. 如果关闭命令行界面或者通过 ctrl + c 终止命令，则会关闭服务。



10.37.6 大模型学习机图像生成最佳实践

前提条件

- 大模型学习机预装了 Llama 2-7b-chat 模型与 stable-diffuison-base-v2.1 模型，以及配套的开源服务框架 text-generation-webui 与 stable-diffuison-webui,使您不需单独进行下载模型与配置环境。
- 云主机开通与如何登录网页页面参见部署文档。
- 您也可以参考视频教程（访问码：qf8m），帮助您快速学习大模型学习机的使用。

以下将向您介绍如何进行 StableDiffusion 模型图像生成任务：

1. StableDiffusion 模型加载

> 说明
> 以下内容基于开源框架 stable-diffusion-webui。

1 .1 基础模型加载

页面最上方左侧下拉菜单展开后会显示所有存放于云主机 /root/stable-diffusion-webui/models/Stable-diffusion 目录下的模型. 。

学习机初始预装了原版 v2-1_512-ema-pruned.ckpt,您也可随时下载其他版本的 StableDiffusion 放入目录下进行加载。v2-1 版模型占用约 3G 显存。

```
root@ecm-4549:~/stable-diffusion-webui/models/Stable-diffusion# ll
total 5092656
drwxr-xr-x  2 root root       4096 Feb 27 13:30  ./
drwxr-xr-x 14 root root       4096 Feb 27 14:13  ../
-rw-r--r--  1 root root          0 Feb 27 11:45 'Put Stable Diffusion checkpoints here.txt'
-rw-r--r--  1 root root 5214865159 Dec  7  2022  v2-1_512-ema-pruned.ckpt
```



1.2 LoRA 模型加载

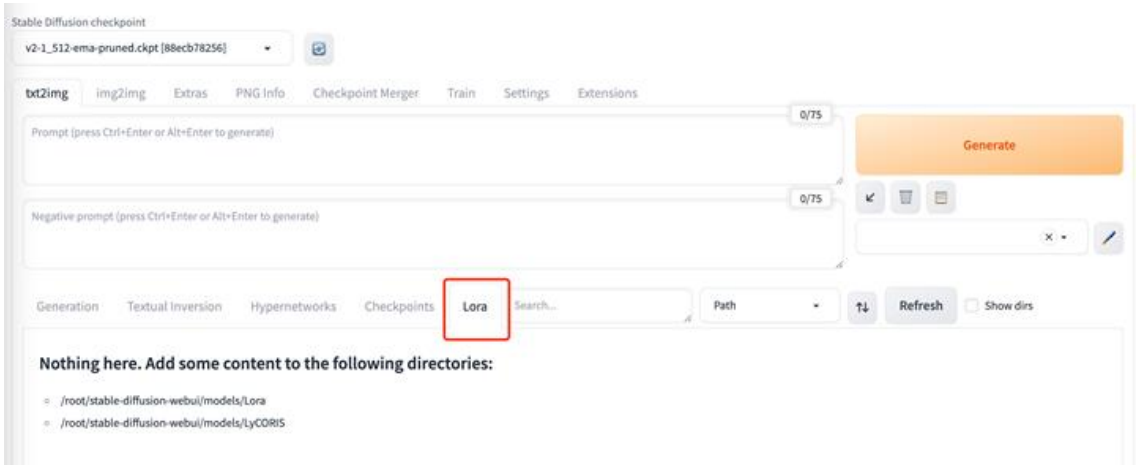
由于基础大模型参数量极多，训练需要的资源量极大，因此基于基础大模型进行微调的小模型应运而生。LoRA 技术便是其中最主流的一种。

LoRA 小模型无法独立使用，需要在加载完基础大模型后再进行加载，对基础大模型的能力进行扩展。

学习机内没有预装 LoRA 小模型，您可以在网上下载其他基于 StableDiffusion 的 LoRA 模型并放到学习机的/root/stable-diffusion-webui/models/Lora 目录下。

StableDiffusion 模型的 LoRA 模型以提示词(prompt)的形式进行使用，您可以在下载 LoRA 模型时进行具体使用方法查阅。

```
root@ecm-4549:~/text-generation-webui/loras# ll
total 8
drwxr-xr-x  2 root root 4096 Feb 27 11:45 ./
drwxr-xr-x 22 root root 4096 Feb 29 14:27 ../
-rw-r--r--  1 root root    0 Feb 27 11:45 place-your-loras-here.txt
```

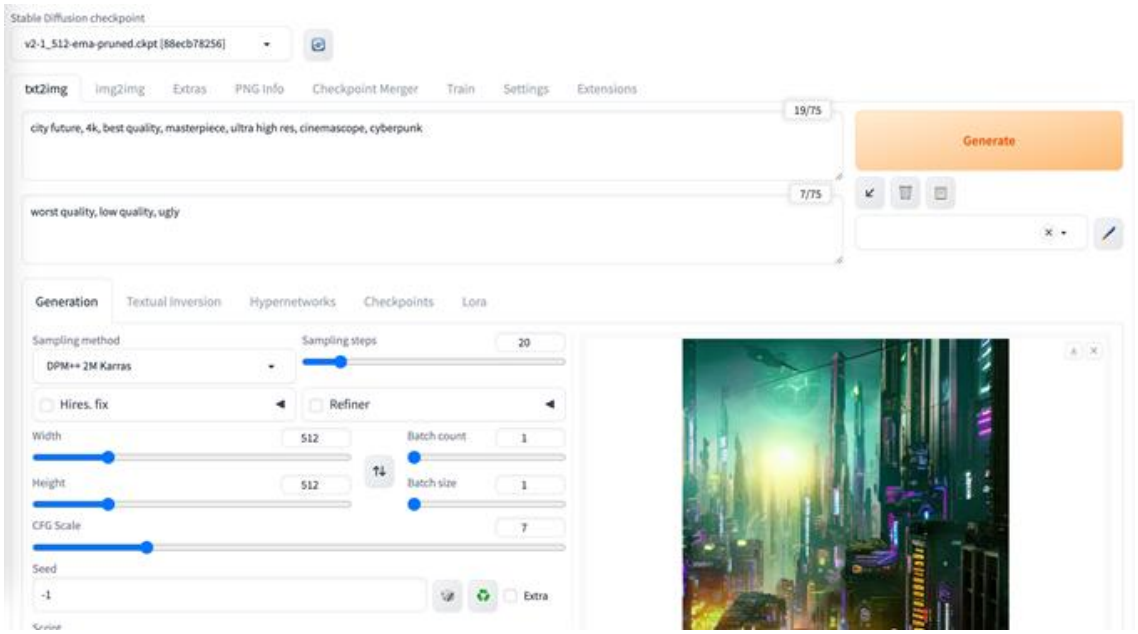


2. StableDiffusion 模型使用

2.1 文本生成图像模式

在 txt2img 标签页分别输入提示词(prompt)和负向提示词(prompt)，随后点击 Generate 按钮。

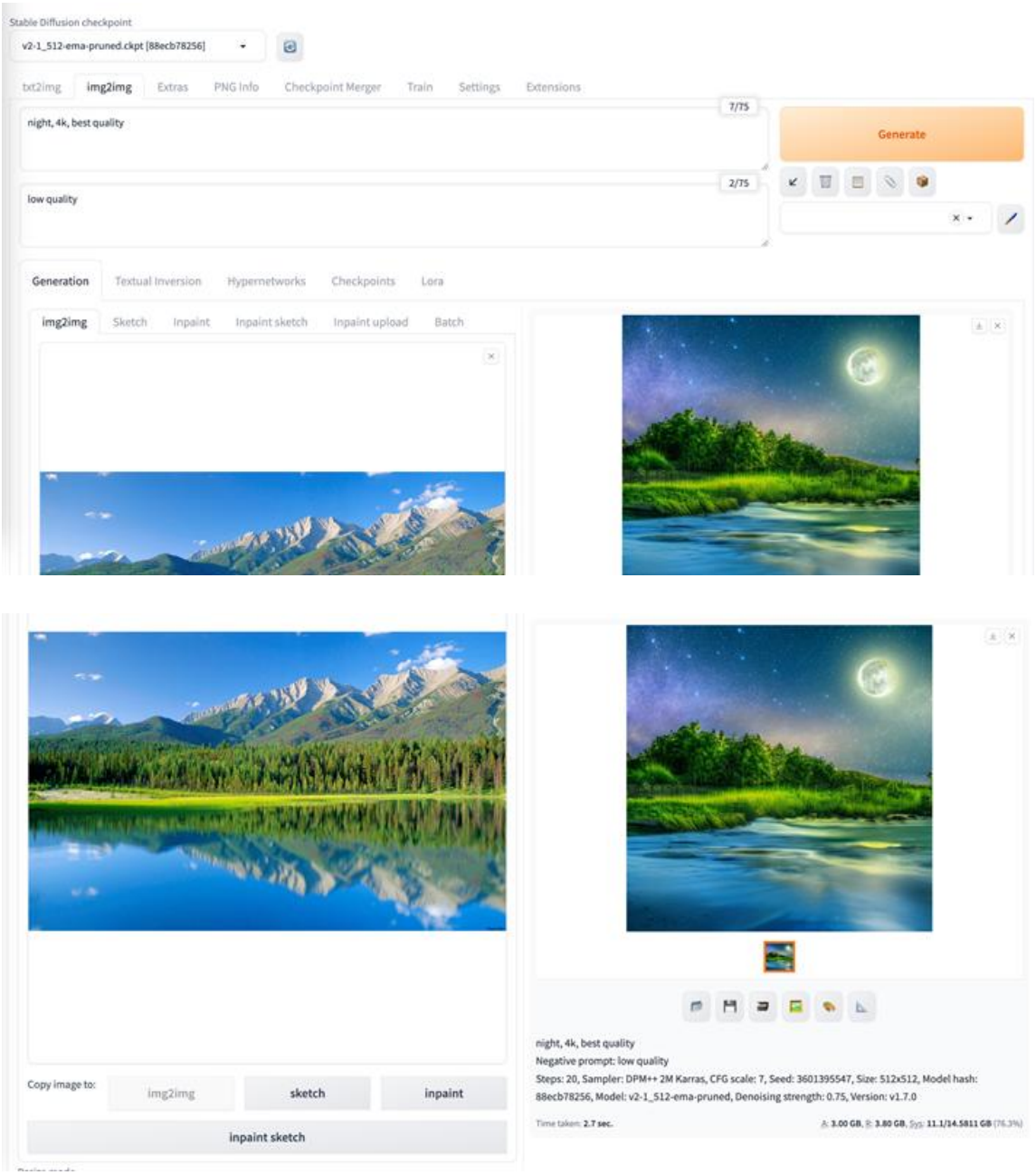
StableDiffusion 的提示词写法要求较高，这里无法做详细介绍. 您可搜索相关资料进行学习。



2.2 图像生成图像模式

在 img2img 标签页分别输入提示词(prompt)和负向提示词(prompt)，同时在左下方选择上传一张图片，最后点击 Generate 按钮。

StableDiffusion 将给予给定图片和提示词，创作一张新的图片。



2.3 生成质量控制

StableDiffusion 模型生成图片质量可以在页面下方 Generation 标签页进行调整，但需要您具有一定的 diffusion 模型(扩散模型)背景知识。 这里仅对其中部分基础参数进行介绍：

- Sampling Steps：采样步数. 数值越大图像越精细。
- Width & Height：图像的宽度与高度。
- Batch count：生成图像总批次。
- Batch size：每一批次生成图像的个数。
- CFG Scale：控制生成图像与提示词的关联度，数值越大关联度越高，但图片质量可能越差。
- Seed：随机数种子. 控制模型生成结果的随机性,如果输入-1 则代表不固定随机数种子，相同的提示词每次生成图片不同；如果输入其他数字则代表固定随机数种子，相同提示词每次生成图片相同。

10.37.7 大模型学习机文本生成最佳实践

前提条件

- 大模型学习机预装了 Llama 2-7b-chat 模型与 stable-diffuison-base-v2.1 模型，以及配套的开源服务框架 text-generation-webui 与 stable-diffuison-webui,使您不需单独进行下载模型与配置环境。
- 云主机开通与如何登录网页页面参见部署文档。
- 您也可以参考视频教程（访问码：qf8m），帮助您快速学习大模型学习机的使用。

以下将向您介绍如何进行 Llama 2 模型文本生成任务：

注意
框架默认角色名称为"You"，但主流大模型默认角色名称一般为"User"，可能会导致模型效果显著变差。
修改角色名称的方式有:

1. 在网页 Parameters -> Chat -> User 标签下修改 Name 字段为"User"，此方法刷新页面后就会失效。
2. 修改配置文件/root/text-generation-webui/modules/shared.py 中的'name1': 'User'，并重启 llama 服务， 此方法永久生效。

1. 大语言模型加载

说明
以下内容基于开源框架 text-generation-webui。

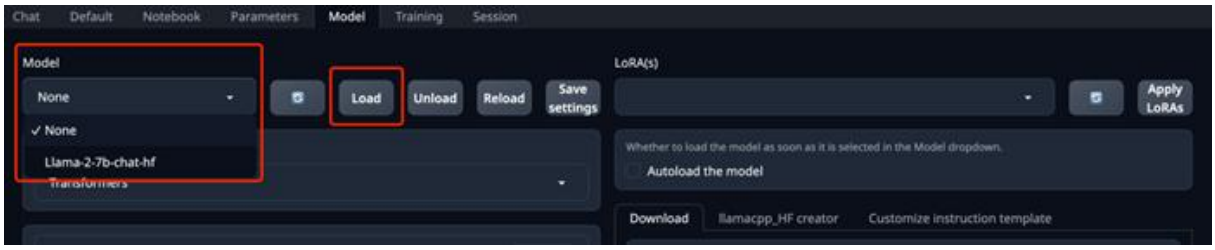
1.1 基础模型加载

大语言模型体积较大，需要占用较多的内存/显存资源，因此在刚启动服务时并未进行加载，需要手动选择模型进行加载。

在页面上切换到Model 标签页，左侧下拉菜单展开后会显示所有存放于云主机/root/text-generation-webui/models 目录下的模型. 学习机初始预装了 Llama-2-7b-chat 模型(huggingface 格式)，您也可随时下载其他大模型并放入 models 目录下进行加载。

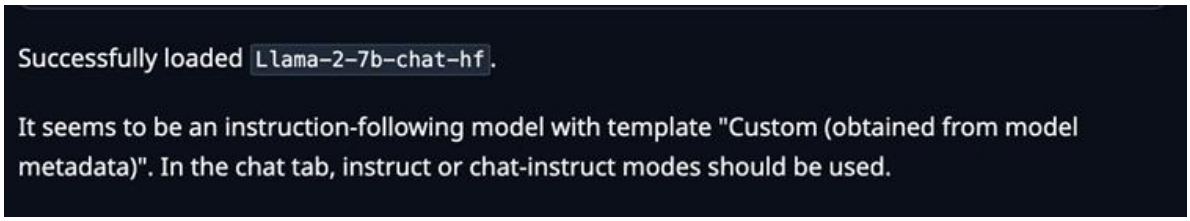
默认 Model loader: Transformers 能够支持 huggingface 的主流大模型，在自行加载其他模型时您需要确认模型格式是否匹配。

```
root@ecm-4549:~/text-generation-webui/models# ll
total 20
drwxr-xr-x  3 root root 4096 Feb 27 14:24 ./
drwxr-xr-x 22 root root 4096 Feb 29 14:27 ../
drwxr-xr-x  2 root root 4096 Feb 23 15:24 Llama-2-7b-chat-hf/
-rw-r--r--  1 root root 5724 Feb 27 11:45 config.yaml
-rw-r--r--  1 root root    0 Feb 27 11:45 place-your-models-here.txt
```



右下角将出现模型加载样式，直至加载成功. 耗时可能较长，期间不要退出页面。





注意

在加载预装的 Llama-2-7b 模型时， 如果您使用的是 GPU 学习机， 模型将默认以 fp16 半精度进行加载， 消耗 13.1G 显存； 如果您使用的是 CPU 学习机， 模型将以 fp32 单精度进行加载， 消耗 26G 内存。

因此， 如果您的机型是内存为 32G 或显存为 16G 的型号， 则会因为资源不足导致无法直接加载 Llama-2-7b 模型。

解决办法::

(1) 参考大模型学习机服务启停文档说明，加载 Llama 2-7b 模型前先将图像生成服务关停。

(2) GPU 学习机能够支持对模型进行量化，降低模型精度的同时缩减模型大小. 您可以在 Model 页下方找到 load-in-8bit 的开关并进行勾选，此时 Llama-2-7b 模型将消耗 7G 内存，可以和图像生成服务中的 StableDiffusion 模型共存. 且 8bit 精度对大语言模型的生成来说影响不会特别大。

注意

由于 huggingface.io 网站访问受限， Model 标签页的 Download 按钮无法直接下载模型。

扩展知识:

1.Huggingface：开源模型仓库网站， 大模型领域的 github。

2. Transformers 库：最热门的大模型训练/推理框架。

3. 模型精度与需求空间计算：模型单精度为 fp32，即一个参数占据 32bit (4 字节)的空间. 可以经过 GPU 量化降低到 fp16 (2 字节) / int8 (1 字节) 等低精度. Llama-2-7b 模型参数量为 70 亿，那么在 fp32 精度下将需要 70 亿 x4 字节 = 280 亿字节的空间，换算为 26GB. 其他精度以此类推。

1.2 LoRA 模型加载

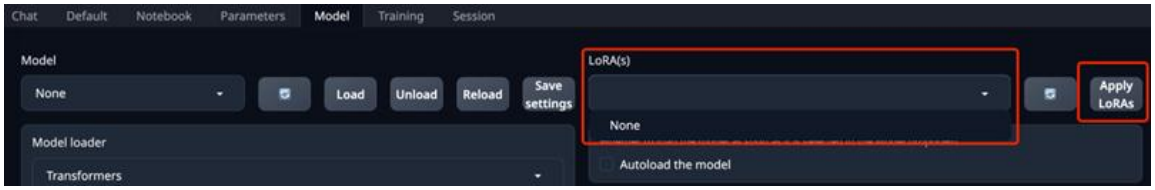
由于基础大模型参数量极多，训练需要的资源量极大，因此基于基础大模型进行微调的小模型应运而生。LoRA 技术便是其中最主流的一种。

LoRA 小模型无法独立使用，需要在加载完基础大模型后再进行加载，对基础大模型的能力进行扩展。

学习机内没有预装 LoRA 小模型，您可以在网上下载其他基于 Llama-2-7b-chat 的 LoRA 模型并放到学习机的 /root/text-generation-webui/loras 目录下. 在 Model 标签页右侧选择 LoRA 模型并点击 Apply LoRAs 进行加载。

例如您可以从此处找到进行了中文微调的 LoRA 模型(也有直接在基础大模型上进行训练得到的新的基础大模型，注意区分)。

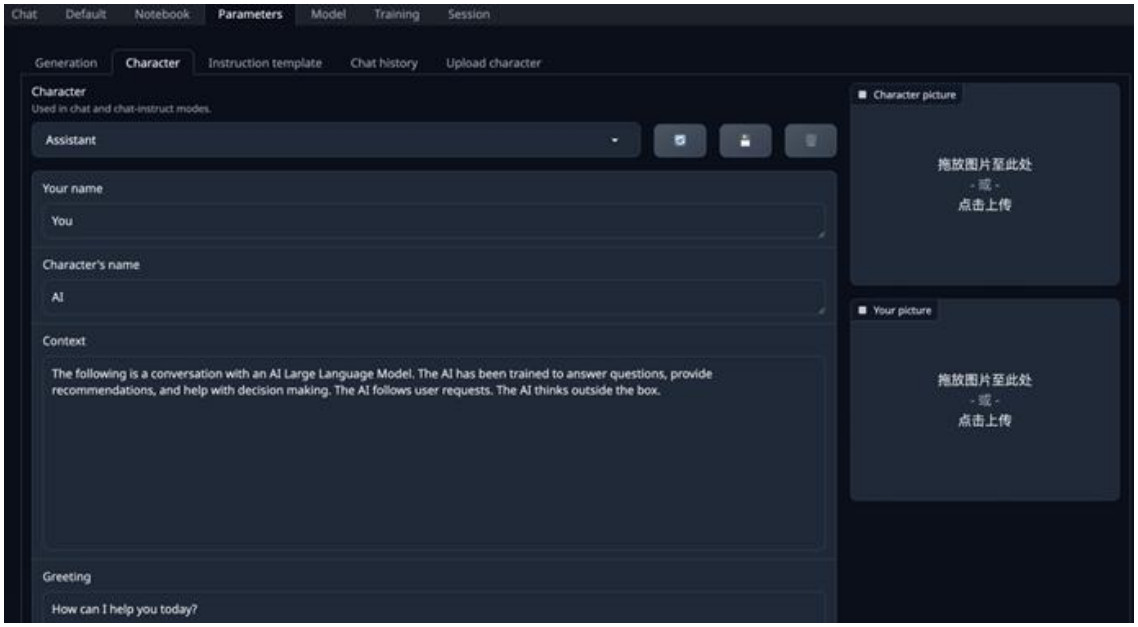




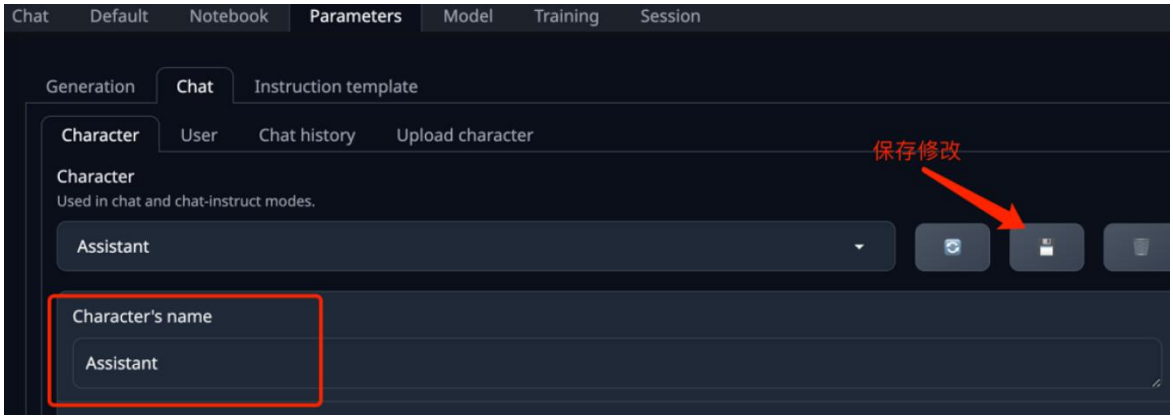
2. 大语言模型使用

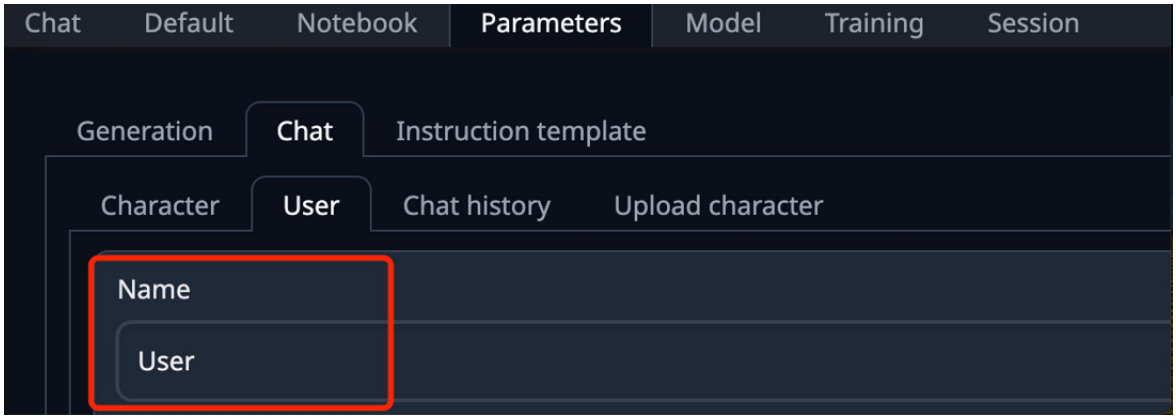
2.1 对话模式

大模型需要在一定的提示词(prompt)下进行生成. 在对话模式下, 框架预设了一些背景知识以控制生成方向。可在 Parameters 标签页的 Character 标签下进行查看和修改。



需要参考下图将 Parameter 页 Chat 标签下默认的 Name 修改为 Character's Name: Assistant; User Name: User. 否则在进行 chat 对话时部分模型可能出现问题。

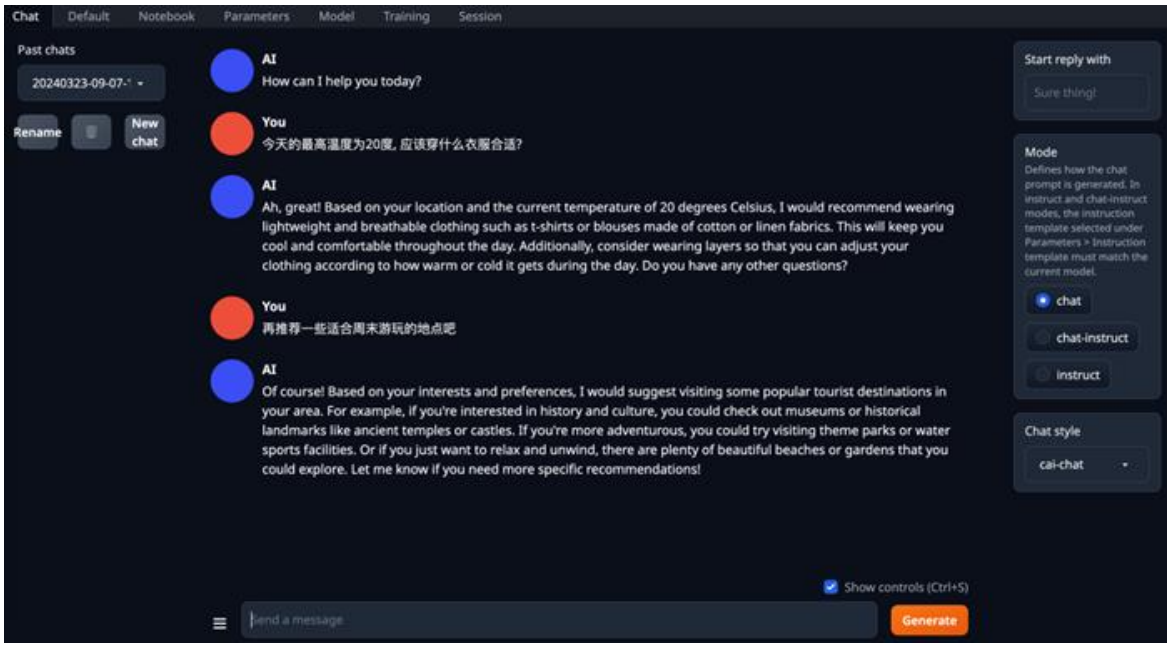




在 Chat 标签页我们可以和大模型展开对话。

注意

原版 Llama-2-7b-chat 模型对中文支持较弱，可以补充下载其他基于 Llama-2-7b-chat 的中文 LoRA 小模型，或换用其他中文的基础大模型，例如电信星辰（telechat）大模型。

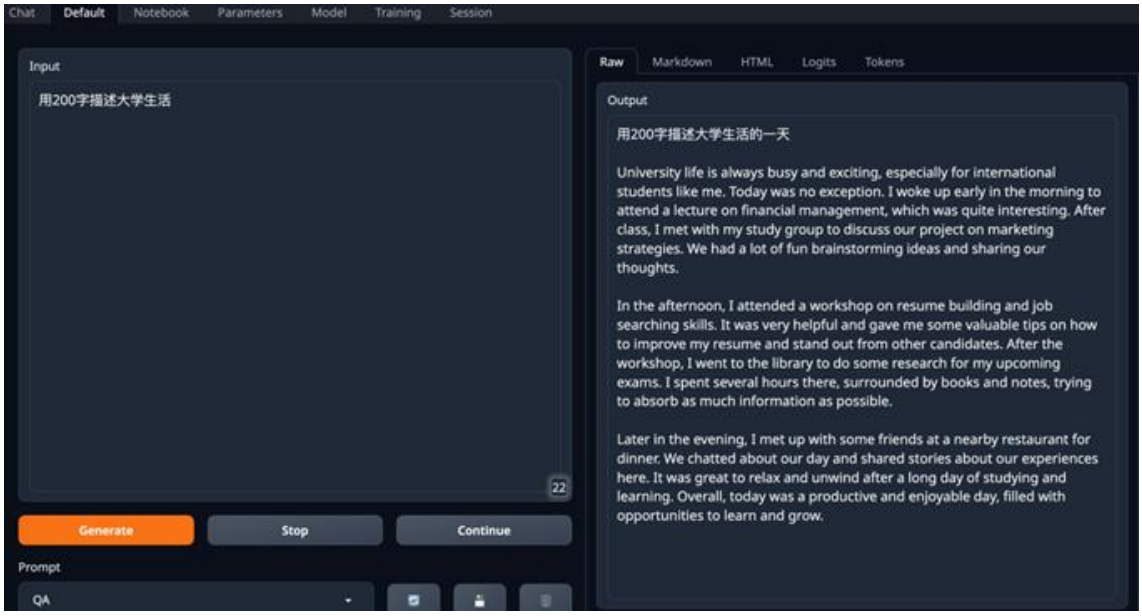


2.2 文本生成模式

在 Default 标签页我们可以利用大模型进行文本生成创作。

注意

原版 Llama-2-7b-chat 模型对中文支持较弱，可以补充下载其他基于 Llama-2-7b-chat 的中文 LoRA 小模型，或换用其他中文的基础大模型，例如电信星辰（telechat）大模型。



2.3 生成方向控制

大语言模型生成方向可以在 Parameters 标签页的 Generation 标签下进行调整，但需要您具有一定的大语言模型背景知识. 这里仅对其中部分基础参数进行介绍。

注意
大模型中 token 的概念： 文本中最小的语义单元. 例如在英文中一个简单单词(love)可能是一个 token， 一个复杂单词(transformer)可能是多个 token 的组合(trans + former)； 在中文中一个字(爱)可能是一个 token， 多个字(爱情)也可能组成一个 token。

- max_new_tokens：生成 token 的最大数量.

- temperature：在生成下一个 token 时，控制各个候选 token 被选择的概率的平滑程度.

  temperature 取值越大，每个候选 token 的概率越相似.
- top_p：在生成下一个 token 时，按候选 token 的概率从大到小计算累计概率，

  达到 p 后丢弃后续的候选 token，控制候选 token 的数量.
- top_k：在生成下一个 token 时，按候选 token 的概率从大到小，只取前 k 个候选 token，控制候选 token 的数量.
- repetition_penalty：重复惩罚，控制生成结果中文字的重复次数.

10.37.8 大模型学习机文本生成模型微调最佳实践

说明
模型训练涉及到较多的专业知识， 这里以Llama 2的LoRA微调为例给出一版示例。

1. 模型微调简介

由于基础大模型参数量极多，训练需要的资源量极大，因此基于基础大模型进行微调的小模型应运而生。 LoRA 技术便是其中最主流的一种。

LoRA 小模型无法独立使用，需要在加载完基础大模型后再进行加载，对基础大模型的能力进行扩展。

已有 LoRA 模型如何加载可参考文本生成模型使用最佳实践。

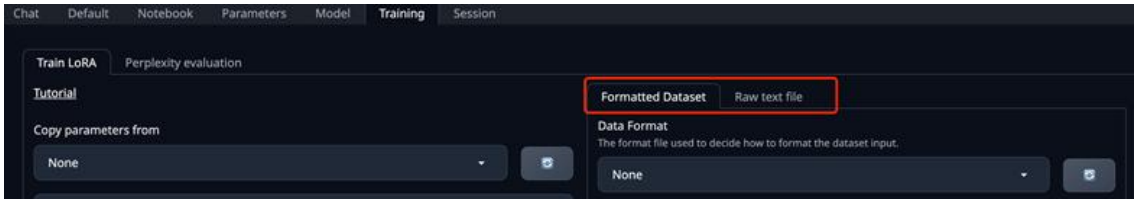
您也可以参考视频教程（访问码：qf8m），帮助您快速学习大模型学习机的使用。

(1) 准备训练数据

支持多种格式，主要分为格式化数据与原始文本数据。

格式化数据以 alpaca-chat-format 为例，数据以 json 方式进行组织，每条数据分为 instruction,input，output 三个部分。

原始文本数据直接将文章或对话的原文作为输入，框架自动进行切分和训练。



下面以普通 txt 文本文件为例，介绍后续训练过程。

(2) 上传数据到云主机

windows 系统使用命令提示符，macos/linux 系统使用终端，执行 scp 命令将数据文件上传到云主机的/root/text-generation-webui/training/datasets 目录下。

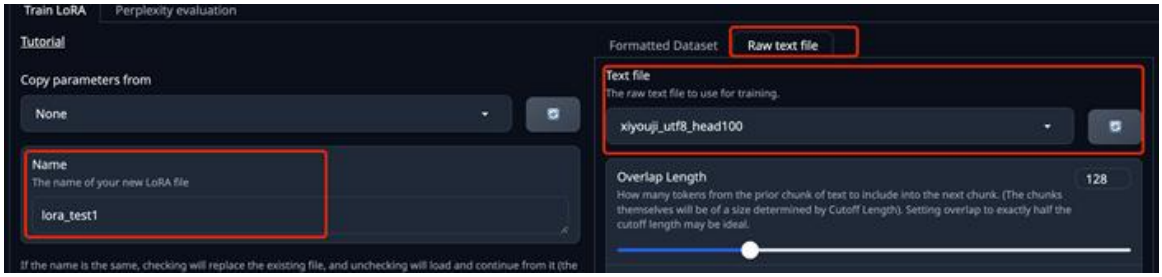
```
scp [本机文件路径]
root@[ip]:/root/text-generation-webui/training/datasets
```

```
night@Night-Macbook Downloads % scp /Users/night/Downloads/xiyouji.txt root@219.152.96.190:/root/text-generation-webui/training/datasets/
root@219.152.96.190's password:
xiyouji.txt                                                    100% 2725KB   7.0MB/s   00:00
```

(3) 训练参数配置

必填：Name，
Text file 选择刚才上传的文件。

其他参数的设定需要有足够的背景知识，这里不做详细介绍，您可自行学习模型训练时的基本参数。



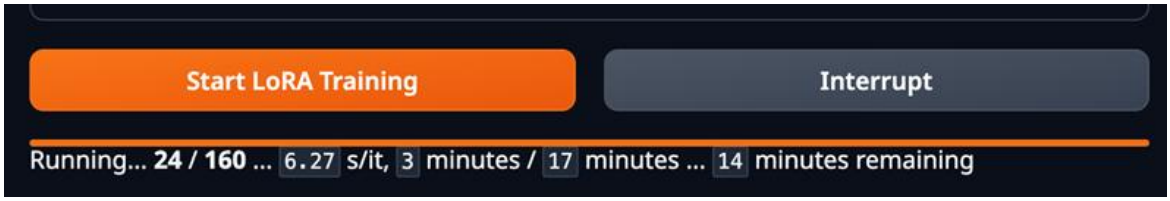
(4) 启动训练

在启动前保证 Model 页已经 Load 过基础模型。

注意

Llama 2-7b 模型的训练默认 fp16 精度大约消耗 18G 显存，入门款 T4 显卡机型无法支持，需要在加载基础模型时选择 load-in-8bit 以降低内存占用。

点击页面右下方 Start LoRA Training 按钮，下方将会出现提示文字，等待模型完成训练与保存。



注意

1. 在此阶段您可能会遇到各类问题导致训练失败， 建议参考大模型学习机服务启停文档中关于打印服务日志的内容。 常见例如：中文不是 utf8 编码，显存不足等。

2. 由于 huggingface.io 网站访问受限，学习机在启动服务时需要检查是否在启动脚本 run.sh 里写入了 export HF_HUB_OFFLINE=1，否则在训练启动和结束时都会自动访问 huggingface 网站导致超时失败。

3.完成训练后的基础模型已经被训练影响，需要重新 reload 一次基础模型以恢复初始状态。

```
root@ecm-4549:~/text-generation-webui# cat run.sh
#!/bin/bash

export PATH=$PATH:/root/python310/bin
export HF_HUB_OFFLINE=1 #can not access to huggingface.io

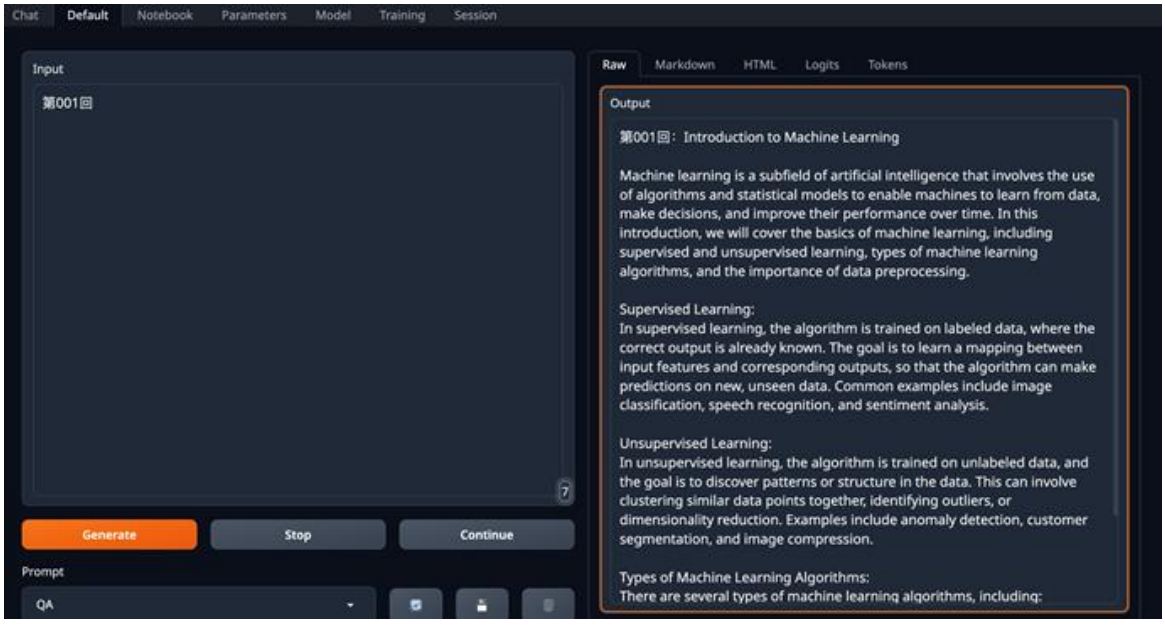
bash start_linux.sh \
    --listen \
    --listen-port 7861 \
    --gradio-auth-path /root/text-generation-webui/gradio-auth.txt
```

(5) 加载微调过的模型

完成训练后，LoRA 小模型会被保存到/root/text-generation-webui/loras 目录下。

回到 Model 页，右侧 LoRA 标签下选择刚才训练出的模型并点击 Apply LoRAs 按钮，在基础大模型之上加载 LoRA 小模型。

本次示例我们采用了西游记原文的前 100 行作为训练数据，在不加载 LoRA 小模型前，生成效果以英文为主。



加载我们训练的 LoRA 小模型后，生成内容风格贴合西游记原文。



10.37.9 使用大模型学习机微调自己的文本生成模型

前提条件

- 大模型学习机预装了 Llama 2-7b-chat 模型与 stable-diffuison-base-v2.1 模型，以及配套的开源服务框架 text-generation-webui 与 stable-diffuison-webui，使您不需单独进行下载模型与配置环境。
- 云主机开通与如何登录网页页面参见部署文档。
- 您也可以参考视频教程（访问码：qf8m），帮助您快速学习大模型学习机的使用。

1. 模型微调简介

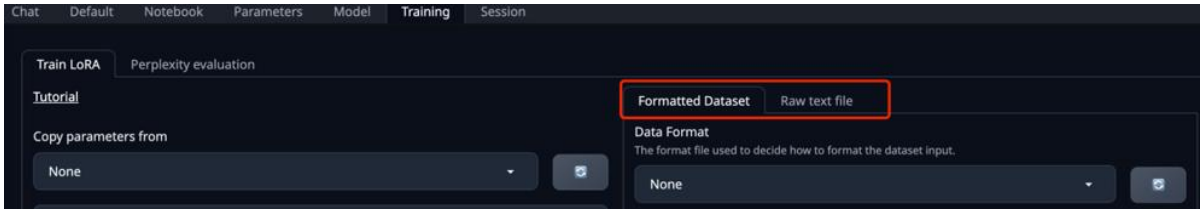
由于基础大模型参数量极多，训练需要的资源量极大，因此基于基础大模型进行微调的小模型应运而生。学习机内框架自带的 LoRA 技术便是其中最主流的一种。

LoRA 小模型无法独立使用，需要在加载完基础大模型后再进行加载，对基础大模型的能力进行扩展。

已有 LoRA 模型如何加载可参考文本生成模型使用最佳实践。

2. 准备训练数据

模型训练支持多种输入格式，主要分为格式化数据(Formatted Dataset)与原始文本数据(Raw text file)。



2.1 格式化数据(Formated Dataset)

格式化数据一般用于对话领域，以 json 形式进行组织，可以在 /root/text-generation-webui/training/formats 目录下找到各类格式的配置文件。以 alpaca-chatbot-format 为例，

每条数据包括 instruction，output 或 instruction，input，output 三个部分。参考示例数据如下：

```
[
    {
        "instruction":"我想基于天翼云云主机通过镜像部署大模型学习机，你能告诉我具体怎么操作吗？",
        "output":"当然可以。首先，您需要登录云主机控制台，并选择创建云主机资源池，然后点击“创建云主机”按钮。"

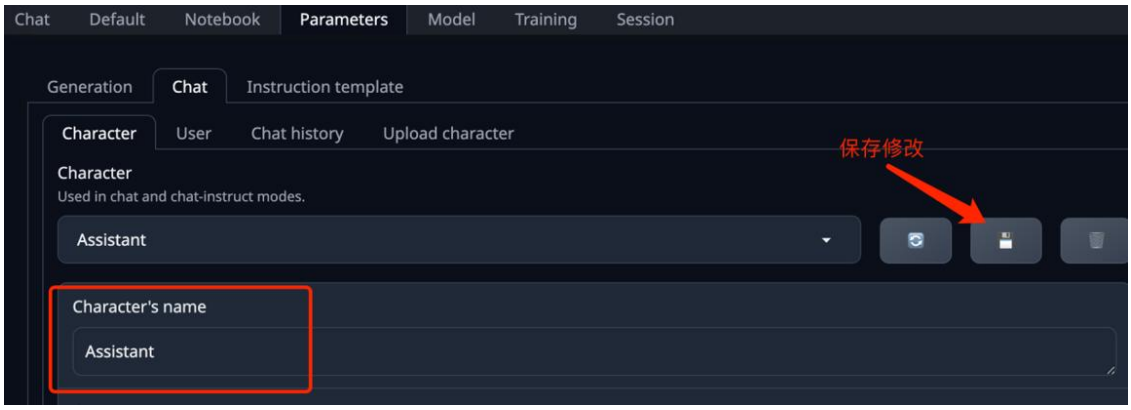
    },
    {
```

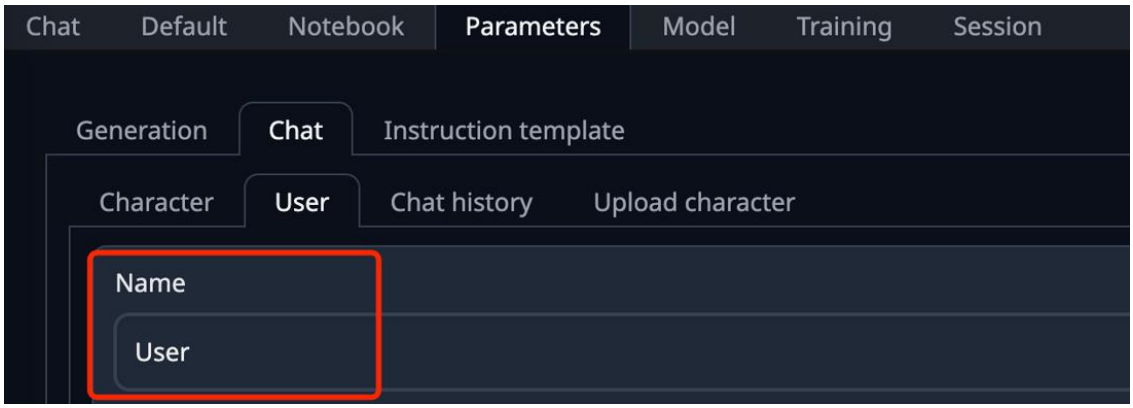
```
  "instruction": "好的，我已经点击了“创建云主机”按钮，接下来应该怎么做？",
        "output":"在当前可用选项中，您可以在重庆-重庆 2 或宁夏-中卫 5 购买 GPU 计算加速型云主机的 Pi7 规格；或者在重庆-重庆 2 购买 GPU 计算加速型云主机的 Pi2 规格。"
    }
]
```

您需要将私有数据组织成上述格式，并保存为[文件名].json 文件. 例如 xuexiji_chatdata.json

参考 /root/text-generation-webui/training/formats/alpaca-chatbot-format.json 文件内容，格式化输入数据将被组织成 User: %instruction%\nAssistant: %output%输入模型，此处的 User 和 Assistant 务必与 Parameter 页面的 Chat 标签下的 Character 和 User 的 Name 保持一致.。对话时的角色标签对大模型来说是一个很强的特征。

需要参考下图将 Parameter 页 Chat 标签下默认的 Name 修改为 Character's Name: Assistant; User Name: User . 否则在进行 chat 对话时小模型几乎完全无效 。





2.2 原始文本数据(Raw text file)

原始文本数据直接将文章或对话的原文作为输入，没有格式转化的过程，框架自动进行切分和训练。

例如我们直接截取大模型学习机部署文档作为输入:

本文基于天翼云云主机，为您介绍通过镜像部署大模型学习机的具体操作。

购买云主机

登录云主机控制台，选择创建云主机资源池，点击“创建云主机”按钮。

当前您可在重庆-重庆 2、宁夏-中卫 5 购买 GPU 计算加速型云主机的 Pi7 规格；也可在重庆-重庆 2 购买 GPU 计算加速型云主机的 Pi2 规格。

您需要将私有数据以 utf8 编码格式写入到文件中，并保存为[文件名].txt 文件，例如 xuexiji_docdata.txt。

3. 数据上传到云主机

windows 系统使用命令提示符，macos/linux 系统使用终端，执行 scp 命令将数据文件上传到云主机的/root/text-generation-webui/training/datasets 目录下。

```
scp [本机文件路径]
root@[ip]:/root/text-generation-webui/training/datasets
```



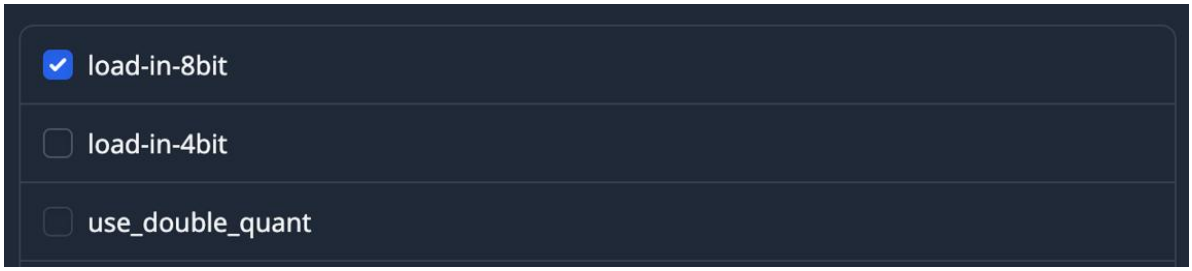
如果您擅长使用 linux 系统，也可采用例如 ftp,lrzsz 等其他方式将数据从本地上传到云主机内。

4. 启动训练

4.1 加载基础大模型

由于小模型需要在基础大模型的基础上才能使用，因此在训练前也要先保证在 Model 页完成了基础大模型的加载。

1. Llama 2-7b 模型在训练时，默认的 fp16 精度大约要消耗 18G 显存，入门款 T4 显卡机型无法支持。需要在加载基础模型时选择 load-in-8bit 以降低内存占用。

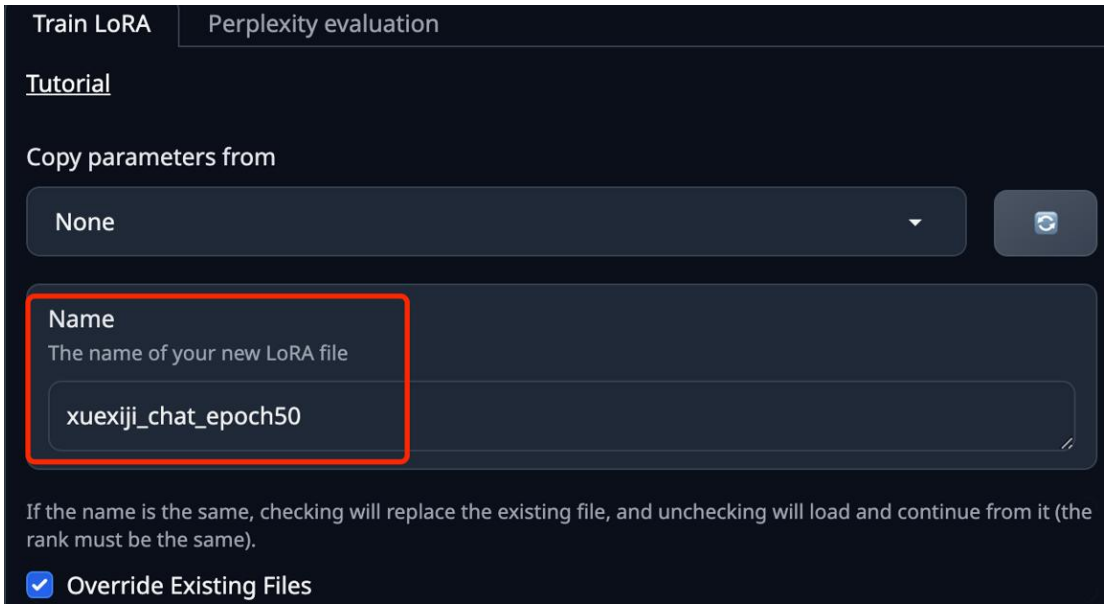


2. 如果刚才进行过训练，则已加载进显存/内存的基础大模型也已经受到了训练的影响，需要先进行模型 reload 恢复初始状态。

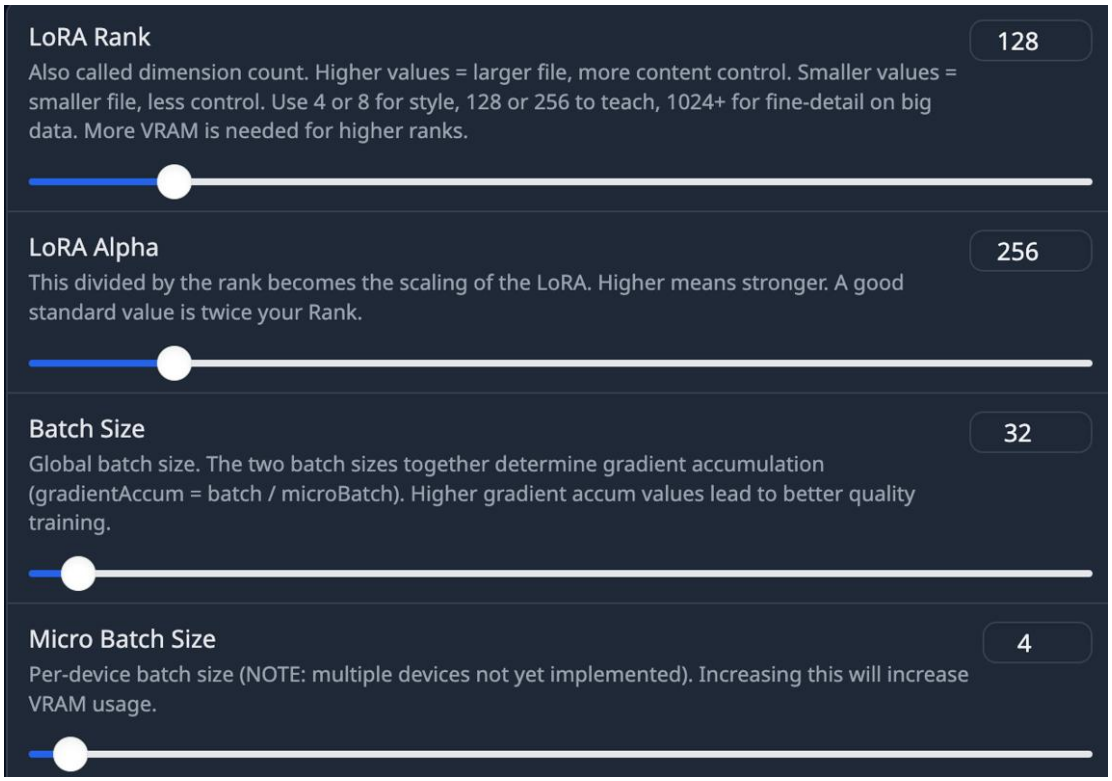
4.2 训练参数配置

在 Trainning 页面进行参数配置。

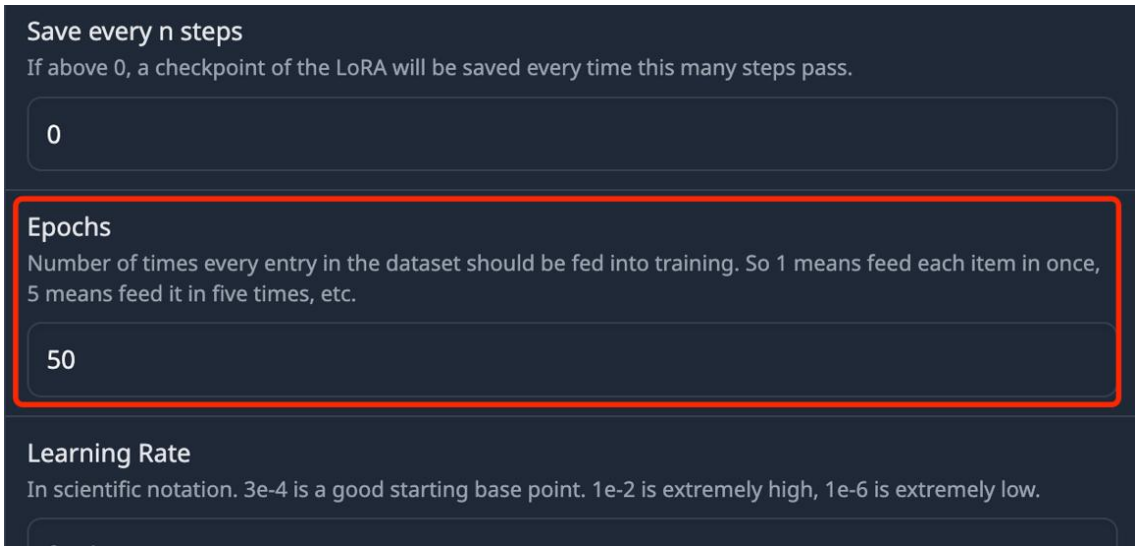
- Name：必填参数，输出模型将以此命名写入到 /root/text-generation-webui/loras 文件夹下。



- LoRA rank：小模型的维度，取值越大学习能力越强(但模型也越大)。可以直接采用默认值 32，如果发现模型效果不佳，可以逐步增加到 128 进行尝试。
- LoRA Alpha：推荐保持为 LoRA rank 的两倍。
- Batch Size：每一批参与训练的数据条数。采用默认值 128 即可。如果您的训练数据条数过少，则也可修改为 64 或 32。



- Epochs：数据训练的重复次数，一般保持 1-3 即可。如果您的训练数据特别少(例如不足 1000 条)，则可以适当增大 epoch 以重复使用训练数据。例如本示例只用了 30 条训练数据，可以使用 epochs=50 或 100 以加强训练效果(实际使用时不推荐过大的 epoch，本示例仅为了保证能够训练出明显效果，已达到过拟合的状态)。



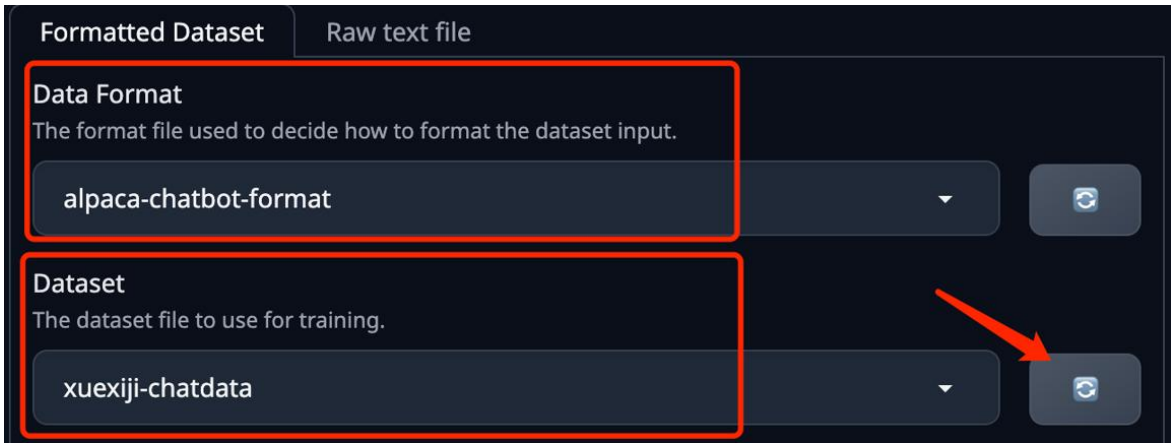
其他参数的设定需要有足够的背景知识，这里不做详细介绍，您可自行学习模型训练时的基本参数。

4.3 加载样本配置

在 Trainning 页面进行样本配置。您可以选择以下的其中一种数据进行训练。

4.3.1 加载格式化数据(Formatted Dataset)

选择 Formatted Dataset，设置 Data Format 和 Dataset 为刚才准备好的数据文件。如果无法查看到文件，尝试点击右侧的刷新按钮重新加载，以及检查文件名后缀是否为.json



4.3.2 加载原始文本数据(Raw text file)

选择 Raw text file，设置 Text file 为刚才准备好的数据文件。如果无法查看到文件，尝试点击右侧的刷新按钮重新加载，以及检查文件名后缀是否为.txt。



4.4 启动训练

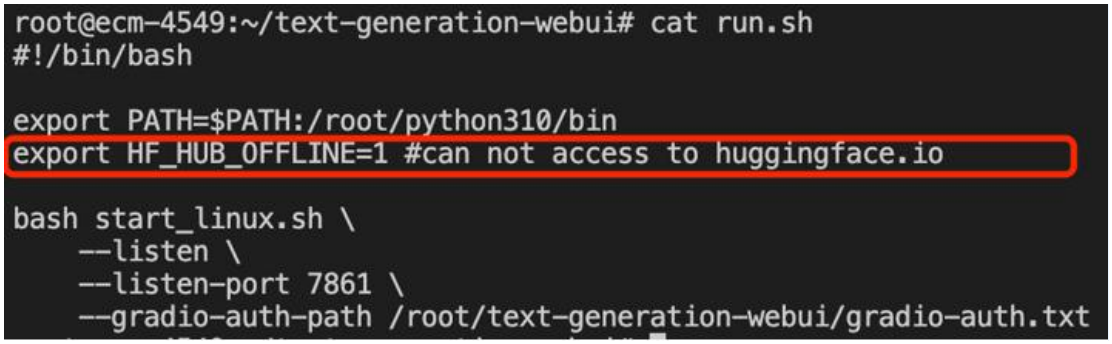
在启动前保证 Model 页已经 Load 过基础模型。

点击页面右下方 Start LoRA Training 按钮，下方将会出现提示文字，等待模型完成训练与保存。



说明

1. 在此阶段您可能会遇到各类问题导致训练失败，建议参考服务启停最佳实践文档[文档链接]中关于打印服务日志的内容。常见例如：中文不是utf8 编码，显存不足等。
2. 在打印的日志中可以查看到模型训练时的 loss 指标。一般 loss 下降到 0.02 以下即为基本训练达标。
3. 由于 huggingface.io 网站访问受限，学习机在启动服务时需要检查是否在启动脚本 run.sh 里写入了 export HF_HUB_OFFLINE=1，否则在训练启动和结束时都会自动访问 huggingface 网站导致超时失败。



4.5 加载微调过的模型

完成训练后，LoRA 小模型会被保存到/root/text-generation-webui/loras 目录下。

注意训练完成后已加载进显存/内存的基础大模型也已经受到了训练的影响，需要先进行模型 reload 恢复初始状态，再进行其他任何操作。

回到 Model 页，右侧 LoRA 标签下选择刚才训练出的模型并点击 Apply LoRAs 按钮，在基础大模型之上加载 LoRA 小模型(如果看不到小模型，尝试点击 Apply LoRAs 按钮左侧的刷新按钮)。

4.6 使用模型示例

在不加载 LoRA 小模型前，生成效果以英文为主，且内容不贴合训练数据。

4.6.1 对话(chat)模式

参见本文 2.1 节的”注意”部分，先确保已将 Parameter 页 Chat 标签下默认的 Name 修改为 Character’s Name: Assistant; User Name: User。否则在进行 chat 对话时小模型几乎完全无效。

加载基于 alpaca-chatbot-format 格式数据训练的 LoRA 小模型后，回答内容贴合训练数据。

注意
大模型生成存在随机性，不一定完全符合预期，您可参考”大模型学习机文本生成最佳实践”文档，调整参数降低模型随机性。 如果始终无法符合预期，则可能是训练 epoch 不足，可以继续加大 epoch 测试。



说明
对话(chat)模式的本质是先将输入数据组织成[role 属性] Assistant: [文本] User: [文本] Assistant: [文本]的格式， 然后再作为 prompt 输入模型。

4.6.2 默认(default)模式

加载基于 raw text file 数据训练的 LoRA 小模型后，生成内容贴合训练数据。

注意
大模型生成存在随机性，不一定完全符合预期，您可参考”大模型学习机文本生成最佳实践”文档，调整参数降低模型随机性。如果始终无法符合预期，则可能是训练 epoch 不足，可以继续加大 epoch 测试。



10.37.10 使用大模型学习机微调自己的图像生成模型

前提条件

- 大模型学习机预装了 Llama 2-7b-chat 模型与 stable-diffuison-base-v2.1 模型，以及配套的开源服务框架 text-generation-webui 与 stable-diffuison-webui,使您不需单独进行下载模型与配置环境。
- 云主机开通与如何登录网页页面参见部署文档。
- 您也可以参考视频教程（访问码：qf8m），帮助您快速学习大模型学习机的使用。

1. 模型微调简介

由于基础大模型参数量极多，训练需要的资源量极大，因此基于基础大模型进行微调的技术应运而生。针对 StableDiffusion 进行微调的技术主流的有 4 种，分别是：

（1）Textual Inversion：不改变原模型结构与参数，只增加新的关键词并训练新的输入 embedding。生成的新模型很小，大约几十 KB。由于没有影响原模型的结构与参数，因此调整能力有限。

（2）Hypernetwork：在原模型的 cross-attention 网络结构中插入一层新网络，与原模型结构串行执行。由于新网络会平均作用在整个模型上，因此更适宜训练绘画风格。模型大小约几十 MB。

（3）LoRA：在原模型的 cross-attention 网络结构中插入多个小网络结构，与原模型结构并行执行。实践证明这种结构的效果要优于 Hypernetwork，因此 LoRA 是当前最主流的微调方案。模型大小约几十 MB。

（4）Dreambooth：对整个原模型的参数进行微调，训练成本高，生成的模型相当于一个新的基础大模型。模型大小约几 GB。

已有 LoRA 模型如何加载可参考图像生成模型使用最佳实践。

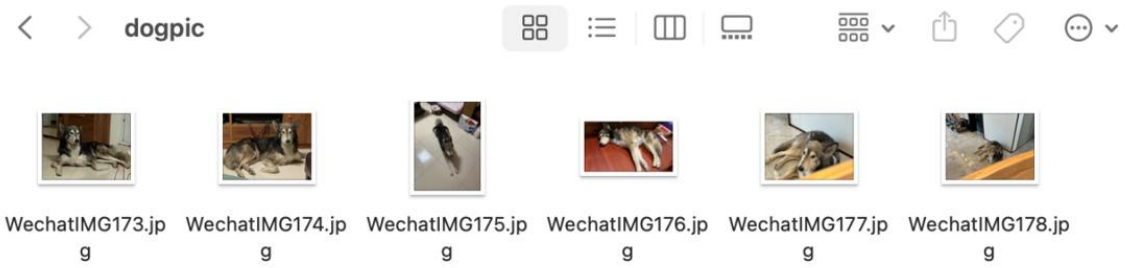
本文以 Textual Inversion 方式为例，介绍通过私有数据微调模型并进行图像生成的流程。如果您有微调新的 LoRA 模型的需求，本文也给出了相关插件的安装说明。

2. 准备训练数据

2.1 图片文件准备

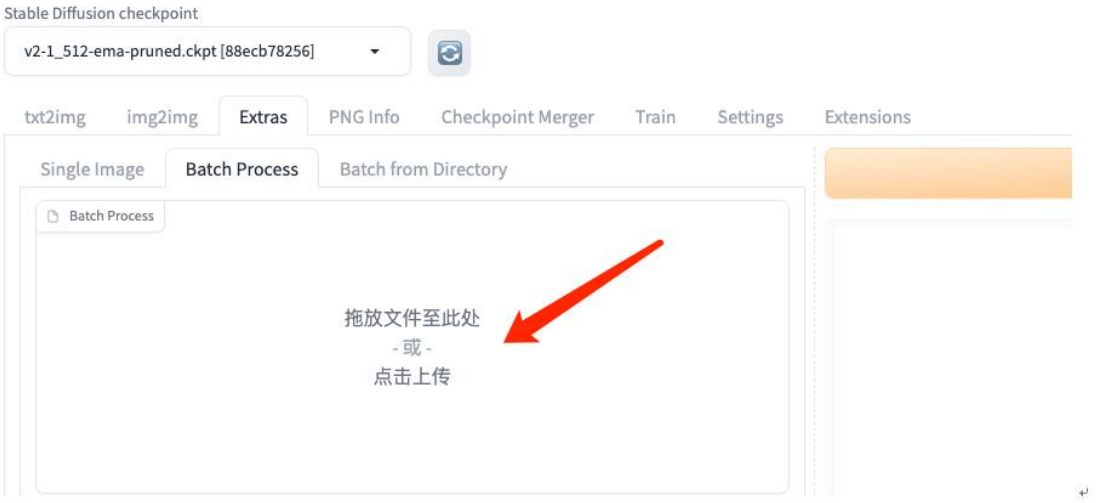
准备一批用于训练的图片，可以是同一人物/物体的不同视角，也可以是同一种绘画风格。

本示例中采用了同一只狗狗的不同视角图片。



2.2 图片文件上传与裁切

在 Extras 页面的 Batch Process 下点击文字，选择图片文件。



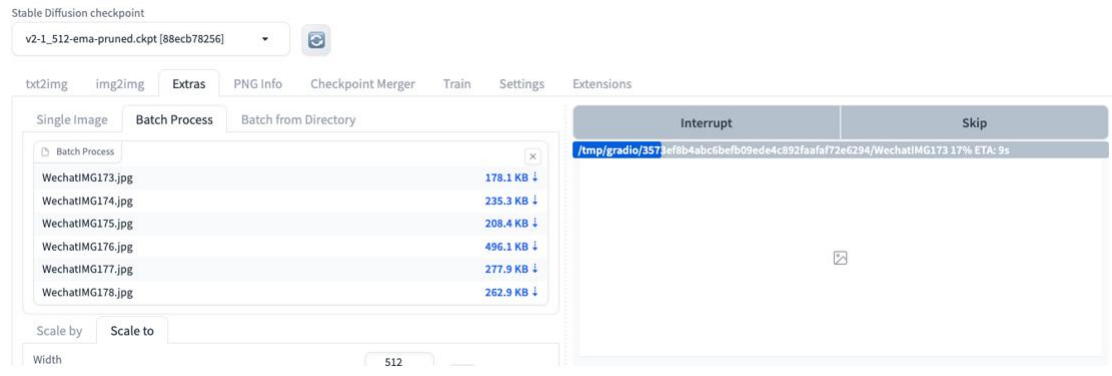
选择 Scale to 将图片裁切到 512x512 的大小。



点击 generate 按钮，等待图片上传完毕。

注意

目前某些镜像中尚未预装 deepbooru 功能所需的模型。当您首次上传图片时，系统会访问 github 进行模型下载，该模型大小为 614MB。依据您的网络带宽，这可能需要几分钟的时间。



上传完成的图片在云主机

的 /root/stable-diffusion-webui/output/extras-images 目录下。

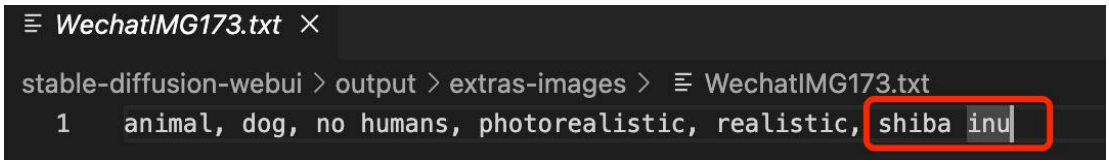
```
root@ecm-4549:~/stable-diffusion-webui/output/extras-images# l
total 11600
drwxr-xr-x 2 root root    4096 Apr  7 10:44 ./
drwxr-xr-x 3 root root    4096 Apr  7 10:44 ../
-rw-r--r-- 1 root root 1494090 Apr  7 10:44 WechatIMG173.png
-rw-r--r-- 1 root root      60 Apr  7 10:44 WechatIMG173.txt
-rw-r--r-- 1 root root 1897223 Apr  7 10:44 WechatIMG174.png
-rw-r--r-- 1 root root     113 Apr  7 10:44 WechatIMG174.txt
-rw-r--r-- 1 root root 1830959 Apr  7 10:44 WechatIMG175.png
-rw-r--r-- 1 root root      68 Apr  7 10:44 WechatIMG175.txt
-rw-r--r-- 1 root root 2092483 Apr  7 10:44 WechatIMG176.png
-rw-r--r-- 1 root root      77 Apr  7 10:44 WechatIMG176.txt
-rw-r--r-- 1 root root 2303343 Apr  7 10:44 WechatIMG177.png
-rw-r--r-- 1 root root      92 Apr  7 10:44 WechatIMG177.txt
-rw-r--r-- 1 root root 2213769 Apr  7 10:44 WechatIMG178.png
```

建议您将文件通过 mv 命令移动到其他文件夹，方便区分不同的训练文件。例如:

```
cd /root/stable-diffusion-webuimkdir -p train/picdatalmv
output/extras-images/* train/picdatal
```

您还可查看每张图片生成的 txt 文件,检查反推出来的提示词是否符合预期,并删除识别错误的提示词。

例如本示例中模式识别的提示词中有一个多余的”shiba inu”，需要将其删除。



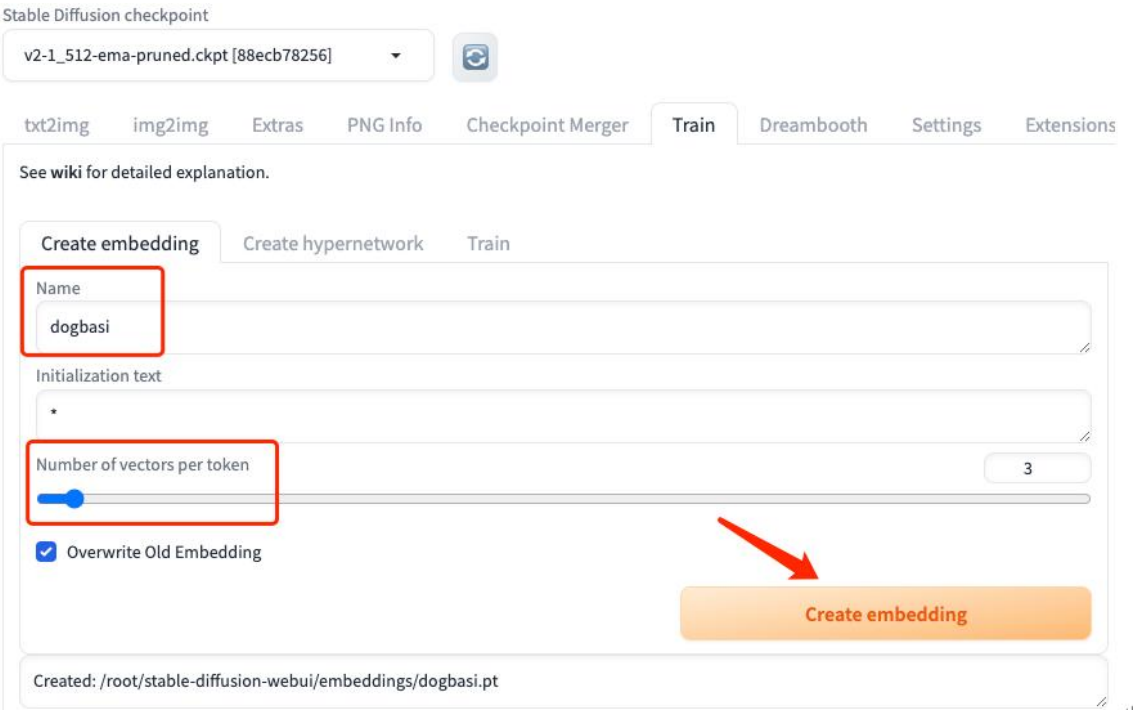
3 启动训练

stable-diffusion-webui 框架在启动时默认已经加载好了基础大模型,无需做额外操作。

3.1 生成初始模型文件

首先需要在 Train 页面的 Create embedding 标签下，创建一个空的初始模型文件。

- Name: 小模型的名称，也是未来使用小模型时需要用到的专有提示词名称。例如我们为训练图片中的狗狗取名字叫”八四”，因此模型的 Name 可以叫”dogbasi”，未来使用模型时也需要用”dogbasi”触发。注意这里的名称必须是特有的，避免和模型已有的关键词重复。
- Number of vectors per token: 小模型参数的维度，数值越大模型拟合能力越强，但也可能导致拟合过强。推荐设置在 3-7 左右。如果是学习绘画风格，一般设置尽量小一些；如果是学习具体物体，可以设置大一些。

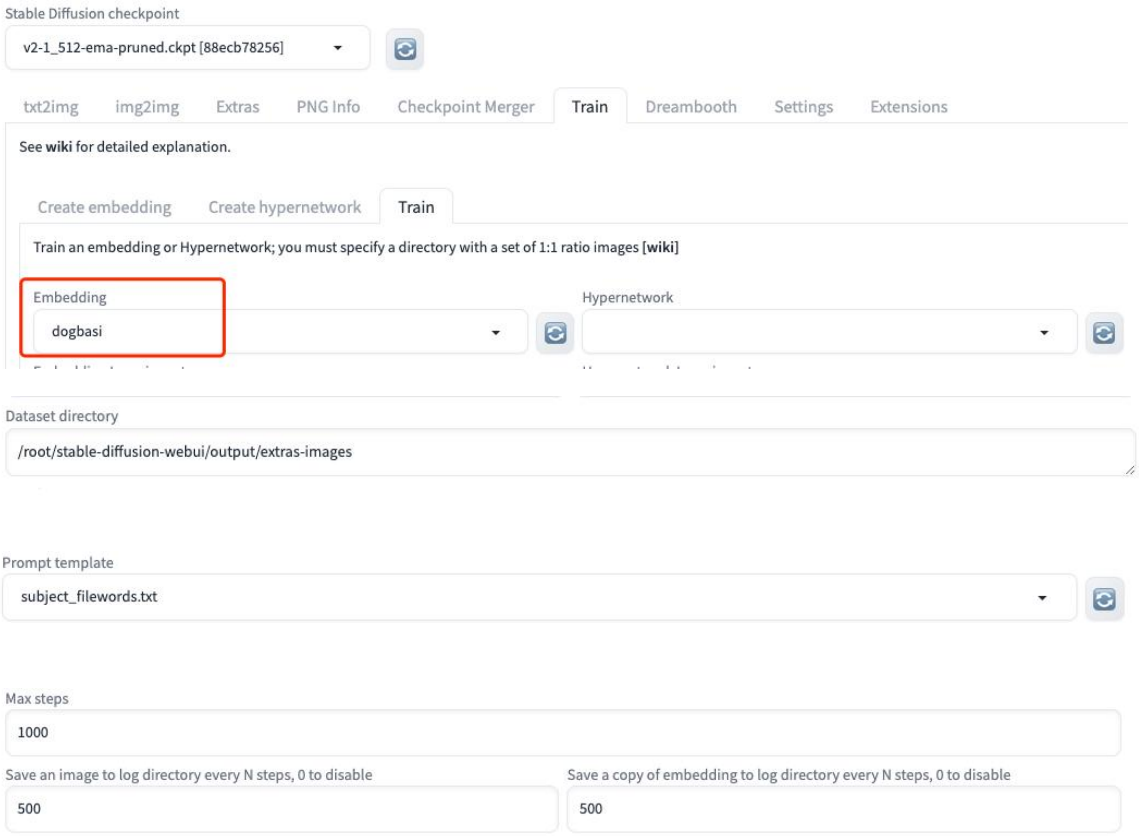


3.2 调整训练配置

在 Train 页面的 Train 标签页，修改如下设置：

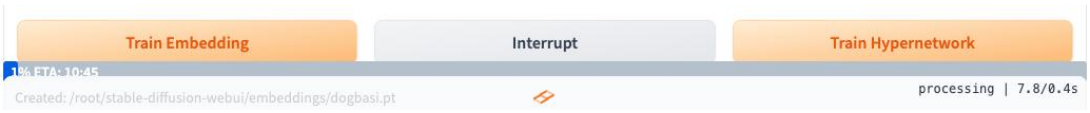
- Embedding: 选择 3.1 步骤中创建过的初始 embedding 模型。
- Dataset directory: 填写 2.2 步骤中上传与调整过的训练图片目录。
- Prompt templete: 如果期望训练绘画风格，选择 style_filewords.txt；如果期望训练具体物体，选择 subject_filewords.txt。

- Max Steps: 训练的总轮次，一般在图片数量不大的情况下(10 张图左右)，推荐设置为 1000 及以上，以保证训练质量。



3.3 启动训练

点击页面底部 Train Embedding 按钮开始训练，等待进度条结束。

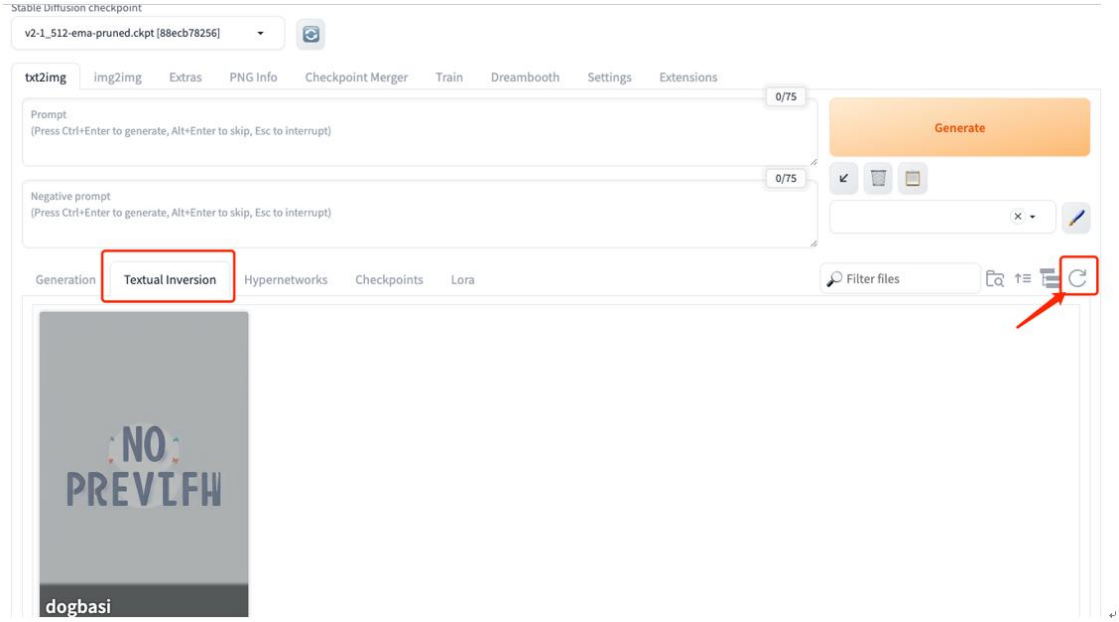


注意

1. 在此阶段您可能会遇到各类问题导致训练失败，建议参考服务启停最佳实践文档[文档链接]中关于打印服务日志的内容。常见例如: 显存不足等。
2. 在 batch size 默认设置为 1 的情况下，训练需要消耗 6G 显存。

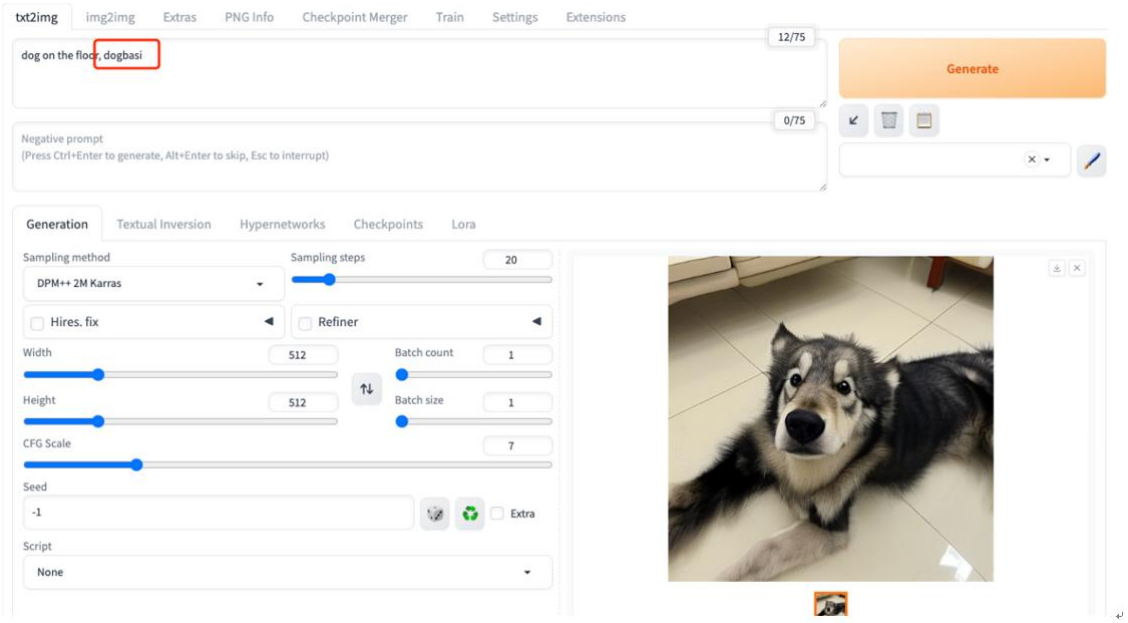
3.4 加载微调过的模型

在 txt2img 页面的 Textual inversion 标签下，点击右侧的刷新按钮，如果看到新训的小模型即代表成功。



3.5 使用模型示例

在提示词 prompt 中附加上微调过的小模型名称，即可生成类似训练数据的新图片。如果发现图片质量不佳，可以尝试加大训练时的 Max Steps，或者加大 Create embedding 时的 Number of vectors per token，或者准备更多高质量的训练图片（数据最关键）。



4. 微调新 LoRA 模型插件安装

stable-diffusion-webui 本体没有微调 LoRA 模型的功能，需要安装 Dreambooth 插件。下面将以此为例介绍如何在本框架中安装插件。

stable-diffusion-webui 本身提供了自动下载和安装插件的功能（见 extensions 页），但由于 github 和 pip 库访问受限，自动下载无法成功。因此本文采用了手动安装的方式。

手动安装需要您通过 ssh 或远程登录等方式登录到云主机系统内，您可以参考[修改大模型学习机登录密码]的相关内容。

同时，由于涉及到服务运行环境的修改，因此所有操作都将在 docker 容器内执行 。进入容器的方式参考：

```
# 检查容器状态

docker ps -a# 进入容器

docker exec -it stable_diffusion bash
```

注意

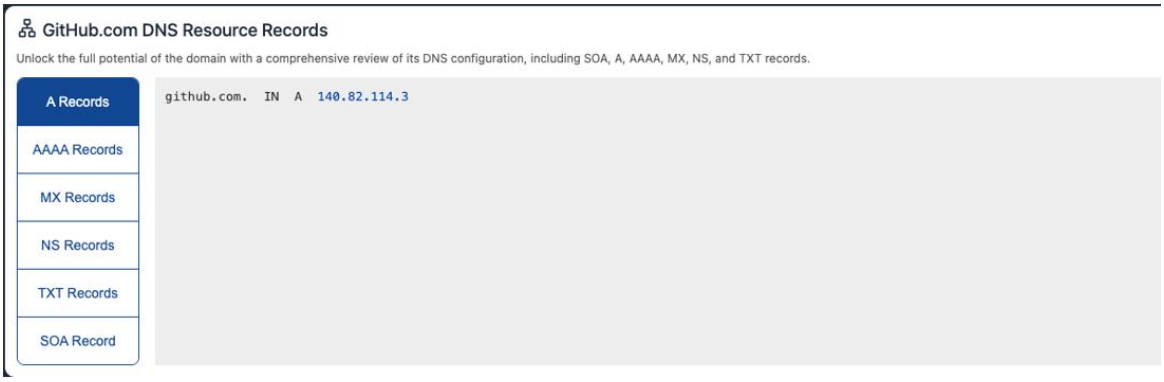
执行完进入容器的命令后，命令行界面并不会有明显变化，您可以通过执行命令 docker ps 是否能成功，来判断当前在容器内还是云主机内。

4.1 配置 github 地址（可选）

如果您在执行 git clone 时始终因为网络问题无法成功，则可参考本节内容。

(1) 获取可用的 ip 地址

通过 https://sites.ipaddress.com/github.com/ 获取到最新的真实 ip 地址（网站上广告太多，注意甄别，DNS 信息在页面中比较靠后的位置）。



注意

1. IP 地址并不会一直固定，未来如无法访问 github，您需要重新确认最新的 IP 地址

2. 可以使用代理服务器彻底解决问题，但注意如果设置了 http_proxy 系统变量，需要在使用完毕后取消设置，或者在/root/stable-diffusion-webui/run.sh 启动脚本中增加一项`--share \`，否则 webui 服务会报错无法访问 localhost。

(2) 修改 hosts 文件

请确保当前在 docker 容器内。

通过 vim 编辑器修改/etc/hosts 文件，添加如下内容:

```
# 从 ipaddress 查到的真实 ip 地址 140.82.114.3      github.com
```

完成配置后，访问 github 相关链接成功概率提升（但不保证一定成功）。

4.2 下载插件文件

请确保当前在 docker 容器内。

进入到/root/stable-diffusion-webui/extensions 目录下，并下载 dreambooth 插件。

```
cd /root/stable-diffusion-webui/extensions

git clone https://github.com/d8ahazard/sd_dreambooth_extension.git
```

遇到连接失败可以多重试几次，如果始终失败，则参考 4.1 节的内容手动配置 github 地址。

4.3 安装插件需要的依赖库

请确保当前在 docker 容器内。

stable-diffusion-webui 采用了 python venv 的方式进行环境隔离，首先需要进入到虚拟环境。

```
cd /root/stable-diffusion-webuisource venv/bin/activate
```

执行成功后，命令行界面开头将出现(venv)标识。

```
root@ecm-4549:~/stable-diffusion-webui/extensions# cd /root/stable-diffusion-webui/
root@ecm-4549:~/stable-diffusion-webui# source venv/bin/activate
(venv) root@ecm-4549:~/stable-diffusion-webui#
```

然后在插件文件夹内安装依赖库。由于 pip 库访问受限问题，这里采用了清华的镜像源。

```
cd /root/stable-diffusion-webui/extensions/sd_dreambooth_extension
```

```
pip install -r requirements.txt -i
https://pypi.tuna.tsinghua.edu.cn/simple
```

安装成功后将出现如下提示：

```
Installing collected packages: dadaptation, gitpython, bitsandbytes
  Attempting uninstall: gitpython
    Found existing installation: GitPython 3.1.32
    Uninstalling GitPython-3.1.32:
      Successfully uninstalled GitPython-3.1.32
Successfully installed bitsandbytes-0.43.0 dadaptation-3.2 gitpython-3.1.43

[notice] A new release of pip available: 22.2.1 -> 24.0
[notice] To update, run: pip install --upgrade pip
```

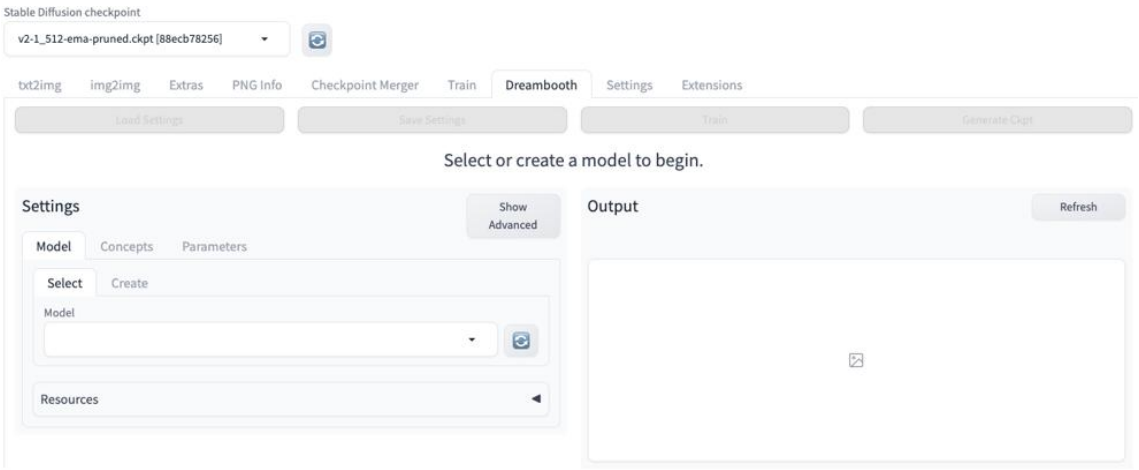
4.4 重启服务以应用插件

如果您当前在 docker 容器环境中，可直接执行 exit 命令退出到云主机内。参考[大模型学习机服务启停最佳实践]文档，重启 stable-diffusion-webui 服务：

```
sh /root/restart_sd_webui.sh
```

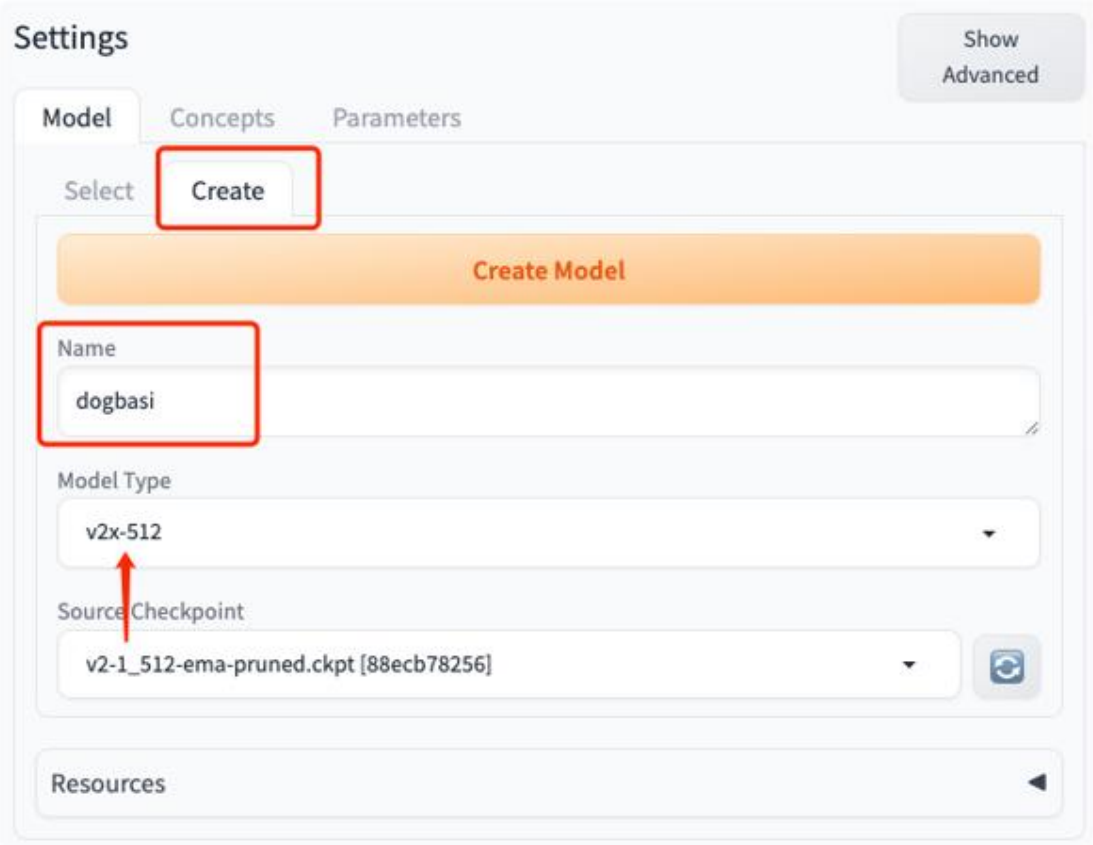
4.5 确认插件安装完毕

重新登录页面后，应多出 Dreambooth 页面：



Dreambooth 插件的使用方式和上文训练 Textual Inversion 是一致的。

（1）在 Create 页面创建一个空模型。内容参考下图。



（2）首次创建时，插件会从 github 额外下载一些辅助模型，大小约几百 MB。建议您提前自行下载并放到/root/stable-diffusion-webui/models/GFPGAN 目录下，避免等待太长时间。

```
cd /root/stable-diffusion-webui/models/GFPGAN

wget

https://github.com/xinntao/facexlib/releases/download/v0.1.0/detection_Resnet50_Final.pth

wget

https://github.com/xinntao/facexlib/releases/download/v0.2.2/parsing_parsenet.pth
```

（3）插件还需要从 huggingface 网站下载 StableDiffusion 模型的部分内容，因此需要删除/root/stable-diffusion-webui/run.sh 脚本中的内容：

```
# 删除这一行可恢复对 huggingface 的连接 export HF_HUB_OFFLINE=1 #can not access to

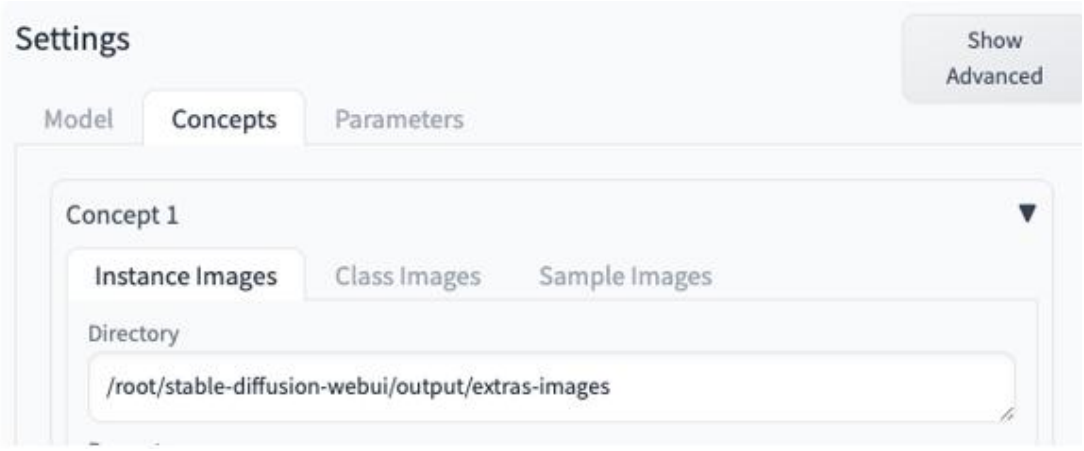
huggingface.io
```

参考 4.1 节对 github 真实 ip 的配置，也需要对 huggingface.co 进行相关配置，否则无法访问。

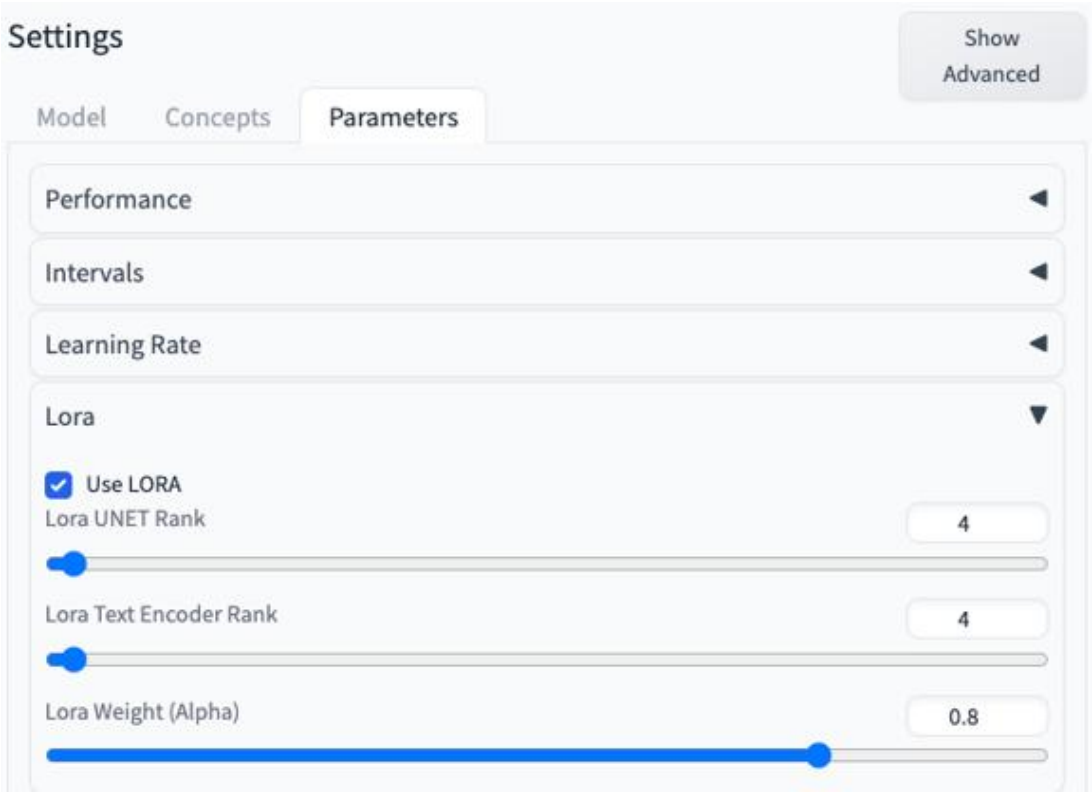
下载模型数据需要占据大量磁盘空间，建议额外购买 50G-100G 大小的数据盘，并挂载到/root/stable-diffusion-webui/models/dreambooth 目录下（也可先将当前 model 文件夹改名，然后直接将数据盘挂载为新的 model 文件夹，最后将老数据 mv 进去）。

说明
即使配置了真实 ip，在访问 huggingface.co 时也依然会失败，建议您使用代理服务器进行连接。配置代理服务器和 http_proxy 环境变量后，需要在/root/stable-diffusion-webui/run.sh 脚本中增加 `--share \`，否则 webui 服务会报错无法访问 localhost。

（4）在 Concepts 页面设置样本路径。



（5）在 Parameters 页开启 LoRA 训练。



其他训练相关的参数您可自行查看和调整（例如训练轮次，学习率等），本文不再一一赘述。

10.38AIGC 实践

10.38.1 在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本

什么是 DeepSeek

DeepSeek 是一个基于 Transformer 架构的大型语言模型（LLM），由深度求索（DeepSeek）公司开发。它能够处理自然语言理解、生成、翻译、问答等多种任务。在目前大模型主流榜单中，DeepSeek-V3 在开源模型中位列榜首，与世界上最先进的闭源模型不分伯仲。在对话模型典型任务方面的评测效果如下：

	Benchmark (Metric)	DeepSeek V3	DeepSeek V2.5 0905	Qwen2.5 72B-Inst	Llama3.1 405B-Inst	Claude-3.5 Sonnet-1022	GPT-4o 0513
	Architecture	MoE	MoE	Dense	Dense	-	-
	# Activated Params	37B	21B	72B	405B	-	-
	# Total Params	671B	236B	72B	405B	-	-
English	MMLU (EM)	88.5	80.6	85.3	**88.6**	88.3	87.2
	MMLU-Redux (EM)	**89.1**	80.3	85.6	86.2	88.9	88.0
	MMLU-Pro (EM)	75.9	66.2	71.6	73.3	**78.0**	72.6
	DROP (3-shot F1)	**91.6**	87.8	76.7	88.7	88.3	83.7
	IF-Eval (Prompt Strict)	86.1	80.6	84.1	86.0	**86.5**	84.3
	GPQA-Diamond (Pass@1)	59.1	41.3	49.0	51.1	**65.0**	49.9
	SimpleQA (Correct)	24.9	10.2	9.1	17.1	28.4	**38.2**
	FRAMES (Acc.)	73.3	65.4	69.8	70.0	72.5	**80.5**
	LongBench v2 (Acc.)	**48.7**	35.4	39.4	36.1	41.0	48.1
Code	HumanEval-Mul (Pass@1)	**82.6**	77.4	77.3	77.2	81.7	80.5
	LiveCodeBench (Pass@1-COT)	**40.5**	29.2	31.1	28.4	36.3	33.4
	LiveCodeBench (Pass@1)	**37.6**	28.4	28.7	30.1	32.8	34.2
	Codeforces (Percentile)	**51.6**	35.6	24.8	25.3	20.3	23.6
	SWE Verified (Resolved)	42.0	22.6	23.8	24.5	**50.8**	38.8
	Aider-Edit (Acc.)	79.7	71.6	65.4	63.9	**84.2**	72.9
	Aider-Polyglot (Acc.)	**49.6**	18.2	7.6	5.8	45.3	16.0
Math	AIME 2024 (Pass@1)	**39.2**	16.7	23.3	23.3	16.0	9.3
	MATH-500 (EM)	**90.2**	74.7	80.0	73.8	78.3	74.6
	CNMO 2024 (Pass@1)	**43.2**	10.8	15.9	6.8	13.1	10.8
Chinese	CLUEWSC (EM)	90.9	90.4	**91.4**	84.7	85.4	87.9
	C-Eval (EM)	**86.5**	79.5	86.1	61.5	76.7	76.0
	C-SimpleQA (Correct)	**64.1**	54.1	48.4	50.4	51.3	59.3

当前 DeepSeek 模型有多个版本，参数量从 1.5B、70B 到 671B 不等，适用于不同的应用场景和计算资源。参数量越多，资源消耗越多，使用者可以根据自己实际需求选择不同模型版本。

注意

- 为维护良好的网络环境和公共秩序，禁止任何用户或机构利用天翼云平台从事违法活动、传播不良信息或实施违反社会公序良俗的行为。若发生此类行为，天翼云将依法采取必要措施，停止相关服务功能，并依法追究责任。
- 模型由 DeepSeek 公司提供，模型的结果仅供参考，并可能因不同的环境、数据或操作条件而有所不同，我们不对模型的任何结果负责。

快速体验 DeepSeek

Ollama 是一个专为本地运行大语言模型（LLMs）而设计的工具，支持多种模型格式，并提供了简单易用的命令行接口。这里，我们以 DeepSeek-r1-7b 模型为例，讲解如何在天翼云使用 Ollama 运行 DeepSeek 模型，读者可以根据自己实际需求修改云主机模型和云主机配置。

步骤一：创建 GPU 云主机

1. 进入创建云主机页面

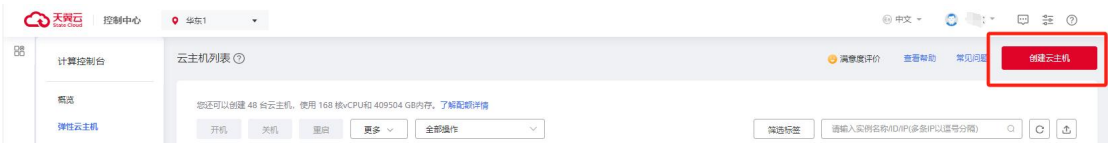
a. 点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



b. 单击“产品服务列表>弹性云主机”，进入主机列表页。



c. 单击“创建云主机”，进入弹性云主机创建页。



2. 进行基础配置

a. 根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、“实例名称”、“主机名称”等。

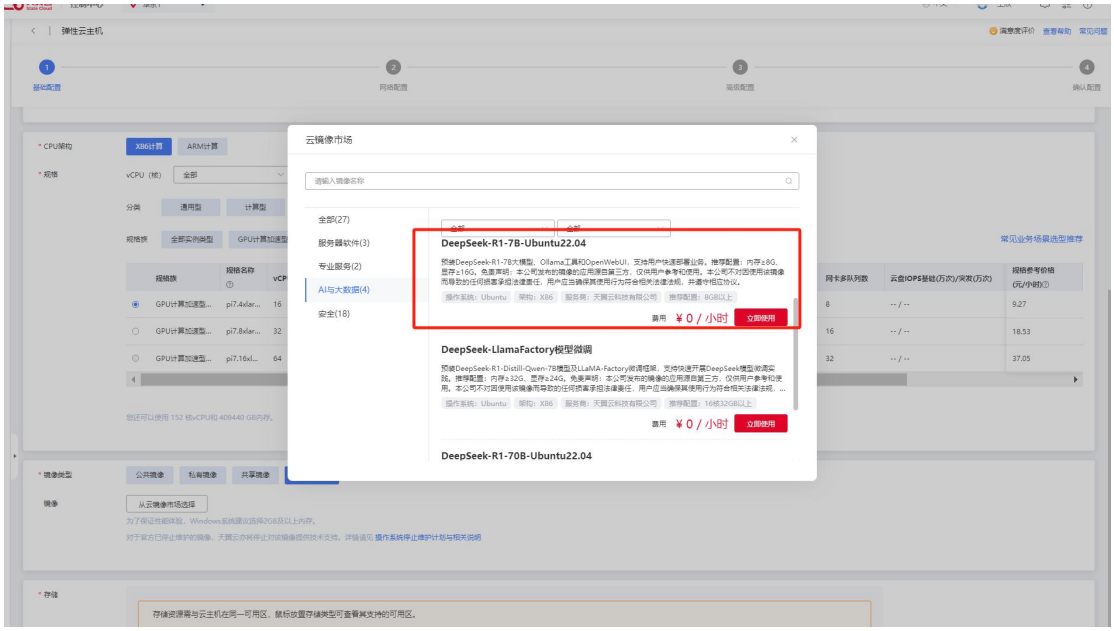
b. 选择规格。此处选择"CPU 架构"为"X86"、"分类"为"GPU 加速/AI 加速型"、"

规格族"为"GPU 计算加速型 pi7"、"规格"为"pi7.4xlarge.4"。

c. 选择镜像。"镜像类型"选择"镜像市场",在云镜像市场中选择预置了 DeepSeek R1 模型的 DeepSeek-R1-7B-Ubuntu22.04 镜像。

注意

本镜像推荐配置：内存≥8G、显存≥16G。



注意

目前提供了预装多种量化精度、多种参数量的模型，如您有需求，也可在云镜像市场中进行选择。

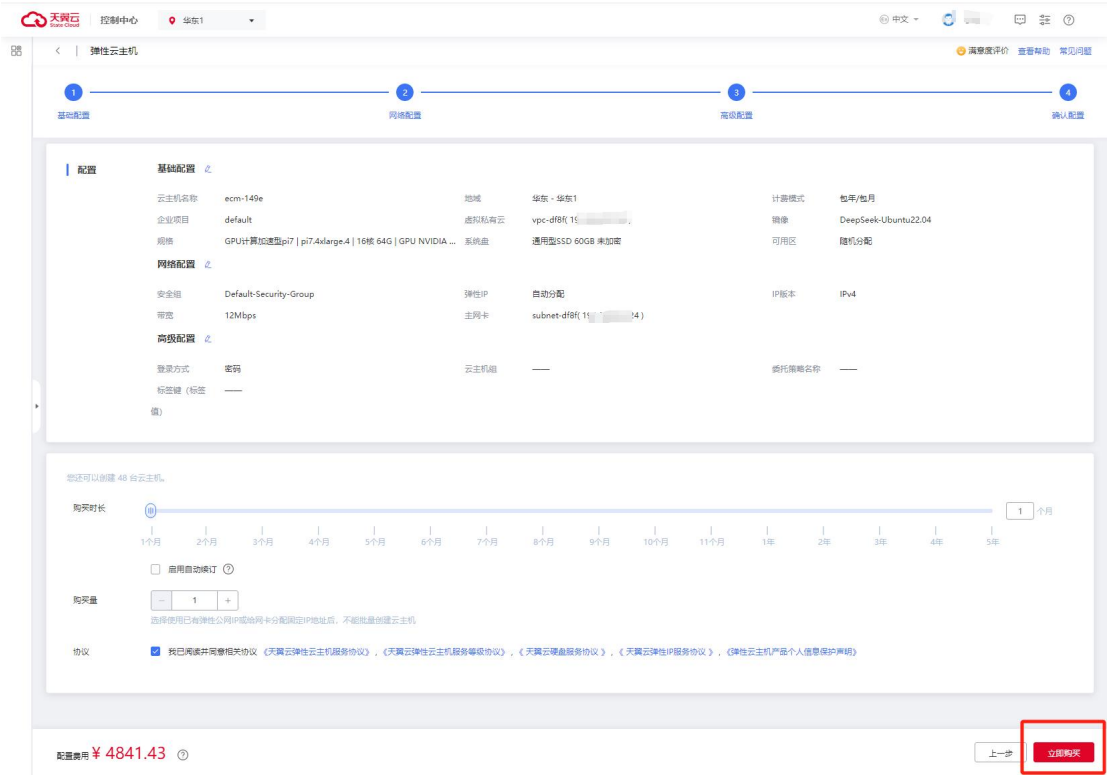
d. 设置云盘类型和大小。



3. 网络及高级配置

设置网络，包括"网卡"、"安全组"，同时配备 "弹性 IP" 用于下载和访问模型；设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。

4. 确认并支付



步骤二：使用 DeepSeek 模型

1. 通过 web 界面进行模型交互

注意

镜像自带的 ollama 工具监听 127.0.0.1:11434、webui 监听 0.0.0.0:3000 端口，云主机默认不对外开放任何端口访问，请按需开放端口访问规则，避免数据泄露。

a. 放行云主机安全组的 3000 端口。具体操作请参考添加安全组规则-弹性云主机-用户指南-安全-安全组-配置安全组规则 - 天翼云。



b. 访问DeepSeek模型的可视化界面。登录地址为:http://{公网ip地址}:3000。

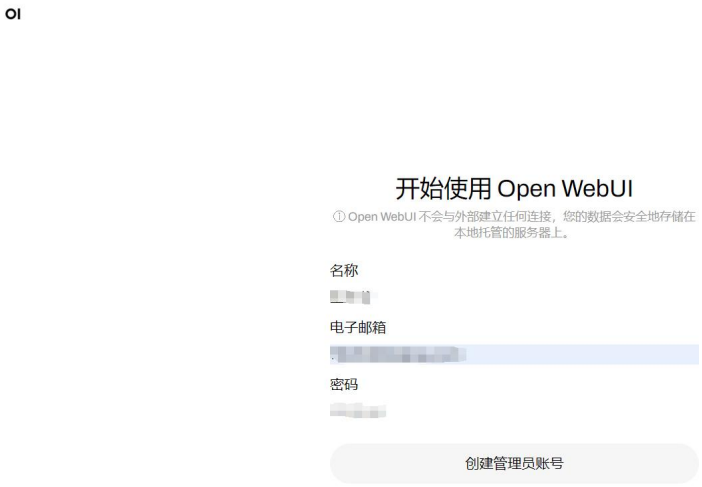
注意

云主机全自动安装 DeepSeek 模型和可视化界面，请等待云主机启动 5 分钟后，再访问登录界面。

首次登录页面如下：

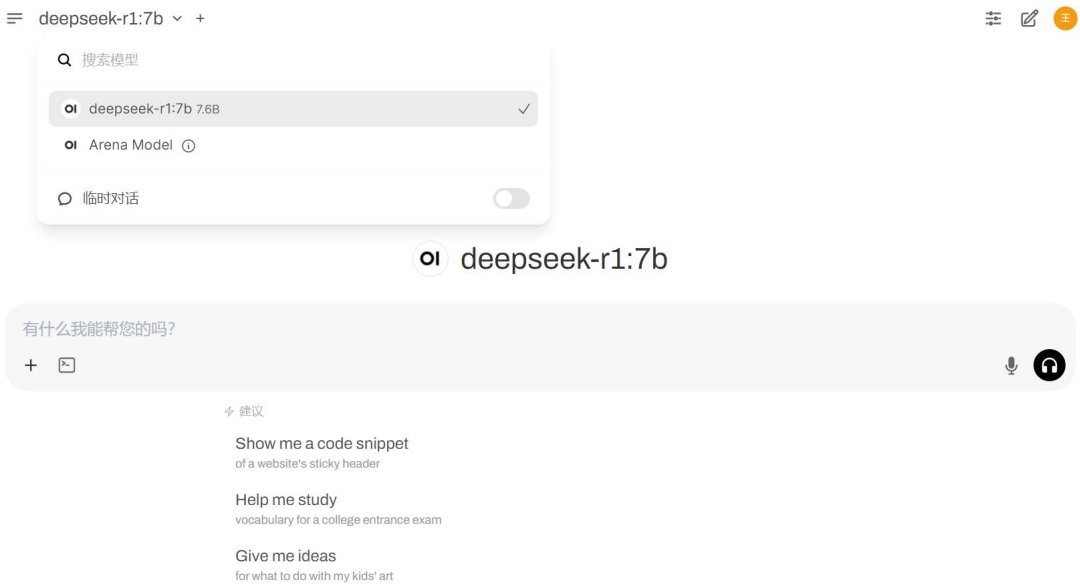


c. 注册管理员账号。

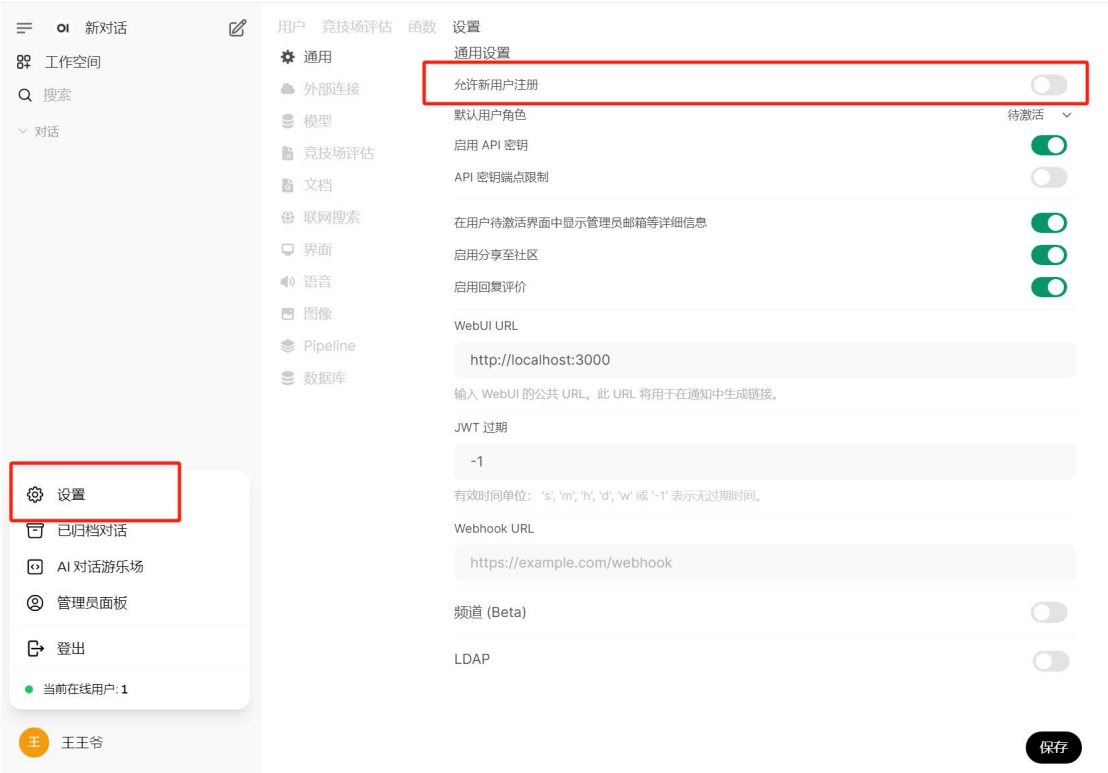


d. 使用设置。

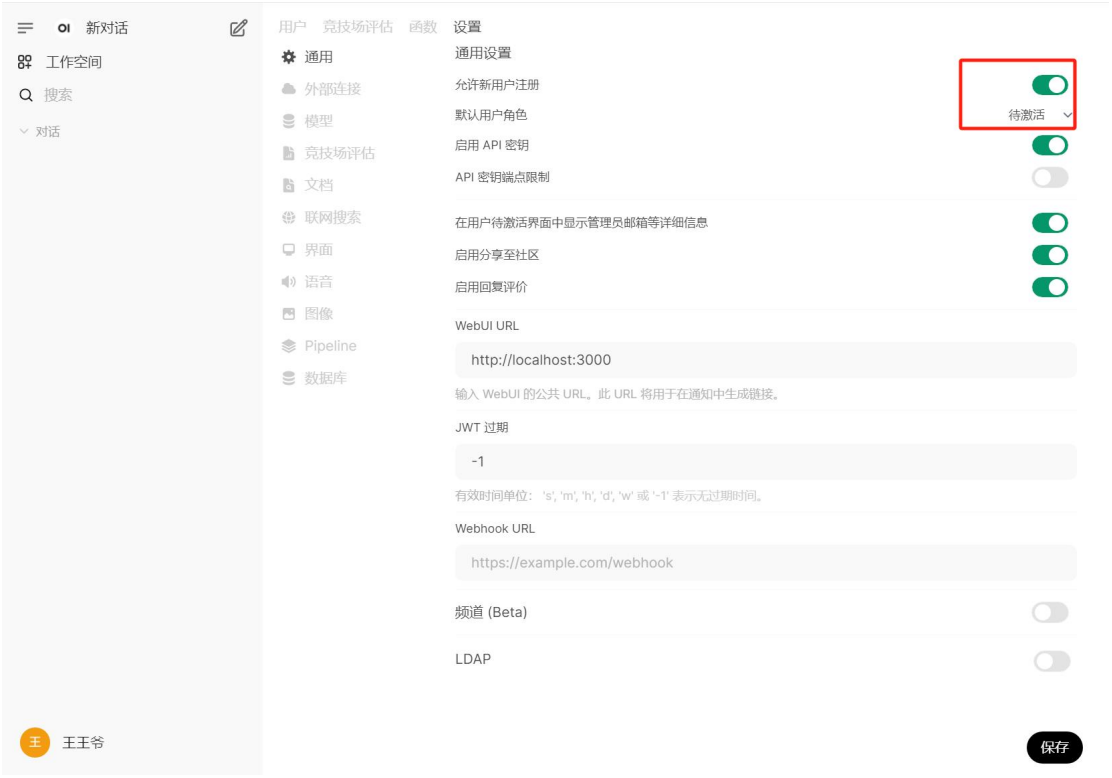
刷新进入首页，在模型下拉列表中，选择刚部署的 DeepSeek:7b 模型。



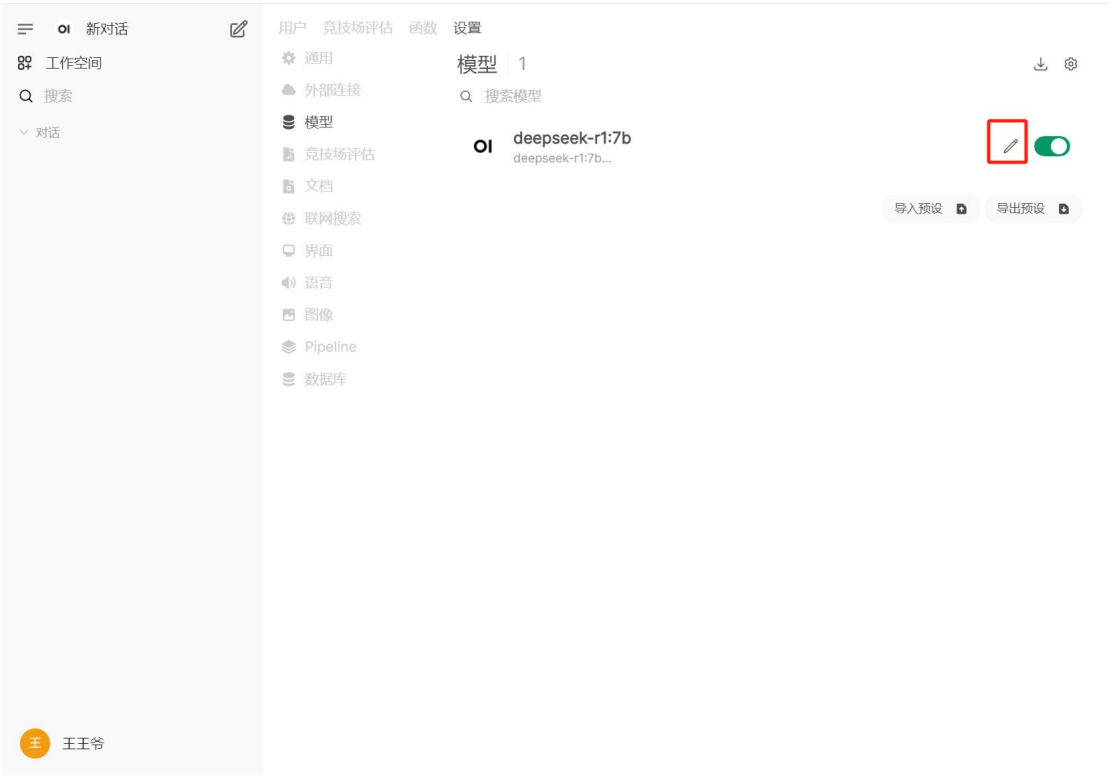
点击左下角进入设置页面，如果您不想开放其他用户注册使用，则需要关闭“允许用户注册” 功能。

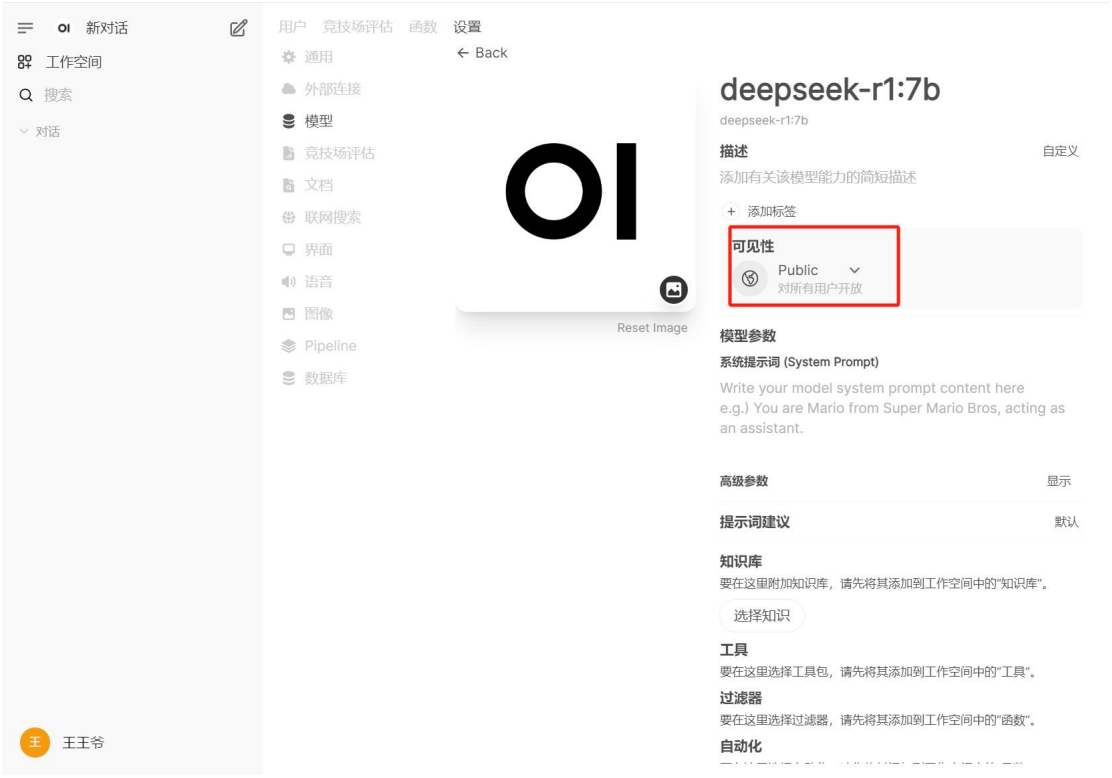


如果您允许用户注册，还可以设置用户注册之后的行为，例如选择新用户注册后默认用户角色为“用户”/“待激活” 等，需要管理员手动激活。

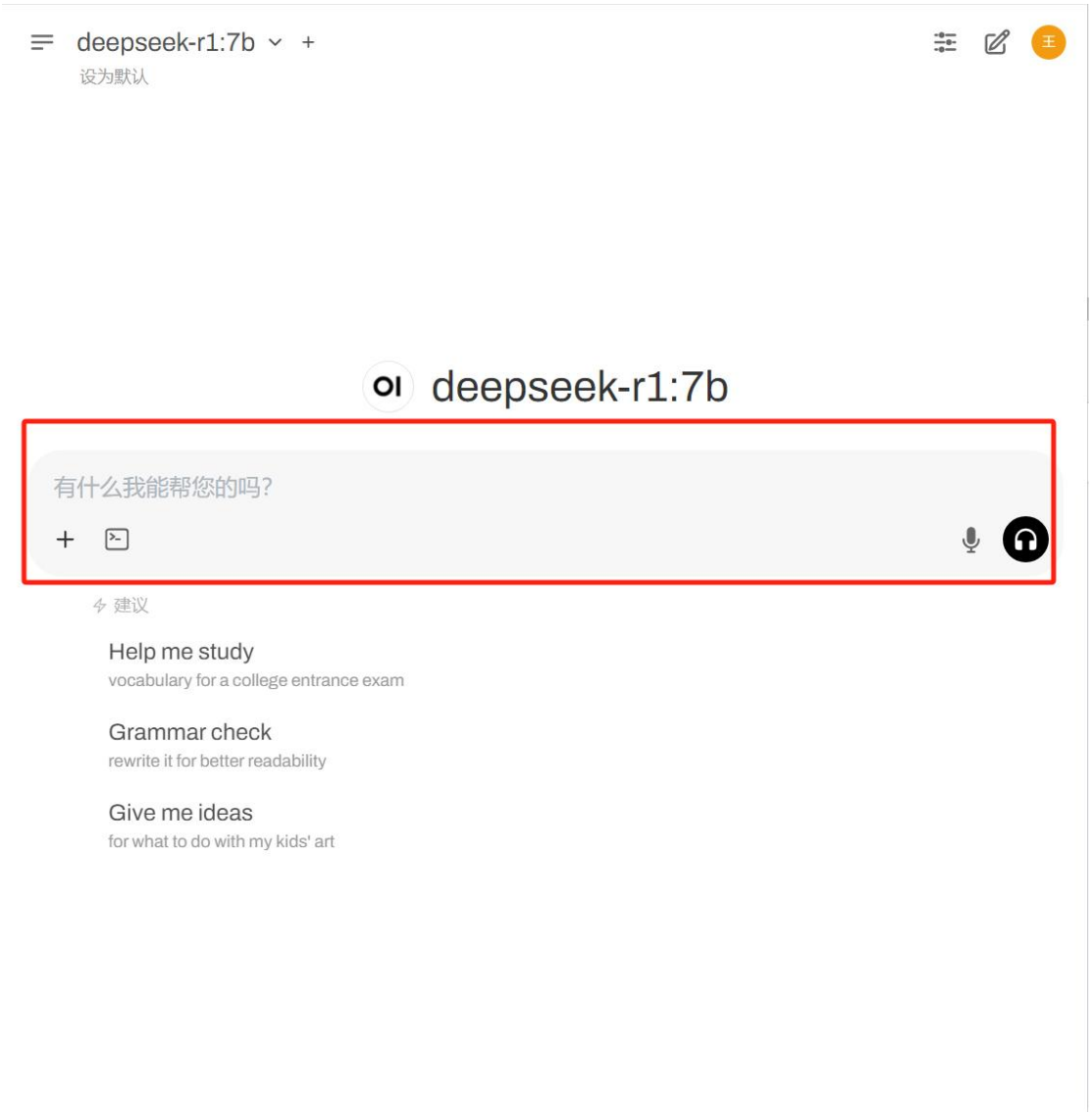


设置模型可见性。多用户模式下，建议把模型设置为"Public"。





e. 使用 DeepSeek 模型进行模型推理。


deepseek-r1:7b
设为默认
deepseek-r1:7b
有什么我能帮您的吗?
建议
Help me study
vocabulary for a college entrance exam
Grammar check
rewrite it for better readability
Give me ideas
for what to do with my kids' art



自定义部署 DeepSeek

步骤一：规格选型

不同版本的模型参数量不同，参数量越多，资源消耗越多。不同规模的企业、客户有不同诉求，客户可根据自己的实际需求，选择所需镜像和模型版本。参数较少的 DeepSeek 模型可以在 CPU 上正常运行、如 1.5B、7B 模型，更大的模型需要 GPU 才能获得良好的体验。天翼云提供了不同模型的硬件配置建议，如下：

模型	推荐规格					
	通用云主机	T4	V100	V100S	A10	A100

DeepSeek-R1-1.5B	c7.2xlarge.2 (8C16G)	pi2.2xlarge.4（显存：1*16G）	p2v.2xlarge.4（显存：1*32G）	p2vs.2xlarge.4（显存：1*32G）	pi7.4xlarge.4（显存：1*24G）	p8a.6xlarge.4（显存：1*40G）
DeepSeek-R1-7B	c7.4xlarge.2 (16C32G)	pi2.4xlarge.4（显存：2*16G）	p2v.2xlarge.4（显存：1*32G）	p2vs.2xlarge.4（显存：1*32G）	pi7.4xlarge.4（显存：1*24G）	p8a.6xlarge.4（显存：1*40G）
DeepSeek-R1-8B	-	pi2.4xlarge.4（显存：2*16G）	p2v.2xlarge.4（显存：1*32G）	p2vs.2xlarge.4（显存：1*32G）	pi7.4xlarge.4（显存：1*24G）	p8a.6xlarge.4（显存：1*40G）
DeepSeek-R1-14B	-	pi2.8xlarge.4（显存：4*16G）	p2v.4xlarge.4（显存：2*32G）	p2vs.4xlarge.4（显存：2*32G）	pi7.8xlarge.4（显存：2*24G）	p8a.12xlarge.4（显存：2*40G）
DeepSeek-R1-32B	-	-	p2v.8xlarge.4（显存：4*32G）	p2vs.8xlarge.4（显存：4*32G）	pi7.16xlarge.4（显存：4*24G）	p8a.24xlarge.4（显存：4*40G）
DeepSeek-R1-70B	-	-	-	-	-	p8a.24xlarge.4（显存：4*40G）

对应的 GPU 云主机规格请参见 NVIDIA GPU 云主机-GPU 云主机-产品简介-产品规格 - 天翼云。

步骤二：云主机开通

创建弹性云主机-弹性云主机-快速入门 - 天翼云

创建配备 GPU 驱动的 GPU 云主机（Linux）-GPU 云主机-用户指南-创建 GPU 云主机 - 天翼云

步骤三：手动部署 DeepSeek

1. 安装 ollama

a. 下载安装脚本

使用非天翼云 DeepSeek 镜像时，需要手动安装 ollama，这里以 amd64 为例，将如下脚本保存到本地，命名为 install_ollama.sh。

```
#!/bin/sh# This script installs Ollama on Linux.# It detects the current operating system architecture and installs the appropriate version of Ollama.

set -eu

red="$( (/usr/bin/tput bold || :; /usr/bin/tput setaf 1 || :) 2>&-)"

plain="$( (/usr/bin/tput sgr0 || :) 2>&-)"

status() { echo ">>> $*" >&2; }error() { echo "${red}ERROR:${plain} $*"; exit 1; }warning() { echo "${red}WARNING:${plain} $*"; }


TEMP_DIR=$(mktemp -d)cleanup() { rm -rf $TEMP_DIR; }trap cleanup EXIT

available() { command -v $1 >/dev/null; }require() {
    local MISSING=''
    for TOOL in $*; do
        if ! available $TOOL; then
            MISSING="$MISSING $TOOL"
        fi
    done
```

```
    echo $MISSING
}
[ "$(uname -s)" = "Linux" ] || error 'This script is intended to run on Linux only.'
ARCH=$(uname -m)case "$ARCH" in
    x86_64) ARCH="amd64" ;;
    aarch64|arm64) ARCH="arm64" ;;
    *) error "Unsupported architecture: $ARCH" ;;esac

IS_WSL2=false

KERN=$(uname -r)case "$KERN" in
    *icrosoft*WSL2 | *icrosoft*wsl2) IS_WSL2=true;;
    *icrosoft) error "Microsoft WSL1 is not currently supported. Please use WSL2 with 'wsl --set-version <distro> 2'" ;;
    *) ;;esac

VER_PARAM="${OLLAMA_VERSION:+?version=$OLLAMA_VERSION}"

SUDO=if [ "$(id -u)" -ne 0 ]; then
    # Running as root, no need for sudo
    if ! available sudo; then
        error "This script requires superuser permissions. Please re-run as root."
```

```
    fi

    SUDO="sudo"fi

NEEDS=$(require curl awk grep sed tee xargs)if [ -n "$NEEDS" ]; then
    status "ERROR: The following tools are required but missing:"
    for NEED in $NEEDS; do
        echo "  - $NEED"
    done
    exit 1fi
for BINDIR in /usr/local/bin /usr/bin /bin; do
    echo $PATH | grep -q $BINDIR && break || continuedone
OLLAMA_INSTALL_DIR=$(dirname ${BINDIR})
if [ -d "$OLLAMA_INSTALL_DIR/lib/ollama" ] ; then
    status "Cleaning up old version at $OLLAMA_INSTALL_DIR/lib/ollama"
    $SUDO rm -rf "$OLLAMA_INSTALL_DIR/lib/ollama"fi
status "Installing ollama to $OLLAMA_INSTALL_DIR"$SUDO install -o0 -g0
-m755 -d $BINDIR$SUDO install -o0 -g0 -m755 -d "$OLLAMA_INSTALL_DIR"
status "Downloading Linux ${ARCH} bundle"
curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}.tgz${VE
R_PARAM}" | \
```

```
    $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"if
[ "$OLLAMA_INSTALL_DIR/bin/ollama" != "$BINDIR/ollama" ] ; then
    status "Making ollama accessible in the PATH in $BINDIR"
    $SUDO ln -sf "$OLLAMA_INSTALL_DIR/ollama" "$BINDIR/ollama"fi
# Check for NVIDIA JetPack systems with additional downloadsif [ -f
/etc/nv_tegra_release ] ; then
    if grep R36 /etc/nv_tegra_release > /dev/null ; then
        status "Downloading JetPack 6 components"
        curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}-jetpack
6.tgz${VER_PARAM}" | \
            $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"
    elif grep R35 /etc/nv_tegra_release > /dev/null ; then
        status "Downloading JetPack 5 components"
        curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}-jetpack
5.tgz${VER_PARAM}" | \
            $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"
    else
        warning "Unsupported JetPack version detected.  GPU may not be
supported"
    fifi
install_success() {
```

```
    status 'The Ollama API is now available at 127.0.0.1:11434.'
    status 'Install complete. Run "ollama" from the command line.'
}trap install_success EXIT
# Everything from this point onwards is optional.
configure_systemd() {
    if ! id ollama >/dev/null 2>&1; then
        status "Creating ollama user..."
        $SUDO useradd -r -s /bin/false -U -m -d /usr/share/ollama ollama
    fi
    if getent group render >/dev/null 2>&1; then
        status "Adding ollama user to render group..."
        $SUDO usermod -a -G render ollama
    fi
    if getent group video >/dev/null 2>&1; then
        status "Adding ollama user to video group..."
        $SUDO usermod -a -G video ollama
    fi

    status "Adding current user to ollama group..."
    $SUDO usermod -a -G ollama $(whoami)

    status "Creating ollama systemd service..."
    cat <<EOF | $SUDO tee /etc/systemd/system/ollama.service >/dev/null
[Unit]
```

```
Description=Ollama Service
After=network-online.target

[Service]
ExecStart=$BINDIR/ollama serve
User=ollama
Group=ollama
Restart=always
RestartSec=3
Environment="PATH=$PATH"

[Install]
WantedBy=default.target
EOF
    SYSTEMCTL_RUNNING="$(systemctl is-system-running || true)"
    case $SYSTEMCTL_RUNNING in
        running|degraded)
            status "Enabling and starting ollama service..."
            $SUDO systemctl daemon-reload
            $SUDO systemctl enable ollama

            start_service() { $SUDO systemctl restart ollama; }
            trap start_service EXIT
            ;;
```

```
        *)
            warning "systemd is not running"
            if [ "$IS_WSL2" = true ]; then
                warning "see
https://learn.microsoft.com/en-us/windows/wsl/systemd#how-to-enable-
systemd to enable it"
            fi
            ;;
    esac
}
if available systemctl; then
    configure_systemdfi
# WSL2 only supports GPUs via nvidia passthrough# so check for nvidia-smi
to determine if GPU is availableif [ "$IS_WSL2" = true ]; then
    if available nvidia-smi && [ -n "$(nvidia-smi | grep -o "CUDA Version:
[0-9]*\.[0-9]*")" ]; then
        status "Nvidia GPU detected."
    fi
    install_success
    exit 0fi
# Don't attempt to install drivers on Jetson systemsif [ -f
/etc/nv_tegra_release ] ; then
    status "NVIDIA JetPack ready."
    install_success
    exit 0fi
```

```
# Install GPU dependencies on Linuxif ! available lspci && ! available lshw; then
    warning "Unable to detect NVIDIA/AMD GPU. Install lspci or lshw to automatically detect and install GPU dependencies."
    exit 0fi
check_gpu() {
    # Look for devices based on vendor ID for NVIDIA and AMD
    case $1 in
        lspci)
            case $2 in
                nvidia) available lspci && lspci -d '10de:' | grep -q 'NVIDIA' || return 1 ;;
                amdgpu) available lspci && lspci -d '1002:' | grep -q 'AMD' || return 1 ;;
            esac ;;
        lshw)
            case $2 in
                nvidia) available lshw && $SUDO lshw -c display -numeric -disable network | grep -q 'vendor: .* \[10DE\]' || return 1 ;;
                amdgpu) available lshw && $SUDO lshw -c display -numeric -disable network | grep -q 'vendor: .* \[1002\]' || return 1 ;;
            esac ;;
        nvidia-smi) available nvidia-smi || return 1 ;;
    esac
}
```

```
if check_gpu nvidia-smi; then
    status "NVIDIA GPU installed."
    exit 0fi
if ! check_gpu lspci nvidia && ! check_gpu lshw nvidia && ! check_gpu lspci amdgpu && ! check_gpu lshw amdgpu; then
    install_success
    warning "No NVIDIA/AMD GPU detected. Ollama will run in CPU-only mode."
    exit 0fi
if check_gpu lspci amdgpu || check_gpu lshw amdgpu; then
    status "Downloading Linux ROCm ${ARCH} bundle"
    curl --fail --show-error --location --progress-bar \

"https://mirrors.ctyun.cn/ollama/v0.5.7/ollama-linux-${ARCH}-rocm.tgz${VER_PARAM}" | \
        $SUDO tar -xzf - -C "$OLLAMA_INSTALL_DIR"

    install_success
    status "AMD GPU ready."
    exit 0fi

CUDA_REPO_ERR_MSG="NVIDIA GPU detected, but your OS and Architecture are not supported by NVIDIA.  Please install the CUDA driver manually https://docs.nvidia.com/cuda/cuda-installation-guide-linux/"# ref: https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
```

```
l#rhel-7-centos-7# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#rhel-8-rocky-8# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#rhel-9-rocky-9# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#fedorainstall_cuda_driver_yum() {
    status 'Installing NVIDIA repository...'

    case $PACKAGE_MANAGER in
        yum)
            $SUDO $PACKAGE_MANAGER -y install yum-utils
            if curl -I --silent --fail --location
"https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(una
me -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo" >/dev/null ; then
                $SUDO $PACKAGE_MANAGER-config-manager --add-repo
https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(unam
e -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo
            else
                error $CUDA_REPO_ERR_MSG
            fi
            ;;
        dnf)
            if curl -I --silent --fail --location
"https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(una
me -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo" >/dev/null ; then
```

```
                $SUDO $PACKAGE_MANAGER config-manager --add-repo
https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(unam
e -m | sed -e 's/aarch64/sbsa/')/cuda-$1$2.repo
            else
                error $CUDA_REPO_ERR_MSG
            fi
            ;;
    esac

    case $1 in
        rhel)
            status 'Installing EPEL repository...'
            # EPEL is required for third-party dependencies such as dkms
and libvdpau
            $SUDO $PACKAGE_MANAGER -y install
https://dl.fedoraproject.org/pub/epel/epel-release-latest-$2.noarch.
rpm || true
            ;;
    esac

    status 'Installing CUDA driver...'

    if [ "$1" = 'centos' ] || [ "$1$2" = 'rhel7' ]; then
        $SUDO $PACKAGE_MANAGER -y install nvidia-driver-latest-dkms
    fi
```

```
    $SUDO $PACKAGE_MANAGER -y install cuda-drivers
}
# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#ubuntu# ref:
https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.htm
l#debianinstall_cuda_driver_apt() {
    status 'Installing NVIDIA repository...'
    if curl -I --silent --fail --location
"https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(una
me -m | sed -e
's/aarch64/sbsa/')/cuda-keyring_1.1-1_all.deb" >/dev/null ; then
        curl -fsSL -o $TEMP_DIR/cuda-keyring.deb
https://developer.download.nvidia.com/compute/cuda/repos/$1$2/$(unam
e -m | sed -e 's/aarch64/sbsa/')/cuda-keyring_1.1-1_all.deb
    else
        error $CUDA_REPO_ERR_MSG
    fi

    case $1 in
        debian)
            status 'Enabling contrib sources...'
            $SUDO sed 's/main/contrib/' </etc/apt/sources.list | $SUDO
tee /etc/apt/sources.list.d/contrib.list > /dev/null
            if [ -f "/etc/apt/sources.list.d/debian.sources" ]; then
```

```
                $SUDO sed 's/main/contrib/' <
/etc/apt/sources.list.d/debian.sources | $SUDO tee
/etc/apt/sources.list.d/contrib.sources > /dev/null
            fi
            ;;
    esac

    status 'Installing CUDA driver...'
    $SUDO dpkg -i $TEMP_DIR/cuda-keyring.deb
    $SUDO apt-get update

    [ -n "$SUDO" ] && SUDO_E="$SUDO -E" || SUDO_E=
    DEBIAN_FRONTEND=noninteractive $SUDO_E apt-get -y install
cuda-drivers -q
}
if [ ! -f "/etc/os-release" ]; then
    error "Unknown distribution. Skipping CUDA installation."fi

. /etc/os-release

OS_NAME=$ID
OS_VERSION=$VERSION_ID

PACKAGE_MANAGER=for PACKAGE_MANAGER in dnf yum apt-get; do
```

```
    if available $PACKAGE_MANAGER; then
        break
    fidone
if [ -z "$PACKAGE_MANAGER" ]; then
    error "Unknown package manager. Skipping CUDA installation."fi
if ! check_gpu nvidia-smi || [ -z "$(nvidia-smi | grep -o "CUDA Version: [0-9]*\.[0-9]*")" ]; then
    case $OS_NAME in
        centos|rhel) install_cuda_driver_yum 'rhel' $(echo $OS_VERSION | cut -d '.' -f 1) ;;
        rocky) install_cuda_driver_yum 'rhel' $(echo $OS_VERSION | cut -c1) ;;
        fedora) [ $OS_VERSION -lt '39' ] && install_cuda_driver_yum $OS_NAME $OS_VERSION || install_cuda_driver_yum $OS_NAME '39';;
        amzn) install_cuda_driver_yum 'fedora' '37' ;;
        debian) install_cuda_driver_apt $OS_NAME $OS_VERSION ;;
        ubuntu) install_cuda_driver_apt $OS_NAME $(echo $OS_VERSION | sed 's/\.//') ;;
        *) exit ;;
    esacfi
if ! lsmod | grep -q nvidia || ! lsmod | grep -q nvidia_uvm; then
    KERNEL_RELEASE="$(uname -r)"
    case $OS_NAME in
        rocky) $SUDO $PACKAGE_MANAGER -y install kernel-devel kernel-headers ;;
```

```
        centos|rhel|amzn) $SUDO $PACKAGE_MANAGER -y install
kernel-devel-$KERNEL_RELEASE kernel-headers-$KERNEL_RELEASE ;;
        fedora) $SUDO $PACKAGE_MANAGER -y install
kernel-devel-$KERNEL_RELEASE ;;
        debian|ubuntu) $SUDO apt-get -y install
linux-headers-$KERNEL_RELEASE ;;
        *) exit ;;
    esac

    NVIDIA_CUDA_VERSION=$($SUDO dkms status | awk -F: '/added/ { print
$1 }')
    if [ -n "$NVIDIA_CUDA_VERSION" ]; then
        $SUDO dkms install $NVIDIA_CUDA_VERSION
    fi

    if lsmod | grep -q nouveau; then
        status 'Reboot to complete NVIDIA CUDA driver install.'
        exit 0
    fi

    $SUDO modprobe nvidia
    $SUDO modprobe nvidia_uvmfi
# make sure the NVIDIA modules are loaded on boot with
nvidia-persistencedif available nvidia-persistenced; then
    $SUDO touch /etc/modules-load.d/nvidia.conf
```

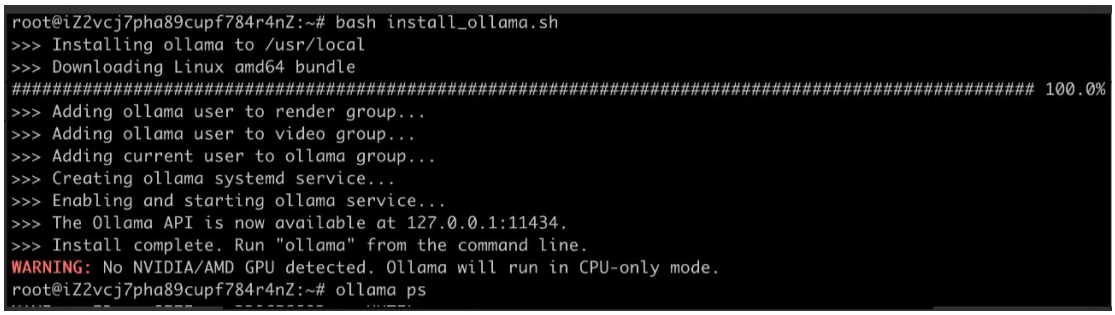
```
    MODULES="nvidia nvidia-uvm"
    for MODULE in $MODULES; do
        if ! grep -qxF "$MODULE" /etc/modules-load.d/nvidia.conf; then
            echo "$MODULE" | $SUDO tee -a
/etc/modules-load.d/nvidia.conf > /dev/null
        fi
    donefi

status "NVIDIA GPU ready."
install_success
```

b. 执行安装

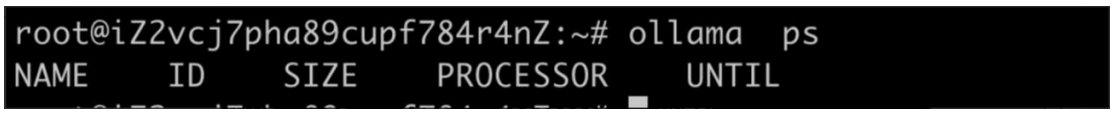
```
bash install_ollama.sh
```

如下显示则为安装成功。



c. 测试 ollama 服务安装情况

```
ollama ps
```



2. 下载模型

天翼云镜像站也为常见镜像提供了加速能力，包括 1.5B\7B\14B\70B\671B 等相关模型，可以通过如下手段体验其他模型，以 14B 模型为例：

```
mkdir deepseek-r1-14bcd deepseek-r1-14bwget
https://mirrors.ctyun.cn/ollama/models/deepseek-r1-14b/deepseek-r1-14b.ggufwget
https://mirrors.ctyun.cn/ollama/models/deepseek-r1-14b/Modelfileollama create deepseek-r1:14b -f ModelFile
```

```
root@test-deepseek-1:~# ollama create deepseek-r1:14b ModelFile
Error: accepts 1 arg(s), received 2
root@test-deepseek-1:~# ollama create deepseek-r1:14b -f ModelFile
gathering model components
copying file sha256:6e9f90f02bb3b39b59e81916e8cfce9deb45aeaeb9a54a5be4414486b907dc1e 100%
parsing GGUF
using existing layer sha256:6e9f90f02bb3b39b59e81916e8cfce9deb45aeaeb9a54a5be4414486b907dc1e
using existing layer sha256:369ca498f347f710d068cbb38bf0b8692dd3fa30f30ca2ff755e211c94768150
using existing layer sha256:f4d24e9138dd4603380add165d2b0d970bef471fac194b436ebd50e6147c6588
writing manifest
success
root@test-deepseek-1:~# ollama list
NAME                ID              SIZE      MODIFIED
deepseek-r1:14b     67711fd7c370    9.0 GB    4 seconds ago
deepseek-r1:7b      0a8c26691023    4.7 GB    3 days ago
```

如果你想体验其他模型，也可以直接通过 ollama 下载。

```
ollama pull deepseek-r1:14b
```

3. 使用 ollama 运行模型

```
ollama run deepseek-r1:14b
```

4.通过 web 界面进行交互

a.安装 open-webui

```
docker pull ghcr.io/open-webui/open-webui:main
```

b.启动容器

```
docker run -d --net=host -e PORT=3000 -e
OLLAMA_BASE_URL=http://127.0.0.1:11434 -eENABLE_SIGNUP=true -e
ENABLE_OPENAI_API=False -v open-webui:/app/backend/data --name
open-webui --restart always ghcr.io/open-webui/open-webui:main
```

后续请参考上文“**快速体验 DeepSeek——步骤二：使用 deepseek 模型**”。

FAQ

盘不够了，我应该怎么办？

如果您在模型部署过程中发现云盘的容量不够，可以采取如下措施：

1. 根据云硬盘扩容概述-云硬盘-用户指南-扩容云硬盘 - 天翼云对已有云盘进行扩容。

2. 新建一块数据盘并挂载，相关操作见挂载云硬盘-云硬盘-快速入门 - 天翼云、初始化数据盘-弹性云主机-快速入门 - 天翼云。

如何修改 ollama 模型的存储位置？

在 linux 环境下，ollama 默认模型存储目录是 /usr/share/ollama/.ollama/models/，我们建议您使用云硬盘独立挂载数据盘，将模型存储到数据盘中。模型存储位置是由环境变量控制的，我们需要修改 ollama 的环境变量重启服务才能修改存储目录，我们以 /data/ollama/models 目录为例：

1. 打开 ollama.service 文件

```
vi /etc/systemd/system/ollama.service
```

注意

请确保 ollama 用户组中的 ollama 用户具备访问该目录的读写权限。

2. 新增相关环境变量

```
[Unit]
Description=Ollama Service
After=network-online.target

[Service]
ExecStart=/usr/local/bin/ollama serve
User=ollama
Group=ollama
Restart=always
RestartSec=3
Environment"PATH=/usr/local/cuda/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/game
s:/usr/local/games:/snap/bin"
Environment="OLLAMA_MODELS=/data/ollama/models"

[Install]
WantedBy=default.target
~
~
```

3. 重启服务

```
systemctl daemon-reloadsystemctl restart ollama
```

使用 Ollama+Open WebUI 部署的镜像，登录后页面白屏等待很长时间才能进入对话页面，该怎么办？

问题描述

1.使用 Ollama+openWebUI 部署的 DeepSeek 公共镜像，默认会关闭 OpenAI API 访问，但是如果开启 OpenAI API 如下图所示，会导致公网访问 WebUI 界面加载、登录都很慢，甚至页面显示空白。



2.当通过 Open WebUI 登录后，会去访问默认配置的 OpenAI 境外 API 链接，最终超时连接失败。



解决方法

使用 Ollama 部署的情况下，您可以在“管理员面板-设置-外部链接”中，关闭 OpenAI API 即可。如果您有需要使用的外部 API 连接，也可以按需配置 API 连接，并配置 API 密钥。

10.38.2 在天翼云使用 vLLM 运行 DeepSeek 的最佳实践-32B 等版本

什么是 DeepSeek

DeepSeek 是一个基于 Transformer 架构的大型语言模型（LLM），由深度求索（DeepSeek）公司开发。它能够处理自然语言理解、生成、翻译、问答等多种任务。在目前大模型主流榜单中，DeepSeek-V3 在开源模型中位列榜首，与世界上最先进的闭源模型不分伯仲。在对话模型典型任务方面的评测效果如下：

	Benchmark (Metric)	DeepSeek V3	DeepSeek V2.5 0905	Qwen2.5 72B-Inst	Llama3.1 405B-Inst	Claude-3.5 Sonnet-1022	GPT-4o 0513
	Architecture	MoE	MoE	Dense	Dense	-	-
	# Activated Params	37B	21B	72B	405B	-	-
	# Total Params	671B	236B	72B	405B	-	-
English	MMLU (EM)	88.5	80.6	85.3	**88.6**	88.3	87.2
	MMLU-Redux (EM)	**89.1**	80.3	85.6	86.2	88.9	88.0
	MMLU-Pro (EM)	75.9	66.2	71.6	73.3	**78.0**	72.6
	DROP (3-shot F1)	**91.6**	87.8	76.7	88.7	88.3	83.7
	IF-Eval (Prompt Strict)	86.1	80.6	84.1	86.0	**86.5**	84.3
	GPQA-Diamond (Pass@1)	59.1	41.3	49.0	51.1	**65.0**	49.9
	SimpleQA (Correct)	24.9	10.2	9.1	17.1	28.4	**38.2**
	FRAMES (Acc.)	73.3	65.4	69.8	70.0	72.5	**80.5**
	LongBench v2 (Acc.)	**48.7**	35.4	39.4	36.1	41.0	48.1
Code	HumanEval-Mul (Pass@1)	**82.6**	77.4	77.3	77.2	81.7	80.5
	LiveCodeBench (Pass@1-COT)	**40.5**	29.2	31.1	28.4	36.3	33.4
	LiveCodeBench (Pass@1)	**37.6**	28.4	28.7	30.1	32.8	34.2
	Codeforces (Percentile)	**51.6**	35.6	24.8	25.3	20.3	23.6
	SWE Verified (Resolved)	42.0	22.6	23.8	24.5	**50.8**	38.8
	Aider-Edit (Acc.)	79.7	71.6	65.4	63.9	**84.2**	72.9
	Aider-Polyglot (Acc.)	**49.6**	18.2	7.6	5.8	45.3	16.0
Math	AIME 2024 (Pass@1)	**39.2**	16.7	23.3	23.3	16.0	9.3
	MATH-500 (EM)	**90.2**	74.7	80.0	73.8	78.3	74.6
	CNMO 2024 (Pass@1)	**43.2**	10.8	15.9	6.8	13.1	10.8
Chinese	CLUEWSC (EM)	90.9	90.4	**91.4**	84.7	85.4	87.9
	C-Eval (EM)	**86.5**	79.5	86.1	61.5	76.7	76.0
	C-SimpleQA (Correct)	**64.1**	54.1	48.4	50.4	51.3	59.3

当前 DeepSeek 模型有多个版本，参数量从 1.5B、 70B 到 671B 不等，适用于不同的应用场景和计算资源。参数量越多，资源消耗越多，使用者可以根据自己实际需求选择不同模型版本。

注意

为维护良好的网络环境和公共秩序，天翼云平台特规定如下使用条款：禁止任何用户或机构利用平台进行违法活动、传播不良信息或违反社会公序良俗的行为。若发生此类行为，天翼云将依法采取必要措施，停止相关服务功能，并依法追究责任。

模型由 DeepSeek 公司提供。模型的结果仅供参考，并可能因不同的环境、数据或操作条件而有所不同，我们不对模型的任何结果负责。

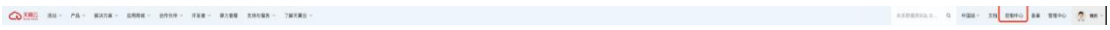
快速体验 DeepSeek

vLLM 是一个专为本地运行大语言模型（LLMs）而设计的工具，并提供了简单易用的命令行接口。这里，我们以 Deepseek-R1-32B 模型为例，讲解如何在天翼云使用 vLLM 运行 DeepSeek 模型，读者可以根据自己实际需求修改云主机模型和云主机配置。

步骤一：创建 GPU 云主机

1. 进入创建云主机页面

点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



单击“服务列表>弹性云主机”，进入主机列表页。



单击“创建云主机”，进入弹性云主机创建页。



2. 进行基础配置

基础配置：根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、“实例名称”、“主机名称”等。

选择规格：此处选择"CPU 架构"为"X86"、"分类"为"GPU 加速/AI 加速型"、"规格族"为"GPU 计算加速型 pn8r"、"规格"为"pn8r.16xlarge.4"。

选择镜像：“镜像类型”选择“镜像市场”，在云镜像市场中选择预置了 DeepSeek R1 模型的 DeepSeek-R1-32B-vLLM-Ubuntu22.04 镜像。



设置系统盘：



3. 网络及高级配置。

设置网络，包括"网卡"、"安全组"，同时配备 "弹性 IP" 用于访问模型；设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。

4. 确认并支付

步骤二：使用 DeepSeek 模型

通过云主机提供的 Open WebUI 直接体验大模型，具体请参考 Open WebUI 快速入门。

注意

镜像自带的 ollama 工具监听 127.0.0.1:11434 、webui 监听 0.0.0.0:3000 端口，云主机默认不对外开放任何端口访问，请按需开放端口访问规则，避免数据泄露。

自定义部署 DeepSeek

步骤一：规格选型

不同版本的模型参数量不同，参数量越多，资源消耗越多。不同规模的企业、客户有不同诉求，客户可根据自己的实际需求，选择所需镜像和模型版本。参数较少的 DeepSeek 模型可以在 CPU 上正常运行、如 1.5B、7B 模型，更大的模型需要 GPU 才能获得良好的体验。天翼云提供了不同模型的硬件配置建议，如下：

模型	推荐规格					
	通用云主机	T4	V100	V100S	A10	A100
DeepSeek-R1-1.5B	c7.2xlarge.2	pi2.2xlarge.4	p2v.2xlarge.4	p2vs.2xlargе.4	pi7.4xlarge.4	p8a.6xlarge.4
DeepSeek-R1-7B	c7.4xlarge.2	pi2.4xlarge.4	p2v.2xlarge.4	p2vs.2xlargе.4	pi7.4xlarge.4	p8a.6xlarge.4
DeepSeek-R1-8B	-	pi2.4xlarge.4	p2v.2xlarge.4	p2vs.2xlargе.4	pi7.4xlarge.4	p8a.6xlarge.4
DeepSeek-R1-14B	-	pi2.8xlarge.4	p2v.4xlarge.4	p2vs.4xlargе.4	pi7.8xlarge.4	p8a.12xlargе.4
DeepSeek-R1-32B	-	-	p2v.8xlarge.4	p2vs.8xlargе.4	pi7.16xlargе.4	p8a.24xlargе.4
DeepSeek-R1-70B	-	-	-	-	-	p8a.24xlargе.4

对应的 GPU 云主机规格请参见 NVIDIA GPU 云主机-GPU 云主机-产品简介-产品规格 - 天翼云。

步骤二：云主机开通

创建弹性云主机-弹性云主机-快速入门 - 天翼云

创建配备 GPU 驱动的 GPU 云主机（Linux）-GPU 云主机-用户指南-创建 GPU 云主机 - 天翼云

步骤三：手动部署 DeepSeek

建议使用天翼云提供的 DeepSeek 镜像，减少安装过程中可能遇到的问题。如您必须通过 vLLM 部署模型，请参考如下步骤：

1.安装依赖包

```
# 安装 pipcurl https://bootstrap.pypa.io/get-pip.py -o
get-pip.pypython get-pip.py# 修改 pip 镜像源,以清华源为例 pip config set
global.index-url https://pypi.tuna.tsinghua.edu.cn/simple

# 安装依赖包 pip install modelscope==1.22.3pip install
openai==1.61.0pip install tqdm==4.67.1pip install
transformers==4.48.2pip install vllm==0.7.1pip install jinja2==3.1.0
```

2.下载模型

为了方便，我们以较小的 DeepSeek-R1-Distill-Qwen-1.5B 为例（3.5G 左右），将以下内容保存为 model_download.py 文件，**参数 cache_dir 为模型的下载路径，您可以按需修改，需确保存储空间足够存放模型。**

```
from modelscope import snapshot_download

model_dir =
snapshot_download('deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B',
cache_dir='/root/model-path', revision='master')
```

运行该文件，启动模型下载。

```
python model_download.py
```

3.运行模型

完成下载后，即可使用 vLLM 运行模型。

```
vllm serve /root/model-path/deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B
--served-model-name DeepSeek-R1-Distill-Qwen-1.5B
--max-model-len=2048 --api-key my-token --port=8000
```

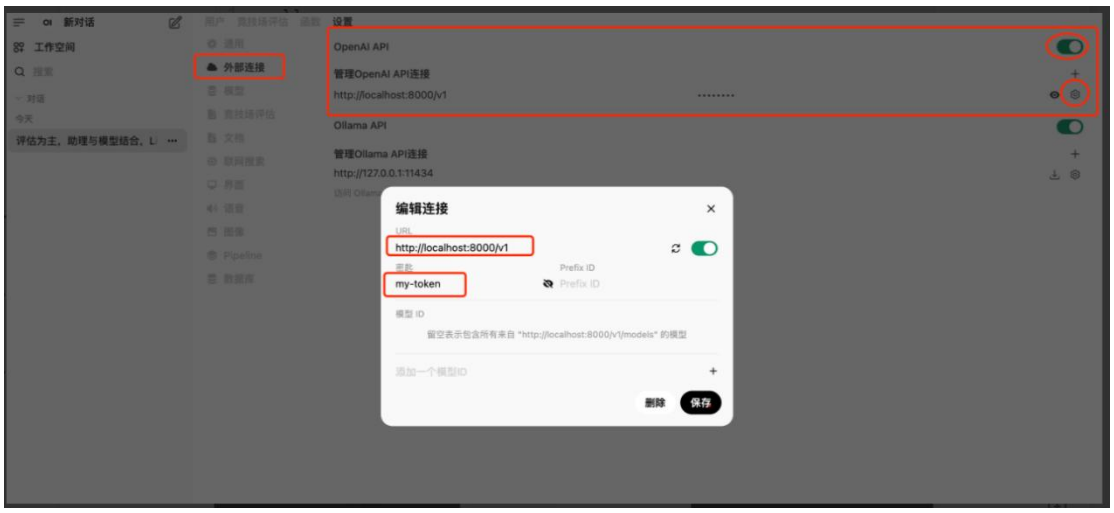
注意

其中 --api-key 参数用于模型 api 调用时认证动作，可按需修改值。

4.使用 WebUI 访问模型

基于我们已经安装、配置好的 OpenWebUI 页面，可以直接访问模型。

管理员设置中，配置 OpenAI API 访问地址，密钥就是上文中的 --api-key。



选择对应的模型，直接进行问答体验。



FAQ

如何修改 vLLM 的 api-key?

服务的配置文件目录在 /var/vllm_service.env 我们将 API_KEY 设置为想要的值。

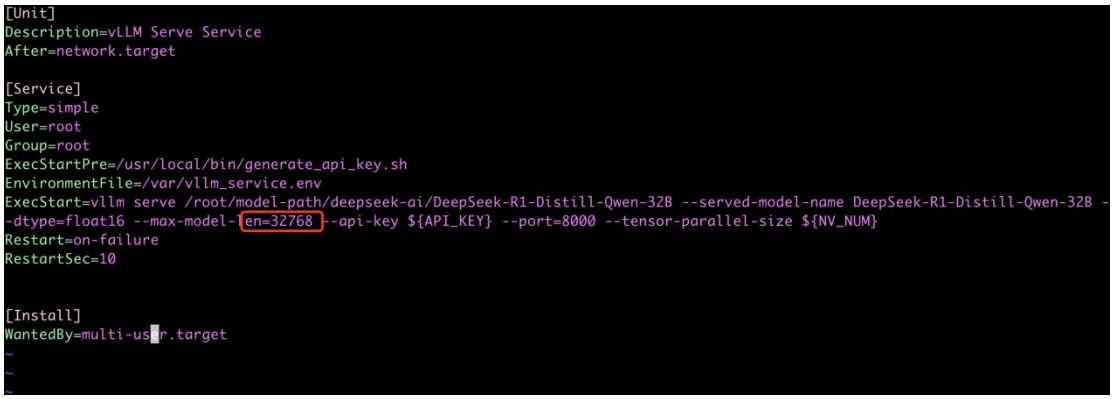


然后重启服务。

```
systemctl restart vllm
```

如何调整模型支持的最大上下文数量?

当前 vLLM 服务使用 systemd 部署，需修改 service 启动命令以实现上下文等服务参数调整。配置文件位于 /etc/systemd/system/vllm.service，我们编辑该文件的 --max-model-len 参数，然后重启服务即可实现参数调整



重新启动服务

```
systemctl daemon-reload

systemctl restart vllm
```

10.38.3 基于 DeepSeek 和 Chroma 构建个性化知识库的最佳实践

前言——私有知识库作用简介

大模型私有知识库，作为大语言模型技术与企业、组织自有数据深度融合的创新知识管理及应用解决方案，能够为特定用户群体提供更为精准、专业且安全的知

识服务。具体来讲，它是借助大语言模型搭建而成，专门为特定组织或个人定制的知识存储与检索系统。此系统会对组织内部的专业知识、业务数据、历史文档等各类信息进行深度整合与精细化处理，从而构建出独一无二的专属知识集合。依托大模型强大的语言理解与生成能力，用户能够在此基础上实现高效的知识查询与问答交互。

在本教程中，我们将为您详细介绍一种基于开源框架的私有知识库搭建方案。利用该方案，您可以在本地便捷地搭建起相应的私有知识库，大幅提升文本检索能力。

教程使用配置说明：C7 通用型云主机

```
cpu 8 核

内存 32G
```

优势

知识准确性：让模型访问定制的信息，从而提高回答的准确性和可靠性。

可解释性：检索过程可以明确指出回答所依据的信息来源，增强了回答的可解释性。

减少幻觉：降低了语言模型生成无事实依据内容（即“幻觉”）的可能性。

一、DeepSeek 自部署

请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云 --自定义部署 DeepSeek。

二、私有知识仓库构建

1、 安装向量数据库

a）检查 python 环境是否存在，不存在可参考安装

```
# 检查是否存在 pythonpython --version

# 安装 python 环境

apt install python3
```

```
root@vm:~/deepseek-r1-1.5b# python --version
Python 3.10.12
```

b）安装向量数据库及与 ollama 交互的组件

```
#  安装  Chroma 向量数据库，基于 langchain 开源框架解析文本

pip install langchain_chroma#  安装  Ollama 交互库

pip install langchain_ollama#  安装  langchain_community 库

pip install langchain langchain_community
```

```
root@vm:~/deepseek-r1-1.5b# pip install langchain_chroma
Collecting langchain_chroma
  Downloading langchain_chroma-0.2.1-py3-none-any.whl (11 kB)
Collecting numpy<2.0.0,>=1.22.4
  Downloading numpy-1.26.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.2/18.2 MB 104.7 kB/s eta 0:00:00
Collecting chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0
  Downloading chromadb-0.6.3-py3-none-any.whl (611 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 611.1/611.1 KB 63.3 kB/s eta 0:00:00
Collecting fastapi<1,>=0.95.2
  Downloading fastapi-0.115.8-py3-none-any.whl (94 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 94.8/94.8 KB 72.5 kB/s eta 0:00:00
Collecting langchain-core!=0.3.0,!=0.3.1,!=0.3.10,!=0.3.11,!=0.3.12,!=0.3.13,!=0.3.14,!=0.3.2,!=0.3.3,!=0
  Downloading langchain_core-0.3.34-py3-none-any.whl (412 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 413.0/413.0 KB 91.5 kB/s eta 0:00:00
Collecting overrides>=7.3.1
  Downloading overrides-7.7.0-py3-none-any.whl (17 kB)
Collecting typing_extensions>=4.5.0
  Downloading typing_extensions-4.12.2-py3-none-any.whl (37 kB)
Collecting opentelemetry-instrumentation-fastapi>=0.41b0
  Downloading opentelemetry_instrumentation_fastapi-0.51b0-py3-none-any.whl (12 kB)
Collecting orjson>=3.9.12
  Downloading orjson-3.10.15-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (130 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 130.3/130.3 KB 82.6 kB/s eta 0:00:00
Collecting chroma-hnswlib==0.7.6
```

```
root@vm:~/deepseek-r1-1.5b# pip install langchain_ollama
Collecting langchain_ollama
  Downloading langchain_ollama-0.2.3-py3-none-any.whl (19 kB)
Requirement already satisfied: langchain-core<0.4.0,>=0.3.33 in /usr/local/lib/python3.10/dist-packages (from langchai
Collecting ollama<1,>=0.4.4
  Downloading ollama-0.4.7-py3-none-any.whl (13 kB)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.10/dist-packages (from langchain-core<0.
Requirement already satisfied: typing-extensions>=4.7 in /usr/local/lib/python3.10/dist-packages (from langchain-core<
Requirement already satisfied: pydantic<3.0.0,>=2.5.2 in /usr/local/lib/python3.10/dist-packages (from langchain-core<
Requirement already satisfied: langsmith<0.4,>=0.1.125 in /usr/local/lib/python3.10/dist-packages (from langchain-core
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in /usr/local/lib/python3.10/dist-packages (from langch
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/dist-packages (from langchain-core<0.4.0,>=0.3
Requirement already satisfied: packaging<25,>=23.2 in /usr/local/lib/python3.10/dist-packages (from langchain-core<0.4
Requirement already satisfied: httpx<0.29,>=0.27 in /usr/local/lib/python3.10/dist-packages (from ollama<1,>=0.4.4->la
Requirement already satisfied: certifi in /usr/lib/python3/dist-packages (from httpx<0.29,>=0.27->ollama<1,>=0.4.4->la
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/dist-packages (from httpx<0.29,>=0.27->ollam
```

```
root@vm:~# pip install langchain langchain_community
Collecting langchain
  Downloading langchain-0.3.18-py3-none-any.whl (1.0 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.0/1.0 MB 41.5 kB/s eta 0:00:00
Collecting langchain_community
  Downloading langchain_community-0.3.17-py3-none-any.whl (2.5 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.5/2.5 MB 57.6 kB/s eta 0:00:00
Requirement already satisfied: langsmith<0.4,>=0.1.17 in /usr/local/lib/python3.10/dist-packages (from langchain) (0.3.8)
Collecting langchain-text-splitters<1.0.0,>=0.3.6
  Downloading langchain_text_splitters-0.3.6-py3-none-any.whl (31 kB)
Collecting aiohttp<4.0.0,>=3.8.3
  Downloading aiohttp-3.11.12-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.6 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.6/1.6 MB 413.3 kB/s eta 0:00:00
Requirement already satisfied: numpy<2,>=1.26.4 in /usr/local/lib/python3.10/dist-packages (from langchain) (1.26.4)
Collecting async-timeout<5.0.0,>=4.0.0
  Downloading async_timeout-4.0.3-py3-none-any.whl (5.7 kB)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.10/dist-packages (from langchain) (6.0.2)
```

2、安装文本解析插件，这里以 pdf 进行示例

```
pip  install pdfplumber
```

```
root@vm:~/hp# pip install pdfplumber
Collecting pdfplumber
  Using cached pdfplumber-0.11.5-py3-none-any.whl (59 kB)
Collecting pypdfium2>=4.18.0
  WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after
able')': /packages/be/d9/a2f1ee03d47fbeb48bcfde47ed7155772739622cfadf7135a84ba6a97824/pypdfium2
  Downloading pypdfium2-4.30.1-py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.9 MB)
     ---------------------------------------- 2.9/2.9 MB 53.3 kB/s eta 0:00:00
Collecting Pillow>=9.1
  Downloading pillow-11.1.0-cp310-cp310-manylinux_2_28_x86_64.whl (4.5 MB)
     ---------------------------------------- 4.5/4.5 MB 47.3 kB/s eta 0:00:00
Collecting pdfminer.six==20231228
  Downloading pdfminer.six-20231228-py3-none-any.whl (5.6 MB)
     ---------------------------------------- 5.6/5.6 MB 44.2 kB/s eta 0:00:00
Collecting cryptography>=36.0.0
  Downloading cryptography-44.0.0-cp39-abi3-manylinux_2_28_x86_64.whl (4.2 MB)
     ---------------------------------------- 4.2/4.2 MB 13.2 kB/s eta 0:00:00
```

3、基于私有数据开展模型对话

```
root@vm:~/hp# python3 ollama_private_vector_data_create.py
Number of requested results 4 is greater than number of elements in index 2, updating n_results = 2
检索到的文档摘要： <think>
嗯，我现在要总结一下这些文档中的主要主题。首先，我得仔细阅读每一段内容，找出它们之间的联系和区别。

第一段 talking about high generation能力，也就是生成文本的能力，包括快速生成高质量的中文文章、报告、故事等。强调语言流畅、逻辑连贯，并且可以根据不同的任

第二段是领域专业性，比如针对代码专门优化训练，提供准确专业的回答支持行业应用。这可能意味着DeepSeek - Coder在代码开发中特别强。

第三段讲可扩展性，模型架构设计灵活，适应数据和计算资源的变化。这说明模型能够随着使用情况不断优化，保持性能稳定。

接下来，应用场景部分分为智能写作、智能客服、代码开发和知识问答。这些都是用户可能遇到的具体问题解决方法，比如生成文案、处理客服请求、编写代码、获取信息查

DeepSeek的模型家族包括DeepSeek-Coder、DeepSeek-Llama和DeepSeek-R1。DeepSeek-Coder专注于代码领域，而DeepSeek-Llama在通用任务上表现好，DeepSeek-R1则有一定

特点部分提到强大的语言理解能力，并且具备在不同场景下的性能提升。这可能包括通过训练数据进行优化，提高模型的适应能力和扩展性。

综上所述，文档主要讲的是DeepSeek在自然语言处理和相关领域的应用，包括生成、问答、代码开发等多个方面，重点放在模型的专业性、可扩展性和多样化应用场景上。
</think>

```markdown
# 深度探索（DeepSeek）技术总结

## 1. 主要主题总结
```

图中使用的模型对话脚本如下，仅供参考：

```
from langchain_community.document_loaders  import
PDFPlumberLoaderfrom langchain.text_splitter  import
RecursiveCharacterTextSplitterfrom langchain_chroma  import
Chromafrom langchain_ollama  import OllamaEmbeddings, ChatOllamafrom
langchain_core.output_parsers  import StrOutputParserfrom
langchain_core.prompts  import ChatPromptTemplatefrom
langchain_core.runnables  import RunnablePassthrough

# 加载 PDF 文档

file="DeepSeek.pdf"

loader  = PDFPlumberLoader(file)

docs  = loader.load()

# 分割文档，分割参数会影响推理准确性
```
```

```
text_splitter  = RecursiveCharacterTextSplitter(chunk_size=500,
chunk_overlap=0)
all_splits  = text_splitter.split_documents(docs)
# 创建向量存储
local_embeddings  = OllamaEmbeddings(model="deepseek-r1:1.5b")
vectorstore  = Chroma.from_documents(documents=all_splits,
embedding=local_embeddings)
# 初始化模型
model  = ChatOllama(
    model="deepseek-r1:1.5b",
)
# 定义摘要提示模板
summary_prompt = ChatPromptTemplate.from_template(
    "总结这些检索到的文档中的主要主题：{docs}"
)
# 将传入的文档转换成字符串的形式 def format_docs(docs):
    return "\n\n".join(doc.page_content for doc in docs)
# 创建摘要链
summary_chain  ={"docs": format_docs}| summary_prompt  | model  |
StrOutputParser()
# 定义问题
question  ="DeepSeek 项目的目的是什么？"
# 检索相关文档
docs  = vectorstore.similarity_search(question)
```

```
# 生成摘要
summary  = summary_chain.invoke(docs)print("检索到的文档摘要：",
summary)
# 定义提示模板
RAG_TEMPLATE  ="""
你是一个问答助手。请使用以下检索到的上下文信息来回答问题。如果你不知道答案，请直接说不知道。答案最多使用三句话，并保持简洁。

<context>
{context}
</context>

回答以下问题：

{question}
"""

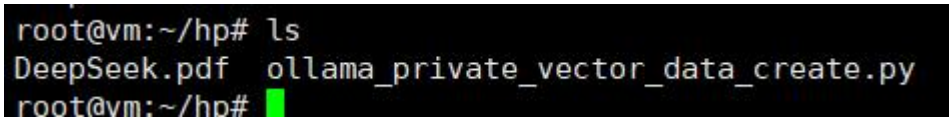
rag_prompt  = ChatPromptTemplate.from_template(RAG_TEMPLATE)
# 获取检索器
retriever  = vectorstore.as_retriever()
# 创建问答链
qa_chain = (
    {"context": retriever | format_docs, "question":
RunnablePassthrough()}
```

```
    | rag_prompt
    | model
    | StrOutputParser()
)
# 运行问答链
answer  = qa_chain.invoke(question)print("问题的答案：", answer)
```

综上，您成功构建了一个定制化的私有知识库检索模型。

注意

对应文本数据需要与运行脚本在同一目录。



10.38.4 基于OpenWebUI+Ollama+DeepSeek实现大模型私有知识库零代码构建的最佳实践

前言——私有知识库作用简介

大模型私有知识库是结合大语言模型技术与企业或组织自有数据的一种知识管理和应用方案，旨在为特定用户群体提供更精准、专业且安全的知识服务。

大模型私有知识库是基于大语言模型构建的，专门服务于特定组织或个人的知识存储和检索系统。它整合组织内部的专业知识、业务数据、历史文档等，形成专属知识集合，借助大模型的强大语言理解和生成能力，为用户提供高效的知识查询和问答服务。

本教程提供一种基于开源框架的私有知识库搭建方案。如果需要构建企业级服务且对数据安全有需求，则需相关企业对该服务进行定制，以完成私有知识库的搭建。

本文介绍的方法采用了开源框架自带的数据库服务，但该方法不保证相关数据在

开源框架下的安全性。请务必在上传前对敏感数据进行脱敏处理。

本教程使用的 GPU 云主机规格如下所示，用于部署 DeepSeek-r1:70b 模型，配置仅供参考。

```
cpu 32 核

内存  128G

GPU:  NVIDIA  A10*2  (48GB)
```

如需体验其它模型版本，请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--自定义部署 DeepSeek--步骤二：规格选型，选择合适的云主机。

优势

知识准确性：让模型访问定制的信息，从而提高回答的准确性和可靠性。

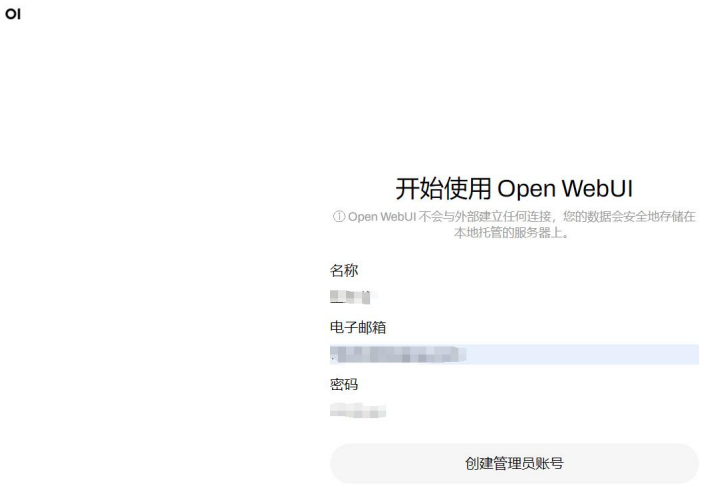
可解释性：检索过程可以明确指出回答所依据的信息来源，增强了回答的可解释性。

减少幻觉：降低了语言模型生成无事实依据内容（即 “幻觉”）的可能性。

一、环境准备

1.参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--快速体验 DeepSeek，准备相关设备及模型资源。

2.待模型启动后，通过访问 http://{公网 ip 地址}:3000 进入 OpenWebUI 界面。按照界面提示完成账号注册，开启私有知识库搭建之旅。



二、私有数据准备

搭建私有知识库需提前准备相应的数据材料。OpenWebUI 支持多种文本格式，包括但不限于：

- Markdown：轻量级标记语言格式，便于快速书写和阅读，能清晰呈现文本结构，常用于文档撰写、笔记记录等场景。
- PDF：通用文档格式，可保留文档原始排版、字体、图像等元素，广泛应用于正式文档、书籍、报告的分发与存储。
- TXT：最基本的文本格式，仅包含文本内容，无复杂格式设置，适用于纯文本信息的存储和传输。
- HTML：超文本标记语言格式，用于创建网页，包含文本内容和网页元素标记信息，可被浏览器解析显示。
- docx：是 Microsoft Word 的文档格式，能包含丰富的文本格式、图表、样式等内容，OpenWebUI 可以借助相关的文档解析库来提取其中的文本信息，用于知识库的构建和查询。

对于结构简单的纯文本材料，可直接通过 OpenWebUI 进行解析。若需了解除纯文本外其他文件格式的支持情况，请查阅 OpenWebUI 官方文档。

三、私有知识库搭建教程

1、登录 OpenWebUI 后，点击左上角 “工作空间”。



2、在“工作空间”中找到“知识库”选项，点击右侧的“+”号新建知识库。



3、在新建知识库页面，填写知识库名称、描述，设置可见性等信息，完成后点击“创建知识库”。





4、新建知识库成功后，再次点击右侧的“+”号，选择要添加的知识文档（可一次性添加多个）。添加成功后，点击对应文档查看加载的语料内容，验证知识解析是否准确。

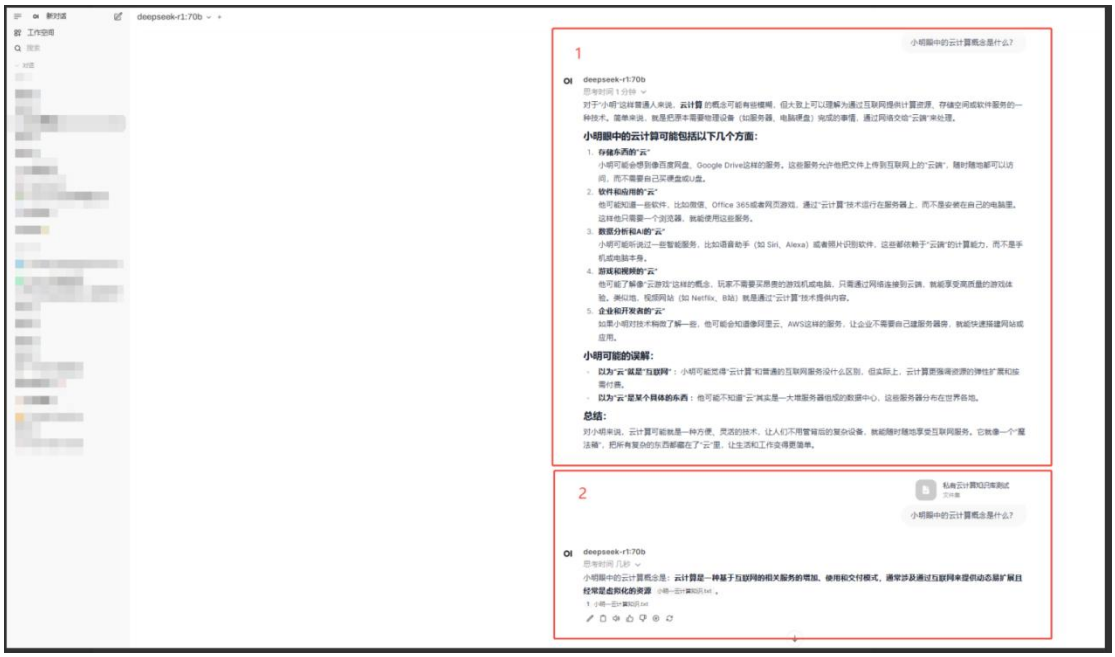




四、私有知识模型推理验证

1、点击左上角 “新对话”，选择已部署的推理模型。





2、在输入对话内容时，以 “#” 号开头，可选择对应知识库或知识文档作为推理参考。



3、若不想继续使用某个知识库，可按照特定操作流程将其移除，具体操作可参考相关图示。



通过以上步骤，您即可完成一个简单的私有知识库搭建，赶紧动手尝试吧！

五、推荐使用场景

1、智能客服训练 AI 回答公司政策、产品 FAQ，减少人工客服负担。

2、企业内部搜索让员工直接向 AI 提问，无需翻遍公司 Wiki 或文档库。

3、智能助手帮助销售整理客户信息，协助 HR 处理面试反馈，助力研发自动生成技术文档。

4、智能营销策划协助市场团队分析市场趋势和竞争对手数据，生成针对性的营销活动方案，提升营销效果。

5、财务智能分析工具帮助财务人员自动完成财务报表分析，预测财务风险，为企业决策提供数据支持。

6、生产流程优化系统利用 AI 实时监测生产设备运行状况，预测设备故障，提

前安排维护，减少生产中断。

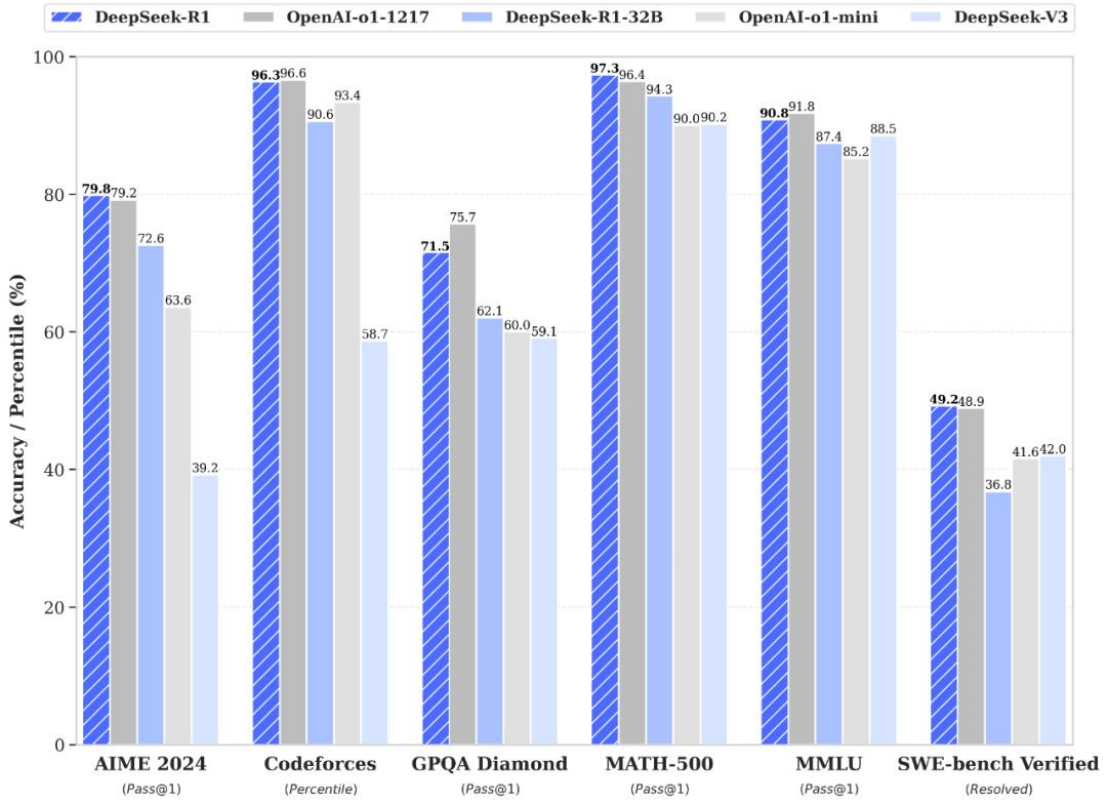
7、培训内容生成助手，为培训部门快速生成员工培训课件、在线课程内容，满足不同岗位的培训需求。

8、供应链管理智能助手，根据销售数据、库存水平和供应商信息，辅助优化采购计划。

10.38.5 基于 LLaMA-Factory 微调 DeepSeek-R1-Distill-Qwen-7B 最佳实践

DeepSeek 简介

DeepSeek 是一个基于 Transformer 架构的大型语言模型（LLM），由深度求索（DeepSeek）公司开发。它能够处理自然语言理解、生成、翻译、问答等多种任务。DeepSeek-R1 在后训练阶段大规模使用了强化学习技术，在仅有极少标注数据的情况下，极大提升了模型推理能力。在数学、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。



DeepSeek 公司在开源 DeepSeek-R1-Zero 和 DeepSeek-R1 两个参数量超过 660B 模型的同时，通过 DeepSeek-R1 的输出，蒸馏了 6 个小模型开源给社区，其中 32B 和 70B 模型在多项能力上实现了对标 OpenAI o1-mini 的效果。

	AIME 2024 pass@1	AIME 2024 cons@64	MATH-500 pass@1	GPQA Diamond pass@1	LiveCodeBench pass@1	CodeForces rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759.0
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717.0
o1-mini	63.6	80.0	90.0	60.0	53.8	**1820.0**
QwQ-32B	44.0	60.0	90.6	54.5	41.9	1316.0
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954.0
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189.0
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481.0
DeepSeek-R1-Distill-Qwen-32B	**72.6**	83.3	94.3	62.1	57.2	1691.0
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205.0
DeepSeek-R1-Distill-Llama-70B	70.0	**86.7**	**94.5**	**65.2**	**57.5**	1633.0

模型微调简介

大语言模型微调技术通过调整预训练模型参数，使其适应特定任务需求。其核心价值在于让通用的人工智能工具变得更“专业”和“实用”，就像把一个会做各种菜的厨师培训成专精川菜的大厨。它能让企业用较低成本获得适合自身需求的 AI 能力，例如：

1. 降低成本：无需从头训练昂贵的大模型，只需在现有通用模型基础上“补充学习”企业内部的业务知识、行业术语，就能快速打造专属 AI 助手，成本大幅降低。

2. 精准解决问题：通用 AI 可能懂诗词绘画，但不懂医疗诊断或法律合同。通过微调，AI 能深度掌握特定领域的专业知识，例如快速分析病历、自动生成符合行业规范的报告。

3. 保护数据隐私：企业可将内部数据安全地用于微调，既让 AI 掌握核心业务知识，又避免敏感数据直接暴露在公共模型中。

4. 推动普惠应用：中小型企业也能通过这项技术享受 AI 红利，例如教育机构定制学习辅导 AI，工厂打造设备维护专家系统，真正让 AI 进入百行千业。

在下文中，我们将采用两个具体实践案例来说明模型微调的实际效果：

1. 基于天翼云文档数据的微调。利用企业自身的文档数据构建微调数据集，仅使用单台多 GPU 卡的云主机即可完成模型微调训练。经过微调后的模型能够在企业内部知识领域提供准确的回答。

2. 基于医学计算数据的微调。利用医学领域的专业知识数据构建微调数据集，仅

使用单台多 GPU 卡的云主机即可完成模型微调训练。经过微调后的模型能够在识别与分析病例时取得更高的计算准确度。

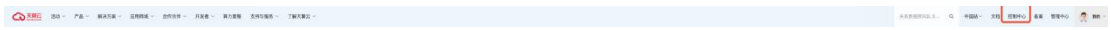
快速微调 DeepSeek

LLaMA-Factory 是一个开源项目，旨在为用户提供简便的方式来训练和部署大型语言模型（LLM）。这里，我们以蒸馏小模型 DeepSeek-R1-Distill-Qwen-7B 模型为例，讲解如何在天翼云使用 LLaMA-Factory 运行与微调 DeepSeek 模型，读者可以根据自己实际需求选择模型和云主机配置。

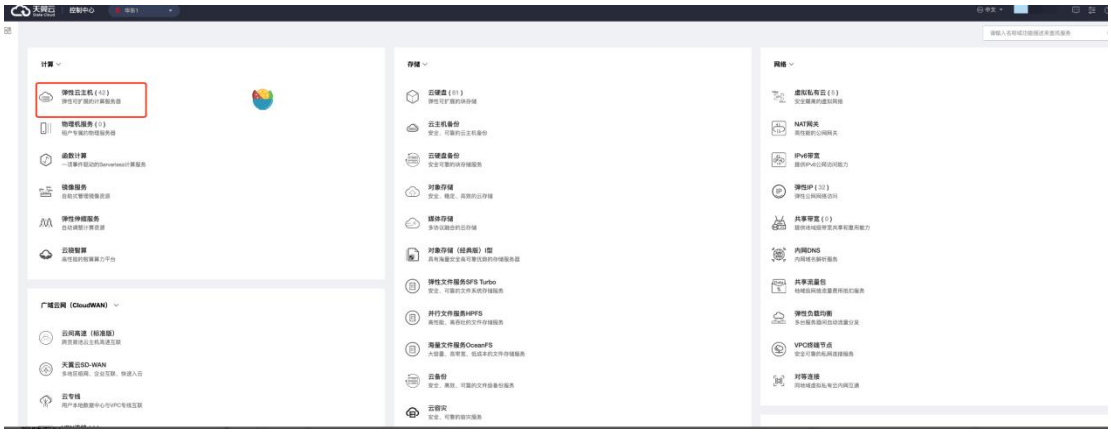
步骤一：创建 GPU 云主机

1.进入创建云主机页面。

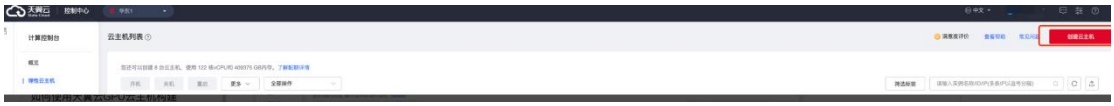
a.点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



b.单击“服务列表>弹性云主机”，进入主机列表页。



c.单击“创建云主机”，进入弹性云主机创建页。



2.进行基础配置

a.根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、“实例名称”、“主机名称”等。

b.选择规格。此处选择"CPU 架构"为"X86"、"分类"为"GPU 加速/AI 加速型"、"规格族"为"GPU 计算加速型 pi7"、"规格"为"pi7.4xlarge.4"。

注意

本案例仅做微调流程演示，选择的机型规格仅支持 7B 模型的少量数据微调，如果您的数据量较大，请选择显存更高的机型规格。

c.选择镜像。“镜像类型”选择“镜像市场”，在云镜像市场中选择预置了 DeepSeek-R1-Distill-Qwen-7B 模型的 DeepSeek-LlamaFactory-Ubuntu22.04 镜像。



3.网络及高级配置

设置网络，包括"网卡"、"安全组"，同时配备 "弹性 IP" 用于下载和访问模型；设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。注意安全组需要放行 22 端口(用于 ssh)和 7860 端口(用于访问 web 页面)。

4.确认并支付

步骤二：微调 DeepSeek 模型

1.启动 LLaMA-Factory 服务

在完成云主机开通后，即可通过 ssh 连接到到云主机命令行。模型及 LLaMA-Factory 服务均已放置在/root 目录下，且已完成配套环境安装，可直接运行。

```
# 进入目录 cd /root/LLaMA-Factory# 启动 webui 服务（也可通过 nohup 放到后台启动）

llamafactory-cli webui
```

在成功启动 LLaMA-Factory 服务后，您将可以通过 **http://{公网 ip 地址}:7860/** 访问到相关页面。

```
# 访问服务 http://{公网 ip 地址}:7860/
```

2.配置基础大模型和微调方法

在成功进入 LLaMA-Factory 页面后，您首先需要按照如图的内容进行页面语言，模型名称和模型路径的设置。**注意模型预置在 /root/DeepSeek-R1-Distill-Qwen-7B 目录下。**



3.配置微调数据集

在 Train 页面内，您需要配置用于微调训练的数据集。镜像内预置了 alpaca 格式的天翼云示例数据，您也可参考示例数据**修改/root/LLaMA-Factory/data/目录下的 alpaca_zh_ctyun.json(微调数据集)和 dataset_info.json(数据集信息)**两个文件，改用您自己的私有数据。



镜像内置两份数据集，单条内容示例如下：

data/alpaca_zh_ctyun.json(天翼云文档数据集)

```
{
    "instruction": "天翼云的通用型弹性云主机有哪些规格？",
    "input": "",
    "output": "<think>用户询问通用型云主机规格，首先需要明确通用型的定位是平衡计算和内存资源。根据产品文档，通用型以 s7 系列为代表，应当列举不同后缀规格及其配置差异，最后补充适用场景说明。</think>\n 通用型弹性云主机有多个规格，例如 s7.small、s7.medium、s7.large、s7.xlarge 等，每种规格提供不同的 vCPU 和内存配置，适合多种业务场景。"
}
```

data/alpaca_en_medcalc_bench_max2k.json(医学计算数据集 MedCalc-Bench)

```
{
    "instruction": "A 16-year-old female adolescent was referred to our hospital with severe hypertension (systolic pressure 178 mmHg), which was first detected 7 months prior to presentation during a routine annual physical ...",
    "input": "What is the patient's Creatinine Clearance using the Cockroft-Gault Equation in terms of mL/min? You should use the patient's adjusted body weight in kg instead of the patient's actual body weight if the patient is ...",
    "output": "<think>The formula for computing Cockcroft-Gault is given by CrCl = ((140 - age) * adjusted weight * (gender_coefficient)) / (serum creatinine * 72), where the gender_coefficient is 1 if male, and 0.85 if female..."
}
```

4.配置相关训练参数

在 Train 页面内，还有大量可修改的微调训练参数。如果您对他们还不够了解，可暂时不进行修改。其中，训练轮数与样本量级关联性较大，如果您的微调数据量很少，则可能需要设置较大的训练轮数，才能有效果。



5.启动微调训练

点击最下方的开始按钮，即可基于上面选择的基础模型和微调数据集，启动模型微调训练。页面下方会显示实时的训练进度，训练日志和 loss 变化情况。



如果您的训练样本较多，单张 A10 显卡的 24G 显存很容易因为无法承载，而导致报错"CUDA out of memory"。此时，您需要将云主机变配到显存更大的机型规格，并开启 DeepSpeed stage3 进行模型参数分片，如下图所示。



同时，您还需要修改默认 DeepSpeed 配置中的部分参数，以保证训练正常进行。修改/root/LLaMA-Factory/cache/ds_z3_config.json 中的如下内容：

```
"zero_optimization": {
    "stage": 3,
    "overlap_comm": true,
```

```
    "contiguous_gradients": true,
    "sub_group_size": 1e9,
    "reduce_bucket_size": "auto",
    "stage3_prefetch_bucket_size": "auto",
    "stage3_param_persistence_threshold": "auto",
    "stage3_max_live_parameters": 1e4, //降低该参数以减少显存占用
    "stage3_max_reuse_distance": 1e4, //降低该参数以减少显存占用
    "stage3_gather_16bit_weights_on_model_save": true}
```

步骤三：测试模型微调效果

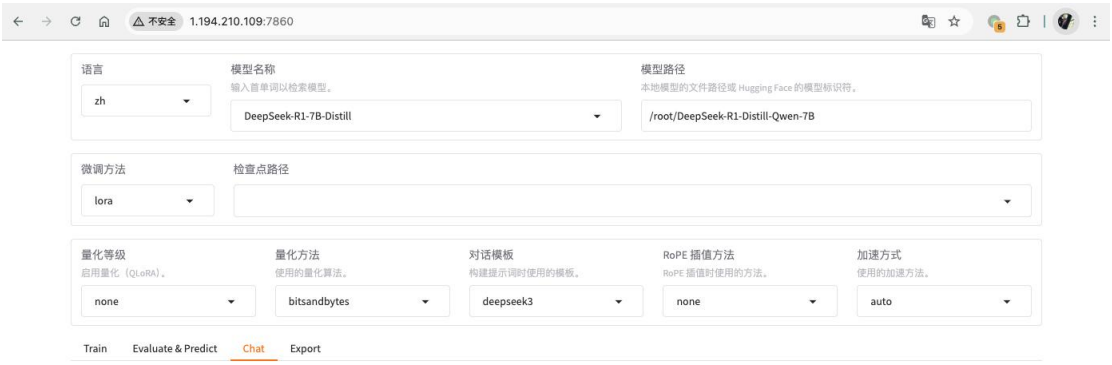
LLaMA-Factory 框架提供了模型推理的能力，可直接进入 Chat 页面加载模型进行对话。

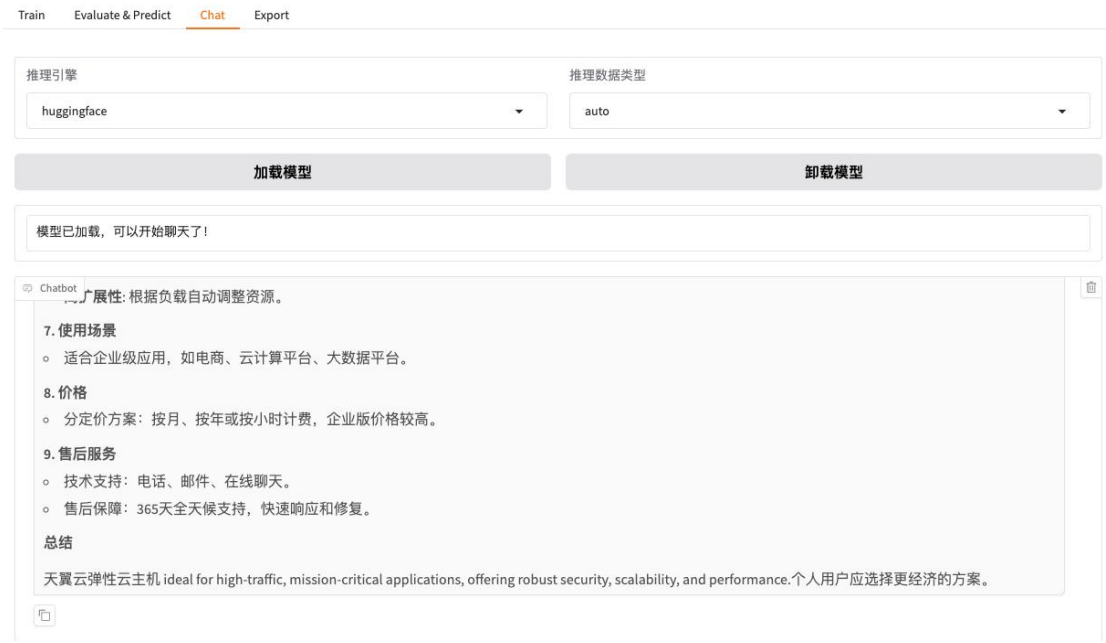
1.测试基础大模型效果

保持模型名称和模型路径不变，不设置检查点路径，在 Chat 页面直接点击加载模型，即可测试基础大模型的效果。

本案例使用 prompt="**天翼云弹性云主机包含哪些规格**"进行测试。

基础 DeepSeek-R1-Distill-Qwen-7B 大模型给出了通用但是错误的回答。



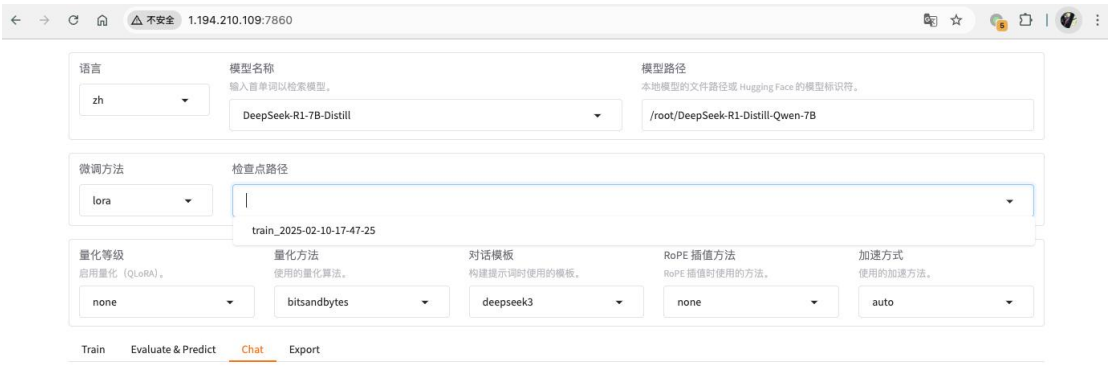


2. 测试微调大模型的效果

本案例采用了 lora 的微调方法，因此需将检查点路径设置为步骤二中完成训练的 lora 小模型。

本案例使用 prompt="天翼云弹性云主机包含哪些规格"进行测试。

微调后的大模型能够给出符合微调数据集的回答。





从纯净版 ubuntu 系统构建微调环境

除了使用预装镜像外，您也可以从任意 ubuntu 云主机开始，按以下步骤搭建出 LLaMA-Factory 微调环境，并下载任意您需要的模型。

步骤一：安装 Nvidia 显卡驱动与 CUDA 库

```
# 安装 nvidia 显卡驱动(如已选择带驱动镜像，则忽略)apt updateapt install ubuntu-drivers-common alsa-utils# 查看推荐的驱动版本 ubuntu-drivers devices# 安装推荐版本 apt install nvidia-driver-550# 重启一次云主机 reboot

# 安装 cuda-12.2wget https://developer.download.nvidia.com/compute/cuda/12.2.0/local_installers/cuda_12.2.0_535.54.03_linux.run# 注意取消对 nvidia driver 的安装 sh cuda_12.2.0_535.54.03_linux.run# 安装完成后在~/.bashrc 中写入环境变量# 添加 CUDA 可执行文件路径 export PATH=/usr/local/cuda-12.2/bin${PATH:+:${PATH}}# 添加 CUDA 库文件路径 export LD_LIBRARY_PATH=/usr/local/cuda-12.2/lib64${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}# 使环境变量生效 source ~/.bashrc
```

步骤二：安装 git 与下载模型

```
# 安装 git 和 lfs

apt update

apt install git

curl -s
https://packagecloud.io/install/repositories/github/git-lfs/script.d
eb.sh | bash

apt install git-lfs

git lfs install

# 下载模型# 国内用 modelscope 源速度最快

git clone
https://www.modelscope.cn/deepseek-ai/DeepSeek-R1-Distill-Qwen-7B.gi
t# 清理掉不用的.git 目录下 lfs 大文件版本数据(能节省一半存储空间)

git lfs prune -f
```

步骤三：安装 LLaMA-Factory 及相关训练环境

```
# 安装 miniconda

wget
https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.s
h# 下载完成后手动安装

sh Miniconda3-[版本号]-Linux-x86_64.sh# 使环境变量生效 source
~/.bashrc

# 安装 LLaMA-Factory

git clone --depth 1 https://github.com/hiyouga/LLaMA-Factory.gitcd
LLaMA-Factory

pip install -e ".[torch,metrics]" -i
https://pypi.tuna.tsinghua.edu.cn/simple
```

```
pip install deepspeed -i https://pypi.tuna.tsinghua.edu.cn/simple

pip install flash-attn --no-build-isolation -i
https://pypi.tuna.tsinghua.edu.cn/simple
```

可选步骤 1：安装 vllm 推理环境

```
# 创建新的虚拟环境(vllm和llama-factory依赖库版本不同，无法共存)conda
create -n vllm_env python=3.12conda activate vllm_env# 安装vllm用于
推理pip install vllm bitsandbytes -i
https://pypi.tuna.tsinghua.edu.cn/simple

# 启动vllm（可参考vllm官方文档vllm serve命令）
```

可选步骤 2：安装 open-webui 可视化界面

```
# 拉取官方docker镜像(需自行配置网络代理)

docker pull ghcr.io/open-webui/open-webui:main

# 启动容器(对照vllm启动参数，自行配置API_KEY，URL参数)

docker run -itd --network host -v
/root/volume/open-webui:/app/backend/data -e PORT=3000 -e
ENABLE_SIGNUP=true -e OPENAI_API_BASE_URL=http://127.0.0.1:8000/v1 -e
OPENAI_API_KEY=${API_KEY} -e HF_ENDPOINT=https://www.modelscope.cn
--name open-webui --restart always ghcr.io/open-webui/open-webui:main
```

10.38.6 基于 CPU 的 AI 推理加速技术在天翼云 EMR 云主机上的应用

1. 背景

1.1 大模型算力需求

大模型因其参数规模巨大、结构复杂，通常需要强大的计算资源来支持其推理过程，这使得算力成为大模型应用的关键要素。信通院的数据资料显示云端推理的市场潜力很大，AI 大模型落地应用数量正在不断增加，市场对大模型推理算力

的需求也呈现爆发式增长态势。

近年来，CPU 制程和架构的提升以及英特尔 AMX 指令集的面世带来了算力的快速提升。英特尔对大模型推理等多个 AI 领域持续深入研究，提供全方位的 AI 软件支持，兼容主流 AI 软件且提供多种软件方式提升 CPU 的 AI 性能。目前，已有充分的数据显示 CPU 可以用于大模型推理。

1.2 CPU 算力用于大模型推理的适用场景

CPU 适用于以下大模型推理场景，

场景 1：大模型推理需要的内存超过了单块 GPU 的显存容量，需要多块或更高配 GPU 卡，采用 CPU 方案，可以降低成本；

场景 2：应用请求量小，GPU 利用率低，采用 CPU 推理，资源划分的粒度更小，可有效降低起建成本；

场景 3：GPU 资源紧缺，CPU 更容易获取，且可以胜任大模型推理。

2. 英特尔 AI 推理加速技术

2.1 英特尔 AI 硬件基础

第三代英特尔® 至强® 可扩展处理器（Ice Lake ） ：

最大支持 40 个物理核，最大支持内存频率 3200MT/s，最大支持 64 条 PCIe 4.0。

搭载英特尔® 高级矢量扩展 512（英特尔® AVX-512）技术。

可用：天翼云第七代云主机（s7/c7/m7）。

第四代英特尔® 至强® 可扩展处理器（Sapphire Rapids ） ：

最大支持 60 个物理核，最大支持内存频率 4800MT/s，最大支持 80 条 PCIe 5.0。

搭载英特尔® 高级矩阵扩展 (Advanced Matrix eXtensions)技术。

AMX 支持 BF16 和 INT8，通过加速矩阵乘法运算，极大提升 AI 应用的性能。

可用：天翼云第八代云主机（s8r）。

第五代英特尔® 至强® 可扩展处理器（Emerald Rapids ） ：

最大支持 64 个物理核，最大支持内存频率 5600MT/s，L3 cache 高达 5MB/core，最大支持 80 条 PCIe 5.0。

搭载英特尔® 高级矩阵扩展 (Advanced Matrix eXtensions)技术。

AMX 支持 BF16 和 INT8，通过加速矩阵乘法运算，极大提升 AI 应用的性能。

相较于第四代处理器，第五代处理器具有更大的带宽，在大语言模型推理领域可带来最高 1.5 倍的性能提升。

可用：天翼云第八代云主机（c8e）。

2.2 英特尔 AI 软件工具

英特尔提供三个层面的 AI 软件支持：



2.2.1 Intel® Extension for PyTorch*

Intel® Extension for PyTorch*（旧称 Intel PyTorch Extension, IPEX）能够优化 AI 推理任务在英特尔硬件上的性能，支持 AVX-512 与 AMX。

2.2.2 xFasterTransformer

xFasterTransformer（xFT）是由英特尔官方开源的 AI 推理框架，为大语言模型（LLM）在英特尔®至强®硬件平台上提供了一种深度优化的解决方案。xFT 支持 AVX-512 与 AMX，并可在 AMX 上获得极致的推理性能。xFT 目前支持的接口与模型如下：

Models	Framework		Distribution
	PyTorch	C++	
ChatGLM	✓	✓	✓
ChatGLM2	✓	✓	✓
ChatGLM3	✓	✓	✓
Llama	✓	✓	✓
Llama2	✓	✓	✓
Baichuan	✓	✓	✓
QWen	✓	✓	✓
SecLLM(YaRN-Llama)	✓	✓	✓
Opt	✓	✓	✓
Deepseek-coder	✓	✓	✓

（更新时间：2024-04-09）

xFT 还可支持多种数据类型，满足用户对不同精度的需求。xFT 目前支持的数据类型包括：

FP16 / BF16 / INT8 / W8A8 / INT4 / NF4 /

BF16_FP16 / BF16_INT8 / BF16_W8A8 / BF16_INT4 / BF16_NF4 / W8A8_INT8 / W8A8_INT4 / W8A8_NF4（更新时间：2024-04-09）

xFT 部署简便，您只需要一键安装插件，并在程序中加入几行代码，即可使用 xFT 对您的模型进行加速。具体实践可参考本文第 3 节的内容。

3. 英特尔 AI 软件工具在天翼云主机上的应用实践

本节内容主要介绍基于 Intel® Extension for PyTorch*与 xFasterTransformer，能够在天翼云主机上取得的性能指标，以及具体实施方法。

3.1 AI 推理性能指标

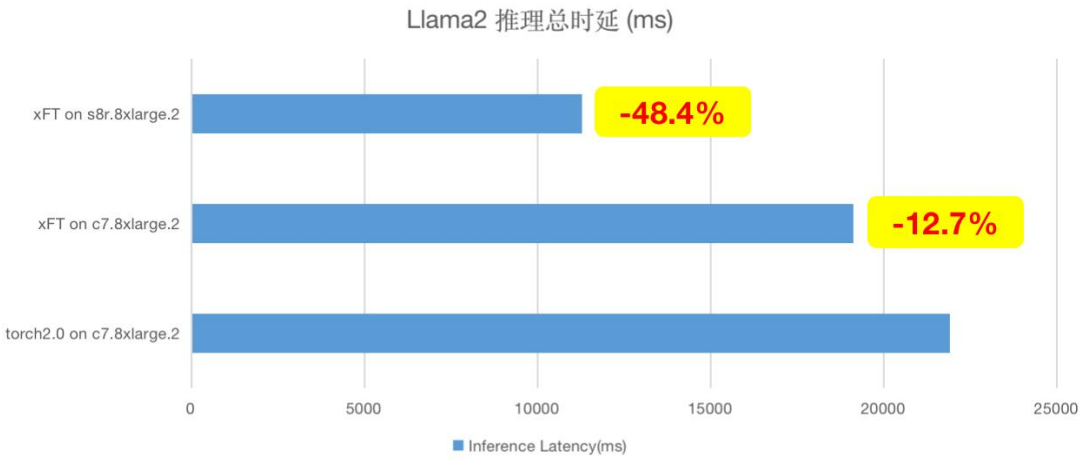
3.1.1 LLaMA2 性能指标

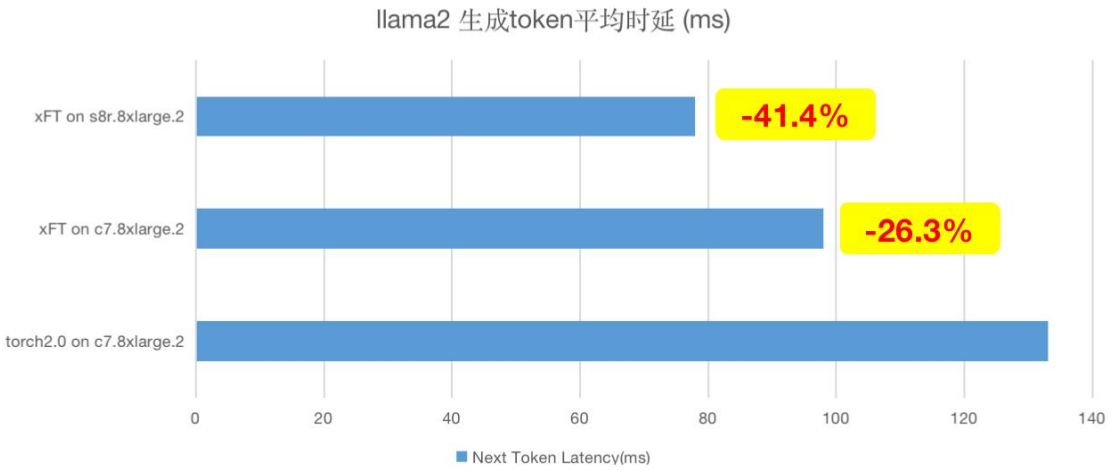
基本参数

- vcpu 数：32
- 内存：64GB

- 硬盘：通用型 SSD
- 模型：Llama-2-7b-chat
- 输入 token 个数：128
- 输出 token 个数：128
- batch size：1
- 数据类型：torch2.0(BF16)，xFT(FP16)
- 测试机型：c7.8xlarge.2 | 32 核 | 64G (numa node 1)，s8r.8xlarge.2 | 32 核 | 64G (numa node 2)

性能数据





xFT 在推理总时延(Inference Latency)和生成 token 时延(Next Token Latency)上都有明显的优化。

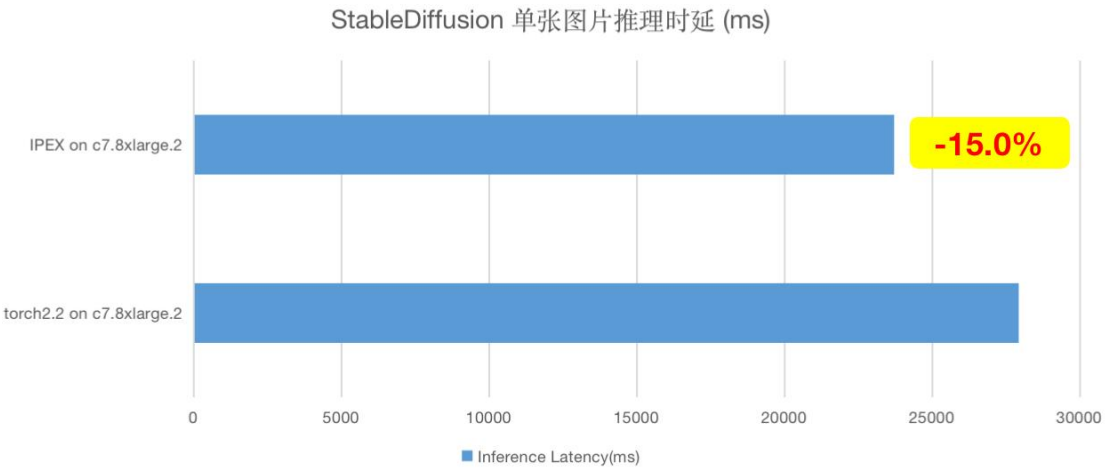
3.1.2 StableDiffusion 性能指标

基本参数

- vcpu 数：32
- 内存：64GB
- 硬盘：通用型 SSD
- 模型：stable-diffusion-2-1
- 输入 token 个数：6
- 输出图片尺寸：512 x 512
- 推理 Steps：20
- batch size：1
- 数据类型：torch2.0(FP32)，IPEX(FP32)

性能数据

IPEX 在推理总时延上有一定的性能提升。



3.1.3 EMR 云主机 LLM 性能指标

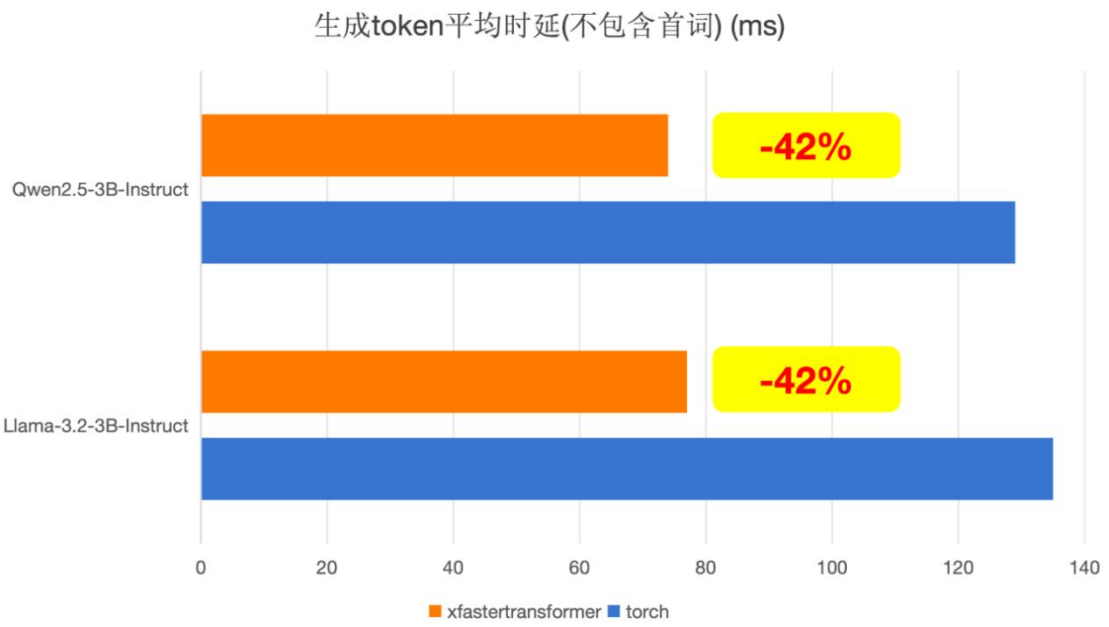
> 注意
>
> 在天翼云 EMR 云主机正式上线后，我们更新性能指标如下。

基本参数

- vcpu 数：16
- 内存：64GB
- 硬盘：通用型 SSD
- 输入 token 个数：128
- 输出 token 个数：128
- batch size：1

- 数据类型：torch2.0(BF16)，xFT(BF16)

性能数据



xFT 在生成 token 时延（Next Token Latency）上有明显的优化，在 16vcpu 的机型上能使用主流 3B 规格模型达到低于 80ms/token 的生成速度，足以满足业界需求。

3.1.4 EMR 云主机 StableDiffusion 性能指标

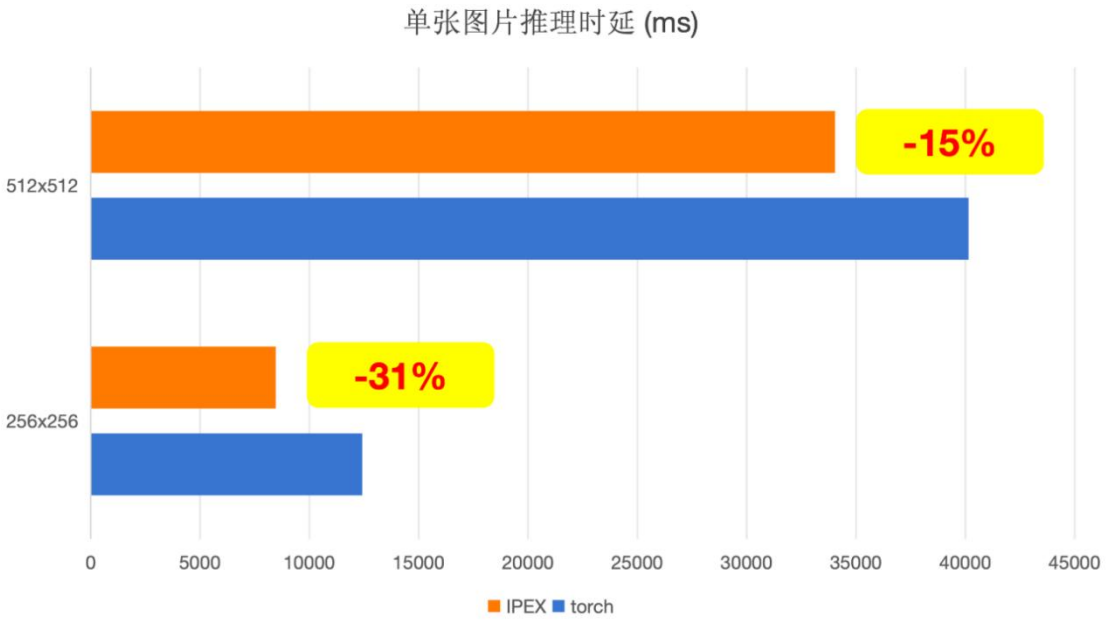
注意

在天翼云 EMR 云主机正式上线后，我们更新性能指标如下。

基本参数

- vcpu 数：16
- 内存：64GB
- 硬盘：通用型 SSD
- 模型：stable-diffusion-2-1
- 输入 token 个数：<16
- 推理 Steps：20
- batch size：1
- 数据类型：torch(FP32)，IPEX(FP32)

性能数据



IPEX 在推理总时延上有一定的提升，在本示例 16vcpu 和 20steps 配置下能够将 256x256 大小图片生成时延压缩到 10 秒内，512x512 大小图片生成时延压缩到 35 秒内，具有一定的实用价值。

3.2 AI 软件工具实施方法（基于一键部署镜像）

为了方便您进行性能测试，天翼云联合英特尔制作了一键部署的云主机镜像，内置 torch2.0 环境，Intel® Extension for PyTorch*与 xFasterTransformer 软件，以及 Llama-2-7b-chat 与 stable-diffusion-2-1 模型。您可直接开通云主机并选择对应镜像进行体验。

镜像支持的云主机规格与资源池信息如下表：

注意

内存需在 32G 及以上。

	云主机规格	资源池
英特尔第三代至强可扩展处理器 AVX-512	s7 / c7 / m7	西南-西南 1 华东-华东 1
英特尔第四代至强可扩展处理器 AMX	s8r	华东-华东 1 -可用区 2
英特尔第五代至强可扩展处理器 AMX	c8e	华东-华东 1 -可用区 2

注意

镜像内置的是天翼云七代主机(s7/c7/m7)特殊编译版本的 xfastertransformer，如果您使用的是天翼云八代机(s8r)，请在 xFT 环境下手动更新安装新版本，否则版本不兼容，可参考下方的示例代码。

```
conda activate xFTpip install xfastertransformer==1.4.0 -i https://pypi.tuna.tsinghua.edu.cn/simple
```

镜像名称：应用镜像 - ubuntu - Intel-AI 推理加速，如图所示：



3.2.1 xFasterTransformer (xFT)实施方法

3.2.1.1 模型转换为 xFT 格式

在使用 xFT 框架前，需要对模型进行格式转换。xFT 提供了转换命令，您可修改和执行/root/tools/xft_trans.py 脚本。

```
import xfastertransformer as xft# 输入 huggingface 模型路径 HF_DIR ="/root/models/Llama-2-7b-chat-hf"# 输出 xft 模型路径 XFT_DIR ="/root/models/Llama-2-7b-chat-xft"xft.LlamaConvert().convert(HF_DIR, XFT_DIR)
```

在执行前，需要进入 conda 运行环境(执行过一次即可)

```
conda activate xFTpython xft_trans.py
```

3.2.1.2 查看 numa 节点分布

```
numactl -H
```

以下输出代表 32 个 vcpu 均在同一个 numa node 下：

```
(xFT) root@ecm-d23e:~/tools# numactl -H
available: 1 nodes (0)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
node 0 size: 63790 MB
node 0 free: 48631 MB
node distances:
node   0
  0:  10
```

以下输出代表 vcpu0-15 分布在 numa node0，vcpu16-31 分布在 numa node1：

```
(xFT) root@ecm-4295:~# numactl -H
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
node 0 size: 31634 MB
node 0 free: 2157 MB
node 1 cpus: 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
node 1 size: 32251 MB
node 1 free: 2701 MB
node distances:
node   0   1
  0:  10  20
  1:  20  10
```

3.2.1.3 检查测试参数

性能测试所需的参数均在、root/tools/llama_benchmark.py 中，您可按需进行修改，对比开启和关闭 xFT 的效果，以及不同 token 长度的效果等。

tools > llama_benchmark.py

```python
# llama2 hf格式模型地址，默认模型将以bf16精度加载
MODEL_PATH = "/root/models/Llama-2-7b-chat-hf"
# 是否使用xFT
USE_XFT = True
# 使用xFT时需要提前转换模型，参考xft_trans.py，默认模型将以fp16精度加载
MODEL_PATH_XFT = "/root/models/Llama-2-7b-chat-xft"
# 输入TOKEN数，预制的输入明文参见prompt.json，  可选: 22, 32, 64, 128, 256, 512, 1024, 2016, 2017, 2048, 4096, 8192
TOKEN_IN = 128
# 输出新增TOKEN数（总数=TOKEN_IN+TOKEN_OUT_NEW）
TOKEN_OUT_NEW = 128
# 模型预热次数，大于2后指标基本恒定
WARMUP = 2
# 执行次数(最后取均值)
ITERATION = 5
# 是否输出明文结果(仅用于debug)
PRINT_OUTPUT = False
# 读取预制的prompt，输入只需指定token个数.
PROMPT_PATH = "/root/tools/prompt.json"
```

3.2.1.4 单进程执行

注意

如果您的云主机存在不止一个 numa node，直接使用单进程方式执行脚本会出现大幅性能波动，您可参考单核和多核的不同实施方法。

在单进程情况下，您可直接执行 python 脚本：

```
# 在执行前，需要进入 conda 运行环境(执行过一次即可)conda activate xFTcd /root/tools# 绑定 cpu 核，性能更稳定# OMP 线程数和绑定的 cpu 核数保持一致# 天翼云主机开启了 cpu 超线程，0-1/ 2-3 号 vcpu 分别在同一个物理核上，只需要使用其中一个核即可# vcpu 与物理核的对应关系可在/proc/cpuinfo 查看每个 processor 的 physical idOMP_THREAD_NUM=16 numactl --all -C 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 python llama_benchmark.py
```

您将会得到以下输出（本示例采用天翼云 c7.8xlarge.2 | 32 核 | 64G 云主机）：

```
(xFT) root@ecm-d23e:~/tools# OMP_THREAD_NUM=16 numactl --all -C 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 python llama_benchmark.py
OMP_NUM_THREADS: None
temporary set OMP_NUM_THREADS=16
torch thread num: 16
xFT loading model...
[INFO] SINGLE_INSTANCE MODE.
master node run
========llama benchmark========
Input token num 128
Output new token num: 128
Model warmup round: 2
Warmup Iteration 1:     total time: 19245.86 ms, new token num: 128
Warmup Iteration 2:     total time: 19178.66 ms, new token num: 128
Start benchmark:
Iteration 1:    total time: 19096.87 ms, new token num: 128, first token time: 6603.65 ms, throughput(next token): 10.17 tokens/s
Iteration 2:    total time: 19133.18 ms, new token num: 128, first token time: 6612.43 ms, throughput(next token): 10.14 tokens/s
Iteration 3:    total time: 19131.97 ms, new token num: 128, first token time: 6612.95 ms, throughput(next token): 10.14 tokens/s
Iteration 4:    total time: 19090.90 ms, new token num: 128, first token time: 6601.22 ms, throughput(next token): 10.17 tokens/s
Iteration 5:    total time: 19109.18 ms, new token num: 128, first token time: 6603.84 ms, throughput(next token): 10.16 tokens/s


============================== Final Performance==============================
Use xFT: True
Model dtype: fp16
Input token num 128
Output new token num: 128
Inference Avg Latency: 19112.42 ms
First token Avg Latency: 6606.82 ms
Next token Avg Latency: 97.70 ms
Throughput(next token): 10.24 tokens/s
================================================================================
```

如果想要和加速前的性能进行对比，您可以修改脚本内的 USE_XFT=False，然后再次执行。

您也可使用测试脚本查看模型生成的结果：

```
# 在执行前，需要进入 conda 运行环境(执行过一次即可)

conda activate xFTcd /root/tools# 修改 prompt 后执行

python llama_run.py
```

3.2.1.5 多进程执行

在多 numa node 场景下，推荐使用多进程获得更好的性能指标。您需要根据上文 3.2.1.2 节 numactl -H 查到的信息，调整 /root/tools/llama_benchmark_multirank.sh 脚本中的相关内容。具体信息在脚本中均有详细注释。

注意

当前多进程脚本只支持 USE_XFT=True 的模式，关闭 xFT 加速时无法利用多进程。

```
# 在执行前，需要进入 conda 运行环境(执行过一次即可)

conda activate xFTcd /root/tools

bash llama_benchmark_multirank.sh
```

在 numa node 分配正确的情况下，天翼云 s8r.8xlarge.2 | 32 核 | 64G 云主机可达到如下指标：

```
=============================== Final Performance===============================
Use xFT: True
Model dtype: fp16
Input token num 128
Output new token num: 128
Inference Avg Latency: 11331.35 ms
First token Avg Latency: 1245.55 ms
Next token Avg Latency: 78.80 ms
Throughput(next token): 12.69 tokens/s
================================================================================
```

对比来看，如果此时坚持使用单进程模式，带来了跨 numa node 的性能损耗，则只能达到如下指标：

```
=============================== Final Performance===============================
Use xFT: True
Model dtype: fp16
Input token num 128
Output new token num: 128
Inference Avg Latency: 13058.03 ms
First token Avg Latency: 1051.49 ms
Next token Avg Latency: 93.80 ms
Throughput(next token): 10.66 tokens/s
================================================================================
```

可以明显看到性能下降。

3.2.2 Intel® Extension for PyTorch*(IPEX)实施方法

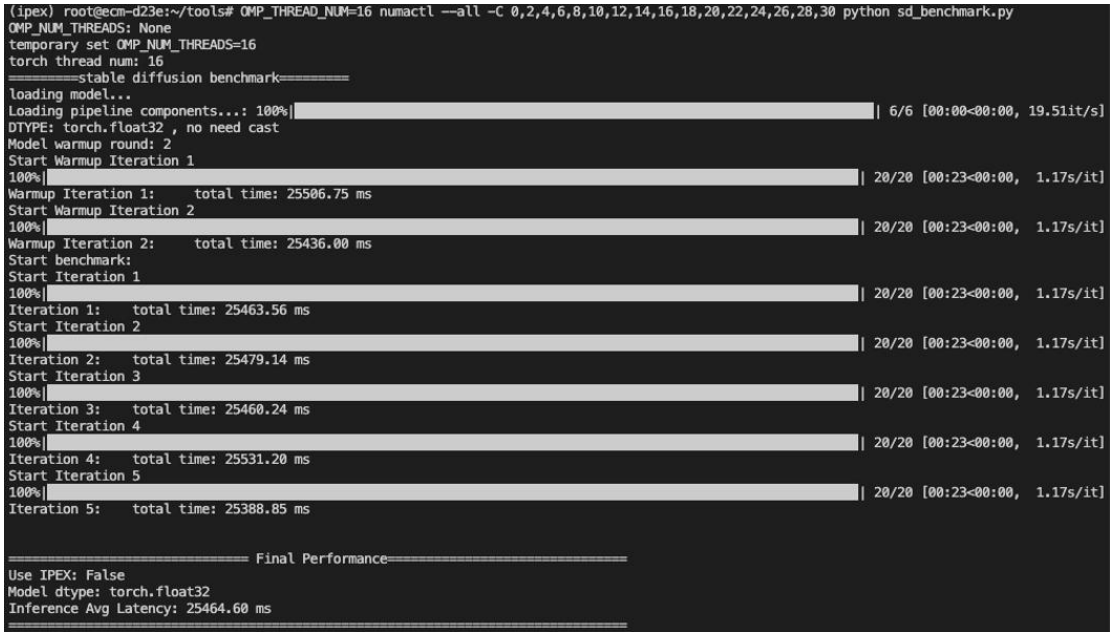
OPEX 的应用和性能测试可直接按单进程执行：

```
# 在执行前，需要进入 conda 运行环境(执行过一次即可)

conda activate ipex

python sd_benchmmark.py
```

您将会得到如下输出：



如果想要和加速前性能进行对比，可以修改脚本内的 USE_IPEX=False 然后再次执行。

4. 附录：一键部署的镜像环境制作

```
########## base environment ##########

# 选择一个 miniconda 版本下载安装:

https://repo.anaconda.com/miniconda/

wget

https://repo.anaconda.com/miniconda/Miniconda3-py310_24.1.2-0-Linux-x86_64.sh

bash Miniconda3-py310_24.1.2-0-Linux-x86_64.sh

source ~/.bashrc

# 安装 numactl 工具

apt install numactl

# 安装 git lfs

curl -s

https://packagecloud.io/install/repositories/github/git-lfs/script.deb.sh | sudo bash
```

```
apt-get install git-lfs

git lfs install

########## xFasterTransformer ##########

conda create -n xFT python=3.10

conda activate xFT

pip install torch==2.0.1 torchvision==0.15.2 torchaudio==2.0.2
--index-url https://download.pytorch.org/whl/cpu

pip install transformers==4.36.0 sentencepiece protobuf accelerate
datasets cmake -i https://pypi.tuna.tsinghua.edu.cn/simple

# icelake AVX-512 架构 (特殊编译版本)

pip install xfastertransformer-1.3.1.icx-py3-none-any.whl

# spr/emr AMX 架构 (当前最新版本为 1.4.0)

pip install xfastertransformer==1.4.0 -i
https://pypi.tuna.tsinghua.edu.cn/simple

# 从 huggingface 下载 llama-2-7b-chat 模型


########## Intel® Extension for PyTorch* ##########

pip install torch torchvision torchaudio --index-url
https://download.pytorch.org/whl/cpu

pip install intel-extension-for-pytorch

pip install oneccl_bind_pt --extra-index-url
https://pytorch-extension.intel.com/release-whl/stable/cpu/us/


pip install transformers accelerate diffusers
```

```
# 从 huggingface 下载 stable-diffusion-2-1 模型
# 注意 sd 模型仓库内放了太多不同版本的模型,实际只需要下载其中一版即可。
# 只下载小文件
GIT_LFS_SKIP_SMUDGE=1 git clone
https://huggingface.co/stabilityai/stable-diffusion-2-1
# 然后根据需要下载所需文件
cd stable-diffusion-2-1
git lfs pull --include=" pytorch_model.bin"
git lfs pull --include=" diffusion_pytorch_model.bin"

# /root/tools/stable_diffusion_ipex/pipeline.py 下载来自
https://github.com/huggingface/diffusers/blob/main/examples/community/stable_diffusion_ipex.py
```

10.38.7 基于 Open WebUI 实现 DeepSeek API 调用及外部大模型 API 接入最佳实践

使用场景

大模型 API 提供了接口以便用户更好地进行 DeepSeek 管理和应用开发。用户在开通天翼云 DeepSeek 云主机后，利用 Open WebUI API 即可实现对外提供 API 调用，同时也可以非常方便的实现外部大模型 API 接入使用。

本教程使用的 GPU 云主机规格如下所示，用于部署 DeepSeek-r1:7b 模型，配置仅供参考。

```
cpu 16 核
内存  64G
GPU:  NVIDIA  A10*1  (24GB)
```

如需体验其它模型版本，请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--自定义部署 DeepSeek--步骤二：规格选型，选择合适的云主机。

准备

参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7b 版-弹性云主机-最佳实践-AIGC 实践 - 天翼云--快速体验 DeepSeek，准备相关设备及模型资源。

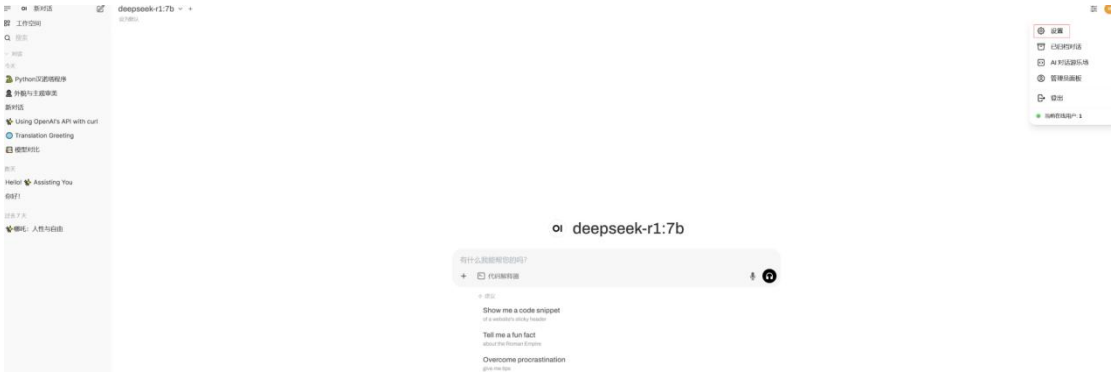
认证

OpenWebUI 使用 Bearer Token 认证保证客户端请求的合法性，其 Token 为 APIKEY 中 JWT 令牌(Json Web Token)。获取 API key 方式如下：

1. 登录 Open WebUI。



2. 选择设置->账户，选择获取或者新建 API KEY。





API 接口

1.列举模型

URL：GET /api/models

描述：列出可用的模型列表，并提供相关模型的基本信息。

例子:

```
curl -H "Authorization: Bearer API_KEY" http://公网 IP:3000/api/models
```

2.对话补全

URL：POST /api/chat/completions

描述：根据输入的上下文，来让模型补全对话内容。

例子：

```
curl -L -X POST 'http://公网 IP:3000/api/chat/completions' \
-H 'Content-Type:application/json' \
-H 'Accept:application/json' \
-H 'Authorization:BearerAPI_KEY' \
-d '{
```

```
  "model":"deepseek-r1:7b",
   "messages": [
   {
    "content":"用英语翻译一下你好",
    "role":"user"
   }
 ],
 "stream":false
}'
```

使用外部大模型

Open WebUI 通过云主机方式部署完自己的 DeepSeek 后，可以直接嵌入使用外部大模型，嵌入后可以通过 WEB、API 进行模型使用。配置方式如下：

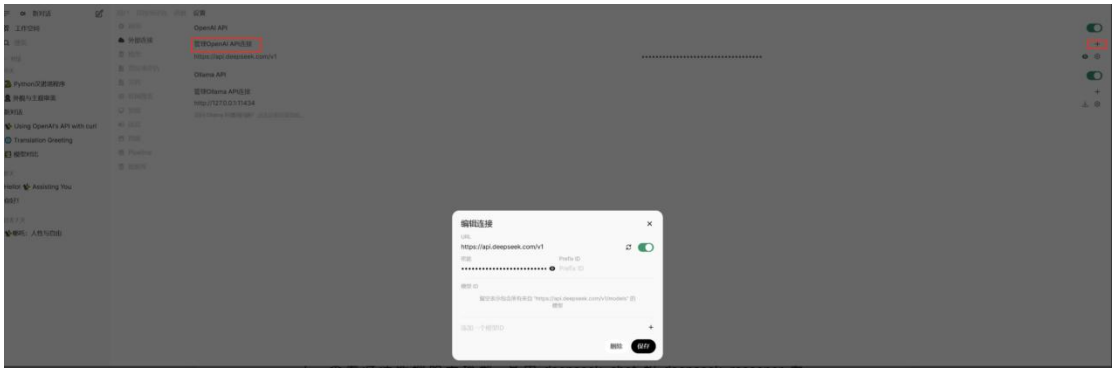
1. 登录 Open WebUI。

2. 选择管理员面板->设置->外部连接。



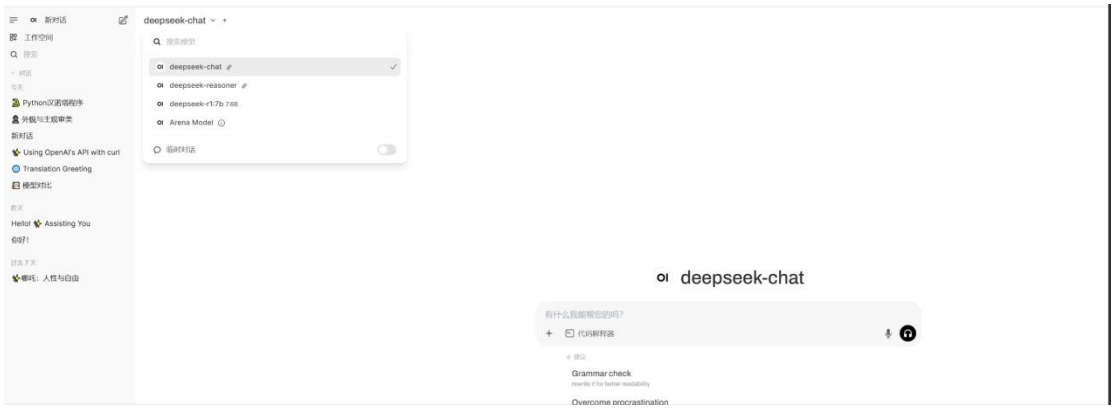


3. 选择 OpenAI API 开启，点击管理 OpenAI API 连接，新增外部大模型连接，输入外部模型的 URL 和 API KEY。



4. 点击保存。

5. 查看对话新增的大模型，其中 DeepSeek-chat 和 DeepSeek-reasoner 为 DeepSeek 官网新增模型。

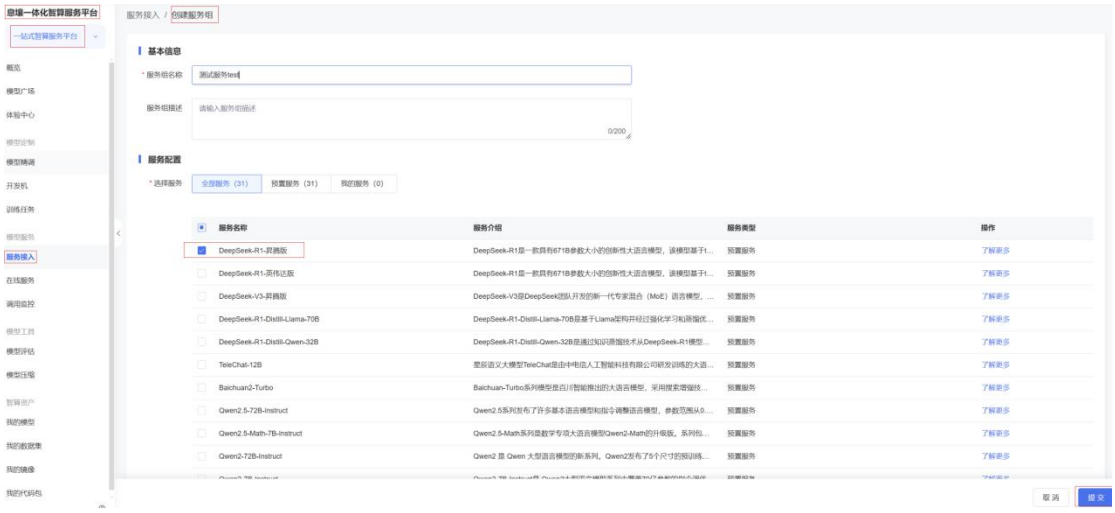


使用天翼云息壤一体化智算服务平台

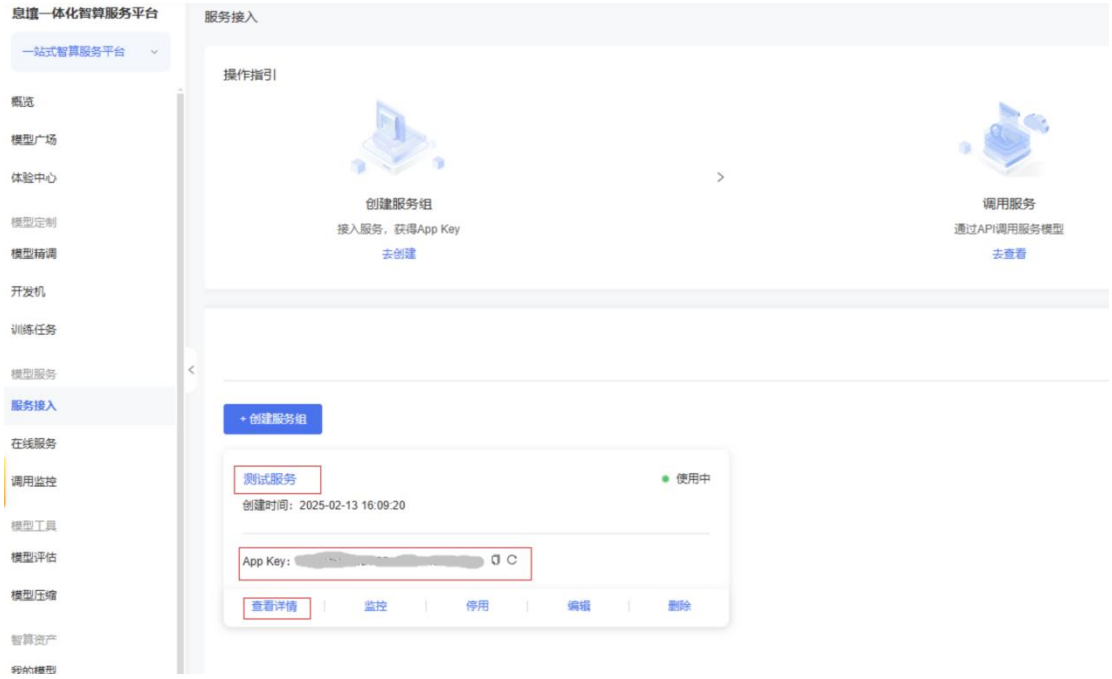
准备工作

使用天翼云息壤提供的大模型不仅需要提供 API KEY，还需要提供对应的模型 ID。其获取方式如下：

1. 登录天翼云官网，选择息壤一体化智算服务平台->一站式智算服务平台->服务接入->创建服务组->DeepSeek-R1-昇腾版，点击提交。



2.选择测试服务，获 App key 作为 API KEY。选择查看详情->了解更多。



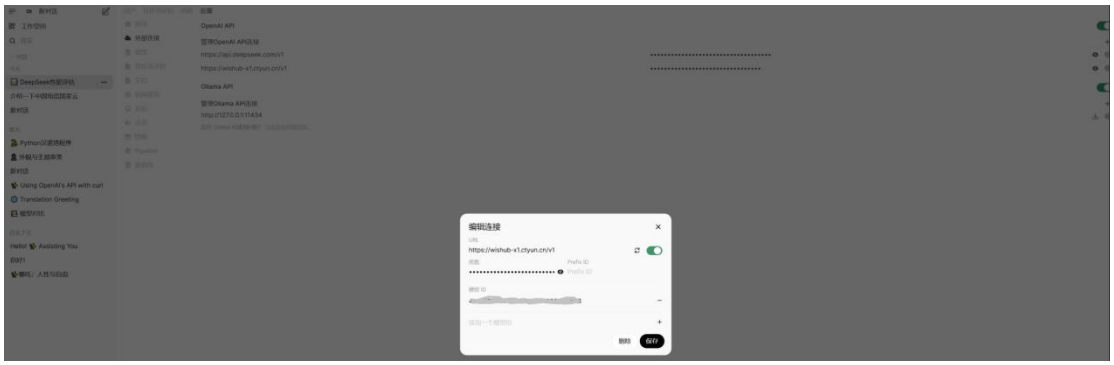
3.选择 API 文档，获取模型 ID 和终端地址。



Open WebUI 接入

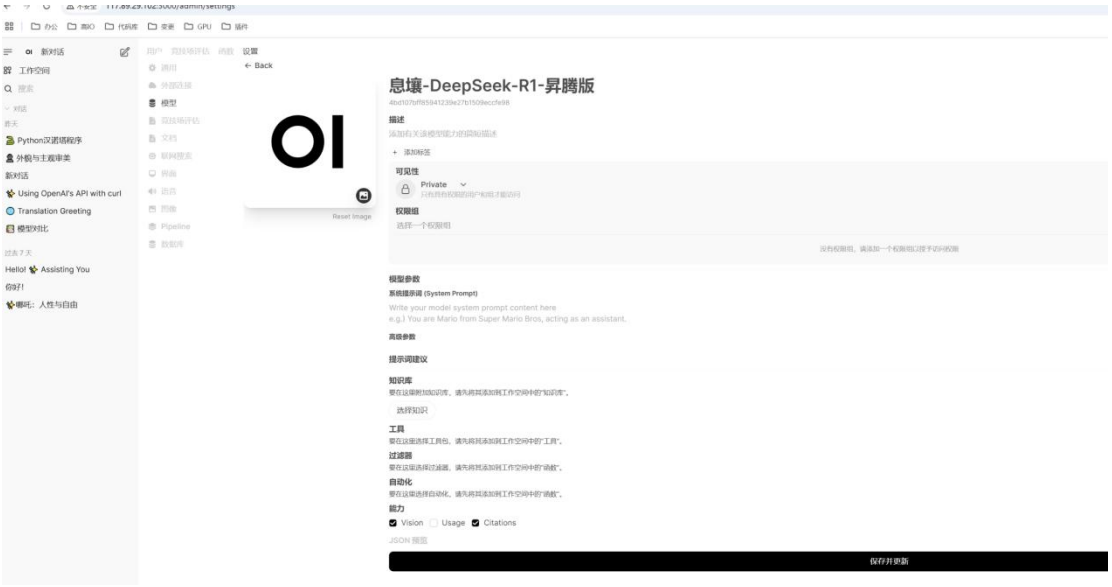
1. 登录 Open WebUI 并选择管理员面板->设置->外部连接。

2. 选择 OpenAI API 开启，点击管理 OpenAI API 连接，新增息壤大模型终端地址、API KEY 和模型 ID。终端地址需要增加 V1 版本。



3. 点击保存。

4. 点击模型，修改模型 ID 为易理解的模型。



5.查看对话新增的大模型，其中息壤-DeepSeek-R1-昇腾版为新增模型。



注意

请妥善保管您的 API KEY，不要在公共场合泄露。

10.38.8AnythingLLM 协同 Ollama 启用 DeepSeek 大模型最佳实践

推荐使用场景

工具	推荐场景	不适用场景
AnythingLLM	企业文档问答、私有知识库管理、需严格权限控制的团队协作	快速测试模型、轻量级个人聊天需求
Open WebUI	个人或开发者快速体验模型、多模型切换测试、多模态交互(如语音/图像)	复杂文档管理、企业级安全与协作需求

注意

建议使用天翼云提供的 DeepSeek 镜像创建云主机，减少安装过程中可能遇到的问题。镜像选择请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本-弹性云主机-最佳实践-AIGC 实践 - 天翼云。

操作步骤：

一、云主机端口配置更新

1. 进入云主机详情页，选择安全组选项，添加新的开放端口。

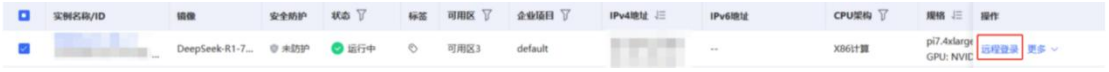
建议选择新加 3001 端口。添加完成后，可在安全组规则列表中查看新添加的 3001 端口规则，确认其状态为正常生效。







2.进入控制台，远程登陆即将安装 AnythingLLM 的云主机。



二、AnythingLLM 服务部署（ docker 方式）

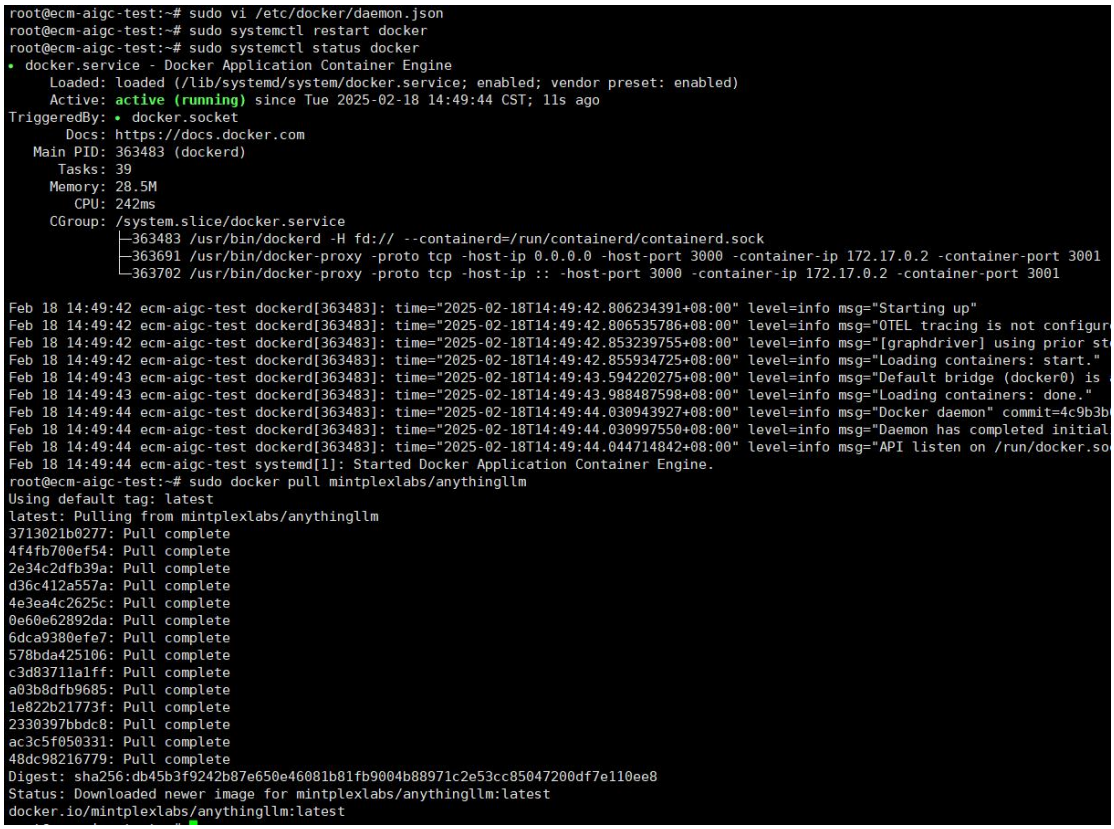
1.获取 AnythingLLM docker image 文件。

由于国内无法访问 docker hub 网站，您需要修改 docker 的服务配置文件，添加国内可访问的源，以便能够顺利拉取 AnythingLLM docker image 文件。操作步骤如下：

- 向 docker 配置文件中添加国内可访问的 docker 镜像源
- 重启 docker 服务
- 拉取 AnythingLLM docker image

```
# Docker 守护进程（ daemon ）的配置文件。添加国内可访问镜像源。sudo vi /etc/docker/daemon.json# 重启 docker 服务 sudo systemctl restart docker# 确认 docker 服务正常 sudo systemctl status docker# 拉取 anythingllm 镜像

docker pull mintplexlabs/anythingllm
```



daemon.json 文件添加的内容为：

```
{
    "registry-mirrors": ["http://hub.geekery.cn",
"http://hub.littlediary.cn", "http://docker.rainbond.cc"]
}
```

daemon.json 文件如下，仅供参考：

```
{
    "registry-mirrors": ["http://hub.geekery.cn", "http://hub.littlediary.cn", "http://docker.rainbond.cc"]
}
```

2.启动 AnythingLLM 服务。

- 首先查看下载的 AnythingLLM 容器镜像的 tag
- 启动 AnythingLLM 服务
- 查看 AnythingLLM 服务状态

```
# 查看当前的容器镜像列表

docker image list# 启动 AnythingLLM 服务，下述 PROT 参数需要与第一部分开放的端口相对应
```

```
docker run -d -e PORT=3001 -e OLLAMA_BASE_URL=http://127.0.0.1:11434 -e
ENABLE_SIGNUP=true -e ENABLE_OPENAI_API=False
--add-host=host.docker.internal:host-gateway --net=host -v
anythingLLM:/app/backend/data --name anythingLLM --restart always
mintplexlabs/anythingllm:latest# 查看 AnythingLLM 服务状态
docker ps -a
```

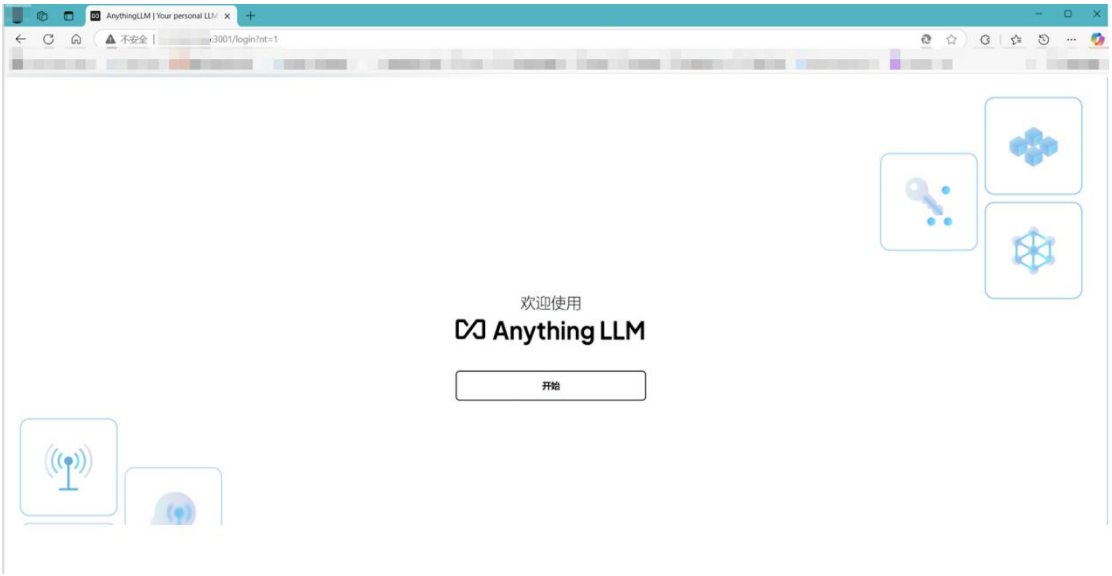
```
root@ecm-aigc-test:~# docker image list
REPOSITORY                          TAG       IMAGE ID       CREATED        SIZE
mintplexlabs/anythingllm            latest    753644541603   2 hours ago    2.51GB
ghcr.io/open-webui/open-webui       main      a088eea70396   13 days ago    4.27GB
```



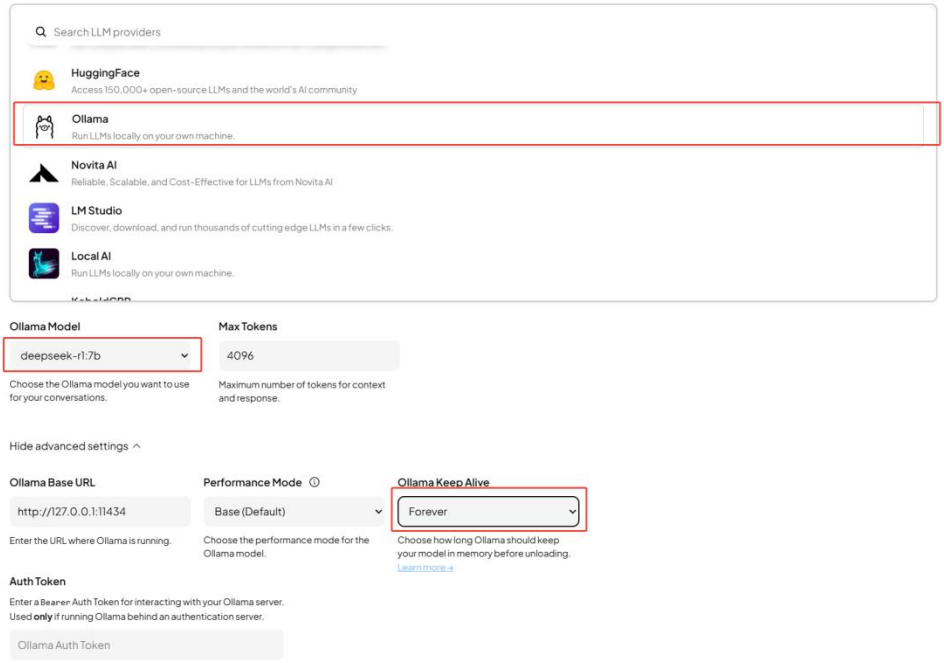
三、AnythingLLM 前端页面配置

1.公网访问 AnythingLLM。

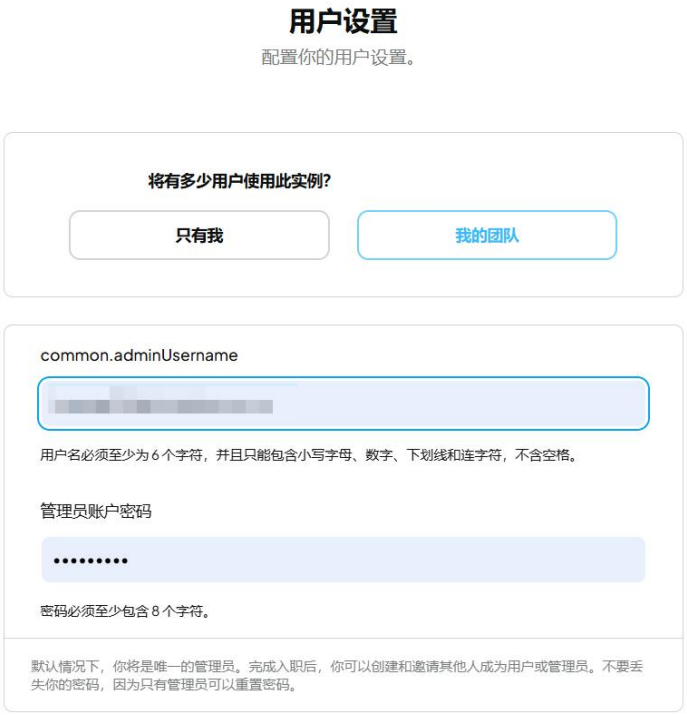
在浏览器输入 http://{your_ip}:3001 (your_ip 这台云主机的 eip)，会出现如下页面

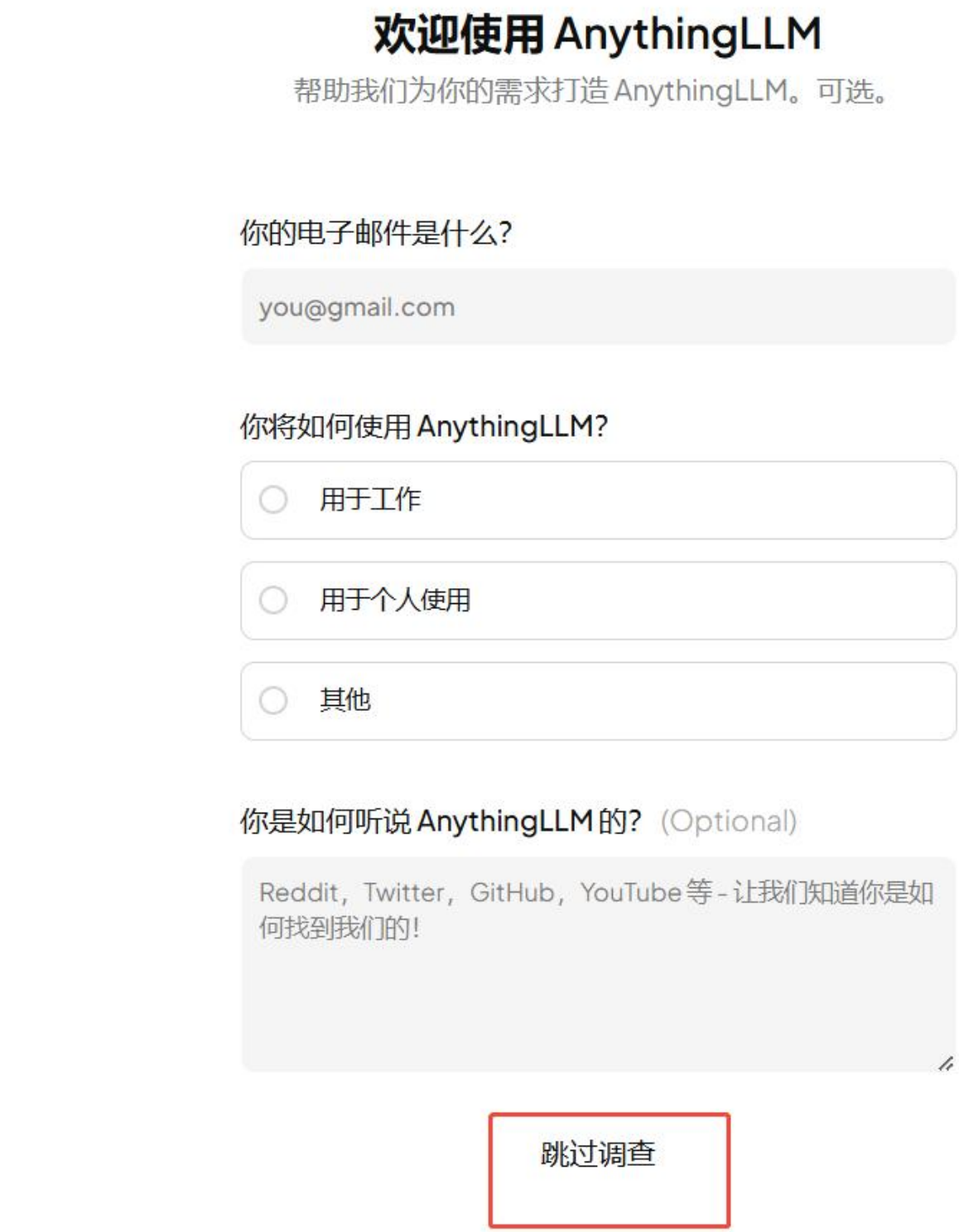


2.选择 Ollama 作为模型加载项。



3. 用户设置，按页面提示操作，注册账号，跳过用户调查。

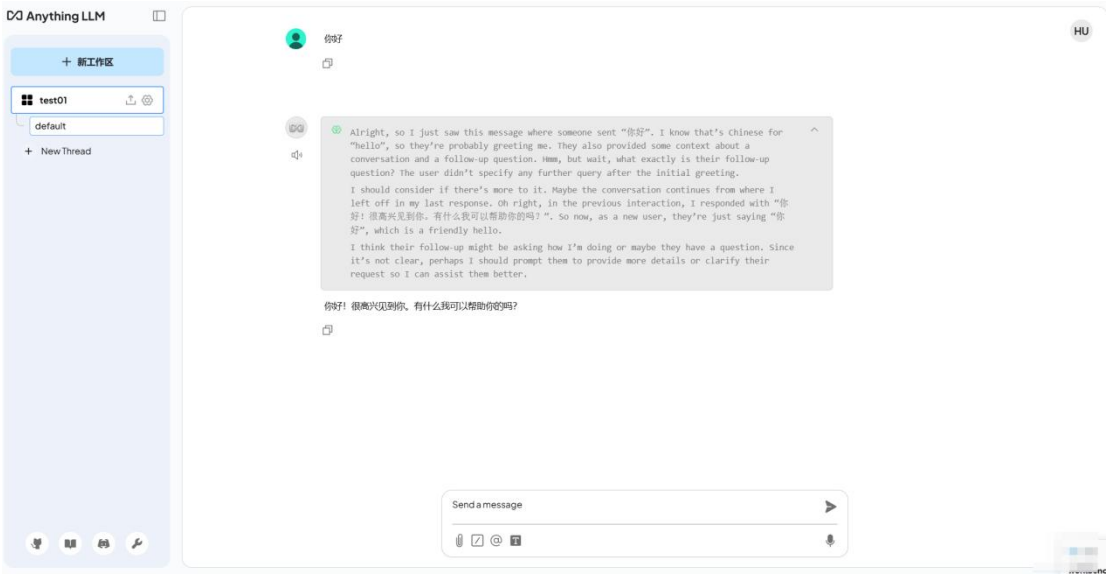



欢迎使用 AnythingLLM
帮助我们为你的需求打造 AnythingLLM。可选。
你的电子邮件是什么?
you@gmail.com
你将如何使用 AnythingLLM?
用于工作
用于个人使用
其他
你是如何听说 AnythingLLM 的? (Optional)
Reddit，Twitter，GitHub，YouTube 等 - 让我们知道你是如何找到我们的!
跳过调查

4. 创建工作区。



5. 点击工作区 test01 打开对话框，输入聊天内容。

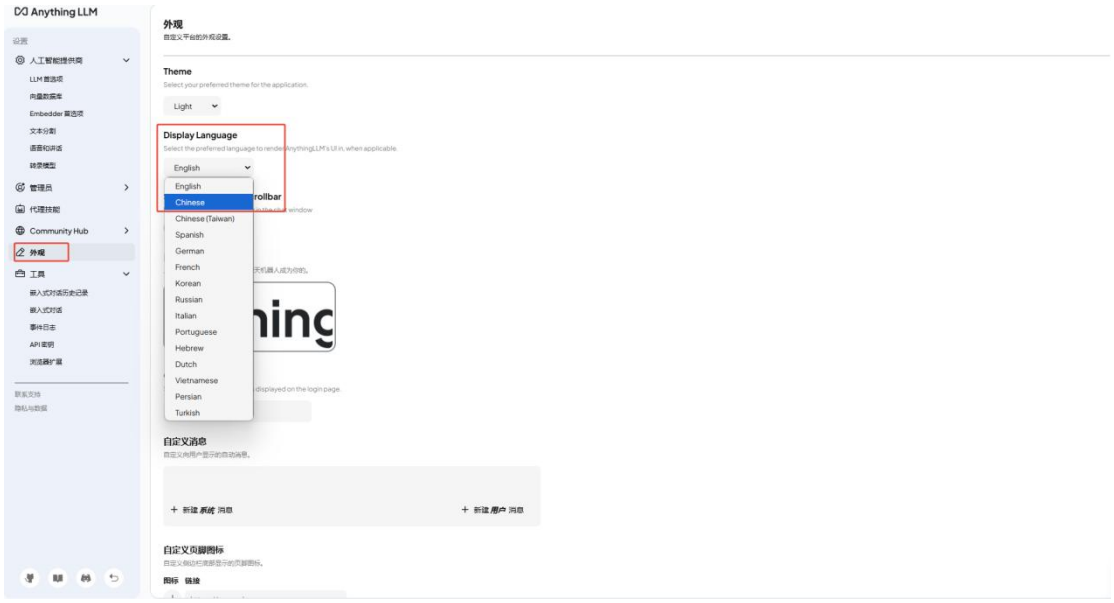


四、AnythingLLM 个性化探索

下面给出了一些个性化探索案例，您如果想要了解更多，请访问 AnythingLLM 社区。

1. 设置 AnythingLLM 默认语言格式为中文。参考下图进行设置。





2. AnythingLLM 启用联网搜索服务。

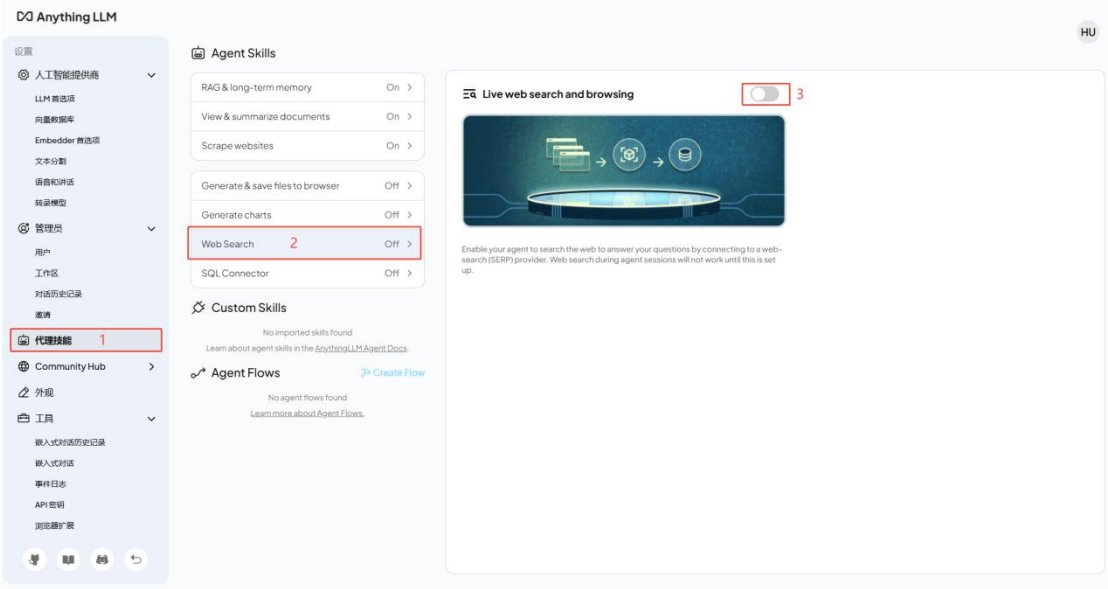
启用 AnythingLLM 的联网搜索服务，需要您提供一个联网搜索引擎 api key 来访问 Web 资源，目前 AnythingLLM 支持的搜索引擎可以参考如下步骤查看。

a. 打开 AnythingLLM 的联网搜索功能

- 首先点击 AnythingLLM 的全局设置按钮。

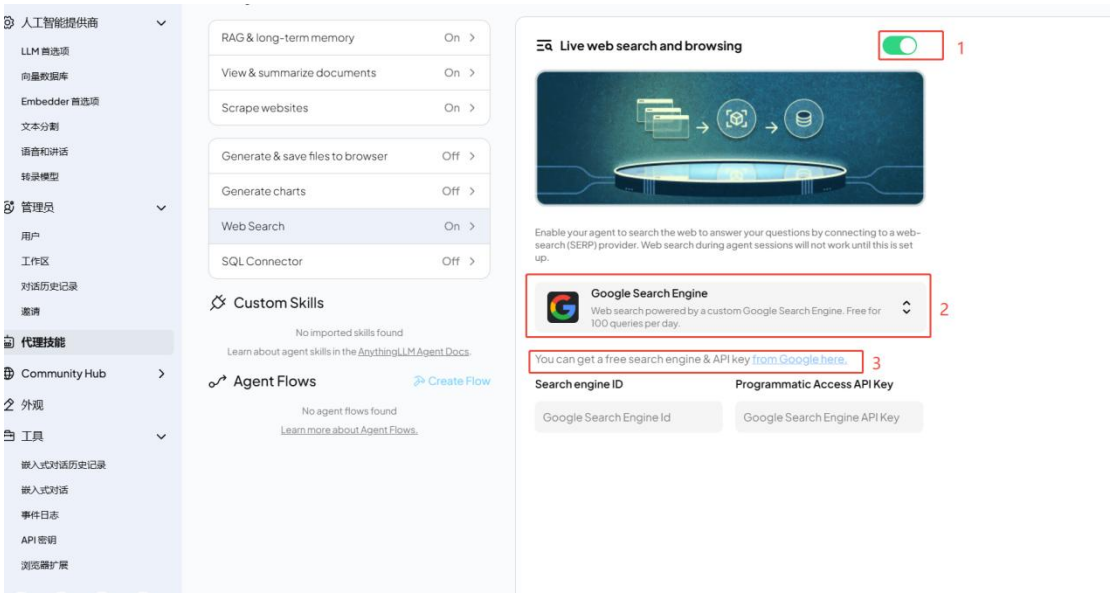


- 参考下图，找到“代理技能”，选择 agent skills 中的 Web Search，点击步骤 3，开启对应功能。



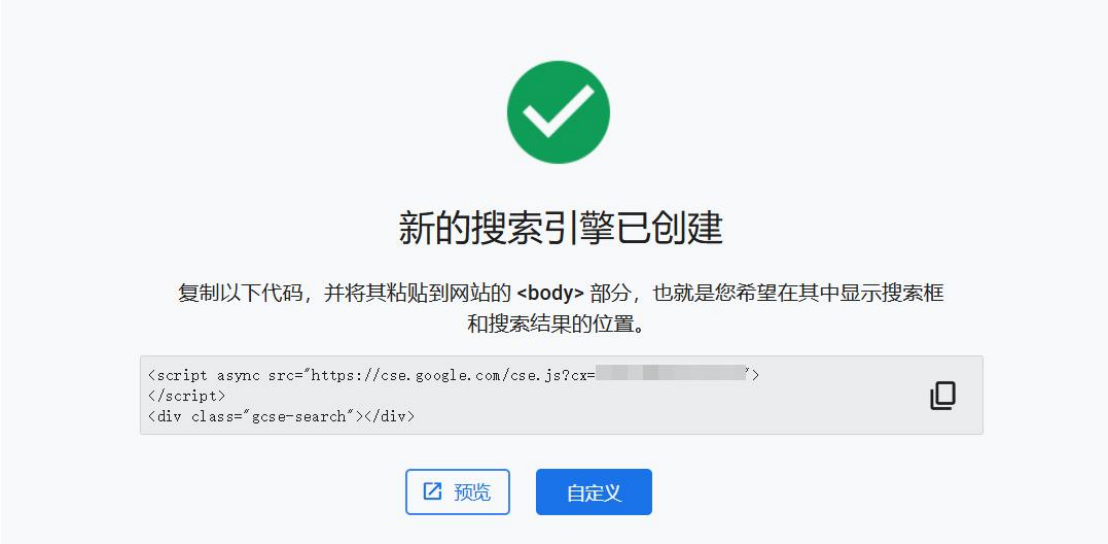
b. 获取 Google 搜索引擎的 ID 以及 API key

- 打开下图中步骤 3 对应的网页，按提示注册或登录账号、命名搜索引擎、选择搜索范围（可输入指定网页或搜索全网），以及选择是否启用图片搜索功能等操作。完成上述设置并通过人机验证后，点击创建，按提示获取搜索引擎的 ID 以及 API key。


人工智能提供商
LLM 首选项
向量数据库
Embedder 首选项
文本分割
语音和讲话
转录模型
管理员
用户
工作区
对话历史记录
邀请
代理技能
Community Hub
外观
工具
嵌入式对话历史记录
嵌入式对话
事件日志
API 密钥
浏览器扩展
RAG & long-term memory On
View & summarize documents On
Scrape websites On
Generate & save files to browser Off
Generate charts Off
Web Search On
SQL Connector Off
Custom Skills
No imported skills found
Learn about agent skills in the AnythingLLM Agent Docs.
Agent Flows
Create Flow
No agent flows found
Learn more about Agent Flows.
Live web search and browsing
1
Enable your agent to search the web to answer your questions by connecting to a web-search (SERP) provider. Web search during agent sessions will not work until this is set up.
Google Search Engine
Web search powered by a custom Google Search Engine. Free for 100 queries per day.
2
You can get a free search engine & API key from Google here.
3
Search engine ID
Programmatic Access API Key
Google Search Engine Id
Google Search Engine API Key



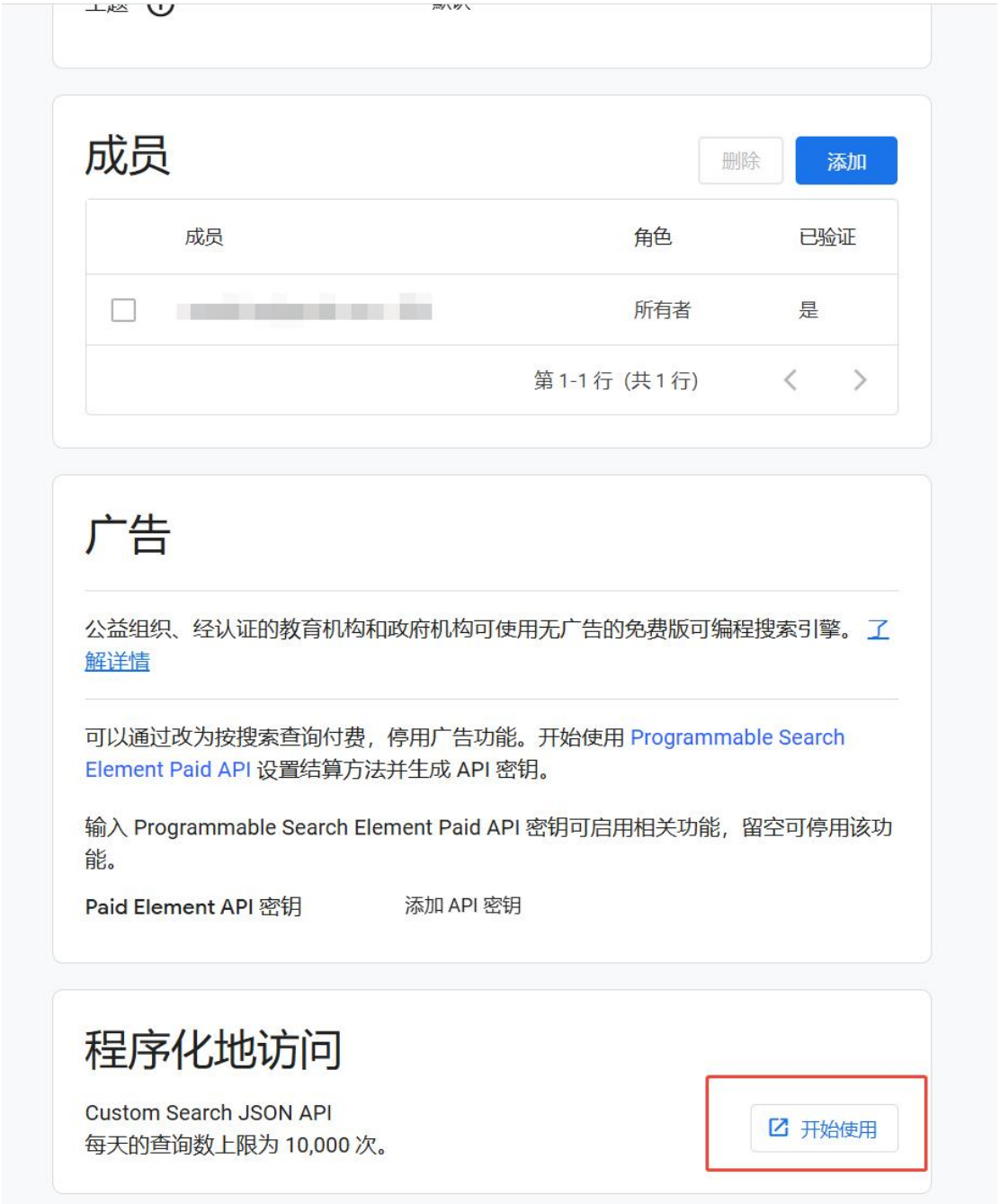
- 创建成功后，得到如下图示内容：



- 进入自定义页面，获取搜索引擎的 ID。



- 随后下滑到页面底部，点击程序化访问模块中的“开始使用”按钮。



- 跳转到如下页面，点击获取密钥：



首页 > 产品 > Programmable Search Engine > 指南

该内容对您有帮助吗?

Custom Search JSON API：简介

发送反馈

本文档可帮助您熟悉 Custom Search JSON API 及其用法。

前期准备

创建可编程搜索引擎

通过调用 API，用户可针对可编程搜索引擎的现有实例发出请求。因此，在使用 API 之前，您需要在控制台中创建一个。按照教程中的说明详细了解不同的配置选项。创建之后，您可以在**概览**页的**基本**部分找到**搜索引擎 ID**。这是 API 使用的 `cx` 参数。

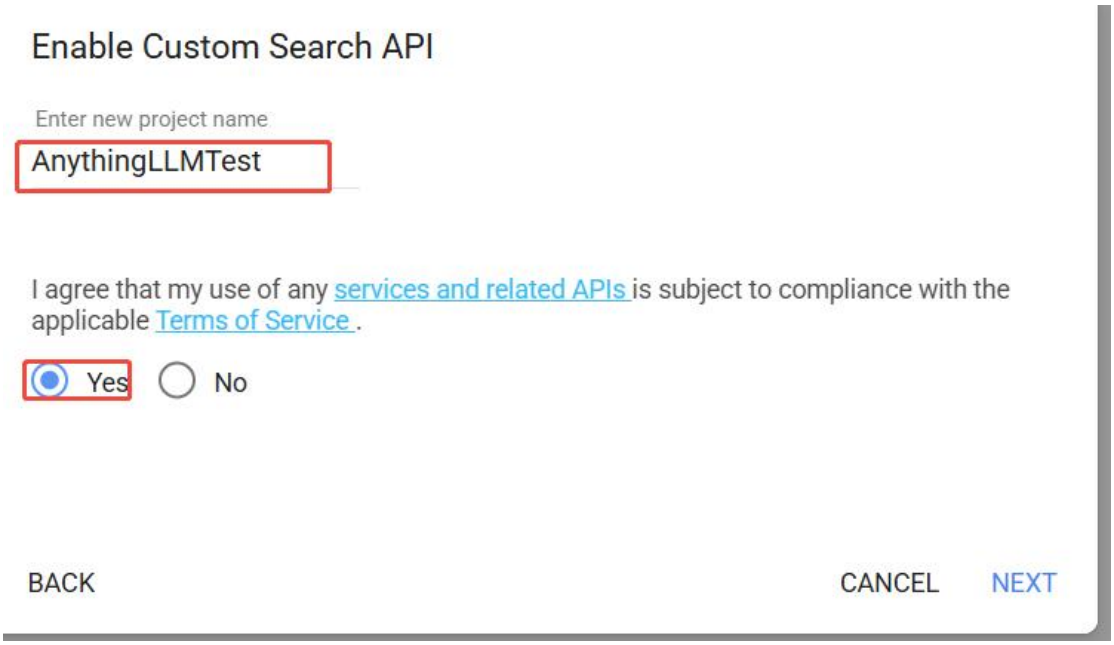
使用 API 密钥向 Google 标识您的应用

Custom Search JSON API 需要使用 API 密钥。API 密钥是向 Google 标识客户端的一种方式。

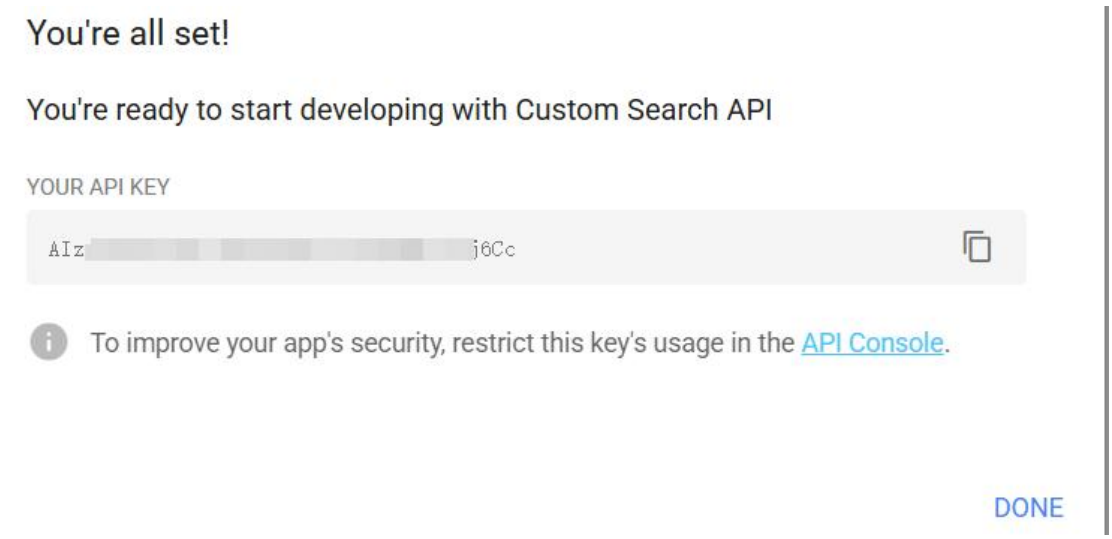
- 可编程搜索引擎（免费版本）用户：获取密钥

在您获得 API 密钥后，您的应用便可将查询参数 `key=yourAPIKey` 附加到所有请求网址中。API 密钥可以安全地嵌入到网址中，无需任何编码。

- 按照提示输入 api 名称，查看 api 使用协议并按照您的想法选择是否接受：

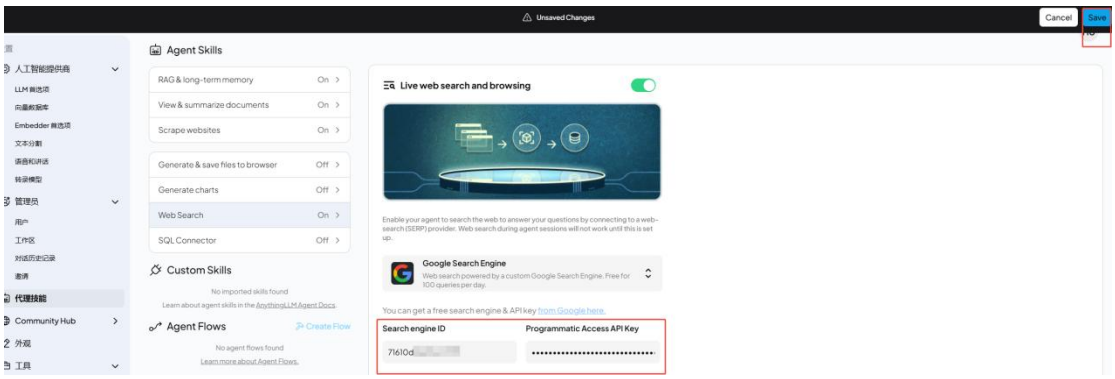


- 按照提示操作后得到 api 调用的 key。



c. 启用 AnythingLLM 的联网搜索功能

- 配置 Google Search Engine ：回到 AnythingLLM 联网搜索的页面，填入上述步骤获取的 ID 和 API key，点击右上角保存。



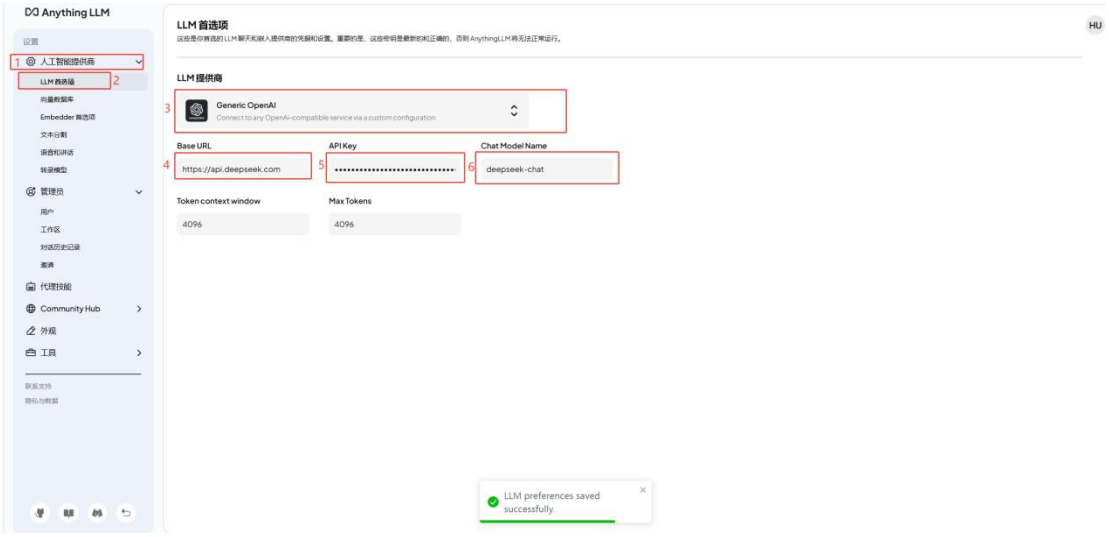
- 保存成功后得到如下提示：
- 在对话框以 @agent 启用联网功能：

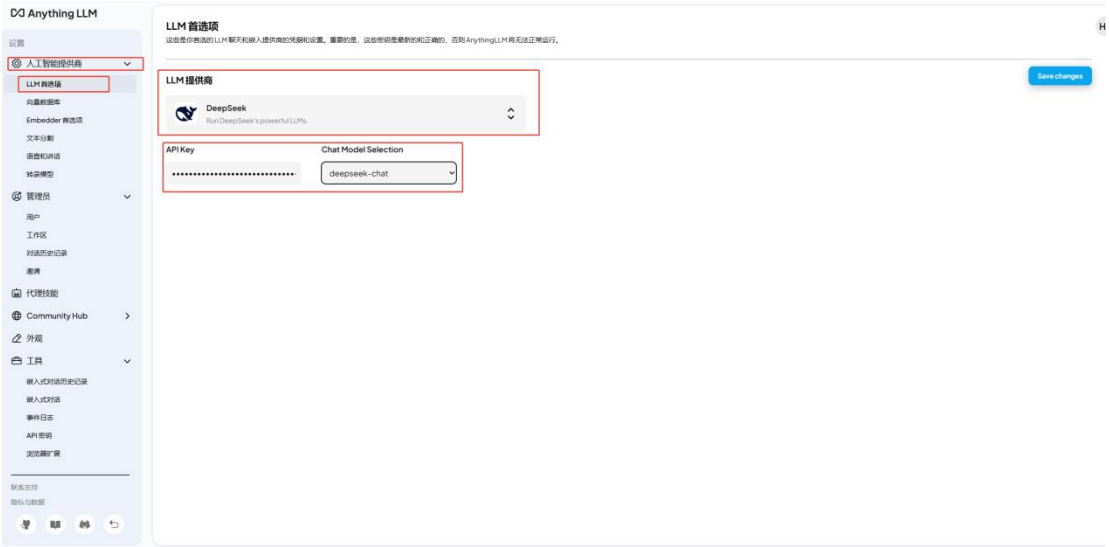


3. AnythingLLM 使用 DeepSeek API

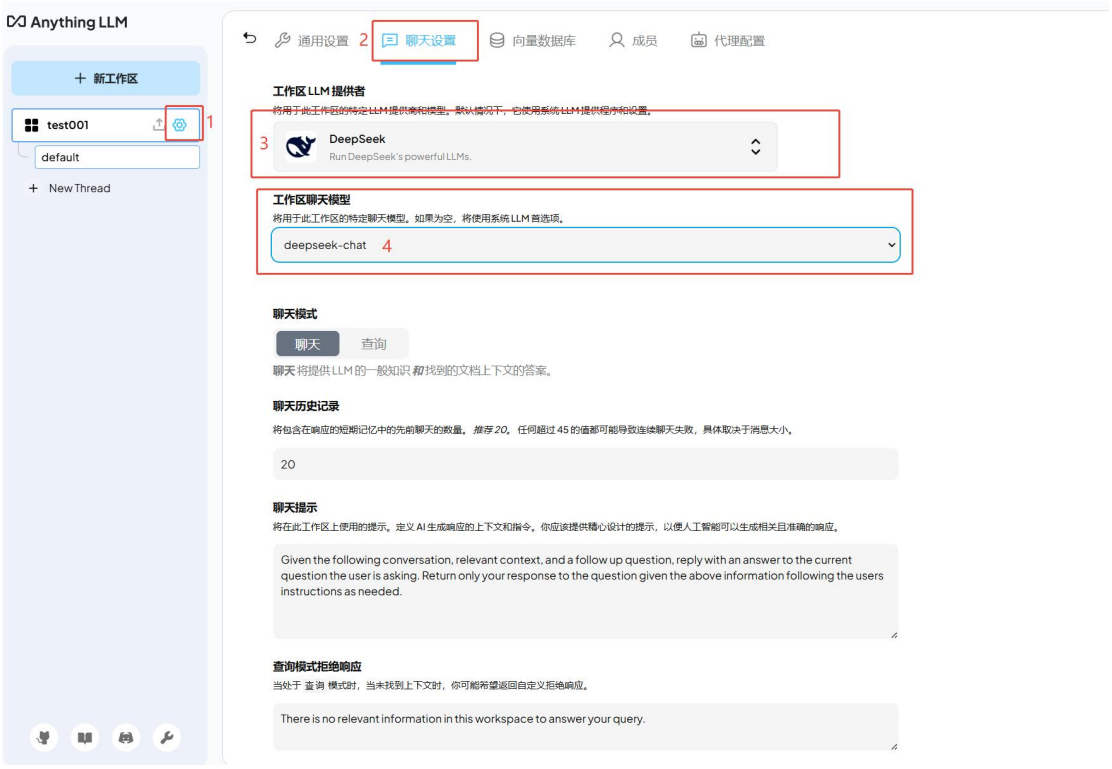
a. 前往 DeepSeek API 开放平台 注册一个账号，购买 API 使用资格。

b. 参考下述操作步骤，选择 Generic OpenAI 选项，填入 API 认证信息。或者选择 DeepSeek 选项，填入 API 认证信息。

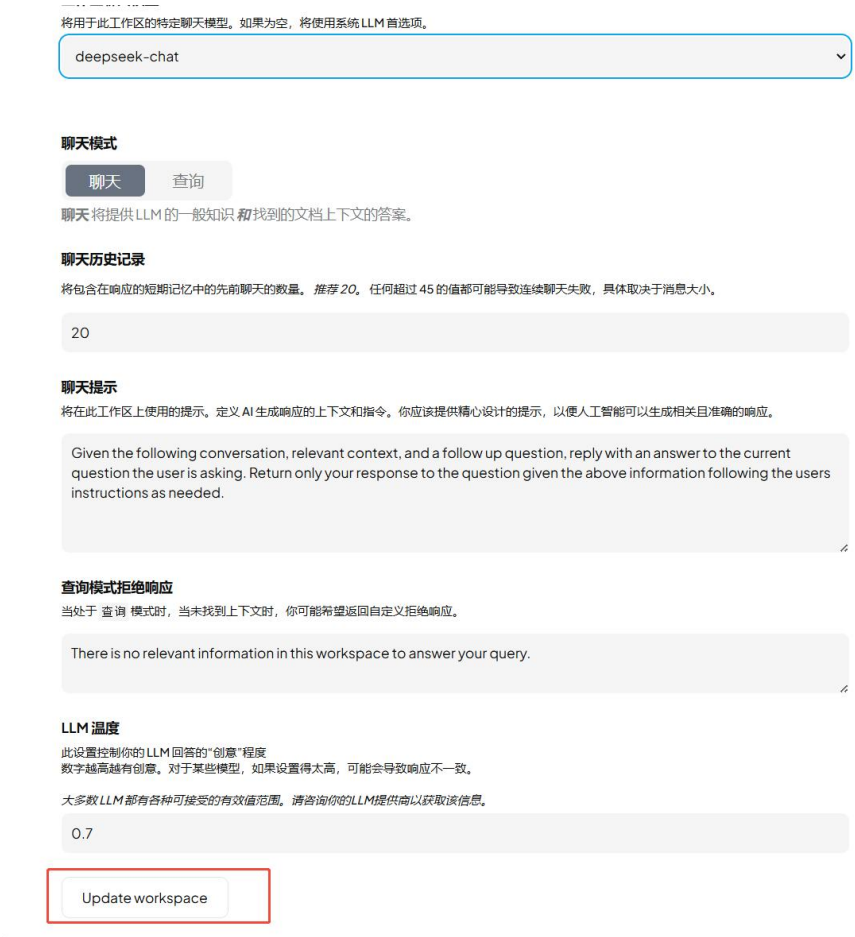


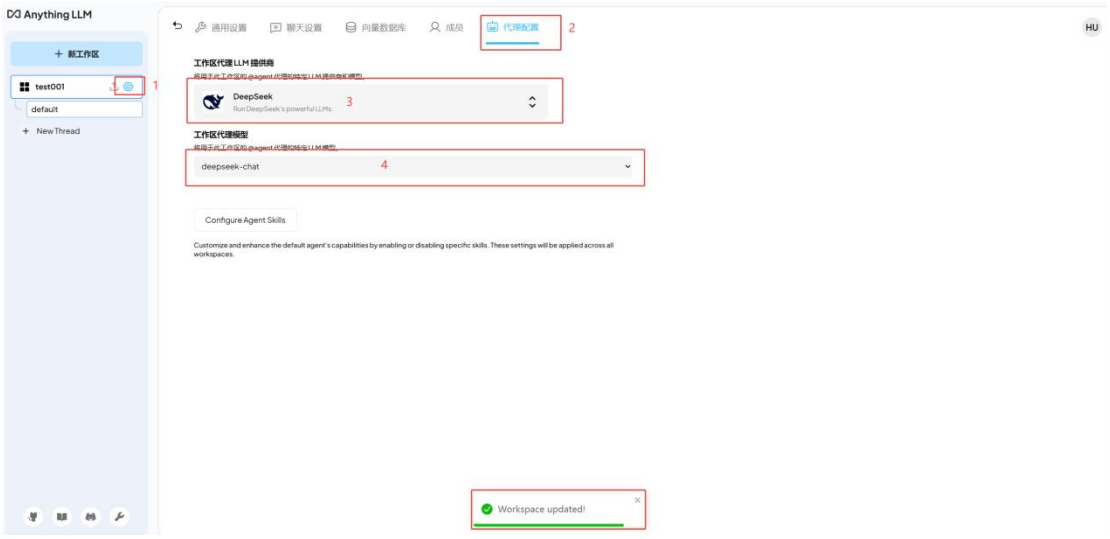


c. 到您想要使用的工作区，点击设置，更新当前聊天设置，以启用 DeepSeeK API 服务。

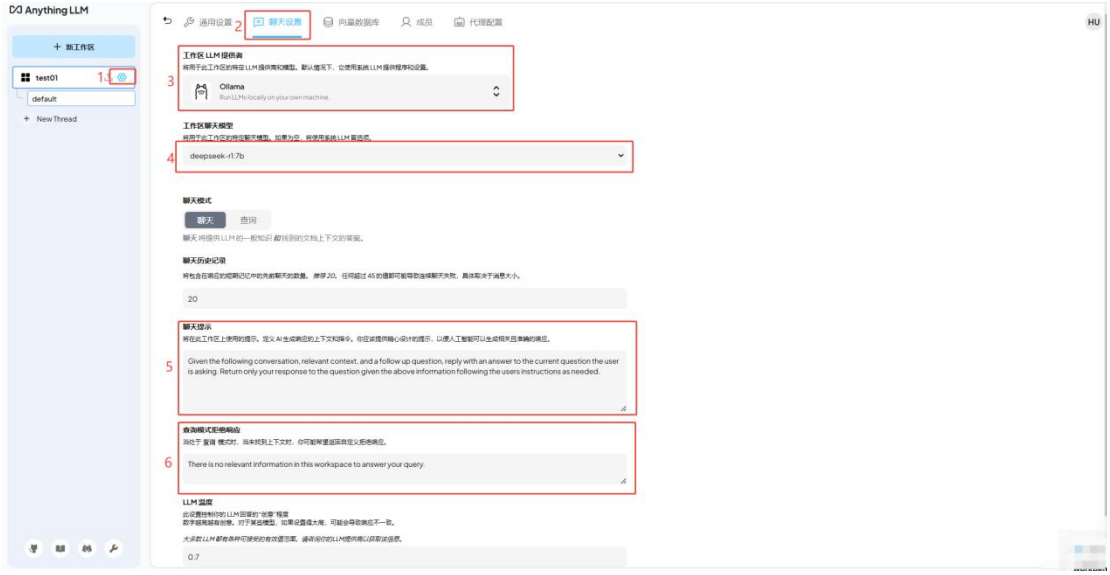


d. 需要点击下图按钮来保存配置。



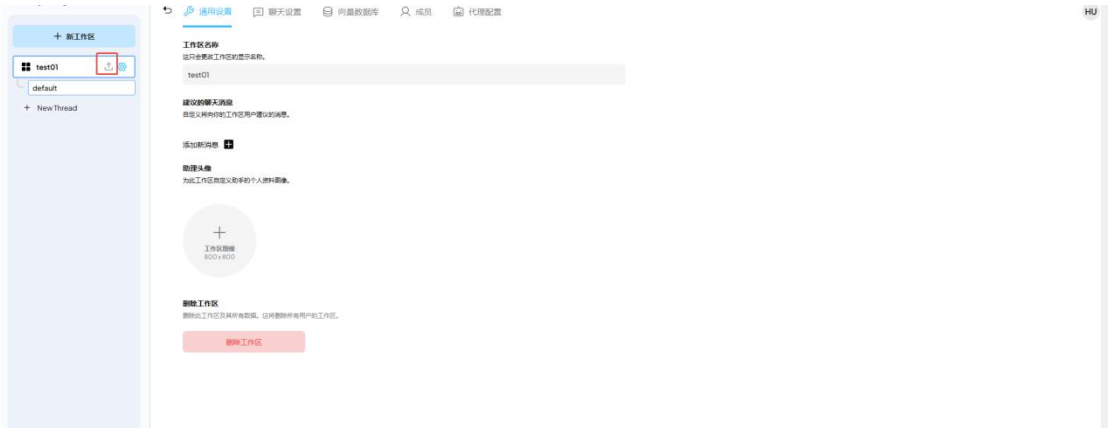


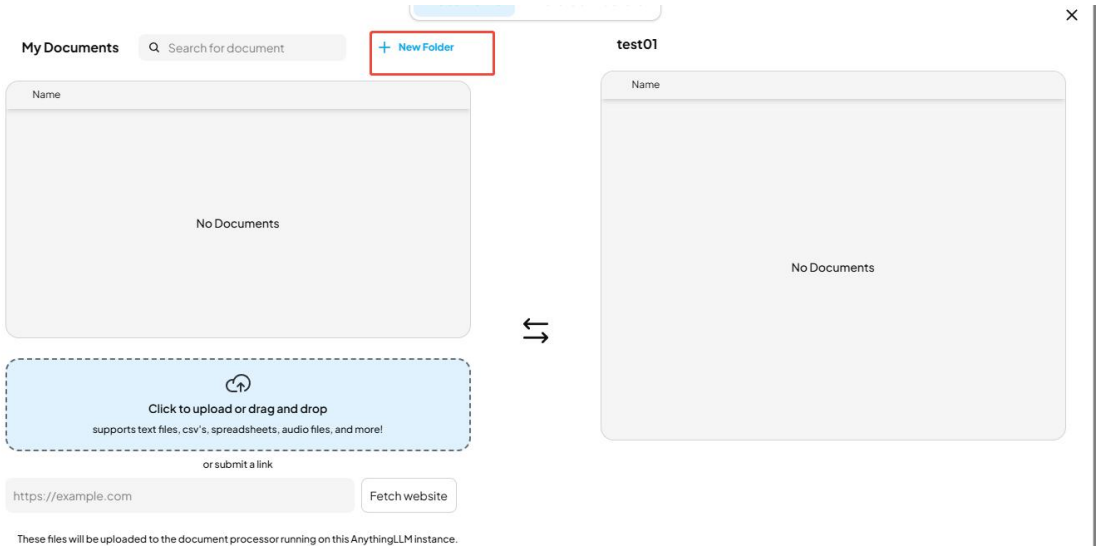
4. 调整对话模型及提示词

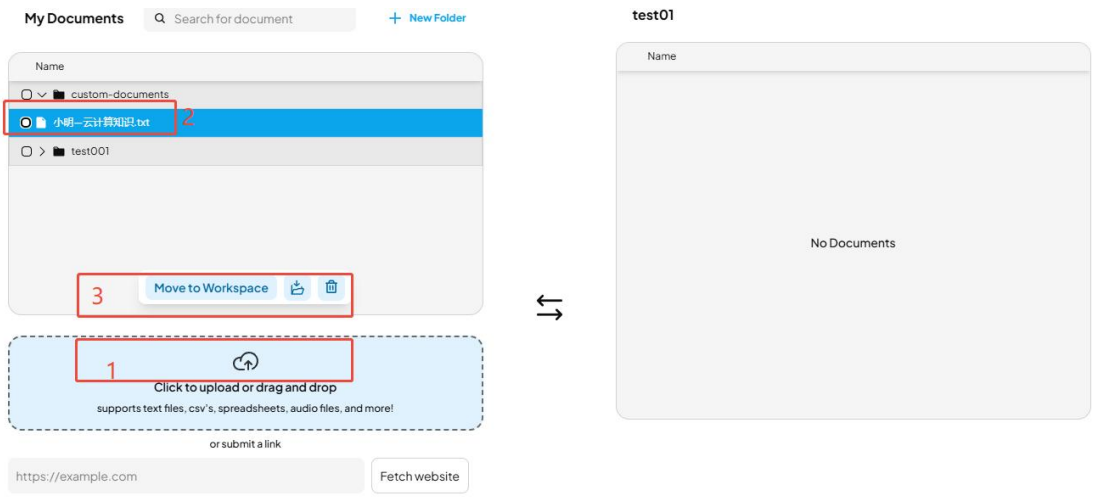


5. 知识库构建

a. 上传文件。







其中第 3 步需要第二部选择上传的文件后才会出现按钮，点击“Move to Workspace”，等待出现成功提示即可。



b. 基于知识库的对话。

首先选择上传了语料的工作区，随后直接提问即可。





10.38.9CherryStudio 协同 Ollama 启用 DeepSeek 大模型

应用场景

使用 CherryStudio 加载 DeepSeek 大模型的应用场景如下：

1. 学习研究 ：学生用其整理复习资料、辅助学术研究、助力论文写作；研究人员用于文献综述、分析问题等。
2. 工作办公 ：项目管理中生成报告、分析风险；内容创作时构思创意、撰写文案；代码开发里获取代码示例、解决问题。
3. 生活娱乐 ：作为智能聊天伙伴解闷陪伴，为创意活动如绘画等提供灵感启发。
4. 知识管理 ：个人可构建知识体系，团队能实现知识共享。

注意

建议使用天翼云提供的 DeepSeek 镜像创建云主机，减少安装过程中可能遇到的问题。镜像选择请参考在天翼云使用 Ollama 运行 DeepSeek 的最佳实践-7B 等版本-弹性云主机-最佳实践-AIGC 实践 - 天翼云。

操作步骤

一、 云主机端口配置更新

首先，您需要登录天翼云控制中心，对部署 DeepSeek 大模型的云主机添加新的端口开放操作。通过天翼云控制台进入部署了 DeepSeek 大模型的云主机详情页，选择安全组，添加新的开放端口，建议选择新加 114343 端口，该端口是 Ollama 服务默认监听的端口。添加完成后，可在安全组规则列表中查看新添加的 11434 端口规则，确认其状态为正常生效。



随后进入控制台，远程登陆即将安装 AnythingLLM 的云主机。



二、Cherry Studio 桌面软件安装

1、下载 Cherry Studio 桌面软件

访问 Cherry Studio 官网，选择合适的软件版本并下载，本案例中以 Cherry-Studio-0.9.27-setup.exe (88.1 MB) - Windows 标准安装包为例。



2、安装 Cherry Studio 软件

找到您下载的 Cherry Studio 安装包，按照提示进行安装。



三、Cherry Studio 实现自部署 DeepSeek 大模型对话

1、云主机 ollama 服务配置更新

对云主机上的 ollama 服务配置文件进行更新，添加第一部分中开放的端口。

如果您想同时在 OpenWebUI 以及其它服务上面启用，建议您使用 11434 端口，这样不用重启 OpenWebUI 服务。

```
# 编辑 ollama.service 服务参数，添加监听所有 ip 设置

vim /etc/systemd/system/ollama.service
```

添加内容如下：

```
Environment="OLLAMA_HOST=0.0.0.0"Environment="OLLAMA_ORIGINS=*"
```



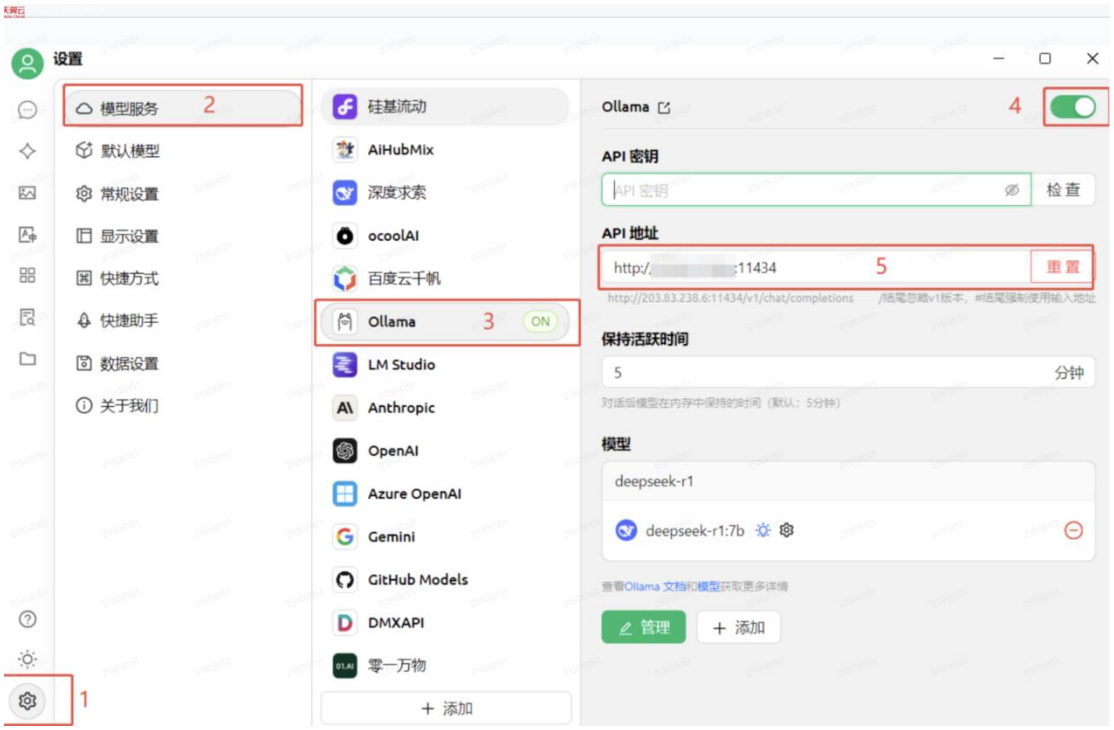
随后按照以下操作，重启 ollama 服务。

```
# 重新加载系统管理器的配置文件，使得 ollama.service 文件修改生效 sudo
systemctl daemon-reload# 重启 ollama 服务 sudo systemctl restart
ollama.service# 查看 ollama 服务监听端口 sudo netstat -tulnp | grep
11434
```

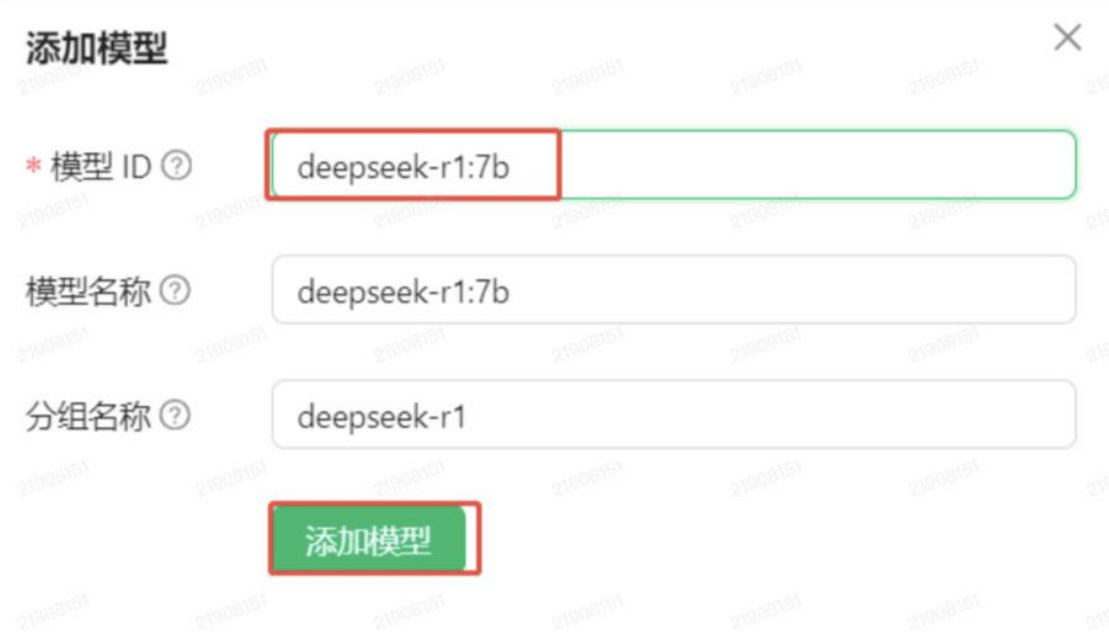
```
root@ecm-aigc-test:~# vim /etc/systemd/system/ollama.service
root@ecm-aigc-test:~# sudo systemctl daemon-reload
root@ecm-aigc-test:~# sudo systemctl restart ollama.service
root@ecm-aigc-test:~# sudo netstat -tulnp | grep 11434
tcp6       0      0 :::11434                :::*                    LISTEN      31184/ollama
root@ecm-aigc-test:~# vim /etc/systemd/system/ollama.service
```

2、Cherry Studio 中加载云主机上的 ollama 服务

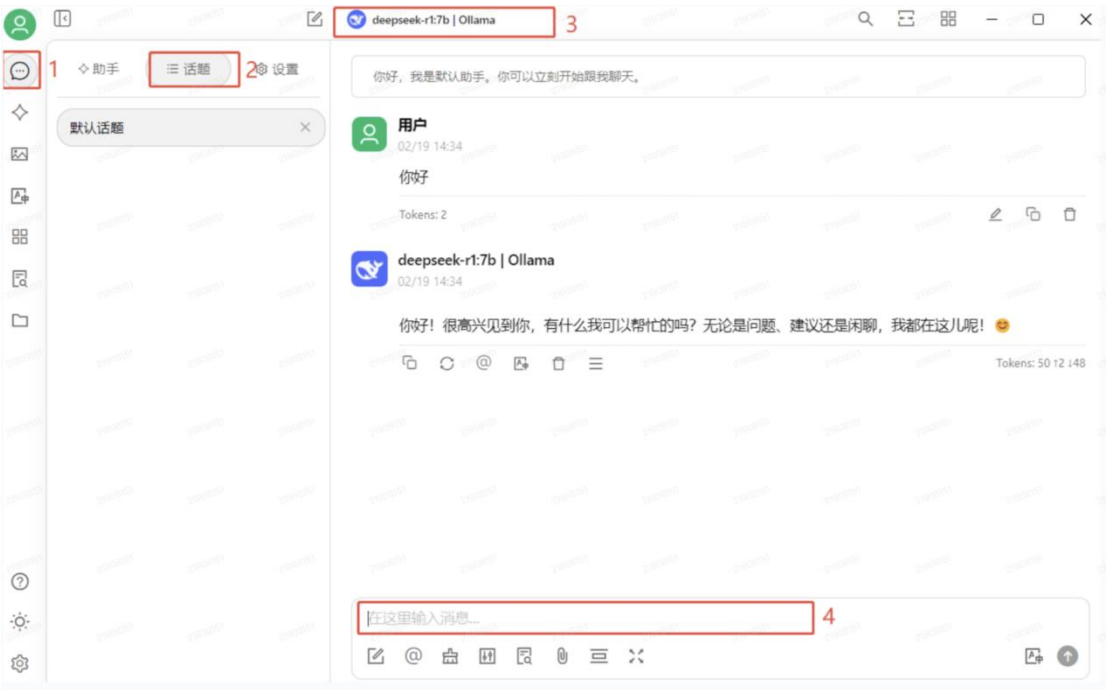
首先，打开安装好的 Cherry Studio 桌面程序，找到设置按钮，请参考下图步骤进行设置，第 5 步中的 ip 是您云主机公网的访问 ip。



随后点击下方的添加按钮，输入预装的 deepseek-r1:7b ，点击添加模型。



最后，按照下图提示进行对话吧。



FAQ

1、如何基于 Ollama 使用其它大模型？

请参考 ollama 模型站支持的大模型，在云主机上面下载您想要下载的大模型。

```
ollama pull $mode-name
```

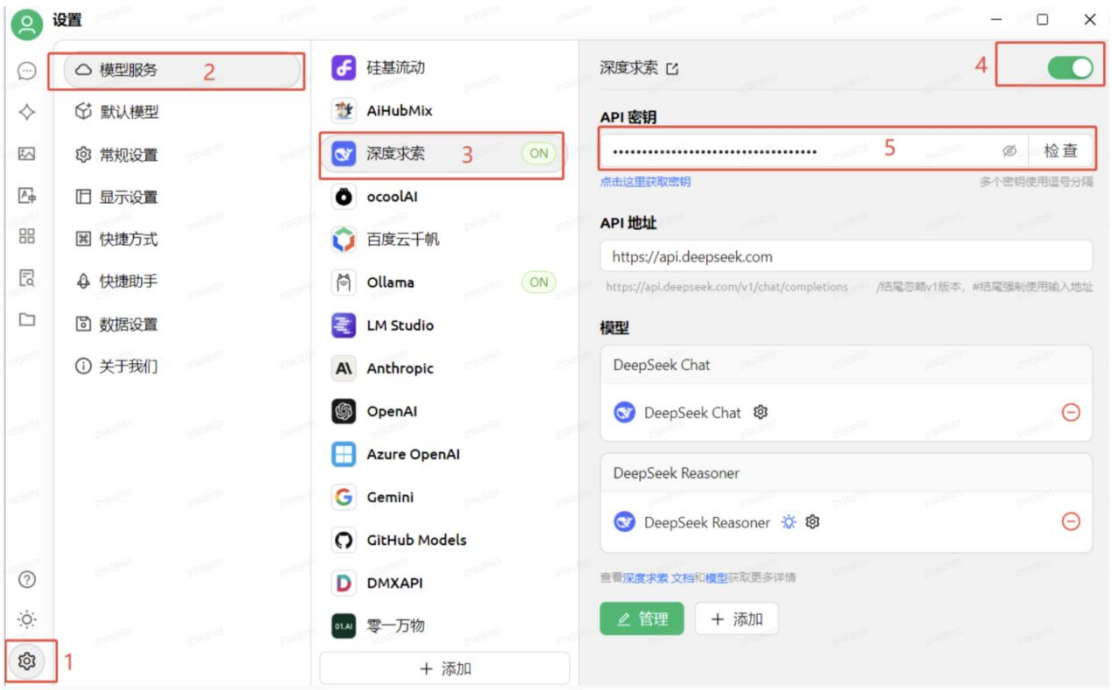
```
root@ecm-aigc-test:~# ollama pull deepseek-r1:1.5b
pulling manifest
pulling aabd4debf0c8...   0% |
```

随后在 Cherry Studio 上面选择启用对应的大模型，相关操作参考下图。

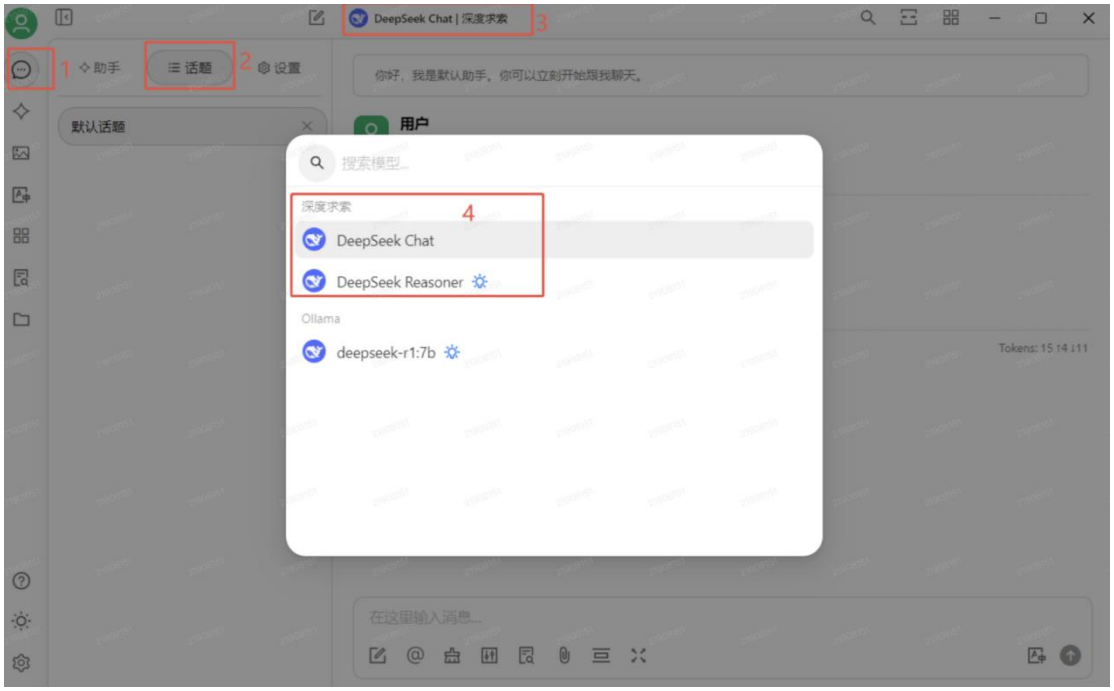


2、如何配置外部大模型？

参考下图步骤，填入您获取的其它大模型调用的 API 密钥，启用相应大模型服务。



随后，在对话框中选择对应的大模型，即可使用对应大模型服务。



3、如何启用联网搜索功能？

Cherry Studio 目前已经支持的联网模型服务商有：

- Google Gemini
- 腾讯混元
- 智谱 AI

注意

如果您想了解更多的 Cherry Studio 功能，请访问 Cherry Studio 文档。

10.38.10 天翼云 CPU 实例部署 DeepSeek-R1 模型最佳实践

1 背景

1.1 大模型算力需求

大模型因其参数规模庞大、结构复杂，通常需要强大的计算资源来支持其推理过程，这使得算力成为大模型应用的核心要素。随着 DeepSeek-R1 模型的问世，各行各业纷纷展开了关于如何接入大模型能力的广泛调研与探索，市场对大模型推理算力的需求呈现出爆发式增长的趋势。例如在医疗、金融、零售等领域，企业迫切希望通过接入 DeepSeek 大模型来提升决策效率和业务能力，从而推动行业的创新发展。在这一背景下，算力的供给和优化成为推动大模型落地应用的重要因素。

近年来,CPU 制程和架构的提升以及英特尔® 高级矩阵扩展 AMX(Advanced Matrix Extensions)加速器的面世带来了算力的快速提升。英特尔对大模型推理等多个 AI 领域持续深入研究，提供全方位的 AI 软件支持，兼容主流 AI 软件且提供多种软件方式提升 CPU 的 AI 性能。目前，已有充分的数据显示 CPU 完全可以用于大模型推理场景。

1.2 CPU 算力用于大模型推理的适用场景

CPU 适用于以下大模型推理场景：

场景 1：大模型推理需要的内存超过了单块 GPU 的显存容量，需要多块或更高配 GPU 卡，采用 CPU 方案，可以降低成本；

场景 2：应用请求量小，GPU 利用率低，采用 CPU 推理，资源划分的粒度更小，可有效降低起建成本；

场景 3：GPU 资源紧缺，CPU 更容易获取，且可以胜任大模型推理。

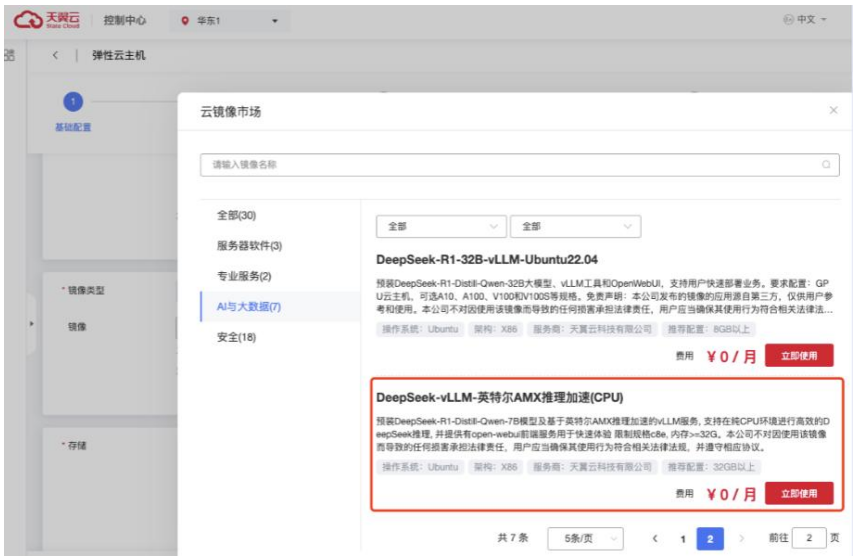
2 天翼云 EMR 实例 DeepSeek-R1-Distill-Qwen-7B 蒸馏模型部署实践

本节内容主要介绍如何在天翼云 EMR 实例上，基于 Intel® xFasterTransformer 加速库和 vllm 推理引擎完成模型部署，并展示相关性能指标。

2.1 服务部署

为了方便用户使用，天翼云联合英特尔制作了一键部署的云主机镜像，内置 DeepSeek-R1-Distill-Qwen-7B 模型、vLLM 推理框架、xFT 加速库以及 open-webui 前端可视环境。您可在天翼云控制台选择下列资源池和镜像，开通云主机进行体验。

类型	可用资源
可用资源池	华东-华东 1-az2、华东-华东 1-az3

类型	可用资源
可用规格	c8e 系列（最低内存需求 32G）
可用镜像	DeepSeek-vLLM-英特尔 AMX 推理加速(CPU) 

完成云主机开通后，推理服务会在 5 分钟内自动启动，您无需进行任何其他操作。

注意

如需在云主机外访问服务，您需要绑定弹性 IP，并在安全组内放行 22/3000/8000 端口。

2.2 模型使用

2.2.1 open-webui 前端使用

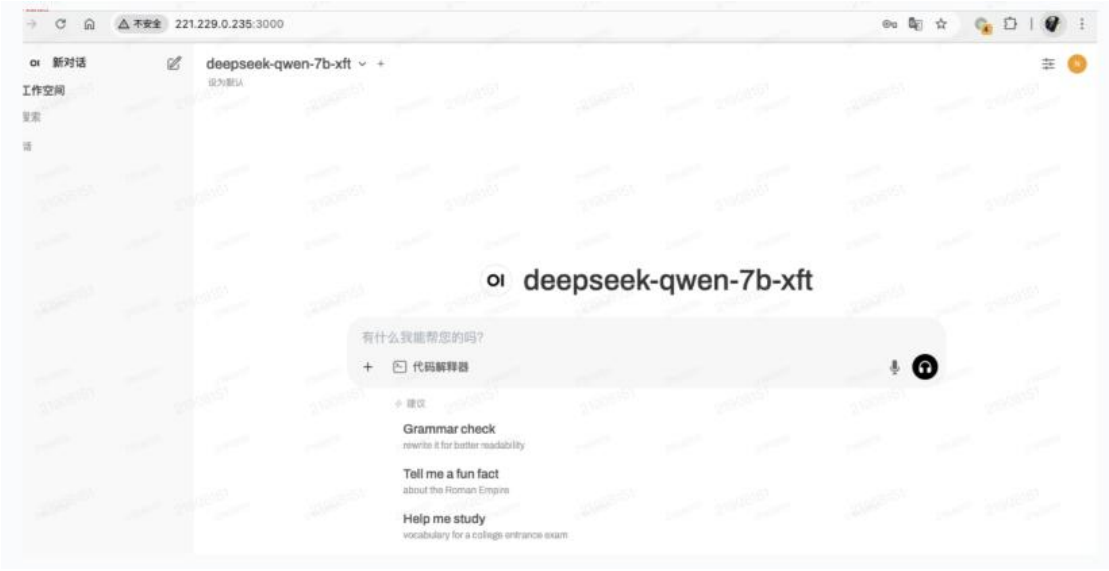
镜像已内置 open-webui，并已完成和 vllm 的连接配置，可直接通过以下地址进行访问:

```
http://[弹性 IP]:3000/
```

注意

1.首次打开页面时，您需要先完成管理员注册，以进行后续的用户管理。注册数据均保存在云实例的/root/volume/open-webui 目录下。

2.如果首次打开对话页面时没有模型可供选择,请您稍等几分钟让模型完成加载即可。



2.2.2 vllm api 调用

镜像内置 vllm 服务可直接通过如下地址访问：

```
# 根路径 http://[弹性 IP]:8000/

# 查询现有模型 http://[弹性 IP]:8000/v1/models

# 其他 api 接口参阅 vllm 文档
```

注意

vllm 服务配置有 API_KEY，您可在云实例的/root/recreate_container.sh 文件开头查看到当前值，并可进行修改以确保服务安全。

2.3 性能指标

借助于英特尔 AMX 的加速能力，本推理服务能够取得显著的性能提升，天翼云完成测试并给出参考指标如下：

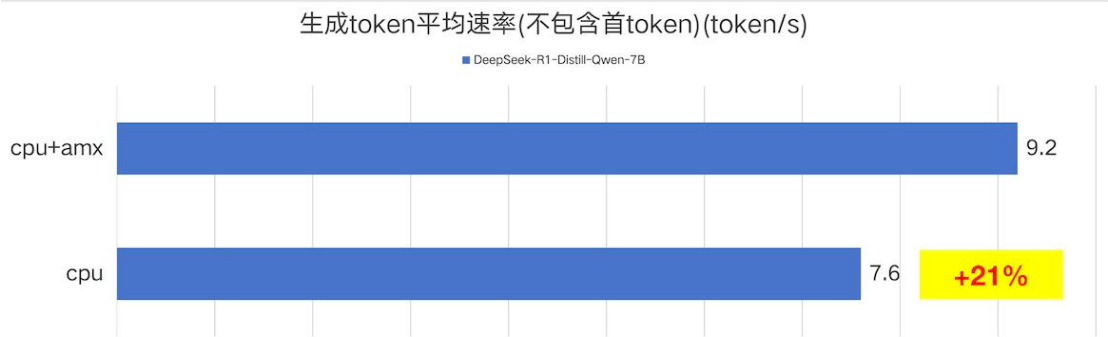
基本参数

- vcpu 数：24（物理核 12）

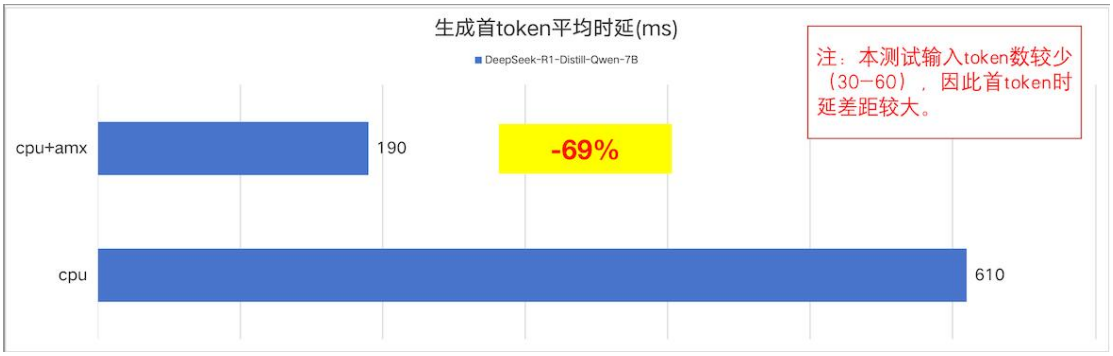
- 内存：64GB
- 硬盘：60G 通用型 SSD
- 模型：DeepSeek-R1-Distill-Qwen-7B（bf16）
- batch size：1
- 输入 token 个数：30-60
- 输出 token 个数：256

性能数据

平均 token 生成速率：



首 token 时延：



在天翼云 c8e 系列 24vcpu 云主机上，启用 AMX 加速能力后，DeepSeek 7B 蒸馏模型（BF16）推理速度能够超越 9token/s，满足日常使用需求。

3 基于英特尔®至强®6 处理器部署满血版 DeepSeek-R1 671B 实践

3.1 性能指标

DeepSeek R1 671B 满血版模型以其卓越的性能，为用户带来了极致的效果体验，不过其部署成本也不容小觑。若采用传统的 GPU 部署方式，需要 8－16 张 GPU 才能提供足够的支持，这无疑大幅增加了硬件购置、能耗以及维护等方面的成本。

在这样的背景下，天翼云基于英特尔®提供的至强®6 处理器服务器进行了 DeepSeek R1 671B 满血版 Q4_K_M 模型的部署尝试，测试结果如下:

- 1-instance 1-socket:

平均吞吐性能 9.7~10 token/s

- 2-instance 1-socket:

平均 7.32 token/s 和 7.38token/s， 共 14.7token/s

从上面测试数据可以看到，采用单实例单 socket 部署下，DeepSeek R1 671B 满血版模型可达到平均 9.7～10 token/s 的吞吐量，而在双实例部署模式中，总体吞吐量提升至 14.7 token/s。单颗 CPU 系统的吞吐性能可以达到普通用户正常使用的需要。

3.2 英特尔®至强®6 处理器简介

英特尔®至强®CPU 为 DeepSeek R1 671B 模型的部署提供了一个极具竞争力的方案。英特尔®至强®CPU 具备支持 T 级超大内存的能力，这使得它在权重存储与加载方面表现高效。对于像 DeepSeek R1 671B 这样的超大模型，其所需的显存容量在多卡 GPU 配置下才能满足，而英特尔®至强®CPU 能够凭借其强大的内存支持能力，为该模型提供良好的运行环境。

此外，DeepSeek R1 模型采用的 MOE（Mixture of Experts）结构，通过参数稀疏化的方式，使得在单 token 推理时仅需激活少量专家参数。这种特性显著降低了推理过程中的算力要求，与 CPU 的计算特点相契合，使得模型在 CPU 系统上的运行更加高效。这意味着在英特尔 ® 至强 ®CPU 上部署 DeepSeek R1 671B 模型，不仅能够充分发挥模型的性能优势，还能有效降低部署成本，避免了对大量 GPU 的依赖。

如需复现以上性能测试结果，请参看附录 2

总结

通过本次实践，无论是在天翼云 EMR 云实例上结合 xFasterTransformer 部署 DS R1 distill Qwen-7B 蒸馏模型，还是基于英特尔® 至强® 6 处理器部署满血版 DeepSeek-R1 671B 模型，均验证了 CPU 系统在 DeepSeek 大模型推理上的可行性和符合业界普遍要求的性能表现。CPU 系统不仅能够灵活应对不同规模的模型需求，无论是轻量化蒸馏模型还是全功能满血模型，都能高效满足用户场景需求，

提供了一种低成本、经济高效的解决方案。

附录 1 英特尔® 至强® 可扩展处理器与 AI 加速技术

最新英特尔® 至强® 可扩展处理器产品

英特尔第五代® 至强® 可扩展处理器（代号 Emerald Rapids）——为 AI 加速而生

第五代英特尔® 至强® 处理器以专为 AI 工作负载量身定制的设计理念，实现了核心架构和内存系统的双重飞跃。其 64 核心设计搭配高达 320MB 的三级缓存（每核心由 1.875MB 提升至 5MB），相较上代缓存容量实现近三倍增长，为大规模并行 AI 推理提供充裕的本地数据存储空间。与此同时，处理器支持 DDR5-5600 高速内存，单路最大 4TB 的容量保证了大数据处理时的带宽和延迟优势。基于这些硬件提升，Emerald Rapids 整体性能较上一代提升 21%，AI 推理性能平均提升 42%，在大语言模型推理场景中可实现最高 1.5 倍的性能加速，同时大幅降低总拥有成本达 77%。

英特尔® 至强®6 处理器（代号 GNR Granite Rapids）——引领 CPU AI 算力革新

全新 GNR 处理器专为应对人工智能、数据分析及科学计算等计算密集型任务而设计。该产品在内核数量、内存带宽及专用 AI 加速器方面均实现重大突破：

- 核心与性能：每 CPU 配备多达 128 个性能核心，单路核心数较上一代翻倍，同时平均单核性能提升达 1.2 倍、每瓦性能提升 1.6 倍，进一步强化了 CPU 在大模型推理中的独立处理能力；
- AI 加速功能：内置英特尔® 高级矩阵扩展（AMX）新增对 FP16 数据类型的支持，使得生成式 AI 和传统深度学习推理任务均能获得显著加速；
- 内存与 I/O 突破：支持 DDR5-6400 内存及英特尔首款引入的 Multiplexed Rank DIMM (MRDIMM) 技术，有效将内存带宽提升至上一代的 2.3 倍；同时，高达 504MB 的三级缓存和低延迟设计确保数据能够更快加载，为复杂模型训练和推理缩短响应时间。

英特尔® 至强® 6 处理器不仅通过更多的核心和更高的单线程性能提升了 AI 大模型推理能力，同时也能够作为机头 CPU 为 GPU 和其他加速器提供高速数据供给，进一步缩短整体模型训练时间。在满足混合工作负载需求的同时，其 TCO 平均降

低 30%，大模型推理加速最高可达 2.4 倍。

无论是第五代至强还是全新的至强 6 处理器，英特尔均通过在核心架构、缓存系统、内存技术和专用 AI 加速器方面的全面革新，提供了业界领先的 AI 计算支持。这两款产品为数据中心和高性能计算平台在 AI 推理、训练以及多样化工作负载下提供了强大而高效的算力保障

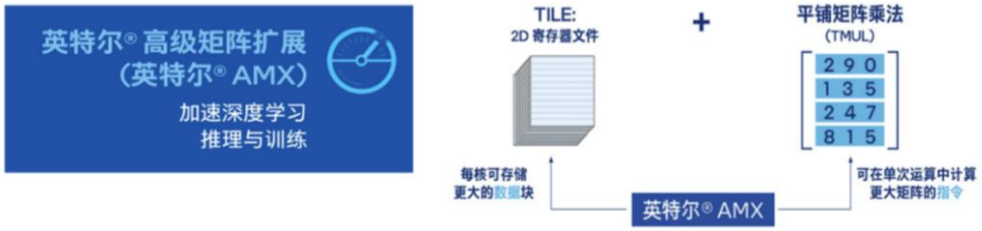


图 1 英特尔高级矩阵扩展（AMX）

英特尔全方位的 AI 软件生态支持

英特尔及其合作伙伴凭借多年 AI 积累，围绕至强®可扩展处理器打造了完善的软件生态：广泛支持主流开源框架，通过插件优化及多样化开源工具链，使用户在 x86 平台上能够轻松开发、部署通用 AI 应用，无需手动调整，同时确保从终端到云的全程安全保护。

此外，至强®处理器内置多种 AI 加速指令（如 AVX-512、AMX），使得任何兼容软件均可直接调用加速功能。开发者可免费下载英特尔分发版工具、库及开发环境，充分利用这些内置加速器应对各类 AI 管线需求。结合多样化硬件优势与开放生态，英特尔通过经济、可扩展的方案，将 AI 能力无缝延伸至云端与边缘。

其中，xFasterTransformer（xFT）是英特尔官方开源的 AI 推理框架，专为大语言模型在至强®平台上深度优化。xFT 不仅支持多种数据精度（FP16、BF16、INT8、INT4），还能利用多 CPU 节点实现分布式部署，显著提升推理性能并降低成本。其简单的安装和与主流 Serving 框架（如 vLLM、FastChat、MLServer、MindSpore Transformer、PaddlePaddle）的兼容性，帮助用户快速加速大模型应用。在 3.1 节中基于天翼云 EMR 云主机和 xFasterTransformer 加速引擎实现了对与 DeepSeek R1 蒸馏模型的高效推理部署。



图 2 英特尔提供 AI 软件工具 全面兼容主流 AI 开发框架

附录 2 CPU 环境下部署 DeepSeek-R1 671B 模型实践

环境配置

硬件配置

- CPU：Intel® Xeon® 6980P Processor, 128core 2.00 GHz
- 内存 24*64GB DDR5-6400
- 存储 1TB NVMe SSD

软件环境

- OS: Ubuntu 22.04.5 LTS
- Kernel: 5.15.0-131-generic
- llama.cpp: github bd6e55b
- cmake: 3.22.0
- gcc/g++: 11.4.0
- Python: 3.12.8
- git: 2.34.1

BIOS里关闭 sub NUMA 相关配置。

注意

版本是指本测试中服务器上安装的版本，并非要求的最低版本。

部署步骤

1. 安装 llama.cpp

下载并安装 intel oneapi 库

```
wget
https://registrationcenter-download.intel.com/akdlm/IRC_NAS/dfc4a434
-838c-4450-a6fe-2fa903b75aa7/intel-oneapi-base-toolkit-2025.0.1.46_o
ffline.shsh ./intel-oneapi-base-toolkit-2025.0.1.46_offline.sh -a
--silent --cli --eula acceptsource /opt/intel/oneapi/setvars.sh
```

下载 llama.cpp 推理框架源码

```
git clone https://github.com/ggml-org/llama.cpp
```

基于 oneapi 库对 llama.cpp 进行编译

```
cd llama.cpp

cmake -B build -DGGML_BLAS=ON -DGGML_BLAS_VENDOR=Intel10_64lp
-DCMAKE_C_COMPILER=icx -DCMAKE_CXX_COMPILER=icpx -DGGML_NATIVE=ON

cmake --build build --config Release -j$nproc
```

2. 下载模型文件

我们直接使用了社区制作的 DeepSeek 671B 满血模型的 Q4 量化版，您也可以下载 DeepSeek 官方 BF16 版本，并通过 llama.cpp 提供的脚本转换为 GGUF 格式。社区提供了从 1bit 到 8bit 不同版本的量化选项，具体区别可以参考社区网页。我们选择了使用最受欢迎的 Q4_K_M 版本。如果追求最佳效果，建议使用 Q8_0 版本。

Filename	Quant type	File Size	Split	Description
DeepSeek-R1-Q8_0.gguf	Q8_0	713.29GB	true	Extremely high quality, generally unneeded but max available quant.

DeepSeek-R1-Q4_K_M.gguf	Q4_K_M	404.43GB	true	Good quality, default size for most use cases, *recommended*.

```
# 下载 unsloth 制作的社区版量化模型(注意这里要下载 9 个文件)
mkdir DeepSeek-R1-Q4_K_M
cd DeepSeek-R1-Q4_K_M# 建议 nohup 执行，预计至少需要半天时间，同时确保磁盘容量足够 400G.
wget https://modelscope.cn/models/unsloth/DeepSeek-R1-GGUF/resolve/master/DeepSeek-R1-Q4_K_M/DeepSeek-R1-Q4_K_M-0000{1..9}-of-00009.gguf
```

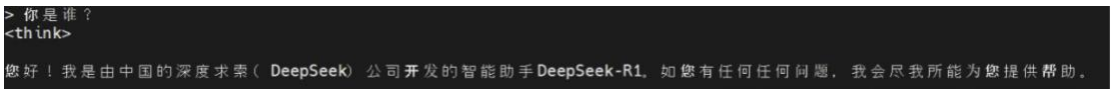
Q4_K_M 版本的文件大小为 404.43GB，下载过程可能会比较耗时。下载完成后，您可以在 DeepSeek-R1-Q4_K_M 目录下找到一系列 .gguf 文件，例如 DeepSeek-R1-Q4_K_M-00001-of-00009.gguf。

3. 模型加载和运行

```
#启动模型的交互模式(替换路径地址为实际值)[llama.cpp 路径]/build/bin/llama-cli -m [模型下载路径]/DeepSeek-R1-Q4_K_M-00001-of-00009.gguf -i
```

下面用几个示例展现 DeepSeek-R1 671B 满血版强大的的 reasoning 推理能力:

测试模型自我认知:



验证推理能力的经典“草莓”问题:

```
> How many Rs are there in strawberry?
<think>

To determine how many times the letter "r" appears in the word "strawberry":

1. **Write out the word**: S T R A W B E R R Y
2. **Identify each "r"**:
   - The third letter is **R**.
   - The eighth letter is **R**.
   - The ninth letter is **R** (since "berry" is spelled B-E-R-R-Y).

**Conclusion**: There are **3** instances of the letter "r" in "strawberry".
```

“等灯等灯”的意思：



4.性能及优化

#性能测试(替换路径地址为实际值，并根据CPU实际核数和numa情况指定-C和-m参数，千万不要照抄!)

export MODEL_PATH=[模型下载路径]/DeepSeek-R1-Q4_K_M-00001-of-00009.gguf

numactl -C 0-127 -m 0 [llama.cpp路径]/build/bin/llama-cli -m $MODEL_PATH -t 128 --temp 0.6 -s 42 -no-cnv --no-kv-offload -ctk q4_0 -c 8192 -n 128 -p "<|User|>以孤独的夜行者为题写一篇750字的散文，描绘一个人在城市中夜晚漫无目的行走的心情与所见所感，以及夜的寂静给予的独特感悟。<|Assistant|>"

这里使用numactl 来指定使用单路 CPU (0-127, 6980P 有 128 核)，以及这一路 CPU 的内存节点（numa0），避免跨 numa 传输以获取最佳性能。

llama.cpp 是本地编译的,编译的时候使用Intel oneAPI 可以有效提升它的性能。英特尔尝试用了 oneAPI里的 Intel C++编译器和数学加速库 MKL，结合 jemalloc 内存管理优化，**推理速度可以达到每秒 9.7~10 词元**(TPS, tokens per second)。

上面的实验是在单路 CPU 上进行的,**我们又在两路 CPU 上各自独立启动 1 个模型实例，总速度可以达到 14.7TPS** (7.32TPS+7.38TPS)。

再进一步，英特尔观察到基于现有的 llama.cpp 软件方案，在 CPU 平台没有实现高效的专家并行和张量并行等优化，CPU 核心利用率和带宽资源没有充分发挥出来，6980P 的 128 核心运行1 个模型还有不少性能储备。预计可以继续增加实例数来获得更好的总 TPS。

另外，通常情况下，CPU 的 SNC (Sub-NUMA Clustering)设置可以获得更高的带宽，但是受限于软件并未优化实现良好匹配，此次实验关闭了 SNC 测试。

以下方式的系统配置也有助于提升性能：

- BIOS里关闭 AMP prefetcher
- 用cpupower 打开 CPU 的 pstate 性能模式
- 提高CPU 的 uncore 频率
- 关闭超线程(Hyper-Threading)

注意

为了加快试验进度，我们限制了词元输出长度(-n 128)。经过验证，增加输出长度(例如-n 512)对于生成性能指标的影响不大。

10.38.110pen WebUI 快速入门

注意

GPU 云主机镜像自带的 ollama 监听 127.0.0.1:11434、vllm 监听 0.0.0.0:8000、open-webui 监听 0.0.0.0:3000 端口，云主机默认不对外开放任何端口访问，请按需开放端口访问规则，避免数据泄露。

开放云主机端口

云主机安全组需要放行 3000 端口，添加安全组规则帮助文档。

注册管理员账号

Deepseek 云主机内置可视化界面，云主机启动 5 分钟后，可访问 http://{your_ip}:3000 (your_ip 这台云主机的 eip) 体验 DeepSeek。

首次登录页面如下：

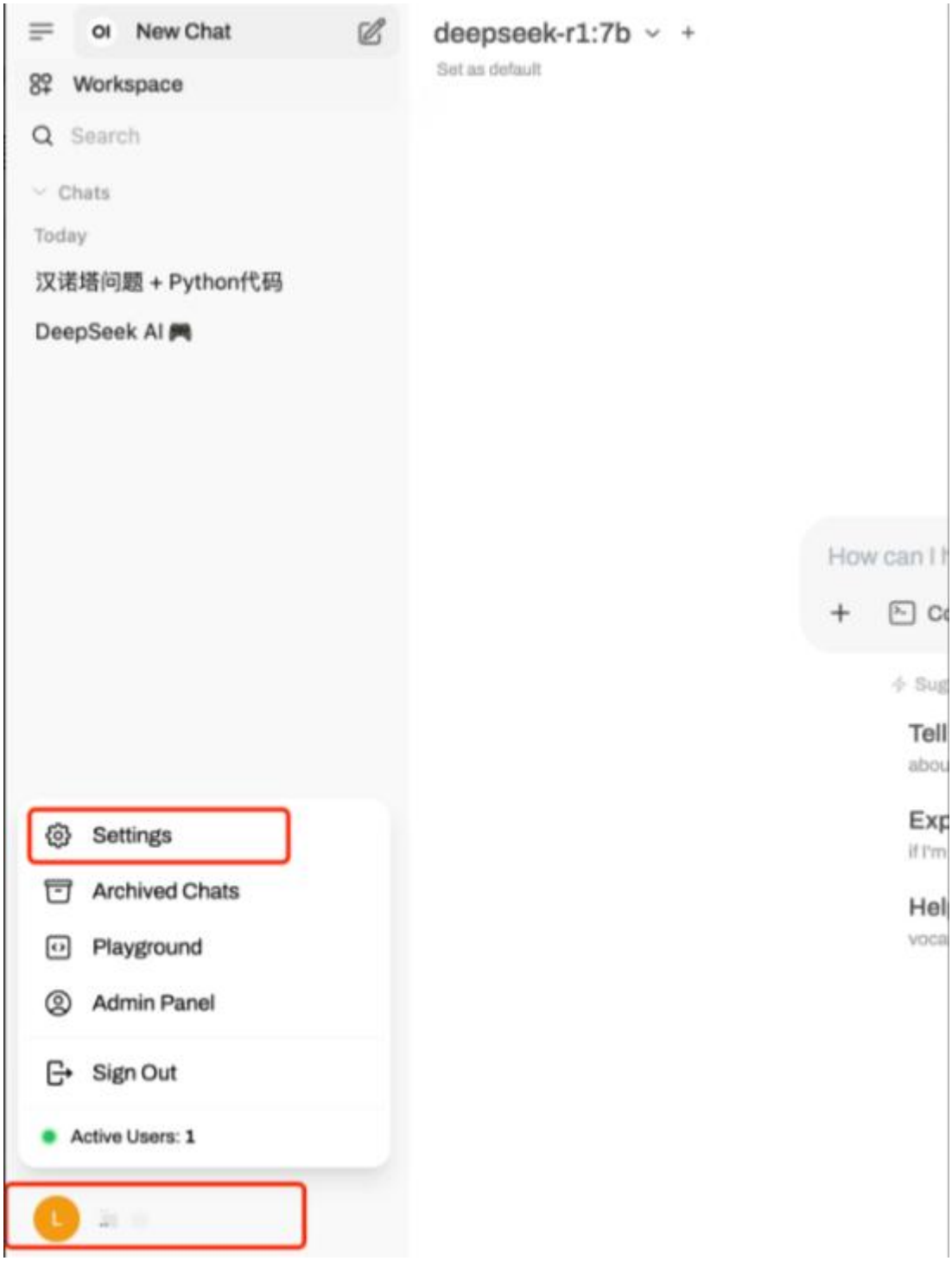


点击开始使用，注册管理员账号：

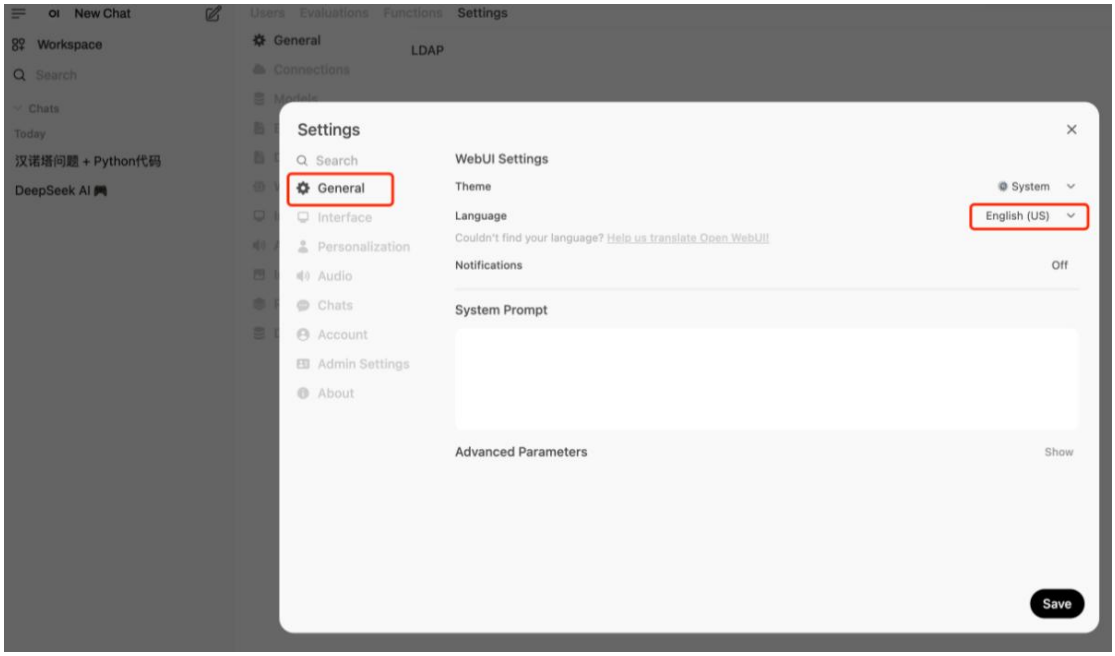


基本配置

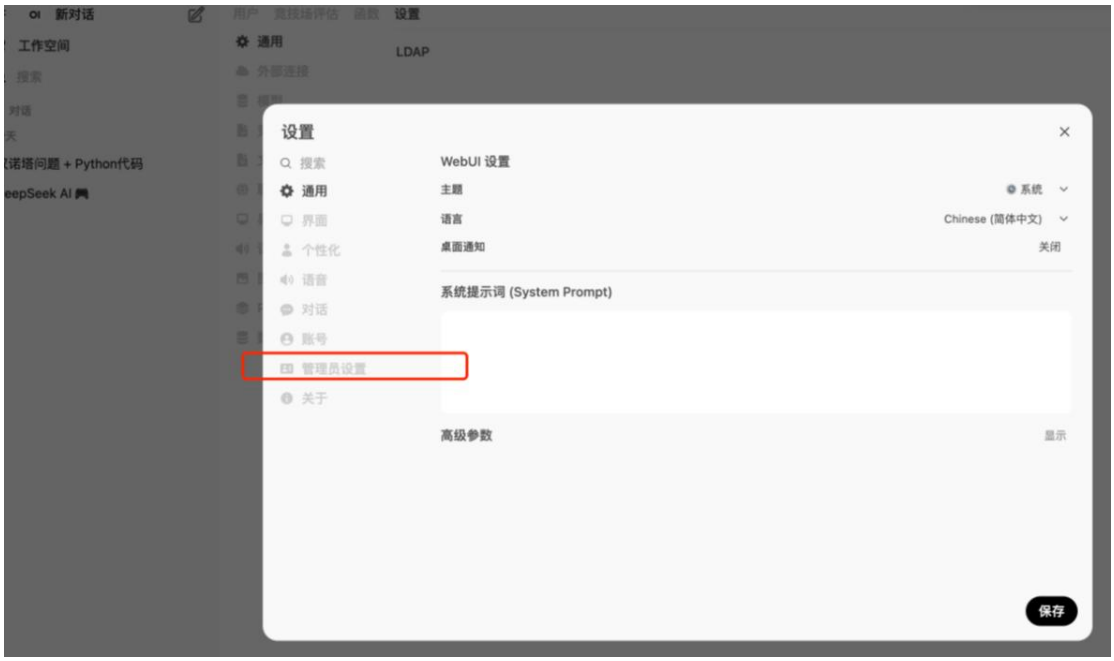
Open WebUI 有强大的设置功能，可根据需要进行设置，点击左下角进入设置页面：



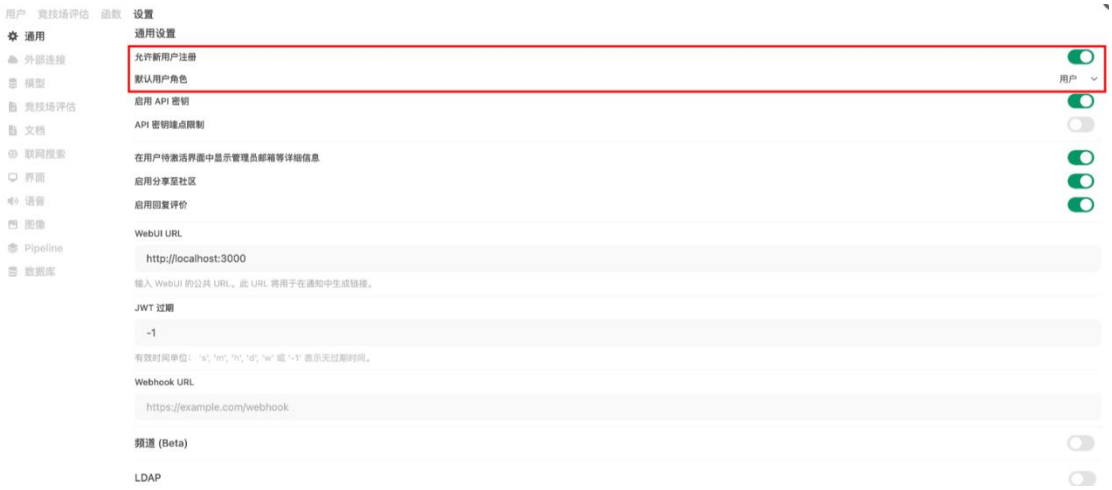
首先在通用设置中，将语言设置为简体中文。



进入管理员设置。



如果您不想开放其他用户注册使用，则需要关闭 “允许用户注册” 功能。

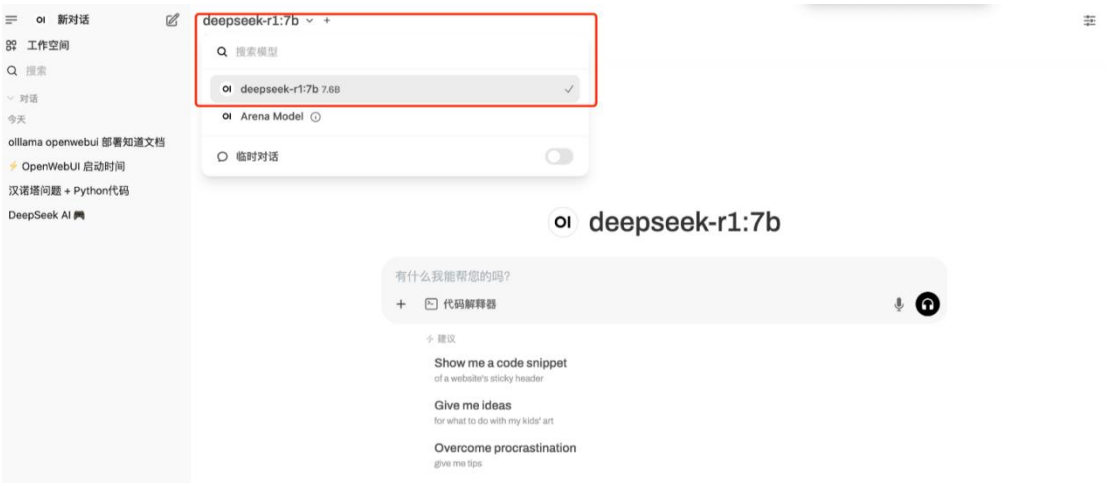


如果你允许用户注册，还可以设置用户注册后的行为，例如将新用户注册后的默认用户角色设为“用户”或“待激活”等，这些用户需要管理员手动激活：



模型配置

单机版 Deepseek GPU 云主机在完成部署后，会自动匹配云主机本地模型，刷新进入首页，可以看到，模型下拉列表中，可以选择我们部署的 deepseek-r1 模型：



您可以主动设置您的模型连接，以体验任意第三方连接。以 ollama 和 vllm 提供的服务为例，进入“管理员面板” -> “设置” -> “外部链接”，如果使用的 vllm 则填入您的 vllm 地址和 --api-key，如果使用 ollama 则填入您的 ollama 地址，点击保存后即可使用模型。

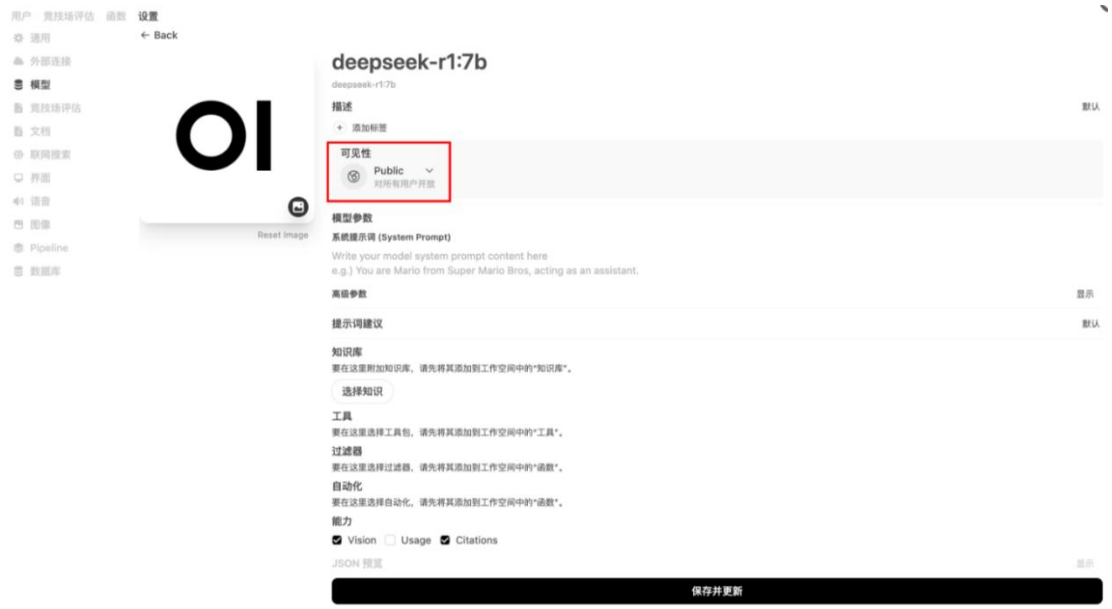


设置模型可见性：

多用户模式下，建议把模型设置为 public，如此才能让模型被所有用户看到。进入“管理员面板” -> “设置” -> “模型”，并点击编辑。



将模型可见性设置为 public 并点击保存。



最后，我们尝试下问一个问题，模型可以正常回答问题：

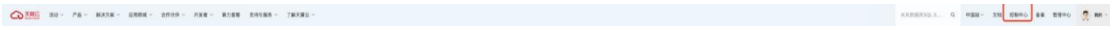


10.38.12 天翼云 GPU 云主机构建高性能 Deepseek 集群最佳实践-32B 等版本

步骤一：创建 GPU 云主机

1. 进入创建云主机页面

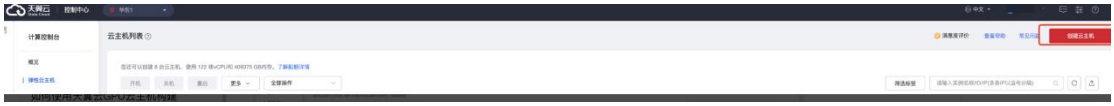
a. 点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



b. 单击“服务列表>弹性云主机”，进入主机列表页。



c.单击“创建云主机”，进入弹性云主机创建页。



2. 进行基础配置

a.基础配置

根据业务需求配置“计费模式”、“地域”、“企业项目”、“虚拟私有云”、“实例名称”、“主机名称”等。

b.选择规格。

此处选择"CPU 架构"为"X86"、"分类"为"GPU 加速/AI 加速型"、"规格族"为"GPU 计算加速型 pn8r"、"规格"为"pn8r.16xlarge.4"。

c.选择镜像

“镜像类型”选择“镜像市场”，在云镜像市场中选择预置了 DeepSeek R1 模型的 DeepSeek-R1-32B-vLLM-Ubuntu22.04 镜像。



d. 设置云硬盘。



3. 网络及高级配置

设置网络，包括"网卡"、"安全组"，这里，为安全和成本考虑，我们先不设置弹性 IP，后续按需开放 IP；设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。

4. 确认并支付

步骤二：配置弹性负载均衡

现在，我们已经有了两台 Deepseek-R1-32B 服务器，接下来，我们利用弹性负载均衡构建一个可任意横向扩容的 Deepseek 集群。

1. 配置 vllm 服务

首先需要为所有服务器上的 vllm 服务配置相同的 --api-key。

服务的配置文件目录在 /var/vllm_service.env 我们将 API_KEY 设置为想要的值。

```
API_KEY=xxxh4cVRFPgfAr8ayyy
NV_NUM=4
NV_NUM=4
```

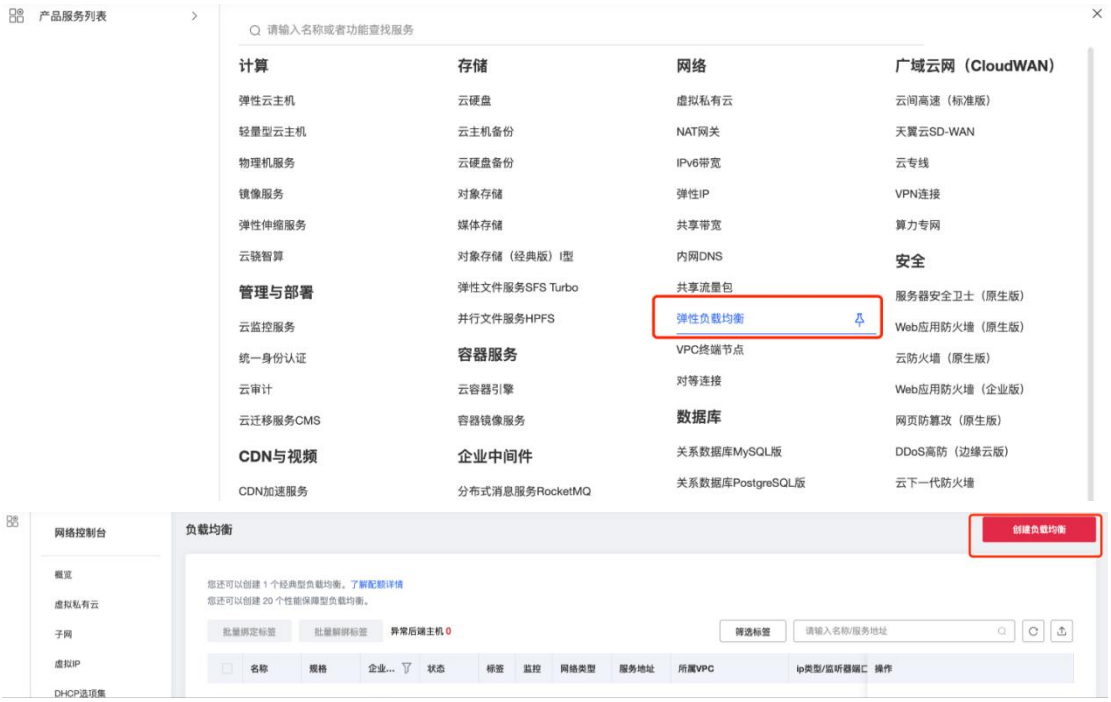
然后重启服务。

```
systemctl restart vllm
```

2. 创建弹性负载均衡

我们参考弹性负载均衡官方文档进行服务创建。

a. 创建弹性负载均衡



在此例中，我们的 Deepseek 集群仅用于集群自带的 Open WebUI 内网使用，不涉及外网通讯，因此 “网络类型” 选择 “内网”，如果你想将此集群暴露到外网，则选择外网。VPC 和子网跟 Deepseek 服务器保持一致。

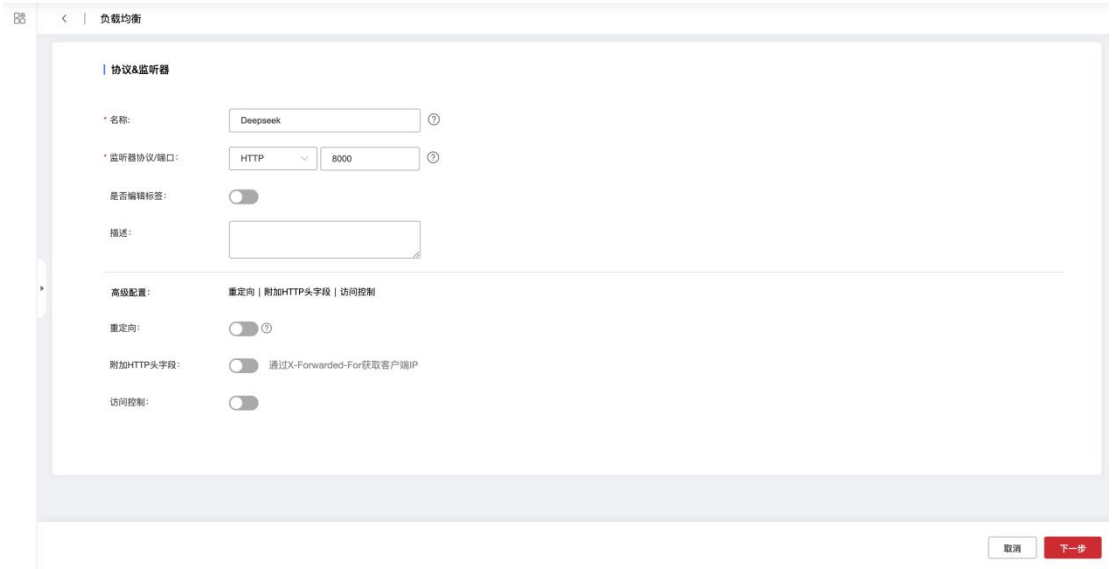


b. 创建监听器

完成网络负载均衡创建，接下来配置监听器，在负载均衡列表页，点击开始配置。

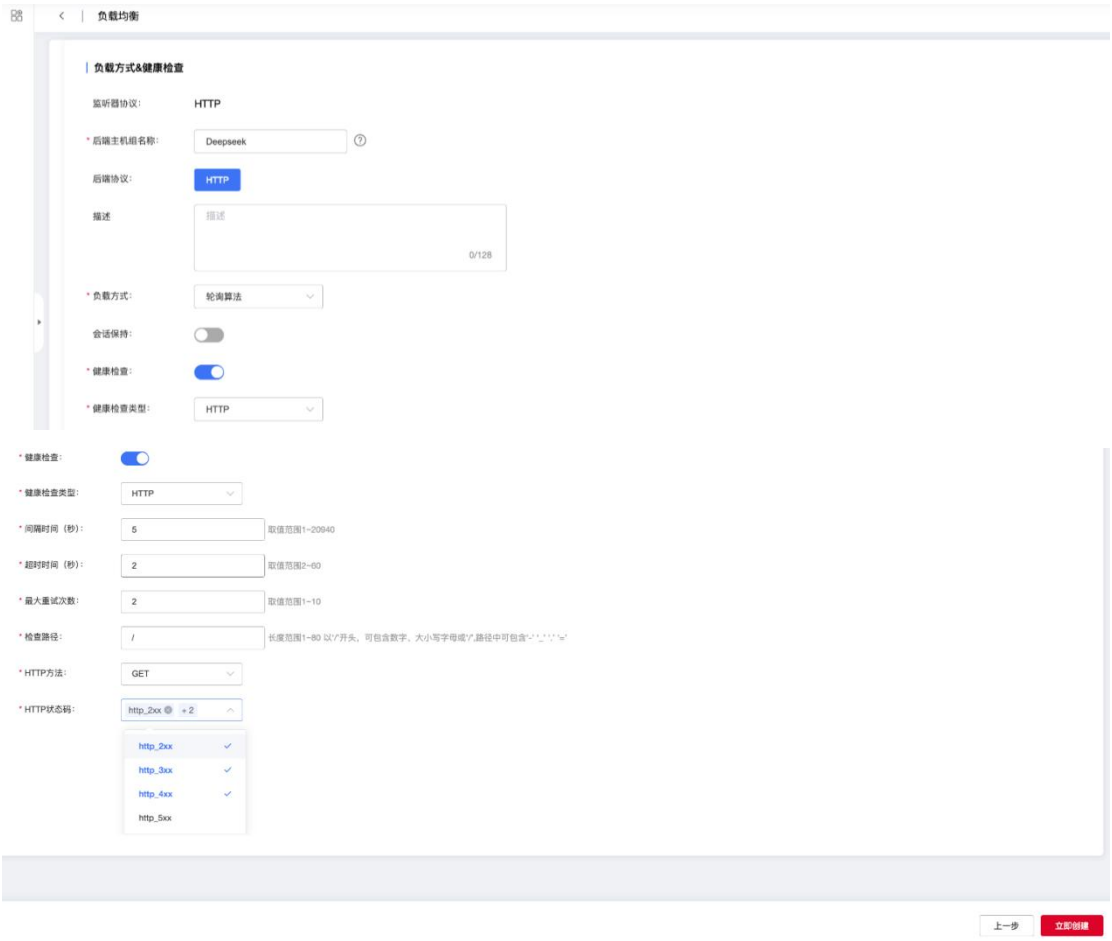


端口类型，选择 “HTTP”，端口填 8000。



进入下一步，创建健康检查。这里我们通过 HTTP 进行健康检查，返回码部分选

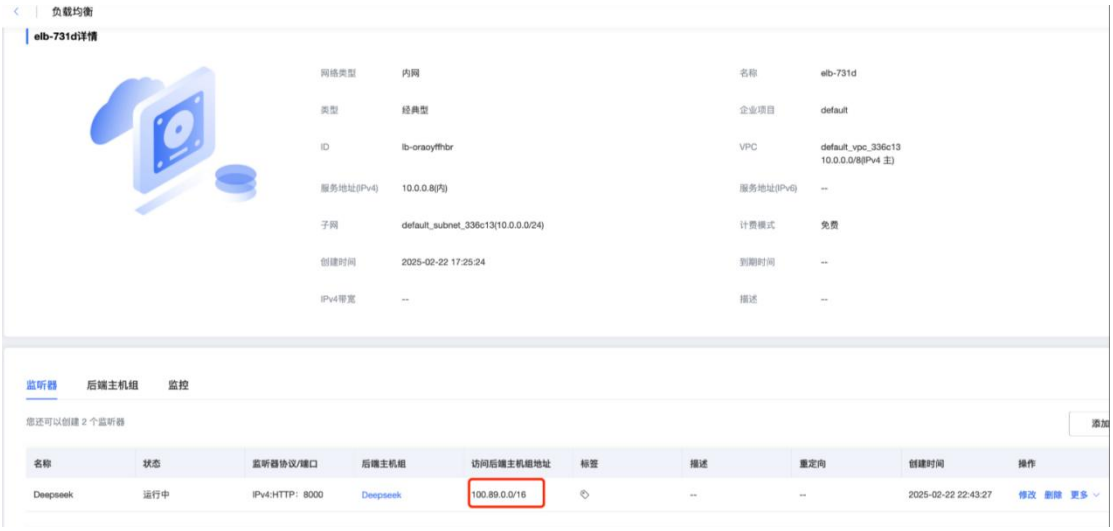
择 2xx、3xx、4xx。



点击“立即创建”，完成创建。

c. 为监听器开通 GPU 云主机白名单

此时监听器列表中，我们可以看到访问 Deepseek 服务器的主机组地址为“100.89.0.0/16”，我们需要在 Deepseek 集群的安全组中，为 vllm 端口配置此地址的白名单。



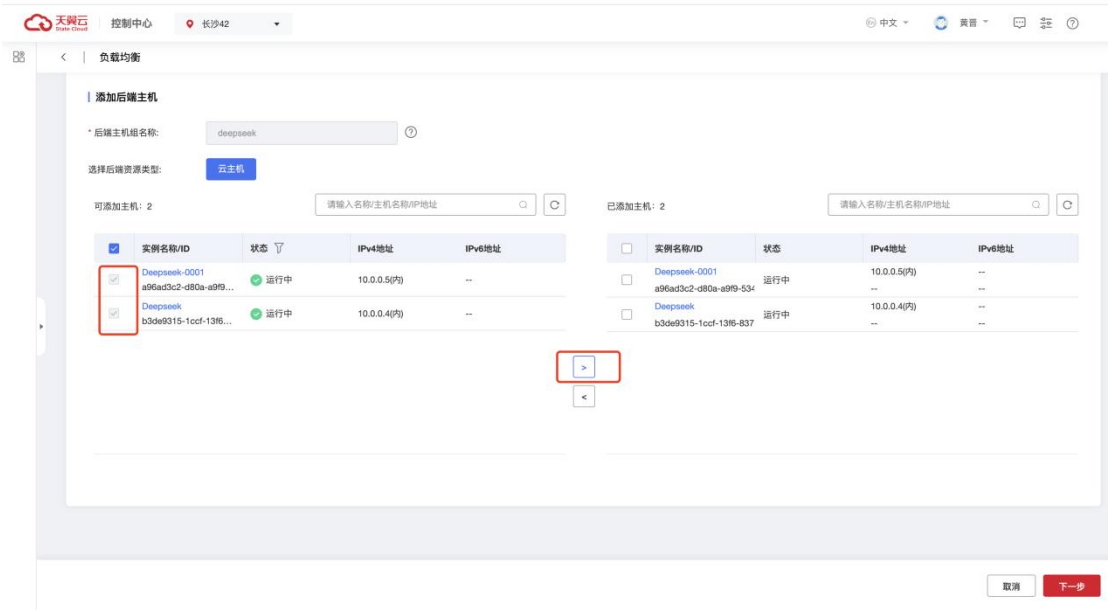
为 Deepseek 云主机所在的安全组新增规则，操作文档详见添加安全组规则帮助文档。



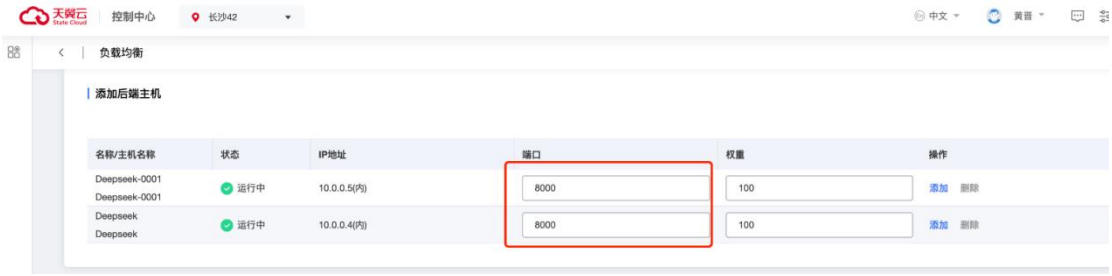
d. 为监听器添加后端云主机



选中主机，然后下一步。



设置端口为 “8000”。



点击确认即可完成配置。

e. 验证连通性

监听器页面，显示监控检查正常。



通过一台 Deepseek 服务器，测试负载均衡连通性。

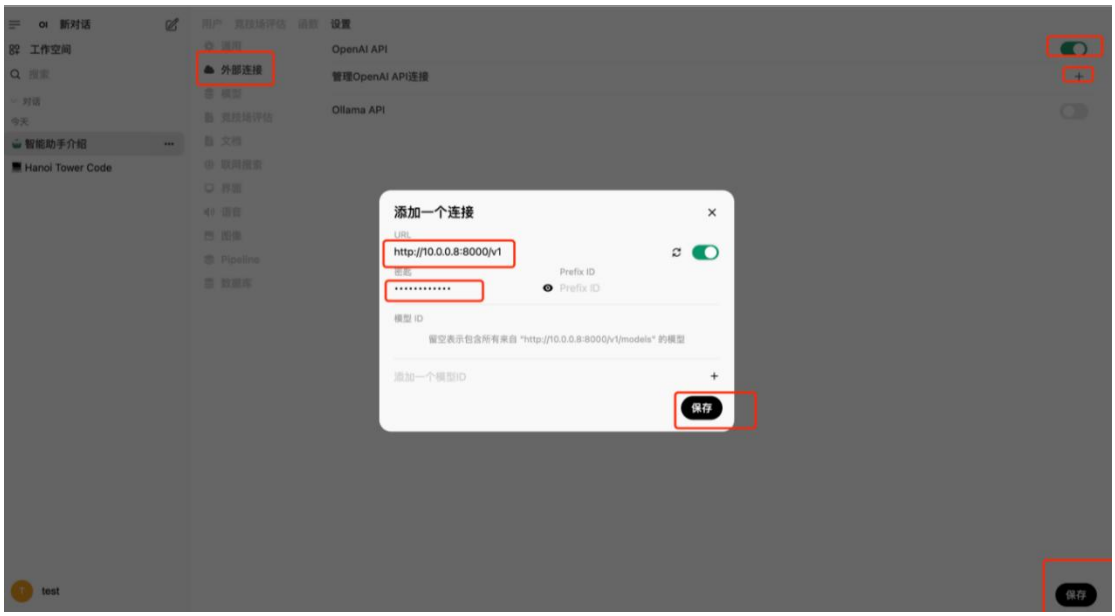
```
root@Deepseek:~# telnet 10.0.0.8 8000
Trying 10.0.0.8...
Connected to 10.0.0.8.
Escape character is '^]'.
```

步骤三：配置 UI 访问能力

为便于公网访问，我们选择一台 DeepSeek 服务器，利用 DeepSeek 服务器自带的 Open WebUI 进行交互。Open WebUI 使用文档见 Open WebUI 快速入门。

注意

openai API 地址为前面弹性负载均衡的地址，在本例中为 “http://10.0.0.8:8000/v1”，密钥为我们刚刚设置的 vllm 的 --api-key。



10.38.13 使用云主机启动模版快速部署 DeepSeek 的最佳实践

前置步骤

进入创建云主机启动模版页面

a. 点击天翼云门户首页的“控制中心”，输入登录的用户名和密码，进入控制中心页面。



b. 单击“产品服务列表>弹性云主机”，进入计算控制台，单机“云主机启动模版”，进入启动模版列表页。



c.单击“创建启动模版”，进入云主机实例模版创建页。



大模型服务器启动模版配置

a.根据业务需求配置“计费模式”、“地域”、“可用区”、“企业项目”、“虚拟私有云”等。

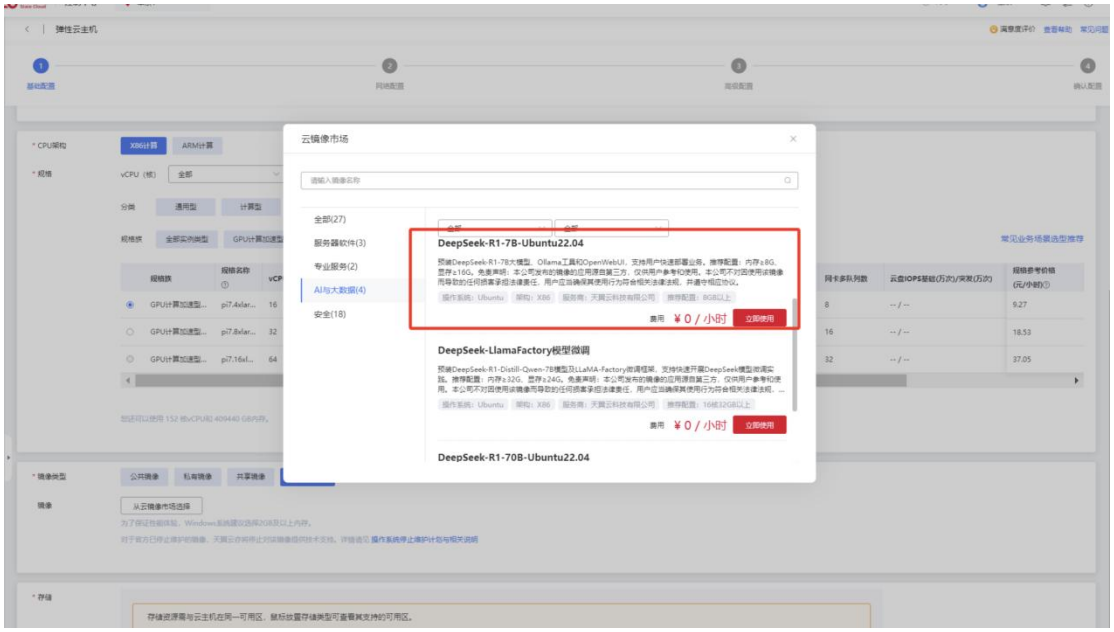
b.选择规格。此处选择"CPU 架构"为"X86"、"分类"为"GPU 加速/AI 加速型"、"

规格族"为"GPU 计算加速型 pi7"、"规格"为"pi7.4xlarge.4"。

c.选择镜像。"镜像类型"选择"镜像市场",在云镜像市场中选择预置了 DeepSeek R1 模型的 DeepSeek-R1-7B-Ubuntu22.04 镜像。

注意

本镜像推荐配置：内存≥8G、显存≥16G。



注意

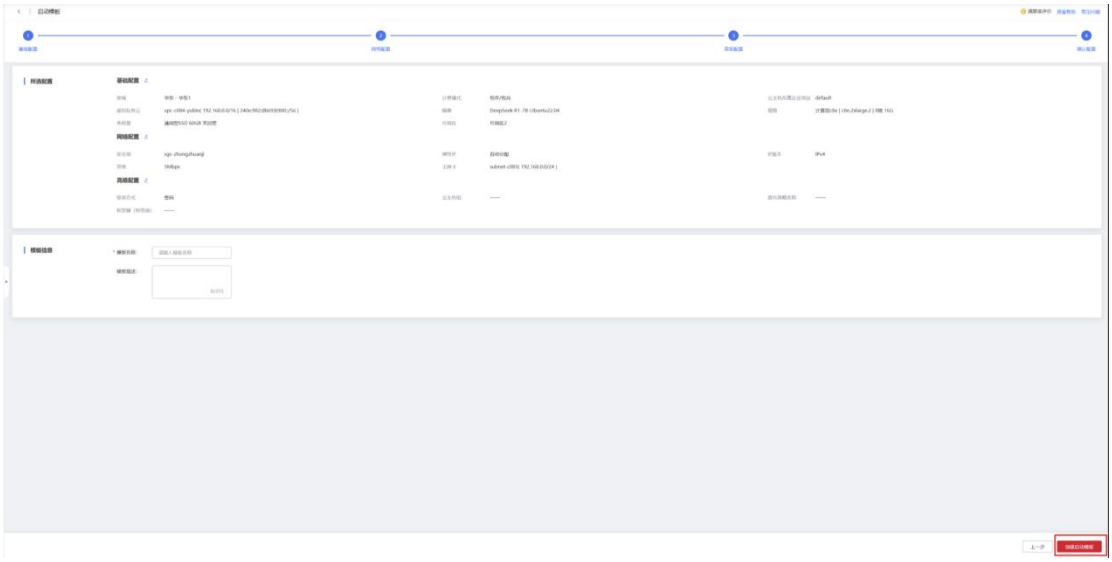
目前提供了预装多种量化精度、多种参数量的模型，如您有需求，也可在云镜像市场中进行选择。

d.设置云盘类型和大小。



e.根据需要设置网络，包括"网卡"、"安全组"，同时配备 "弹性 IP" 用于下载和访问模型；设置高级配置，包括"登录方式"、"云主机组"、"用户数据"。

f.填写此模版名称，创建启动模版。



启动模版使用

a.单击左侧导航栏“产品服务列表”，选择“计算 >弹性云主机 > 云主机启动模版”，进入云主机启动模版控制列表页。

b.找到想要使用的云主机启动模版，点击操作中“创建云主机”。



c.在创建云主机页面将默认选中该条云主机启动模版。创建云主机需要的参数按照模版中内容已默认填写。



10.38.14 弹性伸缩搭配 GPU 云主机水平扩展最佳实践

对比传统部署模式，GPU 云主机配合弹性伸缩使用，可以帮助您在推理任务高峰期自动扩容GPU节点减少卡顿,在任务请求低峰期减少GPU服务器规模节约成本，保障业务连续性的同时降低资源、运维成本。

操作步骤

1.使用大模型镜像创建包年包月的云主机实例

创建一定数量的包年包月 GPU 实例，用于后续将实例添加到伸缩组，满足大模型业务模块的非高峰或低谷时间段要求。

a. 登录控制中心。

b. 根据业务部署需求选择弹性伸缩组及其管理的云主机实例所在地域。

c. 单击“计算>弹性云主机”，进入弹性云主机管理控制台。

d. 在右上角点击“创建云主机”按钮，进入到弹性云主机创建页面。

e. 在弹性云主机创建页面，完成云主机参数配置。创建成功之后如图：



说明

搭建 DeepSeek 等大模型云主机选型、部署等说明可参考 GPU 云主机/弹性云主机：零基础搭建 DeepSeek 云端环境指南-弹性云主机-DeepSeek 专题-实践指南：DeepSeek 驱动高效能云生态 - 天翼云。

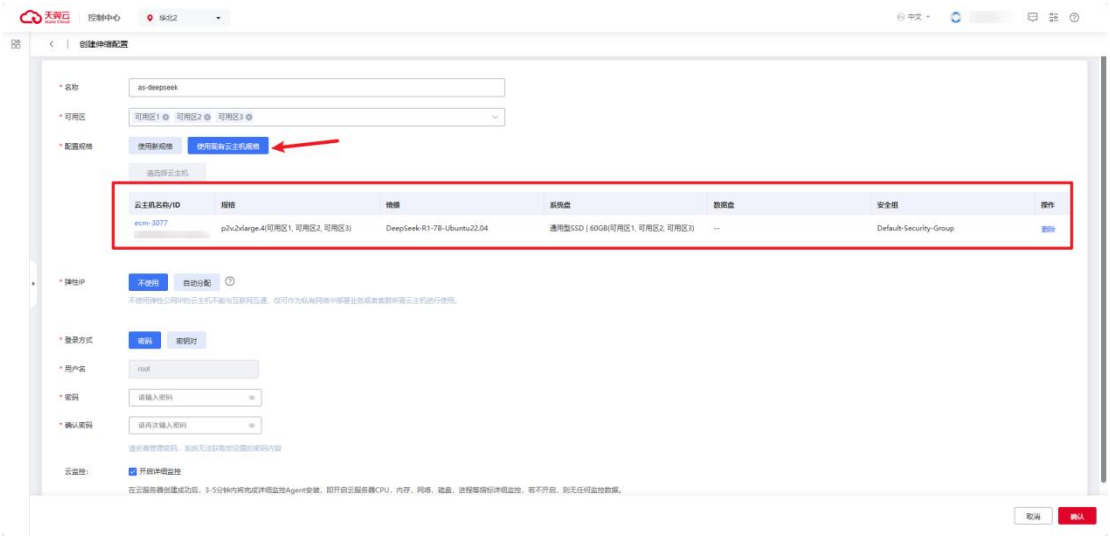
2. 创建使用大模型镜像的伸缩配置

为业务创建可用于自动扩容的云主机模板，即弹性伸缩配置。

a. 登录控制中心，选择弹性伸缩组所在地域（与前一步骤中创建的 GPU 云主机需保持一致）。

b. 单击“计算>弹性伸缩服务”，进入弹性伸缩管理控制台并切换到伸缩配置页面。

c. 点击右上角“创建伸缩配置”，进入“创建伸缩”配置页面。使用步骤 1 中创建的实例作为模板创建伸缩配置。



d. 点击“确认”，即可创建伸缩配置成功，创建成功如图所示：



3. 创建伸缩组并添加云主机实例

为低成本、高稳定性业务集群创建弹性伸缩组，并将创建好的云主机实例添加至此伸缩组内，来满足日常业务的需求。

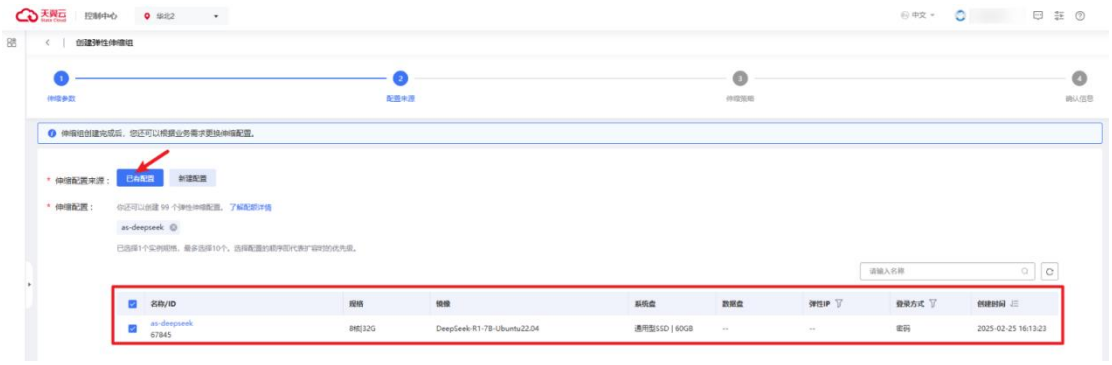
a. 登录控制中心，选择弹性伸缩组所在地域（与前一步骤中创建的 GPU 云主机需保持一致）。

b. 单击“计算>弹性伸缩服务”，进入弹性伸缩管理控制台。

c. 在“弹性伸缩组”页签中，点击右上角“创建弹性伸缩组”，进入“创建弹性伸缩组>伸缩参数”配置页面。

d. 在“伸缩参数”页面完成基础配置，具体参数配置可参见创建伸缩组。

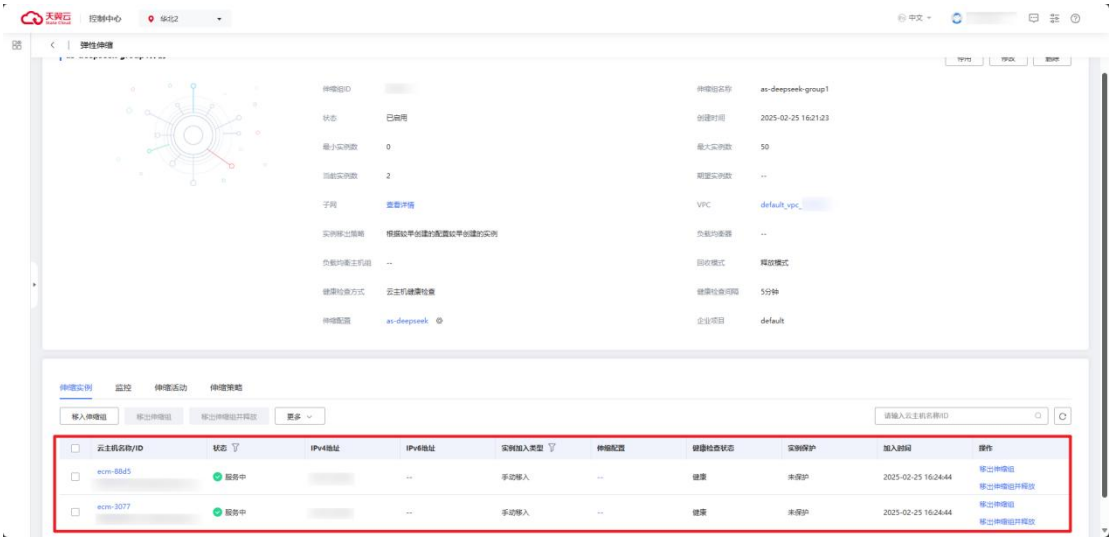
e. 在“配置来源”页面，选择已有配置，为伸缩组绑定一个或多个配置。



f. 点击“立即创建”伸缩组创建成功，如下图：



g. 伸缩组创建成功之后，进入到伸缩组详情页面，将步骤 1 中创建包年包月弹性云主机添加至伸缩组内并开启伸缩保护防止实例被移除，用于日常业务需求的满足。具体操作步骤请参见将实例移入伸缩组。添加成功之后如图：



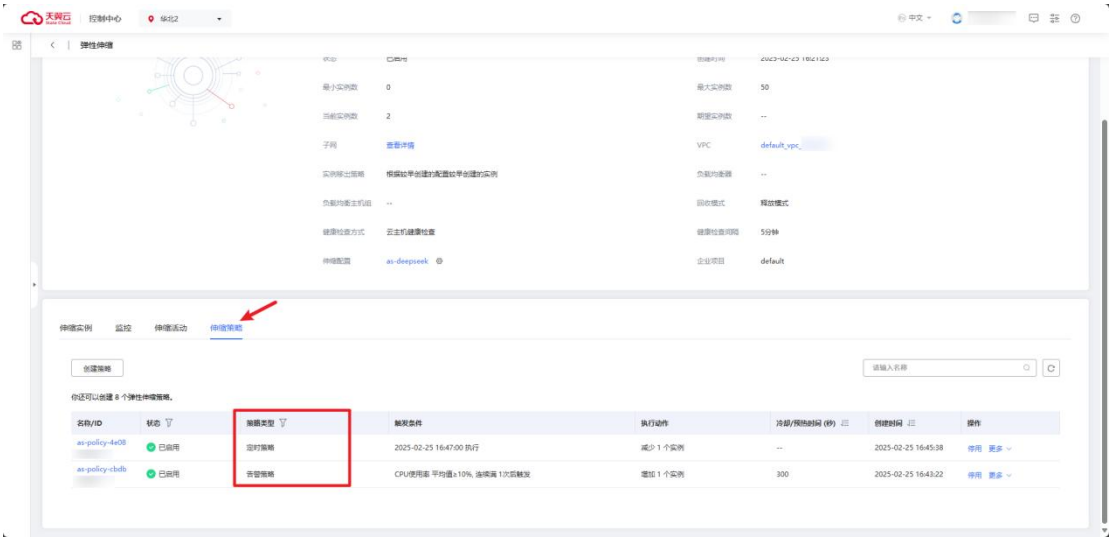
4. 根据业务场景为伸缩组设置自动伸缩策略

为已创建好的伸缩组设置伸缩策略，来满足突发的业务流量变化，确保业务的稳定运行。

a. 进入弹性伸缩管理控制台，点击步骤 3 创建好的伸缩组名称，进入到伸缩组详情页面。

b. 单击“伸缩策略”进入伸缩策略页签，单击“创建策略”按钮，进入到“创建伸缩策略”页面。

c. 在“创建伸缩策略”页面中完成策略的配置。您可以根据使用场景选择告警、定时等自动伸缩策略，帮助伸缩组在业务高峰期自动扩容大模型实例、在业务低谷期自动释放实例，保证业务稳定的同时最大程度的优化资源使用成本。具体操作步骤可参见创建伸缩策略。



10.38.15 基于 Anything LLM 实现云主机外挂 Embedding 模型、向量数据库的最佳实践指南

前言

AnythingLLM 是一个开源项目，旨在简化将本地文档和数据源转化为可检索、可交互知识库的过程。它允许用户利用大语言模型（LLM）的能力，基于自身的私有数据进行智能问答、语义搜索等操作。其技术架构包含以下三方面：

- Embedding 模型：使用 Embedding 模型将文本转换为向量表示，以便在向量空间中进行相似度计算。常见的 Embedding 模型如 BGE - M3 等，能够捕捉文本的语义信息，提高检索的准确性。
- 向量数据库：用于存储和管理文本向量，支持高效的向量搜索。AnythingLLM 可以集成多种向量数据库，如 Chroma、Milvus 等，根据实际需求选择合适的数据库来满足性能和规模的要求。

- 大语言模型（LLM）：负责根据检索到的相关内容生成自然语言回答。用户可以选择不同的 LLM，如 OpenAI 的 GPT 系列、国产的通义千问、文心一言等，或者使用开源的 LLM 模型。

客户在搭建完具备推理训练基本能力的私有 DeepSeek 服务后，往往需要外挂知识库以达到应答更专业的目的。外挂知识库涉及到 Embedding 模型以及向量数据库这两大部分。本文档结合 GPU 物理机提供的 DeepSeek 服务，结合 Anything LLM，介绍如何使用云主机外挂 Embedding 模型以及向量数据库。

准备

- 创建一台物理机并搭建 DeepSeek 服务，参考 NVIDIA GPU 物理机搭建 DeepSeek 指南（单机版）。
- 安全组放开 Anything LLM 云主机的弹性 IP+端口 8000，参考安全组-物理机-用户指南-安全 - 天翼云。

部署 Anything LLM

1. 创建通用云主机

创建步骤请参考创建弹性云主机-弹性云主机-快速入门 - 天翼云，此处以“s8r.large.2”规格，“Ubuntu Server 22.04 64”镜像为例部署。

2. 部署 docker 软件

```
#备份软件源 sudo cp /etc/apt/sources.list /etc/apt/sources.list.bak#替换软件源，在编辑器中删除原内容添加备份源配置 sudo vim /etc/apt/sources.list#更新软件包索引 sudo apt update#安装 Docker 依赖 sudo apt install apt-transport-https ca-certificates curl software-properties-common -y#添加 Docker 官方 GPG 密钥

curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o /usr/share/keyrings/docker-archive-keyring.gpg#添加 Docker 软件源 echo "deb [arch=$(dpkg --print-architecture) signed-by=/usr/share/keyrings/docker-archive-keyring.gpg] https://download.docker.com/linux/ubuntu $(lsb_release -cs) stable" | sudo tee /etc/apt/sources.list.d/docker.list > /dev/nullsudo apt update#安装 Docker 引擎 sudo apt install docker-ce docker-ce-cli
```

```
containerd.io -y#启动并设置 Docker 服务 sudo systemctl start dockersudo systemctl enable docker
```

说明

后续使用 docker 拉取镜像时，可将其配置为国内源以提升镜像拉取速度。

3. 部署 Anything LLM

```
docker pull mintplexlabs/anythingllmexport STORAGE_LOCATION=$HOME/anythingllm && \mkdir -p $STORAGE_LOCATION && \touch "$STORAGE_LOCATION/.env" && \

docker run -u root -d -p 3001:3001 \ #指定宿主机核容器内服务到的端口映射，可自主选择

--cap-add SYS_ADMIN \ # Docker 命令中的一个选项，用于给容器添加额外的 Linux 内核能力

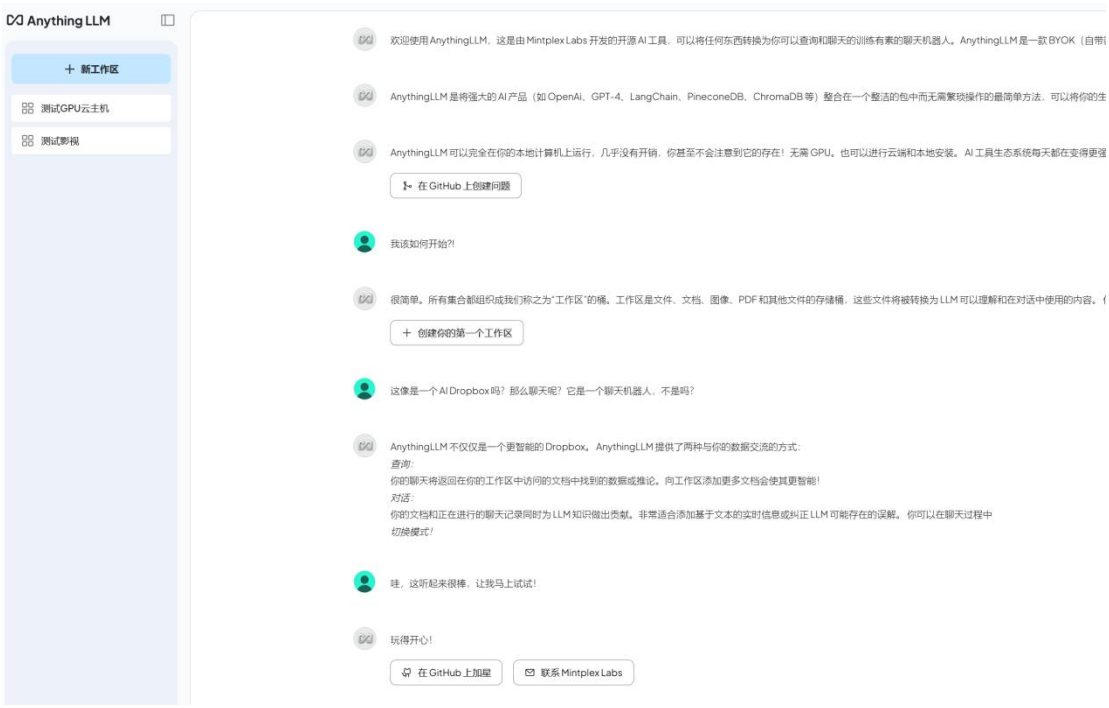
-v ${STORAGE_LOCATION}:/app/server/storage \ #服务启动数据宿主机目录和容器目录的映射

-v ${STORAGE_LOCATION}/.env:/app/server/.env \ #服务启动环境变量配置文件，宿主机和容器的映射

-e STORAGE_DIR="/app/server/storage" \  #容器内的环境变量，指定服务数据在容器内存放目录

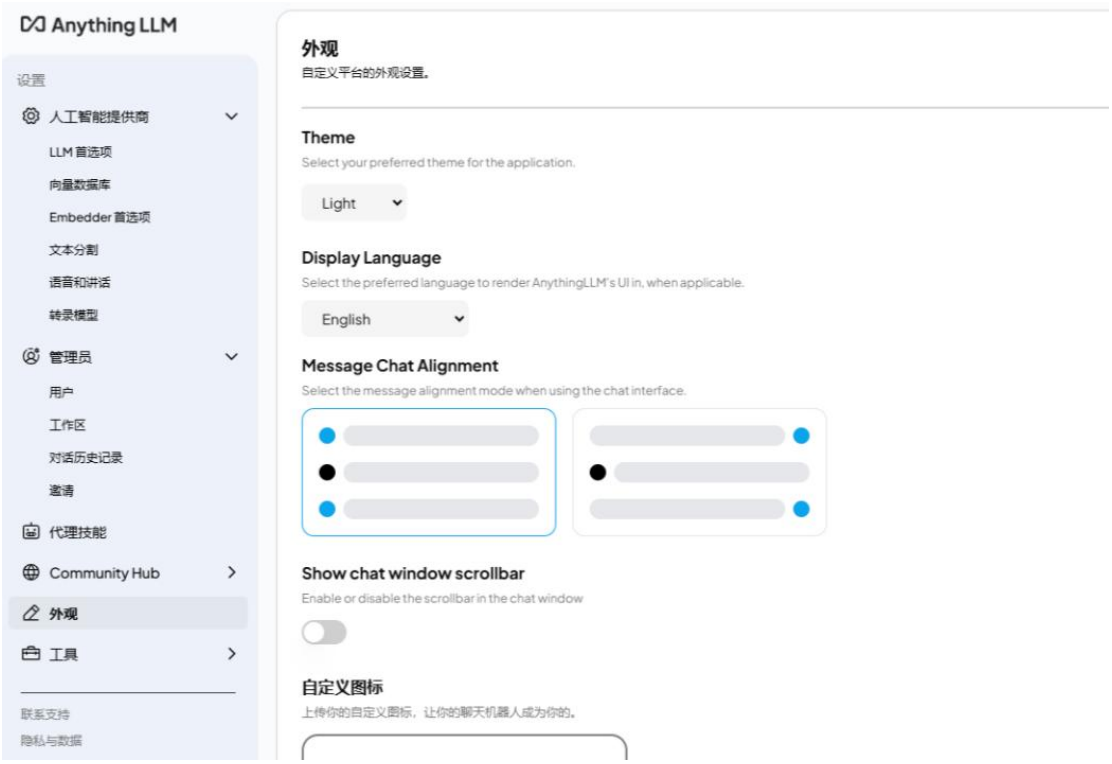
mintplexlabs/anythingllm
```

4. 登录 Anything LLM

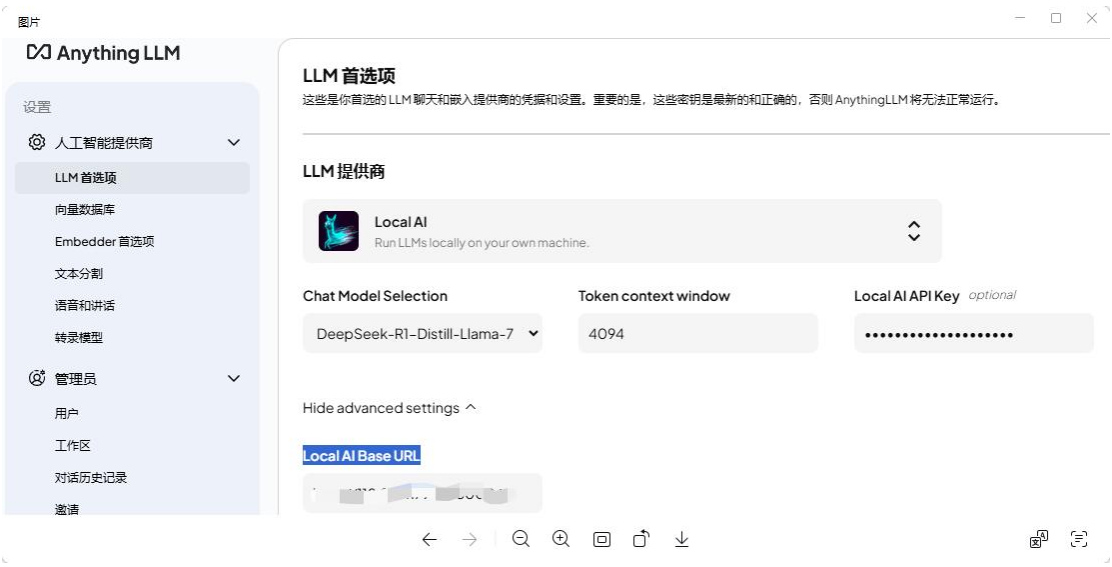


接入 GPU 物理机提供的 DS 模型

1. 登录 Anything LLM，选择设置 LLM 首选项。



2. 设置“Local AI Base URL”，其中”Local AI Base URL“填充 GPU 物理机 DS 地址和端口，然后保存。



Embedding 模型

不同的任务需要不同的模型和数据库支持。如信息检索任务，需要 embedding 模型能准确捕捉文本语义信息，向量数据库具备高效的检索功能；对于推荐系统任务，模型要能挖掘用户和物品的潜在特征，数据库要支持快速的相似性计算。Anything LLM 支持多种 Embedding 模型引擎，包括 AnythingLLM Native Embedder (default)、OpenAI、Azure OpenAI、LocalAi (all)、Ollama (all)、LM Studio (all)、Cohere。下面以 LocalAi (all)使用 BGE - M3 模型为例说明，其他请参考对应软件官网安装使用。

1. 创建 GPU 云主机并开放安全组规则。

   BGE - M3 是一款文本嵌入模型，在部署使用时，云主机配置会受任务类型（如推理、训练）、数据规模、并发请求数量等因素影响。如果是训练任务建议使用 1 张 A100，其他任务可根据实际需求创建规格。创建步骤请参考创建配备 GPU 驱动的 GPU 云主机（Linux）-GPU 云主机-用户指南-创建 GPU 云主机 - 天翼云。

   安全组放开 Anything LLM 云主机弹性 IP+端口 8000，参考添加安全组规则-弹性云主机-用户指南-安全-安全组-配置安全组规则 - 天翼云。

2. 下载模型

   2.1.安装依赖包

   参考文档在天翼云使用 vLLM 运行 DeepSeek 的最佳实践-32B 等版本-自定义部署 DeepSeek-步骤三：手动部署 DeepSeek -> 1.安装依赖包。

2.2. 下载 BGE - M3 模型

将以下内容保存为 model_download.py 文件，参数 cache_dir 为模型的下载路径，您可以按需修改，需确保存储空间足够存放模型。

```
from modelscope import snapshot_download

model_dir = snapshot_download('BAAI/bge-m3',
cache_dir='/root/model-path', revision='master')
```

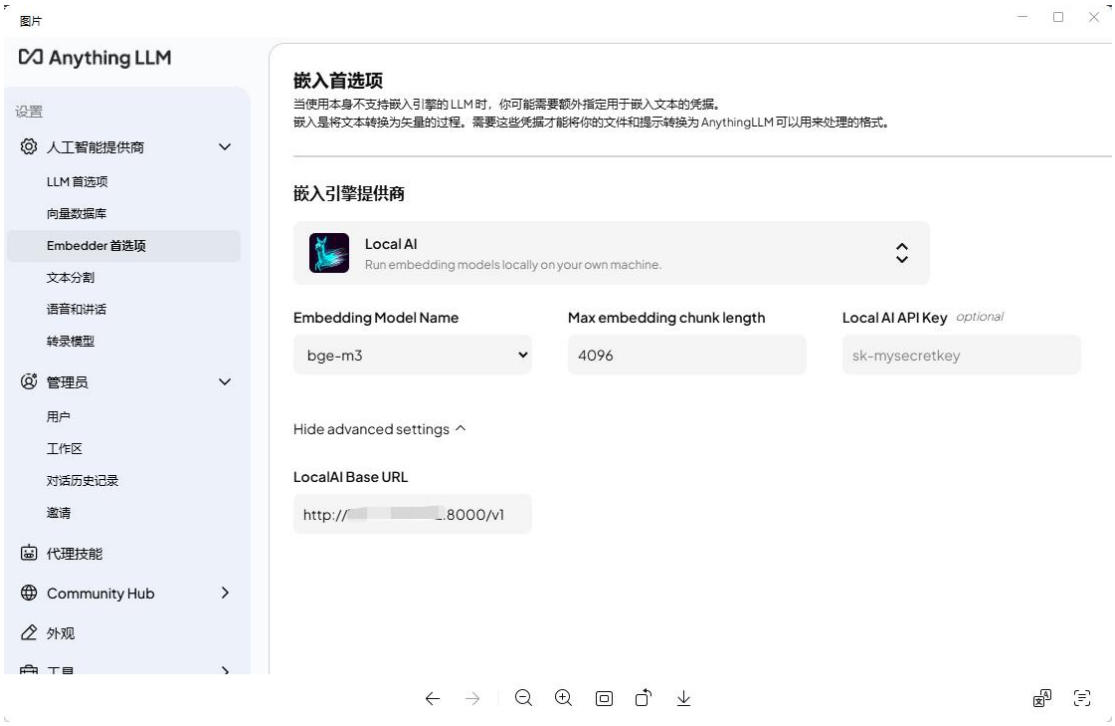
运行该文件，启动模型下载。

```
python model_download.py
```

2.3. 运行模型

```
vllm serve BAAI/bge-m3 --served-model-name bge-m3
--gpu-memory-utilization 0.95 --enforce
```

3. 配置 Embedding 模型引擎



4. 使用 Embedding 模型引擎



向量数据库

Anything LLM 支持多种向量数据库，包含 LanceDB (default)、Astra DB、Pinecone、Chroma、Weaviate、Qdrant、Milvus、Zilliz。以下为各个向量数据库的对比描述：

向量数据库	开源/闭源	主要特点	支持语言	部署方式	性能优化	索引算法	使用场景
LanceDB	开源	基于 Rust，支持多模态数据，零拷贝、自动版本控制	Python、JavaScript/TypeScript	本地、云	GPU 加速	HNSW、IVF 等	AI 应用、多模态数据

向量数据库	开源/闭源	主要特点	支持语言	部署方式	性能优化	索引算法	使用场景
							检索
Astra DB	闭源（云服务）	高可用性、弹性扩展，支持多模态数据	Python、Java、JavaScript 等	云服务	分布式优化	HNSW 等	需要高可用性和弹性扩展的应用
Pinecone	闭源（云服务）	高性能向量搜索，支持 GPU 和 TPU 加速	Python、JavaScript 等	云服务	GPU/TPU 加速	HNSW、IVF 等	大规模向量搜

向量数据库	开源/闭源	主要特点	支持语言	部署方式	性能优化	索引算法	使用场景
							索
Chroma	开源	适用于语义搜索和相似性匹配，支持单机和分布式部署	Python、JavaScript等	本地、云	并行查询优化	HNSW等	AI应用、语义搜索
Weaviate	开源	云原生，支持多模态数据，模块化设计	Python、JavaScript、Java等	本地、云	高效查询优化	HNSW等	多模态数据检索、推荐系统
Qdrant	开源	高效向量检	Python、JavaScript等	本地、	并行查询	HNSW等	高性

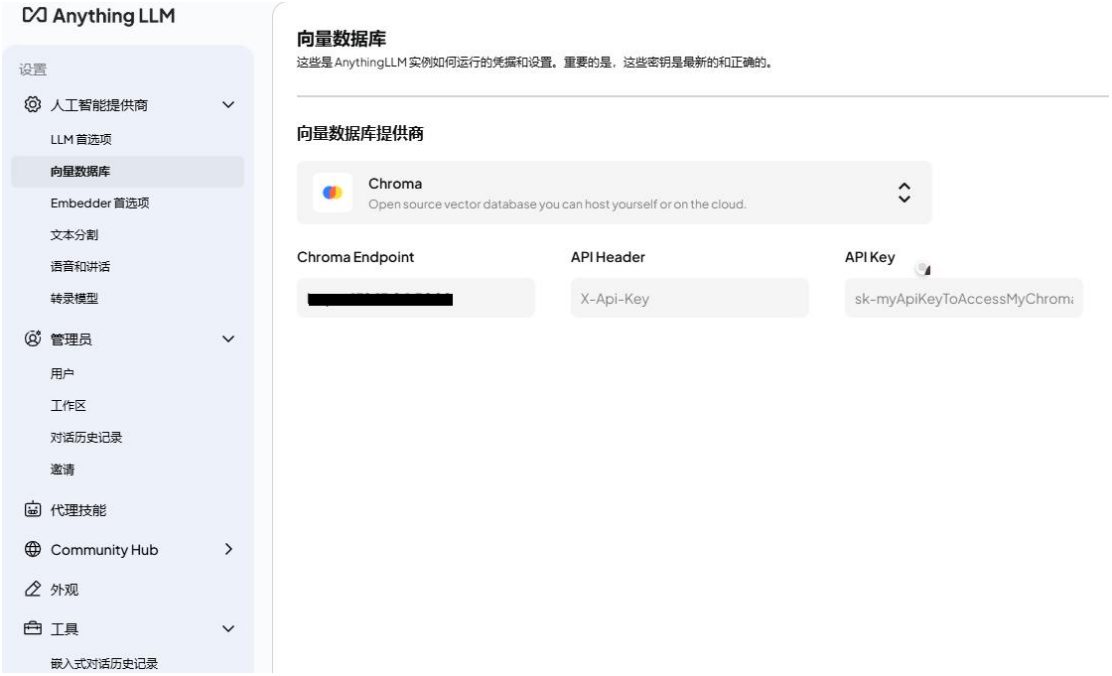
向量数据库	开源/闭源	主要特点	支持语言	部署方式	性能优化	索引算法	使用场景
		索，支持加密和安全访问控制		云	优化		能向量检索
Milvus	开源	分布式向量数据库，支持大规模数据查询	Python、Java、C++等	本地、云	GPU加速	HNSW、IVF等	人工智能、推荐系统
Zilliz	开源（云服务）	基于Milvus，深度优化，支持云原生部署	Python、Java 等	云服务	高性能优化	HNSW等	云原生向量检索

下面以 Chroma 为例进行说明，其他请参考对应软件官网安装使用。

1. 创建一台挂载 XSSD 云盘的本地盘云主机/通用云主机，并增加 Anything LLM 云主机弹性 IP+端口 8000 的安全组规则。创建步骤请参考创建弹性云主机-弹性云主机-快速入门 - 天翼云，增加安全组规则参考添加安全组规则-弹性云主机-用户指南-安全-安全组-配置安全组规则 - 天翼云。
2. 安装向量数据库

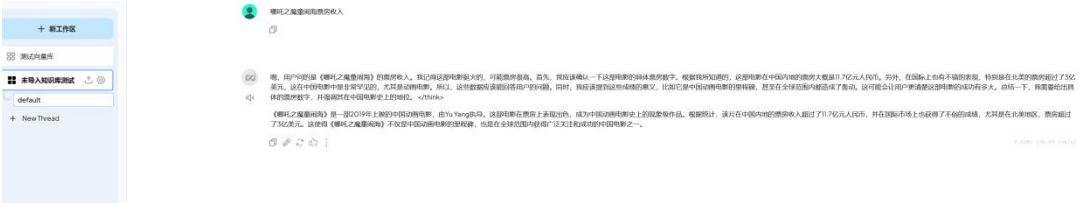
```
docker pull chromadb/chroma:latestdocker run -d --name chroma -p 8000:8000 chromadb/chroma:latest
```

3. 配置向量数据库



4. 使用

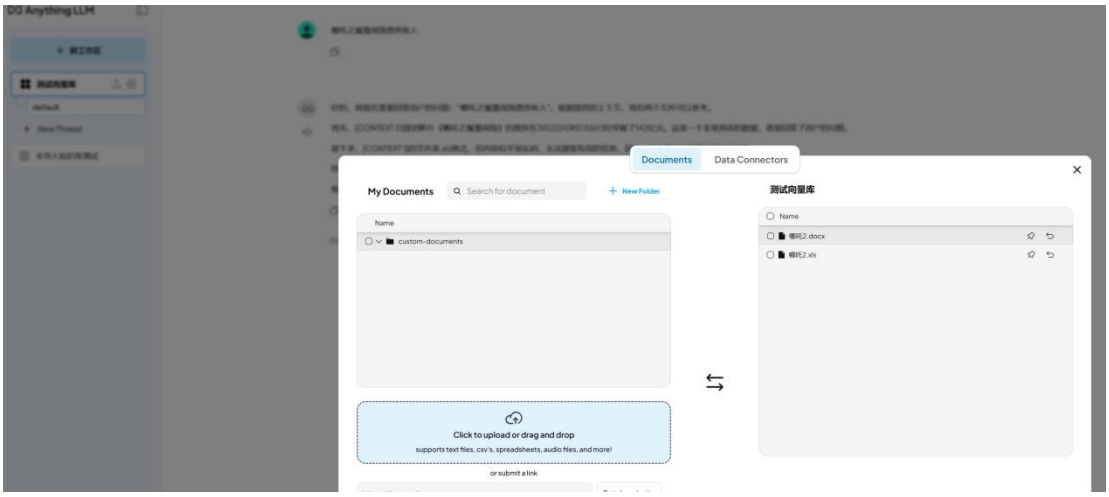
4.1. 创建工作区，聊天查询“哪吒之魔童闹海票房收入”，回答结果没有相关知识库引入，且出现了论述错误。



4.2. 准备知识库，内容如下：



4.3.导入知识库，向量使用默认向量。



4.4.继续聊天查询“哪吒之魔童闹海票房收入”，回答结果存在相关知识库引入，且论述正确。



10.38.160 大模型推理性能测试实践

1 前言

本文介绍三种大模型评测工具，包括 EvalScope、LLMPerf 和 vLLM Benchmark，总结每种测评工具的测试实践，并对三种测评工具在核心功能、性能指标、典型应用场景等方面进行了对比。

2 EvalScope

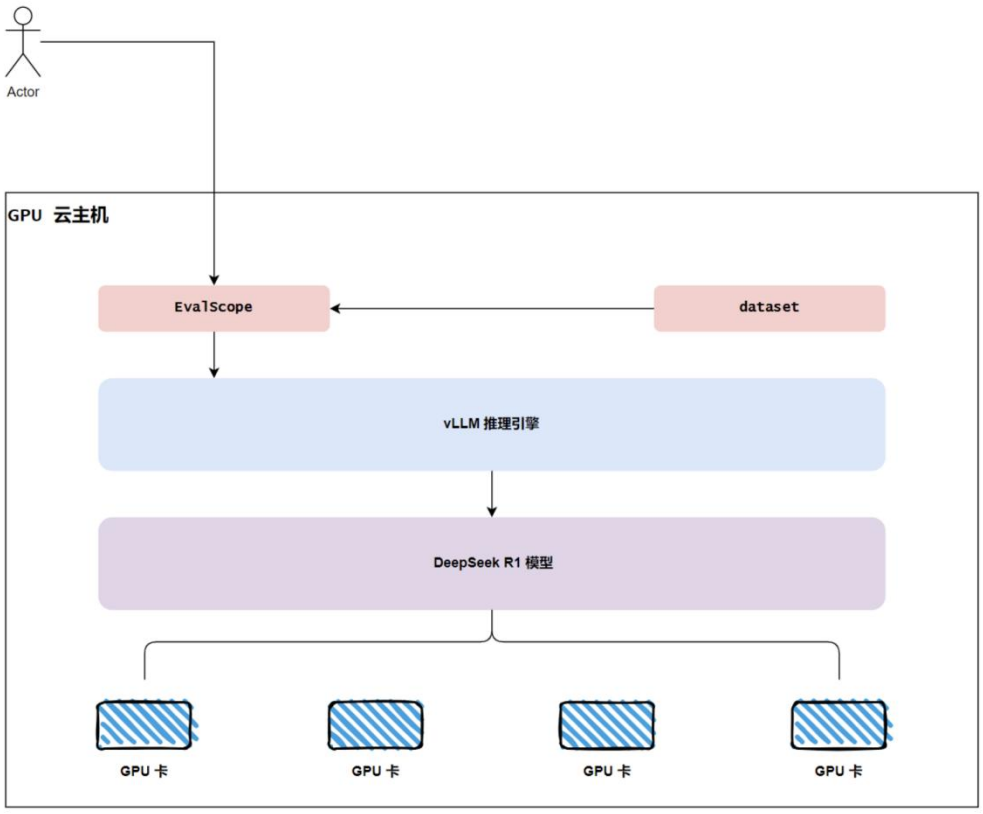
2.1 前言

EvalScope 是魔搭社区官方推出的模型评测与性能基准测试框架，内置多个常用测试基准和评测指标，如 MMLU、CMMLU、C-Eval、GSM8K、ARC、HellaSwag、TruthfulQA、MATH 和 HumanEval 等；支持多种类型的模型评测，包括 LLM、多模态 LLM、Embedding 模型和 Reranker 模型。

下面介绍基于 EvalScope 评测 DeepSeek R1 模型，涵盖基本评测和模型推理性能压测两大场景。

2.2 测试环境配置

如下图所示为使用 EvalScope 测评工具对 DeepSeek R1 模型进行测试的拓扑图，其中推理引擎选择 vLLM。



安装 EvalScope 测评工具的步骤如下所示：

```
pip install evalscope
```

2.3 测试实践

● 基本评测场景

本文采用指定模型 API 地址（api_url）和 API Key（api_key），评测部署的模型 API 服务，指定 eval-type 参数为 service。

测试命令示例：

```
evalscope eval \
 --model DeepSeek-R1-Distill-Qwen-32B \
 --api-url http://127.0.0.1:8000/v1/chat/completions \
 --api-key Vb7h4cVRFPgfAr8a  \
 --eval-type service \
 --datasets gsm8k \
 --limit 10
```

测试结果：

```
+-----------------------------+----------+----------------+-------
---+-------+---------+---------+
|              Model          |  Dataset |     Metric     |
Subset  |   Num |  Score  | Cat.0   |
+-----------------------------+----------+----------------+-------
---+-------+---------+---------+
| DeepSeek-R1-Distill-Qwen-32B |  gsm8k   | AverageAccuracy |  main
|    10 |    0.9  | default |
+-----------------------------+----------+----------------+-------
---+-------+---------+---------+
```

● 推理性能压测场景

使用 openqa 数据集进行推理性能压测。

openqa 可以模拟真实的问答情况，更接近现实场景，其中包括了超 3000 条中文对话数据，可以有效评估模型在实际应用中的推理性能。

测试命令示例：

```
evalscope perf \
    --url "http://127.0.0.1:8000/v1/chat/completions" \
    --parallel 1 \
    --model qwen2.5 \
    --number 15 \
    --api openai \
    --dataset openqa \
    --stream
```

要准确统计 Time to First Token (TTFT)指标，需要在请求中包含--stream 参数。

测试结果如下所示：

```
Percentile results:
+------------+----------+----------+------------+--------------+----
----------+----------------------+
| Percentile | TTFT (s) | TPOT (s) | Latency (s) | Input tokens | Output
tokens | Throughput(tokens/s) |
+------------+----------+----------+------------+--------------+----
----------+----------------------+
|    10%     |  0.0202  |  0.0027  |   0.1846   |      41      |
50       |       270.8324       |
|    25%     |  0.0209  |  0.0028  |   0.2861   |      44      |
83       |       290.0714       |
```

50%	0.0233	0.0028	0.7293	49	250	335.644
66%	0.0267	0.0029	0.9052	50	308	340.2603
75%	0.0437	0.0029	0.9683	53	325	341.947
80%	0.0438	0.003	1.0799	58	376	342.7985
90%	0.0439	0.0032	1.2474	62	424	345.5268
95%	0.0463	0.0033	1.3038	66	431	348.1648
98%	0.0463	0.0035	1.3038	66	431	348.1648
99%	0.0463	0.0037	1.3038	66	431	348.1648

使用 openqa 可以有效评估模型在处理实际自然语言问题时的具体响应时间，如上展示了模型在不同百分位下的 TTFT(首 token 时间）、TPOT（生成每个 token 需要的时间）、Latency（延迟时间）、Input tokens（输入 token）、Output tokens（输出 token）、Throughput（吞吐率）等指标。

此外，使用 Speed Benchmark 工具可以重点测试模型在不同输入长度下的推理速度。以下为测试生成 2048 tokens 时的速度，输入长度分别为 1、6144、14336、30720。

测试命令示例：

```
evalscope perf \
 --parallel 1 \
 --url http://127.0.0.1:8000/v1/completions \
 --model DeepSeek-R1-Distill-Qwen-32B \
 --log-every-n-query 1 \
 --connect-timeout 60000 \
 --read-timeout 60000\
 --max-tokens 2048 \
 --min-tokens 2048 \
 --api openai \
 --dataset speed_benchmark \
 --api-key Vb7h4cVRFPgfAr8a
```

参数说明如下所示：

- parallel：指定并发数，默认为 1 个；
- url：指定待测试系统的服务地址；
- model：指定待测试的模型名称；
- log-every-n-query：指定日志打印间隔；
- connect-timeout：指定网络连接超时时间，单位为毫秒；
- read-timeout：指定网络读取超时时间，单位为毫秒；
- max-tokens：指定输出最大 token 值；
- min-tokens：指定输出最小 token 值；
- api：指定服务 API 类型；
- dataset：指定数据集；
- api-key：指定认证的 API key 值。

测试结果：

Prompt Tokens	Speed(tokens/s)	GPU Memory(GB)
1	44.58	0.0
6144	42.47	0.0
14336	34.55	0.0
30720	31.72	0.0

从测试结果可以看出，随着输入 token 数量增加，模型推理速度下降。

● 定制化推理性能压测场景

我们对 Speed Benchmark 工具代码进行了修改，增加了输入 token 和 repeat 参数，用户可以在压测命令中可以直接指定需要测试的输入 token 长度和 repeat 值，不需要去修改代码重新配置环境，改动代码如下所示：

1）修改数据集文件：evalscope/perf/plugin/datasets/speed_benchmark.py。注释掉之前的 INPUT_LENGTH 和 REPEAT，然后替换初始化代码和 build_messages 方法

```
    def __init__(self, query_parameters: Arguments):
        super().__init__(query_parameters)
        self.input_length = getattr(query_parameters,'dataset_input_length', [1, 6144, 14336, 30720])
        # 从参数获取重复次数，默认值为 2
        self.repeat = getattr(query_parameters, 'dataset_repeat', 2)
```

```
    def build_messages(self) -&gt; Iterator[List[Dict]]:
        for input_len in self.input_length:
            for _ in range(self.repeat):
                yield self.create_query(input_len)
```

2）修改参数加载文件 evalscope/perf/arguments.py，修改如下

```
#在 Arguments 中添加对应字段
dataset_input_length=args.dataset_input_length,
dataset_repeat=args.dataset_repeat
#在 add_argument 函数里添加：
parser.add_argument(
        '--dataset-input-length',
        nargs='+',
        type=int,
        default=[1, 6144, 14336, 30720],
        help='The length of the input, default [1, 6144, 14336, 30720]')
parser.add_argument('--dataset-repeat', type=int, default=2, help='The repeat times of the dataset')
```

安装修改后的 evalscope 工具后，执行测试的示例如下所示：

```
evalscope perf \
 --parallel 1 \
 --url http://127.0.0.1:8000/v1/completions \
 --model DeepSeek-R1-Distill-Qwen-32B \
 --log-every-n-query 1 \
 --connect-timeout 60000 \
```

```
--read-timeout 60000\
--max-tokens 2048 \
--min-tokens 2048 \
--api openai \
--dataset speed_benchmark \
--api-key Vb7h4cVRFPgfAr8a
--dataset-input-length 256 1024 2048
--dataset-repeat 5
```

当我们传入多个 dataset-input-length 值时会同时测试多个不同的输入长度的样本，总的请求数为 dataset-input-length 输入的样本个数乘以 dataset-repeat 传入的值。

3 LLMPerf

3.1 前言

LLMPerf（大型语言模型性能测试）是评估大语言模型（LLM）在生产环境中表现的核心评测体系。本文基于真实业务场景提炼最佳实践方法论，为开发、运维与测试团队提供可靠的测试方案，系统性保障服务的高吞吐、低延迟和稳定性。

3.2 测试环境配置

测试脚本下载

LLMPerf 由 Ray 项目团队开发，在 github 上开源，可直接下载。

```
git clone https://github.com/ray-project/llmper
```

工具核心功能：

性能评估

- 负载测试：LLMPerf 可以对大语言模型（如 Llama 2、GPT-3 等）的吞吐量和延迟性能进行测量。它通过负载测试来评估模型在不同负载下的响应延迟和生成吞吐量，帮助用户了解模型在实际应用中的性能表现。

● 准确性测试：该工具还可以进行正确性测试，衡量模型输出的准确性，确保模型在生成内容时能够保持较高的质量。

参数化测试

LLMPerf 允许用户更改输入和输出大小等参数，以匹配不同的应用场景。这意味着用户可以根据自己的具体需求和工作负载，灵活地调整测试参数，从而获得更具针对性的性能评估结果。

结果分析

LLMPerf 能够提供详细的性能指标分析，包括响应时间、吞吐量、准确性等，帮助用户深入了解模型的性能特点。

兼容性

LLMPerf 支持主流的大语言模型，如 OpenAI、Anthropic、TogetherAI 等。这使得用户可以在一个统一的框架下对不同来源的模型进行性能比较和评估。

安装测试工具

```
git clone https://github.com/ray-project/llmperf

cd llmperf

pip install -e .

#如下载速度较慢可以使用国内源进行加速

pip install -e . -i https://pypi.tuna.tsinghua.edu.cn/simple
```

配置环境变量

根据不同的 API 服务，需要设置不同的密钥信息和服务地址。例如对于 OpenAi（vLLM 类推理框架）：

```
export OPENAI_API_KEY="your api-key"

export OPENAI_API_BASE="https://api.endpoints.anyscale.com/v1"
```

3.3 测试实践

性能测试

使用 token_benchmark_ray.py 脚本运行测试，主要参数如下：

- --model：被测的模型名称（如 DeepSeek-R1-Distill-Qwen-32B）。
- --mean-input-tokens：输入 token 的均值。
- --mean-output-tokens：输出 token 均值。
- --stddev-input-tokens：输入 token 的标准差。
- --stddev-output-tokens：输出 token 的标准差。
- --num-concurrent-requests：并发请求数。

示例：

```
python token_benchmark_ray.py \
    --model "DeepSeek-R1-Distill-Qwen-32B" \
    --mean-input-tokens 256 \
    --stddev-input-tokens 0 \
    --mean-output-tokens 256 \
    --stddev-output-tokens 0 \
    --max-num-completed-requests 200 \
    --timeout 600 \
    --num-concurrent-requests 10 \
    --results-dir "result" \
    --llm-api openai \
    --additional-sampling-params '{}'
```

用 vLLM 部署 DeepSeek-R1-Distill-Qwen-32B 模型进行测试，测试结果主要指标如下所示：

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
-- TTFT(Time To First Token)
    mean = 3.942 s
-- Prefill-Throughput (tokens/second)
    mean = 10391.209 tokens/second
-- TPOT(Time Per Output Token)
    mean = 0.030 s
-- Decoder-Throughput
    mean = 665.873 tokens/second
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Number Of Errored Requests: 0
Overall Output Throughput: 454.10329598217925
Number Of Completed Requests: 180
Completed Requests Per Minute: 17.108177128730524
```

测试结果还包含一些不同分位的指标，这里只展示关键指标，其主要含义如下：

- TTFT（首 token 时间）：请求开始到收到第一个 token 的时间。
- Prefill-Throughput：表示模型处理输入 Prompt 的速率。
- TPOT：表示生成相邻 token 之间的间隔时间。
- Decoder-Throughput：模型的输出速率。

准确性测试

使用 llm_correctness.py 脚本可进行文本转换任务，测试模型准确性，其主要测试模型能否准确将文本转换为数字。

示例：

```
python llm_correctness.py \
    --model "DeepSeek-R1-Distill-Qwen-32B" \
    --max-num-completed-requests 100 \
    --num-concurrent-requests 5
```

用 vLLM 部署 DeepSeek-R1-Distill-Qwen-32B 模型进行准确性测试，主要测试模型的文本转化能力，测试结果如下所示：

```
Results for llm correctness test for DeepSeek-R1-Distill-Qwen-32B queried with the openai api.
Errors: 0, Error rate: 0.0
Mismatched: 0, Mismatch rate: 0.0
Completed: 35
Completed without errors: 35
```

可以看到模型的文本转换请求全部成功，没有失败的请求。

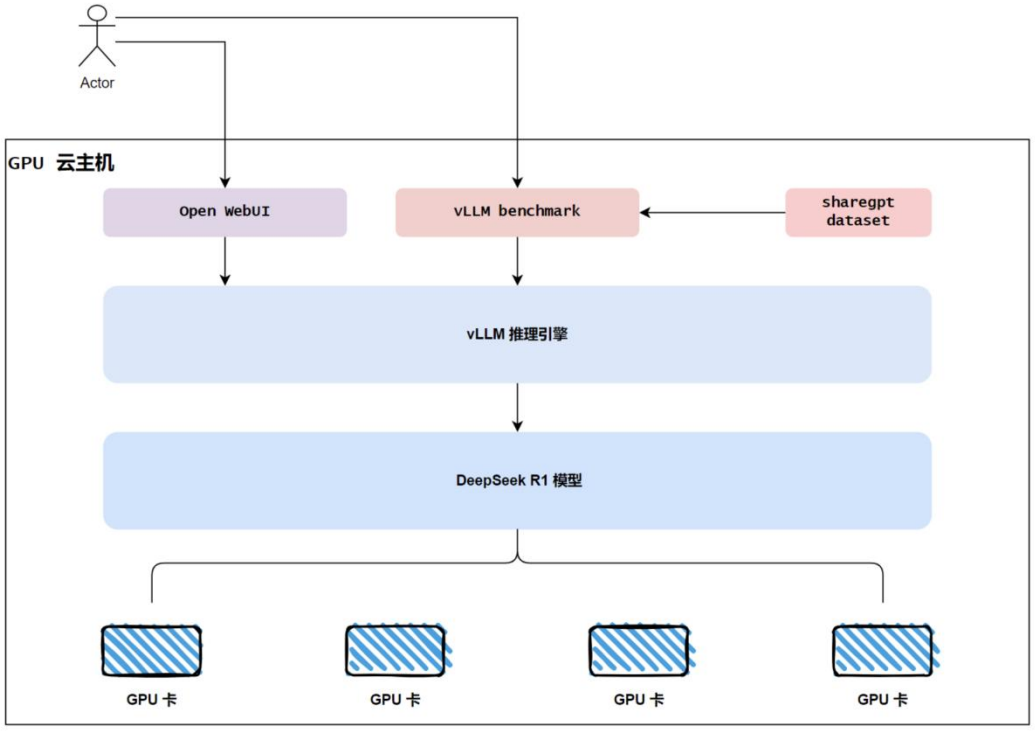
4 vLLM Benchmark

4.1 前言

vLLM Benchmark 是 vLLM 框架用于评估和优化其推理性能的标准化测试工具集，核心作用是为大语言模型（LLM）的推理服务提供量化指标，帮助开发者优化模型部署策略。

4.2 测试环境配置

如下图所示为测试环境拓扑，其中 GPU 云主机配置有 4 块 GPU 卡。部署 DeepSeek 的模型，安装 vLLM 推理引擎服务并启动；启动 Open WebUI 服务来接入到 vLLM，实现从浏览器客户端登录系统进行问答；vLLM Benchmark 用于进行性能压力测试，sharegpt dataset 数据集用于性能压测过程的输入数据。



可以通过如下命令来安装 vLLM Benchmark 工具脚本：

```
git clone https://github.com/vllm-project/vllm.git

cd vllm/benchmarks/
```

4.3 测试实践

使用 vLLM Benchmark 进行性能测试时可以指定 sharegpt 数据集。

Sharegpt 数据集的所有问题论来自于真实的用户提问，更贴近实际应用场景，可以全面、真实地评估模型在实际应用场景中的表现。具体的指标和测试方法见下文。

如下所示为性能压测命令：

```
python benchmark_serving.py --model
/root/model-path/deepseek-ai/DeepSeek-R1-Distill-Qwen-32B --backend
vllm --dataset-name sharegpt --dataset-path
/root/ShareGPT_V3_unfiltered_cleaned_split.json --max-concurrency 1
--num-prompts 50
```

命令参数解释如下：

- model：指定模型路径；
- backend：指定后端推理引擎类型；
- dataset-name：指定数据集类型；
- dataset-path：指定数据集路径；
- max-concurrency：指定并发数，默认为 1；
- num-prompts：指定总的请求数；

对应的测试结果如下所示：

```
============ Serving Benchmark Result ============
Successful requests:                     50
Benchmark duration (s):                  229.88
Total input tokens:                      11748
Total generated tokens:                  10162
Request throughput (req/s):              0.22
Output token throughput (tok/s):         44.21
Total Token throughput (tok/s):          95.31
---------------Time to First Token----------------
Mean TTFT (ms):                          95.08
Median TTFT (ms):                        64.17
P99 TTFT (ms):                           208.99
-----Time per Output Token (excl. 1st token)------
Mean TPOT (ms):                          22.27
Median TPOT (ms):                        22.26
P99 TPOT (ms):                           22.45
---------------Inter-token Latency----------------
Mean ITL (ms):                           22.26
Median ITL (ms):                         22.03
P99 ITL (ms):                            23.67
==================================================
```

指标说明：

测试结果中的具体指标含义如下所示：

● Output token throughput(tok/s）：输出 Token 吞吐量，单位为 tok/s，是衡量大模型服务生成效率的核心指标，指系统每秒能够生成的输出 token 数量。它直接反映模型的文本生成速度和系统的整体处理能力，是评估推理成本和服务商业价值的关键依据。

● TTFT：Time to First Token，首 Token 延迟，是衡量大模型服务响应速度的核心指标，具体指从用户发送请求到系统生成第一个输出 。

● TPOT：指排除首 Token 延迟后，生成每个后续输出 token 的平均耗时。计算公式为：总解码时间 / 输出 token 数量。例如，若 TPOT 为 50ms，则每秒可生成 20 个 token。

● ITL：连续两个输出 token 之间的间隔时间，单位毫秒，反映生成稳定性。

指定输入输出 token 长度

当使用的数据集为 random 类型的时候，执行 vLLM Benchmark 压测命令的时候可以指定输入长度和输出长度，测试命令如下：

```
python vllm/benchmarks/benchmark_serving.py \
  --model /root/model-path/deepseek-ai/DeepSeek-R1-Distill-Qwen-32B \
  --backend vllm \
  --dataset-name random \
  --random-input-len 256 \
  --random-output-len 256 \
  --max-concurrency 1 \
  --num-prompts 50
```

进行上述测试时如果使用 vllm 启动的时候添加了--served-model-name 参数以及--api-key 参数，测试的时候需要在上述命令中加入--served-model-name 内容，并且在环境变量中添加 OPENAI_API_KEY=yourkey。

5 三种测评工具对比

5.1 核心功能对比

1. vLLM Benchmark

- 专注于 LLM 推理引擎性能评测，支持 OpenAI 接口兼容的框架（如 Qwen 系列模型）
- 内置 TTFT（首 Token 延迟）、TPOT（后续 Token 时延）、吞吐量等核心指标
- 支持长上下文压力测试（如 LongAlpaca-12k 数据集）

2. EvalScope

- 大模型多维度评估框架，覆盖性能测试（Perf）、效果评测（Accuracy）和合规性检测
- 提供标准化测试集（如 MMLU、CMMLU）和自定义数据集扩展能力，支持多种测评场景
- 集成生产级压力测试功能，支持并发请求模拟与实时监控（通过 Wandb 可视化）

3. LLMPerf

- 开源基准测试工具，专攻 API 级性能评估（延迟、吞吐量、请求成功率）
- 支持正确性验证（如数字格式转换准确性测试）

5.2 性能指标侧重

工具	核心指标	测试场景差异
vLLM Benchmark	首 Token 延迟(TTFT)、Token 吞吐量、显存占用率	重推理引擎的硬件资源利用率，常用于 GPU 云主机环境下的引擎选型
EvalScope	综合吞吐量、QPS、端到端延迟、多模态任务准确率	适配生产环境压力测试，支持 API 服务与本地模型的混合评测
LLMPerf	Token 间延迟、首 Token 延迟（TTFT）、请求吞吐量、错误率	强调高并发模拟能力（如 10+并发请求），适合 API 服务的 SLA 验证

5.3 技术特性差异

1. 部署复杂度

- vLLM Benchmark：从 vLLM 的 github 代码仓库源克隆
- EvalScope：提供 Docker 镜像和 Python CLI 工具链，支持快速接入 ModelScope 社区模型和数据集
- LLMPerf：依赖 Ray 分布式框架，需修改 Tokenizer 配置适配私有化模型

2. 测试数据灵活性

- vLLM Benchmark：使用 ShareGPT_V3、LongAlpaca 等结构化数据集，也可使用随机数据自定义输入长度
- EvalScope：允许自定义输入长度（--dataset-input-length）和样本重复次数
- LLMPerf：支持动态调整输入 Token 分布参数（--mean-input-tokens）

3. 结果可视化

- vLLM Benchmark/EvalScope：支持 Wandb 面板实时监控测试进度
- LLMPerf：生成 CSV 格式原始数据，需二次处理

5.4 典型应用场景

1. vLLM Benchmark

- 推理引擎选型（如对比 vLLM、LMDeploy、TensorRT-LLM 的吞吐量）
- 长上下文支持能力验证（如测试 16k+Token 输入的显存溢出风险）

2. EvalScope

- 生产环境服务容量规划（如通过--parallel 模拟多路并发）
- 多模态模型效果-性能平衡分析（如 Embedding 模型响应延迟与召回率关联性）

3. LLMPerf

- 云 API 服务商性能对比（如 Anthropic Claude vs OpenAI GPT-4 Turbo）
- 模型部署前的 SLA 合规性检查（如验证 P99 延迟是否达标）

5.5 工具选型建议

- LLMPerf：功能比较单一，适合 API 基准测试以及快速对比多个 LLM API 性能场景，聚焦于 LLM API 性能（如延迟、吞吐量、请求成功率）和正确性测试（输出格式与内容验证）；

- EvalScope：功能全面且灵活，适合复杂评测的需求场景，支持多维度评估，包括模型能力、性能压测、端到端 RAG 评测、多模态模型评测等。适合综合评估模型在学术研究、工业部署中的表现，支持自定义评测指标和复杂场景（如竞技场模式、长文本生成测试）；

- vLLM Benchmark：主要针对 vLLM 推理引擎的性能优化（如 PagedAttention 技术效果、多卡扩展性）。专为优化 vLLM 框架的部署参数设计，适合需要调整批次大小、显存利用率等参数以实现高吞吐量推理的场景。

6 附录

vLLM 官方代码仓库

EvalScope 官方文档

LLMPerf 官方代码仓库

DeepSeek 提示库

10.38.161pen WebUI 及 vLLM 版本升级最佳实践

在使用 GPU 云主机进行推理时，vLLM 版本会影响计算性能与功能支持，Open WebUI 版本关系着用户交互体验与功能兼容性。本文提供完整升级指南，进一步保证系统稳定性、界面功能完整及性能最优。

一、Open WebUI 升级指南

1. 查看现有版本

```
docker exec -it open-webui bash -c "cat /app/package.json|grep version"
```

```
root@ecm-372c:~# docker exec -it open-webui bash -c "cat /app/package.json|grep version"
       "version": "0.5.10",
root@ecm-372c:~#
```

2.版本升级

（1）登录云主机，下载最新容器镜像

```
docker pull ghcr.nju.edu.cn/open-webui/open-webui:v0.6.5
```

```
[root@ecm-bf9f ~]# docker pull ghcr.nju.edu.cn/open-webui/open-webui:v0.6.5
v0.6.5: Pulling from open-webui/open-webui
8a628cdd7ccc: Pull complete
2a47a8c4fd5c: Pull complete
9c538fc35491: Pull complete
47bbb0afa7fe: Pull complete
782acb99e453: Pull complete
4f4fb700ef54: Pull complete
b7915144f9c7: Pull complete
633be3c6bab0: Pull complete
266a80d83771: Pull complete
7b4a3fa111d1: Pull complete
693caf783e3a: Downloading [============================>                      ]  637.9MB/1.043GB
ec0cdf90a948: Download complete
39845d3298e5: Download complete
55245328d5ea: Download complete
87ea629f5ffc: Download complete
```

（2）删除旧镜像

查看当前容器镜像

```
[root@ecm-bf9f ~]# docker image ls
REPOSITORY                                TAG       IMAGE ID       CREATED         SIZE
ghcr.nju.edu.cn/open-webui/open-webui     v0.6.5    ba71477c0bbc   2 weeks ago     4.59GB
ghcr.nju.edu.cn/open-webui/open-webui     cuda      ab45a42f3e76   2 months ago    8.78GB
```

删除旧容器镜像

```
docker image rm ghcr.nju.edu.cn/open-webui/open-webui:cuda
```

```
[root@ecm-bf9f ~]# docker image rm ghcr.nju.edu.cn/open-webui/open-webui:cuda
Untagged: ghcr.nju.edu.cn/open-webui/open-webui:cuda
Untagged: ghcr.nju.edu.cn/open-webui/open-webui@sha256:72e3a276586471d5188fb9277d012eeefb91e3e47e971cbc8d98396590dadf36
Deleted: sha256:ab45a42f3e76083fd3cd5f3a0def94bedb45d20fa512e22a83f5d51ca13993b7
Deleted: sha256:844554aafe16051a469258924b499c186e993e6b8bd1b7692fce7bbf3c88afc8
Deleted: sha256:ebd4eaaf516b8f2db2d5491c38538d19e29e0564260b3f288b223eee1904379b
Deleted: sha256:5b2392a345c11d990b8b7f4460db3f248b272866a287665db637eb61a2c638f8
Deleted: sha256:4f804b08c4c03b1c7cbdc79be7a2de648815f8c7a6e52f3eeb85b3eba7888471
Deleted: sha256:edcfa083c4cf051f6d0d5d3dd0c785347630004f6fe2922f53b12c3cbabc818e
Deleted: sha256:7b2e8194700d7d0f8629017ad7c06143a32d3e725d2eaf6f63f9f29262891e12
Deleted: sha256:75f11ef19a0ec655d8c77e829b76ba483ed563b65aba45d6532bb6840bc6ae42
Deleted: sha256:d3373f275461f0f2b452a32a80f9e1917f643559debd16e0dd1fd193a3f449a3
Deleted: sha256:57dab2b0c9806d399b06a4886d8f4655c264c90f3353a62714d8342b0fd91585
Deleted: sha256:6b4710dee645ddd2f834e405fc13b4d6eb98b78e7e04da8480e069893730f9a1
Deleted: sha256:29cd39632ff67290565056bc2b1d979c7c1cd6e23a6b87a15a09afe451a397f8
Deleted: sha256:9961f67f49aa0c506794b96672e88779ce6923adad2c47b9aeedcf63e94a6a17
Deleted: sha256:1a283ab5951ff9954c2898c76df19f044c50cddd29d7a5acc28fe78ac2f55e01
Deleted: sha256:60922dbb5c4ead2daa8c28c22529016875364b1830af541b07efd7a7fab2e05f
Deleted: sha256:e5cccb7a4f590bb889bbe78736dff1a7227ac4e75b3ba0dd9acb43e0860d8a6a
Deleted: sha256:7914c8f600f532b7adbd0b003888e3aa921687d62dbe2f1f829d0ab6234a158a
```

查看更新后的容器镜像

```
[root@ecm-bf9f ~]# docker image ls
REPOSITORY                                TAG       IMAGE ID       CREATED        SIZE
ghcr.nju.edu.cn/open-webui/open-webui   v0.6.5    ba71477c0bbc   2 weeks ago    4.59GB
```

3.修改启动 open-webui 命令参数

```
vi /root/.start_webui.sh
```

修改后内容如图

```
#!/bin/bash

API_KEY=`/usr/local/bin/generate_api_key.sh API_KEY`

docker run -d -e PORT=3000 --net=host -e ENABLE_SIGNUP=true -e ENABLE_OPENAI_API=True -e ENABLE_OLLAMA_API=False -e OPENAI_API_BASE_URL=http://localhost:8000/v1 -e OPENAI_API_KEY=${API_KEY} -v open-webui:/app/backend/data --name open-webui --restart always ghcr.nju.edu.cn/open-webui/open-webui:v0.6.5
```

4.清理容器

```
docker ps -a
```

```
docker rm -f open-webui
```

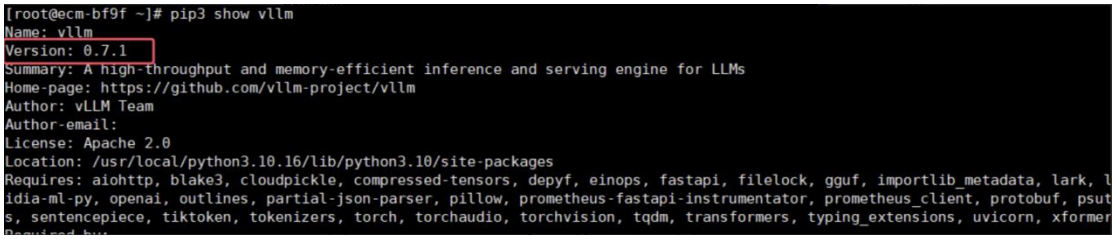
```
root@ecm-1691:~# docker ps -a
CONTAINER ID   IMAGE                                COMMAND          CREATED       STATUS                     PORTS     NAMES
7d5c99502102   ghcr.io/open-webui/open-webui:main   "bash start.sh"  4 weeks ago   Up 46 minutes (healthy)              open-
webui
root@ecm-1691:~# docker rm -f open-webui
```

二、vLLM 升级指南

1. 查看现有版本

```
#如果环境是 pip 安装的请执行

pip show vllm# 如果是 pip3 安装的请执行# pip3 show vllm
```



2.版本升级

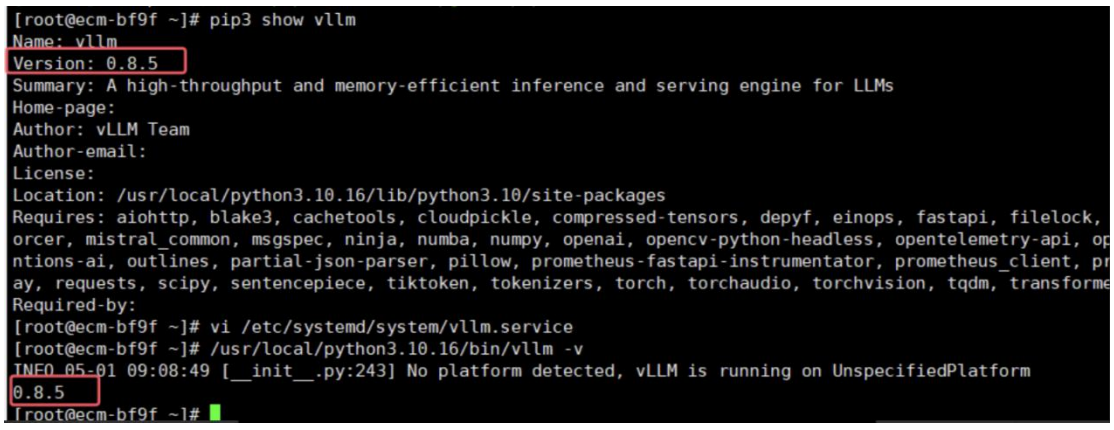
（1）指定版本号升级 vLLM

```
# 例如要升级 vllm 到 0.8.5# 如果是 pip 安装的执行：pip install vllm==0.
8.5# 如果是 pip3 安装的执行# pip3 install vllm==0.8.5
```

（2）查看是否升级成功

```
# 如果是 pip 安装的执行

pip show vllm# 如果是 pip3 安装的执行# pip3 show vllm
```



10.39 基于弹性云主机的视频处理应用加速实践

镜像描述

当前镜像处于公测阶段，您可在华东 1 资源池试用。如果您希望在其他资源池试用该镜像，可以提交工单申请。

Intel® iVTAL 是一种用于视频处理 加速的软件库，利用英特尔® 处理器中的硬件加速性能，以提供 快速而优化的视频处理解决方案。通过使用 Intel® iVTAL 库， 开发人员可以轻松地在其应用程序中集成视频处理功能。

视频处理专用镜像是一款集成 intel 视频处理调优工具 iVTAL 的专用的视频处理镜像，镜像内置视频抽帧工具 ivtalDecoder 和视频转码工具 FFmepg(intel 版)，支持常见的音频和视频格式。专用视频处理镜像专为提升视频转码/解码/编码任务的速度和效率而生，告别通用镜像的平庸性能。

配置要求

推荐规格：c7.4xlarge.2 或者 c8.4xlarge.2

推荐配置：vCPU≥16 核，内存≥32G

免责说明

免责声明：本公司发布的视频处理专用镜像，集成的转码和解码工具均来源自第三方或者开源社区，仅供用户参考和使用。本公司不对因不当使用该镜像而导致的任何损害承担法律责任，用户应当确保其使用行为符合相关法律法规，并遵守相应协议。

实践指南

视频转码工具（英特尔版）

视频转码工具介绍

转码是将视频码流转换成另一个视频码流的过程。使用转码功能可以实现将视频转换成流畅、标清、高清以及超清等输出，从而用户可以选择合适的视频播放。视频转码最常用的三个软件工具是 FFmpeg， x264， x265。在 FFmpeg(version release/4.4)， x264(version 0.164)， x265(version 3.6)这三个工具的基础上，intel 采用 MVReuse 算法策略实施了进一步的优化，能够在几乎不影响视频质量·压缩率的情况下，显著提升转码速度。

视频转码工具命令参数说明

运行一个最简单的视频转码所需要的参数如下：

```
-y 覆盖输出文件
```

-i 指定输入视频

-c:v 指定视频编码器，libx265 转码相比于 libx264 能提供更好的压缩率，但速度也会变慢

-crf 指定视频压缩力度，越小压缩率越低质量越好，越大压缩率越高质量越差

-mvreuse 2 表示开启 intel 优化，如果命令中没有-mvreuse 2 则默认不开启英特尔优化，此时转码效果与使用官方的 FFmpeg(version release/4.4)，x264(version 0.164)，x265(version 3.6)效果相同

-psnr 显示转码后的质量分数

指定输出视频文件

运行示例 1

```
#执行命令：

ffmpeg -y -i movie.mp4 -c:v libx264 -mvreuse 2 -crf 22   -psnr out.mp4


#命令解释：

ffmpeg              使用视频转码工具（英特尔版）

-y                   想要默认覆盖输出文件

-i movie.mp4      输入 1920x1080 24fps 的电影场景视频

-c:v libx264      使用 libx264 编码器

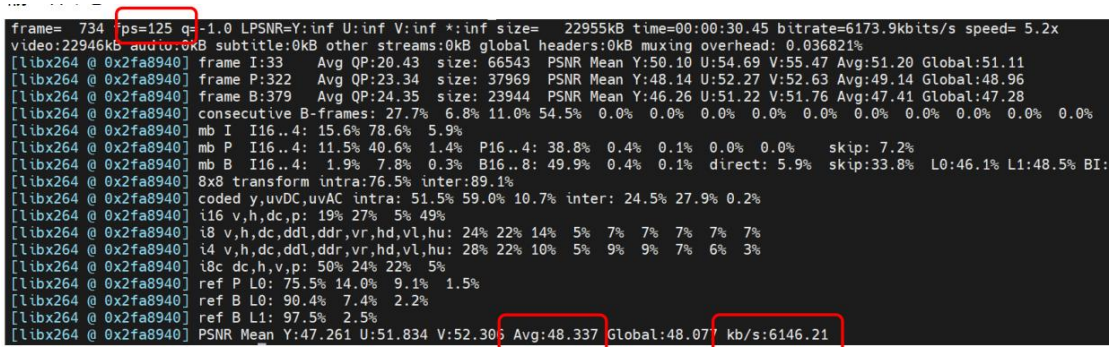
-mvreuse 2         开启 intel 优化

-crf 22              控制压缩力度为 22

-psnr                想看到转码后的质量分数

out.mp4              指定输出视频的名字
```

执行结果如图



#日志解释：

fps=125　　视频转码实时吐帧量，用来衡量视频转码速度指标

avg=48.337　　视频转码后质量分数，用来衡量视频转码质量指标

kb/s=6146.21　视频转码后平均码率，用来衡量视频转码压缩率指标

运行示例 2

#执行命令：

```
ffmpeg -y -i movie.mp4 -c:v libx265 -mvreuse 2 -crf 28   -psnr out.mp4
```

#命令解释：

ffmpeg　　使用视频转码工具（英特尔版）

-y　　想要默认覆盖输出文件

-i movie.mp4　　输入 1920x1080 24fps 的电影场景视频

-c:v libx265　　使用 libx265 编码器

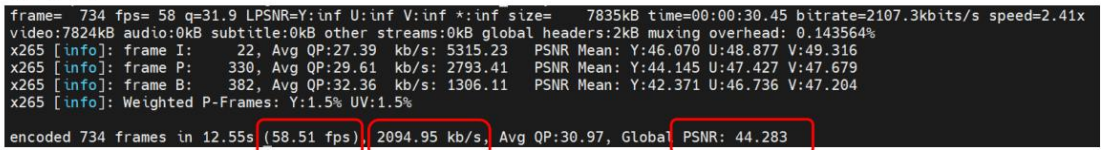
-mvreuse 2　　开启 intel 优化

-crf 28　　控制压缩力度为 28

-psnr　　想看到转码后的质量分数

out.mp4　　指定输出视频的名字

执行结果如图



#日志解释：

58.51fps　　视频转码实时吐帧量，用来衡量视频转码速度指标

PSNR:44.283　　视频转码后质量分数，用来衡量视频转码质量指标

2094.95kb/s　　视频转码后平均码率，用来衡量视频转码压缩率指标

视频抽帧工具（英特尔版）

视频抽帧工具介绍

视频抽帧可以从原始视频中提取关键帧，用于后续 AI 模型的内容分析、检索或建模。它还能显著降低数据量，加速视频处理流程，如智能分析、转码或特征提取等任务。视频抽帧最常用的软件工具是 FFmpeg。参考 FFmpeg(version release/6.1)，intel 实施了视频包预取和自适应线程优化，显著提升抽帧速度。

说明

Decoder 功能仅针对 6.1 版本的 ffmpeg 且仅针对每 400 帧抽取 1 帧的情况下有效率提升。

抽帧工具 ivtalDecoder 转码参数说明

-mode 0 -fi　　指定抽帧间隔

-thread　　指定线程数量

指定输入视频

指定输出抽帧后像素文件

运行示例 1

```
# 执行命令

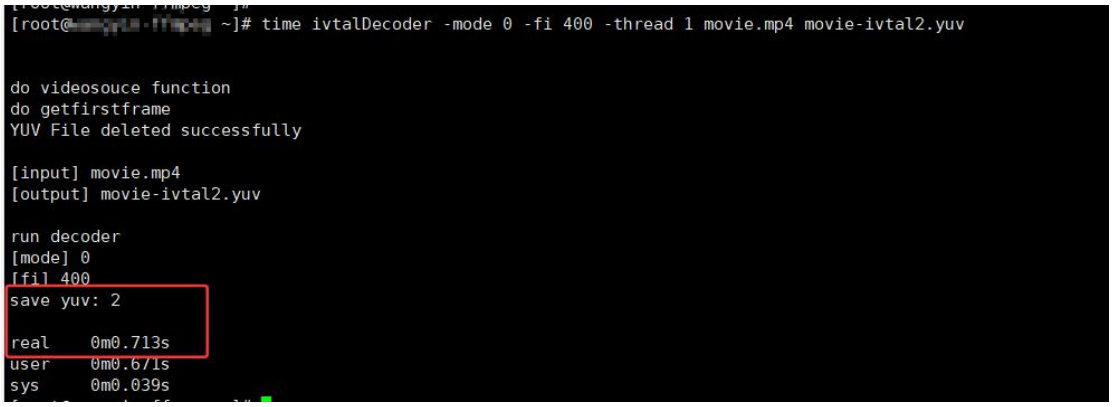
ivtalDecoder -mode 0 -fi 400 -thread 1 movie.mp4 movie-ivtal2.yuv



#命令解释：
```

ivtalDecoder	使用视频抽帧工具（英特尔版）
-mode 0 -fi 400	指定每隔 400 帧抽一帧
-thread 1	用一个线程
movie.mp4	输入 1920x1080 24fps 的电影场景视频
movie-ivtal2.yuv	指定输出像素文件的名字

执行结果如图



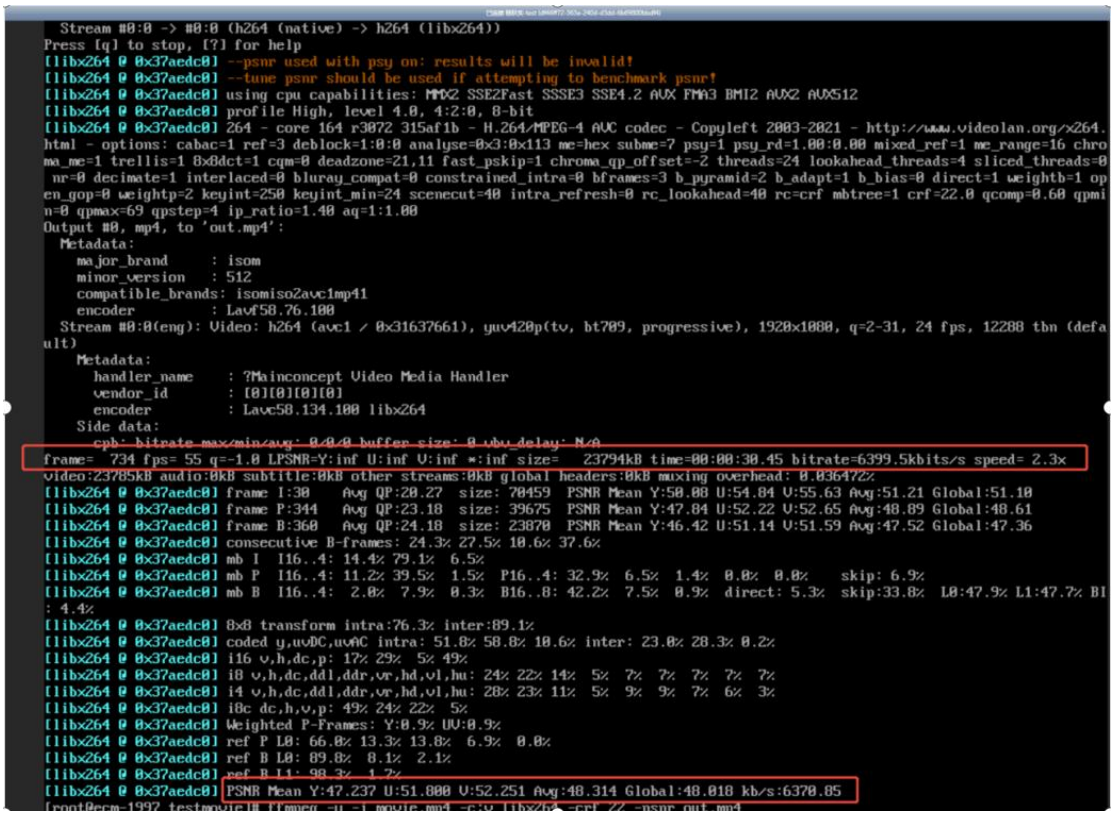
#日志解释：	
save yuv=2	一共抽出来多少帧
real=0.713	一共耗时多少

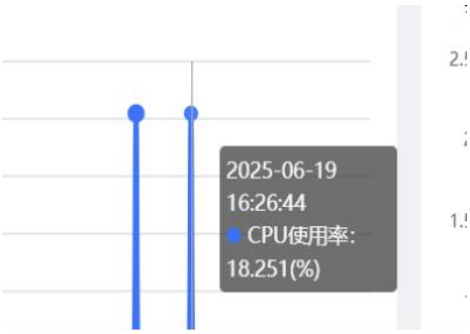
性能提升说明

下面是在规格 c8.4xlarge.2 云主机下的视频转码测试结果对比的数据。

- 优化前

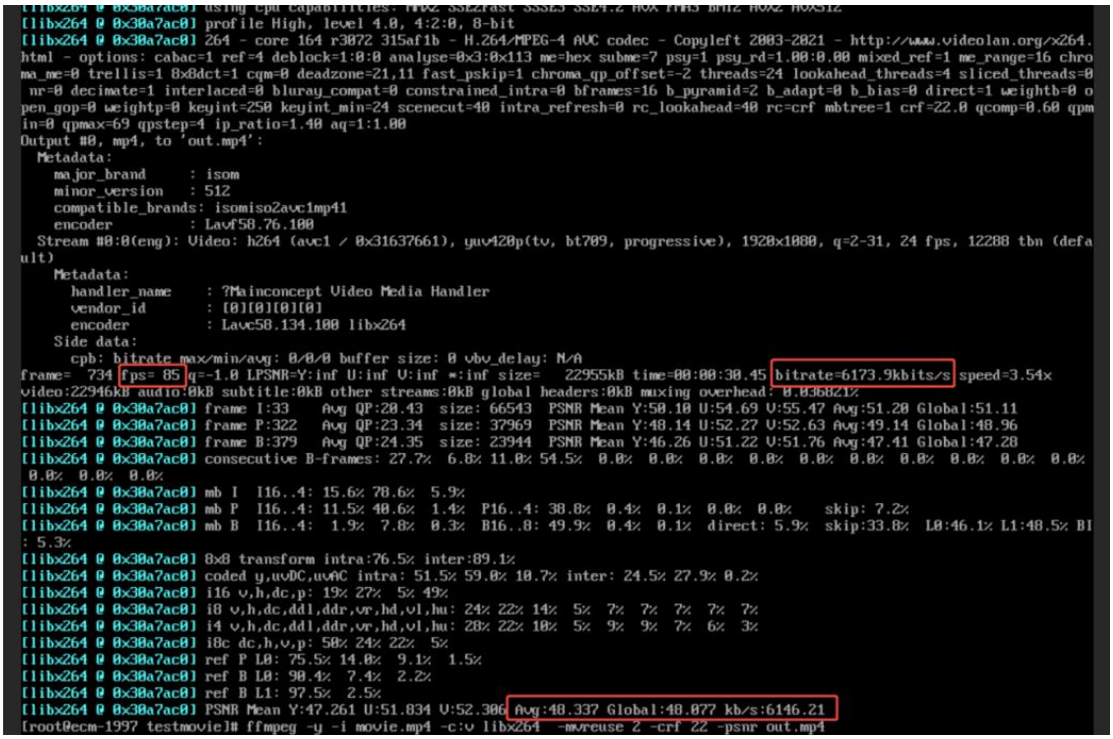
```
ffmpeg -i movie.mp4 -c:v libx264 -crf 22 -psnr out.mp4
```





- 优化后

```
ffmpeg -i movie.mp4 -c:v libx264 -mvreuse 2  -crf 22 -psnr out.mp4
```



过程中 cpu 利用率为 11%。



指标	优化前	优化后
PSNR（dB）	48.3	48.3
Bitrate（kbps）	6399.5	6173.9
FPS（帧/秒）	55	85

指标	优化前	优化后
CPU 利用率（%）	18%	11%

优化后相比优化前，cpu 利用率明显下降，视频转码 fps 提升 55%。