翼 MapReduce 服务

用户操作指南

天 翼 云 科 技 有 限 公 司

1 产品介绍

1.1 产品定义

产品定义

翼 MapReduce（简称："翼 MR"），是基于当前开源新版本大数据组件进行产品化封装，可以为客户提供快速部署、便捷维护的 HDFS、YARN、Spark、Flink、Hive、Doris、Kafka、HBase 等高性能的大数据组件以及运维管理平台，同时产品默认提供强安全验证能力，具备高安全、高扩展、快捷运维等特色，支持批量数据处理、流式数据处理、离线数据分析、在线查询等场景。

产品架构

翼 MR 集群各个版本组件情况请参见版本概述。

详见下图：翼 MR 架构图

翼 MR 架构包括了基础设施和大数据处理流程各个阶段的能力。

- **基础设施**

基于天翼云弹性云主机 CT-ECS 构建的大数据集群，整体集群的高可靠和高安全能力可以得到虚拟化底层的充分保证。

- 虚拟私有云（CT-VPC）为每个租户提供虚拟的内部网络，默认与其他网络隔离，同时通过配套的安全组访问控制确保网络层面的安全性。
- 云硬盘（CT-EVS）提供不同规格和性能表现的高可靠存储能力。

- 弹性云主机（CT-ECS）提供的弹性可扩展虚拟服务器，结合上述的 CT-VPC、安全组、CT-EVS 数据多副本和灾备能力为客户打造一个高效、可靠、安全的业务集群环境。
- 物理机服务（CT-DPS）是基于天翼云软硬结合技术研发的一款拥有极致性能的裸金属服务器，兼具云主机的灵活弹性、物理机的稳定，提供算力强劲的计算类服务，提供专属的云上物理服务器，为大数据、核心数据库、高性能计算等业务提供服务稳定、数据安全、性能卓越的算力服务。

● 数据集成

数据集成层提供了客户的数据集成进翼 MR 集群的能力，包括：Kafka、Logstash、SeaTunnel、Flume，支持各种数据源导入数据到翼 MR 大数据集群中。

● 数据存储

翼 MR 支持结构化和非结构化数据在集群中的存储，并且支持多种高效的格式来满 足不同计算引擎的要求。

- HDFS 是大数据上通用的分布式文件系统。
- Doris 是实时数据仓库服务，具有高并发、低延迟的特点。
- HBase 支持带索引的数据存储，适合高性能基于索引查询的场景。
- Elasticsearch 支持结构化/非结构化数据的检索、分析场景。

● 数据调度和计算处理

- 翼 MR 提供多种主流计算引擎：MapReduce（批处理）、 Spark（内存计算）、Flink（流计算），满足多种离线或实时大数据应用场景，将数据进行结构和逻辑的转换，转化成满足业务目标的数据模型。
- 基于预设的数据模型，使用易用 SQL 的数据分析，用户可以选择 Hive（数据仓库），SparkSQL 以及 Trino 交互式查询引擎。

● 翼 MR Manager

为确保大数据组件服务的高可用性，以 Hadoop 为基础的大数据生态的各种组件均需要以分布式的方式进行部署，涉及其中的部署、管理和运维复杂度要求较高。翼 MR 提供了统一的运维管理平台翼 MR Manager，包括可视化引导式部署集群能力。同时翼 MR Manager 还提供了租户与资源管理能力，以及翼 MR 中各类大数据组件的运维，并提供监控、告警、配置等一站式运维能力。

产品优势

性能优化

- 自研优化实现百亿行、百万列毫秒级的即席查询。

稳定可靠

- 完成对开源组件 100+次的代码及配置优化。

安全可控

- 使用 Kerberos+Ranger 安全技术实现全组件的认证和授权。
- 支持库、表、字段级数据权限管控。

便捷运维

- 全链路可视化操作降低运维门槛，助力实现 90%日常运维场景便捷操作，提升运维效率。

1.2 版本概述

产品发行版本号格式

翼 MapReduce 采用类似翼 MR-a.b.c 格式的版本号，详细说明如下：

- a 代表版本有较大的变动。
- b 代表版本中一些组件的变动。
- c 代表版本中 Bug 修复，可以向前兼容；以及一些较小的变动。

翼 MR 每个版本的捆绑软件和集群创建说明

- 捆绑软件：每个翼 MapReduce 发行版本上捆绑的软件和软件版本都是固定的。
- 集群创建：创建某一个发行版本的翼 MapReduce 集群后，该集群版本不会自动升级。

翼 MR-2.19.0

发布时间：2025 年 7 月

- 支持海光 h3 系列国产化云主机
- 支持八代机云主机
- 支持 XSSD 云硬盘
- 新增企业项目功能
- 升级 Hudi 版本至 0.15.0、升级 Iceberg 版本至 1.6.1、升级 Flink 版本至 1.18.1、升级 KafkaUI 版本至 1.2.0
- 支持 Spark 原生引擎与 Native 引擎切换

组件	版本

Doris	2.1.8
Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.8.2
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.4.1
Trino	440
YARN	3.3.3
ZooKeeper	3.7.1
Flink	1.18.1
Knox	1.6.1
Logstash	7.10.2
Iceberg	1.6.1
Hudi	0.15.0
JeekeFS	1.1.1
Tez	0.10.1
Flume	1.11.0
KafkaUI	1.2.0
SeaTunnel	2.3.7
Pushgateway	1.6.2

TezUI	0.10.1
Amoro	0.7.1
Hue	4.11.0

翼 MR-2.18.0

发布时间：2025 年 5 月

- 新增 Amoro 与 Hue 组件服务
- 升级 Doris 版本至 2.1.8
- 华南 2 资源池新增可用区 2 与可用区 3

组件	版本
Doris	2.1.8
Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.8.2
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.4.1
Trino	440
YARN	3.3.3
ZooKeeper	3.7.1
Flink	1.16.2
Knox	1.6.1

Logstash	7.10.2
Iceberg	1.4.3
Hudi	0.14.0
JeekeFS	1.1.1
Tez	0.10.1
Flume	1.11.0
KafkaUI	1.0.0
SeaTunnel	2.3.7
Pushgateway	1.6.2
TezUI	0.10.1
Amoro	0.7.1
Hue	4.11.0

翼 MR-2.17.0

发布时间：2025 年 3 月

- 支持物理机服务
- 支持麒麟系统
- 新增海光与鲲鹏国产云主机规格
- 数据湖与自定义业务场景新增 TezUI 组件服务
- Doris 组件版本升级至 2.1.6，Trino 升级至 440，Flume 升级至 1.11.0
- 新增 Master 扩容能力
- 新增重置集群主机密码能力
- 对接云日志服务
- 对接云审计服务
- 优化元数据库配置能力

组件	版本
Doris	2.1.6

Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.8.2
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.4.1
Trino	440
YARN	3.3.3
ZooKeeper	3.7.1
Flink	1.16.2
Knox	1.6.1
Logstash	7.10.2
Iceberg	1.4.3
Hudi	0.14.0
JeekeFS	1.1.1
Tez	0.10.1
Flume	1.11.0
KafkaUI	1.0.0
SeaTunnel	2.3.7
Pushgateway	1.6.2
TezUI	0.10.1

翼 MR-2.16.0

发布时间：2024 年 12 月

- 新增自定义业务场景
- 数据湖、实时数据流、自定义业务场景新增组件 KafkaUI（1.0.0）、SeaTunnel（2.3.7）、Pushgateway（1.6.2）；数据湖、自定义场景新增组件 Flume（1.10.0）、Tez（0.10.1）
- Spark 组件版本升级至 3.4.1、Hudi 版本升级至 0.14.0、Kyuubi 版本升级至 1.8.2、Doris 升级至 2.1.5
- 增加新增节点组功能
- 新增拓扑调整功能
- 新增 Hive 资源洞察功能
- 增加云监控服务能力

组件	版本
Doris	2.1.5
Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.8.2
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.4.1
Trino	406
YARN	3.3.3

ZooKeeper	3.7.1
Flink	1.16.2
Knox	1.6.1
Logstash	7.10.2
Iceberg	1.4.3
Hudi	0.14.0
JeekeFS	1.1.1
Tez	0.10.1
Flume	1.10.0
KafkaUI	1.0.0
SeaTunnel	2.3.7
Pushgateway	1.6.2

翼 MR-2.15.2

发布时间：2024 年 9 月

- 数据湖、数据服务、实时数据流场景新增组件 JeekeFS（1.1.1）
- Doris 组件版本升级至 2.1.2
- 新增节点扩容功能
- 新增磁盘扩容功能
- 新增集群服务管理功能
- 新增自动续费功能
- 新增 IPv6 访问资源能力
- 新增反亲和性配置能力

组件	版本
Doris	2.1.2
Elasticsearch	7.10.2
HBase	2.4.12

HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.6.0
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.3.3
Trino	406
YARN	3.3.3
ZooKeeper	3.7.1
Flink	1.16.2
Knox	1.6.1
Logstash	7.10.2
Iceberg	1.4.3
Hudi	0.12.2
JeekeFS	1.1.1

翼 MR-2.14.1

发布时间：2024 年 5 月

- 数据湖场景新增组件 Knox(1.6.1)、Hudi(0.12.2)、Iceberg(1.4.3)
- 数据服务场景新增组件 Knox(1.6.1)
- 云搜索场景新增组件 Logstash(7.10.2)
- 新增实时数据流业务场景

- Doris 组件版本升级至 2.0.4
- Spark 组件版本升级至 3.3.3
- 新增安全组配置通信授权功能
- 新增主机健康检查功能
- 新增上海 36 资源池

组件	版本
Doris	2.0.4
Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.6.0
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.3.3
Trino	406
YARN	3.3.3
ZooKeeper	3.7.1
Flink	1.16.2
Knox	1.6.1

组件	版本
Logstash	7.10.2
Iceberg	1.4.3
Hudi	0.12.2

翼 MR-2.12.2

发布时间：2024 年 5 月

- 数据湖场景新增组件 Flink(1.16.2)
- 新增华北 2 资源池

组件	版本
Doris	1.2.6
Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2
Kyuubi	1.6.0
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.2.2

组件	版本
Trino	406
YARN	3.3.3
ZooKeeper	3.7.1
Flink	1.16.2

翼 MR-2.12.1

发布时间：2024 年 3 月

- 上线 4 个 API 接口
- Bug 修复
- 新增西南 1 资源池（2024 年 4 月 1 日上线新资源池）

翼 MR-2.12.0

发布时间：2023 年 12 月

组件	版本
Doris	1.2.6
Elasticsearch	7.10.2
HBase	2.4.12
HDFS	3.3.3
Hive	3.1.2
Kafka	2.8.1
Kerberos	1.18.2
Kibana	7.10.2

组件	版本
Kyuubi	1.6.0
OpenLDAP	2.4.50
Ranger	2.2.0
Spark	3.2.2
Trino	406
YARN	3.3.3
ZooKeeper	3.7.1

1.3 产品优势

翼 MapReduce（简称翼 MR）服务，历经集团和云公司的大规模集群和业务打磨，提供全年多级用户 SLA 保障。

翼 MR 具有如下优势：

高安全

翼 MR 服务拥有企业级的大数据多租户权限管理能力，拥有企业级的大数据安全管理特性；使用 Kerberos 和 Ranger 安全技术实现全组件的认证和授权；支持库、表、字段级数据权限管控。

优开源

基于 Hadoop3.3.3 等开源最新组件，保证版本性能的同时，优化了底层资源占用，对任务和引擎进行定制化管控。

高可靠

完成对开源组件 100+次的代码及配置优化，实现高 SLA；支持节点组采用反亲和技术，虚拟机分布在不同物理机上，以保证服务高可用。

易运维

翼 MR 提供可视化的大数据集群运维管理平台，全链路可视化操作降低运维门槛，助力实现 90%日常运维场景便捷操作，提升运维效率。

低成本

翼 MR 集群基于多样化的云基础设施，提供了丰富的计算、存储设施的选择，可以用时再创建、用时再扩容，用完就销毁，确保成本最优。

1.4 功能特性

1.4.1 多租户

简介

现代企业的数据集群化正在逐步向中心化、云计算方向发展，多用户在多个集群上执行不同的应用（分析、查询、流式处理等）且存储的数据类型、格式各不相同。某些特定用户（如银行、政府机构等）对数据安全问题非常重视，无法容忍自己的数据和他人数据放在一起。

翼 MapReduce 提供多租户功能，将大数据集群资源分割为相互独立的资源集合，用户可“租用”所需的资源集合，用于执行应用、工作、存储数据。在大数据集群中，提供多个资源集合满足多个用户的不同需求。

优势

合理配置和隔离资源：

租户之间的资源是隔离的，一个租户对资源的使用不影响其他租户，保证了每个租户根据业务需求去配置相关的资源，可提高资源利用效率。

测量和统计资源消费：

系统资源以租户为单位进行计划和分配，租户是系统资源的申请者和消费者，对用户资源消费进行测量和统计。

保证数据安全和访问安全：

多租户场景下，分开存放不同租户的数据、控制用户对租户资源的访问权限，以保证数据安全和访问安全。

1.4.2 安全增强

天翼云大数据平台 翼 MapReduce 支持海量数据存储、海量数据分析、实时处理等行业应用，平台具备高安全性。该产品主要从以下几个方面保障用户可以安全地存储、使用数据以及运行业务。

网络隔离

整个系统部署在用户在公有云上专享的虚拟私有云中，提供安全隔离的网络环境，保证用户大数据集群的业务、管理的安全性。通过结合虚拟私有云的子网划分、路由控制、安全组等功能，可以为用户提供高安全、高可靠的网络隔离环境。

资源隔离

翼MR服务目前支持用户在天翼云4.0资源池内专属化部署，服务使用到的云主机在IaaS层保证资源隔离，资源组合上确保专属计算资源+专属存储资源。

主机安全

翼 MR 支持与天翼云官网安全服务集成，支持漏洞扫描、安全防护、应用防火墙、云堡垒机、网页防篡改等。针对云主机，提供如下安全措施：

- 操作系统内核安全加固
- 更新操作系统最新补丁
- 操作系统权限控制
- 操作系统端口管理
- 操作系统协议与端口防攻击
- 云主机监控
- 防 DDoS 攻击
- 定期备份数据
- 增加登录密码强度

应用安全

通过如下措施保证大数据业务正常运行：

- 身份鉴别与认证
- Web 应用安全

- 访问控制
- 审计安全
- 密码安全

数据安全

针对海量用户数据，提供如下措施保障客户数据的机密性、完整性和可用性。

- 容灾：翼 MR 产品当前数据存放在天翼云云硬盘产品中，云硬盘采用三副本冗余机制，保障上层翼 MR 服务数据的高容灾型。
- 备份：翼 MR 产品当前数据存放在天翼云云硬盘产品中，云硬盘具备定期备份、数据恢复等特性。

数据完整性

通过数据校验，保证数据在存储、传输过程中的数据完整性。

- 用户数据保存在 HDFS 上，HDFS 默认采用 CRC32C 校验数据的正确性。
- HDFS 的 DataNode 节点负责存储校验数据，如果发现客户端传递过来的数据有异常（不完整）就上报异常给客户端，让客户端重新写入数据。
- 客户端从 DataNode 读数据的时候会同步检查数据是否完整，如果发现数据不完整，尝试从其它的 DataNode 节点上读取数据。

数据保密性

天翼云大数据平台 翼 MapReduce 产品中的分布式文件系统采用 Apache Hadoop 3.3.3 版本，提供对文件内容的加密存储功能，避免敏感数据明文存储，提升数据安全性。业务应用只需对指定的敏感数据进行加密，加解密过程业务完全不感知。在文件系统数据加密基础上，Hive 实现表级加密，HBase 实现列族级加密，在创建表时指定采用的加密算法，即可实现对敏感数据的加密存储。

- 从数据的存储加密、访问控制来保障用户数据的保密性。
- HBase 支持将业务数据存储到 HDFS 前进行压缩处理，且用户可以配置 AES 和 SMS4 算法加密存储。
- 各组件支持本地数据目录访问权限设置，无权限用户禁止访问数据。
- 所有集群内部用户信息提供密文存储。

安全认证

基于用户和角色的认证统一体系，遵从帐户/角色 RBAC（Role-Based Access Control）模型，实现通过角色进行权限管理，对用户进行批量授权管理。

- 支持安全协议 Kerberos，翼 MR Manager 使用 OpenLDAP 作为帐户管理系统，并通过 Kerberos 对帐户信息进行安全认证。
- 对登录 Manager 的用户进行审计。

1.4.3 组件 WEB UI 便捷访问

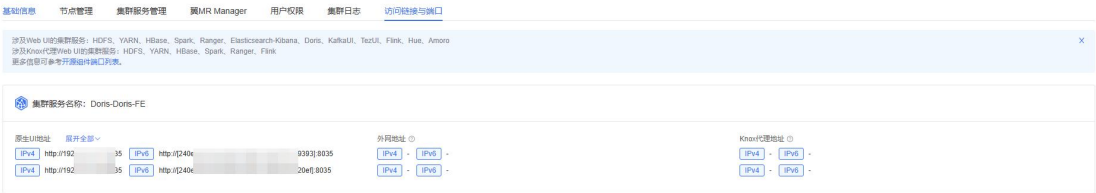
大数据组件都有自己的 WEB UI 页面管理自身系统，但是由于网络隔离的原因，用户并不能很简便地访问到该页面。如访问 HDFS 的 WEB UI 页面，传统的操作方法是需要用户创建 ECS，使用 ECS 远程登录组件的 UI，这使得组件的页面 UI 访问很是繁琐，对于很多初次接触大数据的用户很不友好。

翼 MR 提供了基于内网 IP 地址来便捷访问开源组件 UI。

操作步骤

1.登录翼 MR 控制台，点击正常运行的集群名称，进入集群详情页面。

2.点击“访问链接与端口”。



3.集群服务名称：具有 WEB UI 的集群服务有 HDFS、YARN、HBase、Spark、Ranger、Doris、Elasticsearch—Kibana 等，详情请参考开源组件 Web 站点。

4.原生 UI 地址：当集群部署了具备 WEB UI 的集群服务后，在“原生 UI 地址”列会默认显示相关集群服务的信息，原生 UI 地址展示的是“内网地址：端口号”的格式。

5.获取外网访问地址：外网地址展示的是“外网地址：端口号”的格式。要成功访问开源组件 UI，需要在安全组页面开启该端口并放开出、入方向规则，同时需要开通部署了该集群服务的节点的外网 IP，如何绑定外网 IP 请参考“绑定/解绑弹性 IP”。

6.访问 WEB UI 地址：成功开启安全组和外网 IP，在“外网地址”列会展示出可以访问的外网地址。通过访问外网地址就可以便捷访问开源组件 UI。

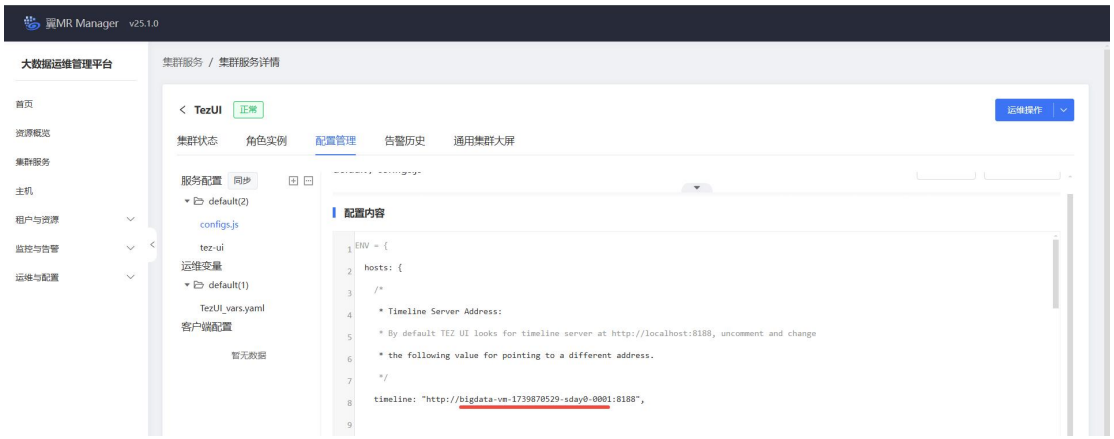
7. WebUI 登录：组件密码可前往 Manager—配置管理的 vars.yaml 高级配置中查询。LDAP 用户名与密码可前往 Manager—LDAP 租户管理中查询。

> ✧ **注意**
>
> 使用时需要先开通安全组端口才能访问；使用后需要及时关闭，避免安全泄漏风险。安全组访问规则配置方法可参加“添加安全组规则”。

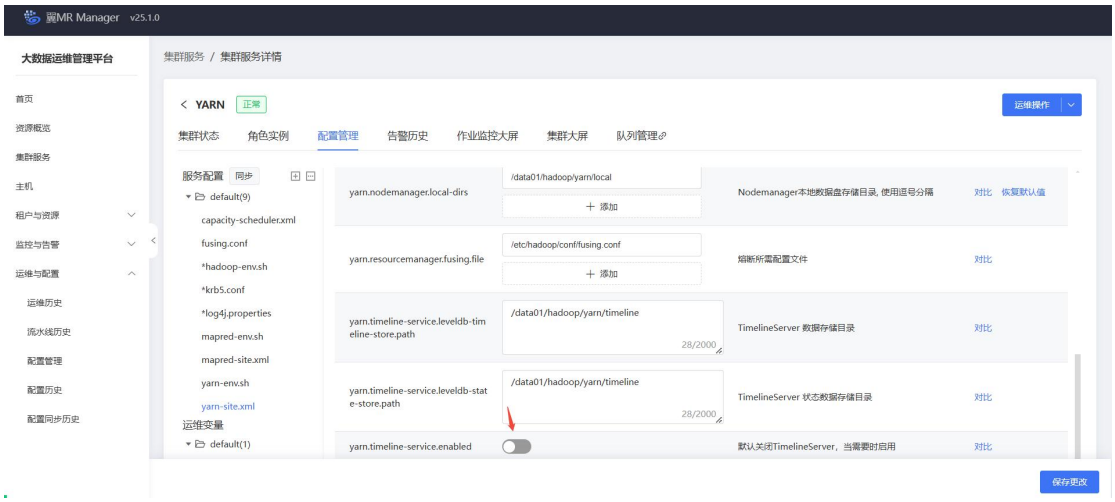
说明：

访问 TezUI 站点时，请参考下述步骤更新配置。

1、前往 Manager，在 TezUI 的配置管理中，将 configs.js 文件中的 timeline 地址更换为外网 IP。即下图中划红线部分替换为控制台—节点管理中 TezUIServer 所在节点的外网 IP。修改后保存并同步配置，同步成功后重启 TezUIServer 实例。



2、前往 YARN 的配置管理页面，在 yarn—site.xml 配置中开启 yarn.timeline—service.enabled 开关。保存更改并同步配置后，重启 YARN 与 Hive 服务。



1.4.4 可靠性增强

天翼云大数据平台 翼 MapReduce 是天翼云在 Apache Hadoop 基础上推出的一站式开源大数据平台产品，具备大数据基础存储计算平台和大数据运维管理平台，主要在大数据组件的可靠性、性能调优等方面进行了优化和提升。

系统可靠性

管理节点均实现 HA

Hadoop 开源版本的数据、计算节点已经是按照分布式系统进行设计的，单节点故障不影响系统整体运行；而以集中模式运作的管理节点可能出现的单点故障，就成为整个系统可靠性的短板。

天翼云大数据平台 翼 MapReduce 对所有组件的管理节点都提供了类似的双机的机制，包括 HDFS NameNode、HiveServer2、HBase HMaster、YARN ResourceManager、KerberosServer、OpenLDAPServer 等，全部采用主备或负荷分担配置，有效避免了单点故障场景对系统可靠性的影响。

异常场景下的可靠性保证

通过可靠性分析方法，梳理软件、硬件异常场景下的处理措施，提升系统的可靠性。

- 保障意外掉电时的数据可靠性，不论是单节点意外掉电，还是整个集群意外断电，恢复供电后系统能够正常恢复业务，除非硬盘介质损坏，否则关键数据不会丢失。
- 硬盘亚健康检测和故障处理，对业务不造成实际影响。
- 自动处理文件系统的故障，自动恢复受影响的业务。
- 自动处理进程和节点的故障，自动恢复受影响的业务。
- 自动处理网络故障，自动恢复受影响的业务。

节点可靠性

操作系统健康状态监控

周期采集操作系统硬件资源使用率数据，包括 CPU、内存、硬盘、网络等资源的使用率状态。

进程健康状态监控

翼 MR 提供业务实例的状态以及业务实例进程的健康指标的检查，能够让用户第一时间感知进程健康状态。

硬盘故障的自动处理

天翼云大数据平台 翼 MapReduce 对开源版本进行了增强，可以监控各节点上的硬盘以及文件系统状态。如果出现异常，立即将相关分区移出存储池；如果硬盘恢复正常（通常是因为用户更换了新硬盘），也会将新硬盘重新加入业务运作。这样极大简化了维护人员的工作，更换故障硬盘可以在线完成；同时用户可以设置热备盘，从而极大缩减了故障硬盘的修复时间，有利于提高系统的可靠性。

节点磁盘 LVM 配置

天翼云大数据平台 翼 MapReduce 支持将多个磁盘配置成 LVM（Logic Volume Management），多个磁盘规划成一个逻辑卷组。配置成 LVM 可以避免各磁盘间使用不均的问题，保持各个磁盘间均匀使用在 HDFS 和 Kafka 等能够利用多磁盘能力的组件上尤其重要。并且 LVM 可以支持磁盘扩容时不需要重新挂载，避免了业务中断。

数据可靠性

天翼云大数据平台 翼 MapReduce 可利用弹性云服务器 ECS 提供的反亲和节点组以及放置组的能力，结合 Hadoop 的机架感知能力，将数据冗余到多个物理宿主机上，避免物理硬件的失效造成数据的失效。

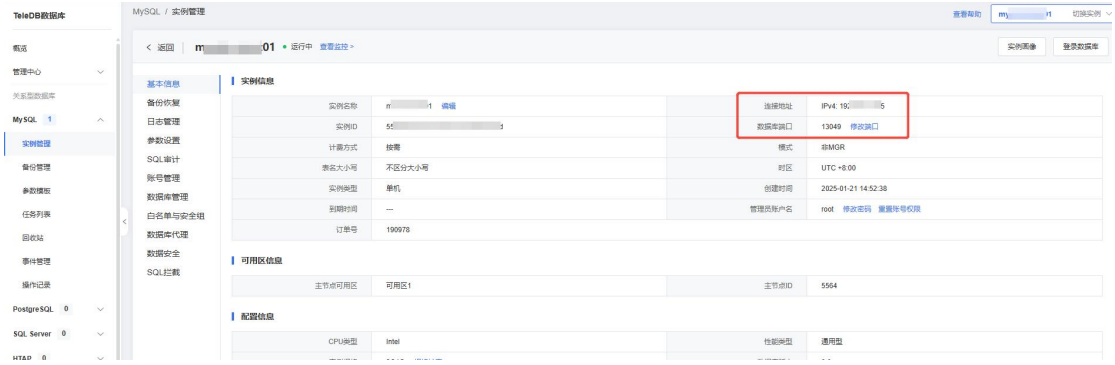
1.4.5 元数据库配置

选择在翼 MR 集群部署 Hive、Ranger、Amoro 或 Hue 集群服务时，翼 MR 提供关系数据库 MySQL 版(CT-RDS MySQL)的元数据存储方式。请选择关联与当前集群同一虚拟私有云(VPC)下的 MySQL 数据库，元数据将存储于关联的数据库中，不会随当前集群的删除而删除，多个翼 MR 集群可共享同一份元数据。

配置信息说明

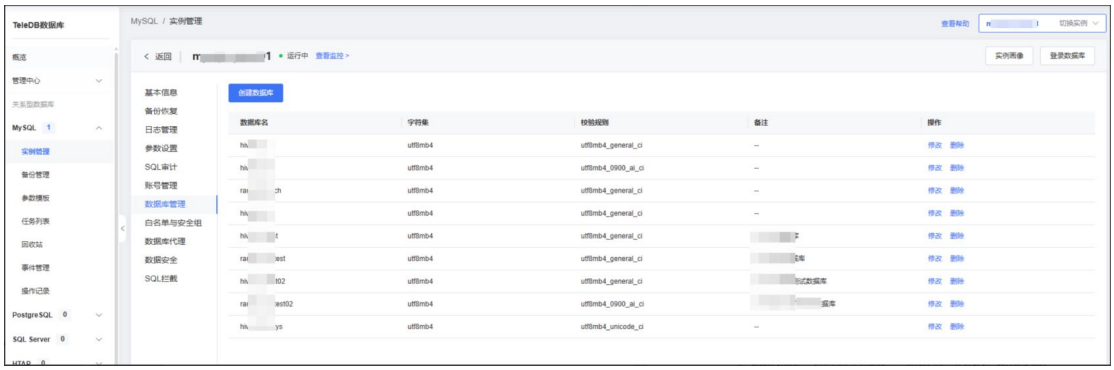
当您选择部署 Hive、Ranger、Amoro 或 Hue 集群服务时，需要填写元数据库有关的五项配置，包括数据库主机、数据库端口、数据库名称、数据库用户名、数据库密码。

- 数据库主机与端口：开通 MySQL 实例后，可从基本信息中获取主机与端口号信息。



- 数据库名称：开通 MySQL 实例后，可在数据库管理页面创建数据库并设置数据库名称。

建库时，字符集建议选择 utf8mb4，校验规则推荐选择 utf8mb4_unicode_ci。



- 数据库用户名与密码：开通实例后，可前往账号管理创建账户，并为该账号授予目标数据库的权限。请确保填写的用户拥有该数据库的读写权限。

说明：

1、当前仅支持通过内网地址进行数据库连接。请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库。

2、配置元数据库信息时，请提前创建数据库，若需要部署多个服务，请创建多个数据库，创建方式可参考创建数据库。

3、元数据配置所需信息可前往数据库控制台查看。

4、请确保配置的数据库用户具有该数据库的读写权限。

5、创建集群时，若元数据库配置错误，将导致 Hive、Ranger、Amoro 或 Hue 异常，但不影响集群的创建与部署，集群创建后，您可前往 Manager 的集群服务，进入相应的集群服务详情，通过运维操作中的“元数据库配置”操作，替换原有元数据库的配置信息并进行初始化。

1.4.6 集群管理

创建集群

支持用户在创建集群时，按需选择集群类型、组件范围、操作系统、主机类型、CPU 类型、各类型的节点数、主机规格、可用区、VPC 网络、认证信息。翼 MR 服务会根据用户选择的配置，帮助客户自动完成企业级大数据平台的安装部署和参数调优。

翼 MR 服务为客户提供完全可控的大数据集群，客户在创建时可设置主机的登录方式，所创建的翼 MR 集群资源完全归客户所用。

翼 MR 集群类型包括数据湖、数据分析、数据服务、云搜索、实时数据流、自定义集群

1. 数据湖集群：提供更高效、灵活的管理集群，更快地运行大数据的计算引擎，更好地提供数据分析能力。
2. 数据分析集群：Apache Doris—开源的 MPP 架构的 OLAP 分析引擎，支持亚秒级的数据查询和多表 join。
3. 数据服务集群：提供更灵活、可靠、高效的数据服务集群。
4. 云搜索集群：为结构化/非结构化数据提供低成本、高性能及可靠性的检索、分析服务能力。
5. 实时数据流集群：提供高效的流式计算、消息队列等组件能力，支持实时数据 ETL 和日志采集分析的业务需求。
6. 自定义集群：提供丰富灵活的服务搭配，支持自行选择业务所需服务。建议不要将多个存储类服务部署在一个节点组上，避免资源争抢。

翼 MR 集群节点类型包括 master 节点、core 节点和 task 节点

1. master 节点：集群中的管理节点，保证集群的调度正常进行；主要部署 NameNode、ResourceManager、HMaster 等进程。集群默认为 HA 模式，master 节点数默认为 3。该类型节点可以通过节点扩容、配置升级与磁盘扩容，以支持更大集群的管理。

2. core 节点：集群中的计算及存储节点，主要部署 DataNode、NodeManager、HRegionServer 等进程。为满足存储数据量或计算量扩展的需求，支持配置升级、节点扩容、新增节点组与磁盘扩容。
3. task 节点：集群中的纯计算节点，主要负责计算数据，不存储数据，支持配置升级、节点扩容、新增节点组与磁盘扩容。

配置升级

当 master、core 或 task 节点实例的规格（vCPU 和内存）无法满足您的业务需求时，您可以使用配置升级功能提升实例规格。

节点扩容

当 master、core 或 task 节点组内的资源无法满足您的业务需求时，您可以使用节点扩容功能增加实例数量。

新增节点组

当存量 core 或 task 节点组的计算或存储资源无法满足您的业务需求时，您可以使用新增节点组功能增加实例组。

磁盘扩容

当 master、core 或 task 节点的数据存储空间无法满足您的业务需求时，您可以使用磁盘扩容功能增加数据盘的空间。

集群运维管理

翼 MR Manager 提供资源概览、集群服务、主机、租户与资源、监控与告警、运维与配置等运维管理功能。

1. 资源概览：展示该集群下所有主机的 CPU、内存、网络等信息，包括 CPU 使用率、磁盘使用率、内存使用率、网络发送速率等。
2. 集群服务：展示当前集群下的所有集群服务，并按组件类型、以列表视图列出，在集群服务列表处支持一键启动所有集群服务、一键停止所有集群服务。
3. 主机：默认展示当前集群下的所有主机列表，可查看当前运维平台的所有主机信息。也可以查看主机上的角色实例分配和告警历史信息。
4. 租户与资源：以集群服务为维度对 LDAP 用户、Kerberos 安全凭证和 YARN 队列进行管理。LDAP 用户管理展示当前集群下的 LDAP 用户和用户组等信息；Kerberos 安全凭证支持新建 Principal、删除 Principal，支持 Keytab 分发与下载，并支持查看 Keytab 的分发记录；YARN 队列管理支持 YARN 队列新建、编辑与删除，

capacity-scheduler.xml 的全局属性配置，支持 YARN 队列的同步生效并支持查看同步生效记录。

5. 监控与告警：支持指标查询和告警历史功能。指标查询支持查询角色实例级、主机级的监控指标，支持指标结果的绘图操作，让用户更直观获取监控项变化；告警历史支持按照集群服务级、角色实例级、主机级查询告警内容。

6. 运维与配置：支持流水线历史、配置管理、配置历史、配置同步历史。流水线历史展示所有流水线的运行历史记录，以及操作人；配置管理支持查看不同集群服务的配置文件，并进行新增、修改、删除配置等操作；配置历史支持查看配置文件不同版本的配置内容、并支持不同版本之前的内容对比；配置同步历史支持查看不同环境的配置同步历史、配置同步操作人，以及配置同步详情。

1.5 应用场景

大数据在人们的生活中无处不在，在金融、交通、互联网、医疗、能源和政府部门等行业均可以使用翼 MR 服务进行大数据处理。

批量数据处理场景

HDFS 集群负责存储海量日志数据。

YARN 集群负责调度离线平台上运行的所有任务。

Hive、Spark、Trino 等主流计算框架从数据加工、数据挖掘到数据分析，快速获取数据洞察力。

分析后的数据回写进 HDFS 集群，为上层数据可视化等产品提供数据支撑。

离线数据分析场景

将海量数据通过导入或者外表等形式引入到 OLAP 分析引擎里，例如，Trino 提供高效、实时和灵活的数据分析能力。

满足用户画像、人群圈选、位置服务、BI 报表和业务分析等一系列的业务场景。

流式数据处理场景

基于 Flink 流式计算框架，对各类业务日志或者消息等实时数据进行分析处理。

相应分析结果同步进 HDFS 集群存储服务中。

在线查询场景

基于 Web 和移动应用程序等生成的 PB 级别的结构化、半结构化或非结构化数据进行在线分析。

方便客户的 Web 应用或者数据可视化产品获取分析结果进行实时展示。

1.6 术语解释

节点

翼 MR 集群中每个节点即为一台云服务器，节点类型及节点功能如下所示。

节点类型	功能
master 节点	翼 MR 集群管理节点，用于管理集群服务，主要负责 ResourceManager 和 NameNode 等控制进程的部署。master 节点组默认采用反亲和技术，以此保证业务高可用性。
core 节点	翼 MR 集群工作节点，主要负责存储和计算数据。
task 节点	翼 MR 集群计算节点，主要负责计算数据，不存储数据（如 HDFS 数据）。默认不开启，按需使用。

Doris

Apache Doris 是一个基于 MPP 架构的高性能、实时的分析型数据库，以极速易用的特点被人们所熟知，仅需亚秒级响应时间即可返回海量数据下的查询结果，不仅可以支持高并发的点查询场景，也能支持高吞吐的复杂分析场景。基于此，Apache Doris 能够较好的满足报表分析、即席查询、统一数仓构建、数据湖联邦查询加速等使用场景，用户可以在此之上构建用户行为分析、AB 实验平台、日志检索分析、用户画像分析、订单分析等应用。

Elasticsearch

Elasticsearch 是一个开源的、高扩展性的分布式全文检索引擎，能够近乎实时地存储、检索数据。它能集中存储您的数据，提供快速搜索、精细调整的相关性和强大的分析能力。

HBase

HBase 提供业务键值数据的结构化存储与检索能力，主要包括键值数据存储、键值数据查询功能，提供键值数据管理和键值数据库监测功能。兼容社区 HBase 接口，提供 Java API，Restful 接口形式。是一个分布式、数据多版本、面向列的 NoSQL 数据库。提供可弹性扩展的多维表格键值存储和即席查询能力。支持上亿行、可扩展列，具备强一致性、高扩展、高可用的特性。

HDFS

HDFS（Hadoop Distributed File System）是 Hadoop 生态系统的一个重要组成部分，是 Hadoop 中的的存储组件。它是一个分布式文件系统，提供对应用程序数据的高吞吐量访问。

Hive

Hive 数据仓库软件通过 SQL 实现对分布式存储中的大型数据集的读写和管理。Hive 提供命令行工具和 JDBC 驱动程序连接用户。Hive 对 SQL 语句编译和解析，生成相应的 MapReduce 任务对数据进行操作。

Kafka

Apache Kafka 是一个优秀的分布式事件流平台，被广泛用于高性能数据管道、流分析、数据集成和任务关键型应用程序中。

Kerberos

Hadoop 使用 Kerberos 作为用户和服务的强身份验证和身份传播的基础。Kerberos 是一种计算机网络认证协议，它允许某实体在非安全网络环境下通信，向另一个实体以一种安全的方式证明自己的身份。Kerberos 是第三方认证机制，其中用户和服务依赖于第三方（Kerberos 服务器）来对彼此进行身份验证。

Kibana

Kibana 是一个开源的数据分析和可视化平台，它被设计用于与 Elasticsearch 协同工作。您可以使用 Kibana 对 Elasticsearch 索引中的数据进行搜索、查看和交互操作。

Kyuubi

Kyuubi 是一个提供 JDBC/ODBC SQL 查询能力的分布式 SQL 引擎管理者，主要是为 Spark Thrift Server 提供多租户以及 HA 能力，同时为其他引擎（例如 Flink 或 Trino 等）提供 SQL 等查询服务。

OpenLDAP

OpenLDAP 是轻型目录访问协议（Lightweight Directory Access Protocol，LDAP），通过 IP 协议提供访问控制和维护分布式信息的目录信息。

Ranger

Ranger 为各组件提供了基于 PBAC（Policy-Based Access Control）的权限管理插件，用于替换组件自身原本的鉴权插件。

Spark

Spark 是一个离线分布式大数据处理引擎，可基于 Spark-SQL 表达语句、Spark API 开发程序、SQL JDBC/ODBC 开发程序、beeline 等方式提交 SQL 作业。Spark 能够部署在各种集群环境，快速的自动实现错误恢复机制，对各种规模大小的数据进行快速计算。

Trino

Trino 是定位在数据仓库和数据分析业务的分布式大数据 SQL 计算引擎，用于查询分布在一个或多个异构数据源上的大型数据集。Trino 是一个存算分离式的计算引擎，数据均存储在远程数据源上，Trino 通过自带的 connector 访问远程数据源进行查询。

YARN

Apache YARN（Yet Another Resource Negotiator）是 Hadoop 集群资源管理器系统，YARN 从 Hadoop 2 引入，最初是为了改善 MapReduce 的实现，但是它具有通用性，同样执行其他分布式计算模式。

ZooKeeper

ZooKeeper 是一个开源的分布式协调服务，主要用于数据订阅/发布，集群管理，配置管理，分布式锁。

Flink

Flink 是大数据的流式计算框架和分布式处理引擎，用于在无边界和有边界数据流上进行有状态的计算。Flink 能在所在常见集群环境中运行，并能以内存速度和任意规模进行计算，支持批处理和流处理数据，有着高速度和高吞吐的特点，是真正的流批处理计算框架，适用于的应用场景：实时监控系统、日志分析系统、推荐系统。

Hudi

Hudi 是数据湖的文件组织层，对 Parquet 格式文件进行管理提供数据湖能力，支持多种计算引擎，提供 IUD 接口，在 HDFS 的数据集上提供了插入更新和增量拉取的流原语。

Iceberg

Iceberg 是一个用户分析的高性能表格式，Iceberg 表格的数据可以存储在 HDFS/S3 等不同存储上，Iceberg 使得能够在大数据系统中更好的使用 SQL 做数据分析，同时可以让多引擎查询同时支持 Iceberg 表格，Iceberg 支持引擎批量/增量的消费，能够降低端到端的耗时。

Knox

Apache Knox Gateway 是一个应用程序网关，用于与 Apache Hadoop 部署的 REST API 和 UI 进行交互。Knox 网关为与 Apache Hadoop 集群的所有 REST 和 HTTP 交互提供了一个单一的访问点。

Logstash

Logstash 是一个流行的开源数据收集引擎，用于从各种来源收集、处理和转发数据。它可以从多种来源（如业务日志文件、消息队列、数据库等）收集数据，对数据进行解析、过滤和转换，最终将处理后的数据输出到目标位置（Opensearch、Elasticsearch、Hadoop、S3 等）。

JeekeFS

JeekeFS 是一个高性能分布式文件系统，用于存储和管理文件与数据。它采用数据与元数据分离的存储架构，数据本身被持久化在对象存储中，元数据则可以按需存储在多种数据库中。

Tez

Tez 组件是 Apache Hadoop 生态系统中的一个计算框架，它利用 DAG（有向无环图）来优化作业执行。通过提供可编程的输入、输出、处理器、任务等组件，Tez 能够高效地执行复杂的数据处理任务，提升 Hadoop 作业的性能和灵活性。

Flume

Flume 组件是一个分布式、可靠且高可用的日志采集、聚合和传输系统，主要用于收集、缓存和传输大量的日志数据到集中存储器中，如 HDFS、HBase 等。它通过 source、channel 和 sink 三个核心组件协同工作，实现数据的高效流动和处理。

KafkaUI

KafkaUI 用于监控和管理 Apache Kafka。它提供直观、轻量级的仪表板，帮助用户追踪 Kafka 集群的关键指标，如 Brokers、Topics、Partitions 等，并支持动态配置主题和多种认证方式。

SeaTunnel

SeaTunnel 组件是一个高性能、分布式的数据集成工具，支持实时数据流处理和离线批处理。它通过插件化的架构设计，可以方便地从多种数据源提取数据，进行清洗、转换后，加载到目标存储系统中，提供了丰富的数据转换功能和任务调度、监控能力。

Pushgateway

Pushgateway 组件是 Prometheus 监控系统的一个功能组件，它接收应用程序推送的监控指标数据，并保存这些数据供 Prometheus 抓取，适用于 Prometheus 无法直接拉取数据的场景。

TezUI

TezUI 是 Apache Tez 的可视化工具，用于监控和分析 Hadoop 数据处理任务的执行过程，展示任务 DAG 结构、资源消耗及进度，帮助用户优化性能、排查错误，提升大数据计算效率。

Amoro

Amoro 是数据湖存储引擎和湖仓管理系统。它支持多种数据源、存储格式及文件系统，能无缝集成到大数据生态系统，提供元数据管理、查询优化、ACID 事务支持等功能，用于高效存储、管理和分析数据。

Hue

Hue 是一款大数据分析 Web 工具，提供可视化界面，简化 Hadoop 生态组件的操作。用户可通过浏览器直接编写并执行查询。降低非技术人员使用大数据平台的门槛，提升数据分析与运维效率。

1.7 与其他产品的关系

与其他产品的关系

服务名称	翼 MR 与其他产品的关系	主要交互功能
虚拟私有云（VPC）	翼 MR 集群创建在虚拟私有云（VPC）的子网内，VPC 通过逻辑方式进行网络隔离，为用户的翼 MR 集群提供安全、隔离的网络环境。	创建虚拟私有云和子网。
弹性云主机（ECS）	翼 MR 服务使用弹性云主机作为集群的节点，每个弹性云主机是集群中的一个节点。	准备运行环境创建集群。
关系数据库 MySQL 版	关系数据库 MySQL 版（RDS）用于存储翼 MR 集群元数据。	配置数据连接。
统一身份认证服务（IAM）	统一身份认证服务（IAM）为翼 MR 提供了鉴权功能。	创建 IAM 子用户并同步用户信息至控制台和 LDAP 用户。
云监控服务	云监控服务支持对 V2.16.0 及以上版本的翼 MR 集群进行主机与组件状态的监控和告警。	创建翼 MR 集群后，可前往云监控服务控制台查看监控与配置告警规则。
云日志服务	云日志服务支持对 V2.17.0 及以上版本的翼 MR 集群进行组件服务的日志监控与管理。	创建翼 MR 集群后，可从翼 MR 控制台内授权开通该集群的云日志服务功能，并前往云日志服务控制台查看日志详情。
云审计	云审计产品支持对 V2.17.0 及以上版本的翼 MR 集群进行操作记录的收集、存储与查询功能，可用于支撑合规审计、安全分析和问题定位等常见应用场景。	创建翼 MR 集群后，可前往云审计服务控制台查看操作记录。

1.8 权限管理

如果您需要将已经创建的翼 MapReduce 服务资源，给企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以结合使用统一身份认证服务（Identity and Access Management，简称 IAM）、角色管理和用户权限进行精细的权限管理。IAM 服务提供用户身份认证和翼 MR 用户与角色的管理权限；角色管理服务提供不同角色的集群菜单权限分配能

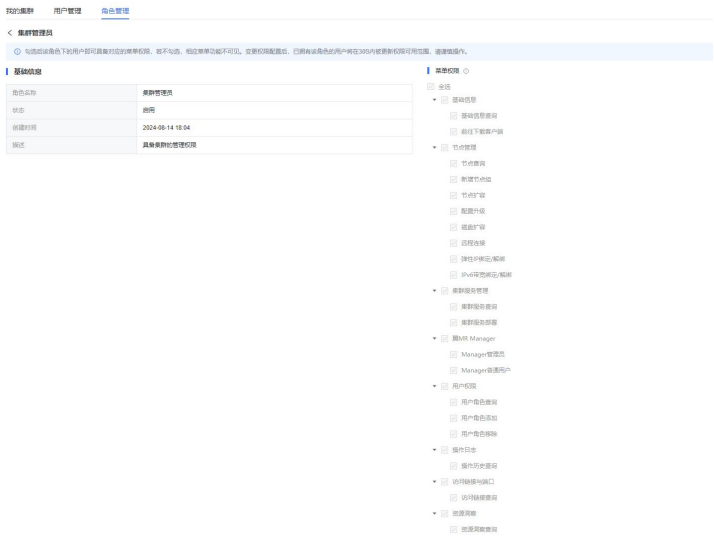
力；用户权限服务赋予不同的 IAM 用户不同的角色，提供集群访问控制，这些可以帮助您安全的控制资源的访问。

如果账号已经能满足您的要求，不需要创建独立的 IAM 用户进行权限管理，您可以跳过本章节，不影响您使用翼 MR 服务的其它功能。

翼 MapReduce 权限说明

默认情况下，管理员创建的 IAM 用户没有任何权限，需要将其授予实际的角色，才能使得 IAM 用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。您可以通过 IAM 与翼 MR 控制台内的角色管理功能为用户进行授权。

- 授予翼 MR 使用权限：您可以通过“账号中心”–“统一身份认证服务”功能为用户授予“翼 MapReduce 管理员”或“翼 MapReduce 用户”权限，并为用户分配企业项目。被授予“翼 MapReduce 管理员”身份的用户可以在翼 MR 控制台内进行 IAM 用户同步、新建角色、编辑角色、删除角色、删除集群、查看所在企业项目的集群等操作；被授予“翼 MapReduce 用户”身份的用户可以在翼 MR 集群列表中查看到所在企业项目的集群。
- 授予集群使用权限：您可以通过翼 MR 控制台内的“角色管理”功能，创建所需角色并为该角色授予集群的使用权限。设定角色后，集群管理员可通过集群内的“用户权限”功能，为 IAM 用户授予创建好的角色，支持被授权用户在该集群内行使相应权限。
  - 角色管理：通过“翼 MR 控制台”–“角色管理”进入页面，如下图所示，您可以按需赋予不同角色以不同的菜单功能，创建的角色可在全部集群中被授予目标用户。



  - 用户权限：在“翼 MR 控制台”–“我的集群”页面选择目标集群，点击“用户权限”进入页面，如下图所示，集群管理员或有“用户角色添加”权限的用户，即可为目标用户授予所选角色，被授权的用户在当前集群内可行使角色相应的权限。

1.9 约束与限制

在使用翼 MapReduce 前，您需要认真阅读并了解以下使用规则。

1. 集群必须创建在 VPC 子网内。
2. 创建集群时，从下拉框中选择已有的安全组。集群创建完成后，请勿随意删除或更改已使用的安全组，否则可能导致集群异常，影响集群的使用。
   a. 集群使用的安全组请勿随意放开权限，避免被恶意访问。
   b. 创建安全组时，出、入方向规则中的端口需要放开，授权策略不能选择“拒绝”，否则会导致不能部署集群和拉起服务。
3. 部署 Hive/Ranger/Hue/Amoro 组件时，元数据配置的数据库与集群需要在同一个 VPC 下，并通过内网地址进行连接。
4. 请根据业务需要规划集群节点的磁盘，如果需要存储大量业务数据，请增加云硬盘数量或存储空间，以防止存储空间不足影响节点正常运行。
5. 集群节点仅用于运行翼 MapReduce 集群，其他客户端应用程序、用户业务程序建议申请独立弹性云服务器部署。

1.10 安全

责任共担

为明确翼 MapReduce 产品在云服务环境下的安全责任边界，保障客户与天翼云协同构建安全可信的服务体系，本文将介绍天翼云与客户在安全性方面各自应承担的责任。

天翼云安全责任

天翼云负责保障云服务自身安全。责任包括：

- 基于天翼云综合安全产品体系，保障翼 MapReduce 产品内部使用的计算、存储、网络等基础设施，以及运行其上的平台服务与应用服务的安全性。
- 基于天翼云账号认证体系，为客户提供云上账户安全管理能力，包括但不限于支持主子账号、分组授权、细粒度授权等账户安全管控手段。

- 基于天翼云的监控与日志管理等能力，为客户提供翼 MapReduce 产品控制台操作及组件服务指标的安全监控能力。
- 未经客户授权，不接触和处理客户的业务数据。

客户安全责任

客户负责云服务内部的安全，正确配置和使用天翼云提供的产品能力和服务，责任包括：

- 负责数据的备份、加密，并对翼 MapReduce 产品中使用的所有密钥进行妥善管理。
- 负责访问权限控制，在账户设置中遵循最小权限原则，保护天翼云账户认证凭证。
- 遵循天翼云提供的安全原则和建议，对翼 MapReduce 产品中所使用的服务器、虚拟私有云、安全组、网络等基础设施进行合理配置，避免因配置不当而引发的安全风险与问题，承担数据安全主体责任。
- 管理安装在翼 MapReducce 集群上的组件服务，对其进行合理配置与使用，避免因不当配置引发的安全风险与问题。
- 不使用天翼云产品和服务从事非法的数据处理活动。

身份认证与访问控制

身份认证

翼 MR 支持安全协议 Kerberos，通过同步 IAM 账号使用 LDAP 作为帐户管理系统，并通过 Kerberos 服务对帐户信息进行安全认证。

Kerberos 安全认证原理和认证机制具体介绍请参见安全认证原理和认证机制。

访问控制

翼 MR 提供基于角色的权限控制模型。

基于角色的权限控制

翼 MR 基于用户和角色的认证统一体系，遵从帐户/角色 RBAC（Role-Based Access Control）模型，实现通过角色进行权限管理，对用户进行批量授权管理，同时提供单点登录能力，统一了系统用户和组件用户的管理及认证。

1.11 技术支持范围

天翼云翼 MapReduce（简称翼 MR）是一款基于云计算平台的数据处理分析服务，帮助客户大数据快速上云，轻松运行 Doris、Elasticsearch、Hive、Kafka、Spark 等高性能大数据组件。

翼 MR 提供用户完全可控的半托管云服务，集群资源归属于用户，用户拥有对集群的完全控制权，集群日常运维管理由用户负责，如在集群运维过程中遇到相关技术问题，可联系技术支持团队获得帮助，该技术支持仅协助分析处理翼 MR 云服务相关求助，不包含云服务以外的求助，例如用户基于大数据平台构建的应用系统等。为了更好地支持用户使用翼 MR 产品，现将翼 MR 产品技术支持服务范围进行说明。

支持的服务

- 翼 MR 管控平台提供的相关功能：
  - 集群的创建与退订销毁流程
  - 集群扩容流程
  - 集群配置变更流程
  - 集群新增服务流程
  - 集群监控与告警功能

协助支持的服务

- 协助排查翼 MR 产品开源组件的缺陷或需求，将根据产品规划迭代排期解决。
- 协助排查翼 MR 产品所依赖的其他天翼云基础产品的缺陷或需求，流转问题至相应产品团队负责解决。依赖的其他天翼云产品包括：
  - 弹性云主机、物理机等计算产品
  - 云硬盘、本地盘、对象存储等存储产品
  - 虚拟私有云、弹性 IP、安全组等网络产品
  - 云日志、云监控、云审计、IAM 等运维管理产品
  - 关系型数据库等用于元数据存储产品

不支持的服务

- 翼 MR 提供服务配置管理、节点管理、服务监控等丰富的运维工具，并保证功能的可用性和可靠性，但不提供具体集群和组件管理的运维操作，包括容量规划、组件性能优化、参数配置修改以及集群上具体的运维操作等。

- 不负责对基于翼 MR 组件之上的客户业务应用开发问题的答疑和处理，例如业务设计、代码开发、作业性能调优和业务迁移等。
- 在集群服务无明显异常或明确产品质量缺陷的情况下，不负责单个作业的异常问题排查。
- 不负责非产品标准能力的支持，例如集群或组件升级、Core 节点缩容和磁盘缩容等。
- 不负责对用户在翼 MR 集群上自建安装的第三方组件的问题排查和解决。
- 不负责因用户进行非产品预期操作，导致集群不稳定甚至集群不可用问题的处理。如重装操作系统、误删除数据、删除服务目录和文件、修改 OS 系统配置和文件权限、删除“/etc/hosts”配置、自定义脚本错误、任务写满磁盘、资源使用不合理、直接后台卸载磁盘、修改节点 IP 地址、删除创建集群时的默认安全组规则等。

支持方式

如需获得相关技术支持，可联系智能客服获取信息。

2 计费说明

2.1 资源节点

目前一类资源节点已在**华东 1、西南 1、华北 2、上海 36、华南 2、武汉 41、呼和浩特 3、南昌 5、杭州 7** 资源池开通，更多资源池规划部署中，因用户权限不同，实际可选资源池请以控制台实际可见区域为准。

您可以在华东 1 资源池中选择不同的可用区去使用：可用区 1、可用区 2、可用区 3。

您可以在西南 1 资源池中选择不同的可用区去使用：可用区 1、可用区 2。

您可以在华北 2 资源池中选择不同的可用区去使用：可用区 1、可用区 2、可用区 3。

您可以在上海 36 资源池中选择可用区 1 去使用。

您可以在华南 2 资源池中选择不同的可用区去使用：可用区 1、可用区 2、可用区 3。

您可以在武汉 41 资源池中选择可用区 1 去使用。

您可以在呼和浩特 3 资源池中选择可用区 1 去使用。

您可以在南昌 5 资源池中选择可用区 1 去使用。

您可以在杭州 7 资源池中选择可用区 1 去使用。

2.2 产品规格

云主机

通用云主机

S7 通用型

规格名称	核数（vCPU）	内存(GB)
s7.2xlarge.4	8	32
s7.4xlarge.2	16	32
s7.4xlarge.4	16	64
s7.8xlarge.2	32	64
s7.8xlarge.4	32	128

C7 通用计算增强型

规格名称	核数（vCPU）	内存(GB)
c7.2xlarge.4	8	32
c7.3xlarge.2	12	24
c7.3xlarge.4	12	48
c7.4xlarge.2	16	32
c7.4xlarge.4	16	64
c7.6xlarge.2	24	48
c7.6xlarge.4	24	96
c7.8xlarge.2	32	64
c7.8xlarge.4	32	128
c7.12xlarge.2	48	96
c7.12xlarge.4	48	192
c7.16xlarge.2	64	128
c7.16xlarge.4	64	256
c7.24xlarge.2	96	192
c7.24xlarge.4	96	384

M7 内存优化型

规格名称	核数（vCPU）	内存(GB)
m7.2xlarge.8	8	64
m7.3xlarge.8	12	96
m7.4xlarge.8	16	128
m7.6xlarge.8	24	192
m7.8xlarge.8	32	256
m7.12xlarge.8	48	384
m7.16xlarge.8	64	512
m7.24xlarge.8	96	768

S8 通用型

规格名称	核数（vCPU）	内存(GB)
s8.2xlarge.4	8	32
s8.3xlarge.4	12	48
s8.4xlarge.2	16	32
s8.4xlarge.4	16	64
s8.6xlarge.2	24	48
s8.6xlarge.4	24	96
s8.8xlarge.2	32	64
s8.8xlarge.4	32	128
s8.16xlarge.2	64	128

规格名称	核数（vCPU）	内存(GB)
s8.16xlarge.4	64	256
s8.24xlarge.2	96	192
s8.24xlarge.4	96	384

C8 通用计算增强型

规格名称	核数（vCPU）	内存(GB)
c8.2xlarge.4	8	32
c8.3xlarge.4	12	48
c8.4xlarge.2	16	32
c8.4xlarge.4	16	64
c8.6xlarge.2	24	48
c8.6xlarge.4	24	96
c8.8xlarge.2	32	64
c8.8xlarge.4	32	128
c8.12xlarge.2	48	96
c8.12xlarge.4	48	192
c8.16xlarge.2	64	128
c8.16xlarge.4	64	256
c8.24xlarge.2	96	192
c8.24xlarge.4	96	384
c8.32xlarge.2	128	256
c8.32xlarge.4	128	512

M8 内存优化型

规格名称	核数（vCPU）	内存(GB)
m8.2xlarge.8	8	64
m8.3xlarge.8	12	96
m8.4xlarge.8	16	128
m8.6xlarge.8	24	192
m8.8xlarge.8	32	256
m8.12xlarge.8	48	384
m8.16xlarge.8	64	512
m8.24xlarge.8	96	768
m8.32xlarge.8	128	1024

国产云主机

海光 hs1 通用型

规格名称	核数（vCPU）	内存(GB)
hs1.2xlarge.4	8	32
hs1.4xlarge.2	16	32
hs1.4xlarge.4	16	64

海光 hc1 计算增强型

规格名称	核数（vCPU）	内存(GB)
hc1.2xlarge.4	8	32
hc1.4xlarge.2	16	32
hc1.4xlarge.4	16	64
hc1.8xlarge.2	32	64
hc1.8xlarge.4	32	128
hc1.16xlarge.2	64	128
hc1.16xlarge.4	64	256

海光 hm1 内存优化型

规格名称	核数（vCPU）	内存(GB)
hm1.2xlarge.8	8	64
hm1.4xlarge.8	16	128

海光 hc3 计算增强型

规格名称	核数（vCPU）	内存(GB)
hc3.2xlarge.4	8	32
hc3.4xlarge.2	16	32
hc3.4xlarge.4	16	64
hc3.8xlarge.2	32	64
hc3.8xlarge.4	32	128
hc3.16xlarge.2	64	128
hc3.16xlarge.4	64	256

海光 hm3 内存优化型

规格名称	核数（vCPU）	内存(GB)
hm3.2xlarge.8	8	64
hm3.4xlarge.8	16	128
hm3.8xlarge.8	32	256

鲲鹏通用型

规格名称	核数（vCPU）	内存(GB)
ks1.2xlarge.4	8	32
ks1.4xlarge.2	16	32
ks1.4xlarge.4	16	64

鲲鹏计算增强型

规格名称	核数（vCPU）	内存(GB)
kc1.2xlarge.4	8	32
kc1.4xlarge.2	16	32
kc1.4xlarge.4	16	64
kc1.8xlarge.2	32	64
kc1.8xlarge.4	32	128
kc1.16xlarge.2	64	128
kc1.16xlarge.4	64	256

鲲鹏内存优化型

规格名称	核数（vCPU）	内存(GB)
km1.2xlarge.8	8	64
km1.4xlarge.8	16	128
km1.8xlarge.8	32	256

物理机

通用物理机

规格名称	核数（vCPU）	内存（GB）	本地盘（系统盘）	本地盘（数据盘）
physical.s5.2xlarge1	2路28核56线程	512	—	—
physical.s5.2xlarge4	2路28核56线程	512	2*480GB（SSD）	2 *3200GB（NVMeSSD）
physical.s5.2xlarge5	2路32核64线程	512	2*480GB（SSD）	2 *3200GB（NVMeSSD）
physical.s5se.xlarge1	2路32核64线程	256	—	—
physical.s5.xlarge3	2路32核64线程	256	—	—
physical.s5.2xlarge24	2路26核52线程	384	—	—
physical.s5.2xlarge2.2	2路28核56线程	1024	—	—
physical.s5.2xlarge19	2路20核40线程	256	—	—
physical.s5.2xlarge6	2路20核40线程	256	2*480GB（SSD）	4*1800GB（SAS）
physical.s5se.2xlarge1.2	2路32核64线程	512	2*960GB（SSD）	4*1920GB（SSD）
physical.s5se.2xlarge1.1	2路32核64线程	512	2*960GB（SSD）	12*12000GB（SATA）

云硬盘

云硬盘的产品规格包括：普通 IO（SATA）、高 IO（SAS）、超高 IO（SSD）、通用型 SSD、XSSD—0、XSSD—1、XSSD—2。

2.3 计费模式

计费项

购买翼 MapReduce 集群的费用包含两个部分：

- 翼 MapReduce 服务管理费用
- IaaS 基础设施资源费用（云主机、物理机、云硬盘）

计费方式

翼 MapReduce 当前支持包年包月的计费方式。

- 包年/包月：根据集群购买时长，一次性支付集群费用。最短时长为 1 个月，实际可订购时长以页面显示为准。

到期

包年/包月集群到期后进入保留期，此时无法在翼 MapReduce 管理控制台进行该集群的操作，相关接口也无法调用，自动化监控或告警等运维也会停止。如果在保留期结束时您没有续费，集群将终止服务，系统中的数据也将被永久删除。

2.4 计费价格

翼 MapReduce 集群费用由翼 MapReduce 服务管理费用和硬件产品费用组成，硬件产品费用由主机费用与硬盘费用组成。

云主机价格

系列	型号	产品	包月价格（元/月）
通用云主机	通用型 s7	vCPU（核）	45.69
		内存（GB）	16.31
	通用计算增强型 c7	vCPU（核）	73
		内存（GB）	14
	内存优化型 m7	vCPU（核）	58.43
		内存（GB）	14
	通用型 s8	vCPU（核）	46
		内存（GB）	17

	通用计算增强型 c8	vCPU（核）	74
		内存（GB）	14
	内存优化型 m8	vCPU（核）	60
		内存（GB）	14
国产云主机鲲鹏	通用型 ks1	vCPU（核）	52
		内存（GB）	15
	计算增强型 kc1	vCPU（核）	104
		内存（GB）	15
	内存优化型 km1	vCPU（核）	104
		内存（GB）	15
国产云主机海光	通用型 hs1	vCPU（核）	45
		内存（GB）	16
	计算增强型 hc1	vCPU（核）	89
		内存（GB）	14
	内存优化型 hm1	vCPU（核）	89
		内存（GB）	14

	计算增强型 hc3	vCPU（核）	89
		内存（GB）	14
	内存优化型 hm3	vCPU（核）	89
		内存（GB）	14

备注：1 年 8.5 折，2 年 7 折，3 年 5 折。

物理机价格

产品规格		包月价格（元/月）
CPU（核）		232
内存（GB）		5
存储	SATA（GB）	0.02
	SAS（GB）	0.04
	SSD（GB）	1

云硬盘价格

产品规格	包月（元/GB）
普通 IO（SATA）	0.3
高 IO（SAS）	0.4
超高 IO（SSD）	1.2
通用型 SSD	0.7
XSSD-0	0.5
XSSD-1	1
XSSD-2	2

备注：1 年 8.5 折，2 年 7 折，3 年 5 折。

翼 MR 包周期服务管理价格

1、通用云主机、国产云主机（鲲鹏、海光 h1 型）、物理机：服务管理费用=主机费用*0.12。

2、国产云主机（海光 h3 型）：服务管理费用=主机费用*0.2。

2.5 购买

操作步骤

1. 注册天翼云官网账号，登录后进入官网首页。
2. 在官网首页，单击左上角“产品”，【大数据>数据计算与分析>翼 MapReduce】。
3. 在【翼 MapReduce】界面，单击【立即开通】。

4. 在翼 MapReduce 服务集群创建页面根据实际需求选择集群类型、节点规格、购买时长等后，勾选相关协议，单击【立即购买】，详情请参考快速创建集群。

5. 按照页面提示，进入支付页面完成付款。

2.6 变更

规格变更

在开通翼 MapReduce 前有多种节点供您选择，您可根据业务需要选择合适的 master、core 和 task 节点实例。当集群启动后，翼 MapReduce 提供如下变更配置的方式。

配置升级：当 master、core 或 task 节点 ECS 实例的规格（vCPU 和内存）无法满足您的业务需求时，您可以使用配置升级功能提升 ECS 实例规格。请参见“配置升级”说明。注：物理机不支持配置升级，请在订购环节规划所需资源。

节点扩容：当 master、core 或 task 节点组内的资源无法满足您的业务需求时，您可以使用节点扩容功能增加实例数量。请参见“节点扩容”说明。

新增节点组：当存量 core 或 task 节点组的计算或存储资源无法满足您的业务需求时，您可以使用新增节点组功能增加 ECS 实例。请参见“新增节点组”说明。

磁盘扩容：当 master、core 或 task 节点的数据存储空间无法满足您的业务需求时，您可以使用磁盘扩容功能增加数据盘的空间。请参见“磁盘扩容”说明。本地磁盘不支持此功能，仅支持对云硬盘进行扩容。

2.7 续订

包年包月续订

手动续订

方法一：

1. 在天翼云官网，点击右上角的“管理中心”。

2. 左侧导航栏点击“订单管理”—“续订管理”，找到需要续订的产品进行操作，详细操作可参考费用中心—手动续订 。

方法二：

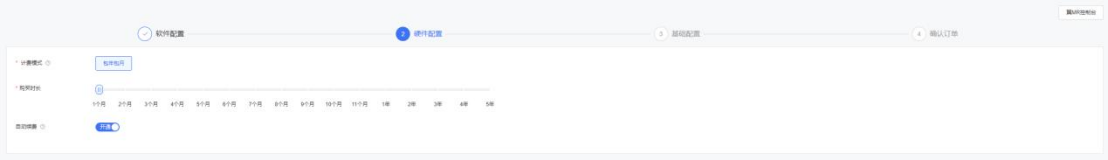
1. 进入翼 MR 控制台–我的集群页面。
2. 选择可续订的集群，在操作列点击高亮显示的续订按钮。

3. 进入所选集群的“续订管理”页面，按提示完成续订操作即可。更多详细信息可参考费用中心–手动续订 。

自动续订

方法一：

1. 订购集群时，在硬件配置页面，开启“自动续费”开关并提交订单，即可完成新集群的自动续订配置。



方法二：

1. 在天翼云官网，点击右上角的“管理中心”。
2. 左侧导航栏点击“订单管理”–“续订管理”，输入筛选条件，找到需要续订的产品，点击操作栏中的“开通自动续订”。
3. 勾选协议并点击“确定提交”即可完成配置。更多详细信息可参考费用中心–自动续订 。

2.8 退订

包周期计费

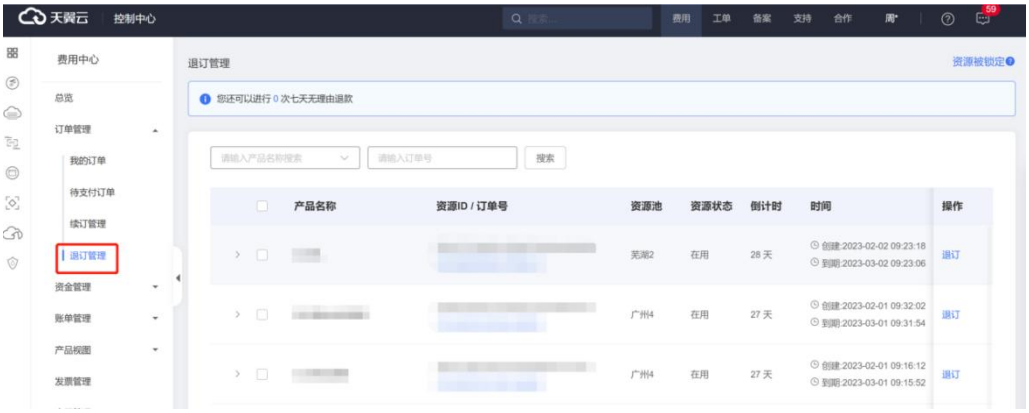
包年包月购买的集群，您可以根据业务需要，手动退订集群来释放资源。翼 MapReduce 集群在申请退订前，请做好数据备份。

注意：

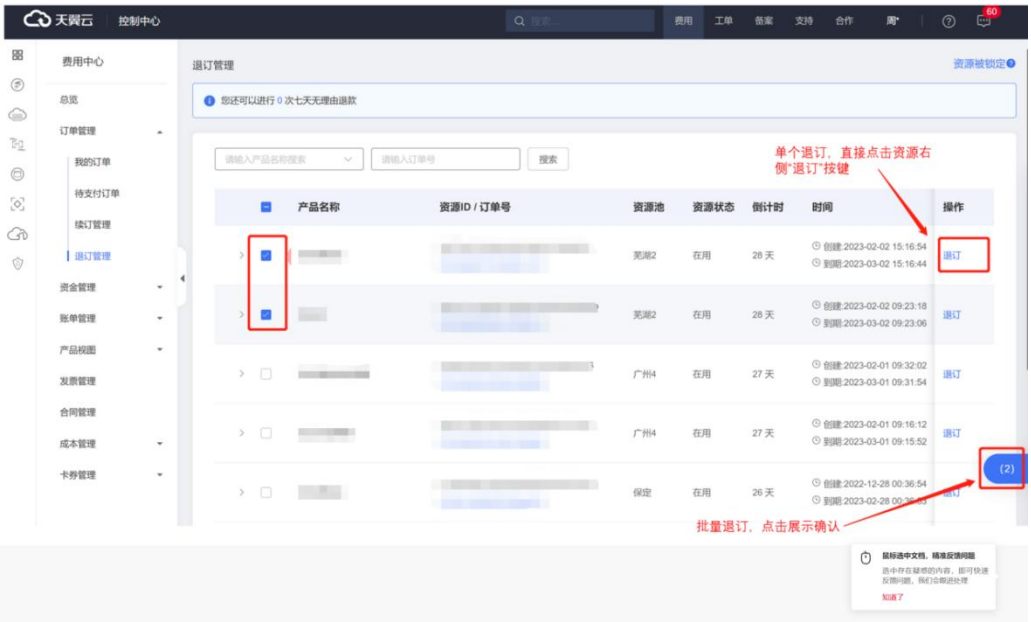
1. 退订翼 MapReduce 集群会导致集群所有数据都被删除，且不可恢复，请谨慎操作。
2. 具体涉及到退订的费用以及规则，详细请参见费用中心—退订规则说明。

方法一、通过订单管理退订：

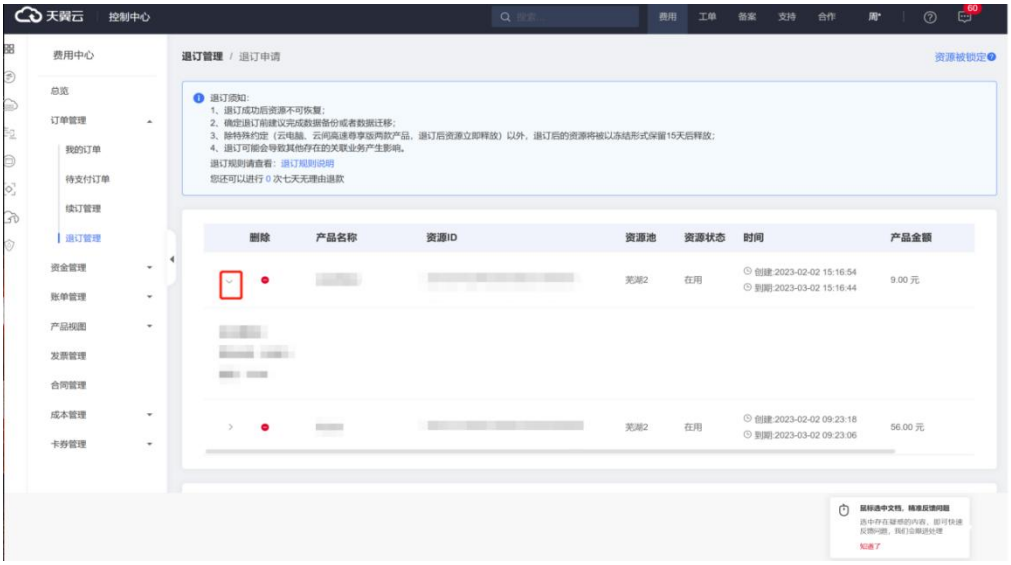
1. 登录天翼云官网，进入“费用中心—订单管理—退订管理”页面。



2. 选择要退订的资源，点击退订，天翼云目前支持单个退订和批量退订。



3. 退订前仔细阅读退订须知，然后确认退订资源信息。



4. 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。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。

方法二、通过翼 MR 控制台退订：

1. 进入翼 MR 控制台—我的集群页面。
2. 选择需要退订的集群，在操作列点击高亮显示的退订按钮。

3. 进入到所选集群的退订管理页面，按照提示步骤完成退订操作即可。

3 快速入门

3.1 如何使用翼 MR

天翼云大数据平台 翼 MapReduce 是一个在云上部署和管理 Hadoop 系统的服务，一键即可开通部署 Hadoop、HBase、Hive、Spark、Flink、Doris 等大数据集群。翼 MapReduce 提供租户完全可控的企业级大数据集群云服务，轻松运行 Hadoop、HBase、Hive、Spark、Flink、Doris 等大数据组件。

翼 MapReduce 使用简单，通过使用在集群中连接在一起的多台计算机，您可以运行各种任务，处理或者存储（PB 级）巨量数据。翼 MapReduce 的基本使用流程如下：

1. 创建集群：进入天翼云官网翼 MapReduce 服务订购页面，选择需要的业务场景，创建集群。用户可以通过业务场景简化大数据集群组合，可以用于数据湖、数据服务、云搜索、数据分析等业务需要。选择完业务场景与服务后，用户可以进一步指定集群的操作系统、主机类型、CPU 类型、实例规格、实例数量、数据盘类型（普通 IO、高 IO、超高 IO 与通用型 SSD 多款类型）等。

2. 管理集群：上述集群购买部署完成后，翼 MapReduce 为用户提供企业级的大数据集群的统一管理平台翼 MR Manager，帮助用户快速掌握集群服务及主机的健康状态，通过图形化的指标监控及时获取系统的关键信息，根据实际业务的性能需求修改服务属性的配置，对集群、服务、角色实例等实现一键启停等操作。

3. 创建翼 MapReduce 操作用户：主用户可以使用统一身份认证服务（Identity and Access Management，简称 IAM）来创建子用户，通过精细的权限管理，在翼 MR 控制台给子用户赋予不同的角色权限。

4. 退订：如果您不想继续使用已有集群，可以通过天翼云官网的订单管理页面，对集群进行退订操作。

5. 续订：如果您的集群即将到期，可以通过续订操作延长集群使用期限。

6. 配置升级：当 master、core 或者 task 节点 ECS 实例的规格（vCPU 和内存）无法满足您的业务需求时，您可以使用配置升级功能提升 ECS 实例规格。

7. 节点扩容：当 master、core 或 task 节点组内的资源无法满足您的业务需求时，您可以使用节点扩容功能增加 ECS 实例数量。请参见[节点扩容](https://www.ctyun.cn/document/10224959/10240786)说明。

8. 新增节点组：当存量 core 或 task 节点组的计算或存储资源无法满足您的业务需求时，您可以使用新增节点组功能增加 ECS 实例。请参见[新增节点组](https://www.ctyun.cn/document/10224959/10240788)说明。

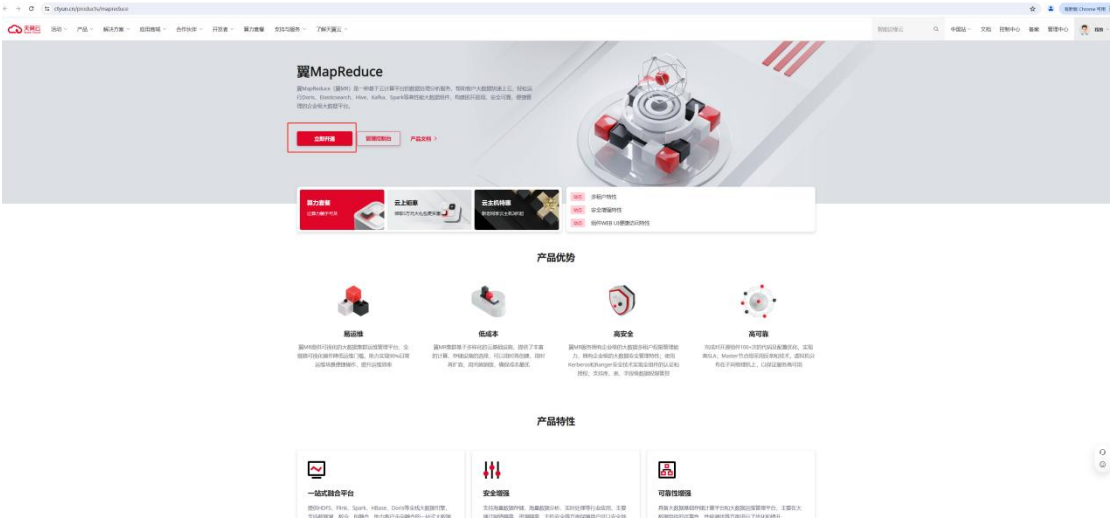
9. 磁盘扩容：当 master、core 或 task 节点的数据存储空间无法满足您的业务需求时，您可以使用磁盘扩容功能增加云硬盘－数据盘的空间。请参见[磁盘扩容](https://www.ctyun.cn/document/10224959/10691200)说明。

3.2 创建集群

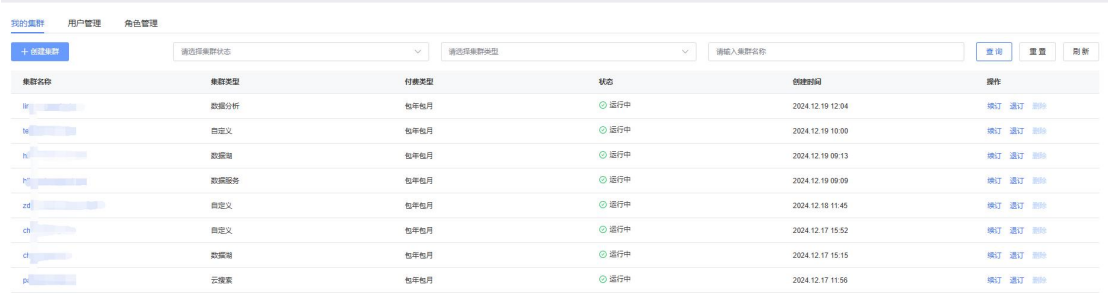
操作步骤

1、进入集群创建页面

方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。

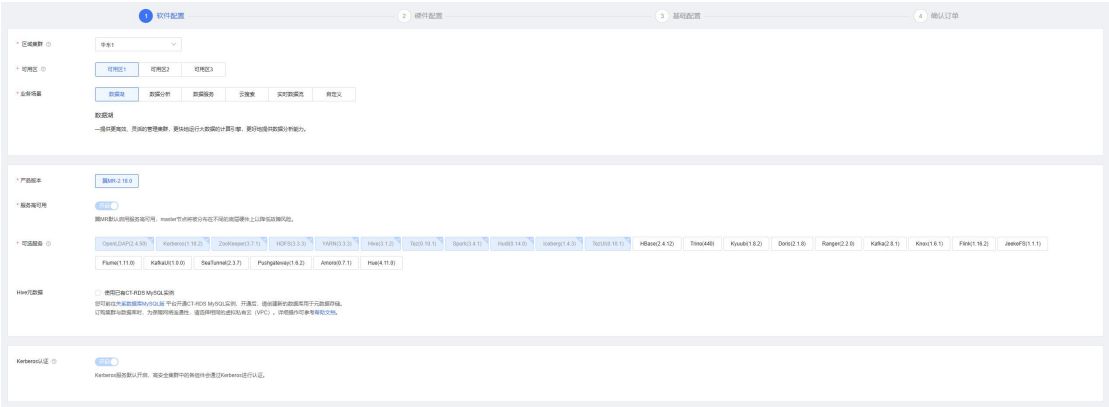


方法二：进入翼 MapReduce 产品一类节点资源池的产品控制台，点击“+创建集群”。



2、软件配置

进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：



- 区域集群与可用区：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。
- 业务场景：选择所需业务场景，此处以“数据湖”场景为例进行说明。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用且不可关闭高可用模式。
- 可选服务：由可选组件和必选组件组成，根据业务场景而定。您可根据自身业务场景对可选组件进行选择。
- 元数据：使用 Hive、Ranger、Amoro 与 Hue 组件服务时，需要配置元数据库。请先确认是否已有 CT-RDS MySQL 实例，若尚未创建，可点击提示链接，前往“关系数据库 MySQL 版”进行开通。若已创建实例，请勾选该选项并进行元数据配置。

- 元数据配置：勾选“使用已有 CT-RDS for MySQL”后，请在展开信息中填写所需的 5 项配置信息。

注意：

1、当前仅支持通过内网地址进行数据库连接。请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库。

2、配置 Hive、Ranger、Amoro 与 Hue 的元数据信息时，请提前创建数据库，若需要部署多个服务，请创建多个数据库，创建方式可参考创建数据库。

3、元数据配置所需信息可前往数据库控制台查看。

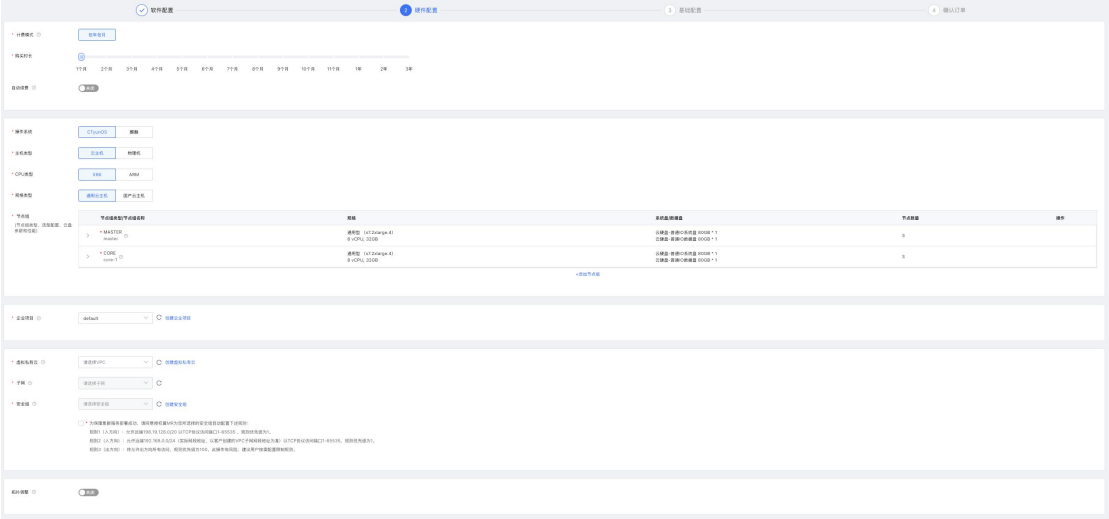
4、请确保配置的数据库用户具有该数据库的读写权限。

5、创建集群时，若元数据库配置错误，将导致 Hive/Ranger/Amoro/Hue 异常，但不影响集群的创建与部署，集群创建后，您可前往 Manager 的集群服务，进入集群服务详情，通过运维操作中的“元数据库配置”操作，替换原有元数据库的配置信息。

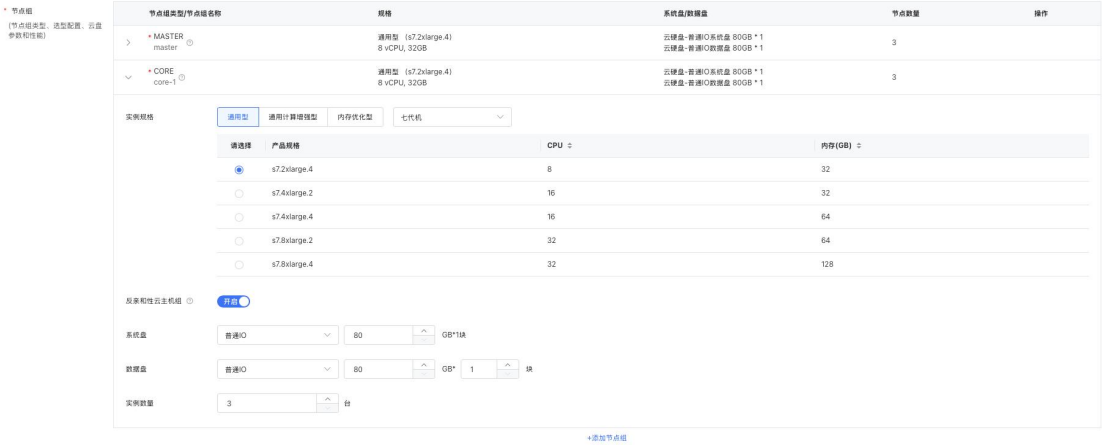
- Kerberos 身份认证：集群中的软件以 Kerberos 安全模式启动，此选项不可关闭。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如需：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 和 task 节点进行增加/删除。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。
- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

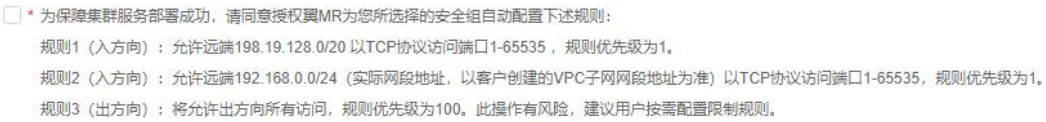
  注：

1）集群和虚拟私有云需在同一企业项目下。

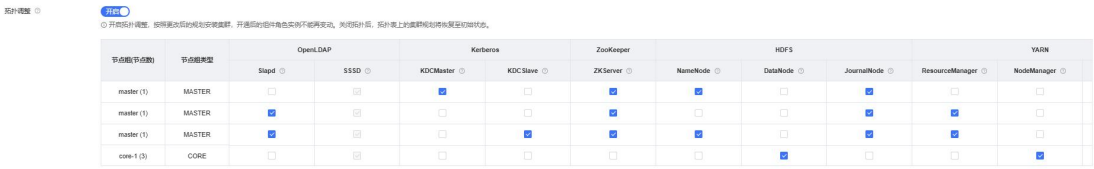
2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。

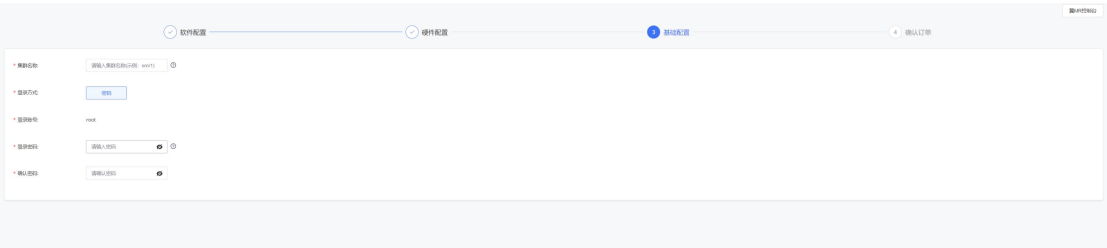


- 拓扑调整：从 V2.16.0 版本起，数据湖、数据服务、实时数据流与自定义场景支持拓扑调整功能，用户可以手动调整已选组件的各角色在节点的部署位置，平均分配各个节点上的资源。系统默认已按照各组件角色的部署原则自动分配实例，您可按需开启该功能进行配置调整。



4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下

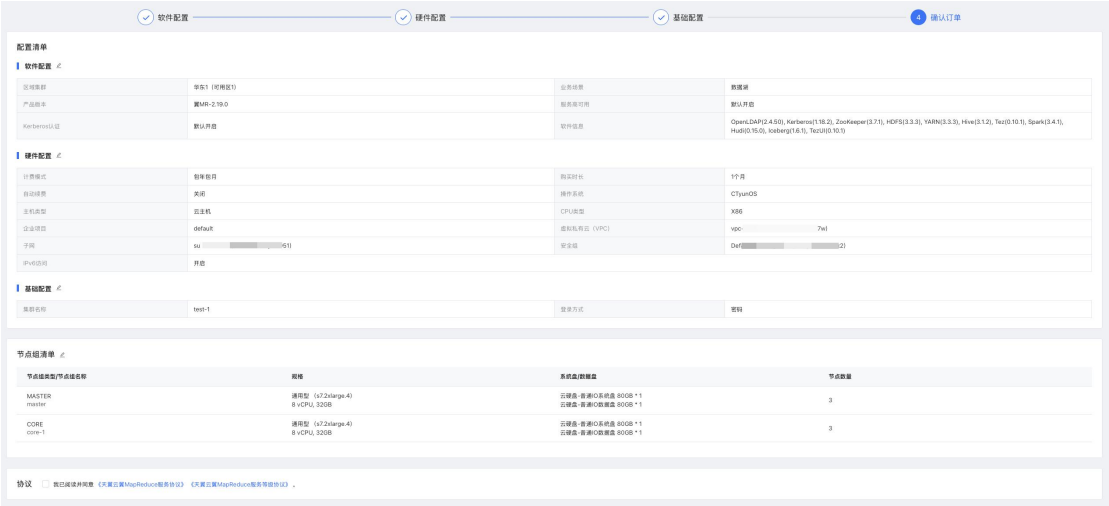


- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。

- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。确认订单页面如下图所示，配置清单说明如下：



- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

6、查看集群创建进度

在翼 MapReduce 的控制台内，可以查看集群信息。集群创建需要时间，集群列表中显示当前创建的进度，请您耐心等待。

3.3 上传示例数据和程序

操作步骤

1. 登录翼 MapReduce 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页签中展开节点组，选择要进入的节点。
4. 单击该节点右侧的“远程连接”。

5. 根据界面提示，**点击输入框**，输入 master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。

6. 在 hdfs 根目录下，创建 input 文件夹。

```
hadoop fs -mkdir /input
```

7. 上传数据。

```
hdfs dfs -put /usr/local/text.txt  /input
```

8. 查看上传的数据。

```
hdfs dfs -ll /input
```

9. 编写一个 WordCount 的 Java 程序，并进行编译打包成 jar 包。

10. 将该 jar 包上传到服务器的/client 目录下：

```
scp 本地 jar 包路径 用户名@Master 的公网 ip 地址:/client
```

11. 执行命令。

```
hadoop jar  jar 包名 WordCount  /input/text.txt
```

3.4 添加作业

用户可将自己开发的大数据应用程序提交到翼 MapReduce 中，执行程序并获取结果。

本章节以 MapReduce 作业为例，MapReduce 作业用于提交 jar 程序快速并行处理大量数据，是一种分布式数据处理模式和执行环境。通过此次的指导，希望用户能够对如何在翼 MapReduce 服务中提交、运行 MapReduce 作业有清晰的认知。

当前翼 MR 服务尚不支持通过前台界面化进行作业提交、运行的功能。本次的介绍仅说明如何通过后台功能来提交作业。

通过后台提交作业

天翼云大数据平台 翼 MapReduce 服务中的大数据集群客户端默认安装路径为“/usr/local/”。

1. 登录翼 MR 管理控制台。

2. 选择“集群列表 > 我的集群”，选中一个运行中的集群并单击集群名，进入集群基本信息页面。

3. 选择“节点管理”，单击展开节点组，显示节点列表。

4. 在目标节点的右侧“操作”列中，单击“远程连接”。

5. 在 VNC 方式新标签页 Web 页面中，可以看到服务器的命令行界面，类似于 Linux 服务器登录模式，输入 root 账号，并输入密码。

6. 当前翼 MR 集群默认都是开启 Kerberos 认证，大数据集群服务及用户使用到的 keytab 文件默认放置在“/etc/security/keytabs/”目录下，执行以下命令做 Kerberos 认证。

```
kinit 翼 MapReduce 集群用户
```

   - 进入“/etc/security/keytabs/”目录下，选择你要使用的 keytab 文件，本文以 hdfs.keytab 为例，执行如下命令，获取到正确的 Principal 名称。

```
klist -kt hdfs.Keytab
```

   - 使用上一步骤获取到的 Principal 名称，执行如下认证命令。

```
kinit -kt hdfs.Keytab Principal 名称
```

7. 执行如下命令提交 pi 计算作业。

```
cd /usr/local/hadoop3/share/hadoop/mapreduce

hadoop jar hadoop-mapreduce-examples-3.3.3.jar pi 10 10
```

```
                Input split bytes=1380
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=339
                Reduce input records=20
                Reduce output records=0
                Spilled Records=40
                Shuffled Maps =10
                Failed Shuffles=0
                Merged Map outputs=10
                GC time elapsed (ms)=515
                CPU time spent (ms)=5660
                Physical memory (bytes) snapshot=14138007552
                Virtual memory (bytes) snapshot=60023365632
                Total committed heap usage (bytes)=15106834432
                Peak Map Physical memory (bytes)=1402277888
                Peak Map Virtual memory (bytes)=5455585280
                Peak Reduce Physical memory (bytes)=292245504
                Peak Reduce Virtual memory (bytes)=5487992832
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1180
        File Output Format Counters
                Bytes Written=97
Job Finished in 23.228 seconds
Estimated value of Pi is 3.20000000000000000000
```

3.5 从零开始使用 Kerberos 认证集群

本章节提供从零开始使用安全集群并执行 MapReduce 程序、Spark 程序和 Hive 程序的操作指导。

本指导的基本内容如下所示：

1. 创建集群。
2. 创建用户。
3. 执行 MapReduce 程序。
4. 执行 Spark 程序。
5. 执行 Hive 程序。

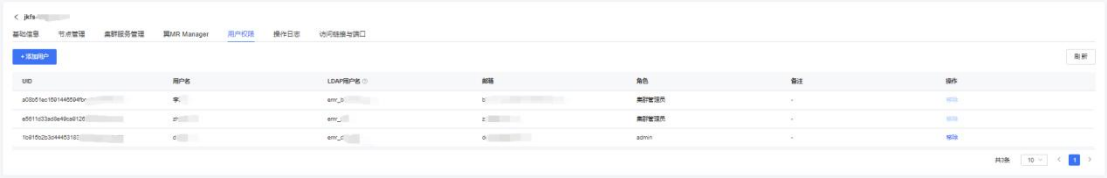
创建集群

1. 创建数据湖、数据服务或实时数据流集群，“Kerberos 身份认证”默认开启，且不支持关闭，详见创建集群相关内容。
2. 登录翼 MR 管理控制台页面。
3. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

创建用户

1. 登录翼 MapReduce 管理控制台页面。

2. 单击“用户权限”，进入到用户管理页面。



3. 单击“添加用户”，输入相关信息，单击“确认”即可。

执行 MapReduce 程序

本小结提供执行 MapReduce 程序的操作指导，旨在指导用户在安全集群模式下运行程序。

前提条件

已编译好待运行的程序及对应的数据文件，如 mapreduce-examples-1.0.jar、input_data1.txt。

操作步骤

1. 采用远程登录软件，通过 ssh 登录到集群的 master 节点。
2. 登录成功后分别执行如下命令，在/opt/client 目录下创建 test 文件夹。

```
cd /opt/client

mkdir test
```

3. 分发 keytab 文件到/opt/client 目录，例如 test.keytab
4. 执行如下命令认证已创建的用户，例如 test。

```
kinit -kt /opt/client/test.keytab test
```

5. 执行如下命令将数据导入到 HDFS 中。

```
cd test
```

```
hdfs dfs -mkdir /tmp/input

hdfs dfs -put intput_data1.txt /tmp/input
```

6. 执行如下命令运行程序。

```
yarn jar mapreduce-examples-1.0.jar WordCount /tmp/input /tmp/output
```

其中：

/tmp/output 指 HDFS 文件系统中 output 的路径，该目录必须不存在，否则会报错。

执行 Spark 程序

本小结提供执行 Spark 程序的操作指导，旨在指导用户在安全集群模式下运行程序。

前提条件

已编译好待运行的程序及对应的数据文件，如 Collection.jar、input_data1.txt。

操作步骤

1. 采用远程登录软件，通过 ssh 登录到集群的 master 节点。
2. 登录成功后分别执行如下命令，在/opt/client 目录下创建 test 文件夹。

```
cd /opt/client

mkdir test
```

3. 分发 keytab 文件到/opt/client 目录，例如 test.keytab
4. 执行如下命令认证已创建的用户，例如 test。

```
kinit -kt /opt/client/test.keytab test
```

5. 执行如下命令将数据导入到 HDFS 中。

```
cd test

hdfs dfs -mkdir /tmp/input
```

```
hdfs dfs -put intput_data1.txt /tmp/input
```

6. 执行如下命令运行程序。

```
cd /opt/client/Spark/spark

bin/spark-submit --class com.tyy.Collection --master yarn-client/opt/client/t

est/Collection-1.0.jar  /tmp/input
```

执行 Hive 程序

本小结提供执行 Hive 程序的操作指导，旨在指导用户在安全集群模式下运行程序。

前提条件

已编译好待运行的程序及对应的数据文件，如 hive-examples-1.0.jar、input_data1.txt。

操作步骤

1. 采用远程登录软件，通过 ssh 登录到集群的节点。
2. 登录成功后分别执行如下命令，在/opt/client 目录下创建 test 文件夹。

```
cd  /opt/client

mkdir test
```

3. 分发 keytab 文件到/opt/client 目录，例如 test.keytab
4. 执行如下命令认证已创建的用户，例如 test。

```
kinit -kt /opt/client/test.keytab test
```

5. 执行如下命令运行程序。

```
chmod +x /opt/hive_examples -R

cd /opt/hive_examples
```

```
java -cp .:
hive-examples-1.0.jar:/opt/hive_examples/conf:/opt/client/Hive/Beeline/lib/:/opt/client/HDFS/hadoop/lib* com.tyy.hive.example.ExampleMain
```

3.6 删除集群

如果作业执行结束后不需要集群，可以删除翼 MapReduce 集群。

背景信息

一般在集群到期 15 天后用户没有续订，或者由用户主动发起退订后，才执行集群删除操作。

操作步骤

1. 登录翼 MapReduce 控制台。
2. 在我的集群页面选择需要删除的集群，在操作列点击删除按钮。

3. 在页面弹出的提示对话框中点击确认即可。

4. 删除后的集群不再显示在我的集群列表中。

3.7 手动修改配置建议书

操作场景

当节点组内 ECS 实例的规格（vCPU 和内存）无法满足您的业务需求时，您可以使用配置升级功能提升 ECS 实例规格。升级配置后需要手动修改 HDFS、YARN 和 Spark 等服务的配置信息。本文为您介绍如何在翼 MR Manager 的“配置管理”页面修改配置项。

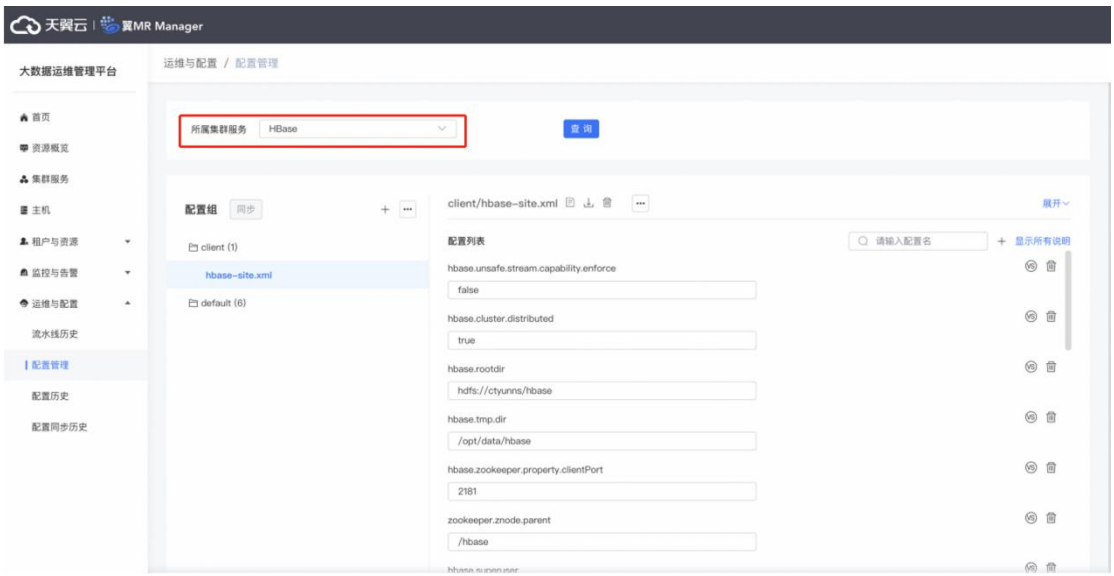
前提条件

已创建集群。

操作步骤

在翼 MR Manager 中，单击“运维与配置”。

1. 单击“配置管理”。
2. 选择“所选集群服务”，点击查询，即可在当前页面修改配置信息。



配置修改建议说明

1. Doris：Doris 所在节点配置升级后，Doris FE 建议使用节点一半内存。
2. Elasticsearch：一个 ES 节点，内存建议不超过 64G。ES 所在节点配置升级后，ES 会自动根据节点情况设置内存值，一般无须用户手动修改，但需要重启集群。如果想手动设置，可以修改 jvm.options 文件配置-Xms30g、-Xmx30g 参数，并重启集群。
3. HBase：HBase 所在节点配置升级后，修改建议如下：

- hbase-site.xml：hbase.regionserver.handler.count

  说明

  一般跟 CPU 核数相同。

- hbase-env.sh：export HBASE_MASTER_OPTS ："-Xmsg -Xmxg"

  说明

  master 不消耗很多内存，一般默认不添加或者分配 2~8G 左右。

- hbase-env.sh：export HBASE_REGIONSERVER_OPTS："-Xmsg -Xmxg "

  说明

  regionserver 需要较多内存，一般配置内存配额的一半或更多。

4. HDFS：HDFS 所在节点配置升级后，可以根据 hadoop-env.sh 参数配置进行，通过调整服务的内存大小调整服务的性能，如-Xmx20g -Xms20g -Xmn4g，然后重启服务。NameNode 建议文件、目录、数据块之和 1 亿，配置 50G。

5. Hive：Hive 所在节点配置升级后，可以通过 hive-env.sh 统一参数配置来进行，也可以在作业提交时使用额外参数指定来进行。通过调整服务的内存大小调整服务的性能，如-Xmx20g -Xms20g -Xmn4g，然后重启服务。内存大小可以根据机器的总内存而定，建议初始值为总内存大小的 10%，后续根据性能需求调整。

6. Kafka：Kafka 所在节点配置升级后，建议配置如下：

   - kafka-env.sh 设置 jvm 配置参数：调整 jvm 堆大小，通过调整参数：export KAFKA_HEAP_OPTS="-Xmx20G -Xms20G -Xmn4g"设置堆大小。

   - server.properties 文件建议修改的配置项：

     - num.io.threads：修改写磁盘的线程数，建议配置为 CPU 核数的 50%；
     - num.replica.fetchers：修改副本拉取线程数，建议配置为 CPU 核数 50%的 1/3；
     - num.network.threads：修改数据传输线程数，建议配置为 CPU 核数的 50%的 2/3；
     - replica.fetch.max.bytes：副本拉取数据量的大小。内存增加，可以适当加大该值；
     - socket.send.buffer.bytes：调整 socket 发送的数据量。内存增加，可以适当加大该值；
     - socket.receive.buffer.bytes：调整 socke 接受的数据量。内存增加，可以适当加大该值；
     - socket.request.max.bytes：socket 请求的数据量。内存增加，可以适当加大该值。

7. Kerberos：建议保持默认值，无需修改配置。

8. Kibana：Kibana 是一个基于 NodeJS 的单页 web 应用，一般情况下，对内存 CPU 占用很少，无须修改内存、CPU 等配置。

9. Kyuubi：Kyuubi 一般情况下，对内存 CPU 占用很少，无须修改内存、CPU 等配置。

10. OpenLDAP：建议保持默认值，无需修改配置。

11. Ranger：Ranger 所在节点配置升级后，修改建议如下：

   - ranger-admin 通过 {installdir}/ews/ranger-admin-services.sh 中变量 ranger_admin_max_heap_size 的值修改 JMX，JAVA_OPTS 修改 Xmx、Xmn 等 JVM 参数，一般设置 1-8g，1K policy 建议设置为 1G，1W policy 建议设置为 8G。

   - ranger-usersync 通过 /{installdir}/ranger-usersync-services.sh 中变量 ranger_usersync_max_heap_size 的值修改 JMX，JAVA_OPTS 修改 Xmx Xmn 等 JVM 参数，一般设置 1-8g，1K policy 建议设置为 1G，1W policy 建议设置为 8G。

12. Spark：Spark 所在节点配置升级后，修改建议如下：

   - spark.history.kerberos.principal 和 spark.history.kerberos.keytab 为 spark 读写 eventLog 的租户，用户如有特殊需求自行更改。

   - spark.yarn.historyServer.address=：说明了 history server 的地址，用户如有特殊需求自行更改。

   - spark.dynamicAllocation.enabled 和 spark.dynamicAllocation.maxExecutors 分别控制动态和动态开启下能使用的最大资源，用户如有特殊需求自行更改。

   - spark.executor.cores 和 spark.executor.memory 确保 spark.executor 的每一个 core 分配到 2~4g 内存，标准是 4g，具体视情况而定，设置 core 的 memory 设置过小 executor 容易 oom。

13. Trino：Trino 的服务包括 coordinator 和 worker。Trino 所在节点配置升级后，可以根据 jvm.config 参数配置进行，通过调整服务的内存大小调整服务的性能，如 -Xmx128g -Xms128g，然后重启服务。

14. YARN：YARN 所在节点配置升级后，可以根据 yarn-env.sh 参数配置进行，通过调整服务的内存大小来调整服务的性能，如-Xmx20g -Xms20g -Xmn4g，然后重启服务。NM 用于集群中作业的内存和 CPU，需要修改 NodeManager 节点的 yarn-site.xml 中的 yarn.nodemanager.resource.memory-mb 的值，该值用于所有作业的最多可用内存；以及 yarn.nodemanager.resource.cpu-vcores 的值，该值用于所有作业的最多可用虚拟 CPU 核数。

15. ZooKeeper：ZooKeeper 所在节点配置升级后，可通过配置 java.env 文件，在其中添加：export ZK_SERVER_HEAP=2048（这里设置的单位默认是 MB）。

3.8 组件依赖关系表

组件依赖关系说明

组件名称	集群部署的依赖组件

组件名称	集群部署的依赖组件
Doris	/
Elasticsearch	/
HBase	HDFS、ZooKeeper、Kerberos、OpenLDAP
HDFS	ZooKeeper、Kerberos、OpenLDAP
Hive	HDFS、YARN、ZooKeeper、Kerberos、OpenLDAP、Hudi、Iceberg、Tez
Kafka	ZooKeeper、Kerberos、OpenLDAP
Kerberos	OpenLDAP
Kibana	Elasticsearch
Kyuubi	Spark、Hive、HDFS、YARN、ZooKeeper、OpenLDAP、Kerberos、Hudi、Iceberg、Tez
OpenLDAP	/
Ranger	HDFS、OpenLDAP、Kerberos、ZooKeeper
Spark	HDFS、YARN、Hive、ZooKeeper、OpenLDAP、Kerberos、Hudi、Iceberg、Tez
Trino	Kerberos、OpenLDAP
YARN	ZooKeeper、Kerberos、HDFS、OpenLDAP
ZooKeeper	Kerberos、OpenLDAP
Flink	HDFS、YARN、Kerberos、OpenLDAP、ZooKeeper、Iceberg、SeaTunnel、Hudi
Hudi	Hive/Spark/Flink
Iceberg	Hive/Spark/Flink
Knox	Kerberos、OpenLDAP
Logstash	/
JeekeFS	Kerberos、OpenLDAP

组件名称	集群部署的依赖组件
Tez	HDFS、YARN、Hive、Kerberos、OpenLDAP、ZooKeeper、Iceberg、Hudi
Flume	/
KafkaUI	Kafka、Kerberos、OpenLDAP、ZooKeeper
SeaTunnel	HDFS、YARN、Flink、OpenLDAP、Kerberos、ZooKeeper、Iceberg、Hudi
Pushgateway	/
TezUI	OpenLDAP、Hudi、Kerberos、Hive、ZooKeeper、Iceberg、Tez、HDFS、YARN
Amoro	Kerberos、OpenLDAP、ZooKeeper、HDFS、YARN
Hue	Kerberos、OpenLDAP

4 用户指南

4.1 准备用户

4.1.1 创建翼 MapReduce 操作用户

如果您需要对您所拥有的翼 MapReduce 服务进行精细的权限管理，您可以使用统一身份认证服务（Identity and Access Management，简称 IAM），创建 IAM 子用户，并给子用户赋予不同的角色权限及企业项目权限。

如果主帐号已经能满足您的要求，不需要创建更多的子用户，您可以跳过本章节，不影响您使用翼 MapReduce 服务的其它功能。

创建用户

1、管理员登录天翼云官网，点击页面右上角用户名，在下拉列表中点击“账号中心”。

2、在账号中心页面，点击左侧菜单栏中的“统一身份认证”进入 IAM 用户管理页面。

3、点击左侧菜单栏栏中的“用户”，在页面中点击“创建用户”按钮开启创建。

说明：具体操作可参考文档创建用户。

用户授权

默认情况下，管理员创建的 IAM 用户没有翼 MR 操作权限，需要为其赋予角色权限后，才能使该用户获得集群操作权限。您可以通过下表中的功能，为新增用户授予权限。

功能	操作	权限作用范围
账号中心—统一身份认证服务	授予翼 MR 与企业项目权限	官网产品
翼 MR 控制台-用户管理	同步 IAM 账号至翼 MR	翼 MR 产品
翼 MR 控制台-角色管理	管理集群角色（含角色的新增、编辑与删除）	翼 MR 产品
翼 MR 集群-用户权限	在指定集群内，为用户授予集群角色，支持被授权用户使用集群	翼 MR 单集群

1、账号中心—统一身份认证服务

授予翼 MR 权限

管理员可以通过“账号中心”–“统一身份认证服务”功能，为用户组授予“翼 MapReduce 管理员”或“翼 MapReduce 用户”权限。被授予“翼 MapReduce 管理员”身份的用户可以在翼 MR 控制台内进行 IAM 用户同步、新建角色、编辑角色、删除角色、删除集群、查看所在企业项目的集群等操作；被授予“翼 MapReduce 用户”身份的用户可以在翼 MR 集群列表中查看到所在企业项目的集群。具体操作可参考文档用户组授权 。

名称	类型	作用范围	描述	操作
翼MapReduce 用户	系统策略	全局	具备翼MapReduce的普通用户权限	查看 | 编辑 | 删除
翼MapReduce 管理员	系统策略	全局	具备翼MapReduce的用户管理权限	查看 | 编辑 | 删除

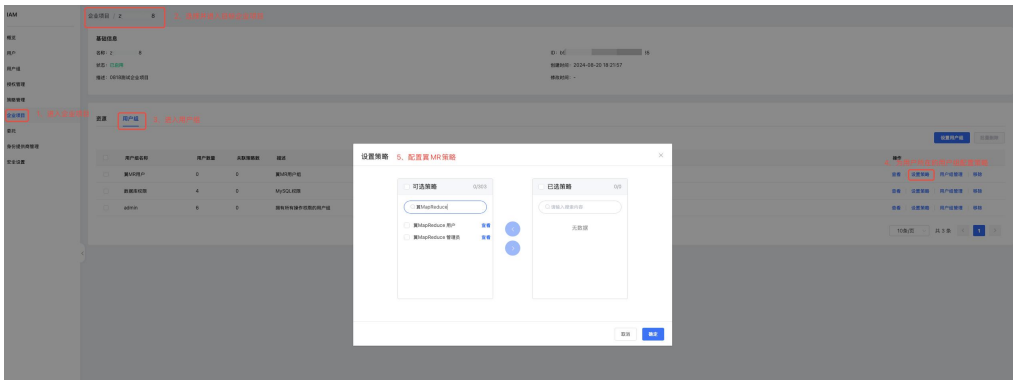
注意：

1）被授予“翼 MapReduce 管理员”权限的用户仅具备翼 MR 用户与角色的管理权限，集群使用权限，需要由具体集群的“集群管理员”或拥有该集群用户管理权限的用户进行授权。

2）如需为用户添加集群的订购、续订、查看账单等权限，可在“统一身份认证服务”中选择合适的策略为用户添加权限，如“resource admin”等。

授予企业项目权限

管理员可在“账号中心”–“统一身份认证服务”功能中，通过用户组授权或基于企业项目完成授权两种方式，为用户授予在目标企业项目下的翼 MR 使用权限。授权后，用户可在翼 MR 控制台看到该企业项目下的集群。



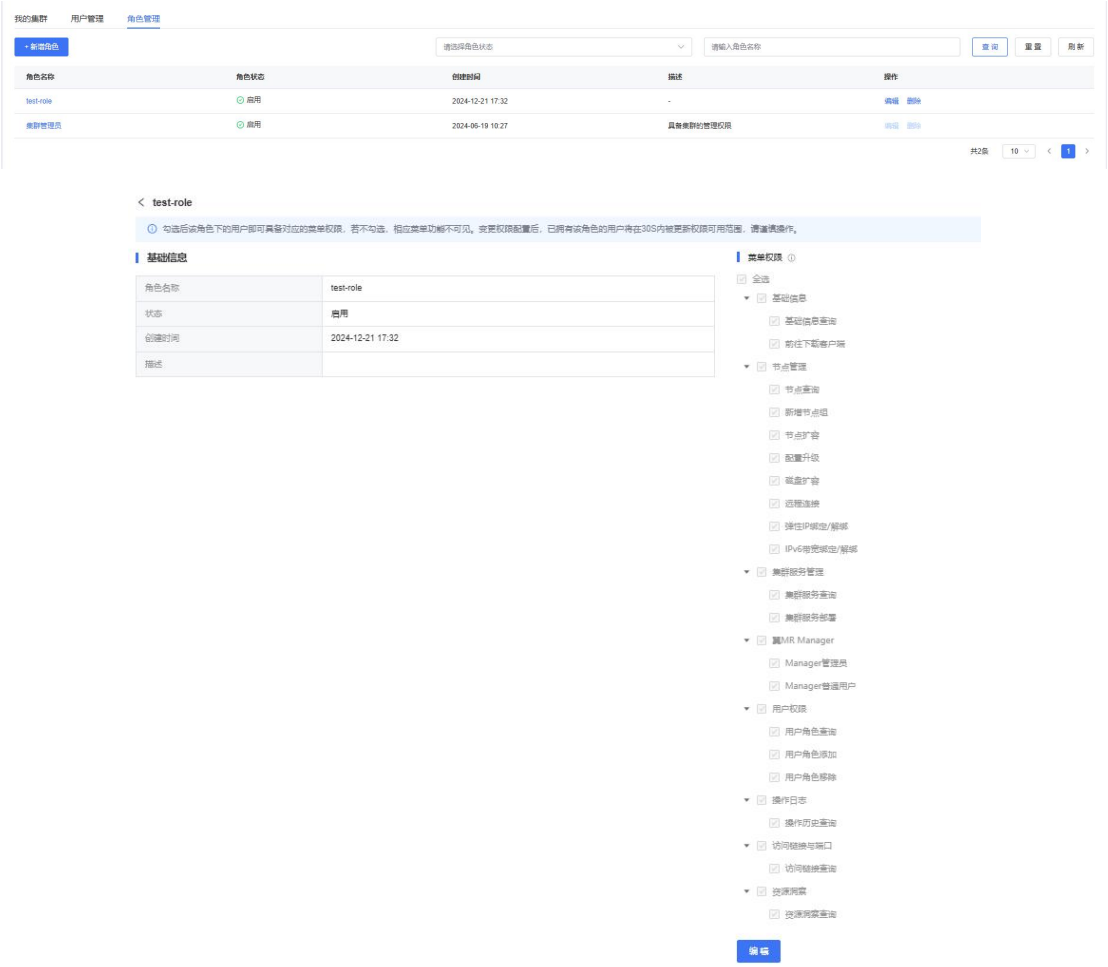
集群创建后，若希望将集群从当前企业项目迁出，并迁入到新的企业项目中，可参考企业项目资源迁出/迁入。迁移时，请同步迁移与集群存在绑定关系的资源，如虚拟私有云、弹性 IP 等产品。

2、翼 MR 控制台–用户管理

为支持 IAM 用户使用翼 MR 集群，需要将新增 IAM 用户的信息同步至翼 MR 产品。翼 MapReduce 管理员可前往“翼 MR 控制台”—“用户管理”页面，点击“IAM 同步”，翼 MR 将从 IAM 中获取最新用户数据，并将状态为启用的子用户同步到翼 MR 控制台的“用户管理”中。

3、翼 MR 控制台–角色管理

翼 MapReduce 管理员可在“翼 MR 控制台”—“用户管理”页面进行集群角色管理。通过创建与编辑角色，可为该角色授予集群的具体操作权限，如查看集群基础信息、管理节点、产看操作日志、下载客户端等。创建的角色可在具体集群内被绑定给目标用户，使得该用户具备集群操作权限。



4、翼 MR 集群–用户权限

集群管理员与拥有“用户角色添加”权限的用户，可在集群内的“用户权限”页面，为已同步至翼 MR 的 IAM 用户授予创建好的集群角色，使得新增用户具备集群使用权限。

4.1.2 LDAP 用户同步

LDAP 用户同步是将 IAM 用户同步到翼 MR Manager 中的“LDAP 用户”服务中，LDAP 用户名以“emr_邮箱前缀”表示。

前提：该集群下需要完成 OpenLDAP 集群服务的部署、配置以及启动，否则无法同步到“LDAP 用户”服务中。

操作步骤

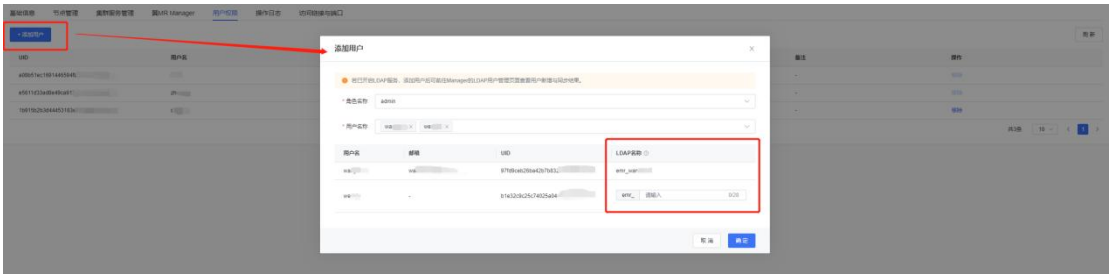
1. IAM 用户同步

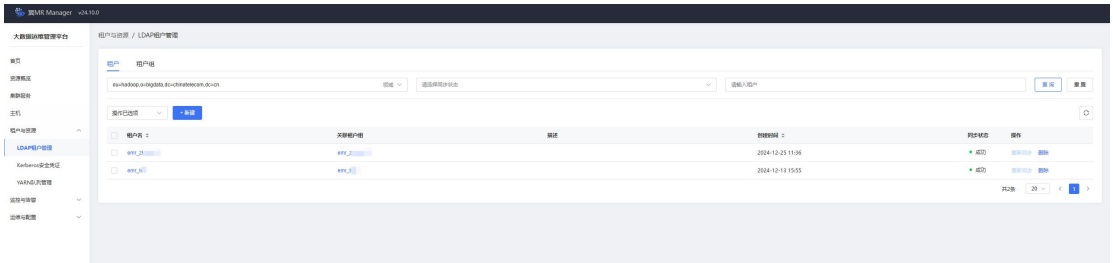
在翼 MapReduce 控制台的“用户管理”页面点击 IAM 同步，将 IAM 子用户信息同步到翼 MapReduce 控制台。

2. LDAP 用户同步

在目标集群的“用户权限”页面添加用户，若该集群已安装 LDAP 服务，选择要添加的用户后，将自动为符合命名规则的新增用户同步 LDAP 账号。

若用户邮箱为空或不符合 LDAP 用户命名规则，添加时需要为该用户重新定义 LDAP 名称。点击“确定”即可将 IAM 用户同步到翼 MR Manager 中的“LDAP 用户”服务中。





4.2 配置集群

4.2.1 创建方式简介

用户可以通过点击翼 MapReduce 产品页的“立即开通”或翼 MR 控制台内的“创建集群”进入订购页面，按需完成软件、硬件与基础配置，通过确认订单并支付完成集群创建，支付后我们将为您尽快拉起集群，您可前往翼 MR 控制台查看集群的创建进度。

创建数据湖集群：数据湖集群提供更高效、灵活的集群管理，更快地运行大数据的计算引擎，更好地提供数据分析能力，帮助您更加方便快捷的对数据湖的计算、存储资源进行构建和管理。

创建数据分析集群：数据分析集群使用开源的 MPP 架构的 OLAP 分析引擎，支持亚秒级的数据查询和多表 join，帮助您更加快捷的进行数据分析工作。

创建数据服务集群：数据服务集群提供高度可靠性、可扩展性的数据服务以满足您的数据需求。

创建云搜索集群：云搜索集群通过分布式计算和存储系统，为结构化/非结构化数据提供低成本、高性能及可靠性的检索、分析服务能力，帮助您更方便的检索大规模数据集。

创建实时数据流集群：实时数据流集群提供高效的流式计算、消息队列等组件能力，支持实时数据 ETL 和日志采集分析的业务需求。

创建自定义集群：自定义集群提供丰富灵活的服务搭配，支持自行选择业务所需服务。建议不要将多个存储类服务部署在一个节点组上，避免资源争抢。

4.2.2 快速创建集群

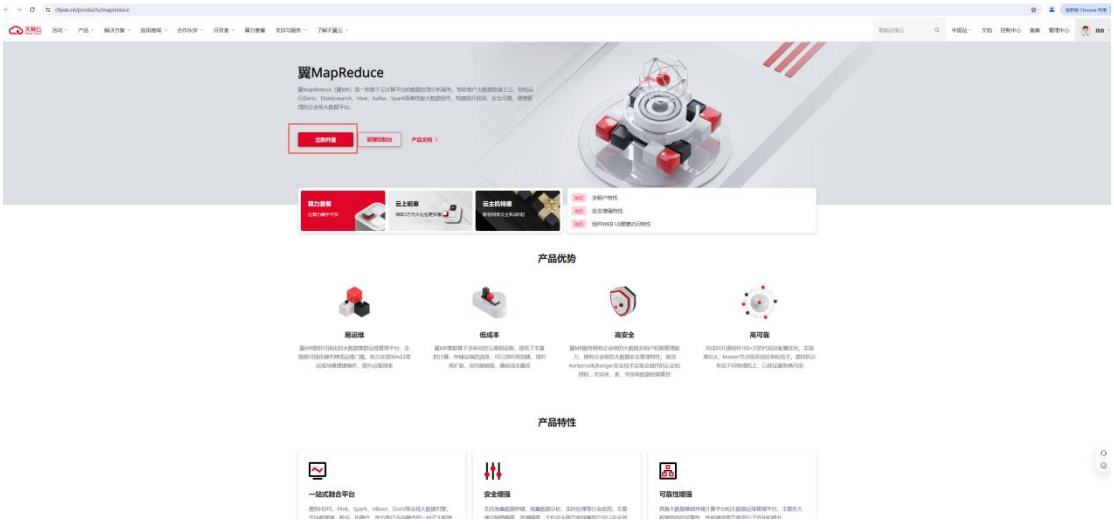
4.2.2 创建数据湖集群

数据湖场景提供更高效、灵活的管理集群，更快地运行大数据的计算引擎，更好地提供数据分析能力。在创建数据湖集群前，需要先创建虚拟私有云。

操作步骤

1、进入集群创建页面

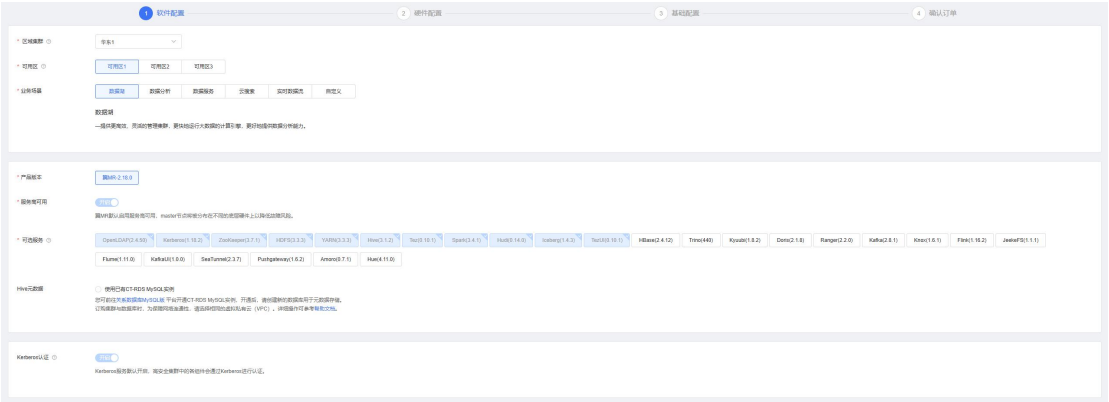
方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。



方法二：进入翼 MapReduce 产品一类节点资源池的产品控制台，点击“+创建集群”。

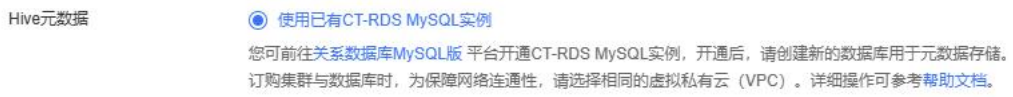
2、软件配置

进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：



- 区域集群与可用区：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。
- 业务场景：选择“数据湖”场景。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用且不可关闭高可用模式。

- 可选服务：由可选组件和必选组件组成，根据业务场景而定。您可根据自身业务场景对可选组件进行选择。
- 元数据：使用 Hive、Ranger、Amoro 与 Hue 组件服务时，需要配置元数据库。请先确认是否已有 CT-RDS MySQL 实例，若尚未创建，可点击提示链接，前往"关系数据库 MySQL 版"进行开通。若已创建实例，请勾选该选项并进行元数据配置。



- 元数据配置：勾选"使用已有 CT-RDS for MySQL"后，请在展开信息中填写所需的 5 项配置信息。

注意：

1、当前仅支持通过内网地址进行数据库连接。请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库。

2、配置 Hive、Ranger、Amoro 与 Hue 的元数据信息时，请提前创建数据库，若需要部署多个服务，请创建多个数据库，创建方式可参考创建数据库。

3、元数据配置所需信息可前往数据库控制台查看。

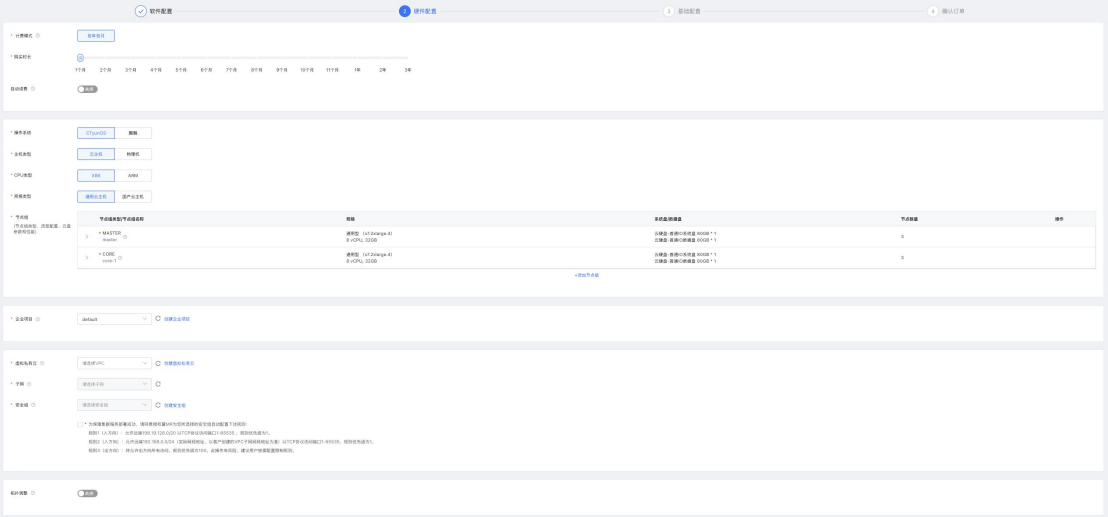
4、请确保配置的数据库用户具有该数据库的读写权限。

5、创建集群时，若元数据库配置错误，将导致 Hive/Ranger/Amoro/Hue 异常，但不影响集群的创建与部署，集群创建后，您可前往 Manager 的集群服务，进入集群服务详情，通过运维操作中的“元数据库配置”操作，替换原有元数据库的配置信息。

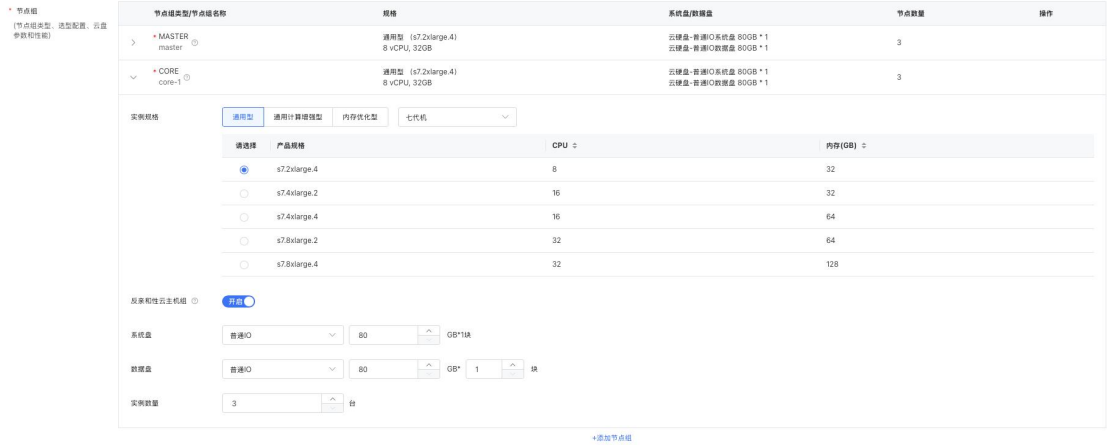
- Kerberos 身份认证：集群中的软件以 Kerberos 安全模式启动，此选项不可关闭。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如下：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 和 task 节点进行增加/删除。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。

- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

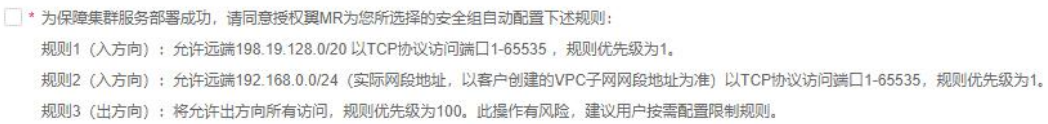
  注：

  1）集群和虚拟私有云需在同一企业项目下。

  2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

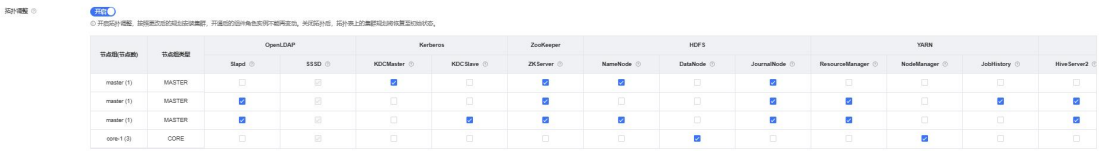
- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。



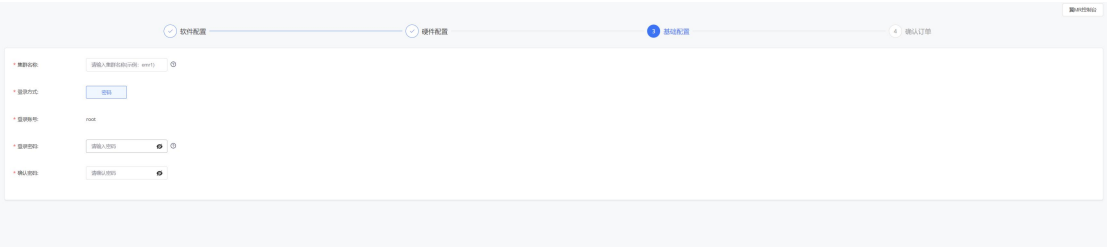
- 拓扑调整：从 V2.16.0 版本起，数据湖、数据服务、实时数据流与自定义场景支持拓扑调整功能，用户可以手动调整已选组件的各角色在节点的部署位置，平均分配各个节

点上的资源。系统默认已按照各组件角色的部署原则自动分配实例，您可按需开启该功能进行配置调整。



4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下：



- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。
- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。配置清单说明如下：

- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

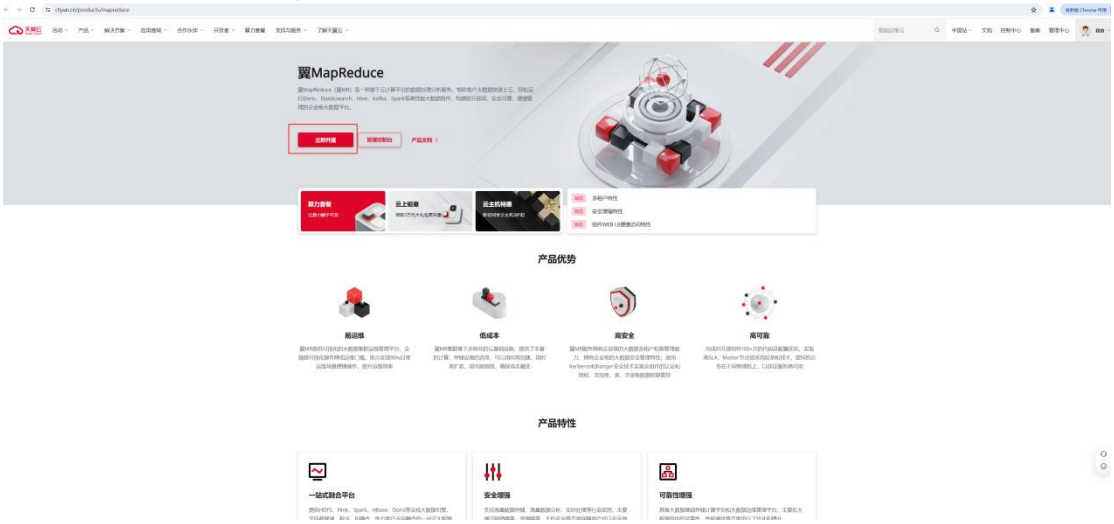
4.2.2 创建数据分析集群

数据分析集群使用 Apache Doris，Apache Doris 是开源的 MPP 架构的 OLAP 分析引擎，支持亚秒级的数据查询和多表 join。在创建数据分析集群前，需要先创建虚拟私有云。

操作步骤

1、进入集群创建页面

方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。



方法二：进入翼 MapReduce 产品一类节点资源池的产品控制台，点击“+创建集群”。

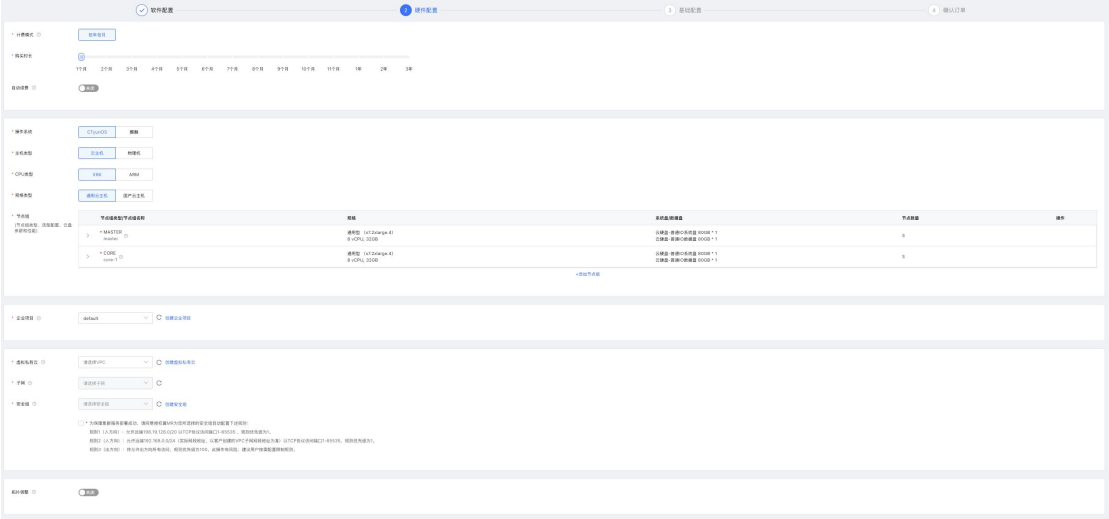
2、软件配置

进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：

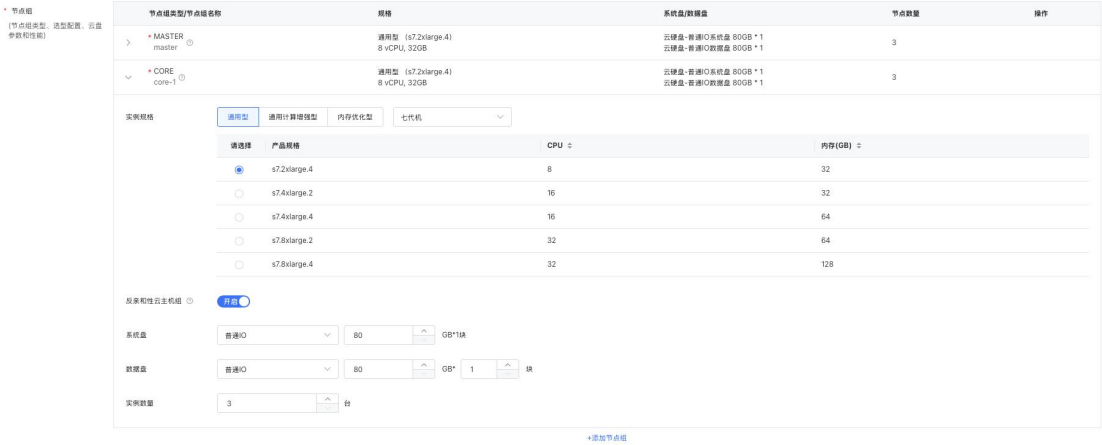
- 区域集群：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。
- 业务场景：选择“数据分析”场景。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用且不可关闭高可用模式。
- 可选服务：由可选组件和必选组件组成，根据业务场景而定。您可根据自身业务场景对可选组件进行选择。数据分析集群场景下默认只有 Doris 一个必选组件。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如下：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 节点进行增加/删除。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。
- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

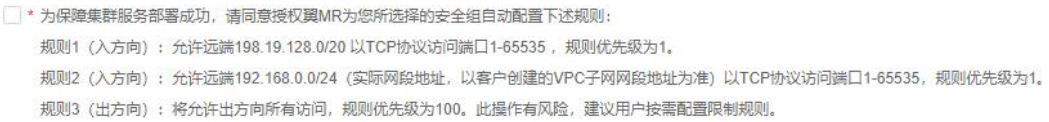
  注：

1）集群和虚拟私有云需在同一企业项目下。

2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

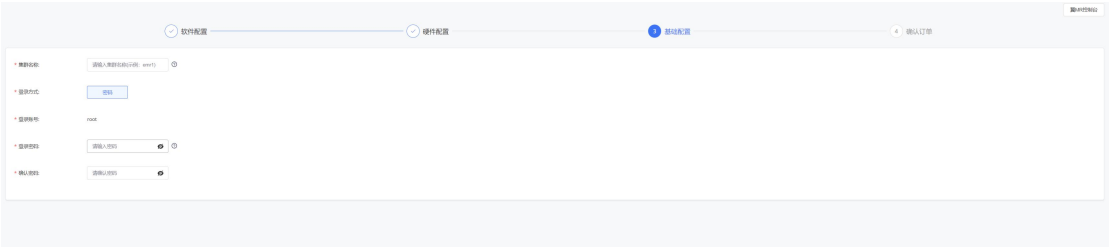
- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。



4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下：



- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。
- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。配置清单说明如下：

- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

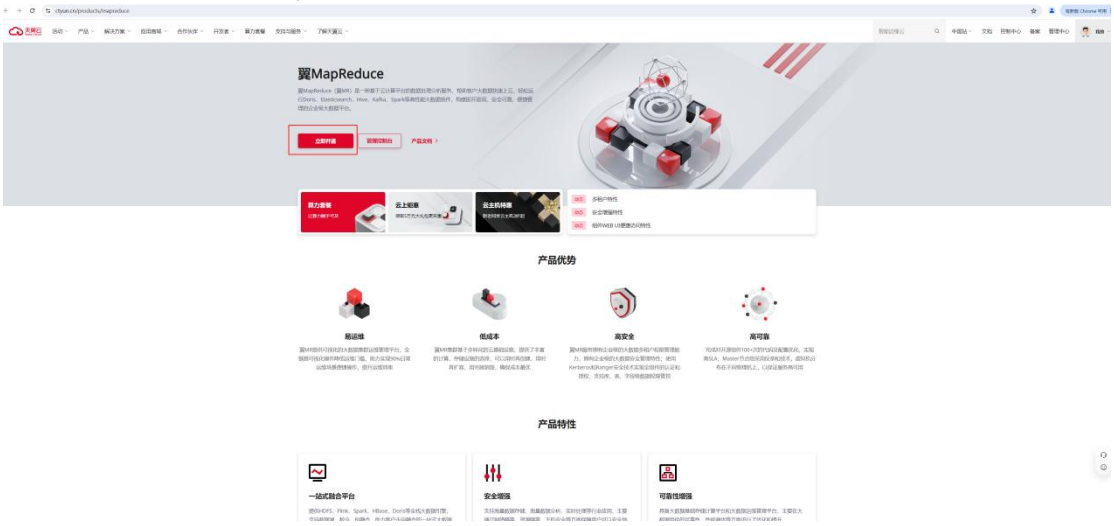
4.2.2 创建数据服务集群

数据服务集群提供更灵活、可靠、高效的数据服务。在创建数据服务集群前，需要先创建虚拟私有云。

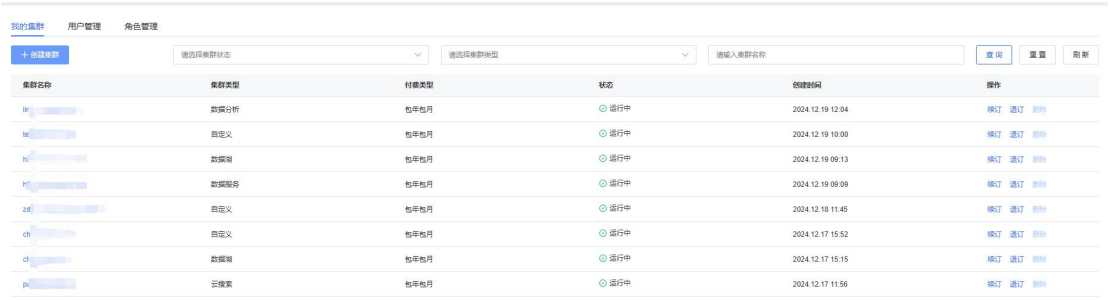
操作步骤

1、进入集群创建页面

方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。

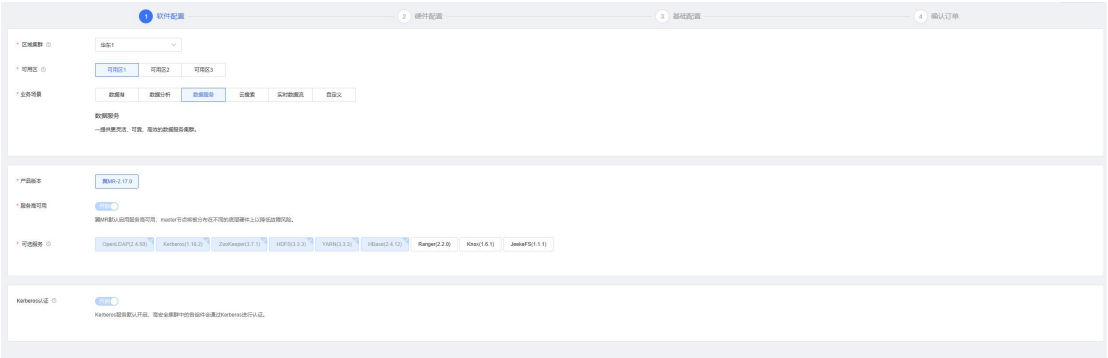


方法二：进入翼 MapReduce 产品一类节点资源池的产品控制台，点击“+创建集群”。

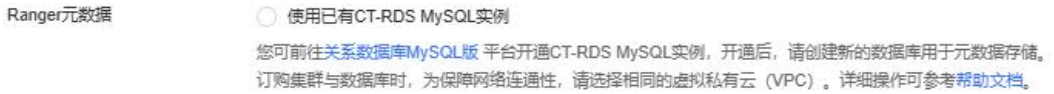


2、软件配置

进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：



- 区域集群：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。
- 业务场景：选择“数据服务”场景。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用且不可关闭高可用模式。
- 可选服务：由可选组件和必选组件组成，根据业务场景而定。您可根据自身业务场景对可选组件进行选择。
- 元数据：如果您选择添加 Ranger 等需要配置元数据库的组件服务，需先确认是否已有 CT-RDS MySQL 实例，若尚未创建，可点击提示链接，前往“关系数据库 MySQL 版”进行开通。若已创建实例，请勾选该选项并进行元数据配置。



- 元数据配置：勾选“使用已有 CT-RDS for MySQL”后，请在展开信息中填写所需的 5 项配置信息。

注意

1、当前仅支持通过内网地址进行数据库连接。请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库。

2、配置元数据库信息前，请先创建数据库，创建方式可参考创建数据库。

3、元数据配置所需信息可前往数据库控制台查看。

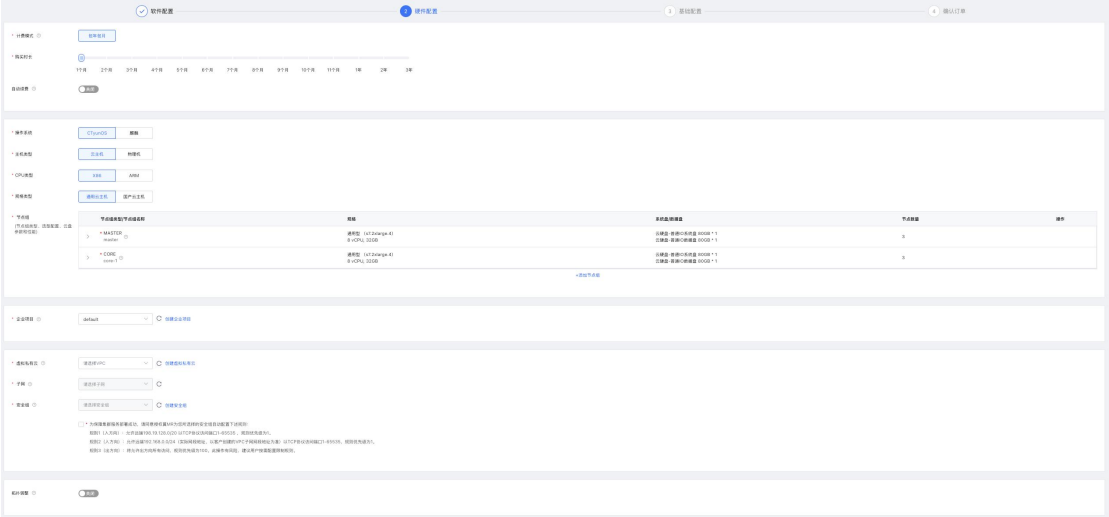
4、请确保配置的数据库用户具有该数据库的读写权限。

5、创建集群时，若元数据库配置错误，将导致服务异常，但不影响集群的创建与部署，集群创建后，您可前往 Manager 的集群服务，进入集群服务详情，通过运维操作中的“元数据库配置”操作，替换原有元数据库的配置信息。

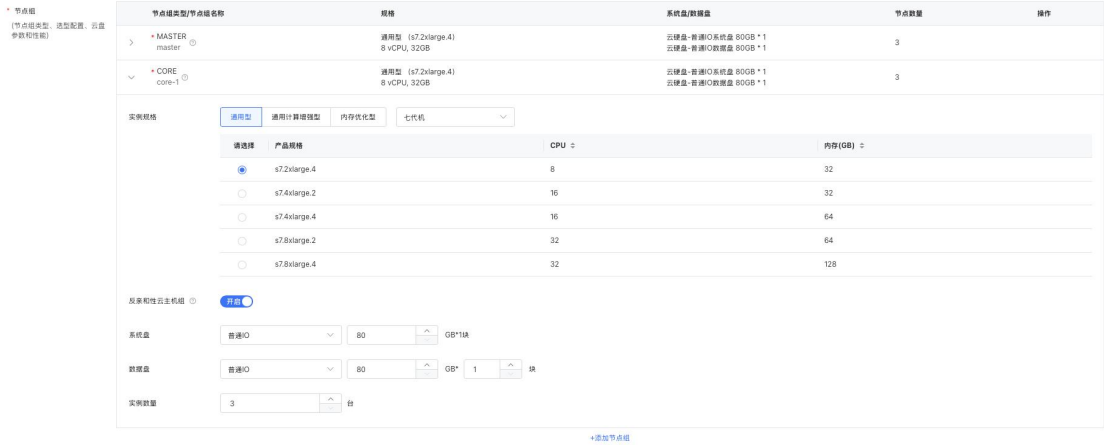
- Kerberos 身份认证：集群中的软件以 Kerberos 安全模式启动，此选项不可关闭。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如下：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 和 task 节点进行增加/删除。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。

- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

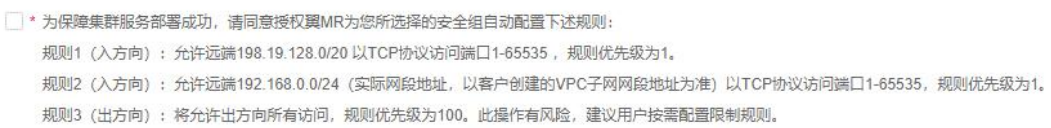
  注：

  1）集群和虚拟私有云需在同一企业项目下。

  2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

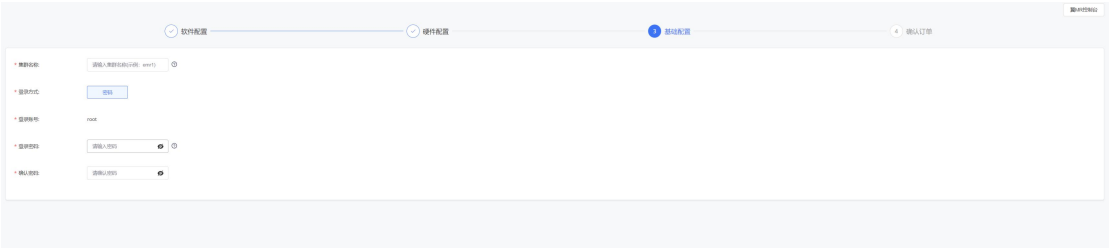
- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。



- 拓扑调整：从 V2.16.0 版本起，数据湖、数据服务、实时数据流与自定义场景支持拓扑调整功能，用户可以手动调整已选组件的各角色在节点的部署位置，平均分配各个节点上的资源。系统默认已按照各组件角色的部署原则自动分配实例，您可按需开启该功能进行配置调整。

4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下：



- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。
- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。配置清单说明如下：

- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

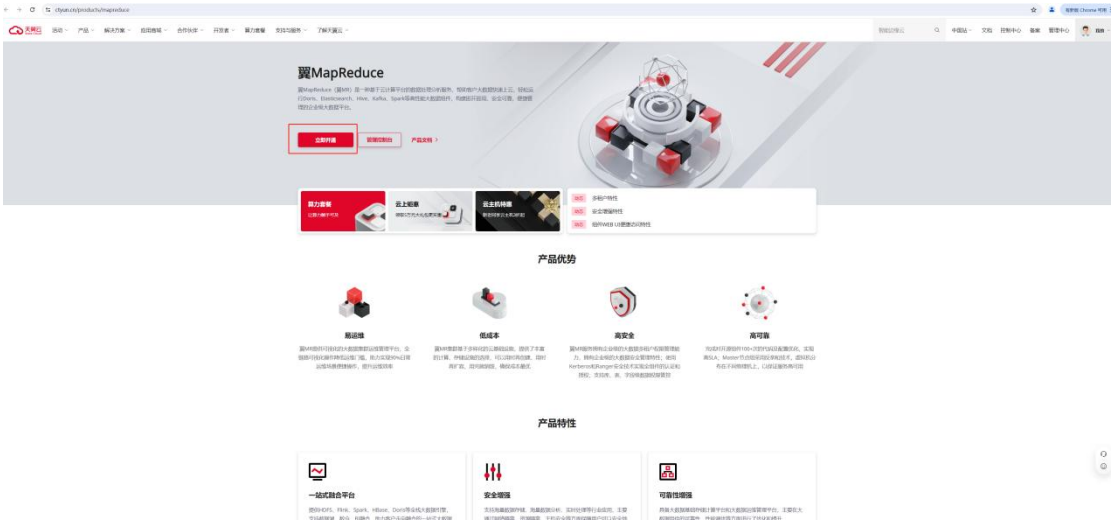
4.2.2 创建云搜索集群

云搜索集群为结构化/非结构化数据提供低成本、高性能及可靠性的检索、分析服务能力。在创建云搜索集群前，需要先创建虚拟私有云。

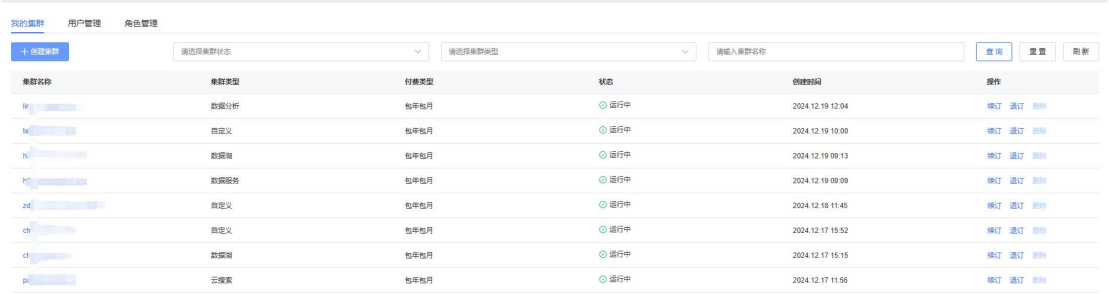
操作步骤

1、进入集群创建页面

方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。



方法二：进入翼 MapReduce 产品—类节点资源池的产品控制台，点击“+创建集群”。



2、软件配置

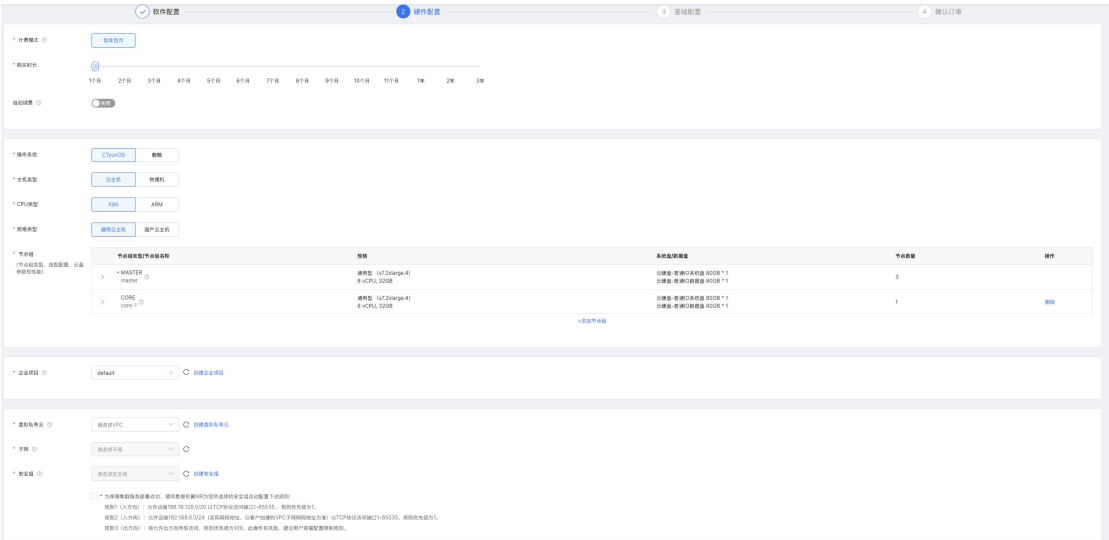
进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：

- 区域集群：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。

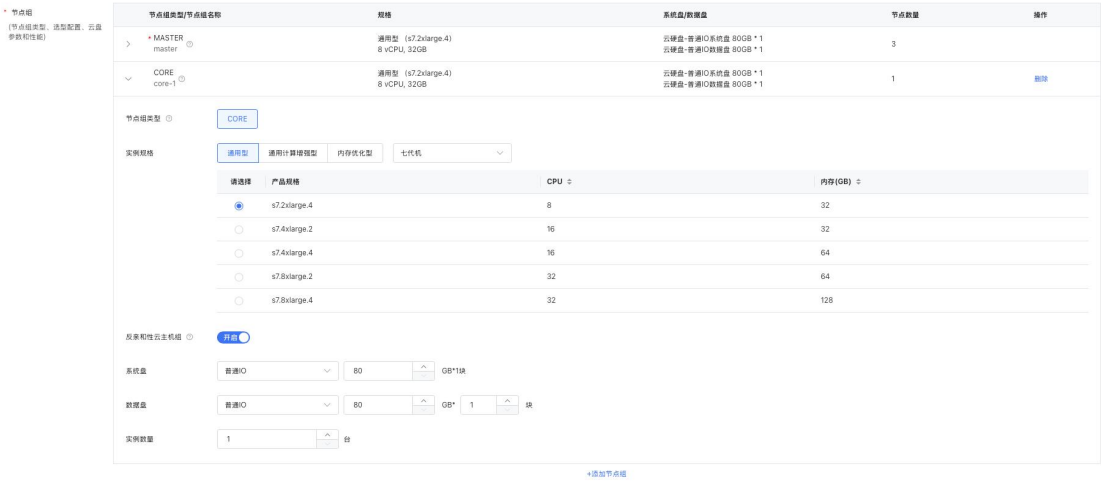
- 业务场景：选择“云搜索”场景。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用模式，且无法关闭。
- 可选服务：由可选组件和必选组件组成，根据业务场景而定。您可根据自身业务场景对可选组件进行选择。云搜索集群场景下默认可选组件如上图所示，浅蓝色表示必选。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如下：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 节点进行增加/删除。。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。

- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

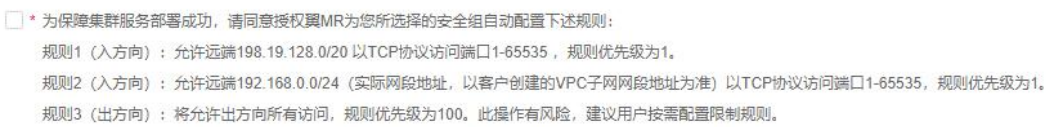
  注：

  1）集群和虚拟私有云需在同一企业项目下。

  2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

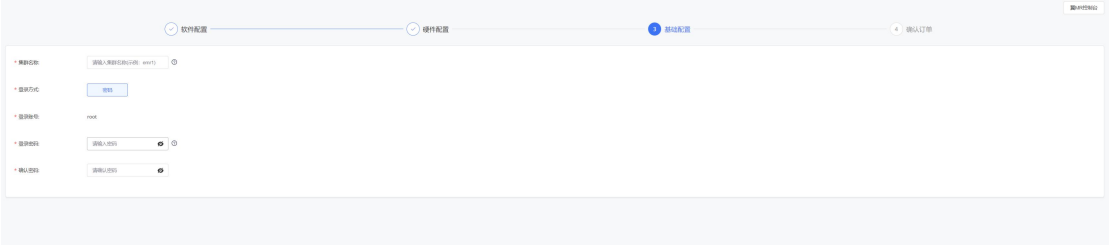
- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。



4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下：



- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。
- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。配置清单说明如下：

- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

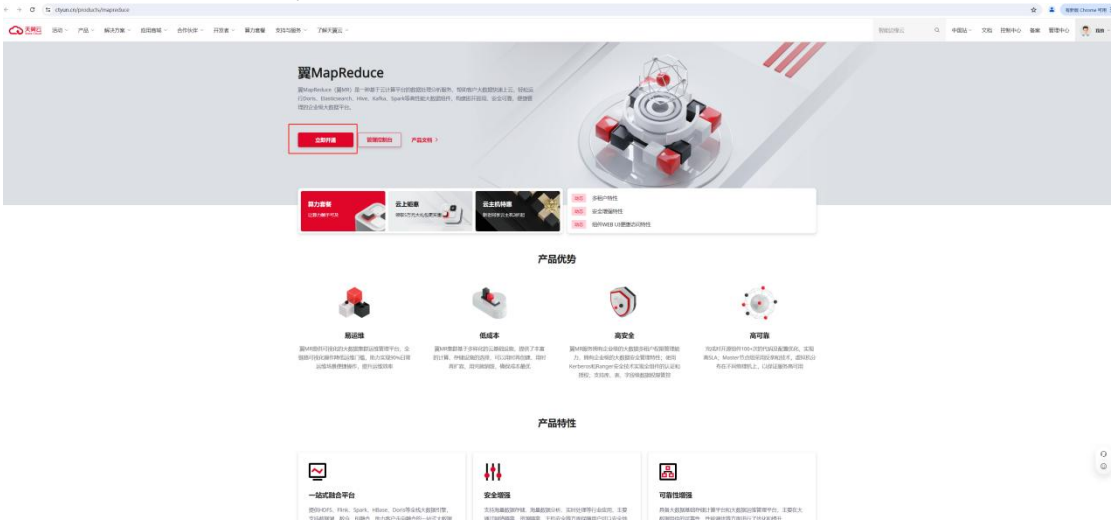
4.2.2 创建实时数据流集群

实时数据流场景提供更高效、灵活的管理集群，更快地运行大数据的实时计算引擎，更好地提供数据分析能力。在创建实时数据流集群前，需要先创建虚拟私有云。

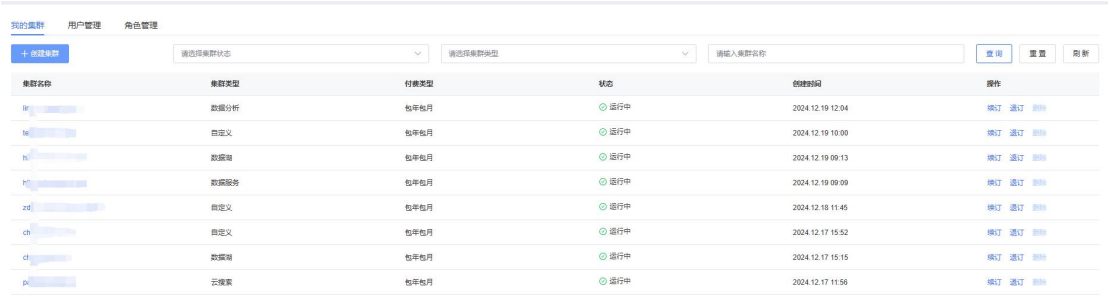
操作步骤

1、进入集群创建页面

方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。

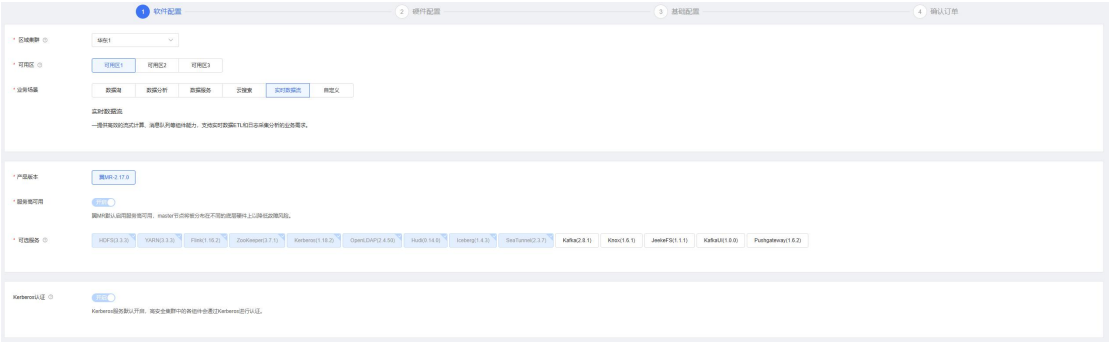


方法二：进入翼 MapReduce 产品一类节点资源池的产品控制台，点击“+创建集群”。



2、软件配置

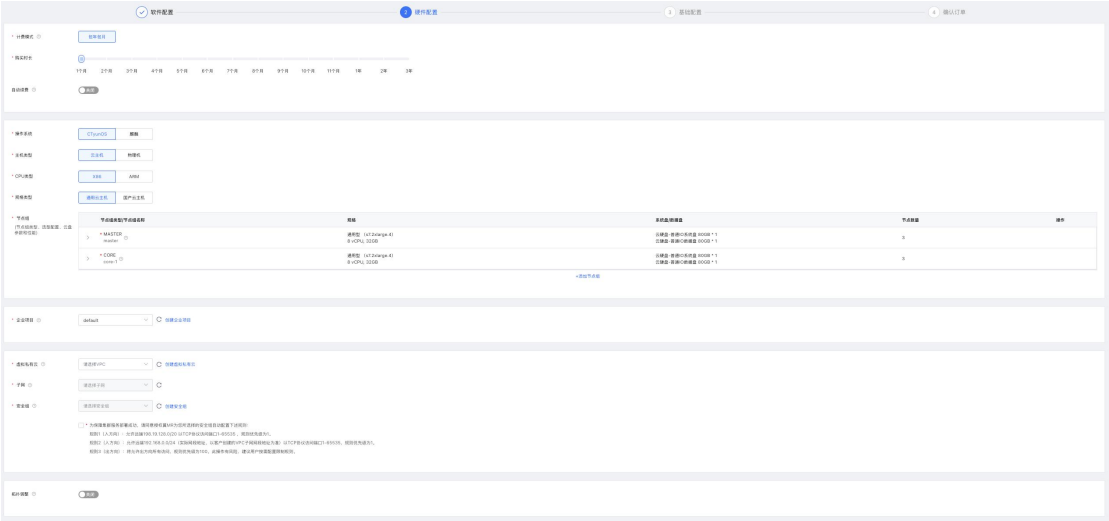
进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：



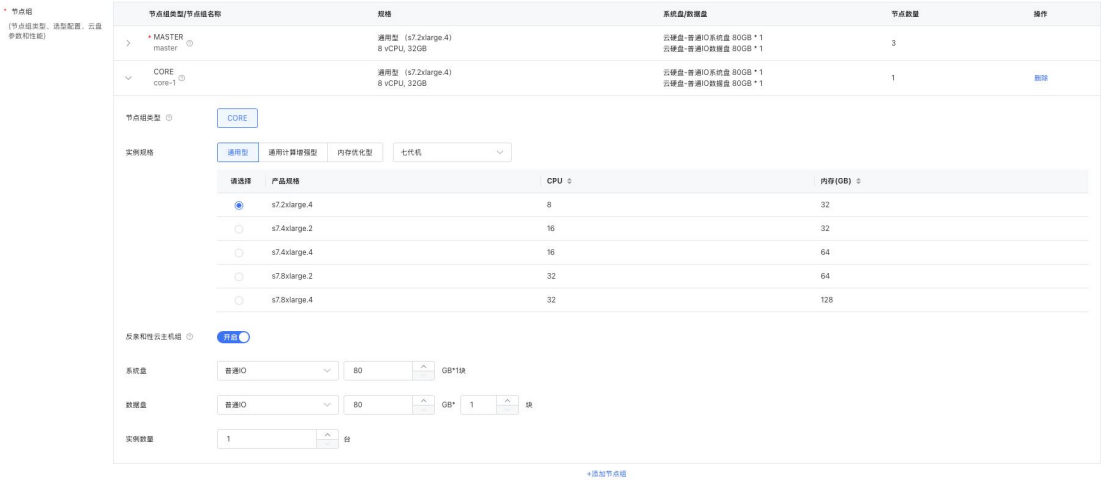
- 区域集群：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。
- 业务场景：选择“实时数据流”场景。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用且不可关闭高可用模式。
- 可选服务：由可选组件和必选组件组成，根据业务场景而定。您可根据自身业务场景对可选组件进行选择。
- Kerberos 身份认证：集群中的软件以 Kerberos 安全模式启动，此选项不可关闭。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如下：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 和 task 节点进行增加/删除。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。

- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

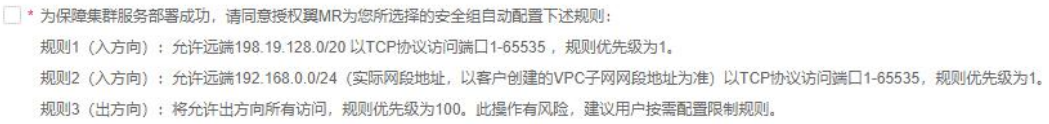
  注：

  1）集群和虚拟私有云需在同一企业项目下。

  2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

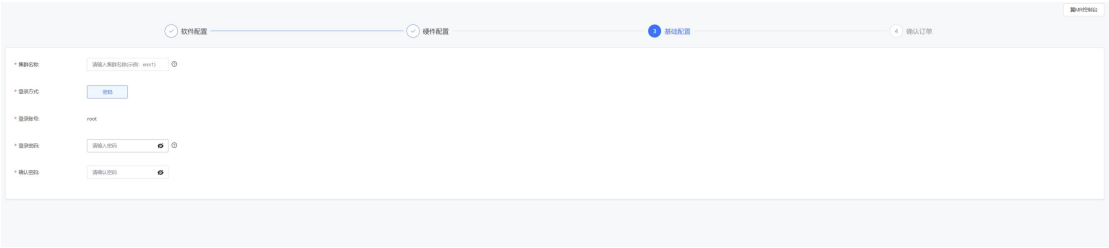
- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。



- 拓扑调整：从 V2.16.0 版本起，数据湖、数据服务、实时数据流与自定义场景支持拓扑调整功能，用户可以手动调整已选组件的各角色在节点的部署位置，平均分配各个节点上的资源。系统默认已按照各组件角色的部署原则自动分配实例，您可按需开启该功能进行配置调整。

4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下：



- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。
- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。配置清单说明如下：

- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

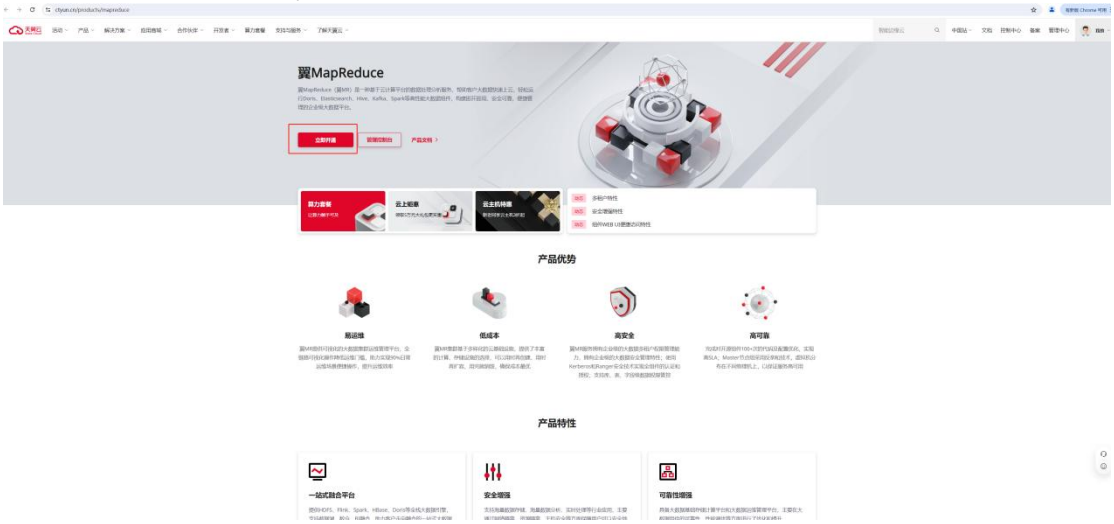
4.2.2 创建自定义集群

自定义集群提供丰富灵活的服务搭配，支持自行选择业务所需服务。建议不要将多个存储类服务部署在一个节点组上，避免资源争抢。

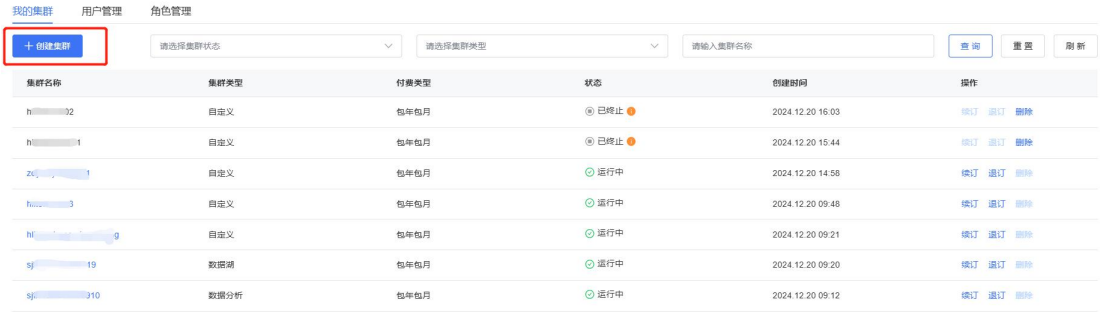
操作步骤

1、进入集群创建页面

方法一：登录翼 MapReduce 产品详情页，直接点击“立即开通”。

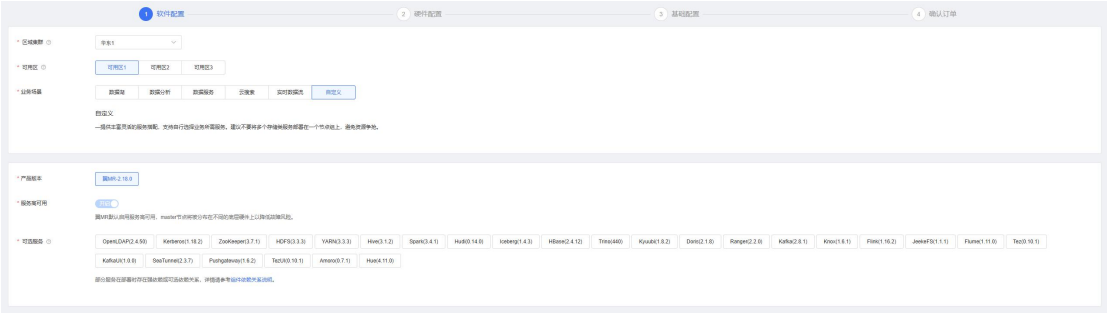


方法二：进入翼 MapReduce 产品一类节点资源池的产品控制台，点击“+创建集群”。



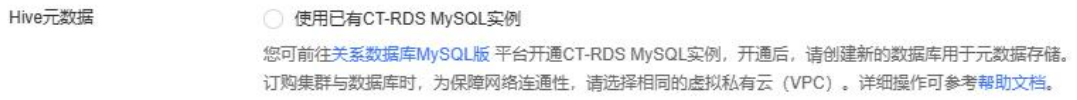
2、软件配置

进入“创建集群”页面进行配置与订购，软件配置页面如下图所示，参数说明如下：



- 区域集群与可用区：集群节点所在的物理位置，根据需要选择区域及可用区，也可以使用默认值。
- 业务场景：选择“自定义”场景。
- 产品版本：选择使用的产品版本，默认值即可。
- 服务高可用：翼 MapReduce 默认启用服务高可用且不可关闭高可用模式。
- 可选服务：您可基于业务诉求选择所需组件，部分组件存在依赖关系，请您按提示勾选。

- 元数据：使用 Hive、Ranger、Amoro 与 Hue 组件服务时，需要配置元数据库。请先确认是否已有 CT-RDS MySQL 实例，若尚未创建，可点击提示链接，前往“关系数据库 MySQL 版”进行开通。若已创建实例，请勾选该选项并进行元数据配置。



- Hive 元数据配置：勾选“使用已有 CT-RDS for MySQL”后，请在展开信息中填写所需的 5 项配置信息。

注意：

1、当前仅支持通过内网地址进行数据库连接。请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库。

2、配置 Hive、Ranger、Amoro 与 Hue 的元数据库信息前，请先创建数据库，若需要部署多个服务，请创建多个数据库，创建方式可参考创建数据库。

3、元数据配置所需信息可前往数据库控制台查看。

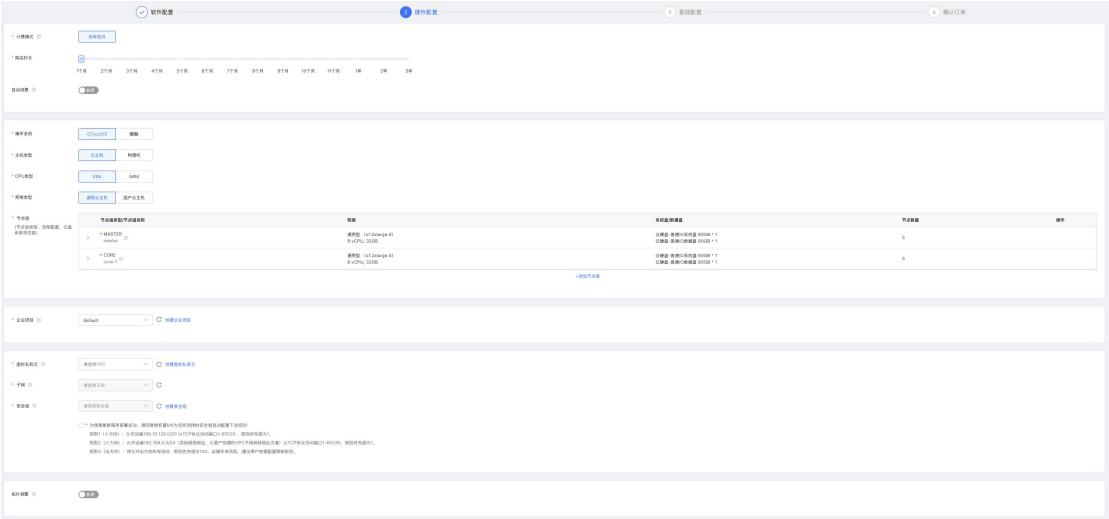
4、请确保配置的数据库用户具有该数据库的读写权限。

5、创建集群时，若元数据库配置错误，将导致 Hive/Ranger/Amoro/Hue 异常，但不影响集群的创建与部署，集群创建后，您可前往 Manager 的集群服务，进入集群服务详情，通过运维操作中的“元数据库配置”操作，替换原有元数据库的配置信息。

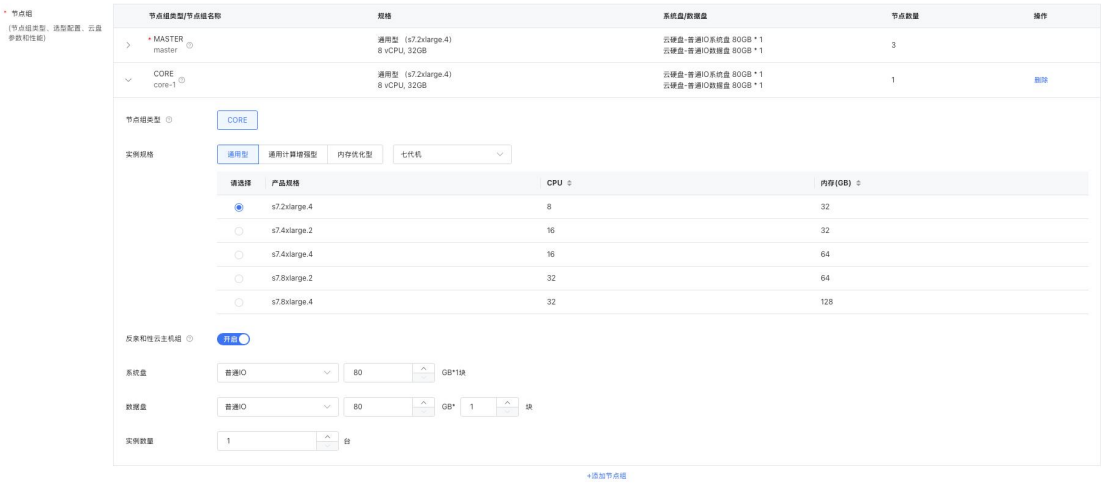
- Kerberos 身份认证：将依据所选组件，开启 Kerberos 安全模式。

3、硬件配置

软件配置选择完成后，点击“下一步”进入硬件配置页面。硬件配置页面如下图所示，参数说明如下：



- 计费模式：默认为包年包月。
- 购买时长：可按需选择订购时长。
- 自动续费：可按需开启自动续费功能。
- 操作系统：可按需选择 CTyunOS 或麒麟系统。
- 主机类型：可按需选择云主机或物理机。
- CPU 类型：可按需选择 X86 或 ARM。
- 规格类型：可按需选择通用主机或国产化主机。
- 节点组：根据您自身需要选择集群节点规格及数量，包括对节点组类型、选项配置、云盘参数和性能的选择，可根据需要对 core 和 task 节点进行增加/删除。



- 企业项目：企业项目提供统一的云资源管理能力，支持对项目及项目内的资源、成员进行管理。此处请根据用户的资源管理需求，选择翼 MR 集群的企业项目归属。仅支持选择用户有权限的企业项目。

- 虚拟私有云：不同虚拟私有云（VPC）网络之间的逻辑彻底隔离。请按需选择需要使用的虚拟私有云。如果目前没有 VPC 可以点击“创建虚拟私有云”跳转到虚拟私有云页面创建。

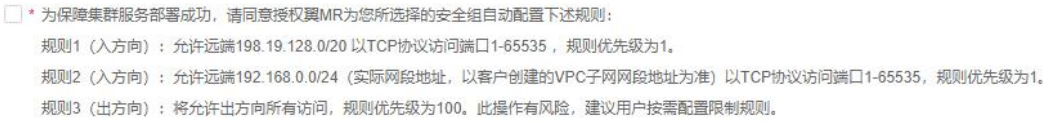
  注：

  1）集群和虚拟私有云需在同一企业项目下。

  2）为保障网络互通，请选择与软件配置中所填数据库相同的虚拟私有云（VPC）。

- 子网：选择虚拟私有云后，子网可以根据需要进行选择。若所选子网已开通 IPv6，可按需选择是否开启 IPv6 访问实例资源的功能。

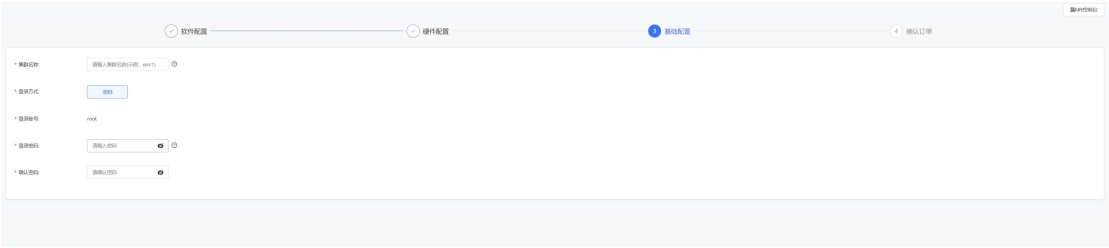
- 安全组：设置集群内实例的网络访问控制。当前天翼云虚拟私有云安全组策略强安全要求，默认服务器内网互相不通，需要客户勾选安全组规则自动配置授权，翼 MR 会默认添加下述安全组中相关的规则。



- 拓扑调整：从 V2.16.0 版本起，数据湖、数据服务、实时数据流与自定义场景支持拓扑调整功能，用户可以手动调整已选组件的各角色在节点的部署位置，平均分配各个节点上的资源。系统默认已按照各组件角色的部署原则自动分配实例，您可按需开启该功能进行配置调整。

4、基础配置

硬件配置选择完成后，左下角会计算当前配置需要费用。点击右下角“下一步”按钮进入基础配置页面。基础配置页面如下图所示，参数说明如下：



- 集群名称：集群名称可由大写字母、小写字母、数字及特殊符号（-_:/）组成，最大 28 字符。
- 登录方式：默认密码。
- 登录账号：用于登录弹性云主机的账号，默认为 root。
- 登录密码：用于登录弹性云主机的密码，密码长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符。
- 确认密码：与登录密码相同。

5、确认订单

基础配置填写完成后，点击右下角“下一步”按钮，进入确认订单页面。配置清单说明如下：

- 确认内容是否有误，点击画笔图标，即可回退到相应页面，修改集群信息。
- 勾选“我已阅读并同意《天翼云翼 MapReduce 服务协议》 《天翼云翼 MapReduce 服务等级协议》 ”，“立即购买”按钮变成高亮显示，点击并进入支付页面，购买后即可租用当前配置的集群。

4.2.3 管理数据连接

配置元数据库

翼 MapReduce 的数据连接是用来管理集群中组件使用的外部源连接，如 Hive、Ranger、Amoro、Hue 服务的元数据使用外部的关系型数据库，可以通过数据连接关联数据库与组件。

选择在翼 MR 集群部署 Hive、Ranger、Amoro、Hue 组件服务时，翼 MR 提供关系数据库 MySQL 版（CT-RDS MySQL）的元数据存储方式。在进行数据连接时，请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库，元数据将存储于关联的数据库中，不会随当前集群的删除而删除，多个翼 MR 集群可共享同一份元数据。

前置操作

1. 登录“关系数据库 MySQL 版”管理控制台。

2. 点击菜单栏“MySQL”—“实例管理”进入页面，选择希望翼 MR 连接的数据库实例。
3. 点击实例名称，进入实例详情页。选择“数据库管理”tab 页，单击“创建数据库”即可创建新的数据库。
4. 可按需前往“账号管理”页，点击“创建账号”，创建非 root 用户账号，并为该账号授予数据库的操作权限。

说明：

1、当前仅支持通过内网地址进行数据库连接。请选择关联与当前集群同一虚拟私有云（VPC）下的 MySQL 数据库。

2、配置元数据库信息时，请提前创建数据库，若需要部署多个服务，请创建多个数据库，创建方式可参考创建数据库。

3、元数据配置所需信息可前往数据库控制台查看。

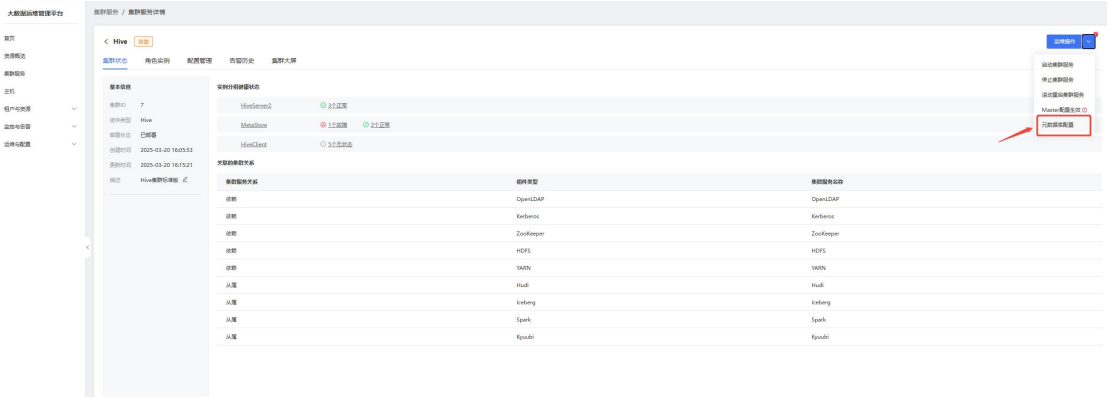
4、请确保配置的数据库用户具有该数据库的读写权限。

配置连接

在创建集群以及在集群服务管理内进行新增服务时，可能涉及 Hive、Ranger、Amoro、Hue 服务的元数据配置工作，您可通过填写前置操作中创建的数据库信息，进行数据连接。Hive、Ranger、Amoro、Hue 配置信息类型相同，所需参数可参考下图：

重置元数据库配置

创建集群时，若元数据库配置错误，将导致 Hive、Ranger、Amoro、Hue 异常，但不影响集群的创建与部署，集群创建后，您可前往 Manager 的集群服务，进入集群服务详情，通过运维操作中的“元数据库配置”操作，替换原有元数据库的配置信息。



配置存算分离

翼 MapReduce 产品基于 JeekeFS 组件服务提供存算分离能力，支持用户将数据存储在天翼云对象存储产品中，同时使用翼 MR 集群做数据计算处理，从而实现更灵活、低成本的数据分析方式。

集群准备

1、开始配置前，请确保购买的翼 MR 集群中已部署 JeekeFS 组件服务。目前，翼 MapReduce 产品主要支持在数据湖、数据服务和实时数据流场景下可选 JeekeFS 组件服务。

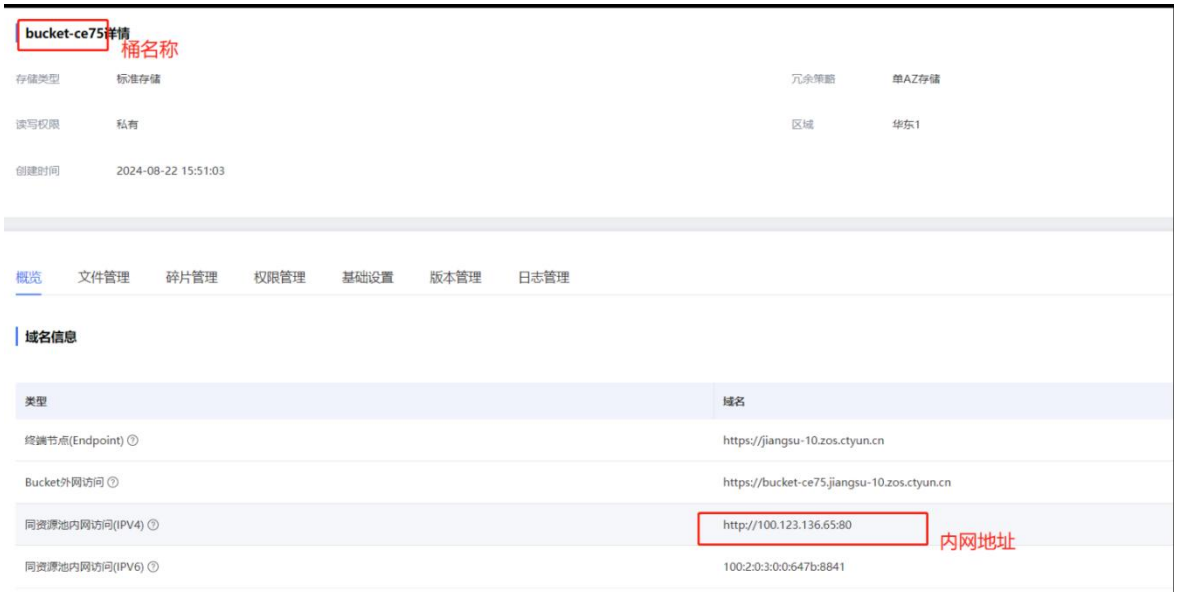
2、请确认已经在相关资源池区域中，购买开通好天翼云对象存储产品，详细操作流程请参照天翼云对象存储产品帮助文档开通对象存储服务章节内容

初始化文件系统

1、前往对象存储产品的存储控制台获取基础信息：Access Key、Security Key。详细操作流程请参照天翼云对象存储产品帮助文档获取访问密钥（AK/SK）中“获取对象存储 AK/SK”章节内容。

2、在对象存储产品中创建 Bucket。详细操作流程请参照天翼云对象存储产品帮助文档创建桶章节内容。

3、获取桶名称和内网地址，以下图为例，对象存储完整内网地址是 http://100.123.136.65/bucket-ce75（注意不要带上最后面的端口号“:80”）。



4、返回翼 MapReduce 产品控制台，进入待配置存算分离能力目标集群的翼 MR Manager 页面，并在集群服务—JeekeFS 的配置管理中获取元数据地址信息，将其格式修改成：tikv://【IP1】:7579，【IP2】:7579，【IP3】:7579 形式，参见下图：

5、通过参照登录集群节点章节内容，登录翼 MR 集群中部署 JeekeFS 组件服务元数据实例角色的节点，基于步骤 1~4 中获取的信息，手工初始化文件系统。

例：假定文件系统名字，s3 上的根目录名称都是 s3fs。

cd /usr/local/jeekefs/bin

./jeekefs format --bucket http://100.123.136.65/bucket-ce75 --access-key xxxxxxx --secret-key xxxxxxx tikv://192.168.0.27:7579/s3fs s3fs

说明：

需要根据实际的对象存储内网地址及 Bucket Name 拼接上述语句中“--bucket”参数的 Value 值。另外，“--access-key”、“--secret-key”和“tikv://192.168.0.27:7579”的值需要同样调整成实际的字符串。

命令执行结果示例：

```
2024/08/23 09:35:22.254885 jeekefs[196865] <INFO>: Meta address: tikv://192.168.0.27:7579/s3fs [interface.go:519]
2024/08/23 09:35:22.255004 jeekefs[196865] <INFO>: TiKV gc interval is set to 3h0m0s [tkv_tikv.go:147]
2024/08/23 09:35:22.271879 jeekefs[196865] <INFO>: Data use s3://bucket-ce75/s3fs/ [format.go:445]
2024/08/23 09:35:22.926268 jeekefs[196865] <INFO>: Volume is formatted as {
  "Name": "s3fs",
  "UUID": "9013872e-543a-45b9-97f[redacted]",
  "Storage": "s3",
  "Bucket": "http://100.123.136.65/bucket-ce75",
  "AccessKey": [redacted]
  "SecretKey": [redacted]
  "BlockSize": 4096,
  "Compression": "none",
  "EncryptAlgo": "aes256gcm-rsa",
  "KeyEncrypted": true,
  "TrashDays": 1,
  "MetaVersion": 1,
  "MinClientVersion": "1.1.0-A",
  "DirStats": true
} [format.go:482]
```

6、创建成功后，即可在对象存储平台看到和文件系统名字一样的目录：

验证方式

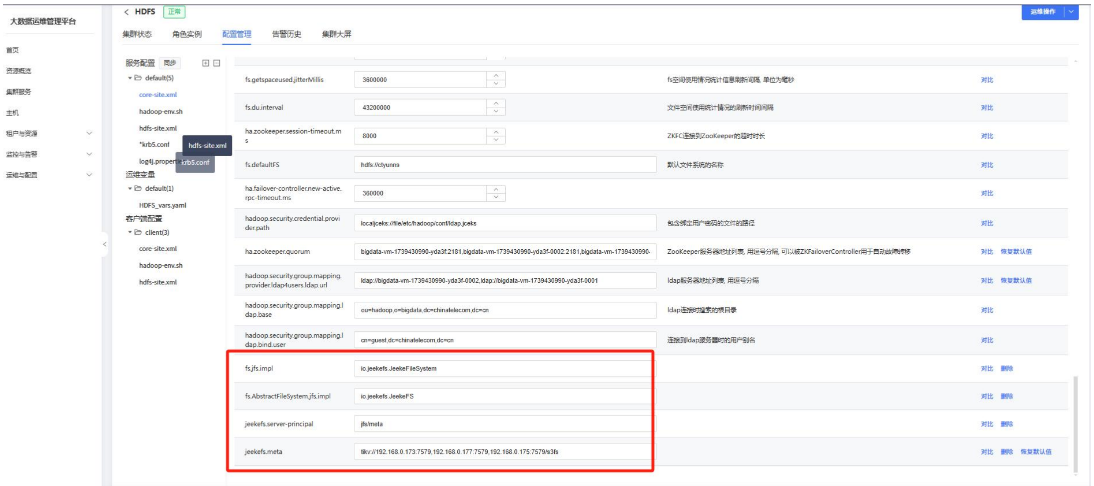
HDFS 集群配置

非缓存模式且混合存储的最小配置如下：

```
<property>
    <name>fs.jfs.impl</name>
    <value>io.jeekefs.JeekeFileSystem</value>
</property>
<property>
    <name>fs.AbstractFileSystem.jfs.impl</name>
    <value>io.jeekefs.JeekeFS</value>
</property>
<property>
    <name>jeekefs.server-principal</name>
    <value>jfs/meta</value>
    <description>元数据服务的 principle</description>
</property>
<property>
```

```
        <name>jeekefs.meta</name>
        <value>tikv://192.168.0.27:7579/s3fs</value>
    </property>
```

将上述的配置，通过使用翼 MR Manager 中的配置管理功能增加配置到翼 MR 集群 HDFS 服务 core-site.xml 配置中，并同步到各个节点，如图所示：



fs.getspaceused.jitterMillis	3600000	fs空间使用情况统计信息刷新间隔，单位为毫秒	对比
fs.du.interval	43200000	文件空间使用统计情况的刷新时间间隔	对比
ha.zookeeper.session-timeout.ms	8000	ZKFC连接到ZooKeeper的超时时长	对比
fs.defaultFS	hdfs://ctyunns	默认文件系统的名称	对比
ha.failover-controller.new-active.rpc-timeout.ms	360000		对比
hadoop.security.credential.provider.path	localjceks://file/etc/hadoop/conf/ldap.jceks	包含绑定用户密码的文件的路径	对比
ha.zookeeper.quorum	bigdata-vm-1739430990-yda3f:2181,bigdata-vm-1739430990-yda3f-0002:2181,bigdata-vm-1739430990-	ZooKeeper服务器地址列表，用逗号分隔，可以被ZKFailoverController用于自动故障转移	对比 恢复默认值
hadoop.security.group.mapping.provider.ldap4users.ldap.url	ldap://bigdata-vm-1739430990-yda3f-0002,ldap://bigdata-vm-1739430990-yda3f-0001	ldap服务器地址列表，用逗号分隔	对比 恢复默认值
hadoop.security.group.mapping.ldap.base	ou=hadoop,o=bigdata,dc=chinatelecom,dc=cn	ldap连接时搜索的根目录	对比
hadoop.security.group.mapping.ldap.bind.user	cn=guest,dc=chinatelecom,dc=cn	连接到ldap服务器时的用户别名	对比
fs.jfs.impl	io.jeekefs.JeekeFileSystem		对比 删除
fs.AbstractFileSystem.jfs.impl	io.jeekefs.JeekeFS		对比 删除
jeekefs.server-principal	jfs/meta		对比 删除
jeekefs.meta	tikv://192.168.0.173:7579,192.168.0.177:7579,192.168.0.175:7579/s3fs		对比 删除 恢复默认值

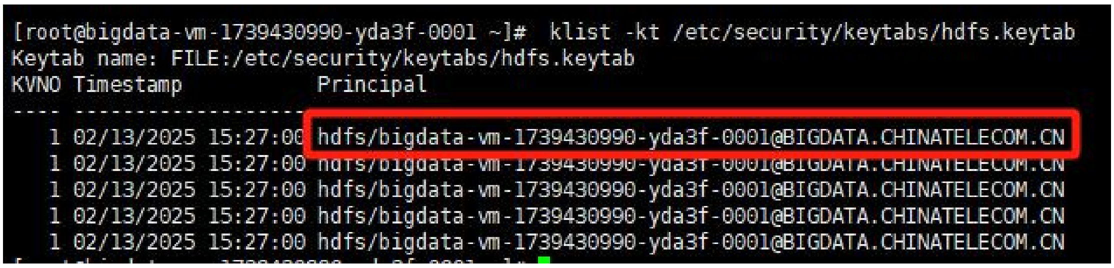
同步成功后，即可登录机器使用 HDFS Shell 命令进行测试。

HDFS Shell 测试

1、初始化 Token

1.1、获取 principal name：

```
klist -kt /etc/security/keytabs/hdfs.keytab
```



```
[root@bigdata-vm-1739430990-yda3f-0001 ~]#  klist -kt /etc/security/keytabs/hdfs.keytab
Keytab name: FILE:/etc/security/keytabs/hdfs.keytab
KVNO Timestamp           Principal
---- ------------------- ------------------------------------------------------
   1 02/13/2025 15:27:00 hdfs/bigdata-vm-1739430990-yda3f-0001@BIGDATA.CHINATELECOM.CN
   1 02/13/2025 15:27:00 hdfs/bigdata-vm-1739430990-yda3f-0001@BIGDATA.CHINATELECOM.CN
   1 02/13/2025 15:27:00 hdfs/bigdata-vm-1739430990-yda3f-0001@BIGDATA.CHINATELECOM.CN
   1 02/13/2025 15:27:00 hdfs/bigdata-vm-1739430990-yda3f-0001@BIGDATA.CHINATELECOM.CN
   1 02/13/2025 15:27:00 hdfs/bigdata-vm-1739430990-yda3f-0001@BIGDATA.CHINATELECOM.CN
```

1.2、kerberos 认证：

```
kinit -kt /etc/security/keytabs/hdfs.keytab 【上一步获取的 principal name】
```

例如：

```
kinit -kt /etc/security/keytabs/hdfs.keytab hdfs/bigdata-vm-1739430990-yda3f-0001@BIGDATA.CHINATELECOM.CN
```

2、访问原生的 HDFS

```
hdfs dfs -ls /
```

```
[root@bigdata-vm-1724327383-uawro-0002 ~]# hdfs dfs -ls /
Found 10 items
drwxrwxr-x   - hive  hadoop          0 2024-08-22 20:05 /apps
drwxrwxrwx   - flink hadoop          0 2024-08-22 20:05 /flink-history
drwxrwxr-x   - hbase hadoop          0 2024-08-22 20:11 /hbase
drwxrwxrwx   - spark hadoop          0 2024-08-22 20:05 /spark3-history
drwxrwxrwt   - yarn  hadoop          0 2024-08-22 20:04 /staging
drwxr-xr-x   - yarn  hadoop          0 2024-08-22 20:06 /system
drwxrwxrwt   - hdfs  hadoop          0 2024-08-22 20:08 /tmp
drwxrwxrwt   - hdfs  hadoop          0 2024-08-22 20:04 /user
drwxrwxr-x   - hive  hadoop          0 2024-08-22 20:05 /warehouse
drwxrwxr-x   - yarn  hadoop          0 2024-08-22 20:04 /yarn
```

3、访问 JeekeFS，并进行文件 CRUD 操作

hdfs dfs -ls jfs://s3fs/

```
[root@bigdata-vm-1724327383-uawro-0002 ~]# hdfs dfs -ls jfs://s3fs/
[root@bigdata-vm-1724327383-uawro-0002 ~]# echo 11111 > test.txt
[root@bigdata-vm-1724327383-uawro-0002 ~]# hdfs dfs -put test.txt  jfs://s3fs/
[root@bigdata-vm-1724327383-uawro-0002 ~]# hdfs dfs -ls jfs://s3fs/
Found 1 items
-rw-r--r--   1 hdfs hdfs          6 2024-08-23 10:45 jfs://s3fs/test.txt
[root@bigdata-vm-1724327383-uawro-0002 ~]# hdfs dfs -cat jfs://s3fs/test.txt
11111
```

```
[root@bigdata-vm-1724327383-uawro-0002 ~]# hdfs dfs -rm jfs://s3fs/test.txt
2024-08-23 10:47:43,295 | INFO | Configuration.deprecation | io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2024-08-23 10:47:43,323 | INFO | fs.TrashPolicyDefault | Moved: 'jfs://s3fs/test.txt' to trash at: jfs://s3fs/user/hdfs/.Trash/Current/test.txt
```

Spark 测试

1、准备工作

kinit -kt /etc/security/keytabs/hdfs.keytab 获取到的实际 principalname

Hive Metastore 收到元数据请求需要识别 JeekeFS，需重启 YARN-ResourceManager 和 Hive-MetaStore 服务。

注意：

翼 MR 集群默认都开启 kerbero 认证。Spark 将应用交给 YARN 之前，应用程序只会初始化 defaultFS 所指向文件系统的安全认证。如果后续作业需要访问其他安全文件系统，会抛出认证不通过的异常。因此，我们这里需要将其他文件系统显式地提供给 Spark 客户端侧配置参数 spark.kerberos.access.hadoopFileSystems=jfs://s3fs。

2、进入 YARN 的 Shell

spark-sql --master yarn --conf spark.yarn.access.hadoopFileSystems=jfs://s3fs

3、创建一个数据库

create database if not exists jfstest location 'jfs://s3fs/jfstest';

4、创建一个数据表

create table jfstest.orcbtb(name string,age int) stored as orc;

5、执行成功后，查看表详情，关注是否是内部表，路径是否在 Jeekefs 上

```
spark-sql> describe extended  jfstest.orcbtb;
name                    string
age                     int

# Detailed Table Information
Database                jfstest
Table                   orcbtb
Owner                   hdfs
Created Time            Fri Aug 23 11:10:34 CST 2024
Last Access             UNKNOWN
Created By              Spark 3.3.3
Type                    MANAGED
Provider                ORC
Location                jfs://s3fs/jfstest/orcbtb
Serde Library           org.apache.hadoop.hive.ql.io.orc.OrcSerde
InputFormat             org.apache.hadoop.hive.ql.io.orc.OrcInputFormat
OutputFormat            org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat
Time taken: 0.14 seconds, Fetched 16 row(s)
```

6、执行增删改查操作

insert into jfstest.orcbtb values('tty',3);

insert into jfstest.orcbtb values('tty',4);

select * from jfstest.orcbtb;

7、清空表

TRUNCATE TABLE jfstest.orcbtb;

8、删除表

drop table jfstest.orcbtb;

说明：

由于路径是 jfs 开头，如果直接创建表，Spark 会识别为外部表。这里先创建数据库，再创建表，才是内部表。只有内部表，删除数据后，才会最终删除 S3。

4.3 管理集群

4.3.1 登录集群

4.3.1 翼 MapReduce 集群节点简介

远程登录

翼 MR 集群节点支持用户远程登录，远程登录包含界面登录、SSH 登录两种方式：

- 界面登录：直接通过翼 MR 管理控制台->节点管理页面中提供的远程登录功能，登录到集群中各个节点的 Linux 界面。
- SSH 登录：仅适用于 Linux 弹性云主机。您可以使用远程登录工具（例如 PuTTY、XSHELL），登录弹性云主机。此登录方式，需要前置依赖绑定弹性 IP 地址，详细步骤请参见绑定/解绑弹性 IP。

✧ **说明**

当您使用密码方式访问集群节点，详细步骤请参见登录集群节点（SSH 密码方式）。

节点类型和功能

翼 MR 集群中每个节点即为一台主机，节点类型及节点功能如下表所示：

节点类型	功能
master 节点	翼 MR 集群管理节点，负责管理和监控集群。在翼 MR 管理控制台选择“我的集群”，选中一个运行中的集群并单击集群名，进入集群信息页面。在“节点管理”中查看节点信息。 节点可以通过点击”远程登录“选择界面的 VNC 方式登录，也可以通过 SSH 方式登录。
core 节点	翼 MR 集群工作节点，负责处理和分析数据，并存储过程数据。
task 节点	翼 MR 集群计算节点，主要负责计算数据，不存储数据（如 HDFS 数据）。默认不开启，按需使用。

4.3.1 登录集群节点

本章节介绍如何使用翼 MR 管理控制台上提供的节点远程连接（VNC 方式）和如何使用密码方式（SSH 方式）登录翼 MR 集群中的节点，远程连接主要用于紧急运维场景，远程登录弹性云主机进行相关维护操作。其他场景下，优先推荐用户使用 SSH 方式登录集群节点。

✧ **说明**

如果需要使用 SSH 方式登录集群节点，需要在集群的安全组规则中手动添加入方向规则：其中源地址为“客户端公网 IPV4 地址/32”，端口为 22，具体请参见“帮助中心>虚拟私有云>安全组>[添加安全组规则]”。

登录主机（VNC 方式）

1. 登录翼 MapReduce 服务管理控制台。
2. 选择“集群列表 > 我的集群”，选中一个运行中的集群并单击集群名，进入集群基本信息页面。
3. 选择“节点管理”，单击展开节点组，显示节点列表。

4. 在目标节点的右侧“操作”列中，单击“远程连接”。
5. 在 VNC 方式新标签页 Web 页面中，可以看到服务器的命令行界面，类似于 Linux 服务器登录模式，输入 root 账号，并输入密码。

✧ **说明**

1、初始密码为创建集群时所设定的集群密码。

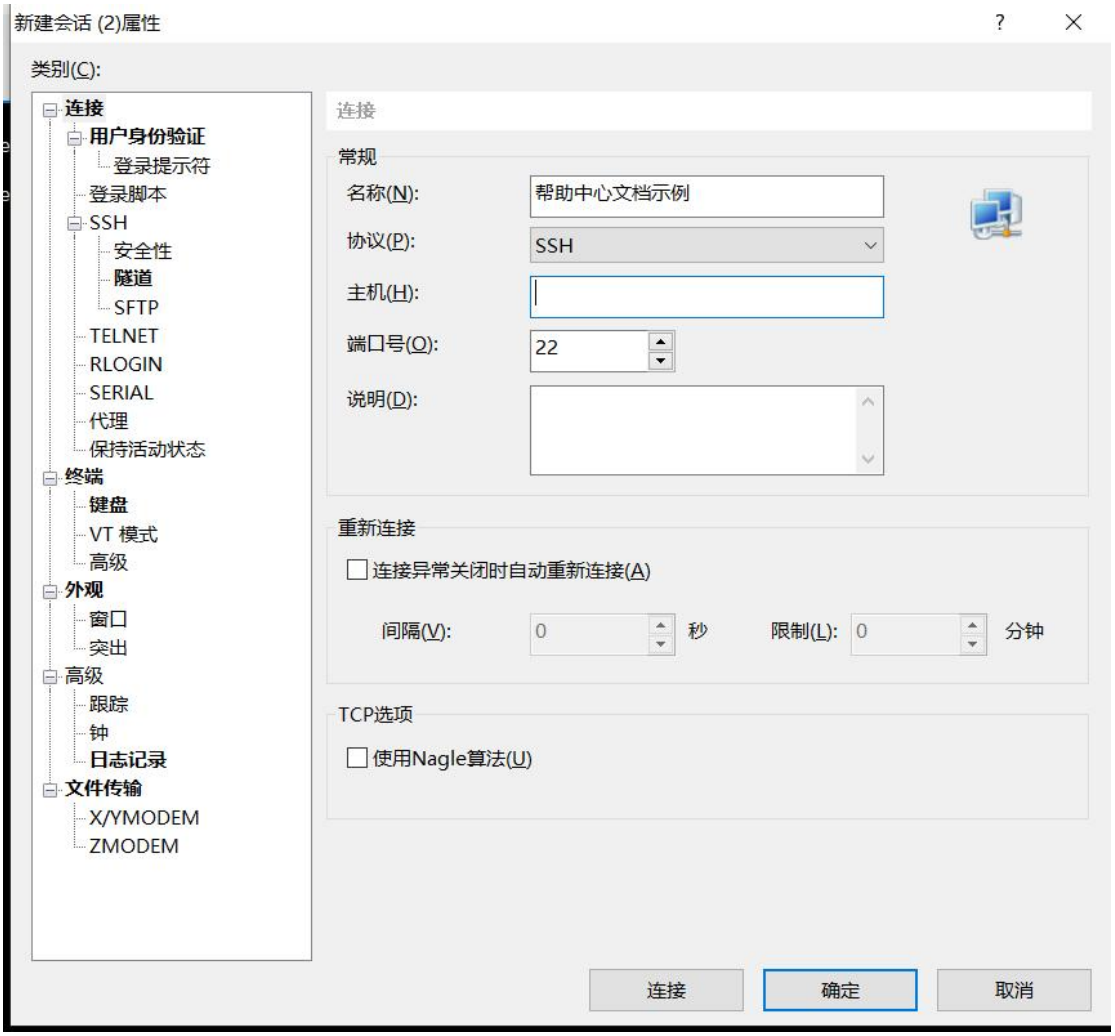
2、自 2.16 版本起，若忘记密码，可通过基础信息页的重置密码功能修改云主机密码，物理机暂不支持修改。

登录主机（SSH 密钥方式）

本地使用 Windows 操作系统

如果您本地使用 Windows 操作系统登录 Linux 主机，可以按照下面方式登录主机。下面步骤以 Xshell 为例。

1. 登录翼 MR 服务管理控制台。
2. 选择“集群列表 > 我的集群”，选中一个运行中的集群并单击集群名，进入集群基本信息页面。
3. 选择“节点管理”，单击展开节点组，显示节点列表。
4. 在目标节点的右侧“操作”列中，单击“更多”，选择“绑定弹性 IP”，可以为该节点选择一个已存在且未绑定使用的弹性公网 IP 或者可以单击“+创建弹性公网 IP”跳转至弹性 IP 控制台完成弹性 IP 购买后，再完成弹性公网 IP 绑定操作，若已绑定弹性公网 IP 请跳过该步骤。
5. 运行 Xshell。
6. 选择“新建会话”。
7. 名称：选择你认为合适的会话名称，用于方便管理。
8. 协议：使用默认的“SSH”。
9. 主机：输入主机所绑定的弹性公网 IP。
10. 端口号：使用默认的“22”，详见下图： 单击“Session”。



11. 单击“连接”，按照提示步骤，分别完成登录用户（默认为 root）、密码的输入并保存，即可完成远程登录操作。

本地使用 Linux 操作系统

如果您本地使用 Linux 操作系统登录 Linux 主机，可以按照下面方式登录。

执行如下命令，登录主机。

```
ssh 默认用户名@弹性公网 IP
```

假设 Linux 主机的默认用户名是 root，弹性公网 IP 为 123.123.123.123，则命令如下：

```
ssh  root@123.123.123.123
```

4.3.2 集群概览

4.3.2 集群列表简介

登录翼 MapReduce 控制台后，可查看账号下拥有的集群列表和当前的集群状态。

集群状态

集群包含的状态如下表所示。

状态	说明
启动中	集群正在创建中，可以查看具体的进程。
运行中	集群创建成功且集群中组件状态处于正常运行中，集群正在正常运行。
异常终止	集群启动失败或因为异常原因而终止，可以查看失败原因。
释放中	表示集群正在退订。
已终止	集群已经到期 15 天以上或者集群已经完成退订。
已冻结	已经到期的集群，资源会保留 15 天，此时状态为已冻结。

4.3.2 查看集群状态

集群创建完成后，登录翼 MapReduce 控制台后，可查看账号下所有集群列表和集群状态。

在“我的集群”中，显示所有您创建过的集群，默认页面上最多显示 10 条集群信息，当您集群数量较多时，可以修改页面右下角当前页面最多显示的集群条数，也可以翻页进行查看。

集群列表参数说明：

参数	参数说明
集群名称	用户自己定义的集群名称。
集群类型	用户创建集群时选择的集群类型。
付费类型	默认为包年包月。
企业项目	集群所属的企业项目。

参数	参数说明
状态	集群当前的状态信息。 启动中：集群正在创建中，通过“查看进程”查看具体的进程。 运行中：集群创建成功且集群中组件状态处于正常运行中，集群正在正常运行。 异常终止：集群启动失败或因为异常原因而终止，通过“失败原因”查看终止原因。 释放中：集群正在退订。 已终止：集群已经到期 15 天以上或者集群已经完成退订。 已冻结：已经到期的集群，资源会保留 15 天，此时状态为已冻结。
创建时间	集群创建的时间。

4.3.2 查看集群基本信息

登录翼 MapReduce 控制台后，进入集群列表页面，点击需要查看集群的名称，即进入集群基本信息界面。

集群基本信息参数主要包括集群信息、软件信息和网络硬件。

集群信息

参数	说明
集群名称	集群的名称。
集群 ID	集群的 ID
创建时间	集群创建的时间。
运行时间	集群运行的时间。
到期时间	集群到期时间。
高可用	默认开启。

参数	说明
付费类型	默认包年包月。
Kerberos 认证	根据业务场景需要，默认开启。
登录方式	默认为密码，V2.16 版本起支持密码重置操作。

软件信息

参数	说明
产品版本	集群使用的翼 MapReduce 版本。
软件信息	集群创建时使用的组件及其版本号。单击需要查看的组件名称，可以进入该组件详情信息。
业务场景	集群创建时选择的业务场景。
客户端	集群涉及组件的客户端安装包、MD5 值和配置文件。当前支持下载客户端的场景有 4 个，分别是：数据湖、数据服务、实时数据流和自定义。

网络硬件

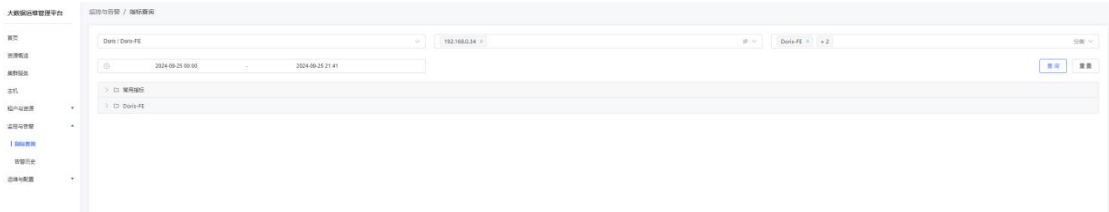
参数	说明
区域集群	集群所在的区域。
硬件配置	集群所有节点 MASTER、CORE、TASK（如有）的硬件信息。
企业项目	集群所属的企业项目。
虚拟私有云（VPC）	集群所使用的虚拟私有云（VPC），单击 复制 VPC ID，单击内容可以跳转并查看该 VPC 详细信息。
安全组	集群所使用的安全组，单击 复制安全组 ID，单击内容可以跳转并查看安全组详细信息。
子网	集群所使用的子网，单击 复制子网 ID，单击内容可以跳转并查看子网详

参数	说明
	细信息。
操作系统	集群所用的操作系统。
主机类型	集群所用主机的类型。
CPU 类型	集群所用主机的 CPU 类型。
IPv6 访问	V2.15 版本起，支持展示集群是否开启 IPv6 访问能力。

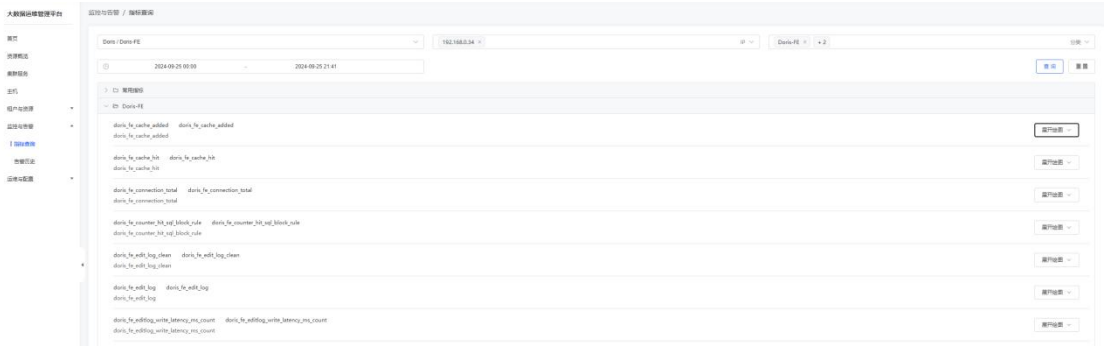
4.3.2 查看集群监控指标

操作步骤

1. 登录翼 MR 控制台。
2. 选择“我的集群 ”，单击集群名称进入集群详情页面。
3. 在集群详情页，选择“翼 MR Manager”点击“前往翼 MR Manager”进入运维管理平台。
4. 进入到翼 MR Manager 以后，选择“监控与告警 ＞ 指标查询” ，即可查看集群监控指标。
5. 在“所属集群服务”选择需要查看的集群。
6. 在“IP”处选择待查询的相关角色实例的节点 IP，可复选。
7. 在“分类”处选择需要设置的指标分类。默认勾选常用指标分类，可自主再添加其他选项，可供选择的选项如下：进程基础、线程基础、网络、CPU、磁盘、内存、edac、文件描述符、系统、文件系统、进程、socket、软中断、时钟、虚拟内存、xfs 文件系统、ARP。
8. 在“时间范围”选择需要查看监控数据的时间段。可供选择的选项如下：昨天、上周、本周、上月、本月、去年、自定义。
9. 点击“查询”并显示所选指标分类。



10. 选择任意一种指标分类，点击左侧箭头展开，显示某一指标分类下的所有指标。

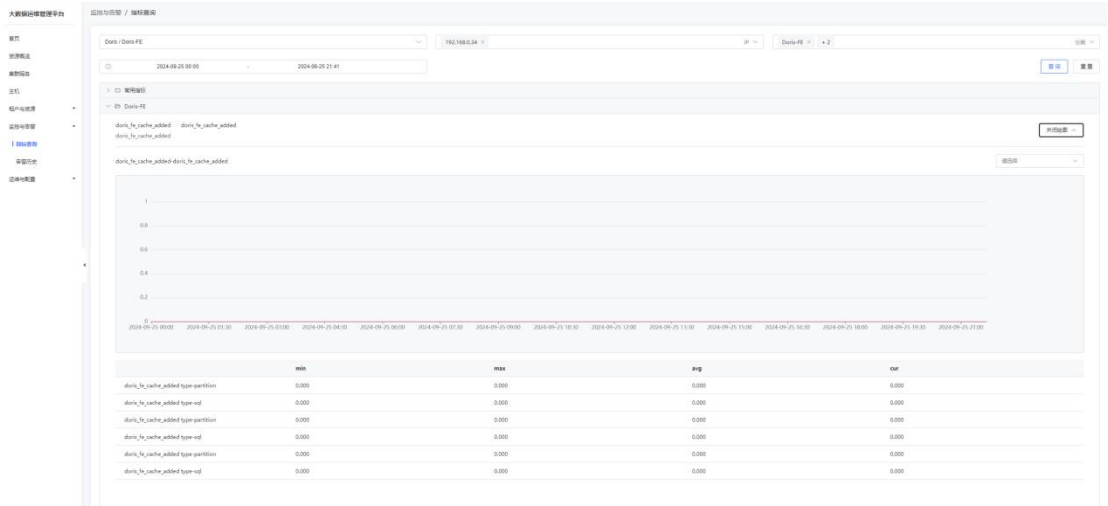


✧ **说明**

常用指标类别共计 18 项，包含如下指标：

- 网卡接收比特率
- 网卡接收丢包率
- 网卡接收错误包率
- 网卡发送比特率
- 网卡发送丢包率
- 网卡发送错误包率
- CPU 使用率
- 磁盘读取平均用时
- 磁盘 IO 使用率
- 磁盘写入平均用时
- 磁盘 IO 操作平均用时
- 内存空闲率
- 1 分钟平均负载
- 5 分钟平均负载
- 15 分钟平均负载
- 磁盘 inode 空闲率
- 磁盘容量空闲率
- Blocked 状态进程数

11. 选择任一类别下的某一指标，点击右侧的“展开绘图”即查看对应集群的监控指标。



4.3.2 管理组件和主机监控

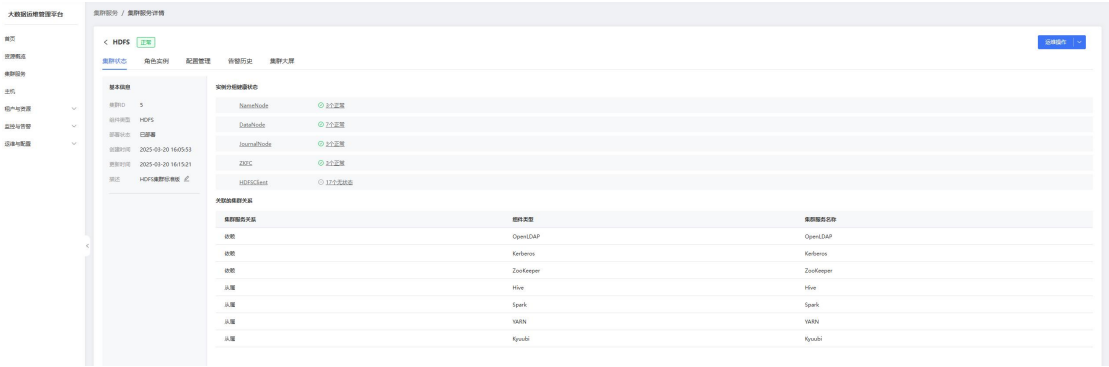
用户在日常使用中，可以在翼 MR Manager 管理所有组件（含角色实例）和主机的状态及指标信息：

- 状态信息，包含运行、健康及角色实例状态统计。
- 指标信息，各组件的主要监控指标项。

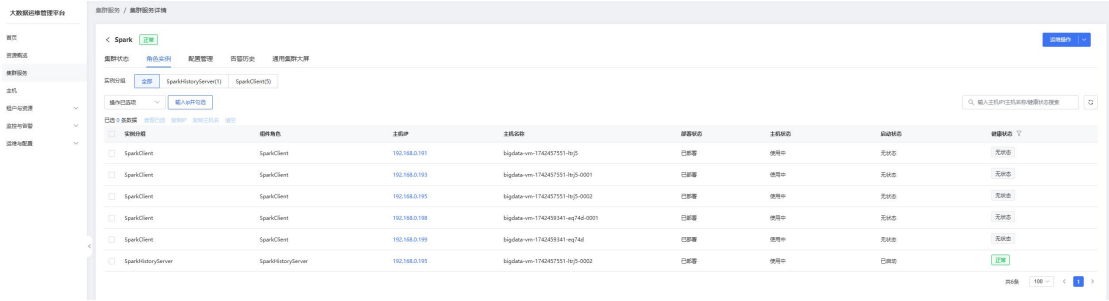
管理组件监控

1. 在翼 MR Manager 页面，点击左侧菜单栏“集群服务”。
2. 在集群服务列表点击任一集群进入。
3. 选择“集群状态”页，展示实例分组健康状态。

状态	描述
告警	集群实例服务曾出现过异常。
故障	集群实例服务产生故障。
正常	集群实例服务已启动。
未监控	集群实例未安装 prometheus，服务未监控。



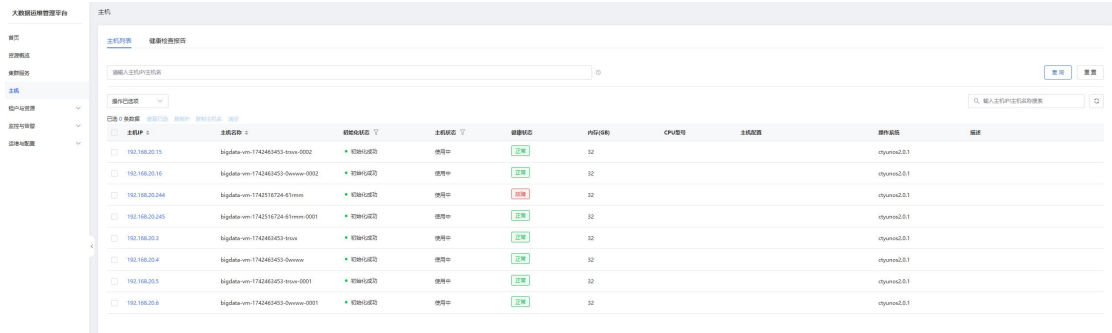
4. 点击任一实例的健康状态后，跳转到角色实例页面，按实例分组显示每个实例组件角色，主机 IP，主机名称，部署状态，主机状态，启动状态和健康状态。



管理主机监控

1. 在翼 MR Manager 页面，点击左侧菜单栏“主机”。

2. 主机页展示主机 IP、主机名称、初始化状态、主机状态、健康状态、内存、CPU 型号、主机配置、操作系统、描述等字段。

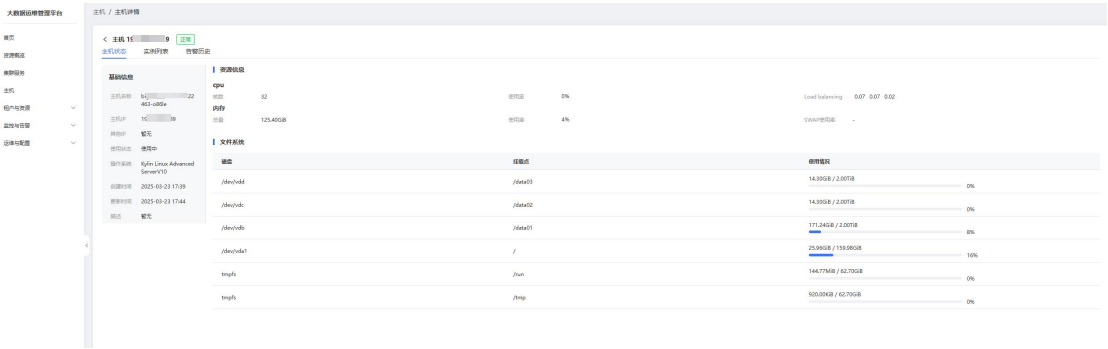


3. 健康状态栏，展示主机健康状态。

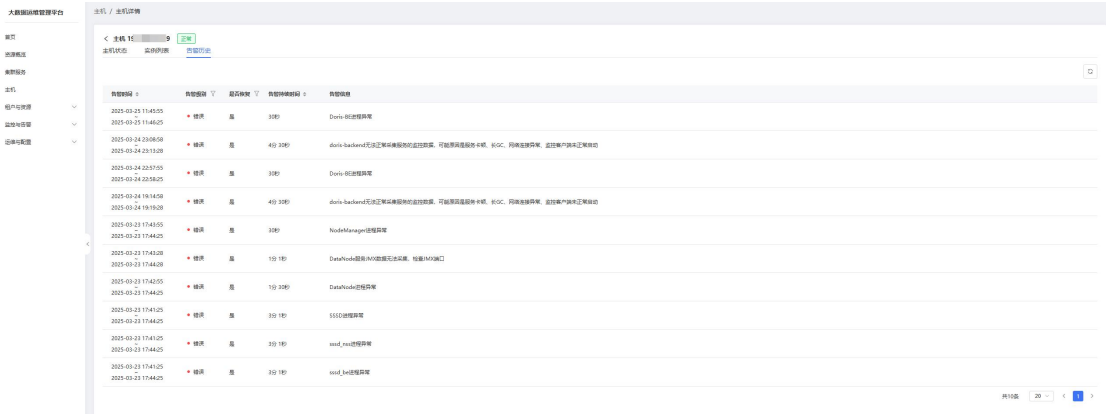
状态	描述
告警	集群实例服务曾出现过异常。
故障	集群实例服务产生故障。
正常	集群实例服务已启动。

状态	描述
未监控	集群实例未安装 prometheus，服务未监控。

4. 点击主机 IP，跳转到主机状态信息栏，展示主机状态、实例列表和告警历史。



5. 点击告警历史，可追溯该主机 IP 下的历史告警。



4.3.2 资源洞察

用户可通过资源洞察功能，快速了解资源的使用情况，如库表数量、存储量、文件量、日增情况、冷热文件信息等。

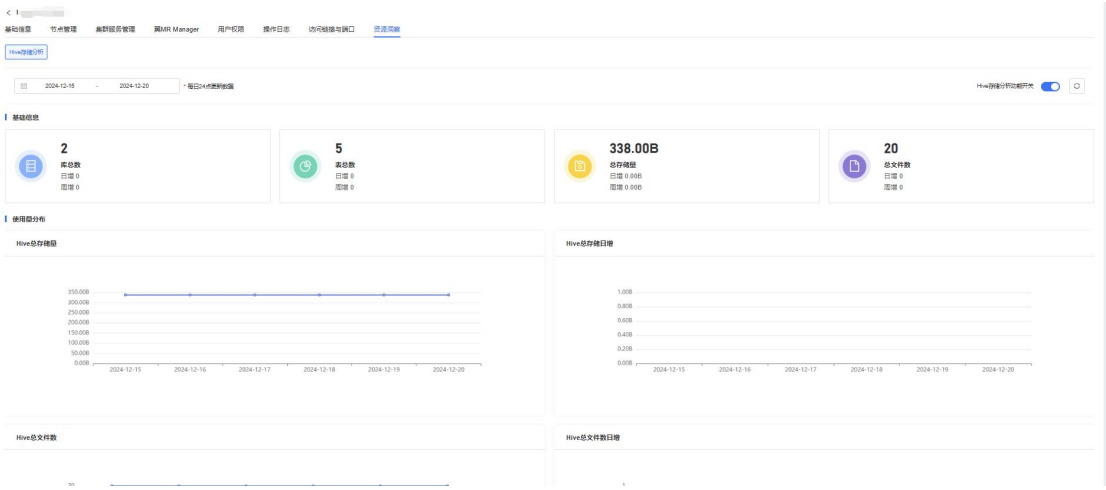
背景信息

1. V2.16 版本起，翼 MR 集群新增资源洞察功能。
2. 当前支持对 Hive 组件服务进行资源洞察，部署 Hive 服务时，可使用该功能。

操作步骤

1. 登录翼 MapReduce 管理控制台。

2. 从“我的集群”中 ，选中一个运行中且已部署 Hive 组件服务的集群并单击集群名称，进入集群信息页面。
3. 选择“资源洞察”，在页面即可查看 Hive 存储分析，含基础信息、使用量分布、库表存储详情等。



注意：

当前仅数据湖与自定义业务场景提供 Hive 可选服务。购买数据湖场景的集群将默认开启资源洞察功能。若选择自定义场景，且订购时未选择 Hive 服务，可在集群创建后按需通过集群服务管理功能添加 Hive 组件服务，添加后将自动开启资源洞察功能。

4.3.2 云监控服务

用户可以通过云监控服务查看翼 MR 集群的主机与组件状态，并支持设置相关监控指标的告警规则，配置短信与邮件告警。如需了解产品详细情况，可参考云监控服务。

操作步骤

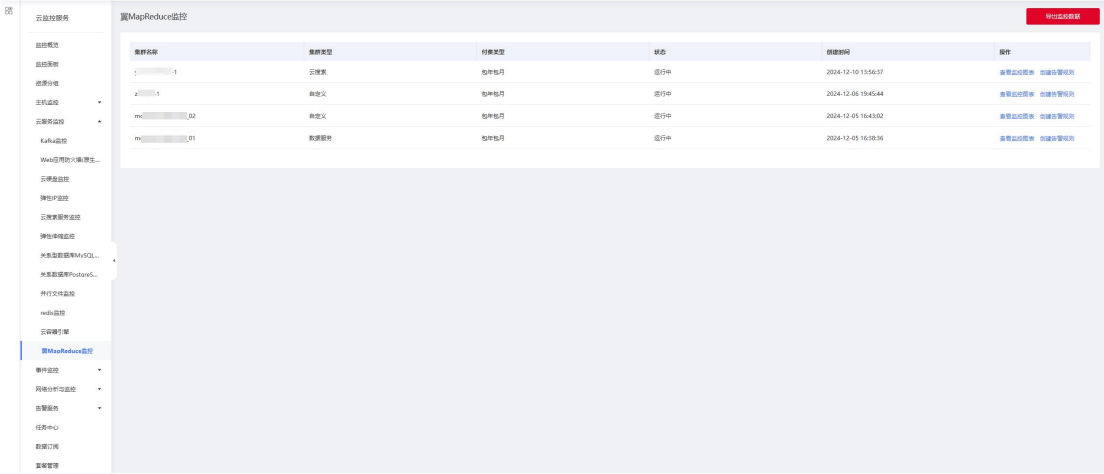
1.登陆天翼云官网，选择“云监控服务”。

2.点击进入云监控服务的管理控制台。

3.在左侧导航栏选择云服务监控中的“翼 MapReduce 监控”，即可查看该用户在当期资源池下的翼 MR 集群监控。

4.点击操作栏中的查看监控图表，可以查看集群相关监控指标，支持自定义查看的时间段。

5.点击操作栏内的创建告警规则，或点击左侧导航栏中告警服务中的“告警规则”，即可配置告警策略、通知与规则。



说明：

- 当前仅华东 1、西南 1、华北 2、上海 36、华南 2、武汉 41、杭州 7 资源池的集群支持使用该服务。
- 仅支持有翼 MR 集群-节点查询权限的用户在云监控服务中查询目标集群的监控数据，管理员可通过翼 MR 的角色管理与用户权限功能为用户赋权。

4.3.3 健康检查

4.3.3 健康检查

健康检查概述

操作场景

该任务指导用户在日常运维中完成集群进行健康检查的工作，以保证集群各项参数、配置以及监控没有异常、能够长时间稳定运行。

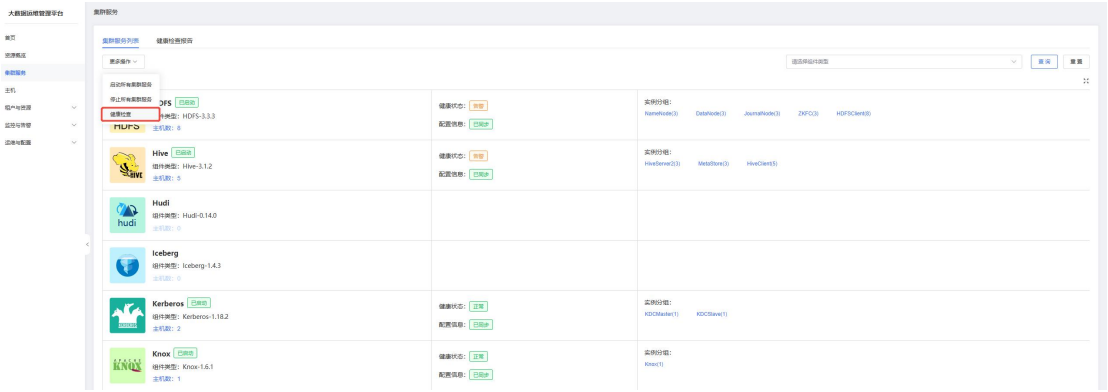
操作说明

集群健康检查的范围包含 Manager 服务级别的各个组件检查：服务级别关注组件是否能够提供正常的服务、告警情况以及各个组件差异化的检查指标。

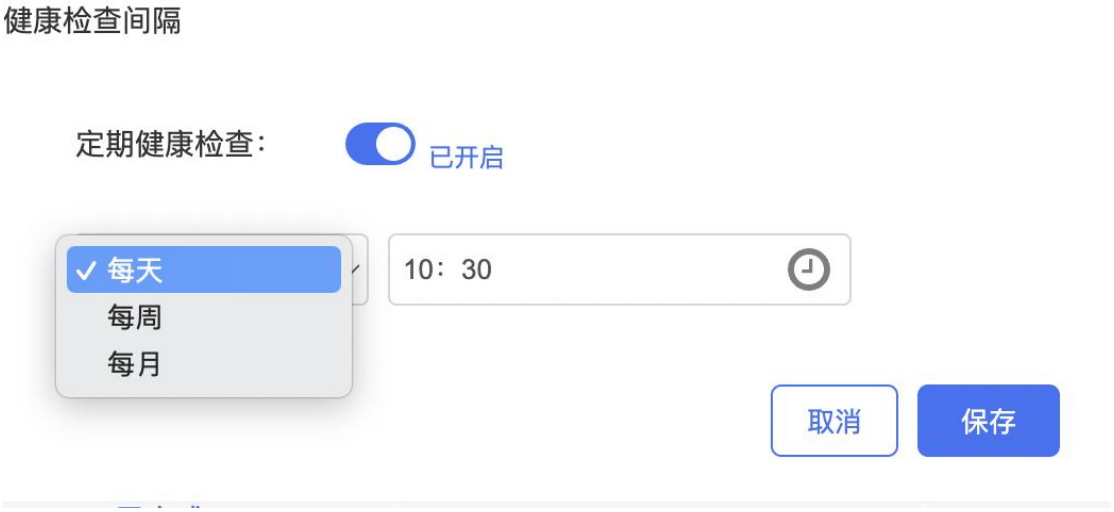
操作步骤

集群健康检查

1. 手动执行所有服务的健康检查。
2. 在集群服务页，单击集群服务列表上方的“更多操作 > 健康检查”。



3. 周期执行所有服务的健康检查。
4. 在集群服务>健康检查报告>点击列表右上方设置按钮>跳出周期任务设置弹框，支持用户根据天/周/月维度进行间隔配置，支持配置具体时刻，创建成功后，对支持的组件服务会定期发起健康检查。



✧ 说明

- 点击健康检查之后页面跳转至健康检查报告 Tab 页面，展示集群的健康检查列表，列表置顶为最新发起的健康检查，点开一级列表下方展示各个组件的检查详情。
- 点击一级列表操作中的导出报告，则支持导出 CSV 文件，用户可本地查阅健康报告。

查看并导出检查报告

操作场景

为了满足对健康检查结果的进一步具体分析，您可以在翼 MR 中查看以及导出健康检查的结果。

操作说明

平台健康检查的范围包含 Manager 服务级别的健康检查。

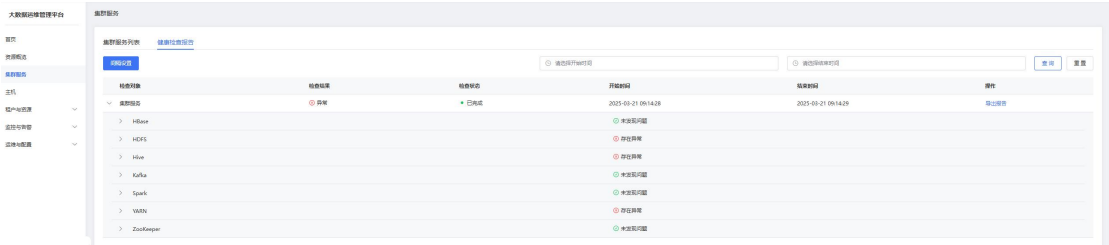
集群健康检查可以包含三方面检查项：各检查对象的服务状态、告警信息、以及每个组件差异化的相关指标。

前提条件

已执行健康检查。

操作步骤

1. 在集群服务页，点击健康检查报告 tab 页。



2. 在健康检查的报告面板上单击“导出报告”导出健康检查报告，下载后可本地查看检查项的完整信息，集群与主机健康检查报告导出格式均为 CSV。

4.3.4 节点管理

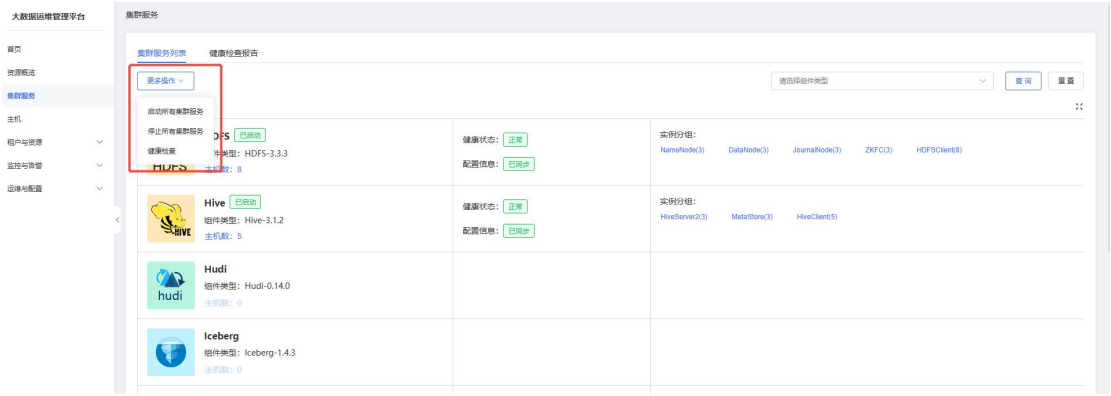
4.3.4 管理节点操作

操作场景

当节点故障异常时，用户可能需要在翼 MapReduce 停止节点上的所有集群服务，对节点进行维护检查。故障清除后，启动节点上的所有集群服务恢复节点业务。

操作步骤

1. 登录翼 MR 控制台，在“我的集群”页面点击正常运行的集群名称，进入集群详情页面。
2. 在集群详情页，单击“翼 MR Manager”，点击”前往翼 MR Manager”。
3. 在 Manager “集群服务”页面，点击“更多操作”，并点击“停止／启动所有集群服务” 执行相应操作。



4.3.4 配置升级

当 master、core 或者 task 节点的规格（vCPU 和内存）无法满足您的业务需求时，您可以使用配置升级功能提升节点规格。

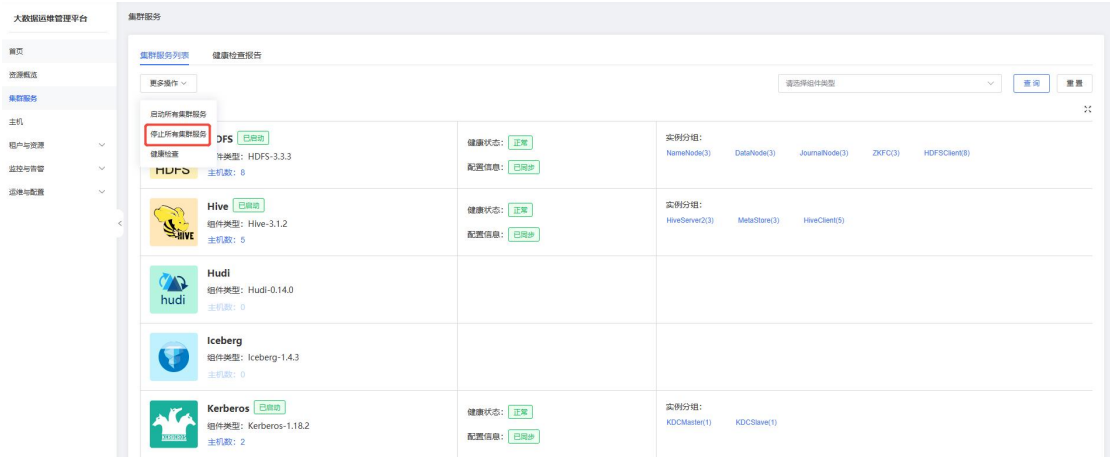
背景信息

配置升级前需要前往 Manager“集群服务—更多操作—停止所有集群服务”，手动关闭大数据集群服务。

操作步骤

1. 登录翼 MapReduce 管理控制台。
2. 从“我的集群”中 ，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 选择“节点管理”，在需要升级配置的节点组的“操作”列并单击“配置升级”，进入升级节点配置页面。只有运行中的集群才能进行配置升级操作。

4. 配置升级前需要前往 Manager “集群服务—更多操作—停止所有集群服务”，手动关闭大数据集群服务，否则无法执行配置升级操作。



5. 在配置升级页面，选择“升级后配置”，确认配置升级金额，勾选协议，并单击确定。

6. 配置升级完成后，您可以按需前往 Manager 的"配置管理”页面进行配置更改。可以参考“配置修改建议说明”。配置修改完成后，要前往 Manager“集群服务—更多操作—启动所有集群服务”手动拉起大数据集群服务。

✧ 说明

- 集群正在升级配置时，不支持进行升级节点配置/节点扩容/磁盘扩容等运维操作。
- 配置升级的操作会被记录在“集群日志”中，可以通过“状态”变化查看配置升级是否执行成功。
- 配置升级成功后，可以在“节点管理”查看集群的节点信息。
- 配置升级前一定要停止集群服务，否则无法执行配置升级操作。
- 配置升级时不能进行 IAM 用户同步操作，若操作 IAM 用户同步会出现同步失败。

4.3.4 节点扩容

当 master、core 或 task 节点组内的计算或存储资源无法满足您的业务需求时，您可以使用节点扩容功能增加实例数量。

背景信息

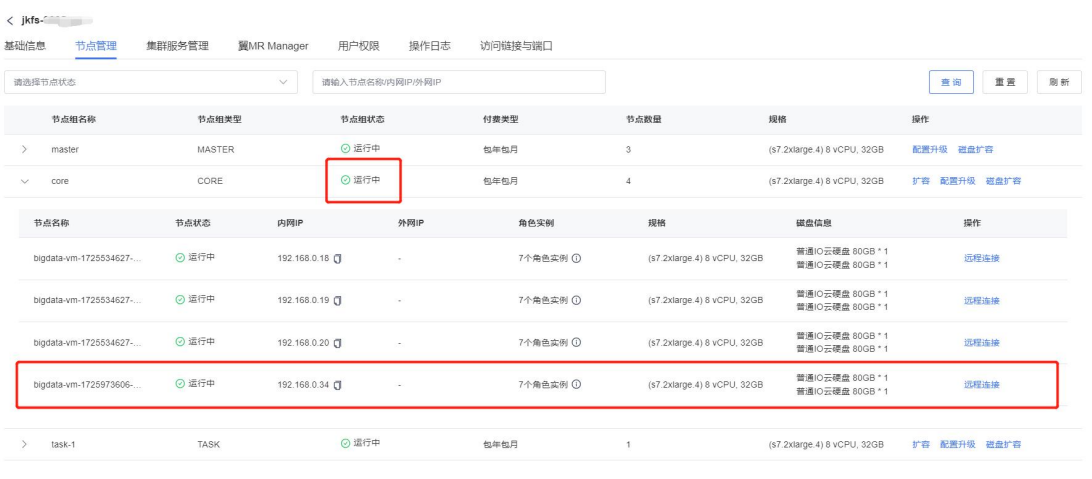
1. V2.15 版本起，数据湖、数据分析、数据服务、实时数据流与自定义场景支持节点扩容功能。
2. V2.17 版本起，云搜索场景支持节点扩容功能。
3. V2.17 版本起，新增 master 扩容功能，支持对 Hive、HDFS、YARN、ZooKeeper、Kyuubi 等服务的部分角色实例进行扩容。数据分析与云搜索场景暂不支持 master 扩容。
4. 仅支持对状态为“运行中”的集群进行节点扩容操作。

操作步骤

Core 与 Task 扩容

1. 登录翼 MapReduce 管理控制台。
2. 从“我的集群”中 ，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 选择“节点管理”，在需要扩容的节点组的“操作”列单击“扩容”，在节点扩容弹窗中进行配置。

4. 选择需要扩容的配置并完成支付后，可在“节点管理”的节点信息中，查看节点扩容结果。当节点组状态从“变更中”变为“运行中”，且展开节点组信息可以看到新增节点，表示扩容成功。



Master 扩容

1、登录翼 MR 控制台，进入目标集群，在节点管理页面，点击 master 节点组操作列中的“扩容”按钮。

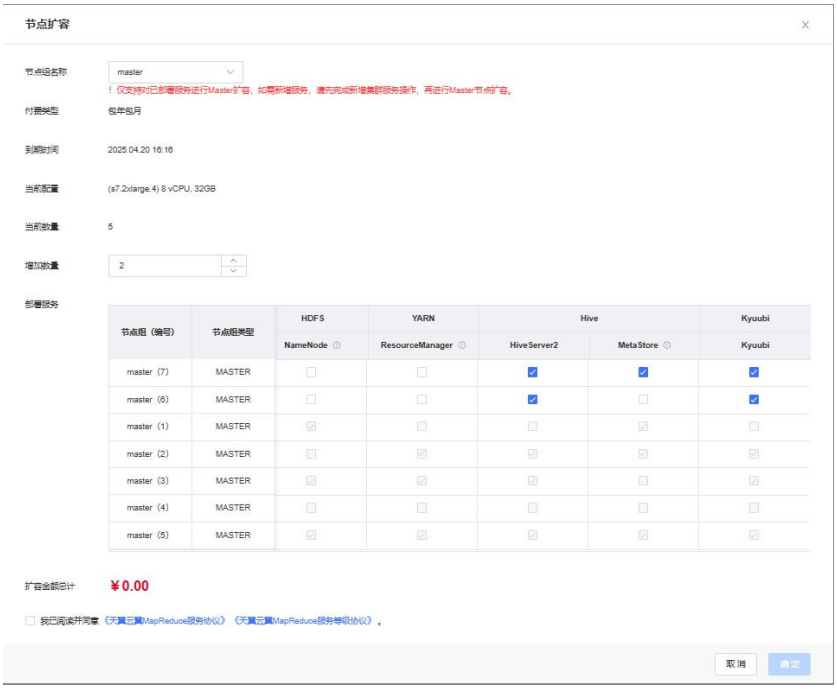
说明：

1）当前仅支持对 Hive、HDFS、YARN、ZooKeeper、Kyuubi 以及 OpenLDAP 集群服务的部分角色实例进行扩容。

2）自定义场景下，若集群已部署上述组件服务，Master 扩容按钮亮起，否则按钮将置灰。

2、在弹窗中，选择需要增加的节点数，并规划新增节点上需要部署的服务。

说明：仅支持对已部署的服务进行 master 扩容，如需新增服务，请先完成新增集群服务操作，再进行 Master 节点扩容。



3、若选择在新增 master 节点上部署 ZooKeeper—ZKServer、HDFS—NameNode、YARN—ResourceManager 或 Hive—MetaStore 角色实例，在控制台显示扩容完成后，需要前往翼 MR—Manager，在相应的集群服务详情中进行“Master 配置生效”操作。

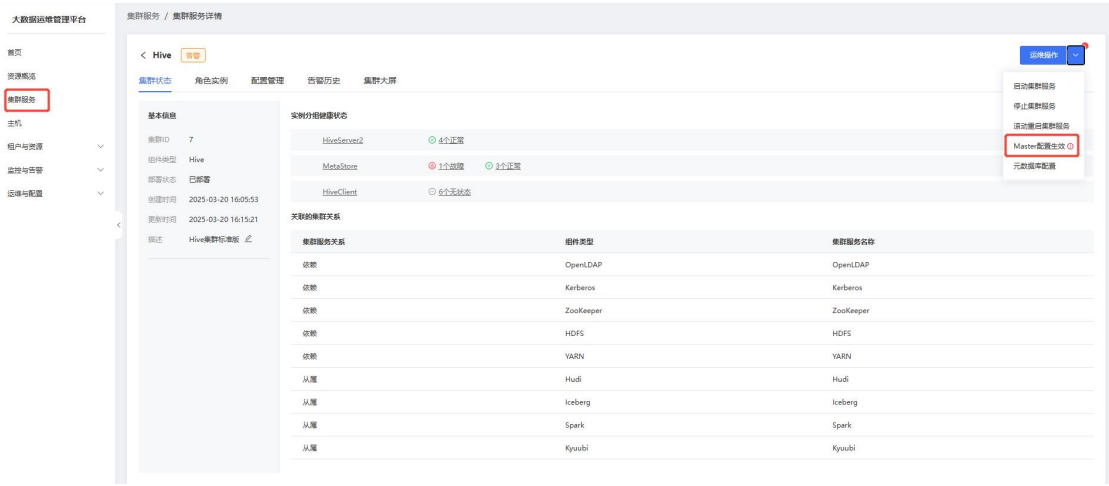
说明：

1）HDFS 扩容后，请对服务进行滚动重启操作。

2）扩容与配置生效过程中，请勿进行其他运维操作，避免影响扩容生效。

3）配置生效过程中可能产生短暂的服务不可用，请合理安排操作时间。

4）进行配置生效前，ZooKeeper、HDFS、YARN 与 Hive 的新增节点无法使用，未避免影响业务开展，请在扩容后尽快进行配置生效操作。



4、对 ZooKeeper、HDFS、YARN、Hive 集群服务进行 master 扩容后，相关联的集群服务可能被影响，请参考 Master 扩容—服务配置修改建议进行操作，避免影响数据业务。

4.3.4 新增节点组

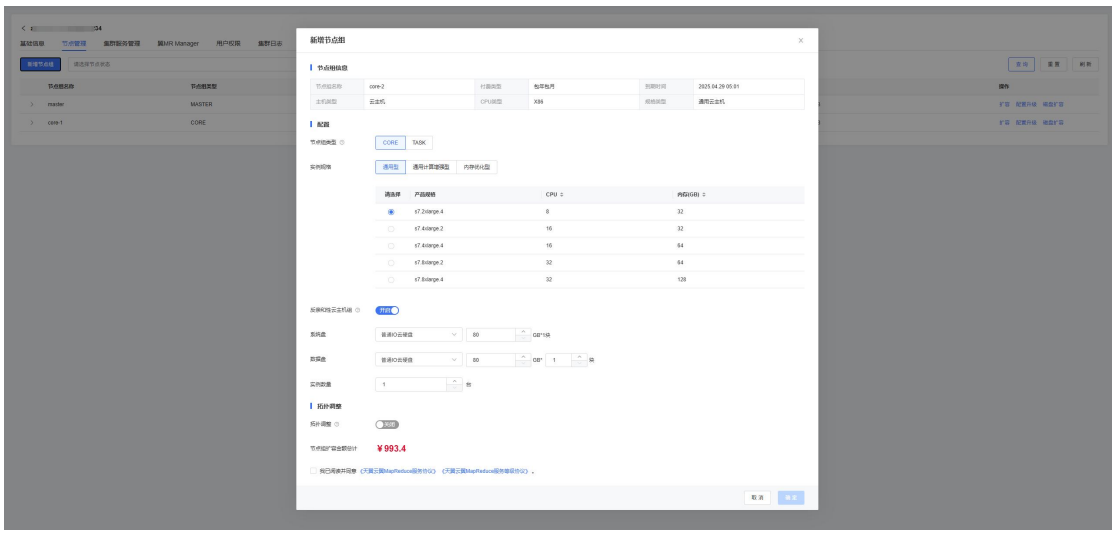
当存量 core 或 task 节点组的计算或存储资源无法满足您的业务需求时，您可以使用新增节点组功能增加 ECS 实例。

背景信息

1. V2.16 版本起，数据湖、数据分析、数据服务、实时数据流与自定义场景支持新增节点组功能。
2. V2.17 版本起，云搜索场景支持新增节点组功能。
3. 当前数据湖、数据服务、实时数据流与自定义业务场景支持新增 core 与 task 节点组，数据分析与云搜索场景支持新增 core 节点组。
4. 仅支持对状态为“运行中”的集群进行新增节点组操作。

操作步骤

1. 登录翼 MapReduce 管理控制台。
2. 从“我的集群”中 ，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 选择“节点管理”，点击“新增节点组”按钮。



4. 选择新增节点组的配置并完成支付后，可在“节点管理”的节点信息中，查看新增节点组结果。当新增节点组状态从变为“运行中”，表示新增成功。

4.3.4 磁盘扩容

当 master、core 或 task 节点的数据存储空间无法满足您的业务需求时，您可以使用磁盘扩容功能增加数据盘的空间。

背景信息

1. V2.15 版本起，翼 MR 集群支持磁盘扩容功能。
2. 当前翼 MapReduce 控制台仅支持对数据盘容量进行扩容，不支持系统盘扩容。
3. 仅支持对状态为“运行中”的集群进行磁盘扩容操作。
4. 当前云硬盘不支持缩容，扩容生效后无法缩小，请按需规划存储空间。
5. 磁盘扩容成功后，需要登录云主机扩展分区和文件系统。

操作步骤

1. 登录翼 MapReduce 管理控制台。
2. 从“我的集群”中 ，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 选择“节点管理”，在需要扩容磁盘的节点组的“操作”列单击“磁盘扩容”，在磁盘扩容弹窗中进行配置。

4. 选择需要扩容的配置并完成支付后，可在“节点管理”的节点信息中，查看磁盘扩容结果，当节点组状态从“变更中”变为“运行中”，且容量已增加，表示扩容成功。

5. 扩容成功后，需要登录云主机扩展分区和文件系统，详细操作方式请参考数据盘扩容磁盘分区和文件系统。

4.3.4 远程连接

操作场景

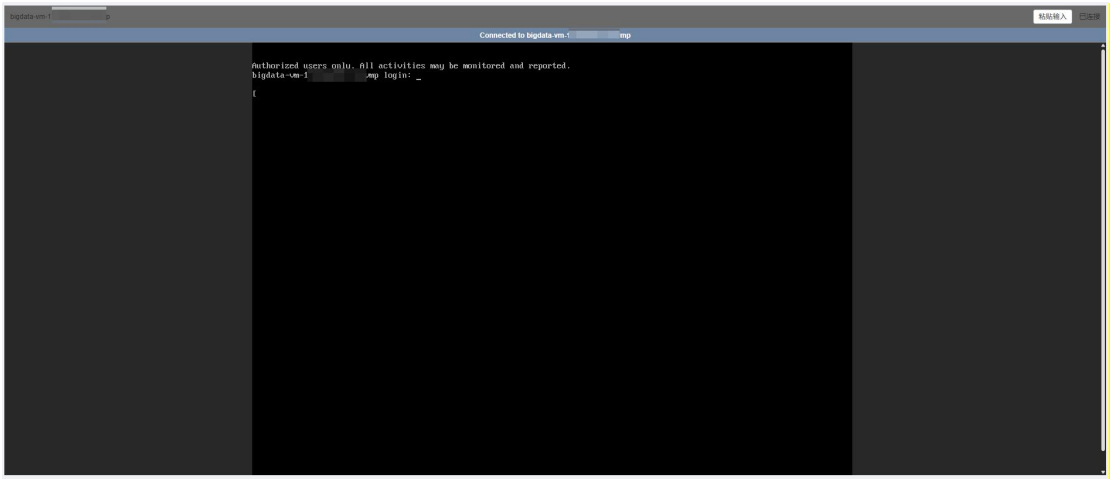
本章节主要介绍如何通过控制台提供的远程登录功能（即 VNC 方式）登录到主机上，查看实例操作系统的运行状态或问题。

操作步骤

1. 登录翼 MR 控制台，点击正常运行的集群名称，进入集群详情页面。

进入“节点管理”页面，点击“节点组名称”列的下拉按钮，展开对应的节点信息，点击“操作”列的“远程连接”，在新的页面上出现 shell 命令窗口。

2. 您可以根据界面提示，通过命令行实现远程登录机器。



✧ 说明

登录弹性云主机时使用的账号与密码为创建集群时设置的登录账密，默认账号为 root。自 2.16 版本起，若忘记密码，可在控制台–基础信息页进行重置密码操作。

4.3.4 绑定/解绑弹性 IP

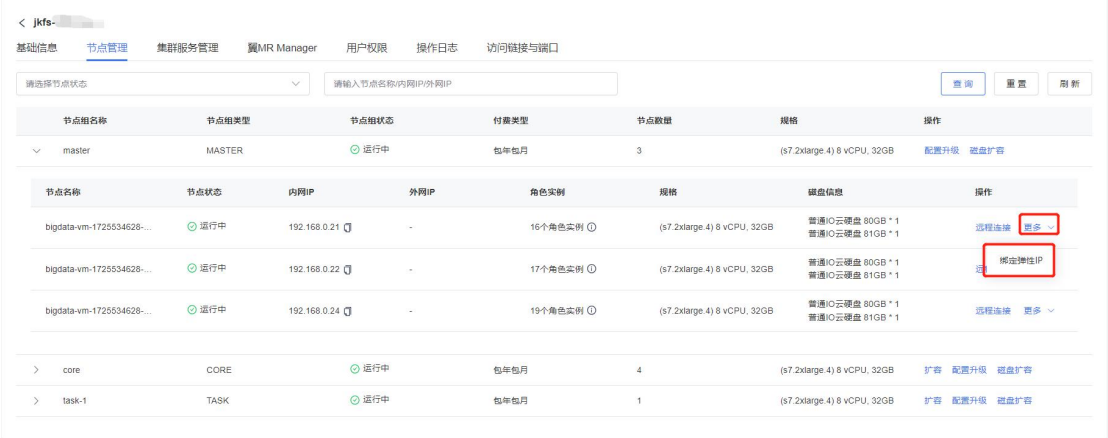
背景信息

创建的翼 MapReduce 集群中所有的节点都会默认被分配内网 IP，外网 IP 需要用户自己去创建。

绑定弹性 IP 操作步骤

1. 登录翼 MapReduce 管理控制台，在我的集群页面，点击集群具体名称，进入集群详情页。

2. 在“节点管理”页面点击“节点组名称”列的下拉按钮，展开对应的节点信息，点击“操作”列的“更多”。



3. 点击“绑定弹性 IP”，在“绑定弹性 IP”的弹框中，如果您账号下没有可用的弹性公网 IP，需要点击“+创建弹性公网 IP”按钮跳转至新页面进行创建；如果您账号下有可用的弹性公网 IP，可以通过下拉“弹性 IP”的选择框选择 IP 后，点击“确认”进行绑定。

解绑弹性 IP 操作步骤

1. 登录翼 MapReduce 管理控制台，在我的集群页面，点击集群具体名称，进入集群详情页。
2. 在“节点管理”页面点击“节点组名称”列的下拉按钮，展开对应的节点信息，点击“操作”列的“更多”中的“解绑弹性 IP”，并在弹窗内进行二次确认，即可完成解绑。

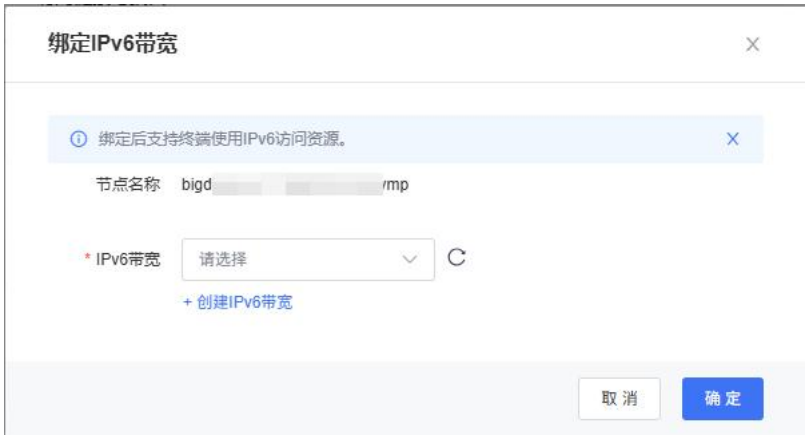
4.3.4 绑定/解绑 IPv6 带宽

若创建翼 MapReduce 集群时所选子网已开通 IPv6，且在订购集群时开启 IPv6 访问，默认所有节点的 IPv6 地址具有内网访问能力，若您希望通过该 IPv6 地址访问公网或被公网访问，需要用户开通并绑定 IPv6 带宽。

绑定 IPv6 带宽操作步骤

1. 登录翼 MapReduce 管理控制台，在我的集群页面，点击集群具体名称，进入集群详情页。
2. 在“节点管理”页面点击“节点组名称”列的下拉按钮，展开对应的节点信息，点击“操作”列的“更多”。

3. 点击“绑定 IPv6 带宽”，在弹框中选择需要绑定的 IPv6 带宽。如果您账号下没有可用的 IPv6 带宽，需要点击“+创建 IPv6 带宽”按钮跳转至新页面进行创建；如果您账号下有可用的 IPv6 带宽，可以通过下拉“IPv6 带宽”的选择框选择，点击“确认”进行绑定。



解绑 IPv6 带宽操作步骤

1. 登录翼 MapReduce 管理控制台，在我的集群页面，点击集群具体名称，进入集群详情页。
2. 在“节点管理”页面点击“节点组名称”列的下拉按钮，展开对应的节点信息，点击“操作”列的“更多”中的“解绑 IPv6 带宽”，并在弹窗内进行二次确认，即可完成解绑。

4.3.5 组件管理

4.3.5 对象管理简介

翼 MapReduce（简称：翼 MR）集群包含了各类不同的基本对象，不同对象的描述介绍如下表所示：

对象	描述	举例
集群服务	可以完成具体业务的一类功能集合	例如：HDFS 集群服务、YARN 集群服务。
服务角色	组成一个完整服务的一类功能实例，一般情况下可使用角色标识	例如：HDFS 由 NameNode、DataNode、HDFSClient、ZKFC、JournalNode 组成，在组件中起到不同的作用，就像扮演着不同的角色，相互通信协调，提供分布式文件存储能力。
实例分组	相同角色的实例组合成实例分组，通常实例分组对应一个角色，名称一般和角色的名称一样	例如：JournalNode 实例分组、DataNode 实例分组。

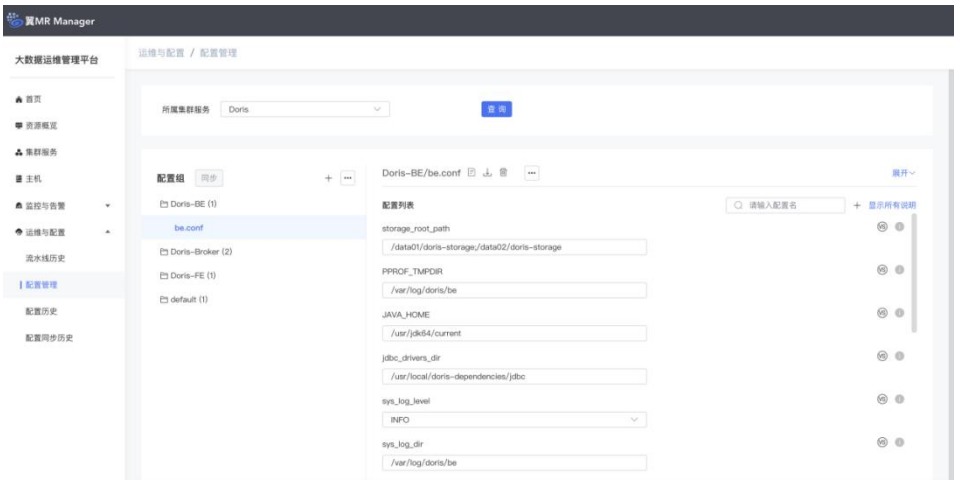
对象	描述	举例
角色实例	服务角色在主机节点上运行的具体实例	例如：运行在 Host2 上的 KerberosAdmin，运行在 Host3 上的 KerberosServer。
主机	用于运行角色实例的物理主机或虚拟主机	例如：Host1～Host5。
集群	由多台主机组成的可以提供多种服务的逻辑实体	例如：名为 Cluster1 的集群由（Host1～Host5）5 个主机组成，提供了 HDFS 和 YARN 等服务。

4.3.5 查看配置

用户可以在翼 MapReduce（简称：翼 MR）上查看集群服务的配置。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务，即可查看该集群服务的相关配置。如图所示：



4.3.5 管理服务操作

用户可以在翼 MapReduce（简称：翼 MR）：

- 启动集群服务

- 停止集群服务
- 滚动重启集群服务

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 选择指定集群服务，单击集群服务 ICON 进入到集群服务的详情页面。
6. 点击操作按钮“运维操作”，单击“启动集群服务”、“停止集群服务”、“滚动重启集群服务”执行相应操作。集群服务之间存在依赖关系。对某集群服务执行启动、停止滚动重启操作时，与该集群服务存在依赖关系的服务将受到影响，具体影响如下：
   a. 启动某集群服务，该集群服务依赖的下层集群服务需先启动，集群服务功能才可生效。
   b. 停止某集群服务，依赖该集群服务的上层集群服务将无法提供功能。
   c. 滚动重启某集群服务，依赖该集群服务且启动的上层集群服务需重启后才可生效。

4.3.5 集群服务管理

集群创建成功后，用户可以根据实际业务场景，新增此前未安装的集群服务，用于满足日渐丰富的业务需求。

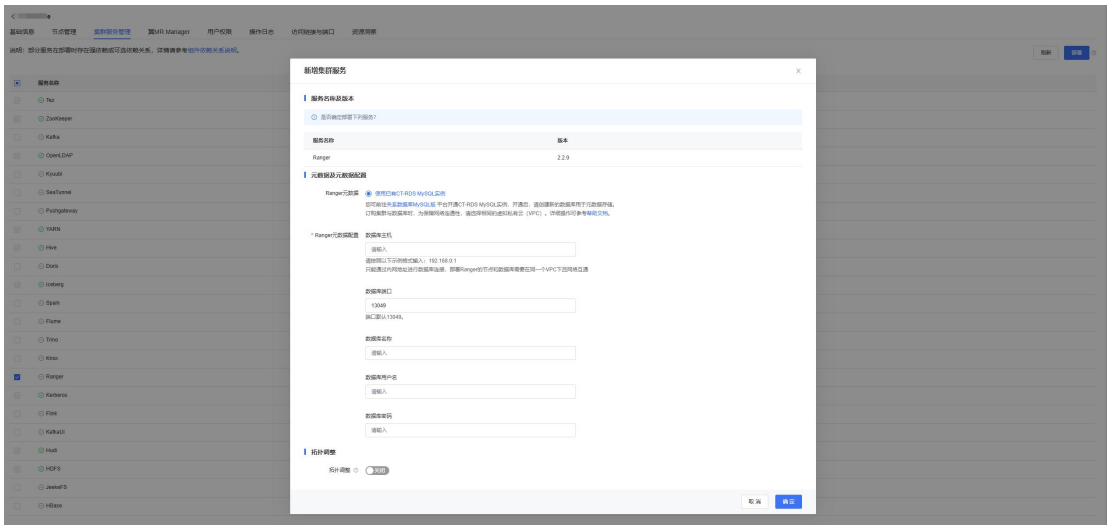
背景信息

1. V2.15 版本起，翼 MR 集群支持集群服务管理功能。
2. 仅支持对状态为“运行中”的集群新增集群服务。
3. 部分组件部署时依赖其他组件，在勾选此类服务时，会同时勾选被依赖组件。详情请参考组件依赖关系说明。

操作步骤

1. 登录翼 MapReduce 管理控制台。
2. 从“我的集群”中 ，选中一个运行中的集群并单击集群名称，进入集群信息页面。

3. 选择“集群服务管理”，在页面中勾选需要新增的服务，并点击“部署”按钮进行新增服务。新增 Hive 与 Ranger 组件时，需填写元数据配置，点击“确定”将自动进行连接性测试，测试无误将启动部署。



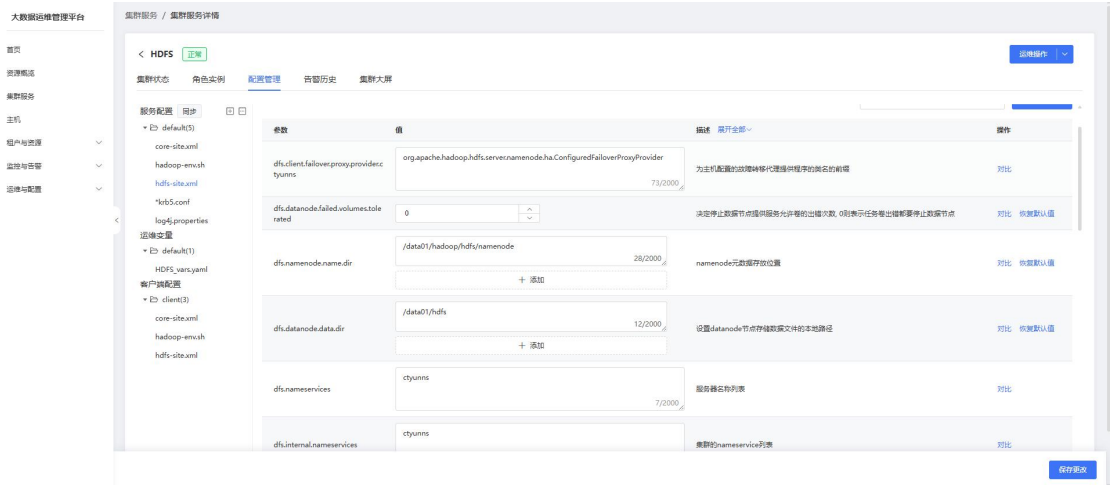
4. 部署完成后，服务状态变为“部署完成”，用户可在“翼 MR Manager”的“集群服务”中查看服务详情。

4.3.5 配置服务参数

用户可以根据实际业务场景，在翼 MapReduce（简称：翼 MR）中快速查看和修改集群服务默认的配置。

修改集群服务配置操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 选择指定集群服务，单击“集群服务”进入到集群服务详情页面。
6. 单击“配置管理”tab，选择要修改的配置文件，修改后，点击右下角“保存更改”按钮。



7. 单击“同步”按钮，对变更的配置文件进行同步。
8. 配置同步完成后，再次进入到该集群服务的详情页面。
9. 单击“运维操作”，单击“滚动重启集群”对该集群服务重启后，修改后的配置即可配置生效。

4.3.5 同步服务配置

操作场景

当用户发现部分集群服务的配置存在待同步配置文件时，可以使用配置同步功能对待同步配置进行同步。

对系统的影响

同步集群服务配置后，需要滚动重启集群服务配置才会生效。重启时对应的集群服务不可用。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 选择指定集群服务，单击集群服务 ICON 进入到集群服务的详情页面。
6. 单击“配置管理”tab。
7. 单击“同步”按钮，对待同步的配置文件进行同步。
8. 配置同步完成后，再次进入到该集群服务的详情页面。

9. 单击“运维操作”，单击“滚动重启集群”对该集群服务重启后，修改后的配置即可配置生效。

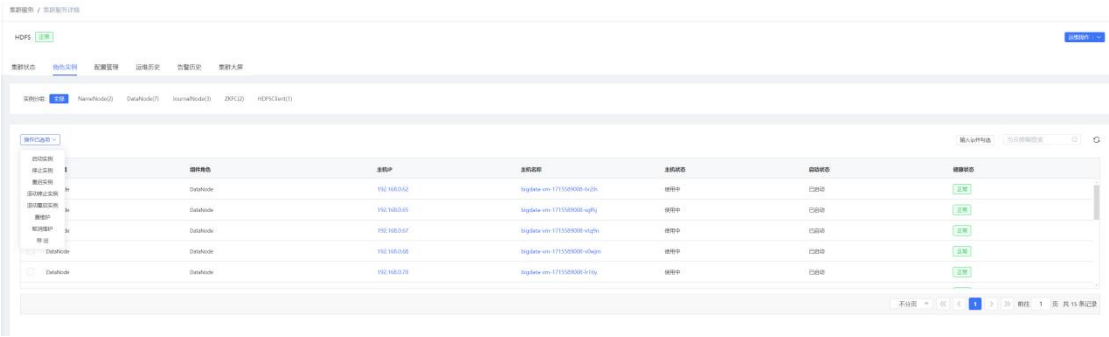
4.3.5 管理角色实例操作

用户可以在翼 MapReduce（简称：翼 MR），对指定角色实例执行：

- 启动实例
- 停止实例
- 重启实例
- 滚动停止实例
- 滚动重启实例
- 置维护
- 取消维护

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 选择指定集群服务，单击集群服务 ICON 进入到集群服务的详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项”，单击“启动实例”、“停止实例”、“重启实例”、“滚动停止实例”、“滚动重启实例”、“置维护”或“取消维护”等，执行相应操作。



4.3.5 启动及停止集群

集群是包含着服务组件的集合。用户可以启动或停止集群中所有集群服务。

启动及停止集群

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 单击“更多操作”，单击“启动所有集群服务”或“停止所有集群服务”执行相应的操作。

4.3.5 下载集群的配置数据

操作场景

为了满足实际业务的需求，用户可以下载配置文件，用于快速更新集群服务配置。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 选择指定集群服务，单击集群服务 ICON 进入到集群服务的详情页面。
6. 单击“配置管理”tab。
7. 单击要下载的配置文件名称，单击下载按钮即可。

4.3.5 支持滚动重启

在修改了大数据组件的配置项后，需要重启对应的服务来使得配置生效，使用普通重启方式会并发重启所有服务或实例，可能引起业务断服。为了确保服务重启过程中，尽量减少或者不影响业务运行，可以通过滚动重启来按批次重启服务或实例（对于有主备状态的实例，会先重启备实例，再重启主实例）。滚动重启方式的重启时间比普通重启时间久。

使用限制

- 请在低业务负载时间段进行滚动重启操作。例如：在滚动重启 Kafka 服务时候， 如果 Kafka 服务业务吞吐量很高（100M/s 以上的情况下），会出现 Kafka 服务滚动重启失败的情况。例如：在滚动重启 HBase 服务时候，如果原生界面上每个 RegionServer

上每秒的请求数超过 1W，需要增大 handle 数来预防重启过程中负载过大导致的 RegionServer 重启失败。

- 重启前需要观察当前 HBase 的负载请求数（原生界面上每个 rs 的请求数如果超过 1W，需要增大 handle 数来预防到时候负载不过来）。
- 在集群 Core 节点个数小于 6 个的情况下，可能会出现业务短时间受影响的情况。

滚动重启集群服务

1. 在控制台集群详情页面，单击“前往翼 MR Manager”。
2. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
3. 选择指定集群服务，单击集群服务 ICON 进入到集群服务的详情页面。
4. 单击“运维操作”，单击“滚动重启集群服务”。
5. 弹出确认滚动重启集群服务操作弹框，单击确定。
6. 进入到滚动重启集群服务操作页面，依次点击“下一步”，滚动重启任务完成后，单击“完成”。

滚动重启实例

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 操作界面，单击“集群服务”菜单。
5. 选择指定集群服务，单击集群服务 ICON 进入到集群服务的详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项”，单击“滚动重启实例”。
9. 弹出确认滚动重启实例操作弹框，单击确定。
10. 进入到滚动重启实例操作页面，依次点击“下一步”，滚动重启任务完成后，单击“完成”。

滚动重启参数说明

滚动重启参数说明如下表所示。

参数名称	描述
单批实例数	按照角色实例维度，每 1 个批次的并发实例数，例如 HDFS-DataNode 默认为 1，输入限

参数名称	描述
	制为 int 的最大值，仅支持输入正整数，当角色实例数小于并发数时，以当前已有实例数为准。 例如：并发数设置为 2，HDFS-DataNode 的实例数为 3，则第一批次执行的实例数是 2，第 2 批次执行的实例数则为 1。
批次间隔时	上个批次成功后与下个批次开始的间隔时间，默认 30 秒，取值范围 1-1800，仅支持输入正整数，单位默认为：秒，不可改。 例如：设置时间为 600S 批次间隔则表示，前 1 个批次运行完成后，等待 600S 后开始下一个批次。
批次等待时间	上个批次开始后，开始下个批次的等待时间默认 30 秒，取值范围 1-1800，仅支持输入正整数；单位默认为：秒，不可改。 例如：设置时间为 600S 批次等待则表示，前 1 个排次开始运行，等待 600S 后不管前一个批次是否完成运行，都开始下一个批次的运行。
实例容错阈值	实例执行失败的容错次数为 2，默认为 0，输入限制为 int 的最大值，仅支持输入正整数。 例如：设置为 0 时，即表示任意一个角色实例的操作失败后，滚动操作终止设置为 2 时，即表示任意 3 个角色实例的操作失败后，滚动操作才会终止。

4.3.6 作业管理

4.3.6 运行 MapReduce 作业

用户可将自己开发的程序提交到翼 MR 中，执行程序并获取结果。

前提条件

用户已经将运行作业所需的程序包和数据文件上传至 HDFS 系统中。

通过后台提交作业

例如安装路径为“/usr/local/hadoop3”。具体以实际为准。

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页选中单击 Master 节点，选择要进入的 Master 节点。
4. 单击该节点右侧的“远程连接”。
5. 根据界面提示，输入 Master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。
6. 集群默认开启 Kerberos 认证，执行以下命令认证当前用户。

```
示例：
Klist -kt /etc/security/keytabs/hdfs.keytab   获取 keytab 的 principalna
me
kinit -kt /etc/security/keytabs/hdfs.keytab_x0005_ keytab 的 principalna
me
```

7. 执行如下命令运行 Example 的 wordcount 作业。

```
hadoop jar /usr/local/hadoop3/share/hadoop/mapreduce/hadoop-ma
preduce-client-jobclient-3.3.3-tests.jar TestDFSIO -Ddfs.replication=1
-write -nrFiles 100 -fileSize 100MB
```

4.3.6 运行 SparkSubmit 作业

用户可将自己开发的程序提交到翼 MR 中，执行程序并获取结果。

前提条件

用户已经将运行作业所需的程序包和数据文件上传至 HDFS 系统中。

通过后台提交作业

例如安装路径为“/usr/local/spark3”。具体以实际为准。

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页选中单击 Master 节点，选择要进入的 Master 节点。
4. 单击该节点右侧的“远程连接”。
5. 根据界面提示，输入 Master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。
6. 集群默认开启 Kerberos 认证，执行以下命令认证当前用户。

```
示例：
Klist -kt /etc/security/keytabs/spark.keytab   获取 spark.keytab 的 princ
 ipalname
kinit -kt /etc/security/keytabs/spark.keytab spark.keytab 的 principalna
 me
```

7. 执行如下命令运行计算圆周率作业。

```
spark-submit --master yarn --deploy-mode client --queue default
 --class org.apache.spark.examples.SparkPi /usr/local/spark3/exampl
 es/jars/spark-examples_2.12-3.2.2.jar 100
```

4.3.6 运行 HiveSQL 作业

用户可将自己开发的程序提交到翼 MR 中，执行程序并获取结果。

Hive Sql 作业用于提交 SQL 语句和 SQL 脚本文件查询和分析数据，包括 SQL 语句和 Script 脚本两种形式，如果 SQL 语句涉及敏感信息，请使用 Script 提交。

前提条件

用户已经将运行作业所需的程序包和数据文件上传至 HDFS 系统中。

通过后台提交作业

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。

3. 在“节点管理”页签中单击 Master 节点，选择要进入的 Master 节点。
4. 单击该节点右侧的“远程连接”。
5. 根据界面提示，输入 Master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。
6. 集群默认开启 Kerberos 认证，执行以下命令认证当前用户。

```
示例：
Klist -kt /etc/security/keytabs/hive.keytab   获取 hive.keytab 的principa
 lname
kinit -kt /etc/security/keytabs/hive.keytab hive.keytab 的 principalname
```

7. 执行 Hive Shell 命令。

```
hive
>show databases;
```

4.3.6 运行 SparkSQL 作业

用户可将自己开发的程序提交到翼 MR 中，执行程序并获取结果。本章节教您在翼 MR 集群后台如何提交一个新的 SparkSQL 作业。SparkSQL 作业用于查询和分析数据，包括 SQL 语句和 Script 脚本两种形式，如果 SQL 语句涉及敏感信息，请使用 Spark Script 提交。

前提条件

用户已经将运行作业所需的程序包和数据文件上传至 HDFS 系统中。

通过后台提交作业

例如安装路径为“/usr/local/spark3”。具体以实际为准。

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页选中单击 Master 节点，选择要进入的 Master 节点。
4. 单击该节点右侧的“远程连接”。
5. 根据界面提示，输入 Master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。

6. 集群默认开启 Kerberos 认证，执行以下命令认证当前用户。

```
示例：
Klist -kt /etc/security/keytabs/spark.keytab  获取 spark.keytab 的 princ
 ipalname
kinit -kt /etc/security/keytabs/spark.keytab spark.keytab 的 principalna
 me
```

7. 打开 spark-sql 命令行，进入 spark-sql 命令行后可执行 SQL 语句，执行命令如下：

```
cd $SPARK_HOME
./bin/spark-sql --conf spark.yarn.principal=default
```

若需要执行 SQL 文件，需要上传 SQL 文件（如上传到"/opt/"目录），上传文件后执行命令如下：

```
cd $SPARK_HOME
./bin/spark-sql --conf spark.yarn.principal=default -f /opt/script.sql
```

4.3.6 运行 Flink 作业

通过后台提交作业

目前 Flink 客户端默认安装路径为"/usr/local/flink"。具体以实际为准。

1. 通过 SSH 方式登录集群。
2. 配置环境变量。 在/etc/profile 添加如下配置：

```
export HADOOP_HOME=/usr/local/hadoop3
export FLINK_HOME=/usr/local/flink
export PATH=PATH:PATH:FLINK_HOME/bin:HADOOP_HOME/bin
export HADOOP_CONF_DIR=HADOOPHOME/bin
```

```
exportHADOOPCONFDIR=HADOOP_HOME/etc/hadoop

export PATH=PATH:PATH:HADOOP_CONF_DIR

export HADOOP_CLASSPATH=hadoop classpath
```

执行如下命令初始化环境变量 *source /etc/profile*

3. 集群默认开启 Kerberos 认证，需要执行以下命令以完成认证。

```
示例：

Klist -kt /etc/security/keytabs/flink.keytab   获取 flink.keytab 的 princip

 alname

kinit -kt /etc/security/keytabs/flink.keytab flink.keytab 的 principalname
```

同时，应配置安全认证。在“/usr/local/flink/conf/flink-conf.yaml”配置文件中的对应配置添加 keytab 路径以及用户名。

```
security.kerberos.login.use-ticket-cache: true

security.kerberos.login.keytab: <user.keytab 文件路径 >

security.kerberos.login.principal: user

security.kerberos.login.contexts: Client,KafkaClient
```

例如：

```
security.kerberos.login.keytab: /etc/security/keytabs/hdfs.keytab

security.kerberos.login.principal: hdfs
```

4. 运行 wordcount 作业。 首先执行上述第 3 步骤中的认证操作，然后启动 Flink 集群。

```
/usr/local/flink/bin/start-cluster.sh
```

- Session 模式

执行如下命令在 session 中提交作业。

```
yarn-session.sh -nm "session-name" --detached flink run /usr/local/flink/exa
mples/streaming/WordCount.jar
```

提交成功后，会返回已提交的 Flink 作业的 YARN Application ID 以及 Web 地址，访问 Web 地址以通过 Web UI 的方式查看作业状态。

- Application 模式

  执行如下命令以 Application 方式提交作业。

```
flink run-application -t yarn-application/usr/local/flink/examples/streaming/W
ordCount.jar
```

提交成功后，会返回已提交的 Flink 作业的 YARN Application ID 以及 Web 地址，访问 Web 地址以通过 Web UI 的方式查看作业状态。

4.3.6 运行 Kafka 作业

用户可将自己开发的程序提交到翼 MR 中，执行程序并获取结果。本章节教您在翼 MR 集群后台如何提交一个新的 kafka 作业。

通过后台提交作业

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页选中单击 Master 节点，选择要进入的 Master 节点。
4. 单击该节点右侧的“远程连接”。
5. 根据界面提示，输入 Master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。
6. 创建 kafka 的 topic。

```
/usr/local/kafka/bin/kafka-topics.sh --zookeeper <Zookeeper_IP>:218
1/kafka --topic <TopicName> --create
```

7. 当前集群默认开启 Kerberos 认证，执行以下命令添加认证信息。

```
export KAFKA_OPTS="-Djava.security.krb5.conf=/etc/krb5.conf -Djava.security.auth.login.config=/usr/local/kafka/config/kafka-jaas.conf"
```

8. 向 topic 中写入消息。

```
/usr/local/kafka/bin/kafka-console-producer.sh --broker-list <KafkaBroker_IP>:9092 --topic <TopicName> --producer.config /usr/local/kafka/config/producer.properties
```

9. 消费 topic 中的消息。

```
/usr/local/kafka/bin/kafka-console-consumer.sh --bootstrap-server <KafkaBroker_IP>:9092 --topic <TopicName> --consumer.config /usr/local/kafka/config/consumer.properties --from-beginning --max-messages 5
```

说明：

--from-beginning：只能消费未被消费的历史数据，已消费数据不会出现。
--max-messages：最多消费多少条数据。

4.3.6 运行 Hue 任务

前提条件

1、使用 Hue WebUI 编辑器前，请前往翼 MR Manager-配置管理，请参考下图修改 hue-ini 配置文件中的 Hive 配置信息，保存并同步配置后，重启 Hue 集群服务。

```
[beeswax]
 hive_discovery_hs2 = true
 hive_discovery_hiveserver2_znode = /hiveserver2
 hive_conf_dir=/usr/local/hive/conf
```

2、当前版本支持使用 LDAP 账号访问 Hue。若使用 admin、hdfs 或 hive 等已有权限的用户登录 Hue，无需进行 Ranger 授权，可以跳过此步骤；如果新建 LDAP 用户，需要通过 Ranger 授予 Hive 与 HDFS 权限，操作步骤如下：

1）前往翼 MR Manager—集群服务中—点击 Ranger—插件启用—启用 Hive 与 HDFS，并重启 Hive 与 HDFS 服务；

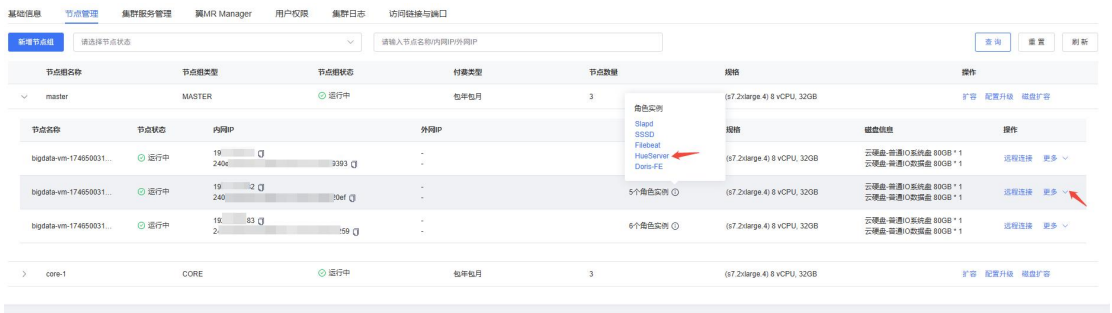
2）登录 Ranger WebUI 界面给需要登录的用户授权 HADOOP SQL/HDFS 等权限。

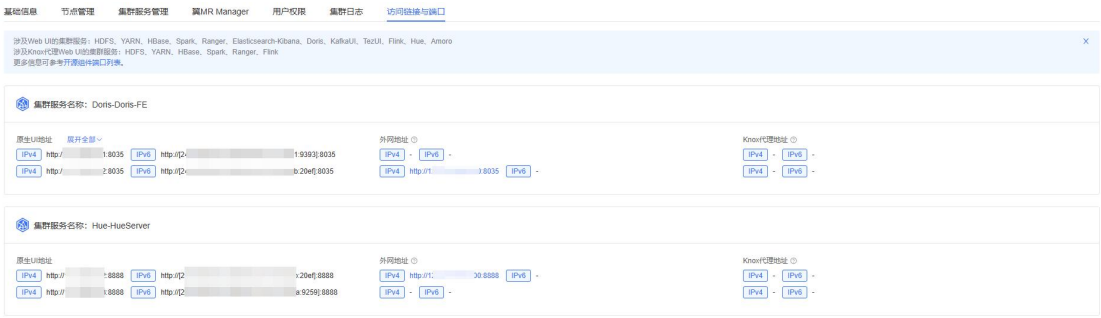
注意：如果没有给登录用户授权，查询 hive 可能报错 Error， Permission denied：user=test，access=WRITE 等问题。

3、修改 core—site.xml 文件，为 Hue 用户配置代理权限。在 HDFS—配置管理中修改 core—site.xml 文件配置，新增 hadoop.proxyuser.hue.hosts 与 hadoop.proxyuser.hue.groups 配置项，配置值可按需设置为*，保存并同步配置后，请重启 HDFS 与 Hive—hiveserver2。

注意：如果没有为 Hue 用户配置代理权限，可能会报错 Failed to validate proxy privilege of hue for test(登录用户)。

4、若通过公网访问 Hue WebUI，需保障网络畅通，您可前往控制台—节点管理，为 Hue 所在的 master 节点绑定弹性 IP，为安全组配置入方向访问规则后，点击访问链接与端口页面中的链接，前往 Hue WebUI 页面。



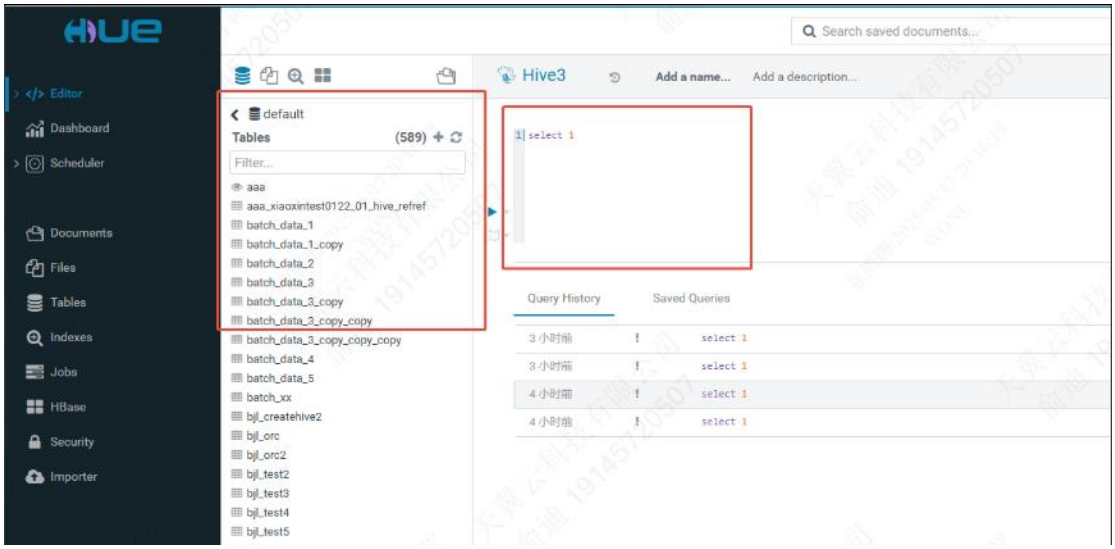


使用 Hive 编辑器

1、在访问链接与端口页面，单击 Hue 服务所在行的链接前往 Hue WebUI 页面，输入 LDAP 账户和密码，即可正常访问 Web UI 页面。

2、在 Hue 的左侧导航栏，单击</>图标，选择 Hive 即可进入 Hive 编辑器。

3、点击列表中上方的刷新按钮，查询集群上的数据库表列。



4、在 Hive 语句编辑区输入语句，即可执行任务。

4.3.6 查看主机和组件日志

操作步骤

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页选中单击 Master 节点，选择要进入的 Master 节点。

4. 单击该节点右侧的“远程连接”。

5. 根据界面提示，输入节点的用户名和密码，用户名为 root。

6. 进入对应组件的日志目录，查看相关日志。

```
[redacted] ~/bigdata/trino-server-427-SNAPSHOT/data/var/log$ll
total 8
drwxr-xr-x  3 [redacted]  96  9 14 15:39 ./
drwxr-xr-x  4 [redacted] 128  9 14 15:39 ../
-rw-r--r--  1 [redacted]  51  9 14 15:39 launcher.log
```

4.3.7 告警管理

4.3.7 查看告警列表

查看告警列表

告警历史显示了集群中的所有告警信息。

告警历史默认按时间顺序排列，时间最近的告警显示在最前端。

告警信息中的各字段说明如下表所示：

参数	参数说明
告警时间	产生告警的时间。
告警级别	灾难：集群服务不可用。 严重：集群服务出现可能影响服务的故障。 错误：集群服务发生异常，需要进一步处理，单集群服务仍提供正常服务。 警告：并不确认告警出现是否影响服务运行，提示可能的风险，需要进一步观测和排查。
是否恢复	告警是否还存在。
告警持续时间	从发生告警到现在的时长。
主机 IP	发生告警的主机 IP 地址。
告警信息	告警信息。
所属集群服务	发生告警的集群服务、实例。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“监控与告警> 告警历史”。
5. 查询条件选择指定集群服务、角色实例，或主机 IP，单击“查询”按钮即可查看相关告警信息。

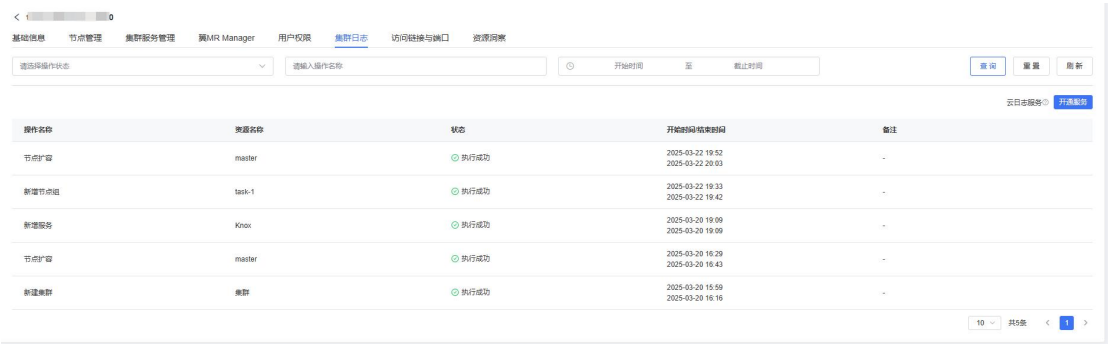
4.3.8 日志与审计

4.3.8 集群日志

翼 MR 控制台提供集群操作日志查询能力，同时，自 2.17 版本起，用户可通过云日志服务与云审计产品查看组件服务日志与操作审计日志。

日志入口

从翼 MR 控制台进入集群信息页面，单击“集群日志”页签即可查看集群操作日志。



日志查询

翼 MR 服务操作日志，提供了筛选框帮助用户快速定位问题原因。用户在选择筛选条件后，单击“查询”按钮进行查询日志，“重置”按钮清空筛选条件。

字段说明

参数	参数说明
操作名称	记录执行的操作名称，包括：新建集群、配置升级等。
资源名称	记录操作的资源名称，包括：集群、core、master 等。
状态	记录操作的状态，包括：执行成功、执行失败、执行中等。

参数	参数说明
开始时间/结束时间	记录操作开始执行时间和执行结束时间。
备注	对操作失败的备注说明。

4.3.8 云日志服务

背景信息

1. V2.17 版本起，用户可通过云日志服务查询翼 MR 集群的组件日志。

操作步骤

1. 从翼 MR 控制台进入集群信息页面，单击“集群日志”页签。
2. 点击右上角“开通服务”，授权翼 MR 开通云日志服务。

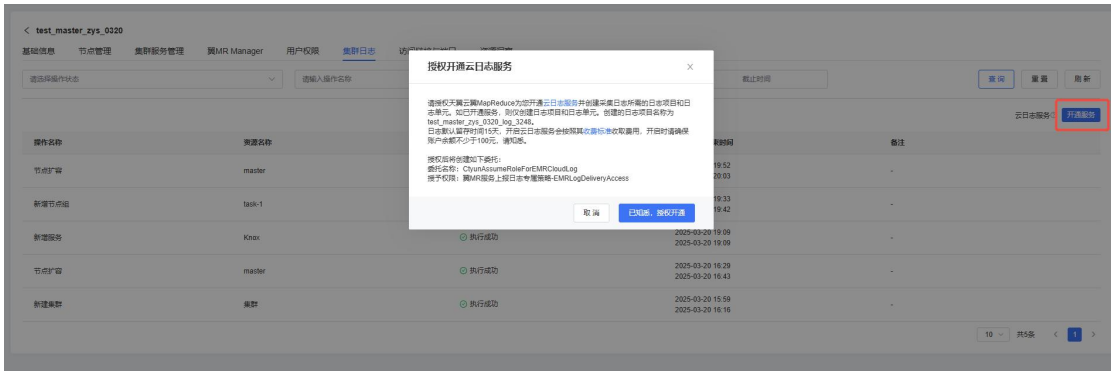
注意：

1)授权开通前请详细阅读云日志服务的收费标准，授权开通后，云日志服务将按标准收取费用。

2)开启云日志服务时，请确保账户余额不少于 100 元，否则将导致开通失败。

3)授权开通时，日志默认留存时间为 15 天，如需调整，可前往云日志服务的控制台进行操作。

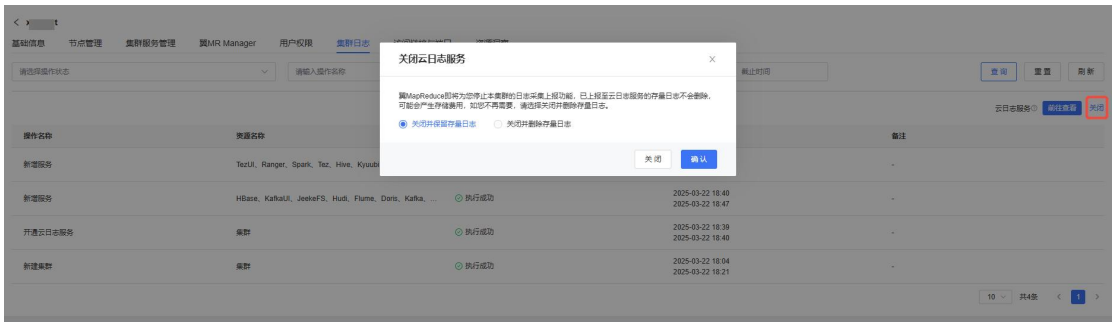
4)云日志服务可创建日志单元数量的上限为 50 个，若存量单元数已达到 50，将导致翼 MR 日志上报失败。



3. 点击【已知悉，授权开通】后，翼 MR 将开始为用户创建该集群的日志项目与日志单元，可点击翼 MR 控制台“集群日志”页面右上角的“前往查看”跳转至云日志服务控制台查看日志详情。

注意：从云日志服务控制台删除翼 MR 所创建的日志项目或日志单元，将导致相应日志信息丢失，请慎重操作。

4. 如需关闭云日志服务，停止日志上报，可点击翼 MR 控制台“集群日志”页面右上角的“关闭”按钮，停止日志上报



5. 云日志服务操作说明，请参考云日志服务帮助文档。

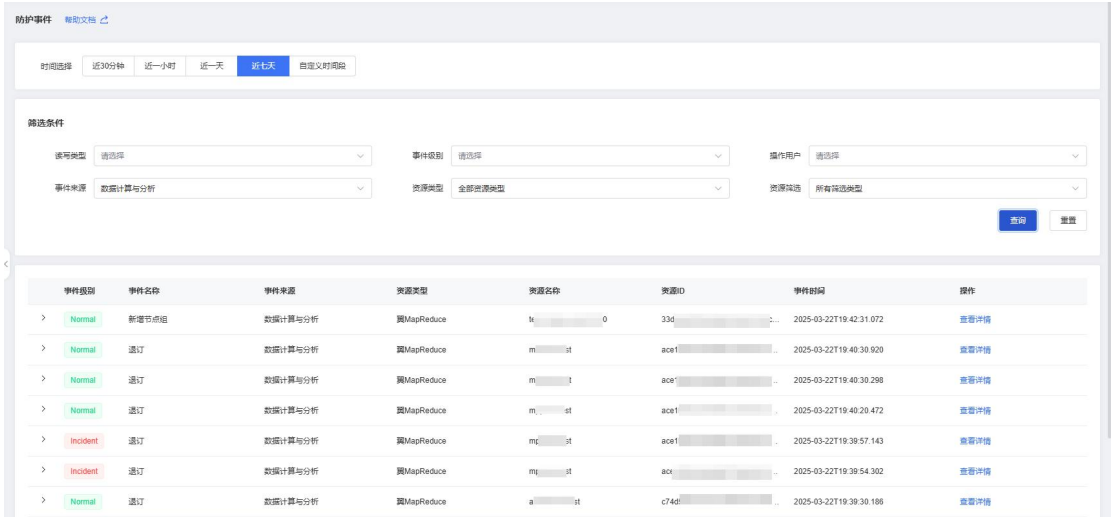
4.3.8 云审计

背景信息

1. V2.17 版本起，用户可通过云审计产品查询翼 MR 集群的组件日志。

操作步骤

1. 请前往云审计-控制台，开通翼 MR 集群所在资源池的云审计产品。
2. 开通云审计产品后，翼 MR 将推送集群日志到云审计平台，用户可在云审计控制台查看



3. 云审计操作说明，请参考云审计帮助文档。

4.4 使用翼 MR 客户端

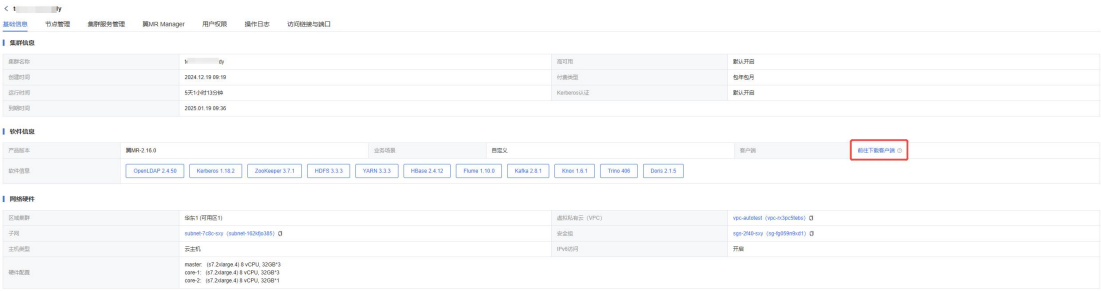
4.4.1 下载客户端

操作场景

未部署服务端的服务器上需要安装组件客户端时，需要下载客户端安装包和配置文件。

操作步骤

1. 登录翼 MR 管理控制台。
2. 在“我的集群”中单击指定的集群名称，进入集群详情页面。
3. 在“基础信息”页面的软件信息中，点击“前往下载客户端”按钮即可下载。当前共 4 个业务场景支持下载客户端，分别是：数据湖、数据服务、实时数据流和自定义场景。



✧ **注意**

- 下载完整客户端包含：客户端安装包、MD5 值和客户端配置文件。
- 客户端安装包包含组件：Hadoop（HDFS、YARN）、Hive、HBase、Flink、Kafka、Spark、ZooKeeper、Trino、Kyuubi、Kerberos。
- V2.17 版本起，客户端安装包中新增 Flume、SeaTunnel。
- 客户端配置文件：包含当前环境公共空间下所有客户端配置文件。
- 文件容量较大，请您耐心等待。

4.4.2 安装客户端

修改集群内组件的服务端配置后，建议重新安装客户端，否则客户端与服务端版本将不一致，可能影响正常使用。

前提条件

- 当安装客户端节点为集群外节点时，该节点必须能够与集群内节点网络互通，否则安装会失败。
- 待安装客户端节点必须启用 NTP 服务，并保持与服务端的时间一致，否则安装会失败。
- 在节点上安装客户端，可以使用 root 用户或任意操作系统用户进行操作，要求该用户对客户端文件存放目录和安装目录具有操作权限，两个目录的权限为 775 。
- 需要允许用户使用密码方式登录 Linux 弹性云服务器（SSH 方式）。

安装客户端

1、请先在服务器上配置时间同步服务（ntpd 或 chronyd），并确认同步状态正常。

```
ps aux | grep 'ntpd\|chronyd'
```

```
[root@yunzhujihll ~]# ps aux | grep 'ntpd\|chronyd'
chrony     619  0.0  0.2  78144  2372 ?        S    16:26   0:00 /usr/sbin/chronyd
root      1770  0.0  0.2  22096  2408 pts/0    S+   16:35   0:00 grep --color=auto ntpd\|chronyd
[root@yunzhujihll ~]#
```

```
[root@yunzhujihll ~]# timedatectl
                Local time: Thu 2024-12-05 16:41:23 CST
            Universal time: Thu 2024-12-05 08:41:23 UTC
                  RTC time: Thu 2024-12-05 08:41:23
                 Time zone: Asia/Shanghai (CST, +0800)
System clock synchronized: yes
               NTP service: active
           RTC in local TZ: no
```

2、服务器上需要安装好 kerberos 客户端相关 rpm 包（krb5-workstation 或 krb5-client），如有 sudo 权限但未安装会自动进行安装。

说明

如果使用 root 用户执行安装脚本，则不需要再安装了，会自行安装

3、服务器上需要安装好 jdk，并新建/usr/jdk64/目录，然后将/usr/jdk64/current 软链到 ${JAVA_HOME}。

- 执行 java -version 确认是否已经安装 jdk。如果使用的是天翼云的 jdk，已经做好软链 不需要额外操作。
- 如何判断是否是天翼云的 jdk？可以执行下面语句，如果能匹配出带有 ccdp 的内容则代表是天翼云的 jdk；否则需要自己软链（下图表示不是天翼云的 jdk）。

```
[root@zhouhua-test1-vmH7iSdANH data01]# rpm -qa | grep jdk
[root@zhouhua-test1-vmH7iSdANH data01]#
```

```
x86_64
[root@yunzhujihll usr]# yum list | grep jdk
copy-jdk-configs.noarch                          4.0-1.ctl3                                   everything
java-1.8.0-openjdk.x86_64                        1:1.8.0.412.b08-5.ctl3                       update
java-1.8.0-openjdk-accessibility.x86_64          1:1.8.0.412.b08-5.ctl3                       update
java-1.8.0-openjdk-accessibility-slowdebug.x86_64  1:1.8.0.412.b08-5.ctl3                     update
java-1.8.0-openjdk-demo.x86_64                   1:1.8.0.412.b08-5.ctl3                       update
java-1.8.0-openjdk-demo-slowdebug.x86_64         1:1.8.0.412.b08-5.ctl3                       update
```

```
[root@yunzhujihll usr]# yum install java-1.8.0-openjdk.x86_64
Last metadata expiration check: 0:00:41 ago on Thu 05 Dec 2024 05:00:41 PM CST.
Dependencies resolved.
================================================================================================================
 Package                          Architecture      Version                          Repository          Size
================================================================================================================
Installing:
 java-1.8.0-openjdk               x86_64            1:1.8.0.412.b08-5.ctl3           update              426 k
Installing dependencies:
 alsa-lib                         x86_64            1.2.5.1-3.ctl3                   everything          454 k
 copy-jdk-configs                 noarch            4.0-1.ctl3                       everything           28 k
```

```
Complete!
[root@yunzhujihll usr]# java -version
openjdk version "1.8.0_412"
OpenJDK Runtime Environment Bisheng (build 1.8.0_412-b08)
OpenJDK 64-Bit Server VM Bisheng (build 25.412-b08, mixed mode)
[root@yunzhujihll usr]#
```

```
lrwxrwxrwx.  1 root root  10 Jan 13  2023 tmp -> ../var/tmp
[root@yunzhujihll usr]# which java
/usr/bin/java
[root@yunzhujihll usr]# echo $JAVA_HOME

[root@yunzhujihll usr]# echo $JAVA_HOME
```

```
[root@yunzhujihll usr]# readlink -f /usr/bin/java
/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.412.b08-5.ctl3.x86_64/jre/bin/java
[root@yunzhujihll usr]#
```

配置环境变量：sudo vim /etc/profile

```
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.412.b08-5.ctl3.x86_64/jre
```

软链：ln -sf {JAVA_HOME} /usr/jdk64/current

```
[root@yunzhujihll CCDP-3.6.0-client]# ll /usr/jdk64/
total 0
lrwxrwxrwx 1 root root 63 Dec  5 18:03 current -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.412.b08-5.ctl3.x86_64/jre
[root@yunzhujihll CCDP-3.6.0-client]#
```

4、/etc/hosts 内需要写入 server 端服务器的 IP 和主机名。

5、将下载的 client_config.tar.gz 和 CCDP--client.tar.gz 两个安装包以及 keytab 文件放在同一目录。

6、tar zxvf CCDP--client.tar.gz 解压，进入 CCDP--client 目录，目录下有 md5 文件和安装包，可用 md5 文件进行校验，例子：md5sum -c CCDP-3.3.3-x86_64-client.tar.gz.md5。

7、再次 tar zxvf CCDP--client.tar.gz 解压，进入 CCDP--client 目录，可看到安装脚本 install.sh 与各组件目录。

8、进入二次解压后的 CCDP-*-client 目录，执行 sh install.sh 安装客户端。可以直接用 root 用户安装。

9、krb5.conf 配置文件拷贝需要 root 或有 sudo 权限的用户。若安装用户没有权限，则需提前使用 root 权限将 krb5.conf 拷贝到/etc/目录，随后使用安装用户执行 sh install.sh -s 跳过 krb5.conf 配置，直接安装大数据客户端。

10、安装运行的日志保存在/tmp/emr_bigdata_client_install.log 中

11、使用客户端前需要执行 source bigdata_env 获取环境变量(要在同一个窗口，不能多个窗口)。source 每次重新登录都要执行。

12、集群默认启用 Kerberos 认证，执行以下命令认证当前用户。请用实际的 keytab 文件的路径替换/path/to/username.keytab。

```
kinit -kt /path/to/username.keytab `klist -kt /path/to/username.keytab |sed -n 4p |awk '{print $NF}'`
```

13、直接执行组件的客户端命令。

例如使用 HDFS 客户端命令查看 HDFS 根目录文件，执行 hdfs dfs -ls /

14、客户端安装完成后，可参考“各组件客户端使用实践”使用客户端。

说明：

- 操作步骤可参考软件包目录下的 README.txt。
- 如需卸载客户端，可直接删除安装目录。

4.4.3 各组件客户端使用实践

4.4.3 使用 YARN 客户端

操作场景

该任务指导用户在运维场景或业务场景中使用 YARN 客户端。

前提条件

已安装客户端。

例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 YARN 客户端

1. 安装客户端。
2. 用户登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。

```
cd /opt/hadoopclient
```

4. 如果集群默认为安全模式，执行以下命令进行用户认证。

```
kinit 组件业务用户
```

5. 直接执行 YARN 命令。例如：

```
yarn application -list
```

4.4.3 使用 Flink 客户端

前提条件

- 翼 MR 集群中已安装 Flink 组件。
- 集群正常运行，已安装集群客户端，例如安装目录为“/usr/local/flink”。以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 Flink 客户端

1. 安装客户端，具体请参考安装客户端章节。
2. 登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。 *cd /opt/hadoopclient*
4. 执行如下命令初始化环境变量。

```
source /opt/hadoopclient/bigdata_env

source ~/.bash_profile
```

5. 若集群开启 Kerberos 认证，需要执行以下命令以完成认证，若集群未开启 Kerberos 认证请跳过该步骤。

```
klist -kt /etc/security/keytabs/hdfs.keytab

kinit -kt /etc/security/keytabs/hdfs.keytab XXXX
```

同时，应配置安全认证。在“/usr/local/flink/conf/flink-conf.yaml”配置文件中的对应配置添加 keytab 路径以及用户名。

```
security.kerberos.login.use-ticket-cache: true

security.kerberos.login.keytab: <user.keytab 文件路径 >
```

```
security.kerberos.login.principal: user

security.kerberossecurity.login.contexts: Client,KafkaClient
```

例如：

```
security.kerberos.login.keytab: /etc/security/keytabs/hdfs.keytab

security.kerberos.login.principal: hdfs
```

6. 运行 wordcount 作业。

   a. **普通集群（未开启 Kerberos 认证）** 首先启动 Flink 集群。*/usr/local/flink/bin/start-cluster.sh*

   b. **Session 模式** 执行如下命令在 session 中提交作业。

   c. *yarn-session.sh -nm "session-name" --detachedflink run /usr/local/flink/examples/streaming/WordCount.jar*

   d. 提交成功后，会返回已提交的 Flink 作业的 YARN Application ID 以及 Web 地址，访问 Web 地址以通过 Web UI 的方式查看作业状态。

   e. **Per-Job 模式**

   f. 执行如下命令以 Per-Job 方式提交作业。

   g. *flink run -t yarn-per-job --detached /usr/local/flink/examples/streaming/WordCount.jar*

   h. 提交成功后，会返回已提交的 Flink 作业的 YARN Application ID 以及 Web 地址，访问 Web 地址以通过 Web UI 的方式查看作业状态。

   i. **Application 模式**

   j. 执行如下命令以 Application 方式提交作业。

   k. *flink run-application -t yarn-application/usr/local/flink/examples/streaming/WordCount.jar*

   l. 提交成功后，会返回已提交的 Flink 作业的 YARN Application ID 以及 Web 地址，访问 Web 地址以通过 Web UI 的方式查看作业状态。

   m. **安全集群（开启 Kerberos 认证）**

   n. 首先执行 5 中的认证操作，再进行和普通集群相同的操作以运行作业。

4.4.3 使用 HBase 客户端

操作场景

该任务指导用户在运维场景或业务场景中使用 HBase 客户端。

前提条件

- 已安装客户端。例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。
- 非 root 用户使用 HBase 客户端，请确保该 HBase 客户端目录的属主为该用户，否则请参考如下命令修改属主。

```
chown user:group -R 客户端安装目录/HBase
```

使用 HBase 客户端

1. 安装客户端，具体请参考安装客户端章节。
2. 用户登录安装客户端的节点。
3. 执行以下命令切换到客户端目录。

```
cd /opt/hadoopclient
```

4. 当前集群默认启用 Kerberos 认证，执行以下命令认证当前用户，当前用户需要具有创建 HBase 表的权限，为用户绑定对应角色。

```
kinit 组件业务用户
```

例如：

```
kinit hbaseuser
```

5. 直接执行 HBase 组件的客户端命令。

```
hbase shell
```

4.4.3 使用 HDFS 客户端

操作场景

该任务指导用户在运维场景或业务场景中使用 HDFS 客户端。

前提条件

- 已安装客户端。
- 例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 HDFS 客户端

1. 安装客户端。
2. 登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。

```
cd /opt/hadoopclient
```

4. 集群默认为安全模式，执行以下命令进行用户认证。

```
kinit 组件业务用户
```

5. 直接执行 HDFS Shell 命令。例如：

```
hdfs dfs -ls /
```

4.4.3 使用 Hive 客户端

操作场景

该任务指导用户在运维场景或业务场景中使用 Hive 客户端。

前提条件

已安装客户端，例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 Hive 客户端

1. 安装客户端，具体请参考安装客户端章节。
2. 用户登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。

```
cd /opt/hadoopclient
```

4. 集群默认开启认证模式，完成 Hive 客户端登录。执行以下命令，完成用户认证并登录 Hive 客户端。

```
kinit 组件业务用户

beeline
```

✧ **说明**

beeline 连接后可以编写并提交 HQL 语句执行相关任务。

4.4.3 使用 Kafka 客户端

操作场景

用户可以在集群客户端完成 Topic 的创建、查询、删除等基本操作。

前提条件

已安装客户端，例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 Kafka 客户端

1. 安装客户端，具体请参考安装客户端章节。
2. 登录安装客户端的节点。
3. 执行以下命令，切换到客户端目录，例如：

```
cd /opt/hadoopclient/Kafka/kafka/bin
```

4. 执行以下命令，配置环境变量。

```
source /opt/hadoopclient/bigdata_env
```

5. 集群默认已启用 Kerberos 认证，执行以下命令认证当前用户。

```
kinit Kafka 用户
```

6. 创建一个 Topic。

```
sh kafka-topics.sh --create --topic 主题名称--partitions 主题占用的分区数 --replication-factor 主题的备份个数 --zookeeper ZooKeeper 角色实例所在节点 IP 地址:clientPort/kafka
```

7. 执行以下命令，查询集群中的 Topic 信息。

```
sh kafka-topics.sh --list --zookeeper ZooKeeper 角色实例所在节点 IP 地址:clientPort/kafka
```

8. 删除创建的 Topic。

```
sh kafka-topics.sh --delete --topic 主题名称 --zookeeper ZooKeeper 角色实例所在节点 IP 地址:clientPort/kafka
```

输入 "y"，回车。

4.4.3 使用 Trino 客户端

操作场景

该任务指导用户使用 Trino 客户端。

前提条件

已安装客户端。

例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 Trino 客户端

1. 将 trino-cli-406-SNAPSHOT-executable.jar 拷贝到/opt/hadoopclient 下。
2. 执行如下命令启动客户端。

```
java -jar trino-cli-406-SNAPSHOT-executable.jar --server Trino 服务的 IP:8080
```

3. 执行成功后的客户端界面。

```
trino:sf1> use tpcds.sf1;
USE
trino:sf1>
```

4.4.3 使用 ZooKeeper 客户端

前提条件

已安装客户端。

例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 ZooKeeper 客户端

1. 安装客户端并安装了 JDK8。
2. 登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。

```
cd /opt/hadoopclient
```

4. 集群默认为安全模式，执行以下命令进行用户认证。

```
kinit 组件业务用户
```

5. 直接执行命令。例如：

```
zkCli.sh -服务器 IP:端口号
```

4.4.3 使用 Kyuubi 客户端

前提条件

已安装客户端并且部署了 ZooKeeper 集群。

例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 Kyuubi 客户端

1. 安装客户端。
2. 登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。

```
cd /opt/hadoopclient
```

4. 集群默认为安全模式，执行以下命令进行用户认证。

```
kinit 组件业务用户
```

5. 直接执行命令。例如：

```
bin/beeline -u 'jdbc:subprotocol://host:port'

- subprotocol:kyuubi or hive2

- host: IP address of the kyuubi server
```

4.4.3 使用 Spark 客户端

前提条件

已安装客户端。

例如安装目录为“/opt/hadoopclient”，以下操作的客户端目录只是举例，请根据实际安装目录修改。

使用 Spark 客户端

1. 安装客户端。
2. 登录安装客户端的节点。
3. 执行以下命令，切换到客户端安装目录。

```
cd /opt/hadoopclient
```

4. 集群默认为安全模式，执行以下命令进行用户认证。

```
kinit 组件业务用户
```

5. 直接执行命令。例如：

```
spark-shell --master spark://IP:port
```

4.5 访问组件 Web 页面

4.5.1 开源组件 Web 站点

场景介绍

翼 MR 集群默认在集群的 Master 节点和 Core 节点创建并托管了不同组件的 Web 站点，用户可以通过这些 Web 站点查看组件相关信息。

访问开源组件 Web 站点步骤：

1. 登录翼 MR 控制台管理页面。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“访问链接与端口”tab 即可。

Web 站点一览

详见下表：开源组件 Web 站点地址

站点类型	WebURL	例子	URL 说明
HDFS	http://nn-ip(nn-hostname):50070	http://127.0.0.1:5007	NameNode

站点类型	WebURL	例子	URL 说明
		0	Web UI
YARN	http://rm-ip(rm-hostname):8088	http://127.0.0.1:8088	ResourceManger Web UI
YARN	http://nm-ip(nm-hostname):8042	http://127.0.0.1:8042	NodeManager Web UI
YARN	http://mrhis-ip(mrhis-hostname):19888	http://127.0.0.1:19888	JobHistory Web UI
HBase	http://hm-ip(hm-hostname):16010	http://127.0.0.1:16010	HMaster Web UI
HBase	http://rs-ip(rs-hostname):16030	http://127.0.0.1:16030	RegionServer Web UI
Spark	http://sparkhis-ip(sparkhis-hostname):18081	http://127.0.0.1:18081	SparkHistory Web UI
Doris	http://fe-ip(fe-hostname):8035	http://127.0.0.1:8035	FE Web UI
Elasticsearch	http://node-ip(node-hostname):9200	http://127.0.0.1:9200	ESNode Web UI
Ranger	http://rgadm-ip(raadm-hostname):6080	http://127.0.0.1:6080	RangerAdmin Web UI
Knox	https://knox-ip(knox-hostname):7743	https://127.0.0.1:7743	Knox Web UI
KafkaUI	http://kafkaui_IP(KafkaUI-hostname):9215	https://127.0.0.1:9215	Kafka Web UI
FlinkUI	http://historyserver_IP(flinkhistoryserver-hostname):8082	https://127.0.0.1:8082	Flink Web UI
TezUI	http://tezui_IP(TezUI-hostname):8822/tez-ui/	https://127.0.0.1:8822/tez-ui/	Tez Web UI
Amoro	https://amoro-ip(amoro-hostname):163	https://127.0.0.1:163	Amoro

站点类型	WebURL	例子	URL 说明
	0	0	Web UI
Hue	https://hue-ip(hue-hostname):8888	https://127.0.0.1:8888	Hue Web UI

4.5.2 开源组件端口列表

背景

随着纳管的组件越来越多，各节点混部时，尤其在测试验证阶段，不同节点完全混部，会出现一些常见端口冲突，导致一些服务无法启动。

方法

通过 netstat 对应 pid 命令、官方配置文件，当前组件配置文件来确认组件常见固定端口。

结果

产出各组件常用端口，通过 Linux 机器初始化时，对这些常用端口进行 net.ipv4.ip_local_reserved_ports 设置，预留这些端口。

常见组件 LIST

组件有 HDFS、YARN、HBase、Hive、Kafka、Spark、ZooKeeper、Kerberos、Trino、OpenLDAP、Doris、Elasticsearch、Kyuubi、Flume、Ranger、Knox 等。

HDFS 常用端口

配置参数	默认端口	当前配置端口	端口说明
dfs.namenode.http-address.x.x	9870	50070	namenode Web UI 绑定端口
dfs.namenode.https-address.x.x	9871	9871	namenode https 绑定端口
dfs.namenode.rpc-address.x.x	8020	54310	namenode 响应 client 请求 RPC 绑定端口

配置参数	默认端口	当前配置端口	端口说明
dfs.namenode.servicerpc-address.x.x	-	53310	namenode 响应内部服务请求 RPC 绑定端口，例如 datanode 等 server 请求
com.sun.management.jmxremote.port	-	8006	namenode jmx 端口
dfs.datanode.address	9866	1004	datanode 数据传输端口
dfs.datanode.http.address	9864	1006	datanode http webUI 端口
dfs.datanode.ipc.address	9867	9867	datanode rpc 响应端口
com.sun.management.jmxremote.port	-	8016	datanode jmx 端口
dfs.datanode.https.address	9865	-	datanode https 端口
dfs.journalnode.rpc-address	8485	8485	journalnode rpc 端口
dfs.journalnode.http-address	8480	8480	journalnode http 端口
dfs.journalnode.https-address	8481	8481	journalnode https 端口
com.sun.management.jmxremote.port	-	8046	journalnode jmx 端口
dfs.ha.zkfc.port	8019	8019	zkfc rpc 端口
com.sun.management.jmxremote.port	-	8056	zkfc jmx 端口
dfs.federation.router.rpc-address	8888	55310	hdfs router 响应 client 请求 rpc 处理端口
dfs.federation.router.admin-address	8111	-	hdfs router 响应 admin 请求 rpc 处理端口
dfs.federation.router.http-address	50071	-	hdfs router http 端口
dfs.federation.router.https-address	50072	-	hdfs router https 端口
com.sun.management.jmxremote.port	-	8106	hdfs router jxm 端口

YARN 常用端口

配置参数	默认端口	当前配置端口	端口说明
yarn.resourcemanager.address	8032	8032	RM 应用管理端口
yarn.resourcemanager.scheduler.address	8030	8030	RM 调度端口
yarn.resourcemanager.webapp.address	8088	8088	RM http webUI 端口
yarn.resourcemanager.webapp.https.address	8090	8090	RM https webUI 端口
yarn.resourcemanager.resource-tracker.address	8031	8031	RM jmx 端口
yarn.resourcemanager.admin.address	8033	8033	RM admin 端口
yarn.nodemanager.address	-	45454	NM container 管理端口
com.sun.management.jmxremote.port	-	8026	RM jmx 端口
yarn.nodemanager.localizer.address	8040	8040	NM 本地化 IPC 端口
yarn.nodemanager.collector-service.address	8048	8048	NM 控制服务端口
yarn.nodemanager.webapp.address	8042	8042	NM http webUI 端口
yarn.nodemanager.webapp.https.address	8044	8044	NM https webUI 端口
yarn.timeline-service.address	10200	10200	timeline server RPC 端口
yarn.timeline-service.webapp.address	8188	8188	timeline server http webUI 端口
yarn.timeline-service.webapp.https.address	8190	8190	timeline server https webUI 端口
yarn.sharedcache.admin.address	8047	8047	SCM 服务 admin RPC 端口
yarn.sharedcache.webapp.address	8788	8788	SCM 服务 webUI 端口
yarn.sharedcache.uploader.server.address	8046	8046	SCM 服务 节点管理 RPC 端口

配置参数	默认端口	当前配置端口	端口说明
yarn.sharedcache.client-server.address	8045	8045	SCM 服务响应 client RPC 端口
yarn.nodemanager.amrmproxy.address	8049	8049	在 yarn federation 下开启 amrmproxy 绑定的服务端口
yarn.router.webapp.address	8089	8089	yarn route http webUI 端口
yarn.router.webapp.https.address	8091	8091	yarn route https webUI 端口
com.sun.management.jmxremote.port	8036	8036	NM jmx 端口
com.sun.management.jmxremote.port	-	8086	mrhistory jmx 端口
mapreduce.jobhistory.address	10020	10020	mrhistory IPC 端口
mapreduce.jobhistory.webapp.address	19888	19888	mrhistory http webUI 端口
mapreduce.jobhistory.webapp.https.address	19890	19890	mrhistory https webUI 端口
mapreduce.jobhistory.admin.address	10033	10033	mrhistory admin RPC 端口
mapreduce.shuffle.port	13562	13562	MR shuffle 端口
spark.shuffle.service.port	7337	7337	spark shuffle 端口

HBase 常用端口

配置参数	默认端口	当前配置端口	端口说明
hbase.master.port	16000	16000	HMaster 服务绑定端口
hbase.master.info.port	16010	16010	HMaster webUI 绑定端口
com.sun.management.jmxremote.port	-	10101	HMaster jmx 端口

配置参数	默认端口	当前配置端口	端口说明
hbase.regionserver.port	16020	16020	regionserver 服务绑定端口
hbase.regionserver.info.port	16030	16030	regionserver webUI 绑定端口
com.sun.management.jmxremote.port	—	10102	regionserver jmx 端口
hbase.zookeeper.peerport	2888	2888	zk 服务端口
hbase.zookeeper.leaderport	3888	3888	zk 选举 leader 端口
hbase.zookeeper.property.clientPort	2181	2181	连接 zk client 端口
hbase.rest.port	8080	8080	hbase rest server 端口
hbase.status.multicast.address.port	16100	16100	hbase 多播端口

Hive 常用端口

配置参数	默认参数	当前配置端口	端口说明
hive.server2.thrift.port	10000	10000	hive.server2.transport.mode 模式为 binary hs2 thrift 端口
hive.server2.webui.port	10002	10002	hs2 webUI 端口
hive.server2.thrift.http.port	10001	10003	hive.server2.transport.mode 模式为 http hs2 thrift 端口
com.sun.management.jmxremote.port	—	9097	hs2 jmx 端口
hive.llap.management.rpc.port	15004	15004	LLAP 管理 rpc 端口
hive.llap.daemon.yarn.shuffle.port	15551	15551	LLAP yarn shuffle 端口
hive.llap.daemon.web.port	15002	15002	LLAP webUI 端口
hive.llap.daemon.output.service.port	15003	15003	LLAP output 服务端口

配置参数	默认参数	当前配置端口	端口说明
hive.zookeeper.client.port	2181	2181	连接 zk client 端口
hive.metastore.port	9083	9083	metastore RPC 绑定端口
com.sun.management.jmxremote.port	—	9093	metastore jmx 端口

Kafka 常用端口

配置参数	默认参数	当前配置端口	端口说明
com.sun.management.jmxremote.port	—	8096	kafka broker jmx 端口
zookeeper.connect	2181	2181	连接 zk client 端口
listeners(SASL_PLAINTEXT)	9092	9092	kafka SASL_PLAINTEXT 端口
listeners(PLAINTEXT)	9091	9091	kafka PLAINTEXT 端口

Spark 常用端口

配置参数	默认参数	当前配置端口	端口说明
spark.yarn.historyServer.address	—	18081	spark history webUI 端口
spark.shuffle.service.port	7337	7337	spark shuffle 端口
spark.ui.port	4040	4040	应用 dashboard UI 端口

ZooKeeper 常用端口

配置参数	默认参数	当前配置端口	端口说明
admin.serverPort	8080	8080	zk admin 服务端口
clientPort	2181	2181	连接 zk client 端口

配置参数	默认参数	当前配置端口	端口说明
com.sun.management.jmxremote.port	2182	2182	zk jmx 端口
server.x	2888:3888;2181	2888:3888;2181	zk 服务端口、选举 leader 端口，client 端口。

Kerberos 常用端口

配置参数	默认端口	当前配置端口	端口说明
kdc_ports	88	88	kdc 端口
iprop_port	744	744	iprop 端口
kpasswd	464	464	kpasswd 端口
admin_server	749	749	kerberos admin server 端口

Trino 常用端口

配置参数	默认端口	当前配置端口	端口说明
erver.http.port	—	9808	trino http 端口

OpenLDAP 常用端口

配置参数	默认端口	当前配置端口	端口说明
provide	389	389	ldap TLS connections 端口
provide	636	636	ldap legacy SSL connections 端口

Doris 常用端口

配置参数	默认端口	当前配置端口	端口修改	端口说明
be_port	9060	9060	—	BE 上 thrift server 的端口号，用于接收来

配置参数	默认端口	当前配置端口	端口修改	端口说明
				自 FE 的请求。
brpc_port	8060	8060	—	BE 上的 brpc 的端口，用于 BE 之间通讯。
edit_log_port	9010	9010	—	bdbje 端口
heartbeat_service_port	9050	9050	—	BE 上心跳服务端口（thrift），用于接收来自 FE 的心跳。
http_port	8030	8030	8035	FE http 端口，当前所有 FE http 端口都必须相同，生产环境被改成了 8035。
https_port	8050	—	—	FE https 端口，当前所有 FE https 端口都必须相同。
query_port	9030	9030	—	Doris FE 通过 mysql 协议查询连接端口
rpc_port	9020	9020	—	FE Thrift Server 的端口
single_replica_load_brpc_port	9070	—	—	单副本数据导入功能中，Master 副本和 Slave 副本之间通信的 RPC 端口。Master 副本 flush 完成之后通过 RPC 通知 Slave 副本同步数据， 以及 Slave 副本同步数据完成后通过 RPC 通知 Master 副本。系

配置参数	默认端口	当前配置端口	端口修改	端口说明
				统为单副本数据导入过程中Master 副本和Slave 副本之间通信开辟了独立的 BRPC 线程池，以避免导入并发较大时副本之间的数据同步抢占导入数据分发和查询任务的线程资源。
single_replica_load_download_port	8050	—	—	单副本数据导入功能中，Slave 副本通过 HTTP 从Master 副本下载数据文件的端口。系统为单副本数据导入过程中 Slave 副本从Master 副本下载数据文件开辟了独立的HTTP 线程池，以避免导入并发较大时 Slave 副本下载数据文件抢占其他 http 任务的线程资源。
webserver_port	8040	8045	8045	BE 上的 http server 的服务端口

Elasticsearch 常用端口

配置参数	默认端口	当前配置端口	端口说明
com.sun.management.jmxremote.port	—	9400	es jmx 端口
http.port	—	9200	es http 端口

配置参数	默认端口	当前配置端口	端口说明
transport.port	—	9300	es transport 端口

Kyuubi 常用端口

配置参数	默认参数	当前配置端口	端口说明
kyuubi.frontend.bind.port	10009	—	kyuubi fe 端口
kyuubi.frontend.mysql.bind.port	3309	—	kyuubi fe mysql 绑定端口
kyuubi.frontend.rest.bind.port	10099	—	kyuubi fe rest 绑定端口
kyuubi.frontend.thrift.binary.bind.port	10009	—	kyuubi thrift fe 绑定端口
kyuubi.metrics.prometheus.port	10019	—	kyuubi 暴露给 Prometheus metrics http 端口

Ranger 常用端口

配置参数	默认参数	当前配置端口	端口说明
ranger.usersync.port	5151	—	ranger usersync 服务端口
com.sun.management.jmxremote.port	—	60081	ranger usersync jmx 端口
ranger.unixauth.service.port	5151	—	ranger unixauth 服务端口
ranger.service.shutdown.port	6085	—	ranger 服务 shutdown 端口
ranger.audit.elasticsearch.port	9200	—	ranger audit 配置的 es 服务端口
ranger.service.http.port	6080	6080	ranger—admin http webUI 端口
ranger.service.https.port	6182	—	ranger—admin https webUI 端口

Knox 常用端口

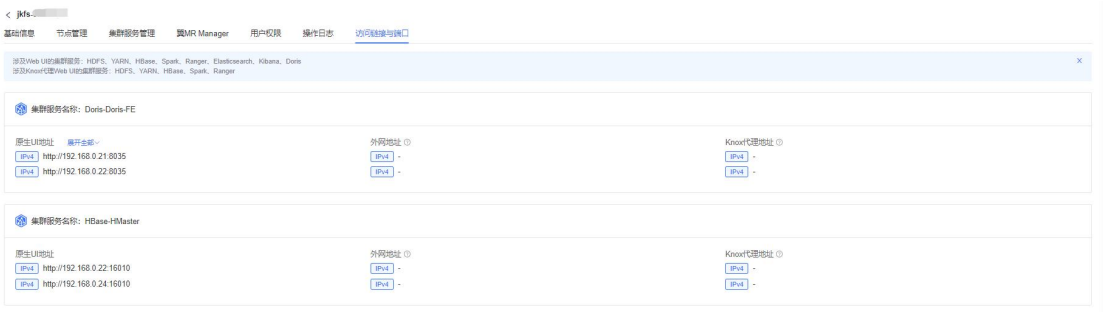
配置参数	默认参数	当前配置端口	端口说明
gateway.port	8443	7743	https webUI 端口

4.5.3 通过弹性公网 IP 访问

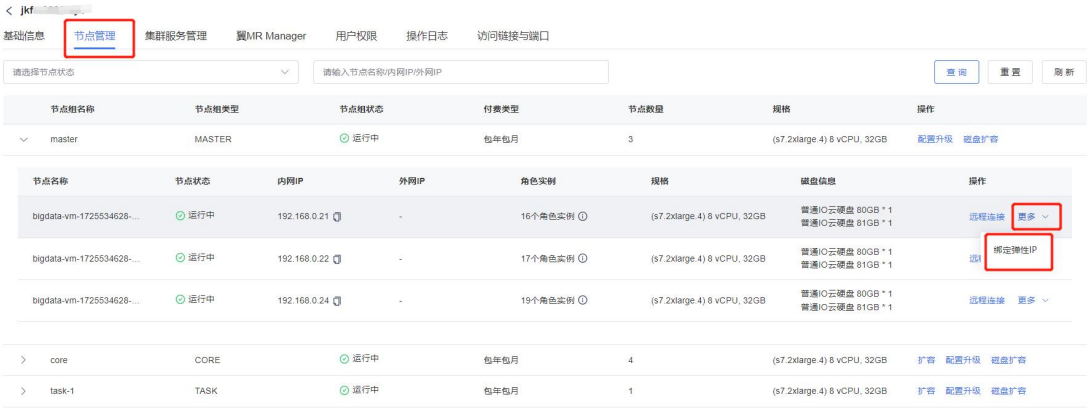
为了方便用户访问开源组件的 Web 站点，翼 MapReduce 集群支持通过为集群绑定弹性公网 IP 的方式，访问集群上托管的开源组件。该方式更加简便易操作，推荐使用该方式访问开源组件的 Web 站点。

绑定弹性公网 IP

1. 在集群列表页面，单击“集群名称”列下需要绑定 IP 的集群名称，进入该集群信息页面。
2. 确定需要开通公网访问组件所在的节点。可以通过单击“访问链接与端口”，在“集群服务名称”中找到需要开通公网访问的组件，其“原生 UI 地址”中包含所在节点内网 IP 地址。



3. 单击“节点管理”，点击“节点组名称”列的下拉按钮，展开对应的节点信息，确定需要开通公网访问的节点，单击“操作”列的“更多”。



4. 单击“绑定弹性 IP”，在“绑定弹性 IP”的弹框中，如果您账号下没有可用的弹性公网 IP，需要点击“+创建弹性公网 IP”按钮跳转至新页面进行创建；如果您账号下有可用的弹性公网 IP，可以通过下拉“弹性 IP”的选择框选择 IP 后，点击“确认”进行绑定。

5. 绑定弹性 IP 后，根据需要对集群安全组规则进行相应的添加或修改，变更端口的访问权限。之后可以通过“公网 IP：端口”的方式访问该开源组件的 Web 站点。

4.6 翼 MR Manager 操作指导

4.6.1 从这里开始

4.6.1 翼 MR Manager 入门指导

概述

翼 MR Manager 由天翼云自主研发，主要为运维工程师提供日常的大数据组件运维管理操作能力。通过可视化、流程化的方式对系统中的各个系统资源和数据资产进行管理，并支持自动化的运维调度、统一的配置文件管理、统一运维监控，支持组件集群服务管理、多租户管理、资源管理等功能。在提高运维人员工作效率的同时，为大数据运维工作者提供专业全面的运维能力，从而大幅降低大数据平台的运维门槛。

- 提供集群服务管理，为用户快速掌握环境、集群、主机、组件服务等数据信息。
- 提供运维自动化管理，提高大数据运维效率，降低人力成本。

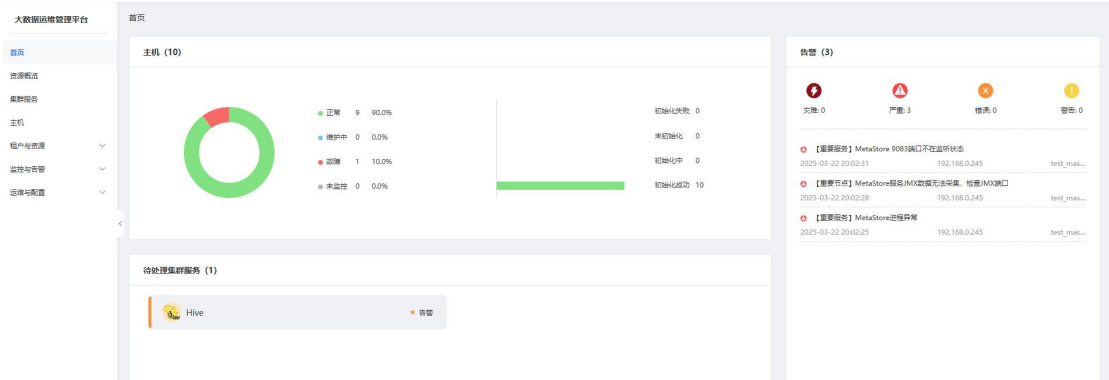
浏览器支持

建议使用 Google Chrome 浏览器 。

系统界面简介

翼 MR Manager 提供统一的集群管理平台，帮助您快捷、直观的完成集群的运行维护。

详见下图：翼 MR Manager 系统界面。



界面左侧为菜单导航区域，右侧为显示区域和操作区域。

菜单导航区域的详细功能如下表所示：

菜单导航	功能描述
首页	展示平台主机健康状态、初始化状态汇总； 故障和告警集群信息； 平台告警信息。
资源概览	展示所有主机的 CPU、内存、网络等信息。
集群服务	展示所有集群，集群列表按组件类型，列表视图列出，在集群列表处一键启动所有集群、一键停止所有集群，支持健康检查； 支持单集群的启动、停止、滚动重启等操作； 支持角色实例的启动、停止、重启、滚动停止、滚动重启等操作； 支持查看集群的告警历史和监控看板等。
主机	展示当前平台的所有主机； 支持对主机进行置维护、取消维护操作； 支持查看单个主机基础信息、主机资源使用信息、主机上的文件系统信息； 支持查看单个主机上安装的角色实例、告警历史。

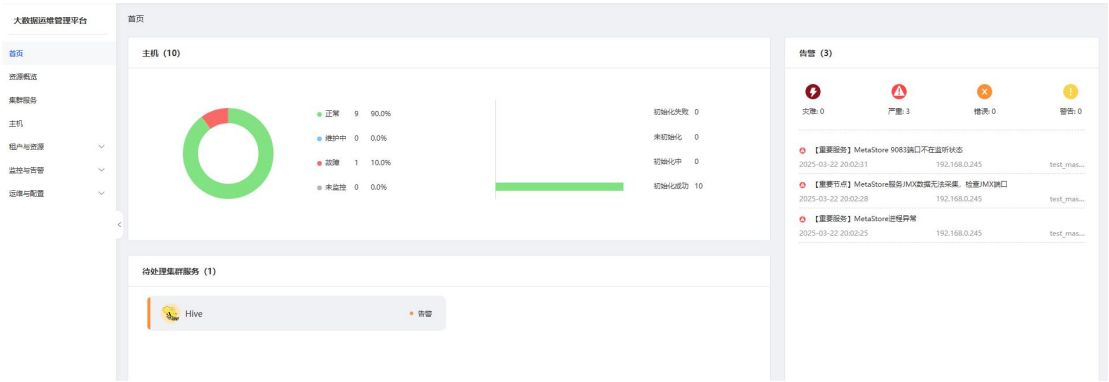
菜单导航	功能描述
租户与资源	支持对 LDAP 租户与 Kerberos 安全凭证进行管理； 提供 YARN 队列管理与 HBase 数据库管理功能。
监控与告警	支持查询角色实例级、主机级的监控指标； 支持指标结果的绘图操作，让用户更直观获取监控项变化； 支持按需创建和导入告警规则； 支持按照集群级、角色实例级、主机级查询告警内容。
运维与配置	支持对各集群服务的配置文件进行修改、同步、回滚等操作； 支持查看所有流水线的运行历史记录，以及操作人。

4.6.2 首页

4.6.2 首页概述

操作步骤如下：

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab。
4. 单击“前往翼 MR Manager”，默认进入到首页，如图所示：



- 主机：展示平台主机健康状态汇总、主机初始化状态汇总。

- 待处理集群服务：展示故障和告警的集群服务信息。
- 告警：展示待处理的告警信息，支持查看不同级别的告警信息。

4.6.3 资源概览

4.6.3 资源概览概述

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab。
4. 单击“前往翼 MR Manager”。
5. 进入翼 MR Manager 后，点击菜单“资源概览”，进入资源概览页面。
6. 默认展示最新 3 小时的数据，可通过右上角时间范围选框，选择其他时间段的监控数据。

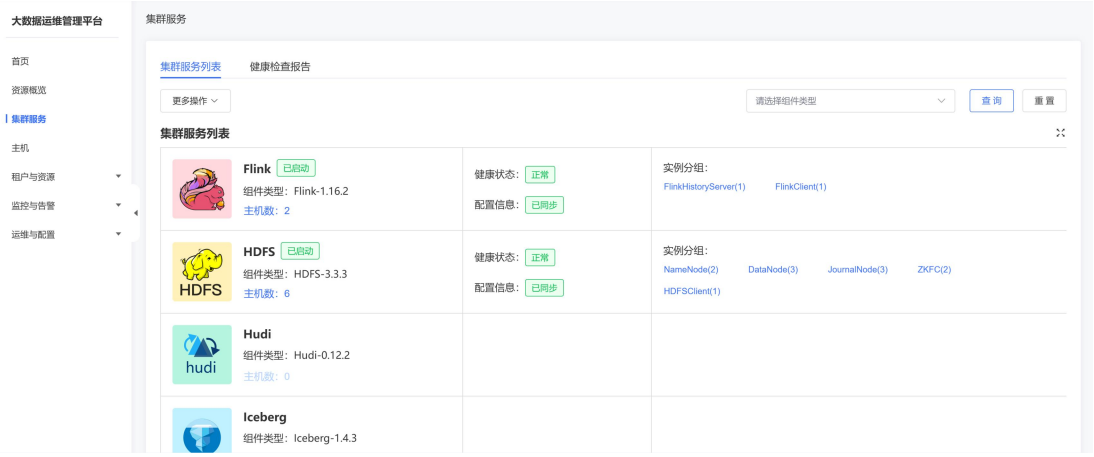
资源概览展示当前平台所有主机最近 5 个小时的 CPU 使用率、磁盘使用率、内存使用率、网络发送速率、网络接收速率信息。如图所示：

4.6.4 集群服务

4.6.4 集群服务管理概述

操作步骤如下：

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab。
4. 单击“前往翼 MR Manager”。
5. 进入翼 MR Manager 后，点击菜单“集群服务”，进入集群服务页面，如图所示：



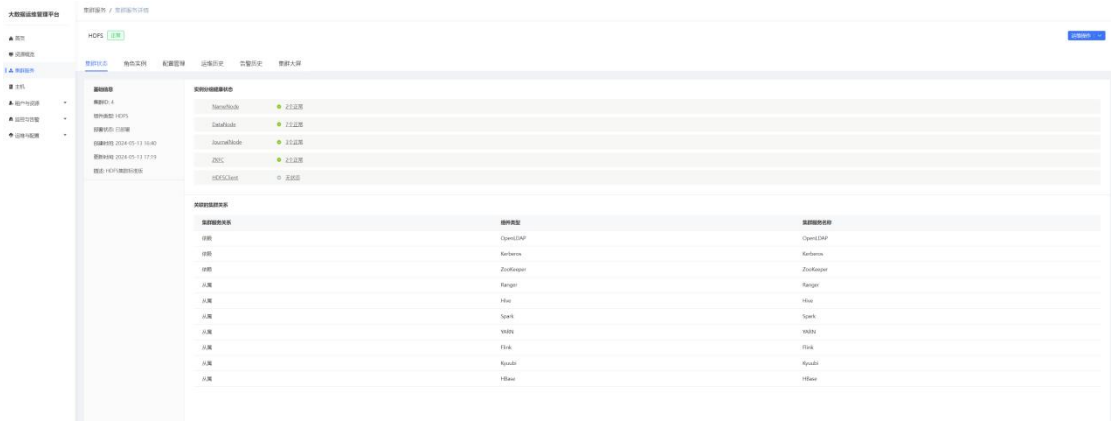
- 集群服务列表上方为查询区域，支持组件类型进行查询。
- 集群服务列表按组件类型，列表视图列出。
- 集群服务列表右上角，可支持以下操作：启动所有集群服务、停止所有集群服务、健康检查。
- 点击集群服务 ICON 或集群服务名称可进入到集群服务详情页面。

- 点击主机数可进入到主机菜单页面。
- 点击实例分组可进入到集群详情-角色实例页面。

4.6.4 集群服务详情页面概述

操作步骤如下：

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab。
4. 单击“前往翼 MR Manager”。
5. 进入翼 MR Manager 后，单击菜单“集群服务”，进入集群服务列表页面。
6. 单击指定集群服务的名称进入集群服务详情页面，如图所示：



- 详情页面上面区域为集群服务名称、集群服务的健康状态。
- 详情页面右上角为集群服务的运维操作，包含：启动集群服务、停止集群服务、滚动重启集群服务等。
- 详情页面上面区域则依次为该集群服务的：集群状态、角色实例、配置管理、运维历史、告警历史、集群大屏等信息。

4.6.4 启动、停止所有集群服务

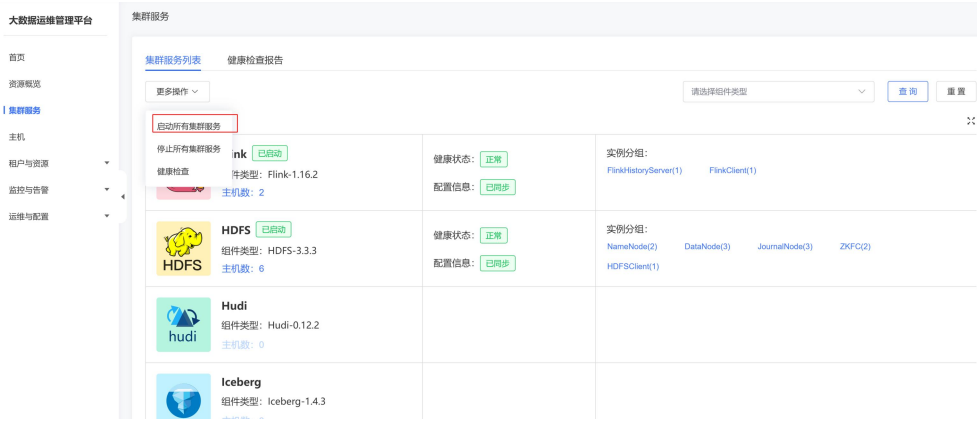
启动所有集群服务

操作场景

一键启动所有集群服务。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 单击“更多操作>启动所有集群服务”，弹出确认操作弹框。
6. 单击“确定”，开始启动所有集群服务，等待所有集群服务完成启动后即可。



注意

1. 启动所有集群服务时按照集群依赖关系串行执行。
2. 在执行过程中，可以取消当前操作，已启动的集群服务无法取消。

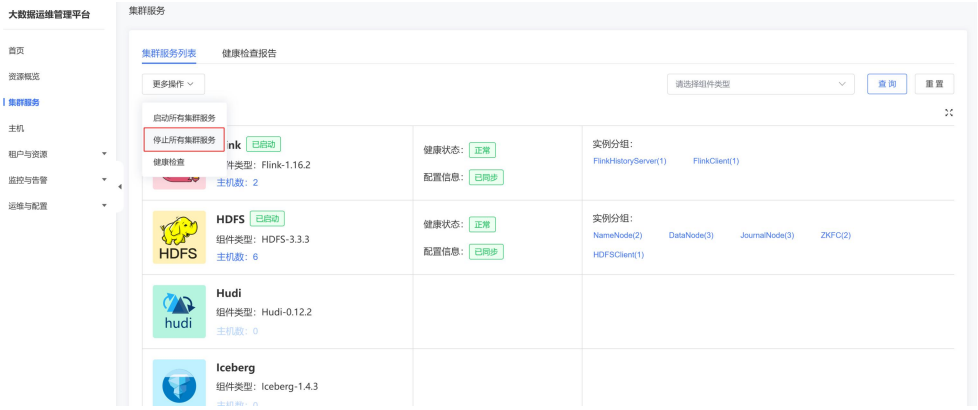
停止所有集群服务

操作场景

一键停止所有集群服务。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 单击“更多操作>停止所有集群服务”，弹出确认操作弹框。
6. 单击“确定”，开始停止所有集群服务，等待所有集群服务完成停止后即可。



注意

1. 停止所有集群服务时按照集群依赖关系串行执行。
2. 在执行过程中，可以取消当前操作，已停止的集群服务无法取消。

4.6.4 启动、停止单集群服务

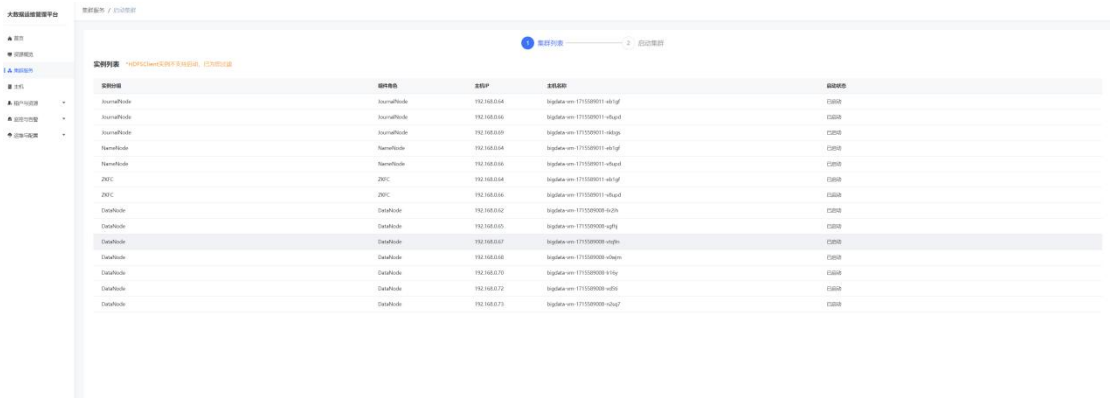
启动集群服务

操作场景

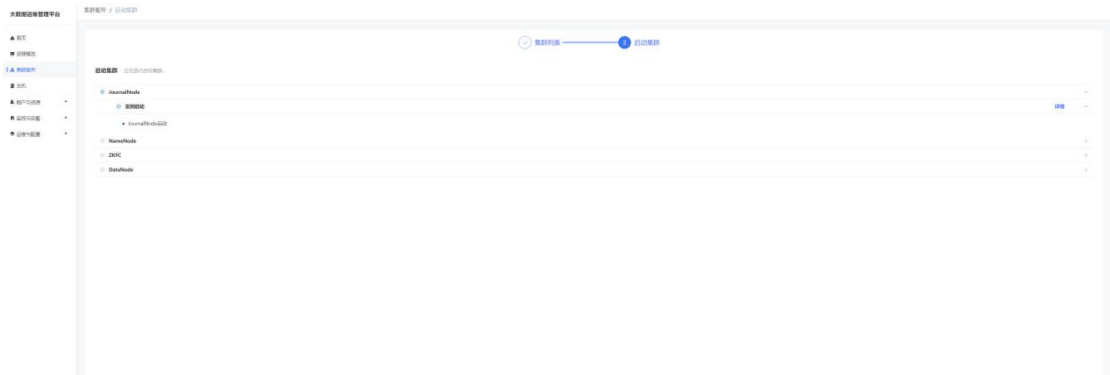
一键启动单个集群服务。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“运维操作>启动集群服务”，弹出确认操作弹框。
7. 单击“确定”，进入到启动集群服务页面，展示当前集群服务支持启动的实例列表。如图所示：



8. 单击“下一步”，按照实例分组顺序进行实例启动，等待所有实例完成启动后即可。如图所示：



停止集群服务

操作场景

一键停止单个集群服务。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“运维操作>停止集群服务”，弹出确认操作弹框。
7. 单击“确定”，进入到停止集群服务页面，展示当前集群服务支持停止的实例列表。
8. 单击“下一步”，按照实例分组顺序进行实例停止，等待所有实例完成停止后即可。

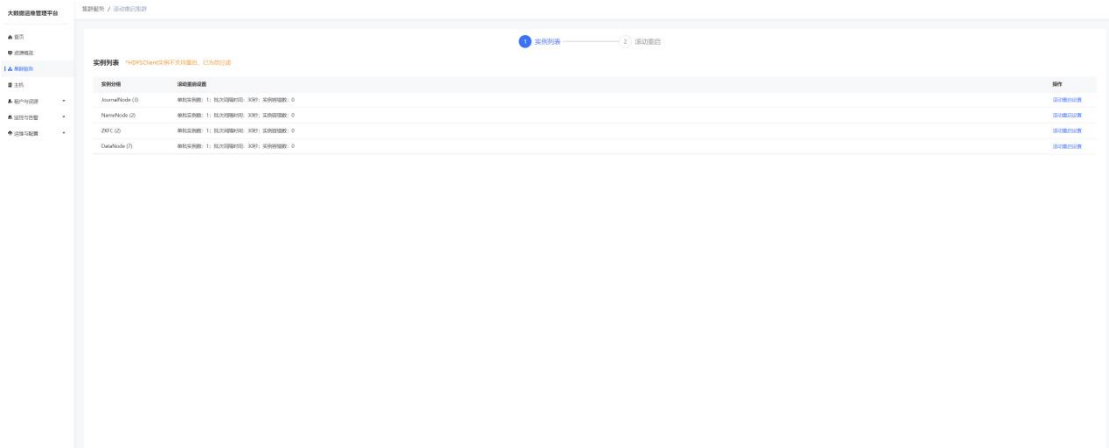
4.6.4 滚动重启单个集群服务

操作场景

一键滚动重启单个集群服务。

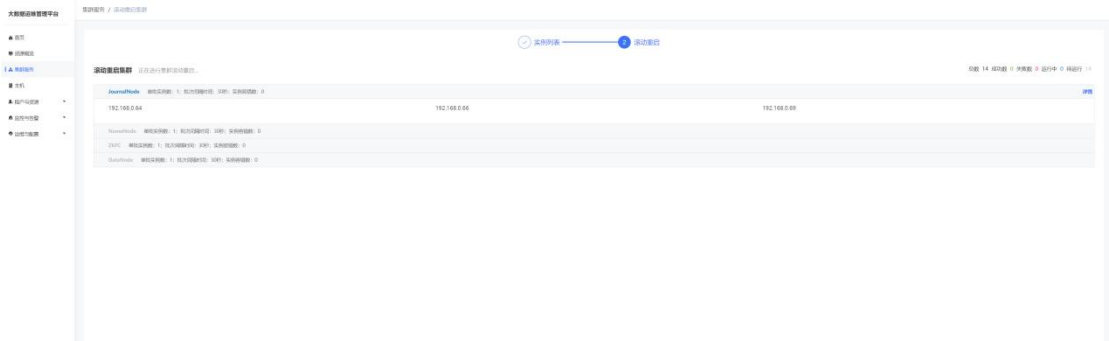
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“运维操作>滚动重启集群服务”，弹出确认操作弹框。
7. 单击“确定”，进入到滚动重启集群服务页面，展示当前集群服务支持滚动重启的实例列表。如图所示：



8. 选择指定实例分组，单击“滚动重启设置”，对默认滚动重启设置进行修改，修改完成后单击“确定”，如图所示：

9. 单击“下一步”，按照实例分组顺序进行滚动重启，等待所有实例完成重启后即可。如图所示：



滚动重启参数说明

滚动重启参数说明如下表所示：

参数名称	描述
单批实例数	1. 按照角色实例维度，每1个批次的并发实例数，例如HDFS-DataNode。 2. 默认为 1，输入限制为 int 的最大值，仅支持输入正整数。 3. 当角色实例数小于并发数时，以当前已有实例数为准例如：并发数设置为 2，HDFS-DataNode 的实例数为 3；则第一批次执行的实例数是 2，第 2 批次执行的实例数则为 1。
批次间隔时间	1.上个批次成功后与下个批次开始的间隔时间。 2. 默认 30 秒，取值范围 1-1800，仅支持输入正整数；单位默认为：秒，不可改。举例：设置时间为 600S 批次间隔则表示，前 1 个批次运行完成后，等待 600S 后开始下一个批次。
批次等待时间	1. 上个批次开始后，开始下个批次的等待时间。 2. 默认 30 秒，取值范围 1-1800，仅支持输入正整数；单位默认为：

参数名称	描述
	秒，不可改。举例：设置时间为 600S 批次等待则表示，前 1 个排次开始运行，等待 600S 后不管前一个批次是否完成运行，都开始下一个批次的运行。
实例容错阈值	1．实例执行失败的容错次数。 2．默认为 0，输入限制为 int 的最大值，仅支持输入正整数例如：设置为 0 时，即表示任意一个角色实例的操作失败后，滚动操作终止设置为 2 时，即表示任意 3 个角色实例的操作失败后，滚动操作才会终止。

4.6.4 启动、停止、重启实例

启动实例

操作步骤

1．登录翼 MR 管理控制台。

2．单击“我的集群”，单击指定的集群名称，进入集群信息页面。

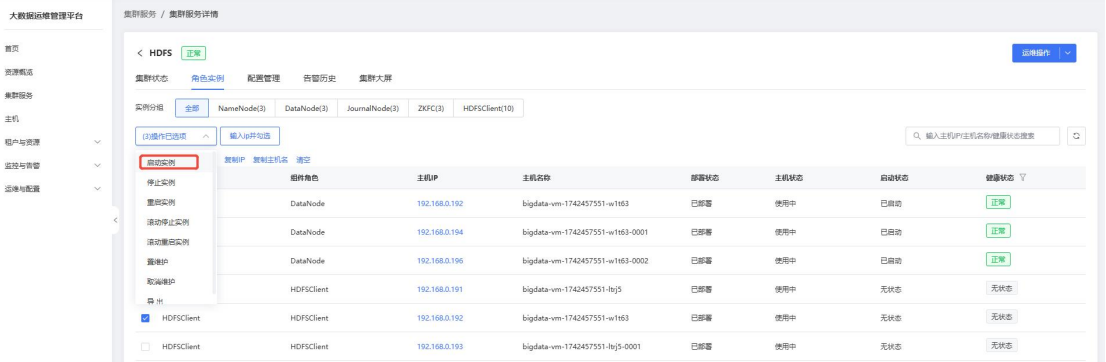
3．单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4．进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。

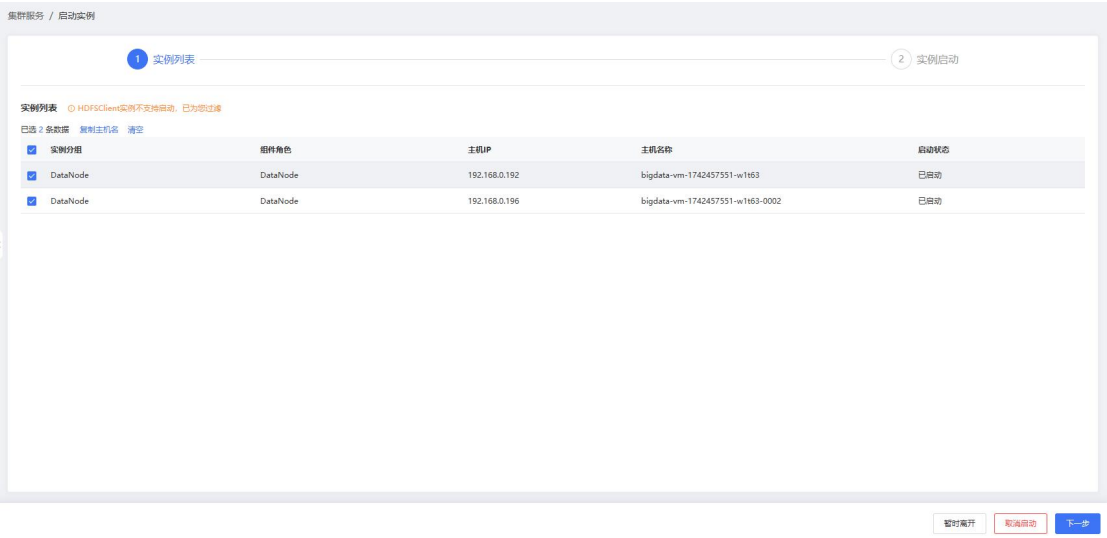
5．选择指定集群服务，单击集群服务名称进入集群服务详情页面。

6．单击“角色实例”tab。

7．勾选需要操作角色实例前的复选框，单击“操作已选项 ＞ 启动实例”。



8．在弹窗中单击“确认”，进入到启动实例页面，展示当前集群服务支持启动的实例列表。如图所示：



9. 单击“下一步”，按照实例分组顺序进行实例启动，等待所有实例完成启动后即可。

停止实例

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项 > 停止实例”，弹出确认操作弹框。
9. 单击“确认”，进入到停止实例页面，展示当前集群服务支持停止的实例列表。
10. 单击“下一步”，按照实例分组顺序进行实例停止，等待所有实例完成停止后即可。

重启实例

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

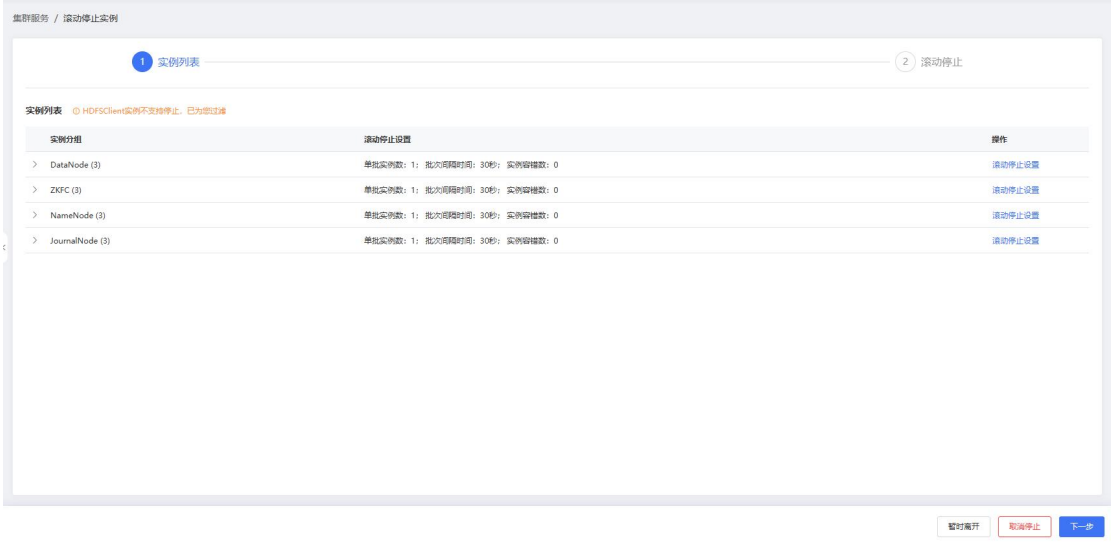
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项 > 重启实例”，弹出确认操作弹框。
9. 单击“确认”，进入到重启实例页面，展示当前集群服务支持重启的实例列表。
10. 单击“下一步”，按照实例分组顺序进行实例重启，等待所有实例完成重启后即可。

4.6.4 滚动停止、滚动重启实例

滚动停止实例

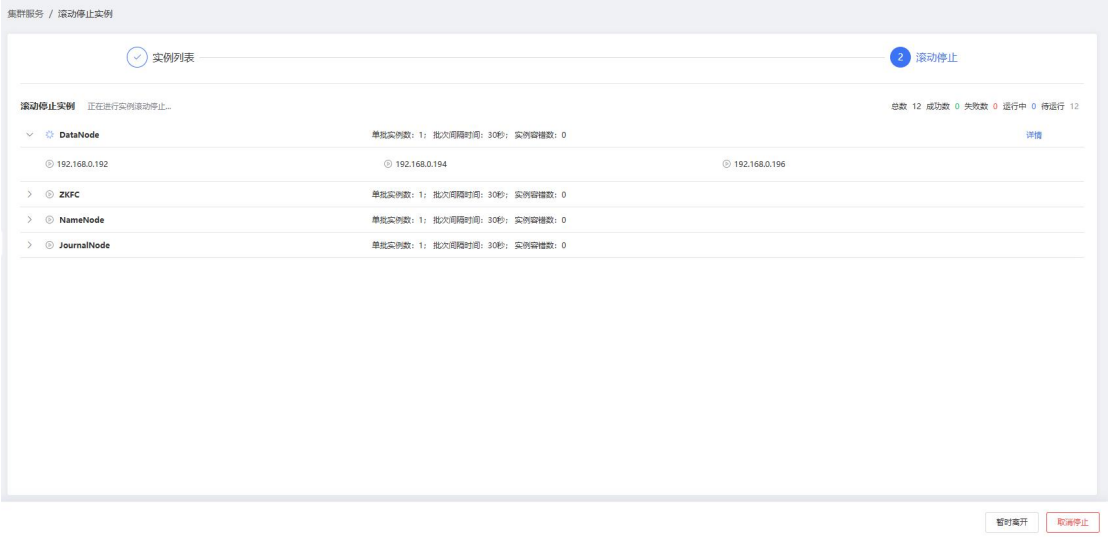
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项 > 滚动停止实例”，弹出确认操作弹框。
9. 单击“确认”，进入到滚动停止实例页面。如图所示：



10. 选择指定实例分组，单击“滚动停止设置”，对默认滚动停止设置进行修改，修改完成后单击“确定”。如图所示：

11. 单击“下一步”，按照实例分组顺序进行滚动停止，等待所有实例完成停止后即可。如图所示：



滚动重启实例

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“角色实例”tab。

7. 勾选需要操作角色实例前的复选框。

8. 单击“操作已选项 > 滚动重启实例”，弹出确认操作弹框。

9. 单击“确认”，进入到滚动重启实例页面。

10. 选择指定实例分组，单击“滚动重启设置”，对默认滚动重启设置进行修改，修改完成后单击“确定”。

11. 单击“下一步”，按照实例分组顺序进行滚动重启，等待所有实例完成重启即可。

滚动设置参数说明

滚动设置参数说明如下表所示：

参数名称	描述
单批实例数	1. 按照角色实例维度，每 1 个批次的并发实例数，例如 HDFS-DataNode。 2. 默认为 1，输入限制为 int 的最大值，仅支持输入正整数。 3. 当角色实例数小于并发数时，以当前已有实例数为准，例如：并发数设置为 2，HDFS-DataNode 的实例数为 3；则第一批次执行的实例数是 2，第 2 批次执行的实例数则为 1。
批次间隔时间	1. 上个批次成功后与下个批次开始的间隔时间。 2. 默认 30 秒，仅支持输入正整数；单位默认为：秒，不可改，举例：设置时间为 600S 批次间隔则表示，前 1 个批次运行完成后，等待 600S 后开始下一个批次。
批次等待时间	1. 上个批次开始后，开始下个批次的等待时间。 2. 默认 30 秒，仅支持输入正整数；单位默认为：秒，不可改，举例：设置时间为 600S 批次等待则表示，前 1 个排次开始运行，等待 600S 后不管前一个批次是否完成运行，都开始下一个批次的运行。
实例容错阈值	1. 实例执行失败的容错次数。 2. 默认为 0，输入限制为 int 的最大值，仅支持输入正整数，例如：设置为 0 时，即表示任意一个角色实例的操作失败后，滚动操作终止设置为 2 时，即表示任意 3 个角色实

参数名称	描述
	例的操作失败后，滚动操作才会终止。

4.6.4 实例置维护/取消维护

置维护

操作场景

用户需要对某个实例进行置维护。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项 > 置维护”，弹出置维护操作弹框。如图所示：

9. 输入备注，单击“确定”即可。

注意

1. 实例置维护后，该实例的健康状态显示为维护中。
2. 实例置维护后，则会屏蔽该实例的告警信息。

取消维护

操作场景

用户需要对某个实例取消置维护。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“角色实例”tab。
7. 勾选需要操作角色实例前的复选框。
8. 单击“操作已选项 > 取消维护”，弹出取消维护操作弹框。如图所示：

9. 输入备注，单击“确定”即可。

4.6.4 查看集群服务配置

操作步骤

1. 登录翼 MR 管理控制台。

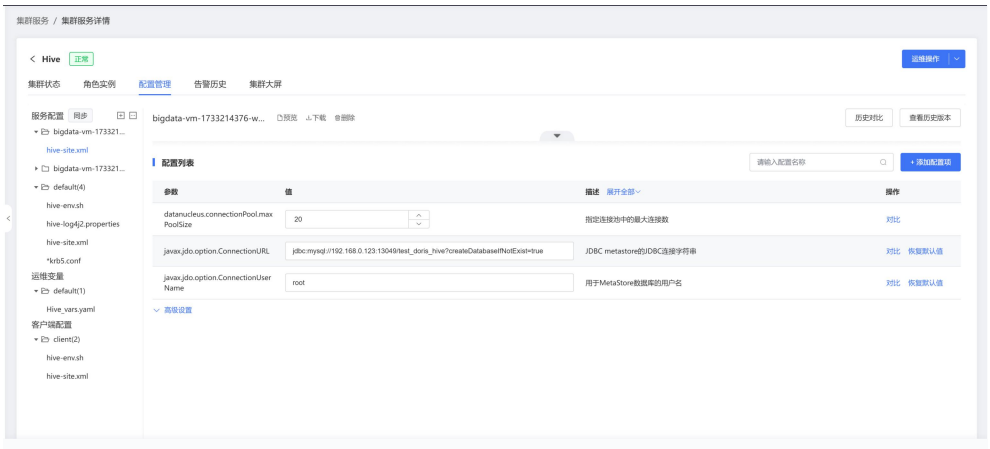
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。

5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。

6. 单击“配置管理”tab，即可查看该集群服务的所有配置。如图所示：



7. 对于隐私配置项，默认将加密展示，点击文本框右侧查看按钮，输入配置主密码即可查看。如果当前集群开通后未设置过主密码，请参考用户手册的配置管理—主密码设置，设置主密码后，进行查看。

说明—运维变量配置

在服务的 default 分组下，以_vars.yaml 后缀结尾的配置文件，为运维变量配置文件。运维变量配置文件不是服务本身的配置文件，而是存放 Manager 运维操作变量的文件，每个运维变量的含义如下表。

组件	运维变量名	是否建议修改	含义说明
HDFS	active_namenode	否	部署时使用，在这台机器上执行 format 操作。
	kdc_master	否	部署和扩容时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	ldapGuestPassword	否	部署和扩容时使用，连接 ldap 的 guest 用户密码，此权限只能查看用户，不能创建用户。
	ldap_uri	否	部署和扩容时使用，配置 ldap 的连接地址。
	dfs_clusterId	否	部署时使用，ns 的默认值。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
YARN	无	—	共享 HDFS 运维变量。
ZooKeeper	kerberosRealm	否	部署时使用，配置 kerberos 的 realm 的信息。
	kdc_master	否	部署和扩容时使用，kerberos 的

			master 节点，用于创建 keytab 文件。
	zk_super_user	否	部署和扩容时使用，配置 zk 的超级用户。
	zk_super_user_passwd	否	部署和扩容时使用，配置超级用户的密码。
	component_user	否	暂未使用。
	dataLogDir	否	部署时使用，需要通过 ansbile 创建目录。
	dataDir	否	部署时使用，需要通过 ansbile 创建目录。
Kerberos	kerberosRealm	否	部署时使用，配置 kerberos 的 realm 的信息。
	kdc_master	否	部署时使用，指定 master 节点，执行 master 节点的操作。
	kdc_slave	否	部署时使用，指定 slave 节点，执行 slave 节点的操作。
	kerberos_database_passwd	否	部署时使用，配置 kerberos 的数据库密码。
	kerberos_admin_user	否	部署时使用，配置 kerberos 的 admin 用户。
	kerberos_admin_passwd	否	部署时使用，配置 kerberos 的 admin 用户密码。
OpenLDAP	ldap_master_hostname	否	部署时使用，指定 master 节点，执行 master 节点的操作。
	ldap_slave_hostname	否	部署时使用，指定 slave 节点，执行 slave 节点的操作。
	olcRootDN	否	部署时使用，服务配置。
	olcSuffix	否	部署时使用，服务配置。
	ldapAdminPwd	否	部署时使用，admin 用户的密码。
	ldapGuestPwd	否	部署时使用，guest 用户的密码。
	ldap_uri	否	部署时使用，ldap 的连接信息。
	ldapBaseUgDN	否	部署时使用，服务配置。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
HBase	kdc_master	否	部署和扩容时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
Spark	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	ldap_master	否	部署时使用，ldap 的 master 节点。
	event_log_dir	否	部署时使用，通过 ansible 创建目录。
	spark_home	否	部署时使用，通过 ansible 创建目录。
Hive	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	ldap_master	否	部署时使用，ldap 的 master 节点。
	database_pass	否	部署时使用，hive 使用的数据库密

			码，需要用户填写。
Kafka	kdc_master	否	部署和扩容时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	kerberos_realm	否	kerberos 的 realm 的信息。
	log_dirs	否	（目前不再使用，通过从配置文件中渲染来创建目录）部署和扩容时使用，通过 ansible 创建目录。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
Kyuubi	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	ldap_master	否	部署时使用，ldap 的 master 节点。
	kyuubi_log_dir	否	目录。部署时使用，通过 ansible 创建。
	kyuubi_event_log_dir	否	目录。部署时使用，通过 ansible 创建。
	kyuubi_work_dir	否	目录。部署时使用，通过 ansible 创建。
	admin_user	否	目录的属主。
	user_group	否	目录的属组。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
Flink	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	ldap_master	否	部署时使用，ldap 的 master 节点。
	logDir	否	目录。部署时使用，通过 ansible 创建。
	HistoryServerDir	否	目录。部署时使用，通过 ansible 创建。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
Elasticsearch	CERT_NODE	否	部署时使用，指定证书节点，在节点上执行生成证书的操作。
	ES_DATA_DIRS	否	目录。部署时使用，通过 ansible 创建。
	ES_DEFAULT_PASSWORD	否	部署时使用。默认的密码，随机密码。
Doris	FE_DEFAULT_LEADER_HOSTNAME	否	部署和扩容时使用，指定 leader 节点，在上面执行 leader 的相关操作。
	FE_DEFAULT_PASSWORD	是	部署时随机生成的 Doris 组件 root 密码。节点扩容需要使用该密码，如用户已修改，需要设置正确的值，否则会影响扩容操作。
	FE_HTTP_PORT	否	服务端口，不推荐修改。
	FE_QUERY_PORT	否	服务端口，不推荐修改。
	FE_EDIT_LOG_PORT	否	服务端口，不推荐修改。
	BE_STORAGE_ROOT_PATH	否	服务端口，不推荐修改。

	BE_HEARTBEAT_SERVICE_PORT	否	服务端口，不推荐修改。
	BROKER_IPC_PORT	否	服务端口，不推荐修改。
Trino	data_dir	否	目录。部署和扩容时使用，通过 ansible 创建。
	kerberos_realm	否	kerberos 的 realm 的信息。
	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	ldap_master	否	部署时使用，ldap 的 master 节点。
	coordinator	否	部署和扩容时使用，指定主节点，在上面执行主节点的相关操作。
	ui_user	否	ui 的用户。
	ui_password	否	ui 的密码，部署时随机生成。
	server_user	否	server 的用户。
	server_password	否	server 的密码，部署时随机生成。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
Ranger	db_host	否	外置数据库信息，需要用户填写。
	db_name	否	外置数据库信息，需要用户填写。
	db_user	否	外置数据库信息，需要用户填写。
	db_password	否	外置数据库信息，需要用户填写。
	db_root_user	否	部署时使用，ranger 使用的外置数据库信息。
	db_root_password	否	部署时使用，ranger 使用的外置数据库信息。
	ranger_password	否	ranger admin 用户密码。
	SYNC_LDAP_BIND_PASSWORD	否	部署时使用，ranger 访问 ldap 的密码。
	SYNC_SOURCE	否	用户同步方式，ldap。
	kerberos_realm	否	kerberos 的 realm 的信息。
	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
	SYNC_LDAP_URL	否	ldap 相关信息。
	POLICY_MGR_URL	否	ranger admin url 信息。
	separate_dba_mode	否	是否单独执行 DBA，默认 true，表示需要提前创建 ranger 使用的数据库和用户，并授权，不推荐修改。
	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
Knox	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
Logstash	pipeline_job_retry_times	是	流水线重试次数。
	job_retry_based_on_pipeline_types	否	自动重试的流水线类型。
JeekeFS	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。
KafkaUI	kdc_master	否	部署时使用，kerberos 的 master 节点，用于创建 keytab 文件。

	ldapGuestPassword	否	部署时使用，OpenIdap 的 guest 用户密码
Pushgateway	PUSHGATEWAY_PASSWORD	否	部署时使用，Pushgateway 服务的密码
	PUSHGATEWY_USERNAME	否	部署时使用，Pushgateway 服务的用户名
TezUI	pipeline_job_retry_times	是	流水线 job 最大重试次数
	job_retry_based_on_pipeline_types	否	基于的流水线类型

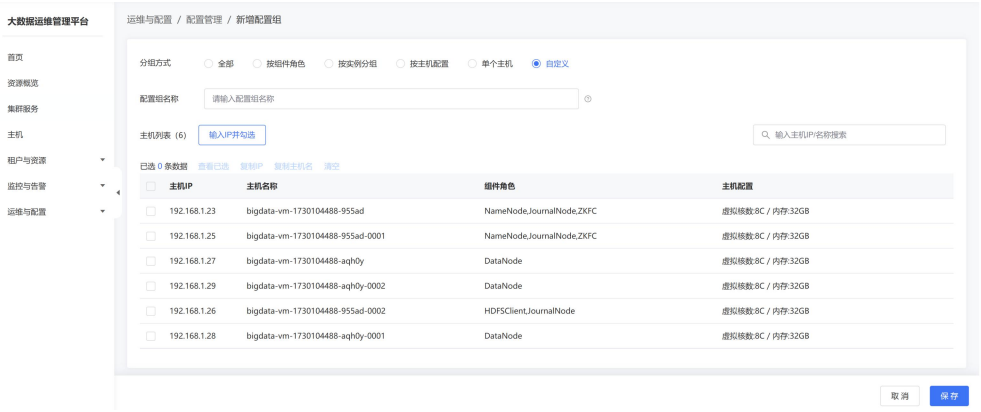
4.6.4 新增配置组

操作场景

用户可以根据不同的分组标准，对集群服务中的主机进行分组，一个配置分组下的主机共享一套配置。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击“新增配置组”，进入新增配置组页面。
8. 根据需要选择分组方式，选择主机，单击右下角“保存”按钮即可。如图所示：



注意

1. 分组方式为：全部、按组件角色、按实例分组、按主机配置时，配置组名称默认且不可改。
2. 分组方式为：单个主机时，需要选择要分组的主机，配置组名称为所选主机的名称且不可改。
3. 分组方式为：自定义时，需要选择分组的主机，配置组名称可以自定义但不能与其他配置组名称相同。

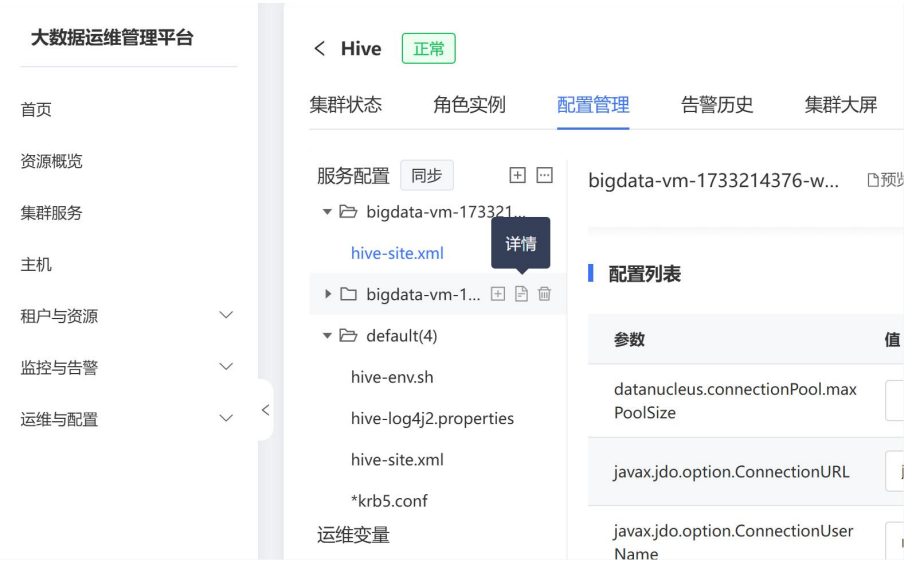
4.6.4 查看配置组详情

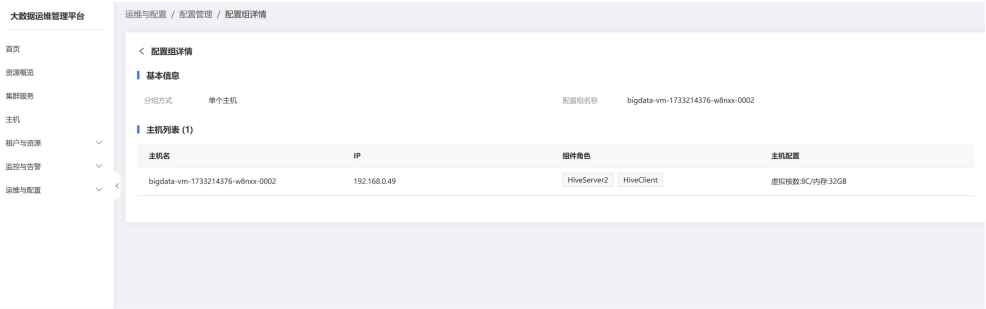
操作场景

用户可以查看不同配置组的详细信息，了解该配置组所包含的主机信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 选择要查看的配置组，鼠标移入到配置组名称上，单击右侧的“详情”图标，进入到配置组详情页面。如图所示：





4.6.4 删除配置组

操作场景

用户可以删除不需要的配置组。

前置条件

配置组下的配置文件数量必须为 0 才能删除。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 选择要删除的配置组，鼠标移入到配置组名称上，单击右侧的“删除”图标，出现删除确认弹框。

8. 单击“确定”即可删除该配置组。

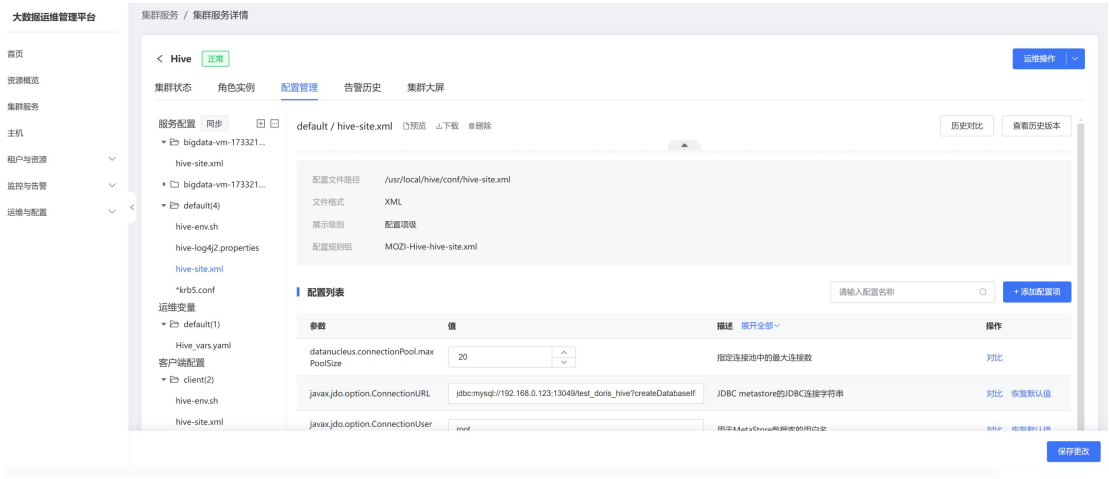
4.6.4 查看配置详情

操作场景

用户可以查看某个配置的详细信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击配置组，单击要查看详情的配置名称，页面右侧展示该配置的详细信息。如图所示：



4.6.4 新增配置

在翼 MR Manager 中支持 4 种方式的新增配置，包括新建、配置模板获取、共享其他集群配置、复制其他组配置。

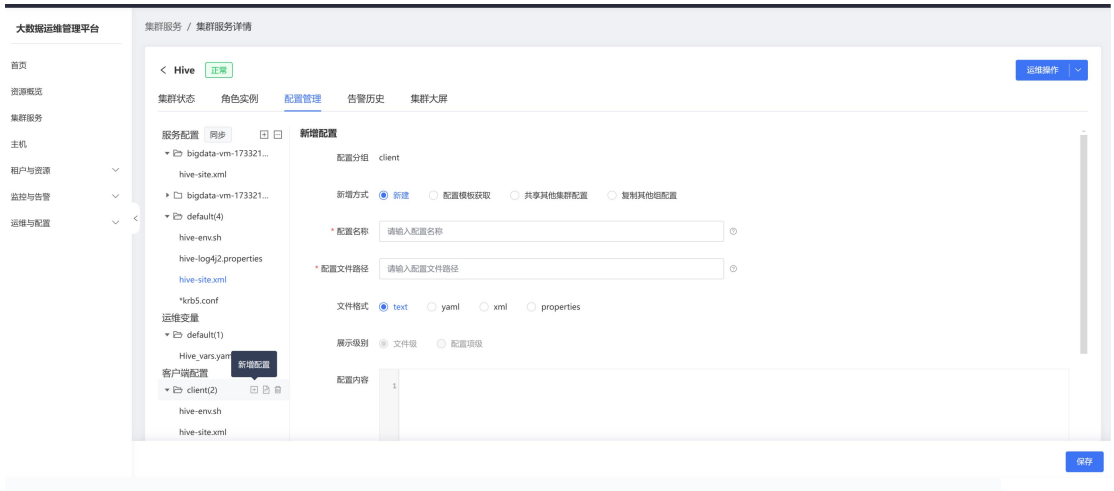
方式 1：新建

操作场景

无需共享其他集群的配置，且配置模板不符合使用要求时，用户为集群服务新增一条全新的配置文件。

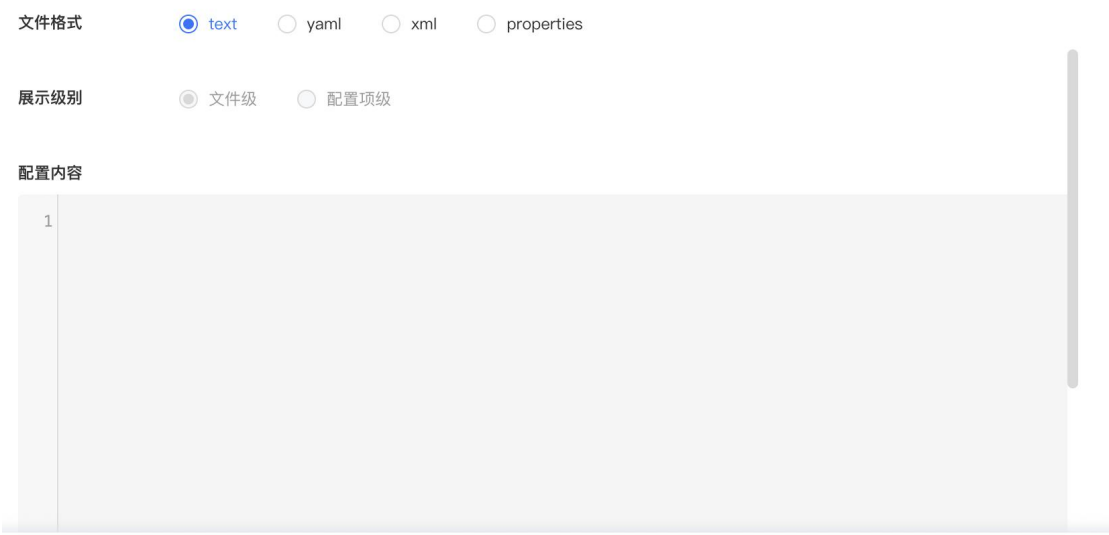
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 选择要新增配置的的配置组，鼠标移入到配置组名称上，单击右侧的“新增配置”图标，页面右侧为新增配置。如图所示：

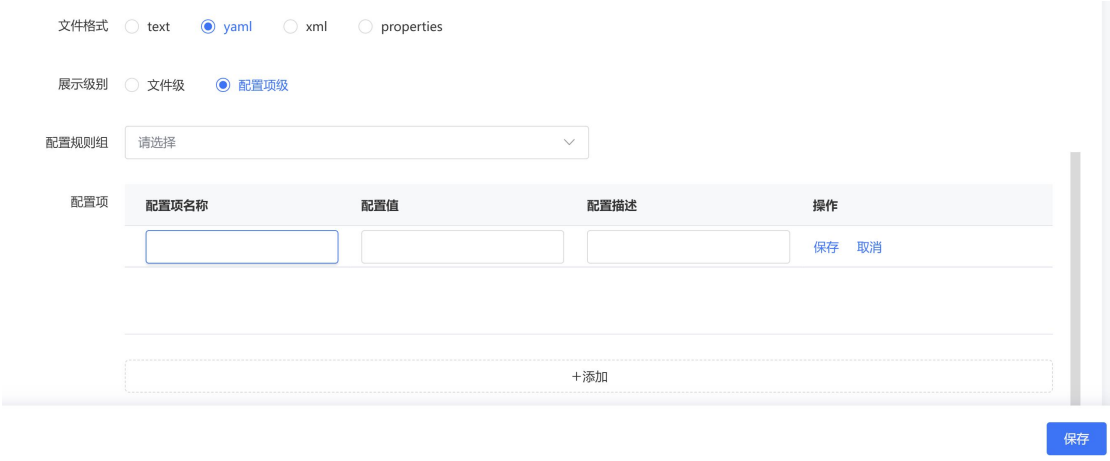


8. 选择新增方式为“新建”，输入配置名称、配置文件路径、选择文件格式，输入配置内容或者添加配置项。

   文件格式为 text 时，仅需要输入配置内容即可。如图所示：



文件格式非 text，为 yaml、xml 或者 properties 时，需要：选择展示级别、选择配置规则组、配置项。如图所示：



9. 单击“保存”完成配置新增。

说明

1. 配置名称：仅支持输入英文、数组、“_-.”，且不能以数字开头。
2. 文件路径：仅支持输入英文、数组、“_-.”；且配置路径必须在白名单中。
3. 新增配置文件时，展示级别选择配置项级，查看配置文件时，将按配置项进行展示。展示级别选择文件级，查看配置文件时，将看到完整的文本文件。
4. 配置规则组：配置规则组包含系统中预设的一组配置规则，可基于现有配置规则组，快速创建配置文件。在新建配置文件时，选择配置规则组后，展示该配置规则组中，默认包含的配置项，可编辑默认配置项的配置值和描述，或继续添加新的配置项。

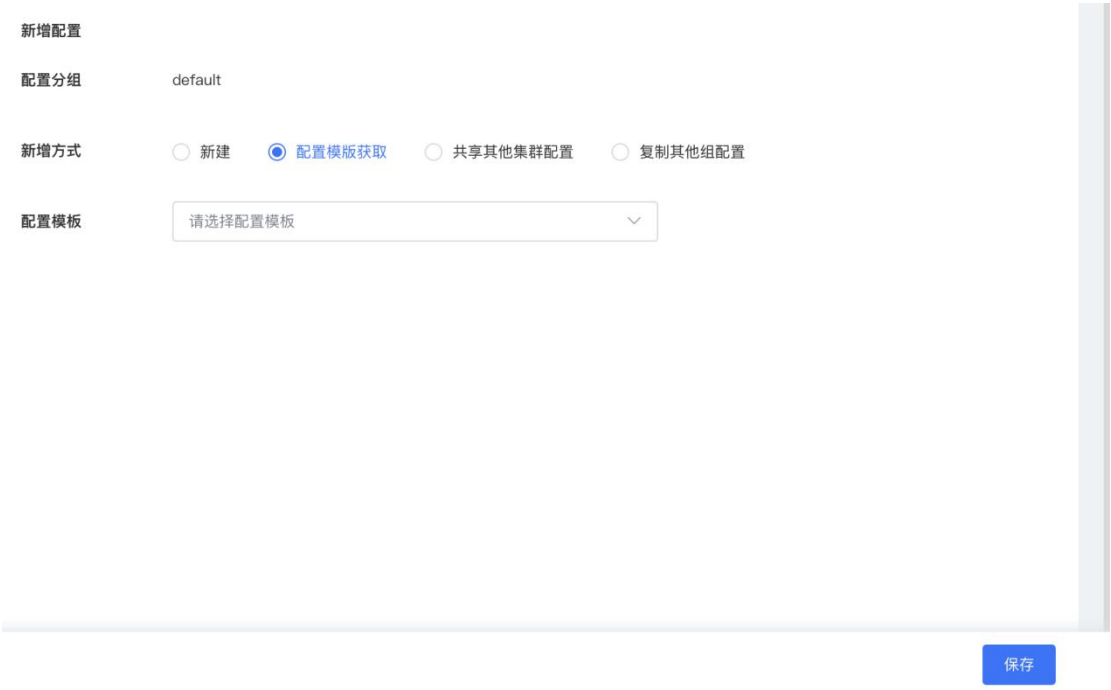
方式 2：配置模板获取

操作场景

直接使用配置模板为集群服务新增配置。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 选择要新增配置的的配置组，鼠标移入到配置组名称上，单击右侧的“新增配置”图标，页面右侧为新增配置。
8. 选择新增方式为“配置模板获取”。如图所示：



9. 选择配置名称。
10. 单击“保存”完成配置新增。

方式 3：共享其他集群配置

操作场景

当前集群服务与其他集群的配置存在依赖管理，可直接共享其他集群服务的配置。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 选择要新增配置的的配置组，鼠标移入到配置组名称上，单击右侧的“新增配置”图标，页面右侧为新增配置。
8. 选择新增方式为“共享其他集群配置”。如图所示：

9. 选择要共享的集群和配置。
10. 单击“保存”完成配置新增。

说明

1. 与其他集群共享的配置，在配置文件名称前会有“*”进行标识。
2. 与其他集群共享的配置，在任意一个集群服务下进行配置修改后，共享该配置的集群服务的配置也会同步进行修改。

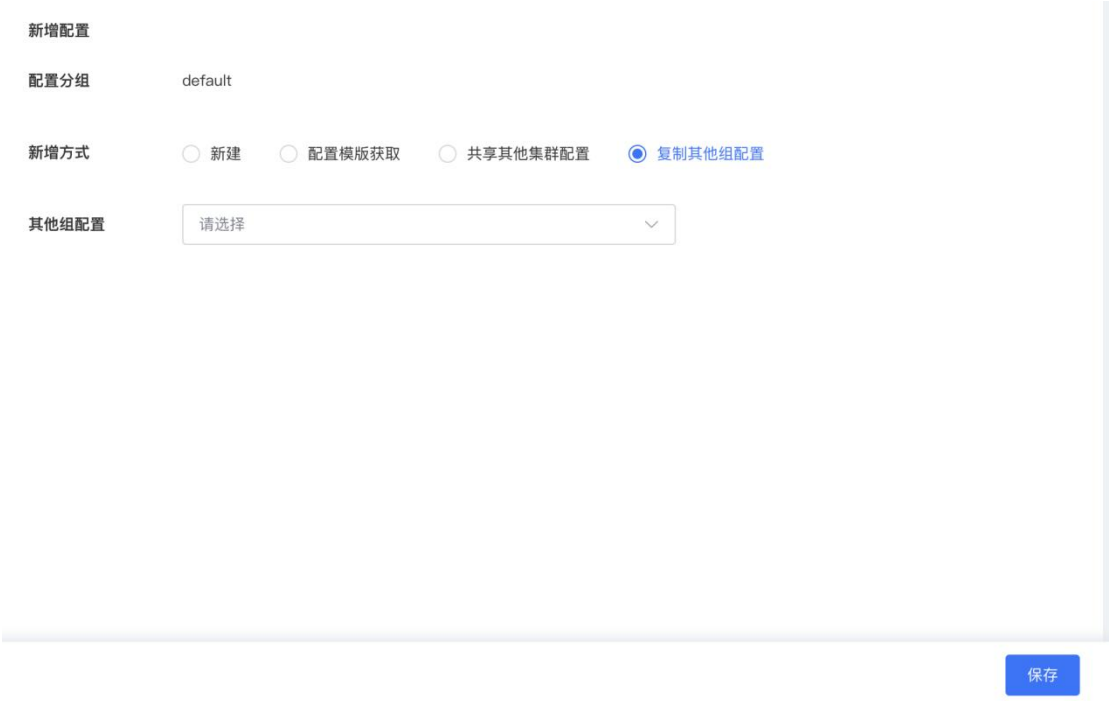
方式 4：复制其他组配置

操作场景

当前配置在其他组已存在，无需再新建，可以直接从其他组复制到当前组。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 选择要新增配置的的配置组，鼠标移入到配置组名称上，单击右侧的“新增配置”图标，页面右侧为新增配置。
8. 选择新增方式为“复制其他组配置”。如图所示：



9. 选择要复制的配置组和配置。
10. 单击“保存”完成配置新增。

> ⓘ说明
>
> 对复制的配置文件进行修改，或者是修改被复制的配置，不会互相影响。

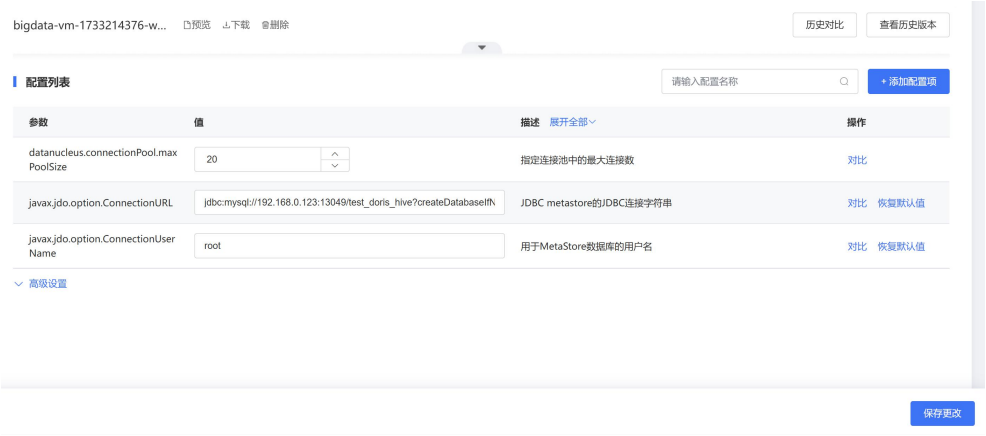
4.6.4 编辑配置

操作场景

用户可以对配置文件进行修改。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击配置组，单击要修改的配置名称，页面右侧展示该配置的详细信息。如图所示：



8. 直接对配置文件修改即可，修改完成后点击“保存更改”。

说明

1. 服务端配置,修改后需要执行同步操作。在配置文件及同步按钮上有数字提示未同步修改。
2. client 分组下配置、default 分组下以 vars.yaml 结尾的配置，不需要同步，修改后不会有数字提示，也不需要执行同步操作。
3. 对于同步按钮，提示数字的含义为待同步配置文件的数量。
4. 对于配置文件前同步提示数字，文件级和配置项级有所区别，具体含义说明如下。
5. 展示级别为文件级：修改后且当前配置文件未同步时，配置名称前展示数字 1。
6. 展示级别为配置项级：修改后且当前配置文件未同步时，配置名称前展示所修改的配置项数，所修改的配置项名称前展示修改标识。

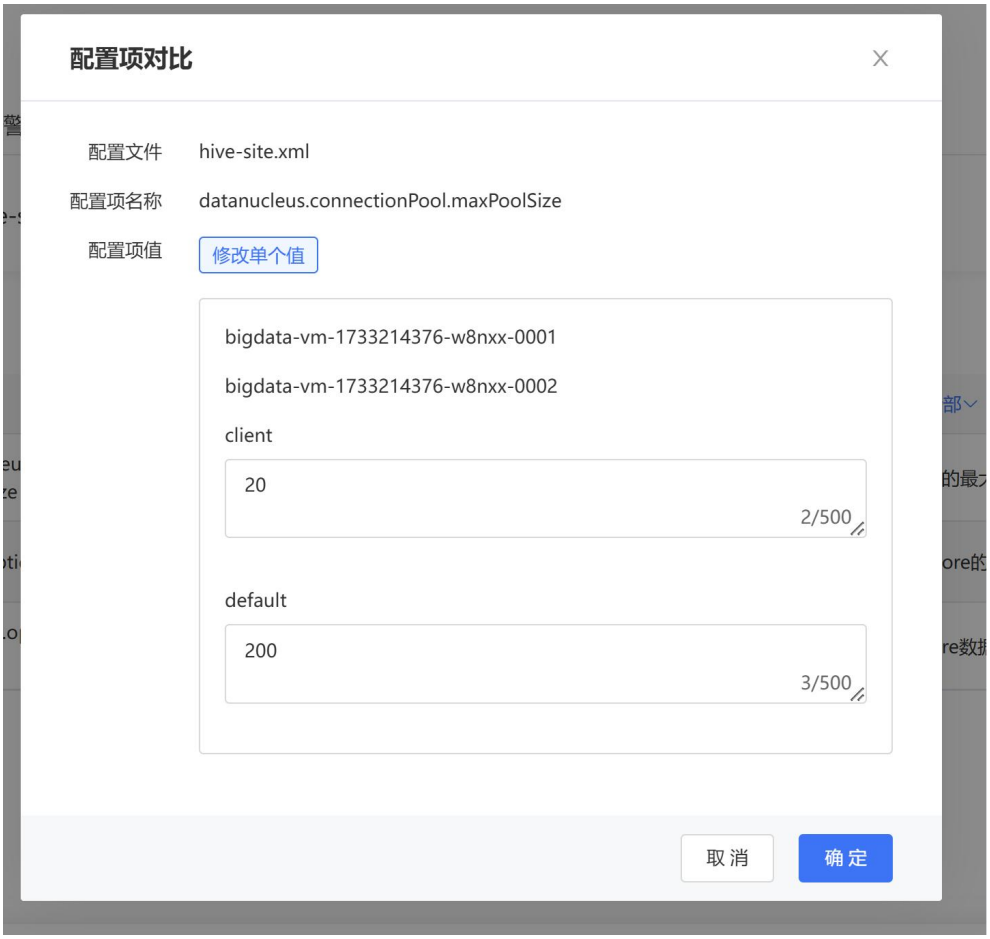
4.6.4 配置项对比

操作场景

用户可以查看同一集群下相同配置项名称在不同配置组中的值。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
8. 选择要对比的配置项，单击该配置项名称右侧的“配置项对比按钮”，出现配置项对比弹框。如图所示：



9. 配置项对比弹框中，当多个配置分组下，存在同一配置项时，支持逐条展示或合并相同值展示与编辑。如图所示，default、client 分组下同一配置项具有相同的值，支持当设置为“修改单个值”或“修改相同值”，进行合并或逐条展示与编辑。

天翼云

配置项对比

配置文件 core-site.xml

配置项名称 io.compression.codec.lzo.class

配置项值 修改单个值

default

client

com.hadoop.compression.lzo.LzoCodec

35/500

取消 确定

配置项对比

配置文件 core-site.xml

配置项名称 io.compression.codec.lzo.class

配置项值 修改相同值

default

com.hadoop.compression.lzo.LzoCodec

35/500

client

com.hadoop.compression.lzo.LzoCodec

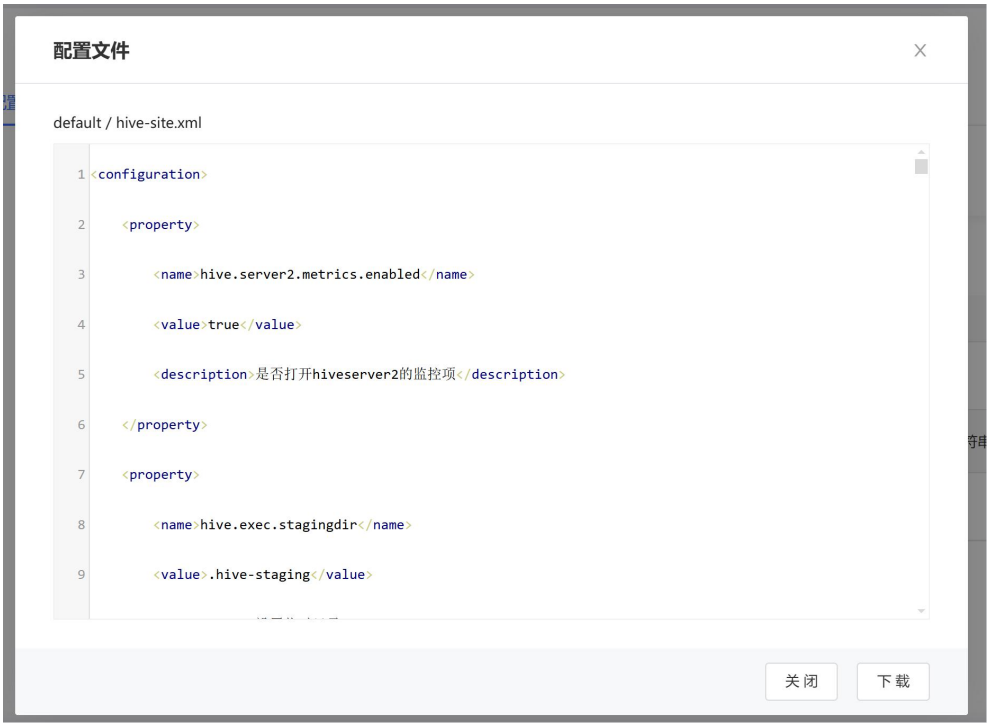
35/500

取消 确定

4.6.4 配置文件预览

操作步骤

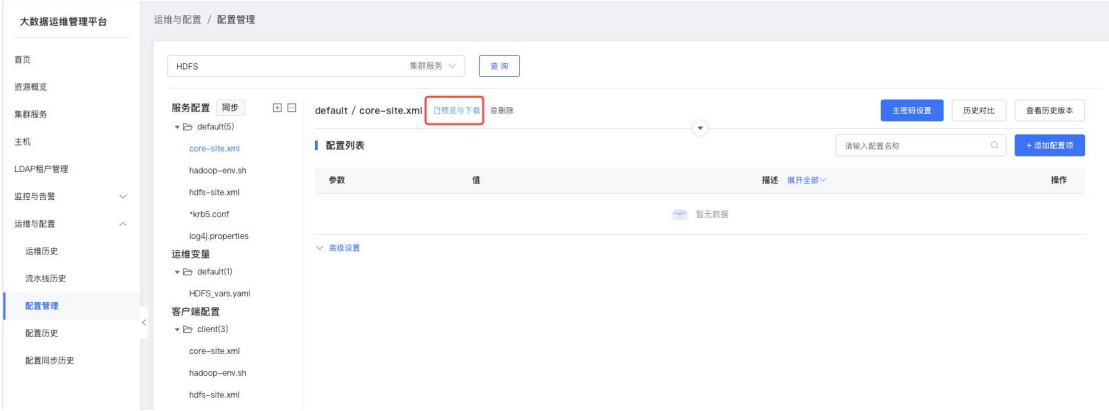
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
8. 单击“预览文件”，出现配置文件预览弹框。如图所示：



4.6.4 下载配置文件

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“预览与下载”，在弹窗中进行下载。如图所示：



4.6.4 配置文件历史对比

操作场景

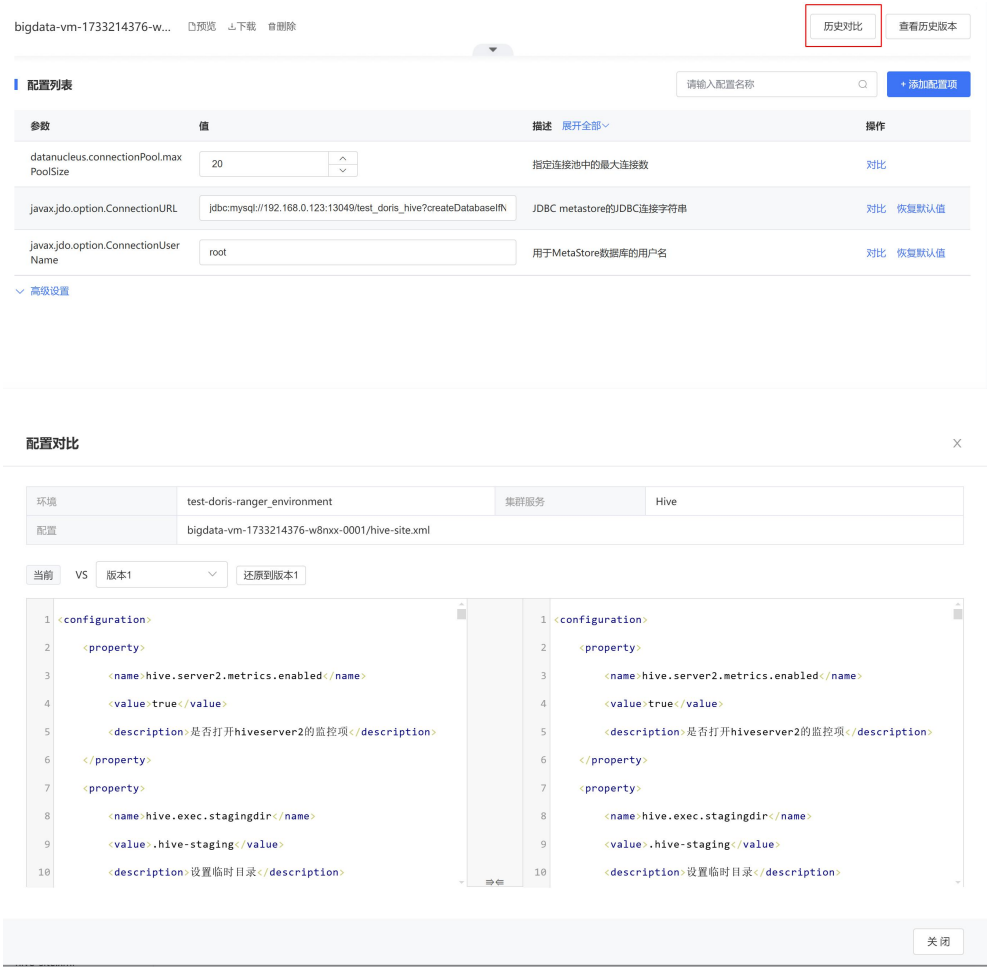
支持用户查看当前配置文件与不同历史版本的差异。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。

7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。

8. 单击“历史对比”，出现历史对比弹框。如图所示：



4.6.4 查看配置历史版本

操作场景

支持用户查看指定配置文件的历史版本，不同版本的配置内容以及差异。

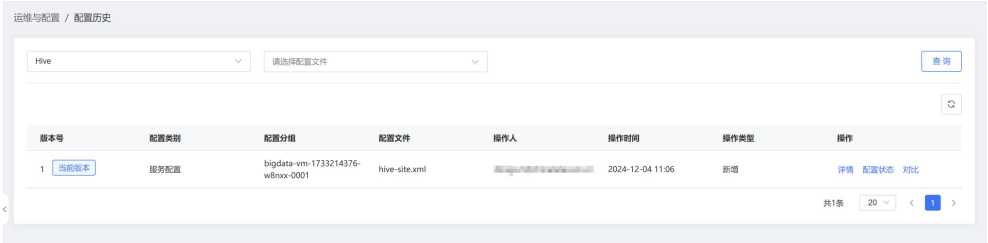
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。

7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。

8. 单击“查看历史版本”。如图所示：

9. 进入到配置历史页面，配置历史页面展示当前配置的所有历史版本。如图所示：



4.6.4 删除配置

操作步骤

1. 登录翼 MR 管理控制台。

2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

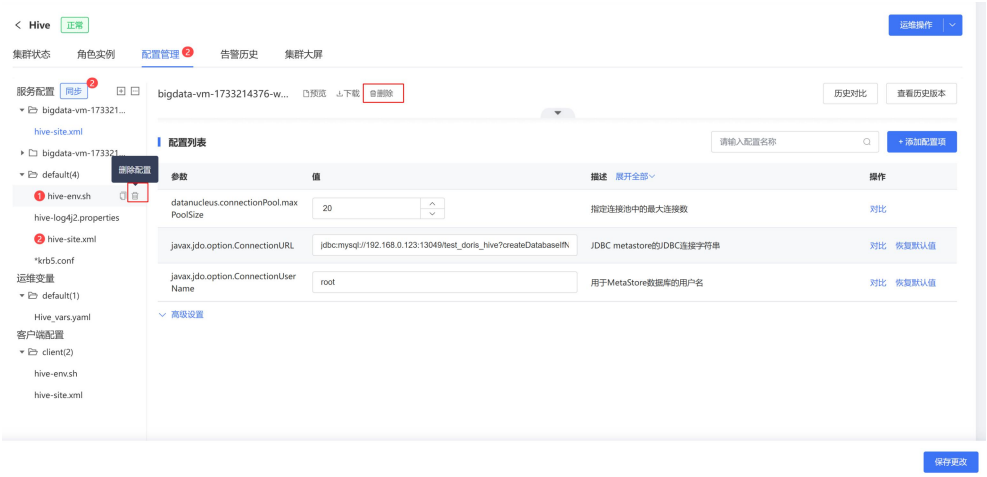
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。

5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。

6. 单击“配置管理”tab。

7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。

8. 点击配置列表或右侧配置文件的“删除”按钮，出现删除确认弹框，单击“确定”即可。如图所示：



注意

当前翼 MR Manager 中不支持删除非自定义配置，只有自定义的配置才支持删除。

4.6.4 配置同步

操作场景

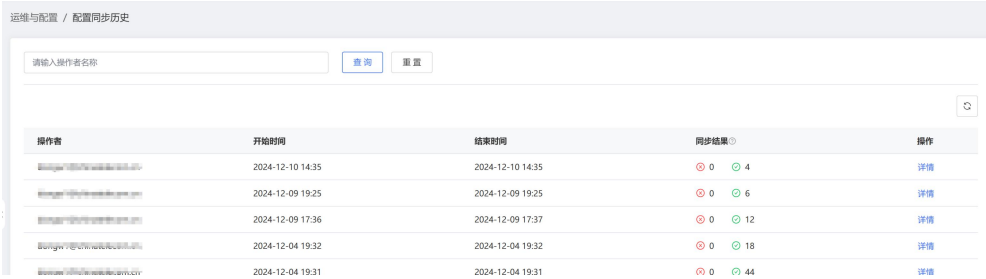
支持将新的配置文件或修改后的配置同步到相应主机上。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 如果配置信息存在变更需要同步生效，“同步”按钮上有数字提示待同步配置文件数。此时单击“同步”按钮，进行配置同步操作。如图所示：

8. 进入待同步配置页面，可查看待同步的配置文件、目标主机等信息。如图所示：

9. 单击“配置同步”按钮，跳转到配置同步菜单页面，点击刷新图标直到同步完成。如图所示：



10. 单击“详情”操作，进入到配置同步详情页面，查看配置同步进度。如图所示：

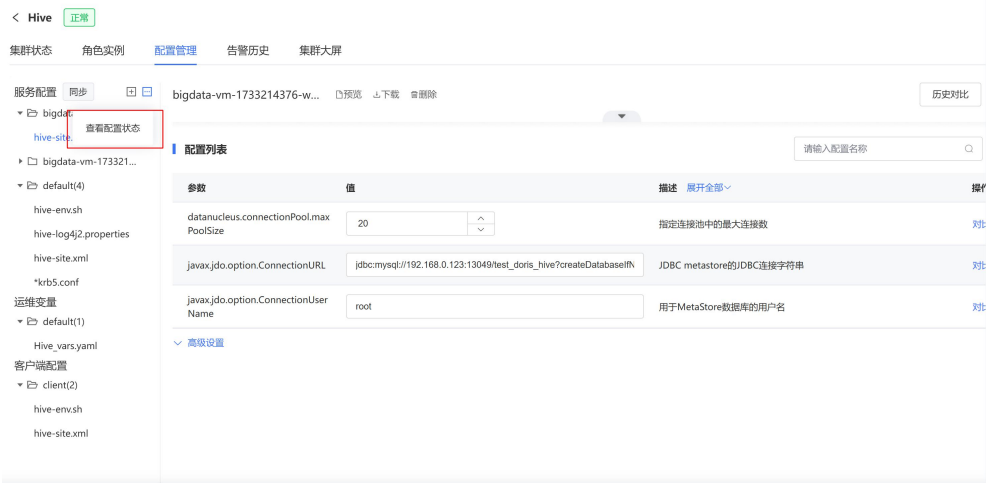
4.6.4 查看配置状态

操作场景

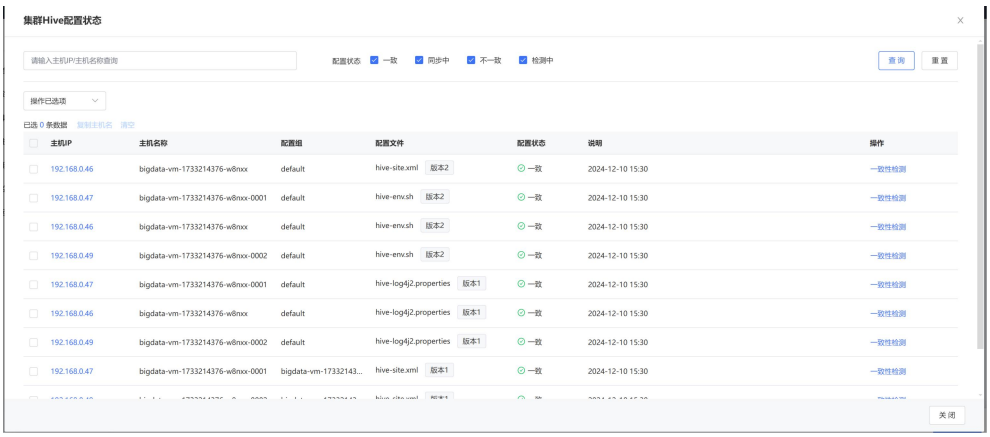
支持查看指定集群服务所有配置文件的同步状态。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击“查看配置状态”按钮。如图所示：



8. 出现配置同步状态弹框，展示所有配置文件的同步状态。如图所示：



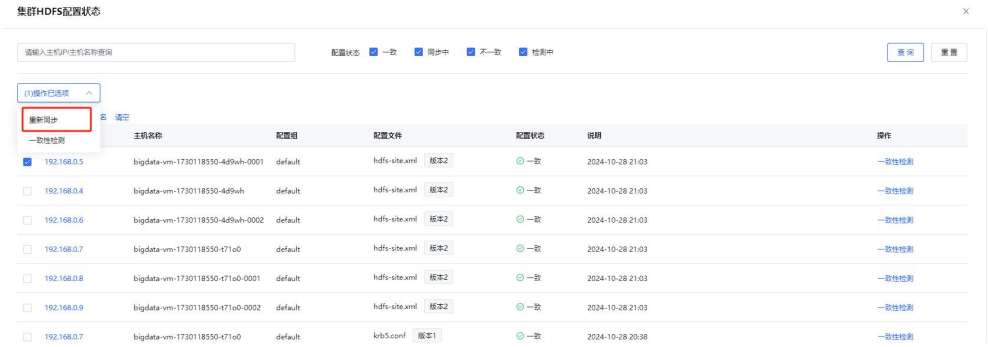
4.6.4 重新同步

操作场景

支持对同步失败的配置或配置状态不一致的进行重新同步。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击“查看配置状态”按钮，出现配置同步状态弹框。
8. 勾选需要操作主机前的复选框。
9. 单击“操作已选项 > 重新同步”，即可支持配置同步操作。如图所示：



4.6.4 配置回滚

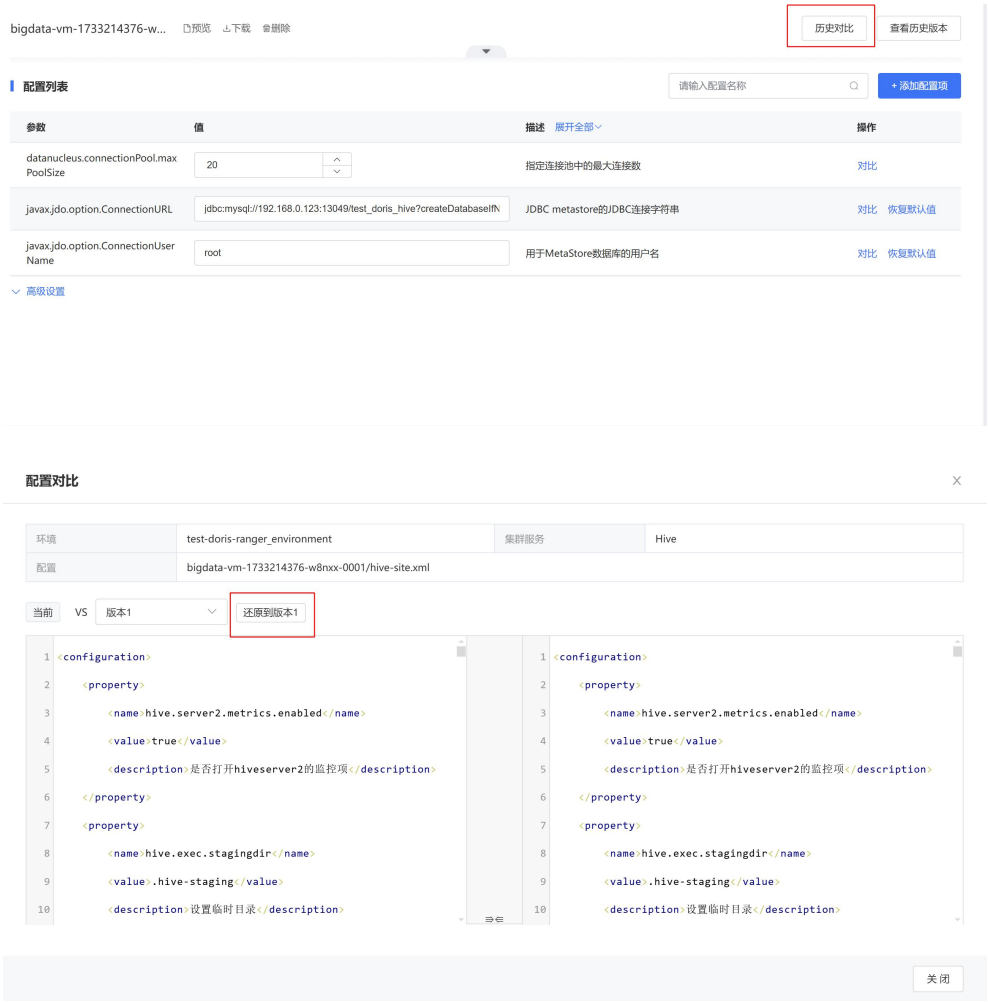
操作场景

支持回退指定配置文件为某个历史版本。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“配置管理”tab。
7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。

8. 单击“历史对比”，出现历史对比弹框。如图所示：



9. 单击“还原到版本 X”按钮，出现确认操作提示弹框，如图所示。单击“确认”按钮，还原配置文件到指定版本。

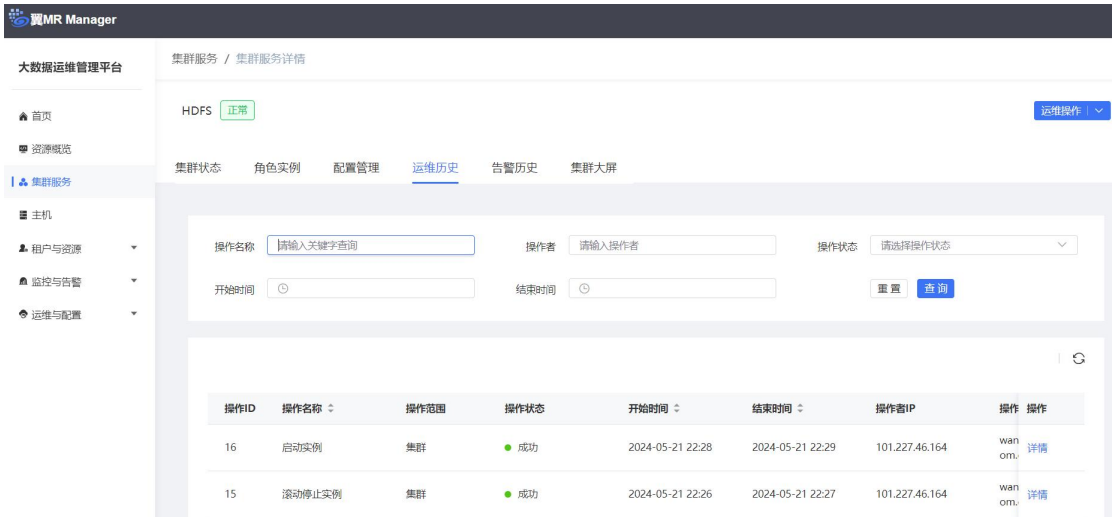
10. 查看配置文件，配置文件内容与所还原的版本内容一致。单击“同步”，对配置文件进行同步操作即可。

4.6.4 查看运维操作历史

仅翼 MR2.14.1 版本，在集群服务详情中，查看当前服务运维历史。翼 MR2.15.2 版本请到运维与配置——运维历史二级菜单中查看。

操作步骤

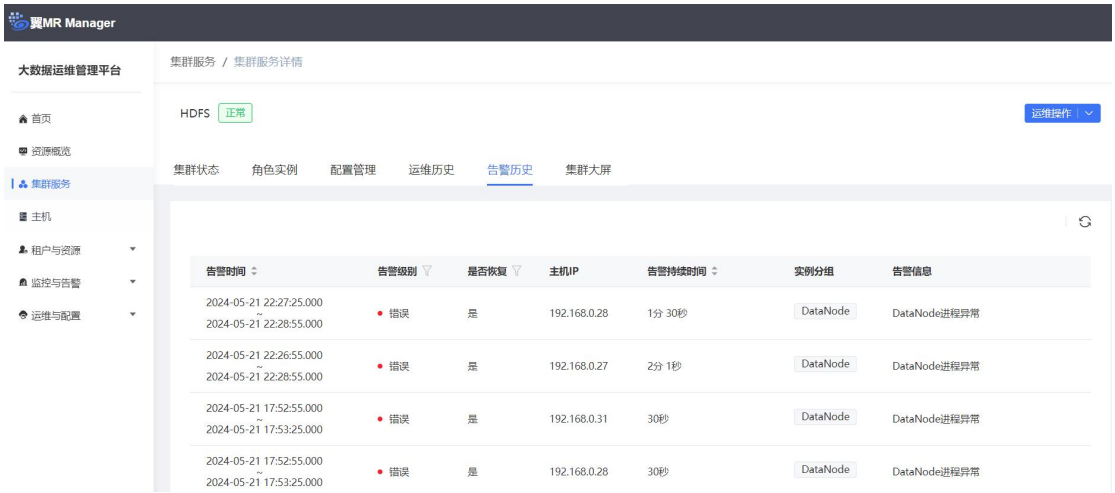
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“运维历史”tab，即可查看该集群服务的所有运维操作历史。如图所示：



4.6.4 查看告警历史

操作步骤

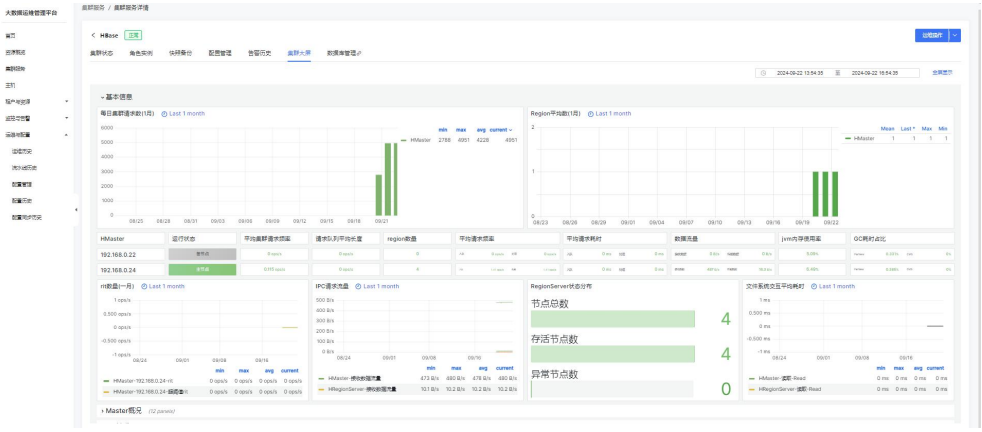
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“告警历史”tab，即可查看该集群服务的所有告警信息。如图所示：



4.6.4 查看集群服务大屏

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
6. 单击“集群大屏”tab，即可查看该集群服务的监控大屏。
7. 默认展示最新 3 小时的监控数据，可通过右上角时间选框，选择查看时间段。如图所示：



4.6.4 集群服务健康检查

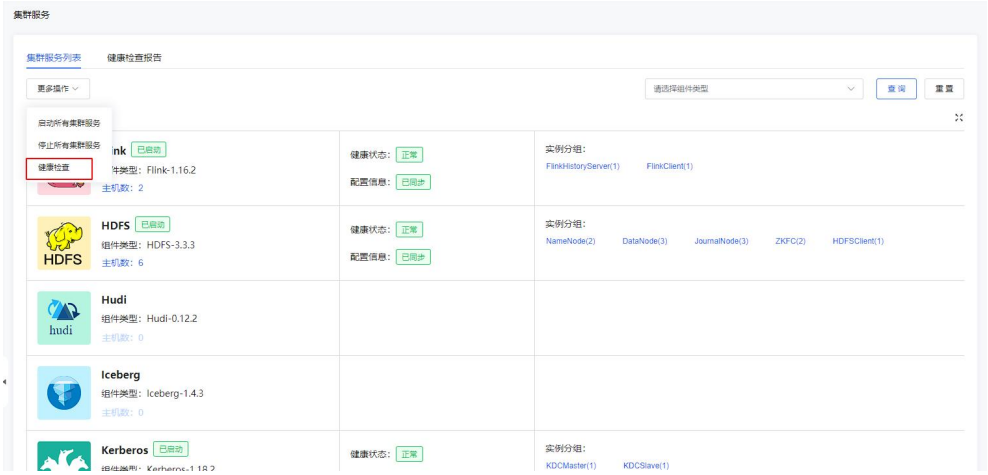
健康检查

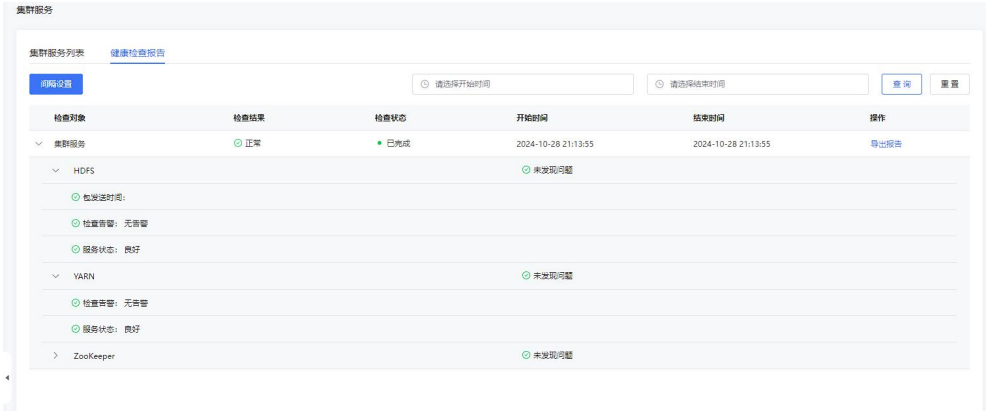
操作场景

对集群服务进行健康检查。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 单击“更多操作>健康检查”，进行集群服务健康检查，并自动跳转到健康检查报告页面，如图所示：





> ⓘ注意
>
> 支持以下组件类型的集群服务进行健康检查：HBase、HDFS、Hive、Kafka、Yarn、Zookeeper、Spark。

导出健康检查报告

操作场景

导出指定健康检查报告到本地。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
5. 单击“健康检查报告”，选择要导出的报告，点击“导出报告”即可，如图所示：

定期健康检查

操作场景

设置定期健康检查。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。

5. 单击“健康检查报告”，单击右侧设置图标，出现健康检查间隔弹框，如图所示：

6. 开启定期健康检查，选择定期健康检查时间，单击“确定”按钮后，会按照所设置的时间对集群服务进行定期健康检查。

ⓘ注意

支持按照每天、每周、每月进行定期健康检查设置。

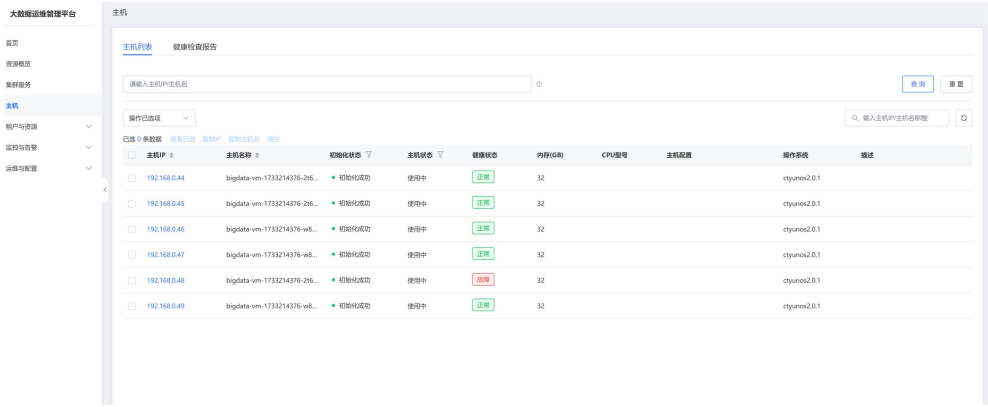
4.6.5 主机

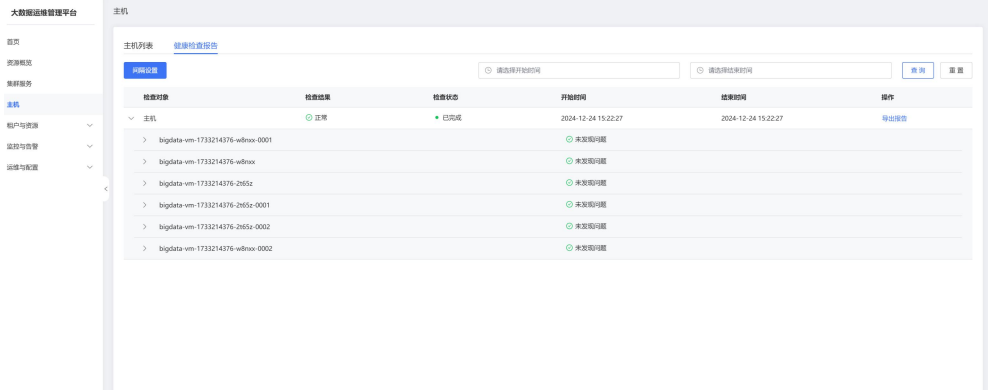
4.6.5 主机管理页面概述

本章节主要介绍主机管理页面功能概述。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab。
4. 单击“前往翼 MR Manager”。
5. 进入翼 MR Manager 后，点击菜单“主机”，进入主机页面，如图所示：





ⓘ说明

1. “主机列表”标签页，概述如下：

   (1) 主机列表上方为查询区域，支持主机 IP/主机名进行查询。

   (2) 主机列表以数据列表视图列出。

   (3) 主机列表在左上角为操作区，可支持操作：置维护、取消维护、导出、健康检查。

   (4) 单击主机 IP 或主机名称可进入到主机详情页面。

2. “健康检查报告”标签页，概述如下：

   (1) 上方查询区域，筛选时间范围内主机健康检查报告记录。

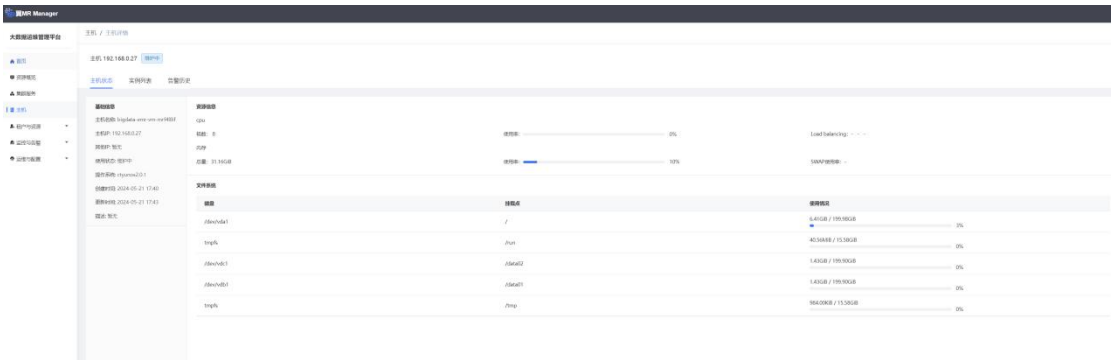
   (2) 左侧“间隔设置”按钮，配置定期主机健康检查。

   (3) 健康检查报告列表以数据列表视图列出。

4.6.5 主机详情页面概述

进入到翼 MR Manager 以后，点击菜单“主机”，进入主机列表页面。

单击指定主机 IP 或主机名称进入到主机详情页面。如图所示：



详情页面下方依次为该主机的：主机状态、实例列表、告警历史信息。

4.6.5 主机置维护、取消维护

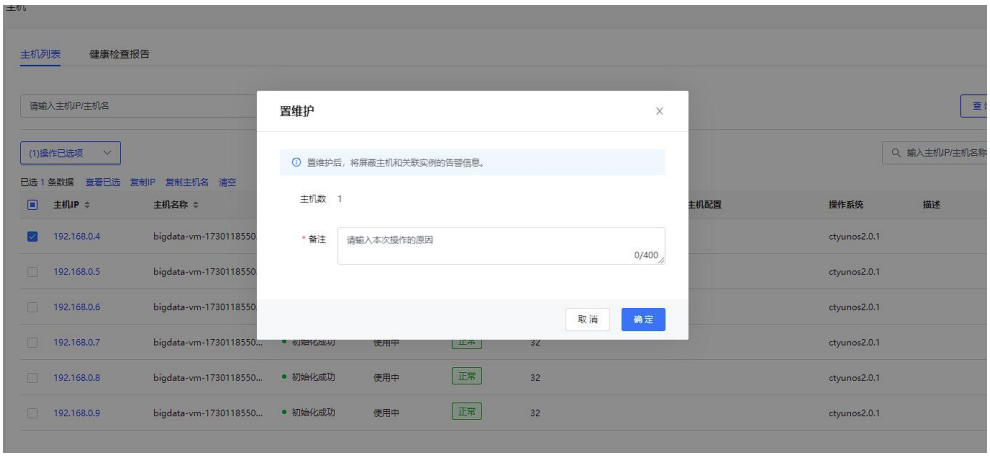
置维护

操作场景

用户需要对某个或一批主机进行置维护。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 勾选需要操作主机前的复选框。
6. 单击“操作已选项 > 置维护”，弹出置维护操作弹框。如图所示：



7. 输入备注，单击“确定”即可。

说明

1. 主机置维护后，该主机的健康状态显示为：维护中。
2. 主机置维护后，将屏蔽主机和关联实例的告警信息。

取消维护

操作场景

用户需要对某个或一批主机进行取消维护。

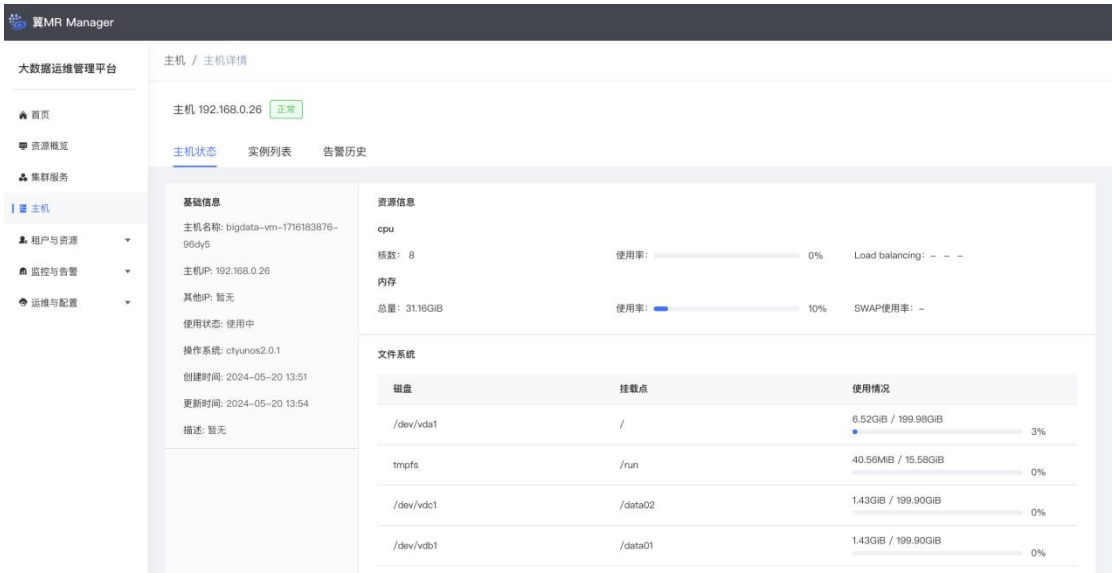
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 勾选需要操作主机前的复选框。
6. 单击“操作已选项 > 取消维护”，弹出取消维护操作弹框。
7. 输入备注，单击“确定”即可。

4.6.5 查看主机状态

操作步骤

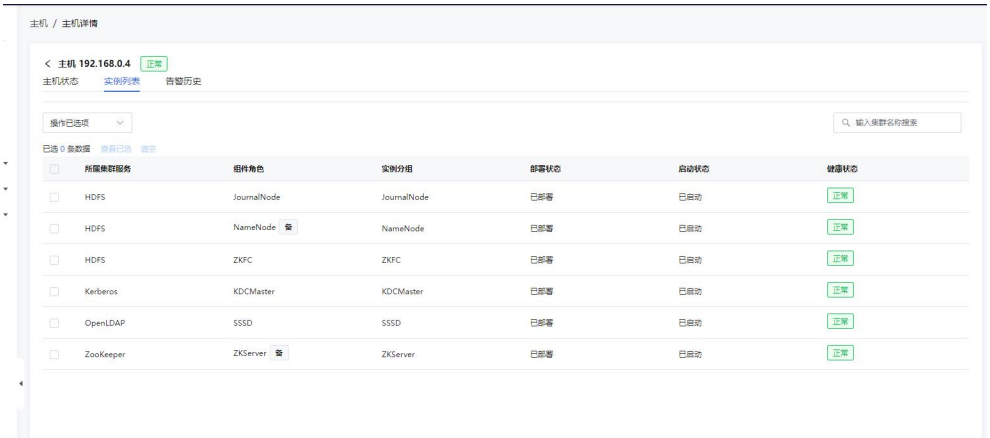
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 单击指定主机的主机 IP 或主机名称进入到主机详情页面，默认进入到主机状态页面。如图所示：



4.6.5 查看主机实例列表

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 单击指定主机的主机 IP 或主机名称进入到主机详情页面。
6. 单击“实例列表”tab，即可查看该主机上安装的实例服务。如图所示：



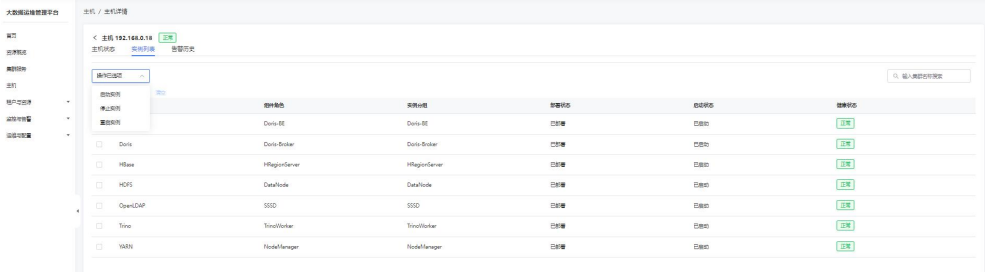
4.6.5 查看主机实例列表—实例操作

操作场景

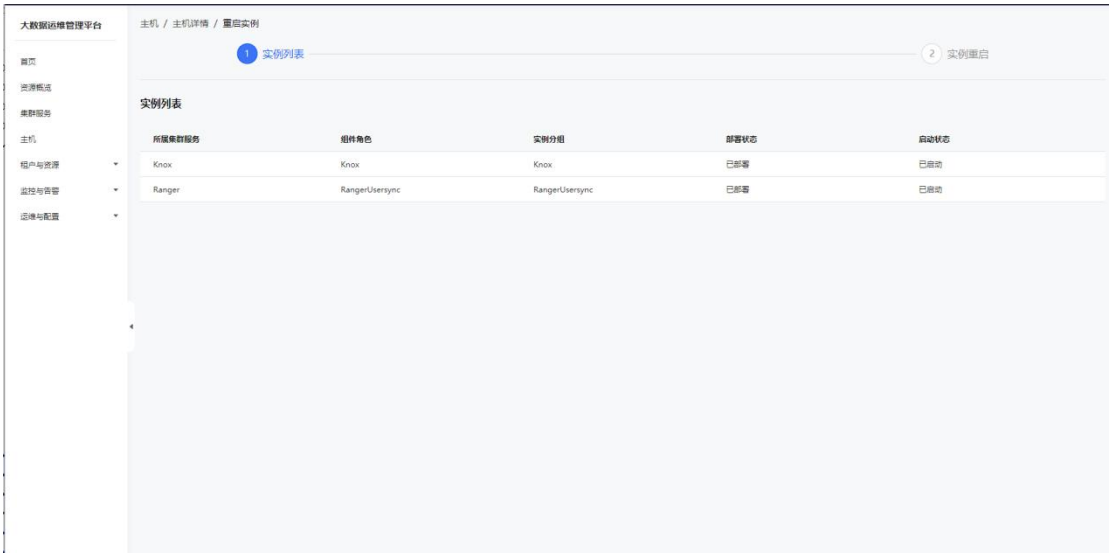
对主机上实例，进行启动、停止、重启等批量操作。

操作步骤

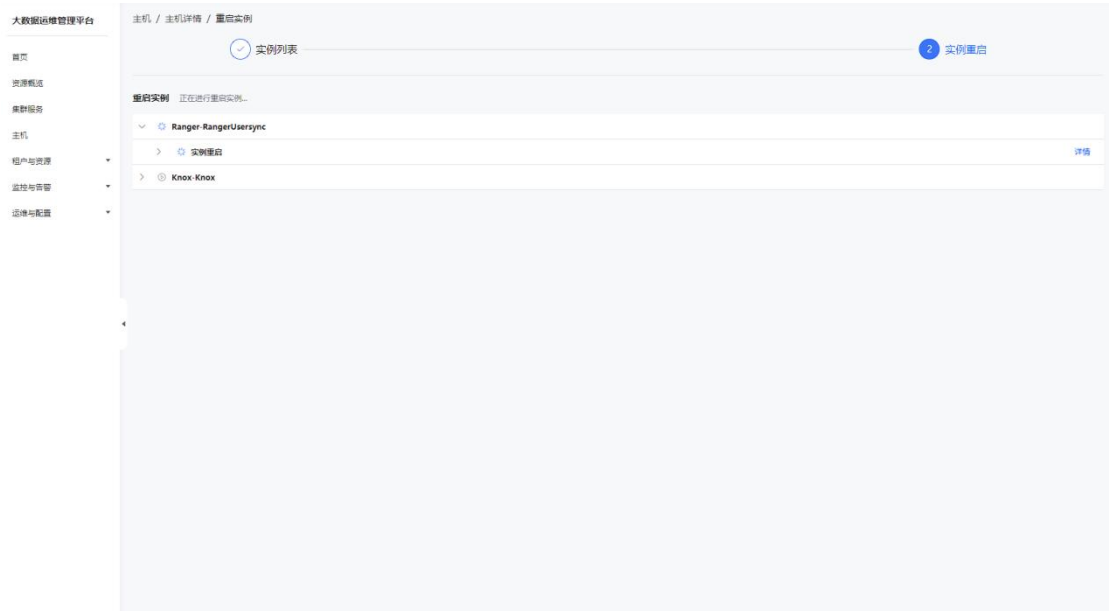
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 单击指定主机的主机 IP 或主机名称进入到主机详情页面。
6. 单击“实例列表”tab，即可查看该主机上安装的实例服务，勾选想要进行批量运维操作的实例。
7. 点击“操作已选项”按钮，选择相应的运维操作，在提示弹窗中点击确定按钮。单次只支持批量进行同一种运维操作。如图所示：



8. 进入引导式运维操作界面，展示待操作的实例列表，点击下一步运行流水线。如图所示：



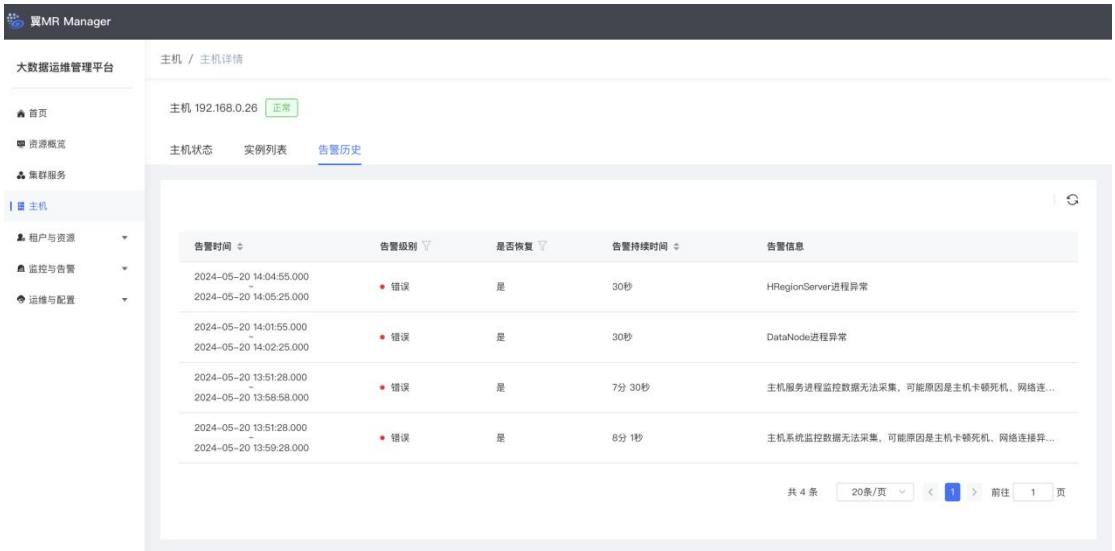
9. 进入流水线运行界面，可查看流水线的运行状态，如图所示。运维操作结束后会自动跳出界面，可在运维历史界面查看运维记录信息。



4.6.5 查看主机告警历史

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 单击指定主机的主机 IP 或主机名称进入到主机详情页面。
6. 单击“告警历史”tab，即可查看该主机的所有告警信息。如图所示：



4.6.5 主机健康检查

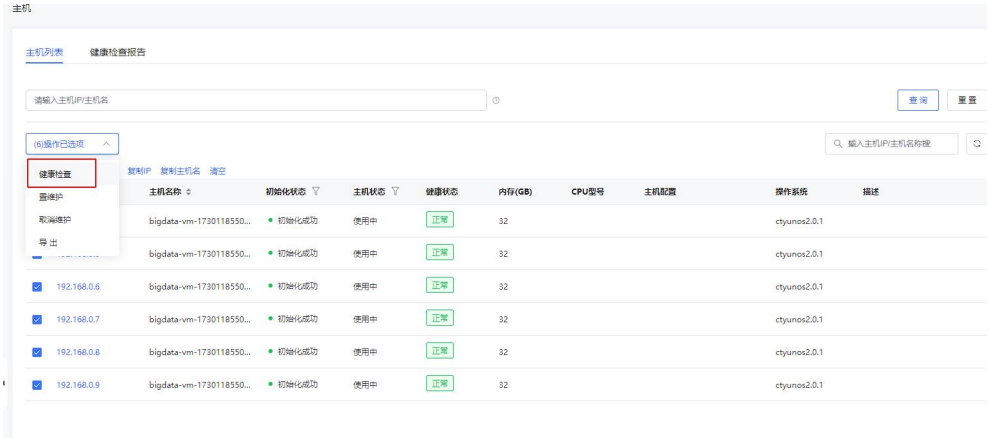
健康检查

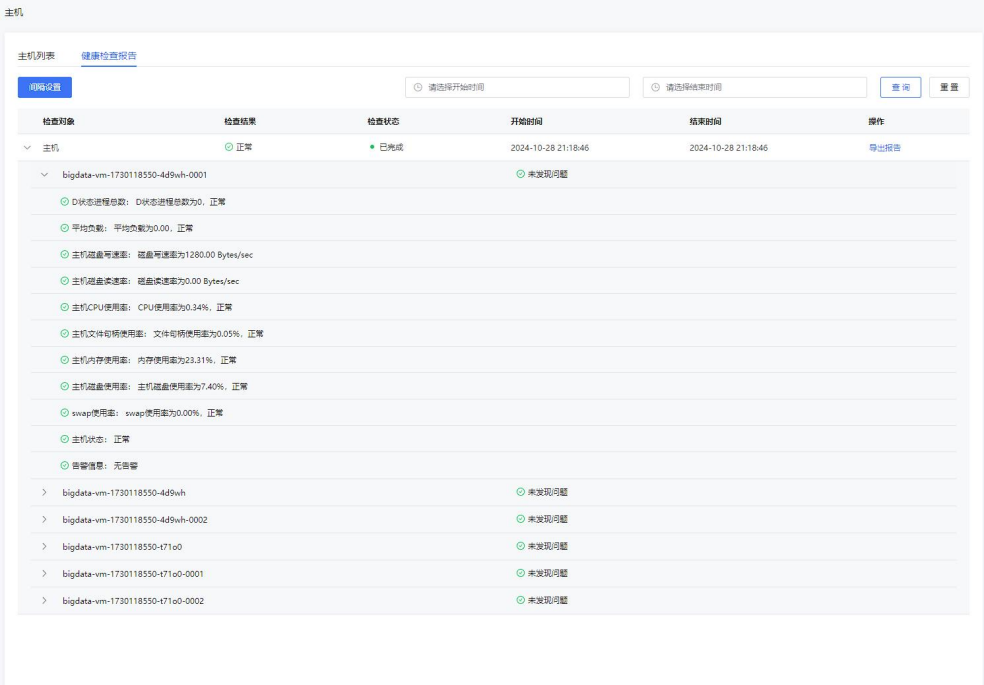
操作场景

对指定主机进行健康检查。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 勾选需要操作主机前的复选框。
6. 单击“操作已选项>健康检查”，进行主机健康检查，并自动跳转到健康检查报告页面，如图所示：





7. 等待健康检查完成即可。

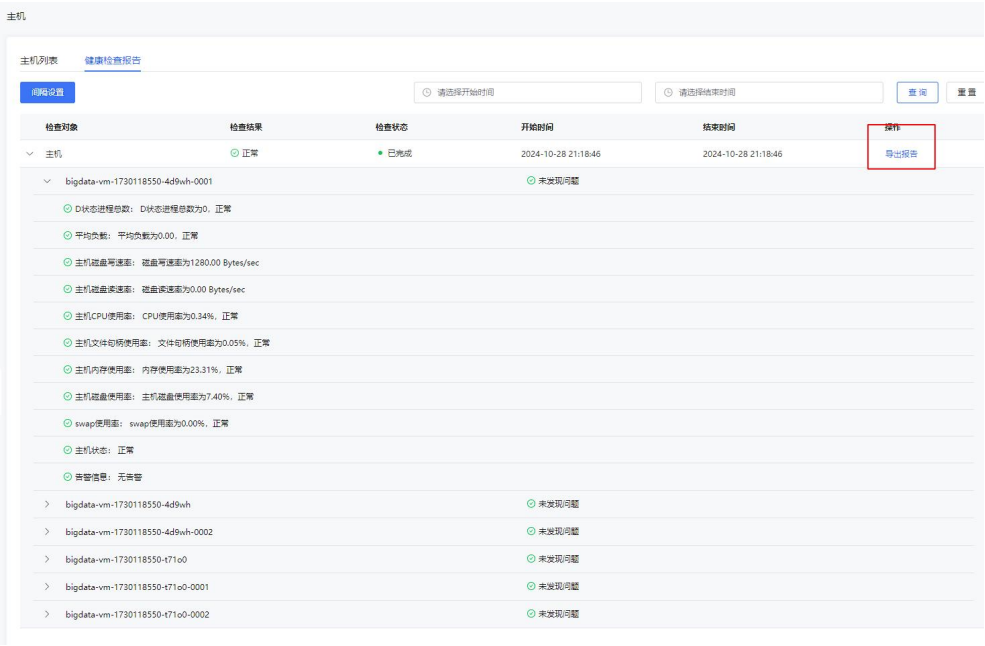
导出健康检查报告

操作场景

导出指定健康检查报告到本地。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 单击“健康检查报告”，选择要导出的报告，点击“导出报告”即可，如图所示：



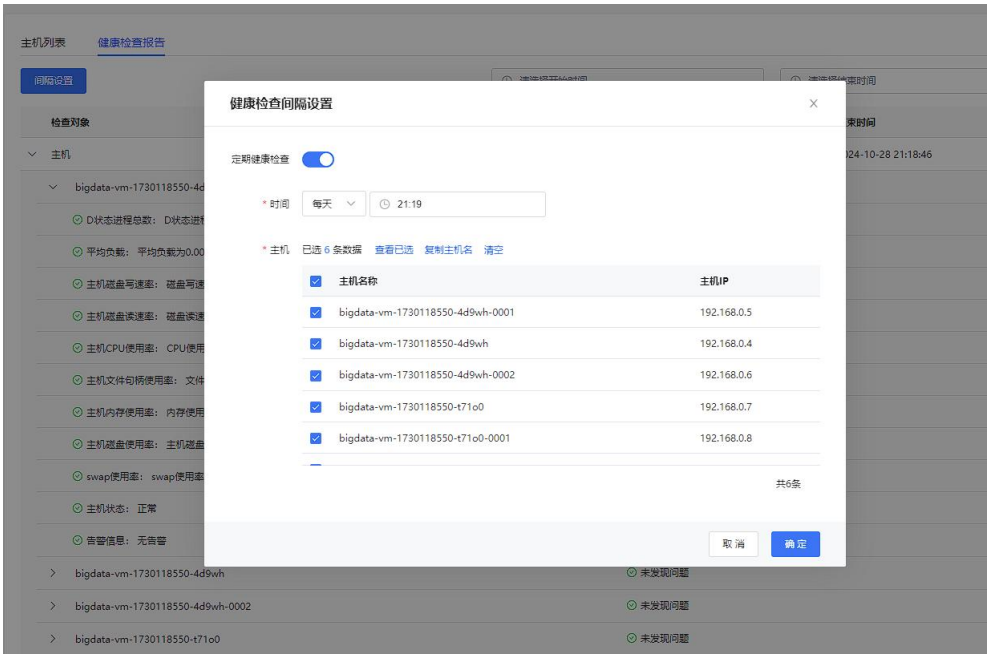
定期健康检查

操作场景

设置定期健康检查。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“主机”，进入主机列表页面。
5. 单击“健康检查报告”，单击右侧设置图标，出现健康检查间隔弹框，如图所示：



6. 开启定期健康检查，选择定期健康检查时间，选择要定期健康检查的主机，单击“确定”按钮后，会按照所设置的时间对指定的主机进行定期健康检查。

> ⓘ说明
>
> 支持按照每天、每周、每月对指定主机进行定期健康检查设置。

4.6.6 租户与资源

4.6.6 租户与资源概述

翼 MR Manager 提供的租户与资源主要包含：LDAP 租户管理、Kerberos 安全凭证、YARN 队列管理、HBase 数据库管理。

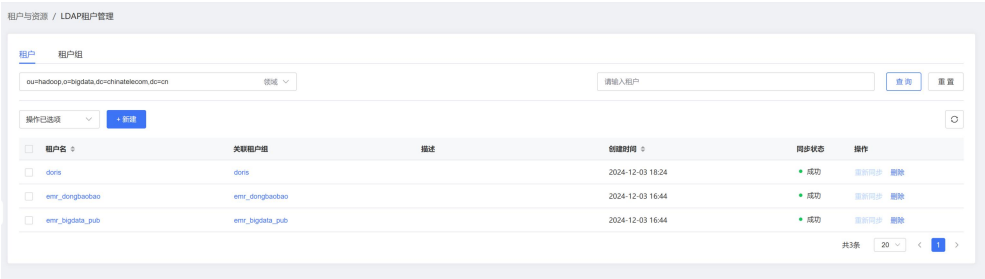
前置操作

使用租户与管理功能，需要进入 Manager 页面，操作步骤如下：

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

LDAP 租户管理

进入到翼 MR Manager 以后，点击菜单“租户与资源 > LDAP 租户管理”，进入 LDAP 租户管理页面。如图所示：



- 支持查看 LDAP 中的租户和租户组。
- 支持新增租户、租户组。
- 支持为租户关联租户组、租户组关联租户以及取消关联关系。
- 支持以租户为维度的 Principal 管理以及 Keytab 的分发。

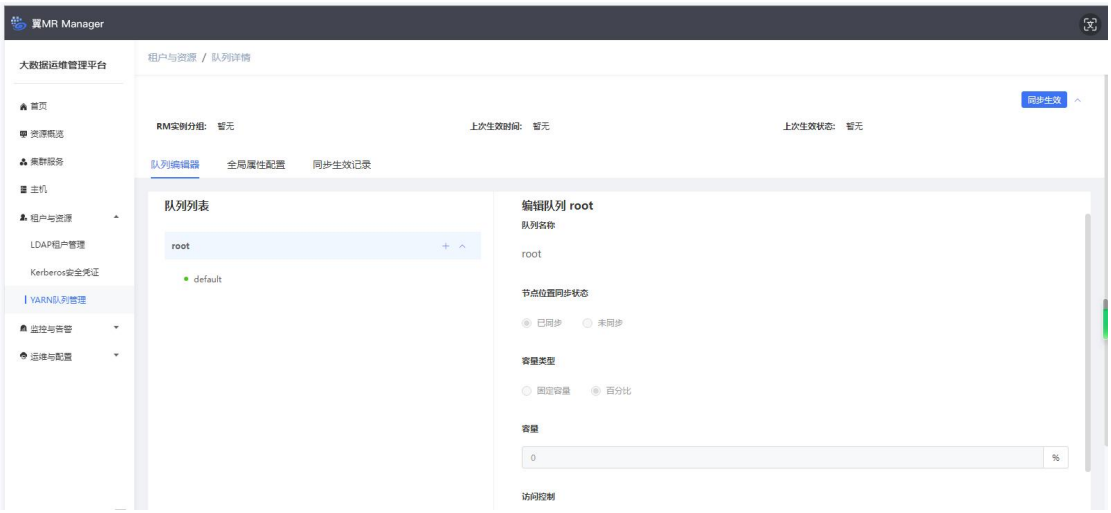
Kerberos 安全凭证

进入到翼 MR Manager 以后，点击菜单“租户与资源 ＞Kerberos 安全凭证”，进入 Kerberos 安全凭证页面。如图所示：

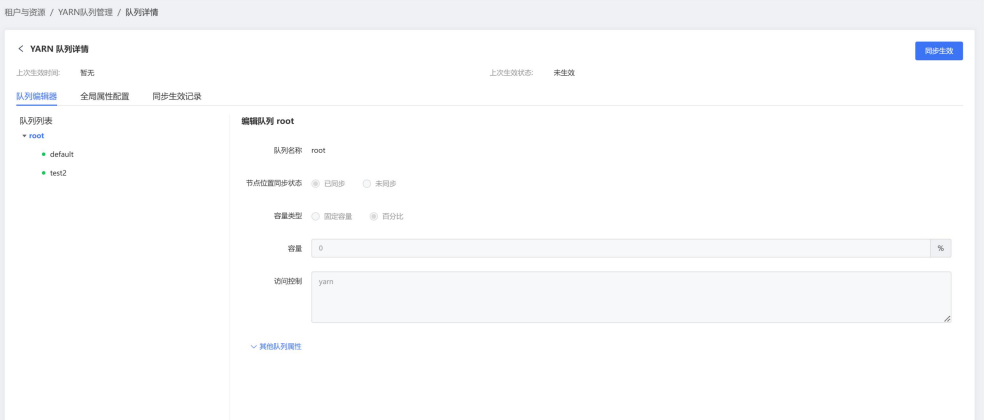
- 支持新建 Principal、删除 Principal。
- 支持 Keytab 分发与下载，并支持查看 Keytab 的分发记录。

YARN 队列管理

进入到翼 MR Manager 以后，点击菜单“租户与资源 ＞ YARN 队列管理”，进入 YARN 队列管理页面。如图所示：

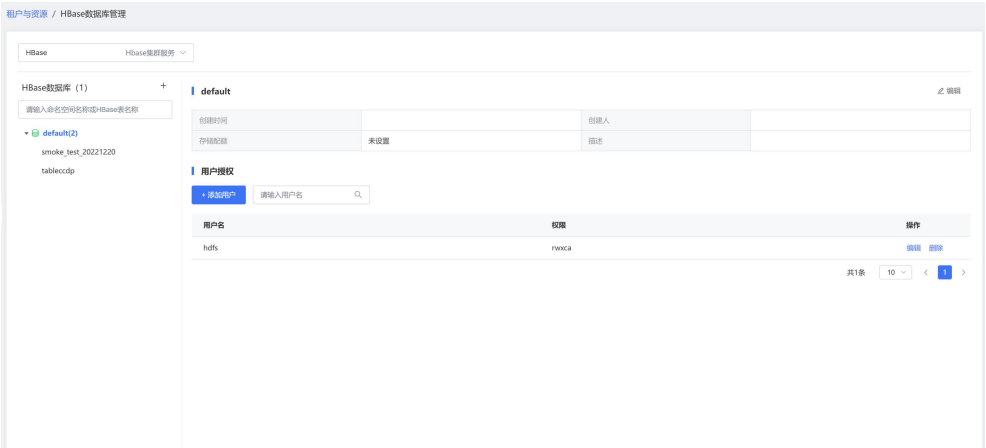


- 支持 YARN 队列新建、编辑与删除。
- 支持 YARN 队列启动与停止。
- 支持 YARN 队列容量设置和 YARN 队列授权。
- 支持 YARN 队列 capacity-scheduler.xml 的全局属性配置。
- 支持 YARN 队列的同步生效并支持查看同步生效记录。



HBase 数据库管理

进入到翼 MR Manager 以后，点击菜单“租户与资源 > HBase 数据库管理”，进入 HBase 数据库管理页面。如图所示：

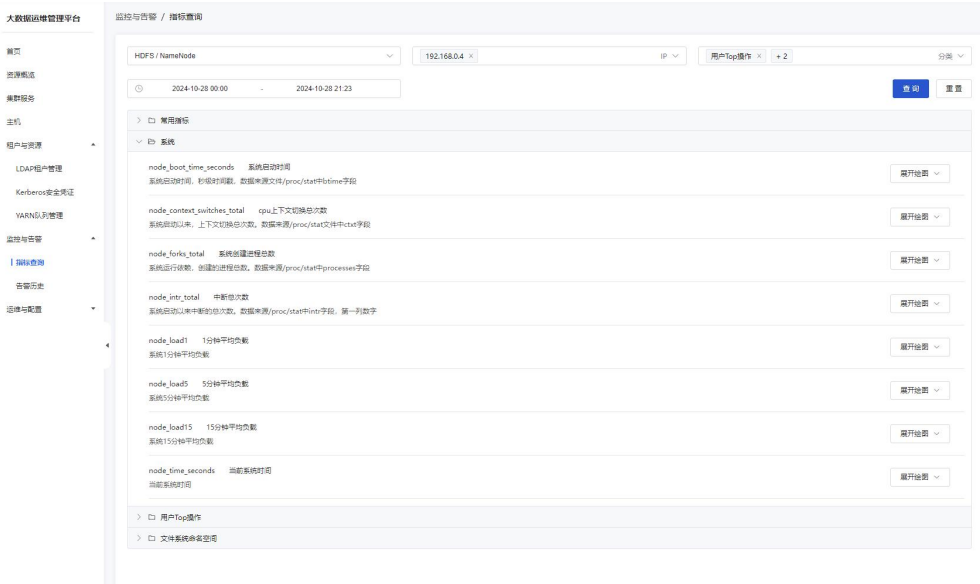


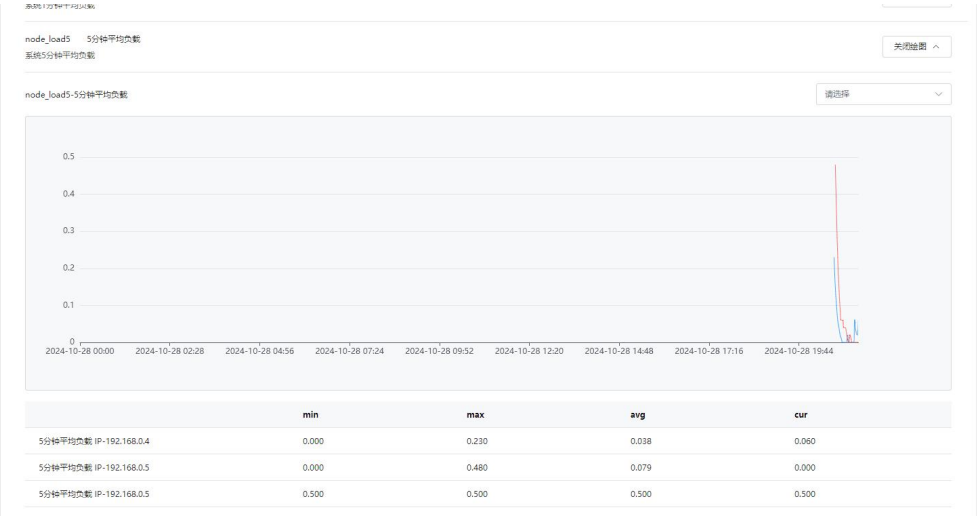
- 支持 HBase 命名空间的新建、删除。
- 支持表的新建、禁用、存储配额设置、列簇设置、分区设置、表用户设置。
- 支持从快照恢复数据。

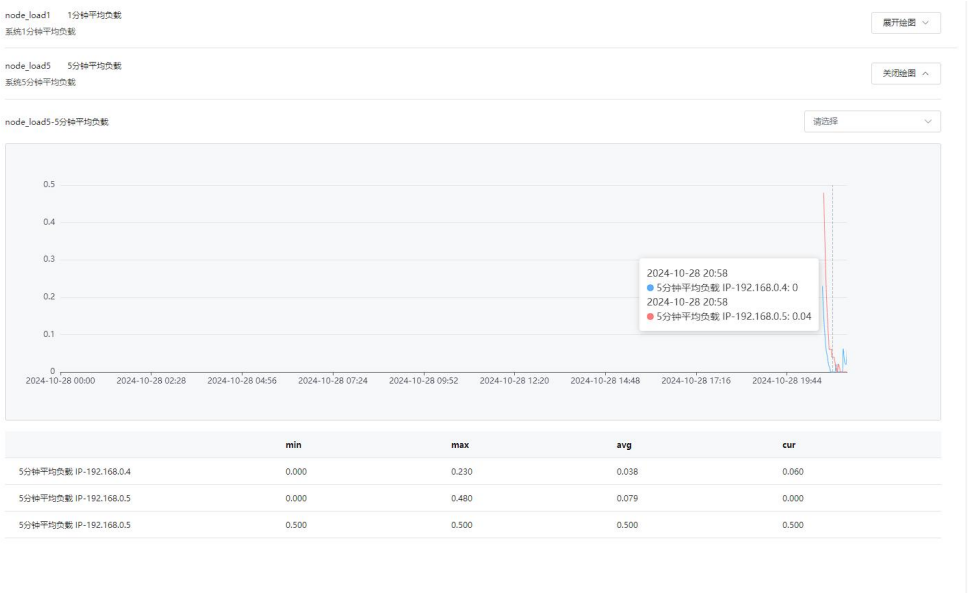
4.6.7 监控与告警

4.6.7 指标查询-指标查询概述

进入到翼 MR Manager 以后，点击菜单“监控与告警 > 指标查询”，进入指标查询页面。如图所示：



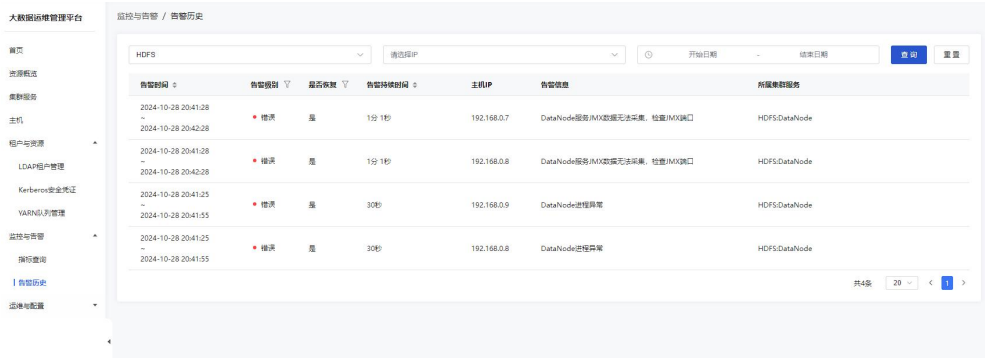




- 页面上方为查询区域，各种查询条件进行组合查询。
- 支持查询角色实例级、主机级的监控指标。
- 支持指标结果的绘图操作，让用户更直观获取监控项变化。

4.6.7 告警历史–告警历史概述

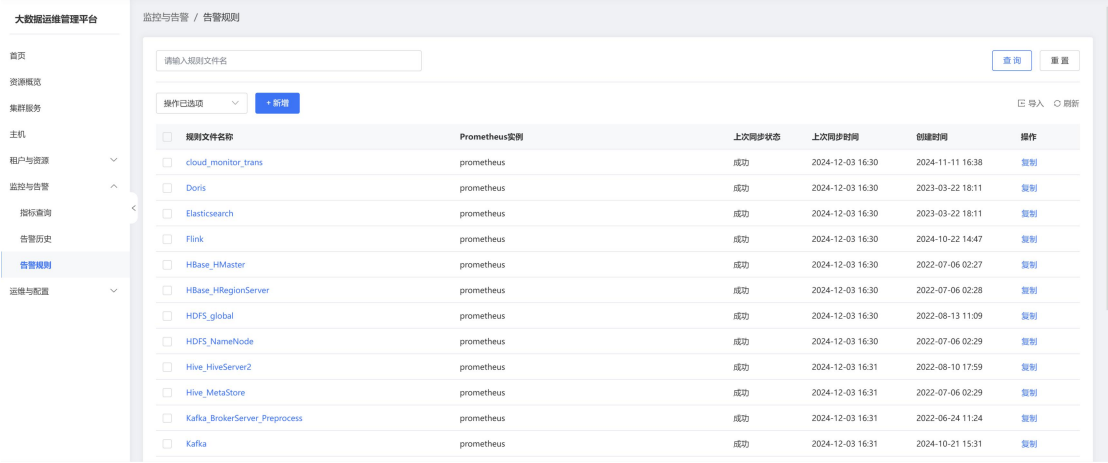
进入到翼 MR Manager 以后，点击菜单“监控与告警 > 告警历史”，进入告警历史页面。如图所示：



- 页面上方为查询区域，各种查询条件进行组合查询。
- 支持查询集群服务级、角色实例级、主机级的告警信息。

4.6.7 告警规则—告警规则概述

进入到翼 MR Manager 以后，点击菜单“监控与告警 ＞ 告警规则”，进入告警规则查询界面。如图所示：



ⓘ说明

1. 页面上方为查询区域，通过文件名，查询告警规则文件列表。
2. 支持新增、导入告警规则文件。
3. 支持勾选告警规则文件，进行同步、删除、导出操作。
4. 支持点击告警规则文件名，跳转到告警规则文件详情，查看配置的告警规则信息。

4.6.7 告警规则—告警规则详情

操作场景

查看告警规则配置信息，新增、修改、删除告警规则。

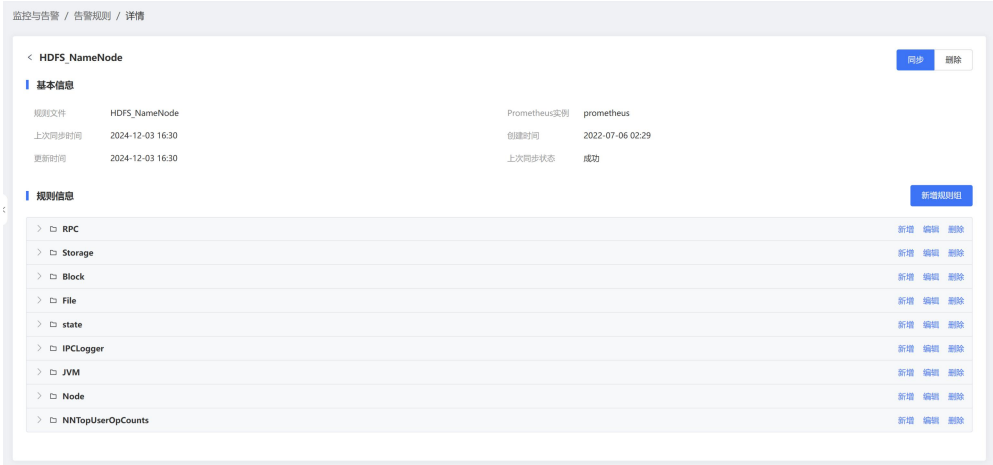
操作步骤

1. 登录翼 MR 管理控制台。

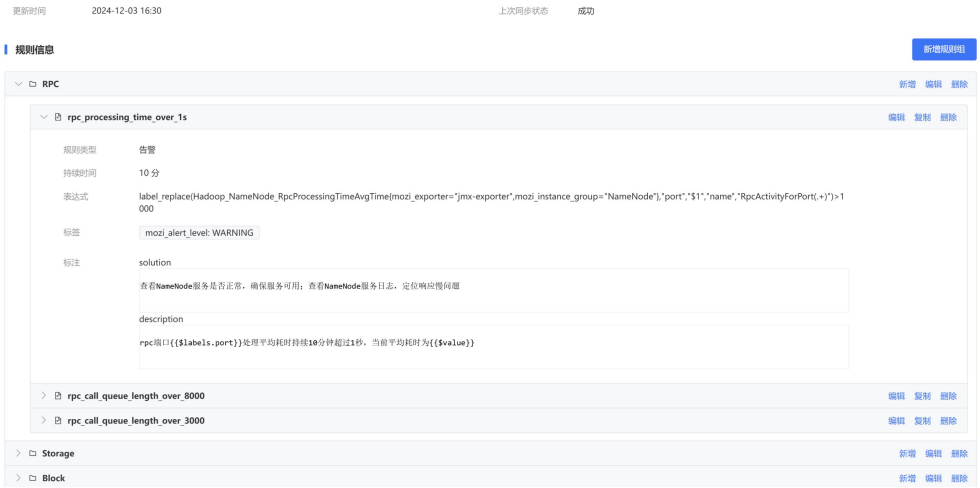
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

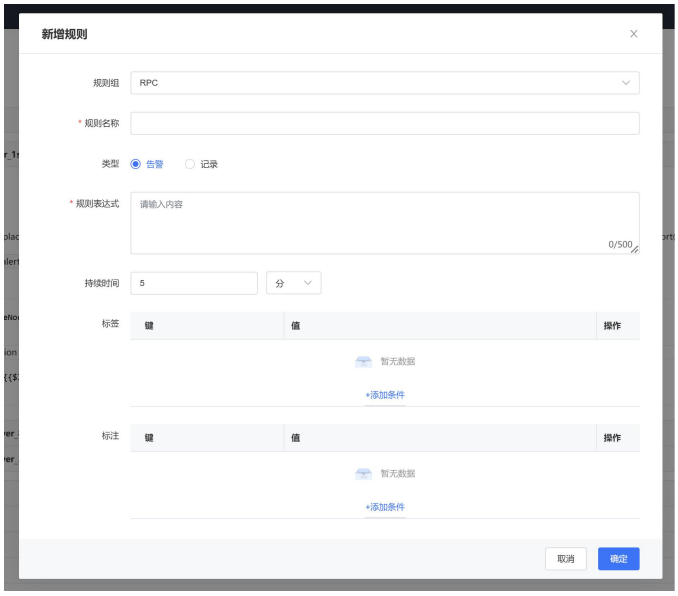
4. 进入到翼 MR Manager 以后，单击菜单“监控与告警 > 告警规则”，点击需要查看或变更的告警规则文件名，跳转至告警规则文件详情，如图所示。



5. “规则信息”一节展示所有告警规则分组，点击告警规则分组名，可展开组下告警规则。继续点击告警规则名，可展开单条告警规则的具体信息，如图所示。

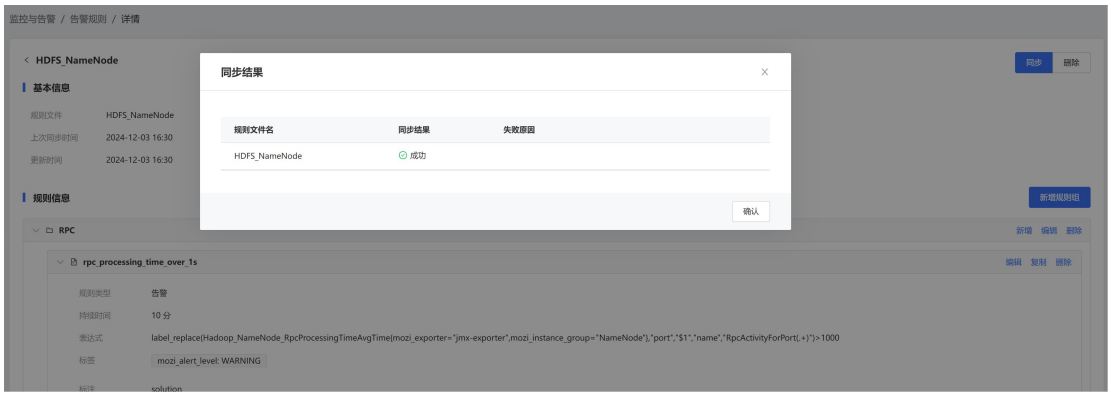


6. 点击告警规则组名右侧新增按钮，可新增告警规则，如图所示。



7. 点击告警规则名右侧编辑按钮，可编辑告警规则信息。点击告警规则名右侧删除按钮，可删除告警规则。

8. 对告警规则的编辑操作，需要点击右上角“同步”按钮，将告警配置同步到 Prometheus 实例并加载生效，如图所示。



说明

1. 告警规则各字段说明如下：
   (1) 规则类型：分为告警、记录两种类型。告警类型用于配置异常告警，当监控指标触发告警表达式时，触发告警。记录类型用于预聚合监控指标，依照表达式聚合计算。
   (2) 持续时间：仅“告警”类型具有，当表达式成立且持续超过所设时间，告警将会触发。如果希望表达式成立后立即触发告警，可设置为 0。需注意，持续时间设置过小，可能导致频繁无意义报警。
   (3) 表达式：填写 PromQL 语法的表达式。
   (4) 标签：预设字段 mozi_alert_level，用于定义告警级别，四个告警级别为 WARNING（警告）、ERROR（错误）、SEVERITY（严重）、DISASTER（灾难）。例如，配置标签 mozi_alert_level 为 SEVERITY，触发告警后，在“告警历史”页面，告警的级别为“严重”。
   (5) 标注：预设标注字段 description，用于定义告警信息，触发告警后，在“告警历史”页面，description 字段对应的信息，将显示为“告警信息”。
2. 预设的记录类型告警规则，与系统功能相关，不建议用户自行修改。
3. 对告警规则组、告警规则的所有新增、修改、删除操作，都需要通过“同步”按钮，同步到监控组件并生效。
4. 可以勾选一批告警规则文件导出为压缩文件，压缩文件可再次用于告警规则导入，导入操作会直接同步到监控组件并生效，不需要额外的同步操作。需要注意，导入过程，同名告警规则文件会被覆盖。
5. 可以勾选一批告警规则文件进行删除，删除操作会直接同步到监控组件并生效，不需要额外的同步操作。

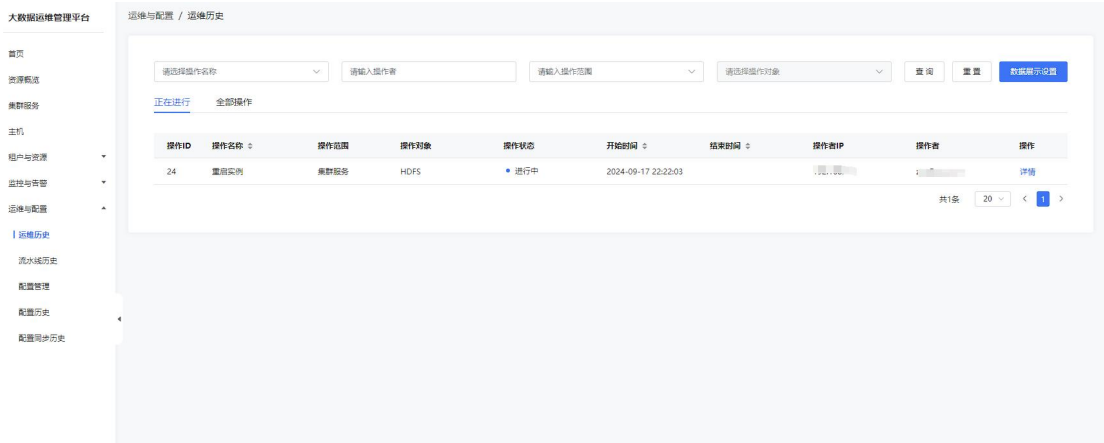
4.6.8 运维与配置

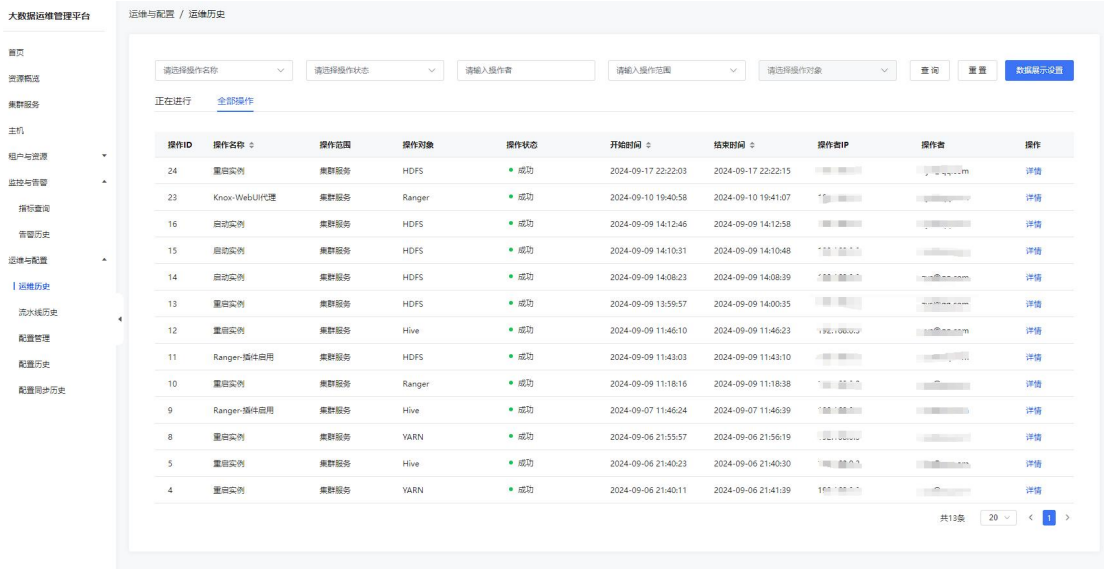
4.6.8 运维历史-查看运维历史

操作步骤

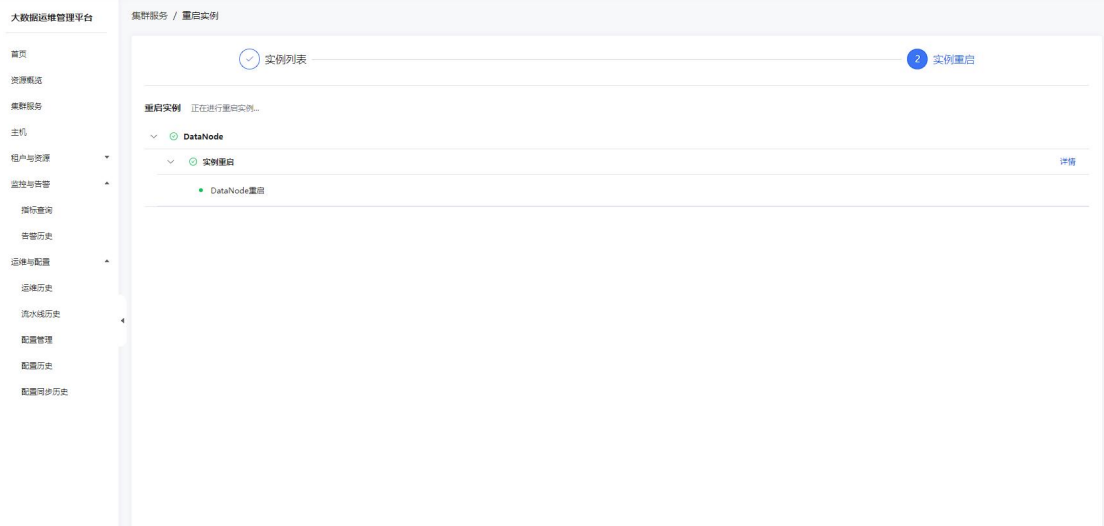
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 ＞ 运维历史”。默认打开“正在进行”标签页，展示正在进行的运维操作。可切换至“全部操作”标签页，展示进行中和已完成的运维操作历史，如图所示。





5. 单击想要查看的某个运维操作的详情按钮，进入到运维操作的详情页面，可查看运维操作的步骤、操作状态。如图所示：



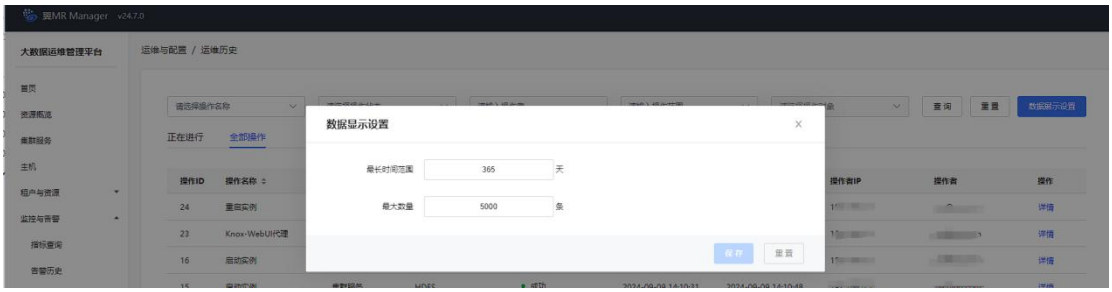
4.6.8 运维历史—数据展示设置

操作场景

设置运维历史展示的最长时间范围、最大数量。

操作步骤

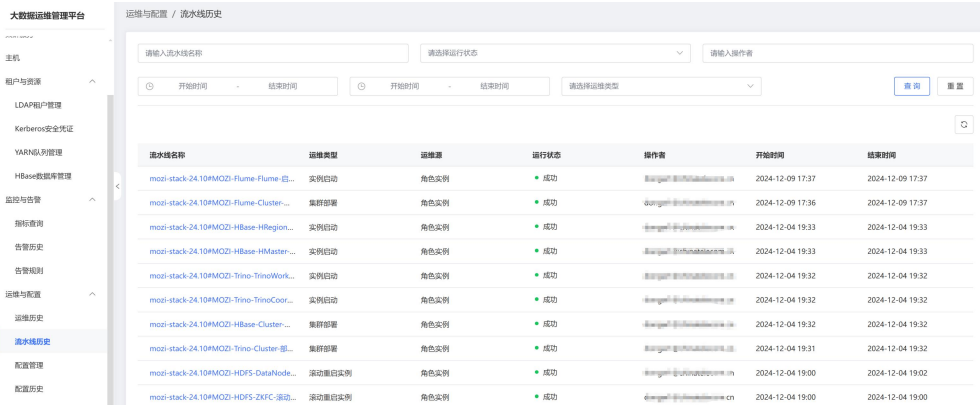
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 运维历史”。点击“数据展示设置”按钮，打开“数据显示设置”弹窗。
5. 可设置展示的最长时间范围、最大数量，如图所示：



4.6.8 流水线历史—流水线历史概述

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，点击菜单“运维与配置 > 流水线历史”，进入流水线历史页面。如图所示：



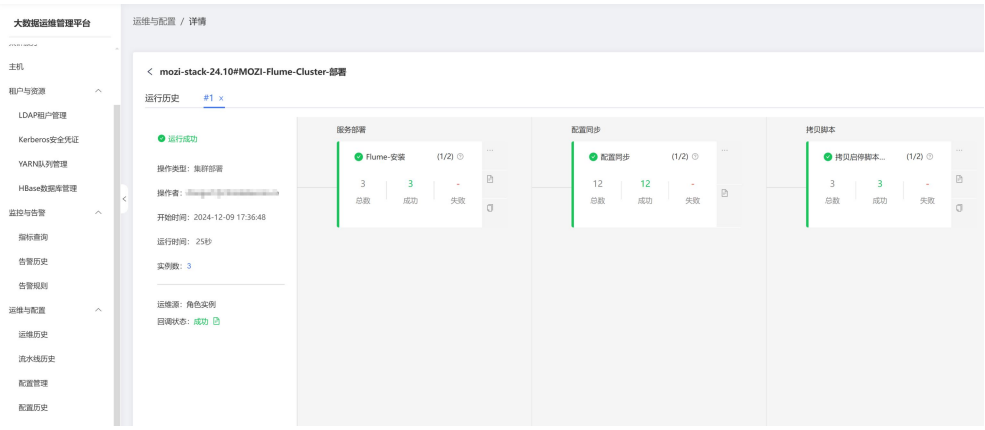
ⓘ说明

1. 页面上方为查询区域，各种查询条件进行组合查询。
2. 展示展示所有流水线的运行历史记录，以及操作人。
3. 单击流水线名称或详情按钮，可查看运行详情。

4.6.8 流水线历史—查看流水线运行详情

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 流水线运行历史”。
5. 单击要查看运行详情的流水线名称或详情按钮，进入到该运行详情页面。如图所示：

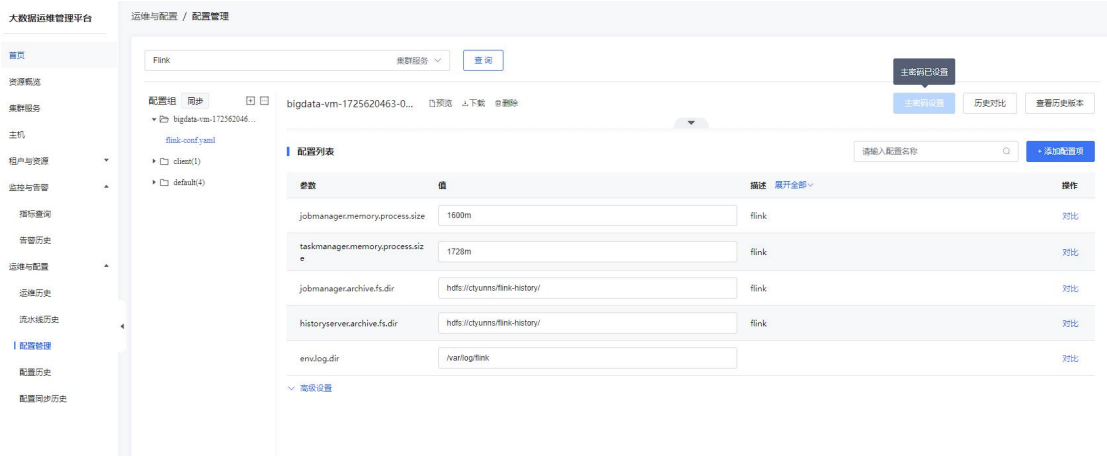


4.6.8 配置管理—主密码设置

主密码是用于查看集群下，隐私配置项的二次校验密码。主密码一经设置无法修改与查看，如果遗忘主密码请通过工单联系平台管理员恢复。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 点击“主密码设置”按钮，在弹窗中完成主密码的设置，主密码一经设置，不支持重置、修改、查看。请妥善保存。
6. 设置主密码成功后，主密码设置按钮置灰，如图所示：



4.6.8 配置管理—配置管理概述

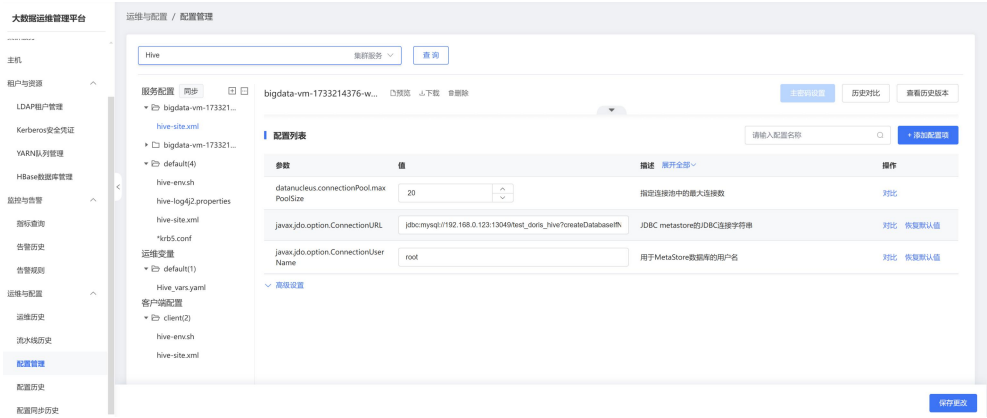
操作场景

用户需要管理集群服务的配置信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。

5. 在集群服务选框中，选择要管理的集群服务名称，点击“查询”按钮，即可查看集群服务配置信息。



ⓘ说明

1. 支持查看不同集群服务的配置。
2. 支持为不同集群服务新增配置组、查看配置组信息、删除配置组。
3. 支持新增配置文件、查看配置文件详情、预览和下载配置文件，以及删除配置文件。
4. 支持按主机维度的配置同步、查看配置状态等。

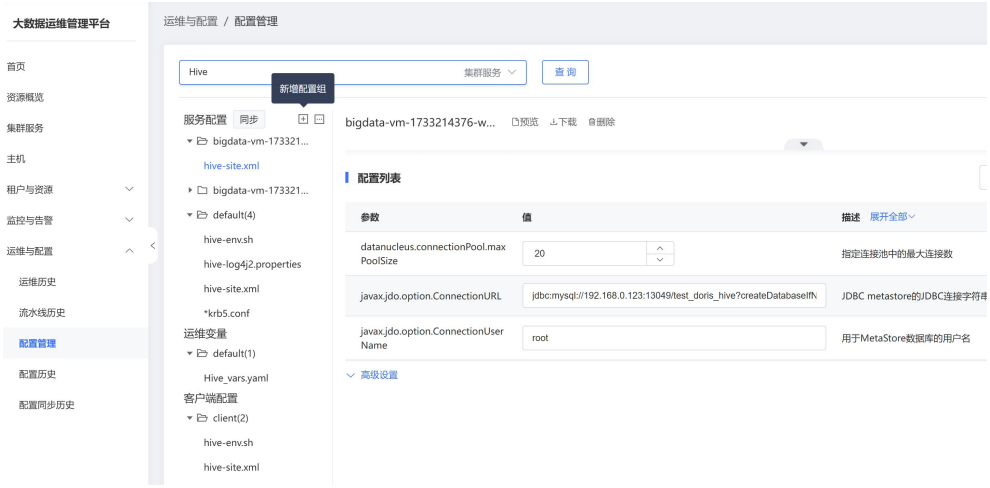
4.6.8 配置管理—新增配置组

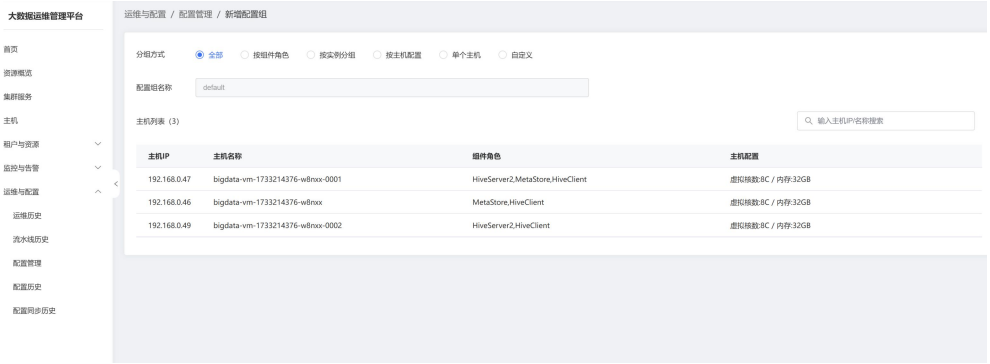
操作场景

用户可以根据不同的分组标准，对集群服务中的主机进行分组，一个配置分组下的主机共享一套配置。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击“新增配置组”，进入新增配置组页面。根据需要选择分组方式，选择主机，单击右下角“保存”按钮即可。如图所示：





ⓘ注意

1. 分组方式为：全部、按组件角色、按实例分组、按主机配置时，配置组名称默认且不可改。
2. 分组方式为：单个主机时，需要选择要分组的主机，配置组名称为所选主机的名称且不可改。
3. 分组方式为：自定义时，需要选择分组的主机，配置组名称可以自定义但不能与其他配置组名称相同。

4.6.8 配置管理—查看配置组详情

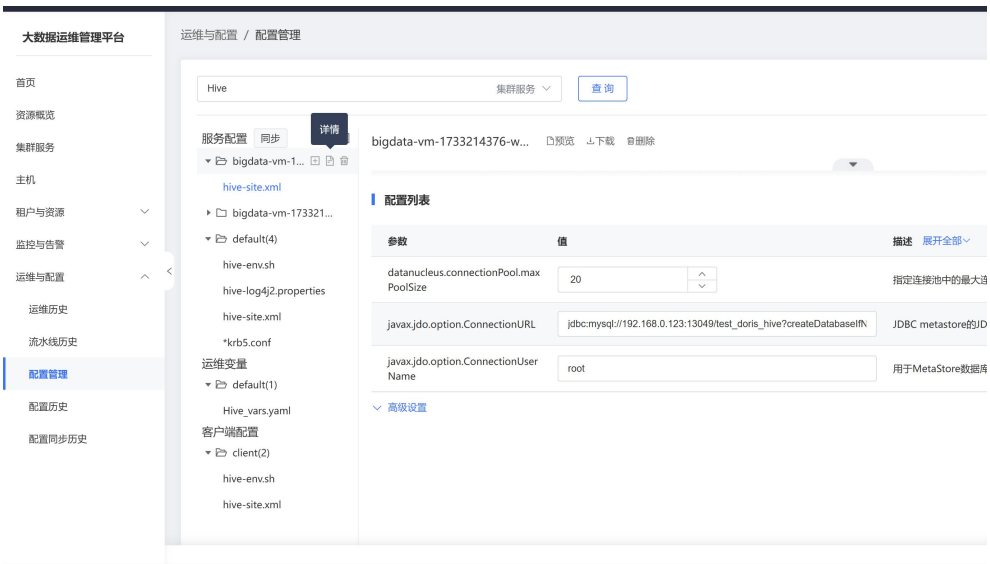
操作场景

用户可以查看不同配置组的详细信息，了解该配置组所包含的主机信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 选择要查看的配置组，鼠标移入到配置组名称上，单击右侧的“详情”图标，进入到配置组详情页面。如图所示：



4.6.8 配置管理—删除配置组

操作场景

用户可以删除不需要的配置组。

前置条件

配置组下的配置文件数量必须为 0 才能删除。

操作步骤

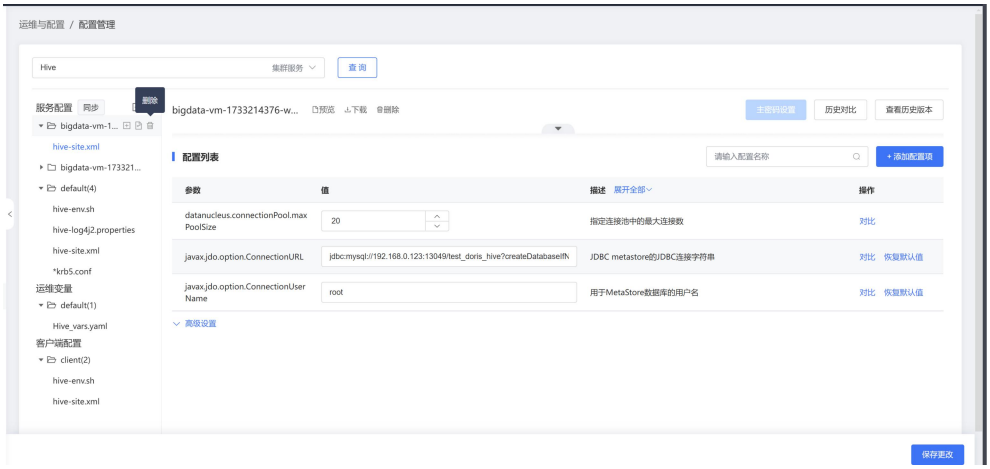
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。

5. 查询条件选择指定集群服务。

6. 选择要删除的配置组，鼠标移入到配置组名称上，单击右侧的“删除”图标，出现删除确认弹框。



7. 单击“确定”即可删除该配置组。

4.6.8 配置管理—新增配置

操作场景

用户可以根据不同的分组标准，对集群服务中的主机进行分组，一个配置分组下的主机共享一套配置。

操作步骤

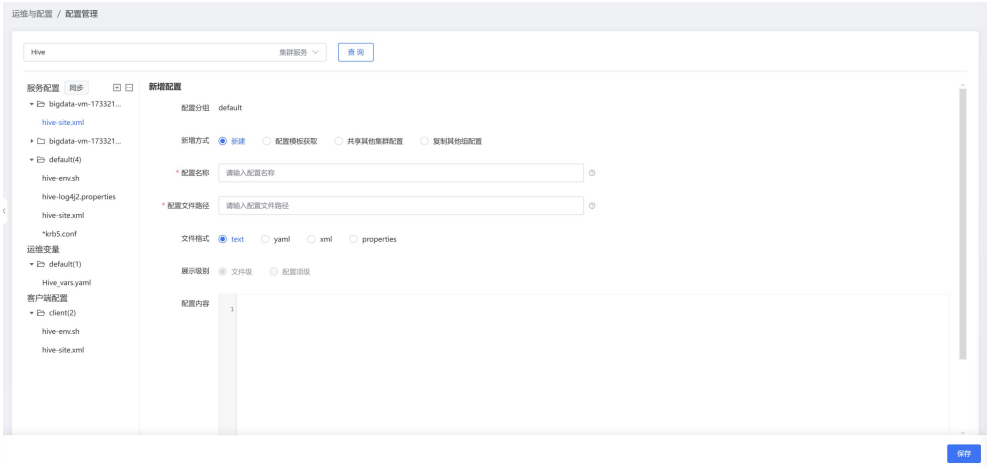
1. 登录翼 MR 管理控制台。

2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。

5. 查询条件选择指定集群服务，查询指定集群配置信息，如图所示：



6. 新增配置操作，同配置管理–新增配置。

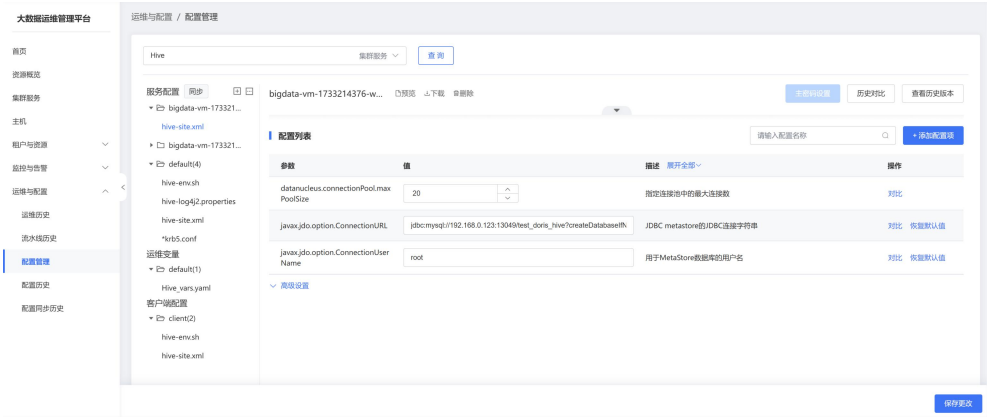
4.6.8 配置管理–查看配置详情

操作场景

用户可以查看某个配置的详细信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击要查看详情的配置名称，页面右侧展示该配置的详细信息。点击三角图标，可以展开配置文件详细信息，如图所示：



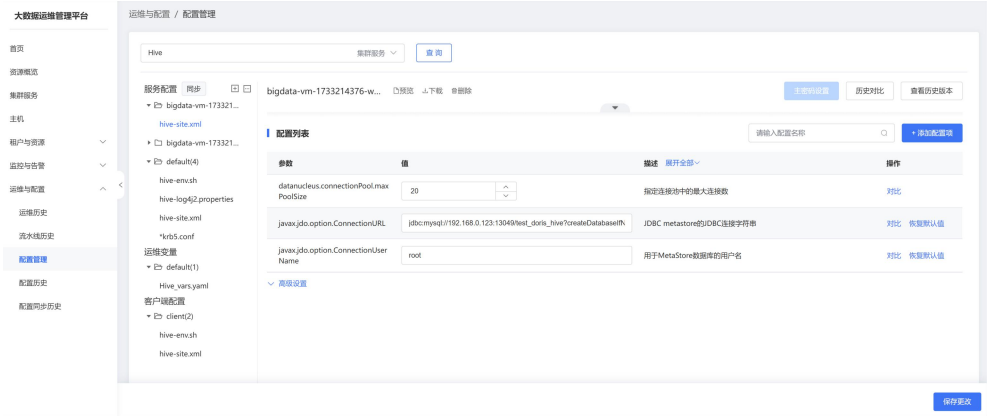
4.6.8 配置管理—编辑配置

操作场景

用户可以对配置文件进行修改。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击要修改的配置名称，页面右侧展示该配置的详细信息。如图所示：



7. 修改完成后点击“保存更改”按钮。

说明

1. 服务端配置，修改后需要执行同步操作。在配置文件及同步按钮上有数字提示未同步修改。
2. client 分组下配置、default 分组下以 vars.yaml 结尾的配置，不需要同步，修改后不会有数字提示，也不需要执行同步操作。
3. 对于同步按钮，提示数字的含义为待同步配置文件的数量。
4. 对于配置文件前同步提示数字，文件级和配置项级有所区别，具体含义说明如下。
   (1) 展示级别为文件级：修改后且当前配置文件未同步时，配置名称前展示数字 1。
5. 展示级别为配置项级：修改后且当前配置文件未同步时，配置名称前展示所修改的配置项数，所修改的配置项名称前展示修改标识。

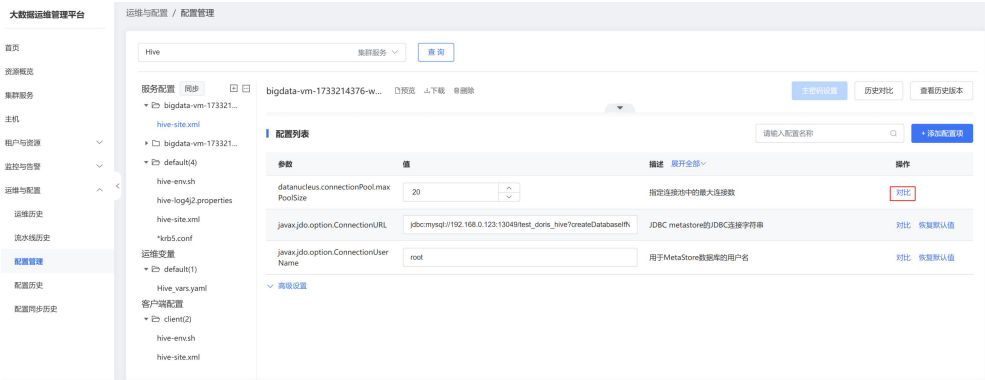
4.6.8 配置管理—配置项对比

操作场景

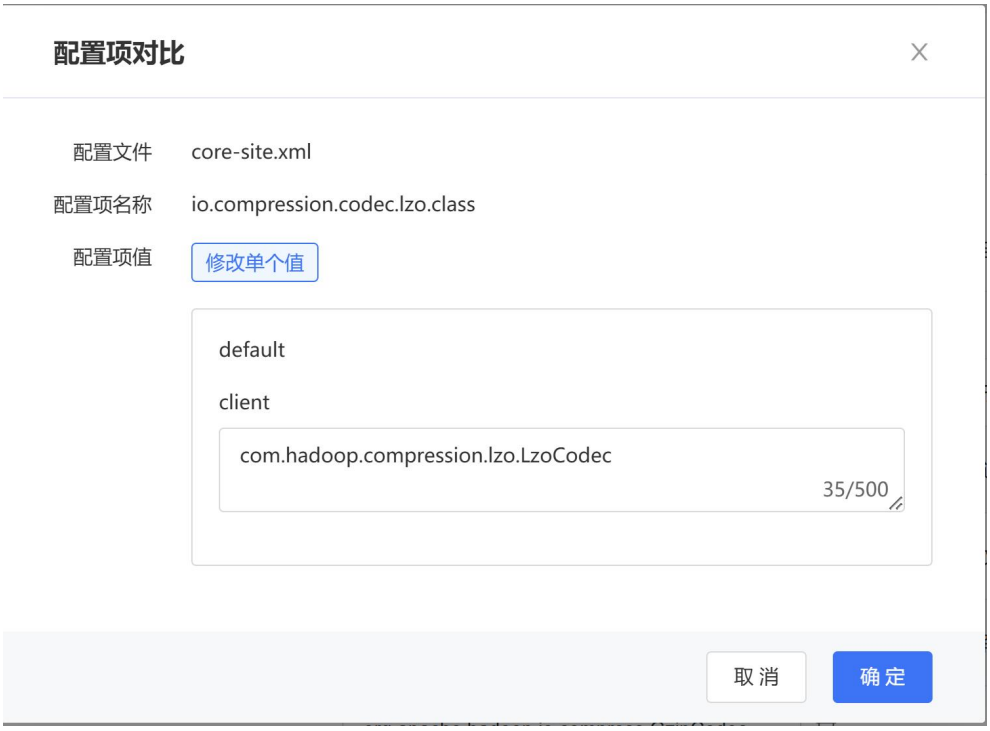
用户可以查看同一集群下相同配置项名称在不同配置组中的值。

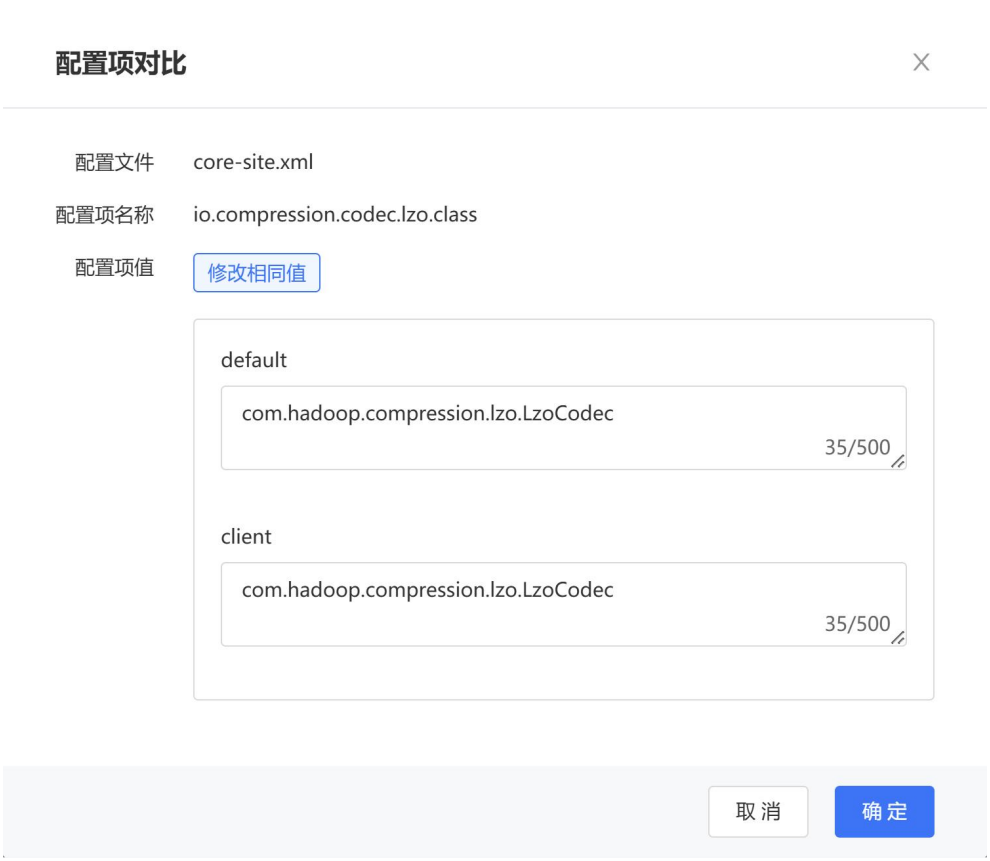
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击“配置管理”tab。
7. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
8. 选择要对比的配置项，单击该配置项名称右侧的“对比”按钮，出现配置项对比弹框。如图所示：



9. 配置项对比弹框中，当多个配置分组下，存在同一配置项时，支持逐条展示或合并相同值展示与编辑。如图所示，default、client 分组下同一配置项具有相同的值，支持当设置为“修改单个值”或“修改相同值”，进行合并或逐条展示与编辑。



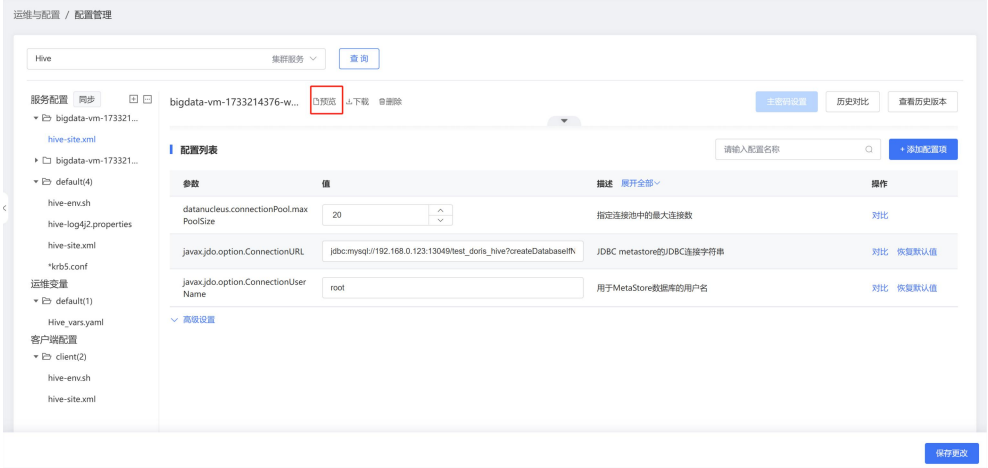


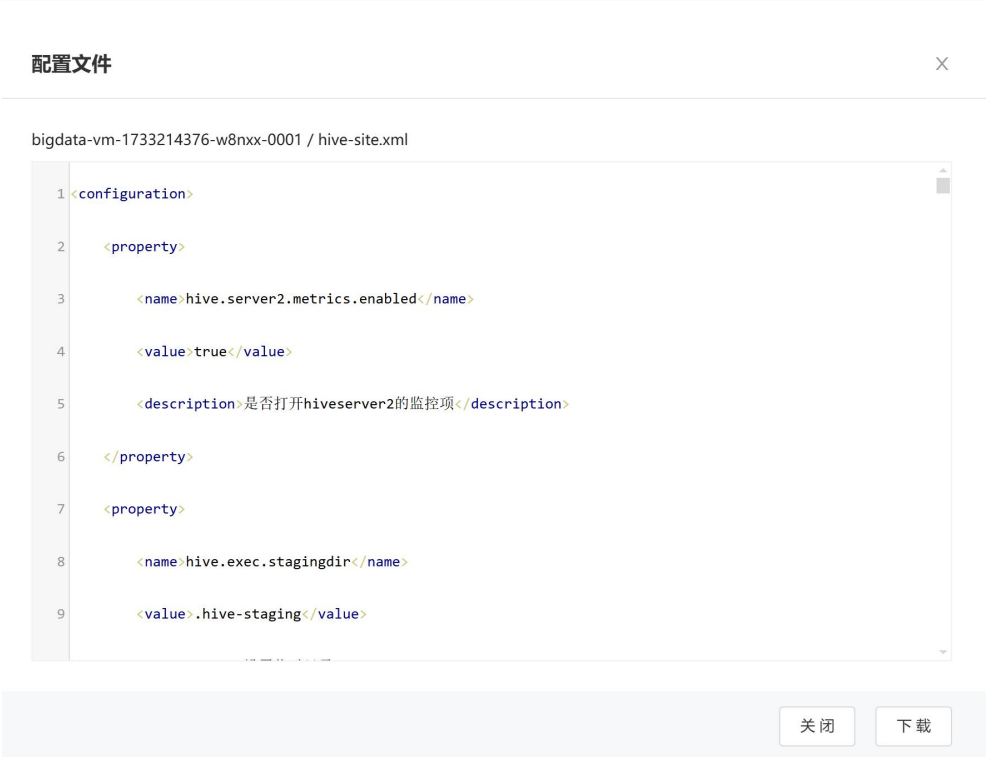
4.6.8 配置管理—配置文件预览

本章节主要介绍翼 MR Manager 的集群服务配置文件预览的操作。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 ＞ 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“预览”，出现配置文件预览弹框。如图所示：

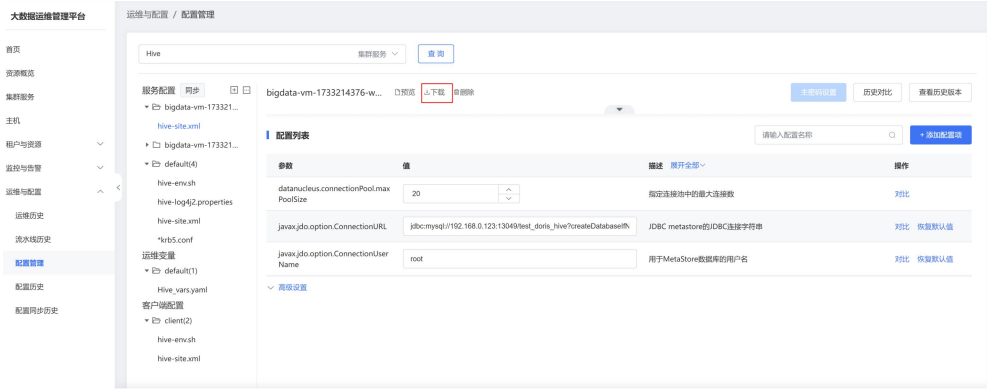




4.6.8 配置管理—下载配置文件

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“下载”，等待下载完成即可。如图所示：



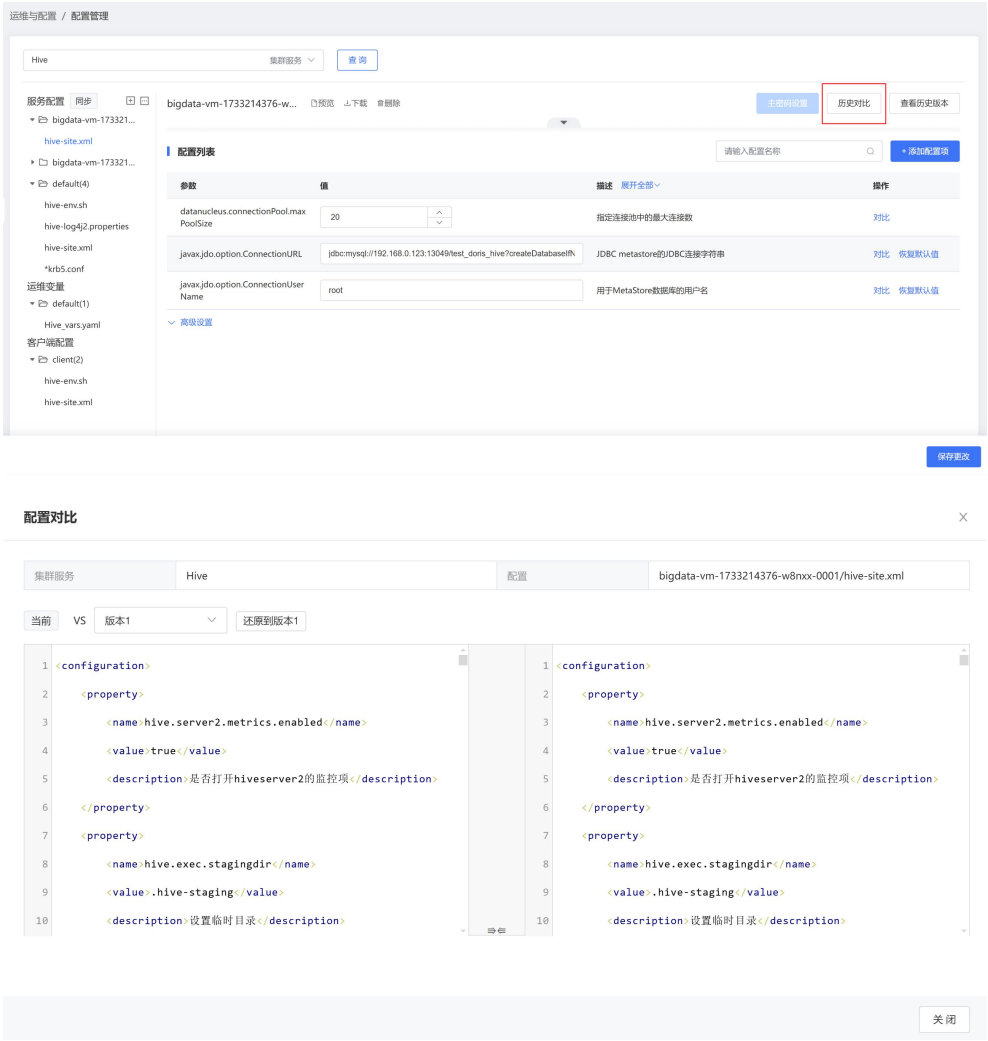
4.6.8 配置管理—配置文件历史对比

操作场景

支持用户查看当前配置文件与不同历史版本的差异。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“历史对比”，出现历史对比弹框。如图所示：



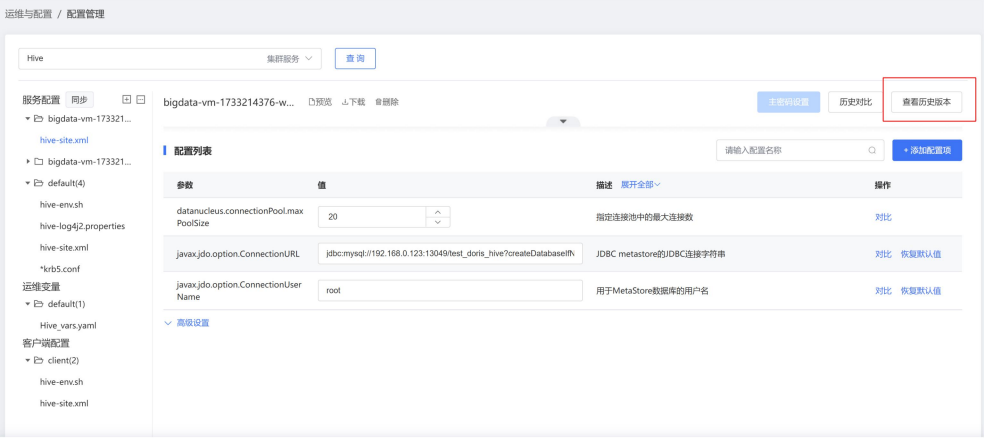
4.6.8 配置管理—查看配置历史版本

操作场景

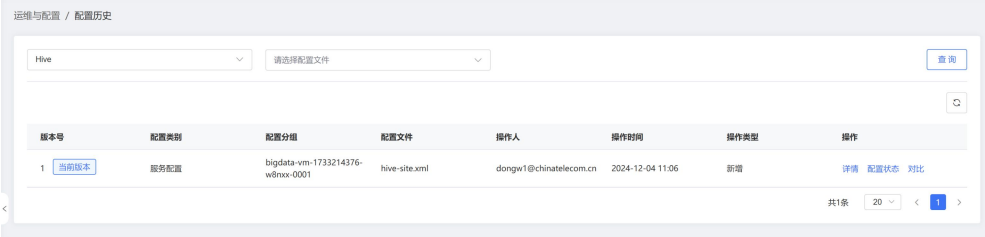
支持用户查看指定配置文件的历史版本，不同版本的配置内容以及差异。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“查看历史版本”。如图所示：



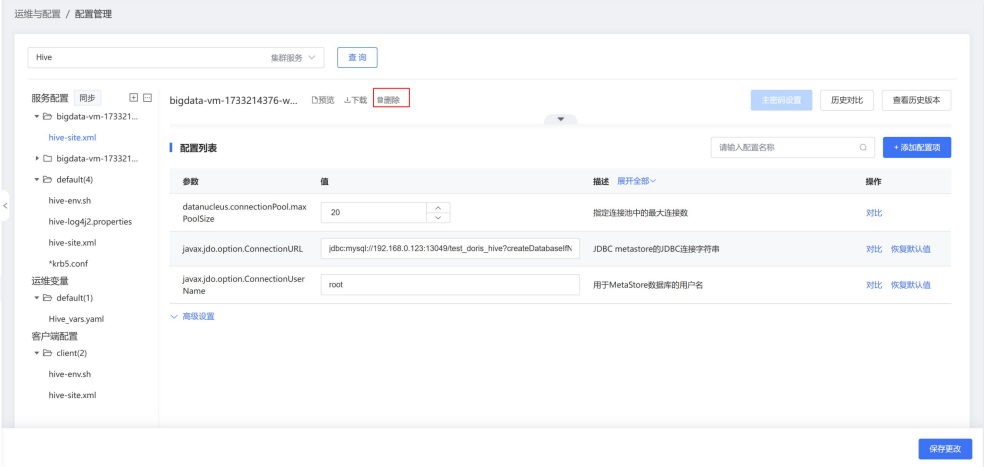
8. 进入到配置历史页面，配置历史页面展示当前配置的所有历史版本。如图所示：



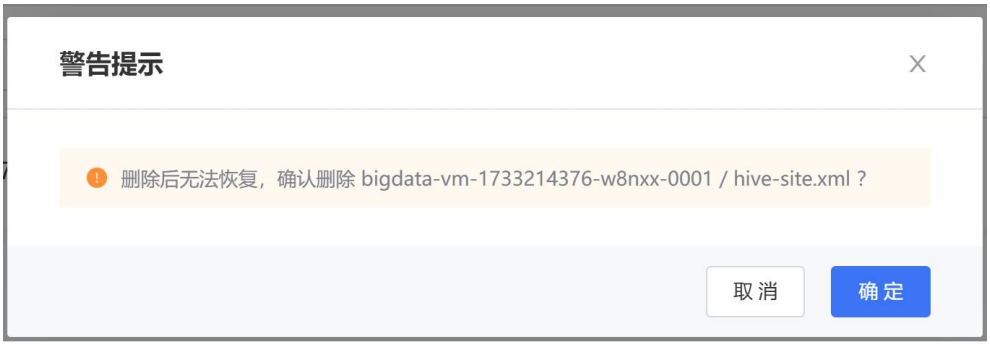
4.6.8 配置管理—删除配置

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“删除”。如图所示：



8. 出现删除确认弹框，单击“确定”即可。如图所示：



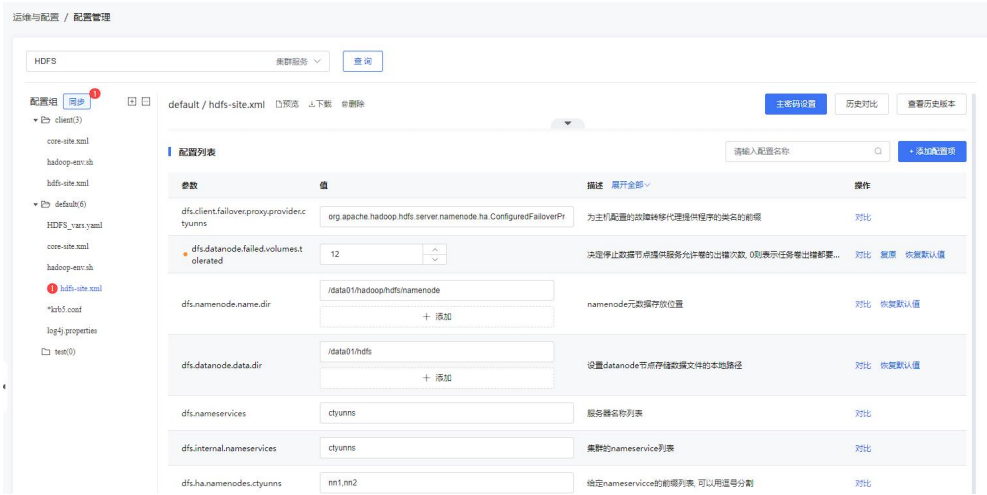
4.6.8 配置管理—配置同步

操作场景

支持将新的配置文件或修改后的配置同步到相应主机上。

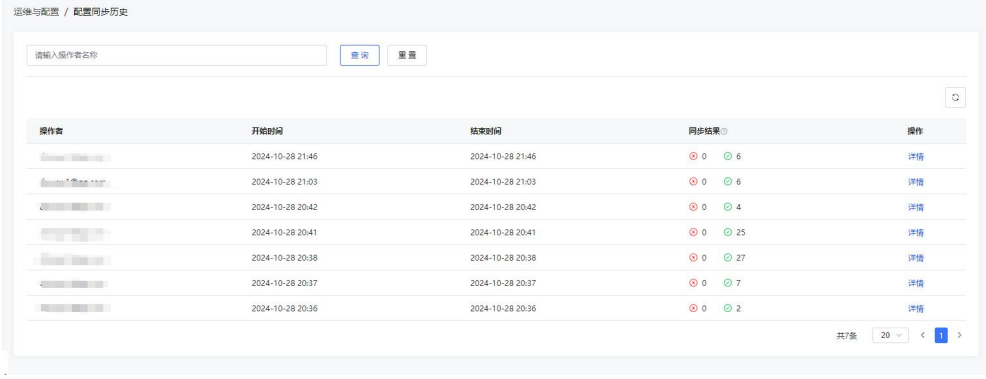
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击“同步”按钮。如图所示：

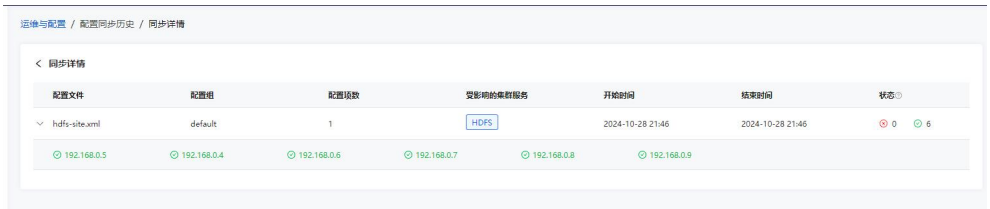


7. 进入待同步配置页面。如图所示：

8. 单击“配置同步”按钮，跳转到配置同步菜单页面。如图所示：



9. 单击“详情”操作按钮，进入到配置同步详情页面，查看配置同步进度，等待同步完成。如图所示：



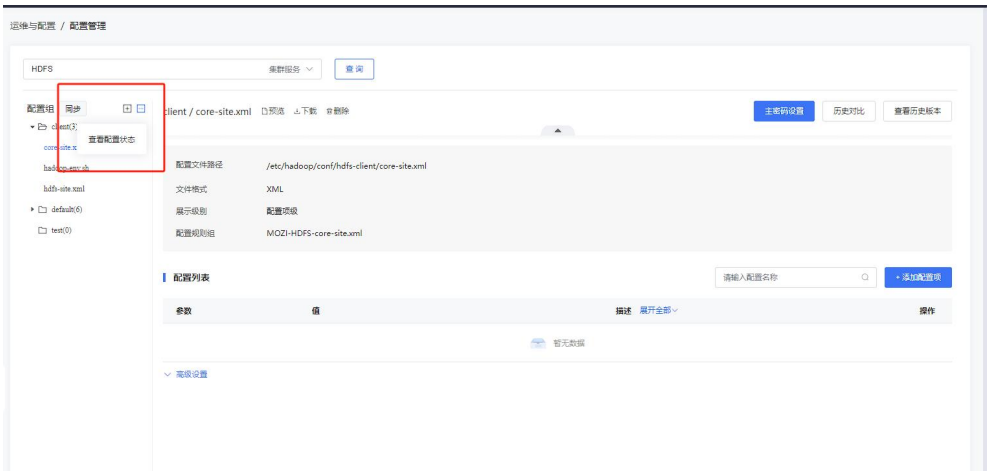
4.6.8 配置管理—查看配置状态

操作场景

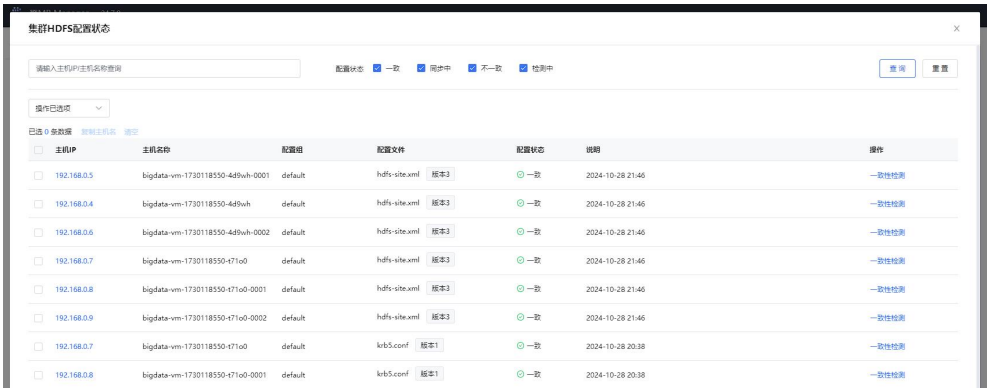
支持查看指定集群服务所有配置文件的同步状态。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 ＞ 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击“查看配置状态”按钮。如图所示：



7. 出现配置同步状态弹框，展示所有配置文件的同步状态。如图所示：



8. 可进行配置一致性校验，勾选需要校验的主机后，点击操作已选项按钮，点击一致性检测，比对主机上配置文件是否与 Manager 上一致。点击校验后，“配置状态”

列将进入“检测中”状态，检测完成会自动刷新。说明列将展示上一次检测配置一致性状态的时间。

4.6.8 配置管理-重新同步

操作场景

支持对同步失败的配置或配置状态不一致的进行重新同步。

操作步骤

1. 登录翼 MR 管理控制台。

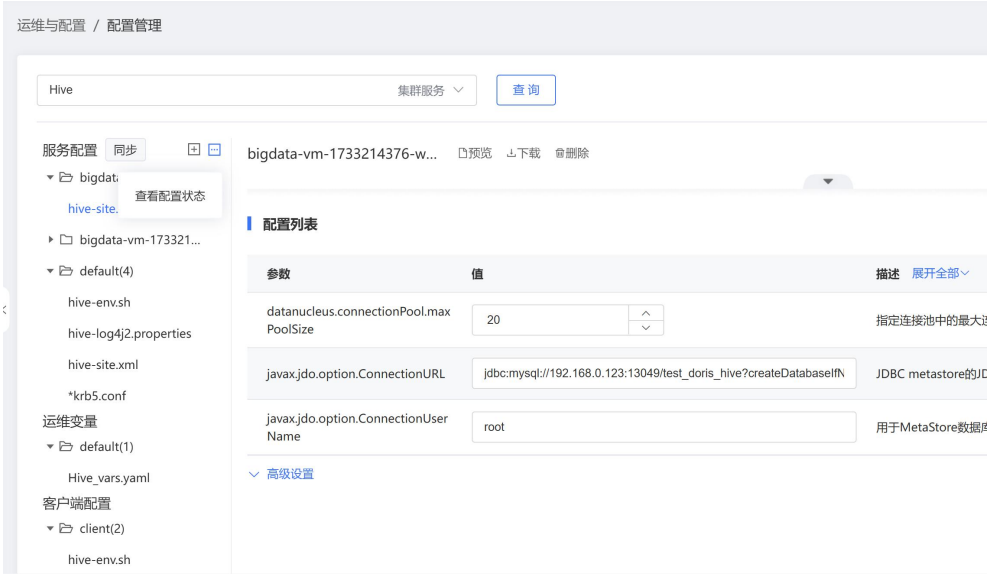
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

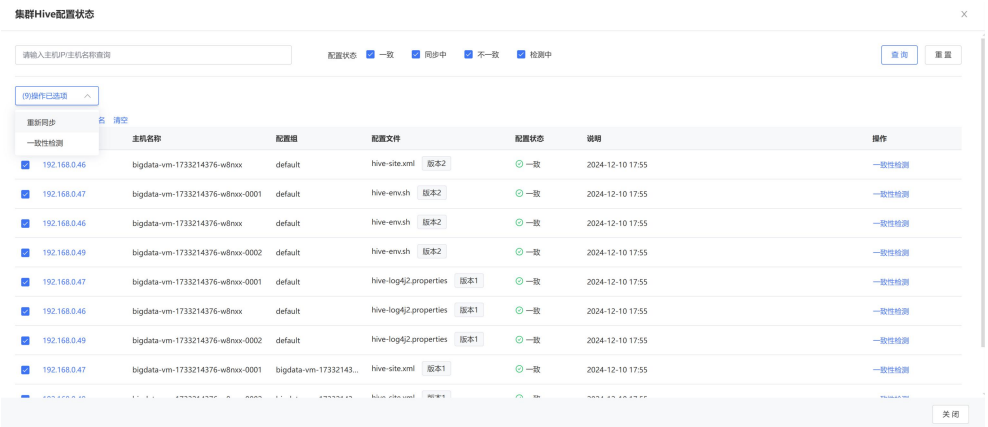
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置管理”。

5. 查询条件选择指定集群服务。

6. 单击“查看配置状态”按钮，出现配置同步状态弹框。



7. 勾选需要操作主机前的复选框。单击“操作已选项 > 重新同步”，即可支持配置同步操作。如图所示：



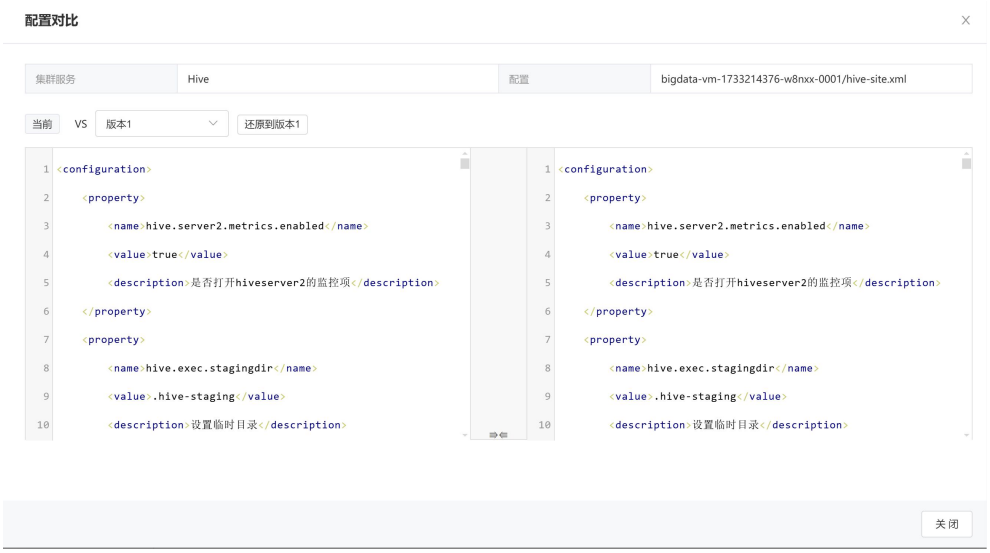
4.6.8 配置管理—配置回滚

操作场景

支持回退指定配置文件为某个历史版本。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 ＞ 配置管理”。
5. 查询条件选择指定集群服务。
6. 单击配置组，单击配置文件名称，页面右侧展示该配置的详细信息。
7. 单击“历史对比”，出现历史对比弹框。如图所示：



8. 单击“还原到版本 X”按钮，出现确认操作提示弹框。单击“确认”还原配置文件版本。如图所示：

9. 单击“关闭”按钮，关闭配置对比弹框。
10. 查看配置文件，则配置文件内容与所还原的版本内容一致。
11. 单击“同步”，对配置文件进行同步操作即可。

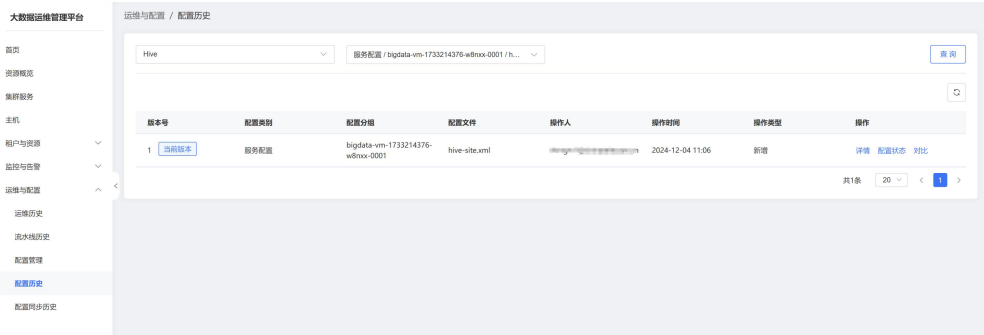
4.6.8 配置历史-配置历史概述

操作场景

用户需要查看配置历史信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置历史”。
5. 选择要查看的集群服务、配置文件，点击“查询”按钮，即可查看配置文件历史，如图所示：



说明

1. 页面上方为查询区域。
2. 支持查看不同配置文件的所有历史版本信息。
3. 支持查看配置文件不同版本的配置内容、并支持不同版本之前的内容对比。
4. 支持查看配置文件当前版本的同步详情。

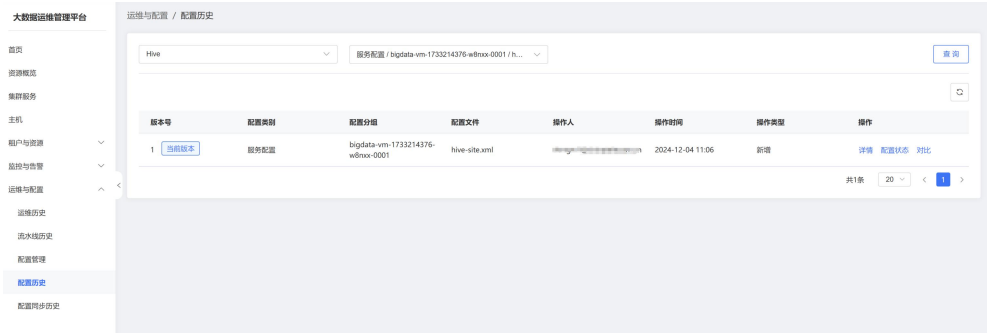
4.6.8 配置历史-查询配置历史

操作场景

用户可以查询某个配置文件的历史版本。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置历史”。
5. 查询条件选择指定集群服务、选择指定配置组、配置文件，下方展示该配置文件的所有历史版本。如图所示：



4.6.8 配置历史-查看配置详情

操作场景

用户在执行配置同步后，用户需要查看不同版本配置文件的详细信息。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置历史”。
5. 查询条件选择指定集群服务、选择指定配置组、配置文件。
6. 选择要查看的版本，单击“详情”按钮，出现该版本配置文件的详情弹框。如图所示：

4.6.8 配置历史-查看配置状态

操作场景

支持查看最新版本的配置状态。

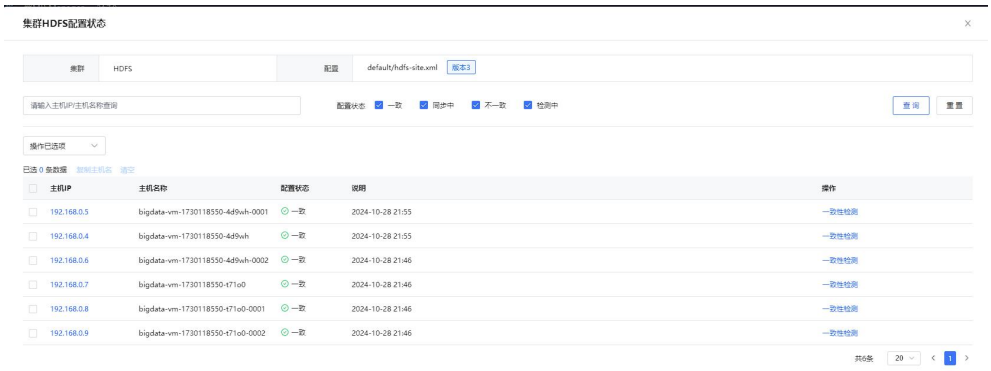
操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置历史”。

5. 查询条件选择指定集群服务、选择指定配置组、配置文件。

6. 单击“配置状态”按钮，出现配置状态弹框。如图所示：



4.6.8 配置历史—查看配置对比

操作场景

支持查看配置文件不同版本之前的差异。

操作步骤

1. 登录翼 MR 管理控制台。

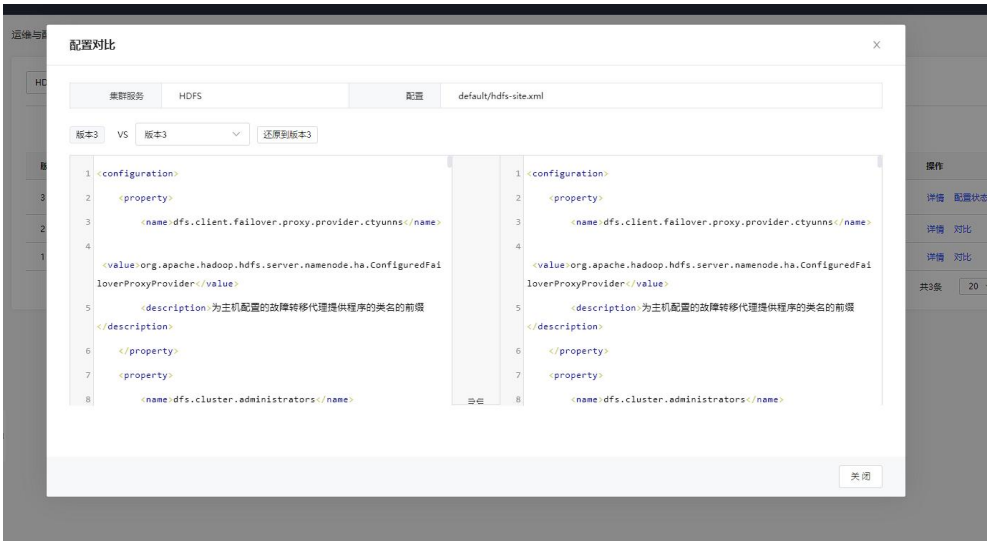
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。

3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。

4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置历史”。

5. 查询条件选择指定集群服务、选择指定配置组、配置文件。

6. 单击“对比”按钮，出现配置对比弹框。如图所示：



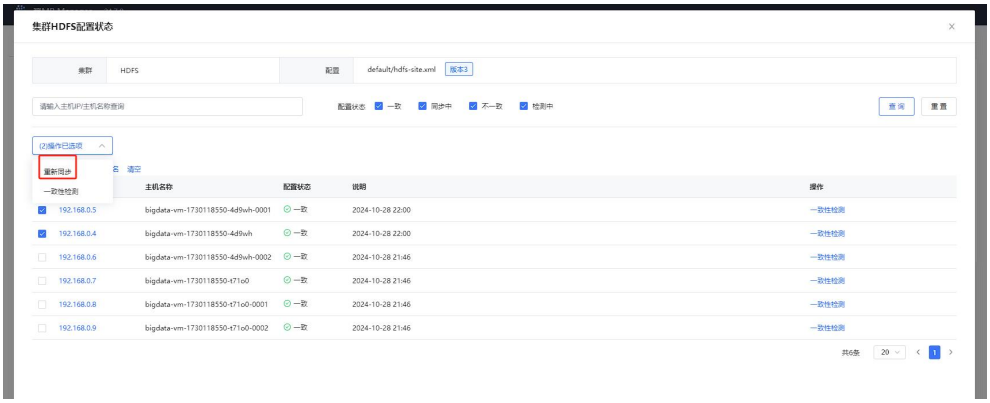
4.6.8 配置历史-重新同步

操作场景

支持查看最新版本的配置状态。

操作步骤

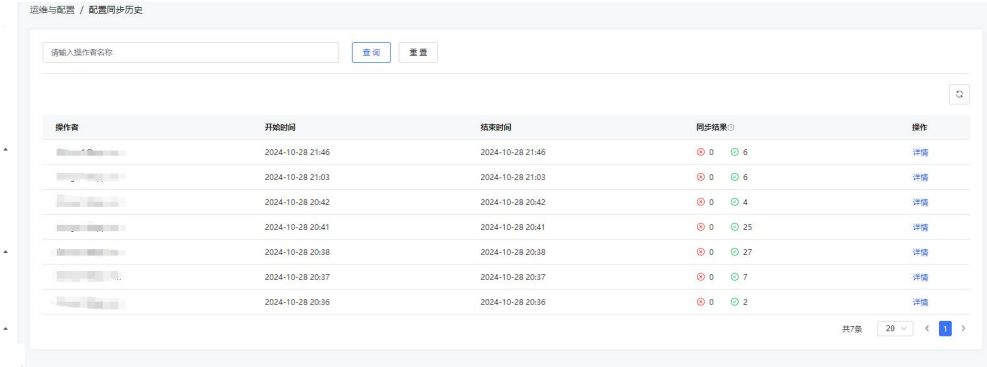
1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置历史”。
5. 查询条件选择指定集群服务、选择指定配置组、配置文件。
6. 单击“配置状态”按钮，出现配置状态弹框。
7. 勾选需要操作主机前的复选框。
8. 单击“操作已选项 > 重新同步”，即可支持配置同步操作。如图所示：



4.6.8 配置同步历史—配置同步历史概述

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，点击菜单“运维与配置 > 配置同步历史”，进入配置同步历史页面。如图所示：



说明

1. 页面上方为查询区域。
2. 展示所有配置同步操作的历史、配置同步操作人，以及配置同步详情。

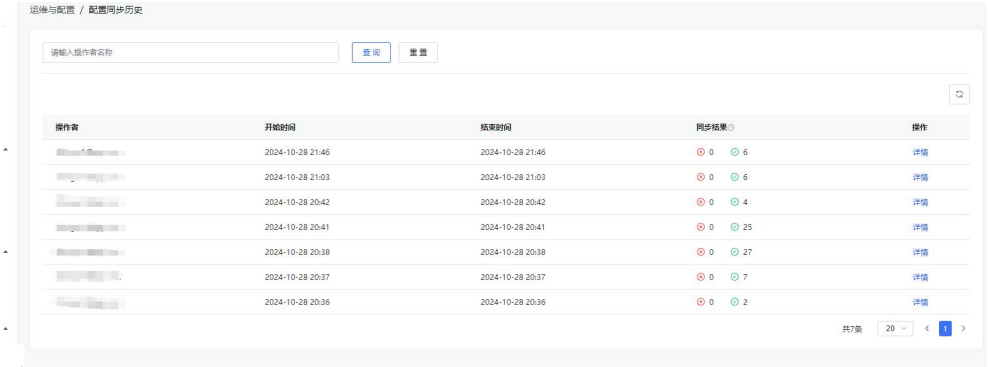
4.6.8 配置同步历史—查看配置同步状态

操作场景

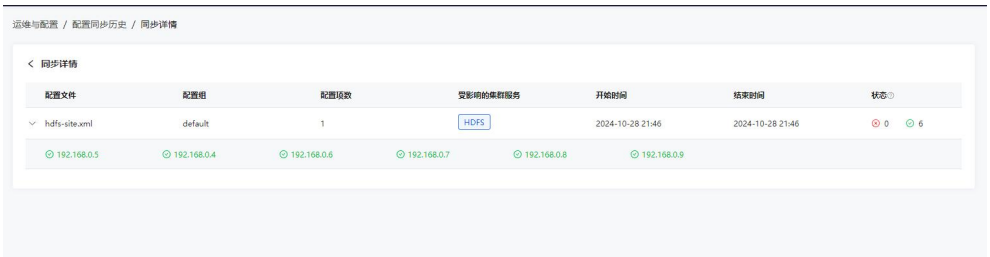
用户在配置管理进行配置同步后，需要查看本次同步操作的同步状态或同步结果。

操作步骤

1. 登录翼 MR 管理控制台。
2. 单击“我的集群”，单击指定的集群名称，进入集群信息页面。
3. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
4. 进入到翼 MR Manager 以后，单击菜单“运维与配置 > 配置同步历史”。如图所示：



5. 选择要查看的操作记录，单击“同步中”或同步结果，进入配置同步详情页面。如图所示：



4.7 安全性说明

4.7.1 安全认证原理和认证机制

功能

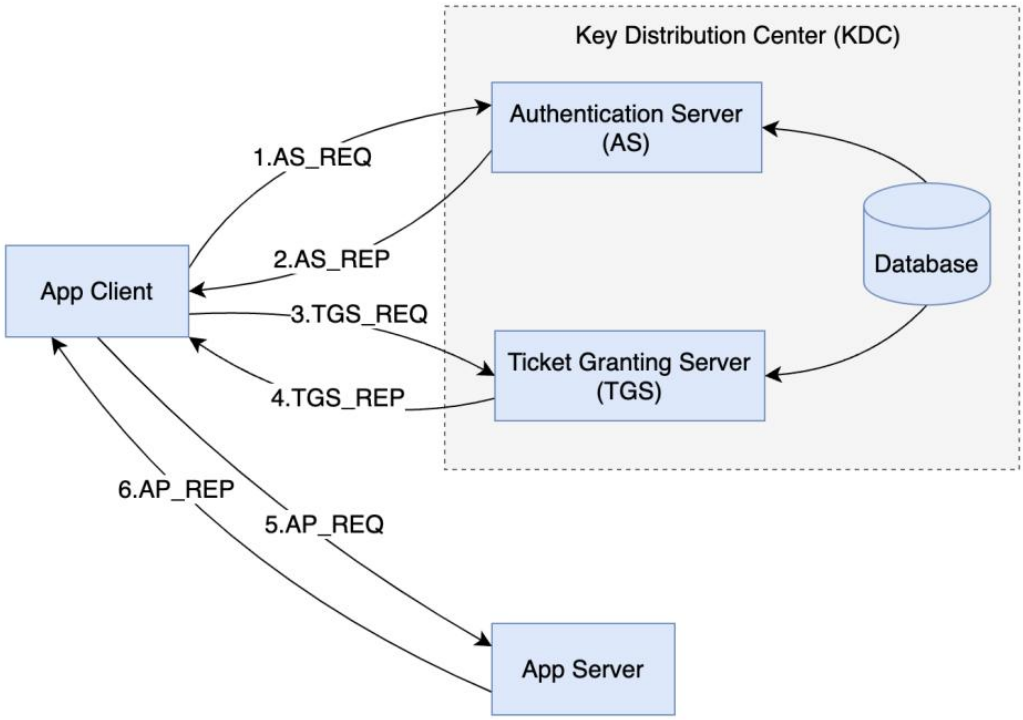
开启了 Kerberos 认证的安全模式集群，进行应用开发时需要进行安全认证。

Kerberos 这一名词来源于希腊神话“三个头的狗——地狱之门守护者”，后来沿用作为安全认证的概念，使用 Kerberos 的系统在设计上采用“客户端/服务器”结构与 AES 等加密技术，并且能够进行相互认证（即客户端和服务器端均可对对方进行身份认证）。可以用于防止窃听、防止 replay 攻击、保护数据完整性等场合，是一种应用对称密钥体制进行密钥管理的系统。

原理

Kerberos 的原理架构和各模块的说明如下图所示：

原理架构



模块说明

- App Client：应用客户端，通常是需要提交任务（或者作业）的应用程序。
- App Server：应用服务端，通常是应用客户端需要访问的应用程序。
- Key Distribution Center（KDC）：提供安全认证的服务。
  - Database：存储 Principal 数据。
  - Authentication Server（AS）：认证服务器，认证客户端身份，发放客户访问 TGS 的票据授权票据（TGT）。
  - Ticket Granting Server（TGS）：票据授予服务器，发放应用客户端访问应用服务端所需的服务票据（ST）。

步骤原理说明

应用客户端（App Client）可以是集群内某个服务，也可以是客户二次开发的一个应用程序，应用程序可以向应用服务提交任务或者作业。

1. AS_REQ：App Client 在提交任务或者作业前，需要向 AS 申请 TGT，用于建立和 TGS 的安全会话。
2. AS_REP：AS 在收到 TGT 请求后，会解析其中的参数来生成对应的 TGT，使用 App Client 指定的用户名的密钥进行加密响应消息。

3. TGS_REQ：App Client 收到 TGT 响应消息后，解析获取 TGT，此时，再由 App Client（通常是 RPC 底层）向 TGS 获取应用服务端的 ST。

4. TGS_REP：TGS 在收到 ST 请求后，校验其中的 TGT 合法后，生成对应的 App Server 的 ST，再使用 App Server 密钥将响应消息进行加密处理。

5. AP_REQ：App Client 收到 ST 响应消息后，将 ST 打包到发给 App Server 的消息里面传输给对应的 App Server。

6. AP_REP：App Server 端收到请求后，使用 App Server 对应的密钥解析其中的 ST，并校验成功后，本次请求合法通过。

✧ **说明**

1. Kerberos 认证时需要配置 Kerberos 认证所需要的文件参数，主要包含 keytab 路径，Kerberos 认证的 principal，Kerberos 认证所需要的客户端配置 krb5.conf 文件。

2. 方法 login()为调用的 UserGroupInformation 的方法执行 Kerberos 认证，生成 TGT 票据。

3. 方法 doSth()调用 hadoop 的接口访问文件系统，此时底层 RPC 会自动携带 TGT 去 Kerberos 认证，生成 ST 票据。

基本概念

- 票据授权票据（Ticket-Granting Ticket，TGT）：由 AS 生成，提供给应用程序与 TGS 建立认证安全会话，该票据的默认有效期为 24 小时，24 小时后该票据自动过期。
- 服务票据（Server Ticket，ST）：由 TGS 生成，提供给应用程序与应用服务建立安全会话，该票据一次性有效。
- 用户身份（Principal）：用于标识用户或者服务，格式一般为<用户名或服务名>/<主机名>，例如 hdfs/host1，user/host2。

样例

以 HDFS 为例，前提：HDFS 集群已经开启 Kerberos 认证。

通过 UserGroupInformation

```
package bigdata.hdfs.examples;

import java.io.IOException;

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.FileStatus; import org.apache.hadoop.fs.FileSyste

m;

import org.apache.hadoop.fs.Path;
```

```
import org.apache.hadoop.security.UserGroupInformation;

public class KerberosTest {

private FileSystem fs;

private Configuration conf;

/**

* initialize Configuration

*/

private void initConf() {

conf = new Configuration();

//add configuration files

//PATH_TO_HDFS_SITE_XML 是 hdfs-site.xml 的路径

//PATH_TO_CORE_SITE_XML 是 core-site.xml 的路径

conf.addResource(new Path(PATH_TO_HDFS_SITE_XML));

conf.addResource(new Path(PATH_TO_CORE_SITE_XML));

}


/**

* login Kerberos to get TGT, if the cluster is in security mode

* @throws IOException if login is failed

*/


private void login() throws IOException {

// not security mode, just return
```

```
if (! "kerberos".equalsIgnoreCase(conf.get("hadoop.security.authentication"))) {
return;
}

//security mode
//PATH_TO_KRB5_CONF 是 krb5.conf 的路径
System.setProperty("java.security.krb5.conf", PATH_TO_KRB5_CONF);

UserGroupInformation.setConfiguration(conf);

//PATH_TO_KEYTAB 是 keytab 的路径
//PRNCIPAL_NAME 是 pincipal 名称

UserGroupInformation.loginUserFromKeytab(PRNCIPAL_NAME, PATH_TO_KEYT
AB);
}

/**
* initialize FileSystem, and get ST from Kerberos
* @throws IOException
*/

private void initFileSystem() throws IOException {
```

```
fs = FileSystem.get(conf);
}

/**
* An example to access the HDFS
* @throws IOException
*/

private void doSth() throws IOException {
Path path = new Path("/tmp");
FileStatus fStatus = fs.getFileStatus(path);
System.out.println("Status of " + path + " is " + fStatus);
//other thing
}

public static void main(String[] args) throws Exception {
KerberosTest test = new KerberosTest();
test.initConf();
test.login();
test.initFileSystem();
test.doSth();
}
}
```

通过 kinit 命令，执行 hadoop 命令

```
kinit -kt /etc/security/keytabs/hdfs.keytab hdfs/host1

hadoop fs -ls /tmp
```

4.8 高危操作一览表

禁用操作

下表中描述了在集群操作与维护阶段，进行日常操作时应注意的禁用操作。

类别	操作风险
严禁删除 ZooKeeper 相关数据目录	HDFS/Yarn/HBase/Hive 等很多组件都依赖于 ZooKeeper，在 ZooKeeper 中保存元数据信息。删除 ZooKeeper 中相关数据目录将会影响相关组件的正常运行。
严禁手动修改 Hive 元数据库的数据（Hivemeta 数据库）	修改 Hive 元数据可能会导致 Hive 数据解析错误，Hive 无法正常提供服务。
禁止对 Hive 的元数据表手动进行 insert 和 update 操作	修改 Hive 元数据可能会导致 Hive 数据解析错误，Hive 无法正常提供服务。
严禁修改 Hive 私有文件目录的权限	修改该目录权限可能会导致 Hive 服务不可用。
严禁修改 Kafka 配置文件中 broker.id	修改 Kafka 配置文件中 broker.id 将会导致该节点数据失效。
严禁修改节点主机名	主机名修改后会导致该主机上相关实例和上层组件无法正常提供服务，且无法修复。
禁止重装节点 OS	该操作会导致翼 MR 集群进入异常状态，影响翼 MR 集群使用。
禁止使用私有镜像	该操作会导致翼 MR 集群进入异常状态，影响翼 MR 集群使用。

高危操作

以下各表分别描述了各组件在操作与维护阶段，进行日常操作时应注意的高危操作。

集群高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
绑定弹性公网IP	该操作会将集群服务的相关主节点如:HDFS NameNode 所在的 master 节点暴露在公网，会增大来自互联网的网络攻击风险可能性。	★★★★★	请确认绑定的弹性公网 IP 为可信任的公网访问 IP，确认是否针对开放的相关端口进行设置安全组规则，只允许可信的 IP 可以访问该端口，入方向规则不推荐设置允许 0.0.0.0 可以访问。	无
开放集群22端口安全组规则	该操作会增大用户利用 22 进行漏洞攻击的风险。	★★★★★	针对开放的 22 端口进行设置安全组规则，只允许可信的 IP 可以访问该端口，入方向规则不推荐设置允许 0.0.0.0 可以访问。	无
删除集群或删除集群数据	该操作会导致数据丢失。	★★★★★	删除前请务必再次确认该操作的必要性，同时要保证数据已完成备份。	无
缩容集群	该操作会导致数据丢失。	★★★★★	缩容前请务必再次确认该操作的必要性，同时要保证数据已完成备份。	无
卸载磁盘或格式化数据盘	该操作会导致数据丢失。	★★★★★	操作前请务必再次确认该操作的必要性，同时要保证数据已完成备份。	无

翼 MR Manager 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改日志级别	如果修改为 DEBUG，会导致 Manager 运行速度明显降低。	★★	修改前确认操作的必要性，并及时修改回默认设定。	无
重启下层服务时，如果勾选同时重启上层服务	该操作会导致上层服务业务中断，影响集群的管理维护和业务。	★★★★	操作前确认操作的必要性，操作时确保同一时间无其它管理维护操作。	观察是否有未恢复的告警产生，观察集群的管理维护是否正常，业务是否正常。
修改 OLDAP 端口	修改该参数时，会重启 LdapServer 和 Kerberos 服务和其关联的所有服务，会影响业务运行。	★★★★★	操作前确认操作的必要性，操作时确保同一时间无其它管理维护操作。	无
重启服务	重启过程中会中断服务，勾选同时重启上层服务会导致依赖该服务的上层服务中断。	★★★	操作前确认重启的必要性。	观察是否有未恢复的告警产生，观察集群的管理维护是否正常，业务是否正常。
修改节点 SSH 默认端口	修改默认端口（22）将导致创建集群、添加服务/实例、添加主机、重装主机等功能无法正常使用，并且会导致集群健康检查结果中节点互信等检查项不准确。	★★★	执行相关操作前将 SSH 端口改回默认值。	无

Flink 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改日	如果修改为 DEBUG，会影	★★	修改前确认操作的必要	无。

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
志级别	响任务运行性能。		性，并及时修改回默认设定。	
修改文件权限	该操作可能导致任务运行失败。	★★★	修改前确认操作的必要性。	观察相关业务操作是否正常。

HBase 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改加密的相关配置项：hbase.regionserver.wal.encryption、hbase.crypto.keyprovider.parameters.uri、hbase.crypto.keyprovider.parameters.encryptedtext	导致服务启动异常。	★★★★	修改相关配置项时请严格按照提示描述，加密相关配置项是有关联的，确保修改后的值有效。	观察服务能否正常启动。
已使用加密的情况下关闭或者切换加密算法，关闭主要指修改 hbase.regionserver.wal.encryption 为 false，切换主要指 AES 和 SMS4 的切换	导致服务启动失败，数据丢失。	★★★★	加密 HFile 和 WAL 内容的时候，如果已经使用一种加密算法加密并且已经建表，请不要随意关闭或者切换加密算法。未建加密表（ENCRYPTION=>AES/SMS4）的情况下可以切换，否则禁止操作。	无。
修改 HBase 实例的启动参数 GC_OPTS、HBASE_HEAPSIZE	导致服务启动异常。	★★	修改相关配置项时请严格按照提示描述，确保修改后的值有效，且 GC_OPTS 与 HBASE_HEAPSIZE 参数值无冲突。	观察服务能否正常启动。
使用 OfflineMetaRepair 工具	导致服务启动异常。	★★★★	必须在 HBase 下线的情况下才可以使用该命令，而且不能在数据迁移的场景中使用该命令。	观察 HBase 服务是否可以正常启动。

HDFS 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改 HDFS 的 NameNode 的数据存储目录 dfs.namenode.name.dir、DataNode 的数据配置目录 dfs.datanode.data.dir	导致服务启动异常。	★★★★★	修改相关配置项时请严格按照提示描述，确保修改后的值有效。	观察服务能否正常启动。
执行 hadoop distcp 命令时，使用-delete 参数	Distcp 拷贝时，源集群没有而目的集群存在的文件，会在目的集群删除。	★★	在使用 Distcp 的时候，确保是否保留目的集群多余的文件，谨慎使用-delete 参数。	Distcp 数据拷贝后，查看目的的数据是否按照参数配置保留或删除。
修改 HDFS 实例的启动参数 GC_OPTS、HADOOP_HEAPSIZE 和 GC_PROFILE	导致服务启动异常	★★	修改相关配置项时请严格按照提示描述，确保修改后的值有效，且 GC_OPTS 与 HADOOP_HEAPSIZE 参数值无冲突。	观察服务能否正常启动。
修改 HDFS 的副本数目 dfs.replication，将默认值由 3 改为 1	1. 存储可靠性下降，磁盘故障时，会发生数据丢失。 2. NameNode 重启失败，HDFS 服务不可用。	★★★★	修改相关配置项时，请仔细查看参数说明。保证数据存储的副本数不低于 2。	观察默认的副本值是否不为 1，HDFS 服务是否可以正常提供服务。
修改 Hadoop 中各模块的 RPC 通道的加密方式 hadoop.rpc.protection	导致服务故障及业务异常。	★★★★★	修改相关配置项时请严格按照提示描述，确保修改后的值有效。	观察 HDFS 及其他依赖 HDFS 的服务能否正常启动，并提供服务。

Hive 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改 Hive 实例的启动参数 GC_OPTS	修改该参数可能会导致 Hive 实例无法启动。	★★	修改相关配置项时请严格按照提示描述，确保修改后的值有效。	观察服务能否正常启动。
删除 MetaStore 所有实例	Hive 元数据丢失，Hive 无法提供服务。	★★★	除非确定丢弃 Hive 所有表信息，否则不要执行该操作。	观察服务能否正常启动。
使用 HDFS 文件系统接口或者 HBase 接口删除或修改 Hive 表对应的文件	该操作会导致 Hive 业务数据丢失或被篡改。	★★	除非确定丢弃这些数据，或者确保该修改操作符合业务需求，否则不要执行该操作。	观察 Hive 数据是否完整。
使用 HDFS 文件系统接口或者 HBase 接口修改 Hive 表对应的文件或目录访问权限	该操作可能会导致相关业务场景不可用。	★★★	请勿执行该操作。	观察相关业务操作是否正常。
导出表数据覆盖写入本地目录，例如将 t1 表中数据导出，覆盖到“/opt/dir”路径下：**insert overwrite local directory '/opt/dir' select * from t1**	该操作会删除目标目录，如果设置错误，会导致软件或者操作系统无法启动。	★★★★★	确认需要写入的路径下不要包含任何文件；或者不要使用 overwrite 关键字。	观察目标路径是否有文件丢失。
将不同的数据库、表或分区文件指定至相同路径，例如默认仓库路径“/user/hive/warehouse”	执行创建操作后数据可能会紊乱，如果删除其中一个数据库、表或分区，会导致其他对象数据丢失。	★★★★★	请勿执行该操作。	观察目标路径是否有文件丢失。

Kafka 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
删除 Topic	该操作将会删除已有的主题和数据。	★★★	采用 Kerberos 认证，保证合法用户具有操作权限，并确保主题名称正确。	观察主题是否正常处理。
删除数据目录	该操作将会导致业务信息丢失。	★★★	请勿手动删除数据目录。	观察数据目录是否正常。
修改数据目录下内容（创建文件、文件夹）	该操作将会导致该节点上的 Broker 实例故障。	★★★	请勿手动在数据目录下创建或修改文件及文件夹。	观察数据目录是否正常。
修改磁盘自适应功能“disk.adapter.enable”参数	该操作会在磁盘使用空间达到阈值时调整 Topic 数据保存周期，超出保存周期的历史数据可能被清除。	★★★	若个别 Topic 不能做保存周期调整，将该 Topic 配置在“disk.adapter.topic.blacklist”参数中。	在 KafkaTopic 监控页面观察数据的存储周期。
修改数据目录“log.dirs”配置	该配置不正确将会导致进程故障。	★★★	确保所修改或者添加的数据目录为空目录，且权限正确。	观察数据目录是否正常。
减容 Kafka 集群	该操作将会导致部分 Topic 数据副本数量减少，可能会导致 Topic 无法访问。	★★	请先做好数据副本转移工作，然后再进行减容操作。	观察分区所在备份节点是否都存活，确保数据安全。
单独启停基础组件	该操作将会影响服务的一些基础功能导致业务失败。	★★★	请勿单独启停 ZooKeeper/Kerberos/LDAP 等基础组件启停基础组件请勾选关联服务。	观察服务状态是否正常。

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
重启/停止服务	该操作将会导致业务中断。	★★	确保在必要时重启/停止服务。	观察服务是否运行正常。
修改配置参数	该操作将需要重启服务使得配置生效。	★★	确保在必要时修改配置。	观察服务是否运行正常。
删除/修改元数据	修改或者删除 ZooKeeper 上 Kafka 的元数据可能导致 Topic 或者 Kafka 服务不可用。	★★★	请勿删除或者修改 Kafka 在 ZooKeeper 上保存的元数据信息。	观察 Topic 或者 Kafka 服务是否可用。
修改元数据备份文件	修改 Kafka 元数据备份文件，并被使用进行 Kafka 元数据恢复成功后，可能导致 Topic 或者 Kafka 服务不可用。	★★★	请勿修改 Kafka 元数据备份文件。	观察 Topic 或者 Kafka 服务是否可用。

KrbServer 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改 KrbServer 的参数 KADMIN_PORT	修改该参数后，若没有及时重启 KrbServer 服务和其关联的所有服务，会导致集群内部 KrbClient 的配置参数异常，影响业务运行。	★★★★★	修改该参数后，请重启 KrbServer 服务和其关联的所有服务。	无。
修改 KrbServer 的参数 kdc_ports	修改该参数后，若没有及时重启 KrbServer 服务和其关联的所有服务，会导致集群内部	★★★★★	修改该参数后，请重启 KrbServer 服务和其关联	无。

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
	KrbClient 的配置参数异常，影响业务运行。		的所有服务。	
修改 KrbServer 的参数 KPASSWD_PORT	修改该参数后，若没有及时重启 KrbServer 服务和其关联的所有服务，会导致集群内部 KrbClient 的配置参数异常，影响业务运行。	★★★★★	修改该参数后，请重启 KrbServer 服务和其关联的所有服务。	无。
修改 Manager 系统域名	若没有及时重启 KrbServer 服务和其关联的所有服务，会导致集群内部 KrbClient 的配置参数异常，影响业务运行。	★★★★★	修改该参数后，请重启 KrbServer 服务和其关联的所有服务。	无。
配置跨集群互信	该操作会重启 KrbServer 服务和其关联的所有服务，影响集群的管理维护和业务。	★★★★★	更换前确认操作的必要性，更换时确保同一时间无其它管理维护操作。	观察是否有未恢复的告警产生，观察集群的管理维护是否正常，业务是否正常。

LdapServer 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
修改 LdapServer 的参数 LDAP_SERVER_PORT	修改该参数后，若没有及时重启 LdapServer 服务和其关联的所有服务，会导致集群内部 LdapClient 的配置参数异常，影响业务运行。	★★★★★	修改该参数后，请重启 LdapServer 服务和其关联的所有服务。	无。
恢复 LdapServer 数据	该操作会重启 Manager 和整个集群，影响集群的管理维护和业务。	★★★★★	修改前确认操作的必要性，修改时确保同一时间无其它管	观察是否有未恢复的告警产生，观察集群的管

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
			理维护操作。	理维护是否正常，业务是否正常。
更换 LdapServer 所在节点	该操作会导致部署在该节点上的服务中断，且当该节点为管理节点时，更换节点会导致重启 OMS 各进程，影响集群的管理维护。	★★★	更换前确认操作的必要性，更换时确保同一时间无其它管理维护操作。	观察是否有未恢复的告警产生，观察集群的管理维护是否正常，业务是否正常。
修改 LdapServer 密码	修改密码需要重启 LdapServer 和 Kerberos 服务，影响集群的管理维护和业务。	★★★★	修改前确认操作的必要性，修改时确保同一时间无其它管理维护操作。	无。
节点重启导致 LdapServer 数据损坏	如果未停止 LdapServer 服务，直接重启 LdapServer 所在节点，可能导致 LdapServer 数据损坏。	★★★★★	使用 LdapServer 备份数据进行恢复。	无。

YARN 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
删除或者修改数据目录 yarn.nodemanager.local-dirs 和 yarn.nodemanager.log-dirs	该操作将会导致业务信息丢失。	★★★	请勿手动删除数据目录。	观察数据目录是否正常。

ZooKeeper 高危操作

操作名称	操作风险	风险等级	规避措施	重大操作观察项目
删除或者修改 ZooKeeper 的数据目录	该操作将会导致业务信息丢失。	★★★	修改 ZooKeeper 目录时候，严格按照扩容指导操作。	观察服务能否正常启动，关联组件能否正常启动。
修改 ZooKeeper 实例的启动参数 GC_OPTS	导致服务启动异常。	★★	修改相关配置项时请严格按照提示描述，确保修改后的值有效。	观察服务能否正常启动。
设置 ZooKeeper 中 znode 的 ACL 信息	修改 ZooKeeper 中 znode 的权限，可能会导致其他用户无权限访问该 znode，导致系统功能异常。	★★★★	修改相关配置项时请严格按照“ZooKeeper 权限设置指南”章节操作，确保修改 ACL 信息，不会影响其他组件正常使用 ZooKeeper。	观察项目观察其他依赖 ZooKeeper 的组件能否正常启动，并提供服务。

4.9 服务配置修改建议

4.9.1 配置升级-服务配置修改建议

操作场景

当节点组内 ECS 实例的规格（vCPU 和内存）无法满足您的业务需求时，您可以使用配置升级功能提升 ECS 实例规格。升级配置后需要手动修改 HDFS、YARN 和 Spark 等服务的配置信息。本文为您介绍如何在翼 MR Manager 的“配置管理”页面修改配置项。

前提条件

已创建集群。

操作步骤

1. 在翼 MR Manager 中，单击“运维与配置”。
2. 单击“配置管理”。
3. 选择“所选集群服务”，点击查询，即可在当前页面修改配置信息。

配置修改建议说明

1. Doris：Doris 所在节点配置升级后，Doris FE 建议使用节点一半内存。
2. Elasticsearch：一个 Elasticsearch 节点，内存建议不超过 64G。Elasticsearch 所在节点配置升级后，Elasticsearch 会自动根据节点情况设置内存值，一般无须用户手动修改，但需要重启集群。如果想手动设置，可以修改 jvm.options 文件配置-Xms30g、-Xmx30g 参数，并重启集群。
3. HBase：HBase 所在节点配置升级后，修改建议如下：
   - hbase-site.xml：hbase.regionserver.handler.count **说明**：一般跟 CPU 核数相同。
   - hbase-env.sh：export HBASE_MASTER_OPTS："-Xmsg -Xmxg" **说明**：master 不消耗很多内存，一般默认不添加或者分配 2~8G 左右。
   - hbase-env.sh：export HBASE_REGIONSERVER_OPTS："-Xmsg -Xmxg " **说明**：regionserver 需要较多内存，一般配置内存配额的一半或更多。
4. HDFS：HDFS 所在节点配置升级后，可以根据 hadoop-env.sh 参数配置进行，通过调整服务的内存大小调整服务的性能，如-Xmx20g -Xms20g -Xmn4g，然后重启服务。NameNode 建议文件、目录、数据块之和 1 亿，配置 50G。
5. Hive：Hive 所在节点配置升级后，可以通过 hive-env.sh 统一参数配置来进行，也可以在作业提交时使用额外参数指定来进行。通过调整服务的内存大小调整服务的性能，如-Xmx20g -Xms20g -Xmn4g，然后重启服务。内存大小可以根据机器的总内存而定，建议初始值为总内存大小的 10%，后续根据性能需求调整。
6. Kafka：Kafka 所在节点配置升级后，建议配置如下：
   - kafka-env.sh 设置 jvm 配置参数：调整 jvm 堆大小，通过调整参数：export KAFKA_HEAP_OPTS="-Xmx20G -Xms20G -Xmn4g"设置堆大小。
   - server.properties 文件建议修改的配置项：
     - num.io.threads：修改写磁盘的线程数，建议配置为 CPU 核数的 50%；
     - num.replica.fetchers：修改副本拉取线程数，建议配置为 CPU 核数 50%的 1/3；
     - num.network.threads：修改数据传输线程数，建议配置为 CPU 核数的 50%的 2/3；
     - replica.fetch.max.bytes：副本拉取数据量的大小。内存增加，可以适当加大该值；
     - socket.send.buffer.bytes：调整 socket 发送的数据量。内存增加，可以适当加大该值；
     - socket.receive.buffer.bytes：调整 socke 接受的数据量。内存增加，可以适当加大该值；
     - socket.request.max.bytes：socket 请求的数据量。内存增加，可以适当加大该值。

7. Kerberos：建议保持默认值，无需修改配置。

8. Kibana：Kibana 是一个基于 NodeJS 的单页 web 应用，一般情况下，对内存 CPU 占用很少，无须修改内存、CPU 等配置。

9. Kyuubi：Kyuubi 一般情况下，对内存 CPU 占用很少，无须修改内存、CPU 等配置。

10. OpenLDAP：建议保持默认值，无需修改配置。

11. Ranger：Ranger 所在节点配置升级后，修改建议如下：

    - ranger-admin 通过 {installdir}/ews/ranger-admin-services.sh 中变量 ranger_admin_max_heap_size 的值修改 JMX，JAVA_OPTS 修改 Xmx、Xmn 等 JVM 参数，一般设置 1-8g，1K policy 建议设置为 1G，1W policy 建议设置为 8G。

    - ranger-usersync 通过 /{installdir}/ranger-usersync-services.sh 中变量 ranger_usersync_max_heap_size 的值修改 JMX，JAVA_OPTS 修改 Xmx Xmn 等 JVM 参数，一般设置 1-8g，1K policy 建议设置为 1G，1W policy 建议设置为 8G。

12. Spark：Spark 所在节点配置升级后，修改建议如下：

    - spark.history.kerberos.principal 和 spark.history.kerberos.keytab 为 spark 读写 eventLog 的租户，用户如有特殊需求自行更改。

    - spark.yarn.historyServer.address=：说明了 history server 的地址，用户如有特殊需求自行更改。

    - spark.dynamicAllocation.enabled 和 spark.dynamicAllocation.maxExecutors 分别控制动态和动态开启下能使用的最大资源，用户如有特殊需求自行更改。

    - spark.executor.cores 和 spark.executor.memory 确保 spark.executor 的每一个 core 分配到 2~4g 内存，标准是 4g，具体视情况而定，设置 core 的 memory 设置过小 executor 容易 oom。

13. Trino：Trino 的服务包括 coordinator 和 worker。Trino 所在节点配置升级后，可以根据 jvm.config 参数配置进行，通过调整服务的内存大小调整服务的性能，如 -Xmx128g -Xms128g，然后重启服务。

14. YARN：YARN 所在节点配置升级后，可以根据 yarn-env.sh 参数配置进行，通过调整服务的内存大小来调整服务的性能，如-Xmx20g -Xms20g -Xmn4g，然后重启服务。NM 用于集群中作业的内存和 CPU，需要修改 NM 节点的 yarn-site.xml 中的 yarn.nodemanager.resource.memory-mb 的值，该值用于所有作业的最多可用内存；以及 yarn.nodemanager.resource.cpu-vcores 的值，该值用于所有作业的最多可用虚拟 CPU 核数。

15. ZooKeeper：ZooKeeper 所在节点配置升级后，可通过配置 java.env 文件，在其中添加：export ZK_SERVER_HEAP=2048（这里设置的单位默认是 MB）。

4.9.2 多数据盘-服务配置修改建议

操作场景

创建翼 MapReduce 集群、新增节点组以及扩容节点时，用户可以为节点配置多块数据盘，但组件配置中默认是 1 块。如需使用多块数据盘，需要用户将配置目录中的信息改为多块，详细操作方法请参见下述信息。说明：该能力适用于 2.15.2～2.18.0 翼 MapReduce 版本。

操作步骤

1、 默认的配置文件

HDFS

配置文件： hdfs-site.xml

```
<property>
<name>dfs.namenode.name.dir</name>
<value>/data01/hadoop/hdfs/namenode</value>
<description>namenode 元数据存放位置</description>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/data01/hdfs</value>
<description>设置 datanode 节点存储数据文件的本地路径</description>
</property>
<property>
<name>dfs.datanode.failed.volumes.tolerated</name>
<value>0</value>
<description>决定停止数据节点提供服务允许卷的出错次数，0 则表示任务卷出错都要停止数据
节点</description>
</property>
<property>
<name>dfs.journalnode.edits.dir</name>
```

<value>/data01/hadoop/hdfs/journal</value>

<description>存储 journalnode edit 文件的目录</description>

</property>

YARN

配置文件： yarn-site.xml

<property>

<name>yarn.nodemanager.log-dirs</name>

<value>/data01/hadoop/yarn/log</value>

<description>Nodemanager 本地数据盘日志存储目录，使用逗号分隔，一般个数与

{nm_local-dirs}对应</description>

</property>

<property>

<name>yarn.nodemanager.local-dirs</name>

<value>/data01/hadoop/yarn/local</value>

<description>Nodemanager 本地数据盘存储目录，使用逗号分隔</description>

</property>

配置文件： mapred-site.xml

<property>

<name>mapreduce.jobhistory.recovery.store.leveldb.path</name>

<value>/data01/hadoop/mapreduce/jhs</value>

<description>Jobhistory 恢复信息本地存储路径</description>

</property>

Kafka

配置文件： Kafka_vars.yaml

log_dirs=/data01/kafka/data

配置文件： server.properties

log.dirs=/data01/kafka/data

2、如何修改配置文件

如果用户在下单阶段选择了多块数据盘，需要如何修改配置？

a. 下单选择的数据盘，开通集群时，数据盘会自动挂载，挂载后的目录按照/data01 /data02 /data03...，依次类推。

b. 修改配置：各个组件使用的目录，需要用户在配置管理中，手动修改，并同步配置，然后重启相关的服务，具体的配置见下方。

c. 创建目录机器上的目录：目前需要用户登录到机器上，手动执行目录的创建命令，并修改成对应的权限。

d. 重启服务：配置，机器上的目录都修改之后，需要重启相关服务，让配置生效。请选择合适的时间进行服务的重启。

假如选择了 2 块数据盘。可以根据下面的方式进行修改组件的配置文件，并在对应的机器上，创建对应的目录.。

HDFS

配置文件： hdfs-site.xml

影响的配置项：

- DataNode 使用的数据盘： dfs.datanode.data.dir 和 fs.datanode.failed.volumes.tolerated 。
- dfs.datanode.data.dir 配置成磁盘的个数，中间使用分号;分割，如 2 块盘则配置成 /data01/hdfs;/data02/hdfs；3 块盘，则配置成/data01/hdfs;/data02/hdfs;/data03/hdfs 。
- dfs.datanode.failed.volumes.tolerated 值 < 数据盘个数。

  当数据盘个数 = 1 时，则 dfs.datanode.failed.volumes.tolerated = 0；

  数据盘=2 时， dfs.datanode.failed.volumes.tolerated 可以配置成 0 或者 1 。

1）修改配置

在 Manager 页面，集群服务 -> HDFS 集群 -> 配置管理，找到 defualt 分组下的 hdfs-site.xml 文件；搜索配置项 dfs.datanode.data.dir 和 dfs.datanode.failed.volumes.tolerated，按下面的配置进行修改。

修改后保存配置文件，然后点击配置同步。

<property>

<name>dfs.datanode.data.dir</name>

<value>/data01/hdfs;/data02/hdfs</value>

<description>设置 datanode 节点存储数据文件的本地路径</description>

</property>

<property>

<name>dfs.datanode.failed.volumes.tolerated</name>

<value>1</value>

<description>决定停止数据节点提供服务允许卷的出错次数，0 则表示任务卷出错都要停止数据

节点</description>

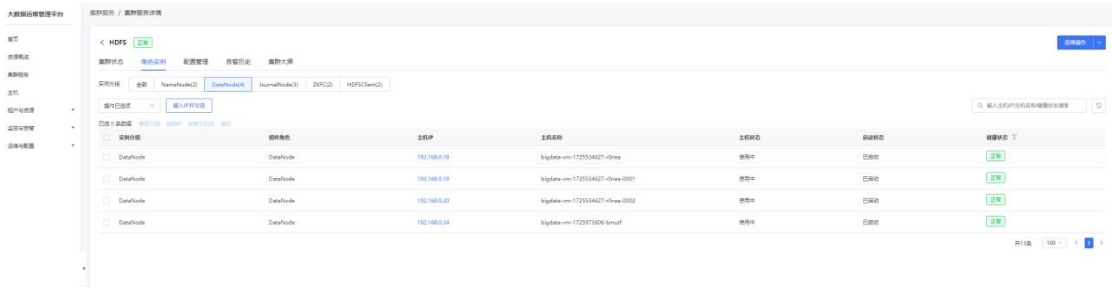
</property>

2）创建目录机器上的目录

```
# 登录每台 HDFS-DataNode 的机器
# 创建目录，并修改权限
sudo mkdir -p /data02/hdfs
sudo chown hdfs:hadoop /data02/hdfs
sudo chmod 700 /data02/hdfs
```

3）重启服务

在 Manager 页面，进入服务管理，选择 HDFS 集群，DataNode，重启 DataNode 服务。请选择合适的时间进行服务的重启。



YARN

配置文件： yarn-site.xml

影响的配置项：

- NodeManager 使用的数据盘： yarn.nodemanager.log-dirs 和 yarn.nodemanager.local-Dirs。
- 配置成磁盘的个数，如有 2 块数据盘， yarn.nodemanager.log-dirs 配置成 /data01/hadoop/yarn/log;/data02/hadoop/yarn/log 。

1）修改配置

在 Manager 页面，集群服务 -> 配置管理，修改 yarn-site.xml 文件，按上述配置修改。

2）创建目录机器上的目录

```
# 登录每台 YARN-NodeManager 的机器
# 创建目录，并修改权限
sudo mkdir -p /data02/hadoop/yarn/log
sudo mkdir -p /data02/hadoop/yarn/local
sudo chown yarn:hadoop /data02/hadoop/yarn/log
sudo chown yarn:hadoop /data02/hadoop/yarn/local
sudo chmod 755 /data02/hadoop/yarn/log
sudo chmod 755 /data02/hadoop/yarn/local
```

3）重启服务

在 Manager 页面，进入服务管理，选择 YARN 集群，NodeManager，重启 NodeManager 服务。请选择合适的时间进行服务的重启。

Kafka

配置文件： Kafka_vars.yaml

修改的配置项：

- log_dirs ，配置成磁盘的个数，如有 2 块数据盘，log_dirs 配置成 /data01/kafka/data;/data02/kafka/data。

配置文件： server.properties

影响的配置项：

- log.dirs ，配置成磁盘的个数，如有 2 块数据盘， log.dirs 配置成 /data01/kafka/data;/data02/kafka/data

log.dirs=/data01/kafka/data;/data02/kafka/data

1）修改配置

在 Manager 页面，集群服务->配置管理，修改 Kafka_vars.yaml 和 server.properties 文件，按上述配置修改。

2）创建目录机器上的目录

```
# 登录每台 Kafka-BrokerServer 的机器
# 创建目录，并修改权限
sudo mkdir -p /data02/kafka/data
sudo chown kafka:kafka /data02/kafka/data
```

3）重启服务

在 Manager 页面，进入服务管理，选择 Kafka 集群，BrokerServer，重启 BrokerServer 服务。请选择合适的时间进行服务的重启。

4.9.3 Master 扩容—服务配置修改建议

4.9.3.1 Zookeeper 扩容

前置条件

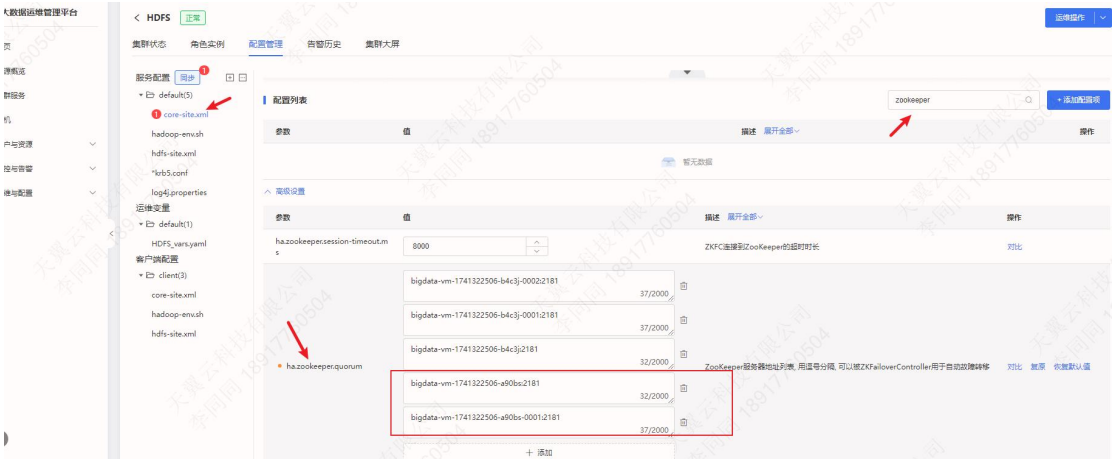
1. ZooKeeper 扩容完成，从 3 节点扩容到 5 节点。
2. ZooKeeper 扩容之后，ZooKeeper 集群正常。

相关服务修改建议

01 HDFS

01 修改配置

1. 前往 Manager 页面，集群服务—> HDFS 集群—> 配置管理—> default 配置组—> 配置文件 core—site.xml —> 配置项 ha.zookeeper.quorum。
2. 将新的 ZK 节点，添加到配置项 ha.zookeeper.quorum 中。



02 配置同步

点击配置同步，推送配置。

02 YARN

01 修改配置

1. 前往 Manager 页面，集群服务—> YARN 集群—> 配置管理—> default 配置组—> 配置文件 yarn—site.xml —> 配置项 hadoop.zk.address。
2. 将新的 ZK 节点，添加到配置项 hadoop.zk.address 中。
3. 相同操作：同时修改 client 配置组的配置项。

02 配置同步

点击配置同步，推送配置。

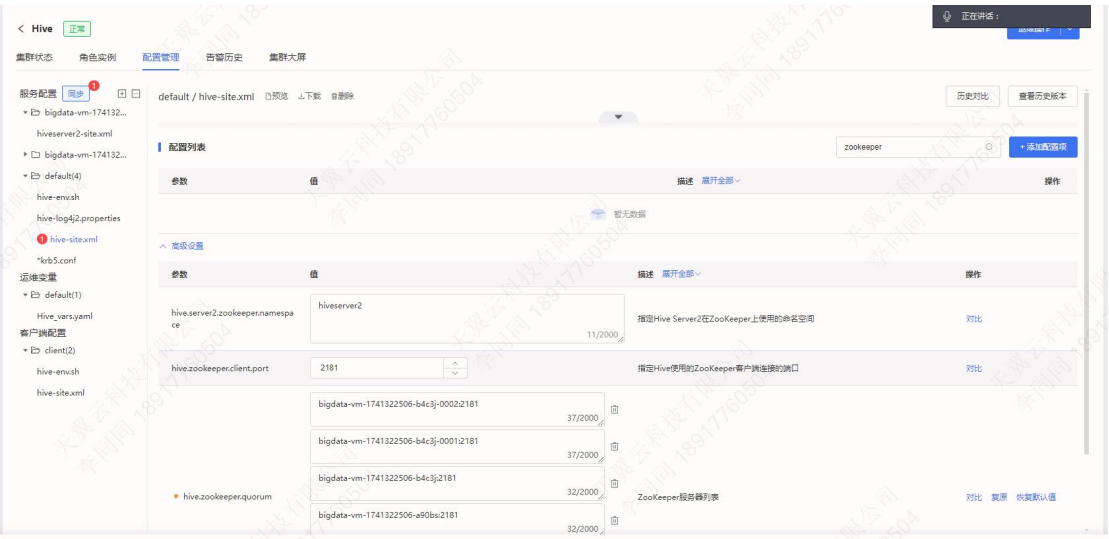
03 重启服务

需要重启 ResourceManager 服务，推荐滚动重启。

03 Hive

01 修改配置

1. 前往 Manager 页面，集群服务–> Hive 集群–> 配置管理–> default 配置组–> 配置文件 hive–site.xml –> 配置项 hive.zookeeper.quorum。
2. 将新的 ZK 节点，添加到配置项 hive.zookeeper.quorum 中。
3. 相同操作：同时修改 client 配置组的配置项。



02 配置同步

点击配置同步，推送配置。

03 重启服务

重启 HiveServer2 服务，推荐滚动重启。

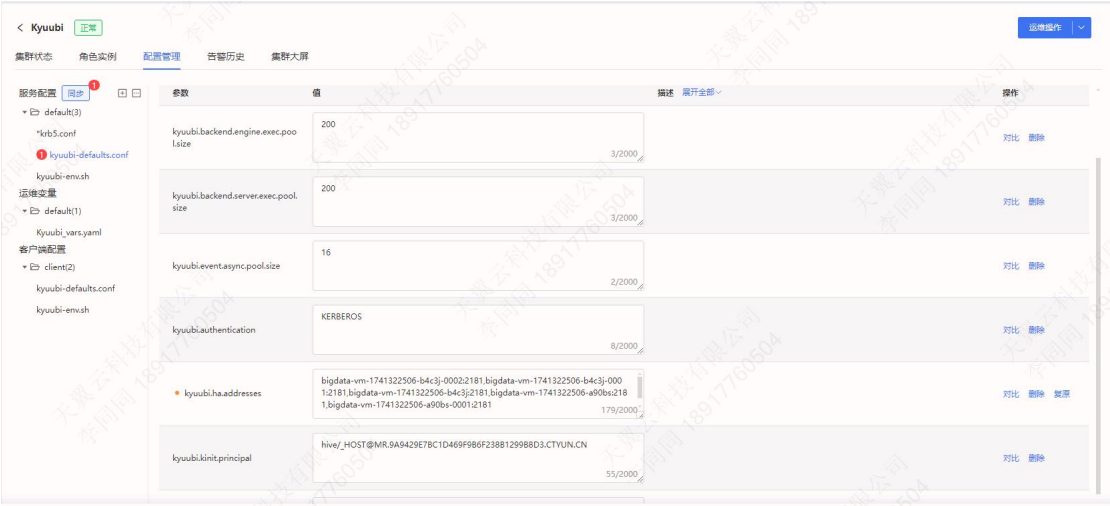
04 Flink

可直接提交到 YARN，不受 ZooKeeper 变化的影响。

05 Kyuubi

01 修改配置

1. 前往 Manager 页面，集群服务-> Kyuubi 集群-> 配置管理-> default 配置组-> 配置文件 kyuubi-defaults.conf -> 配置项 kyuubi.ha.addresses。
2. 将新的 ZK 节点，添加到配置项 kyuubi.ha.addresses 中。
3. 相同操作：同时修改 client 配置组的配置项。



02 配置同步

点击配置同步，推送配置。

03 重启服务

需要重启 Kyuubi 服务，推荐滚动重启。

06 HBase

01 修改配置

1. 前往 Manager 页面，集群服务-> HBase 集群-> 配置管理-> default 配置组-> 配置文件 hbase-site.xml -> 配置项 hbase.zookeeper.quorum。
2. 将新的 ZK 节点，添加到配置项 hbase.zookeeper.quorum 中。
3. 相同操作：同时修改 client 配置组的配置项。

02 配置同步

点击配置同步，推送配置。

03 重启服务

需要重启 HBase 所有服务，包含 HMaster 和 HRegionServer，推荐滚动重启。

07 Kafka

警告：Kafka 不建议重启，所以不建议修改 ZK 节点。

原因：Kafka 每个节点都是有状态的，每个节点的主分区都不一样，重启会导致业务受影响。

01 修改配置

1. 前往 Manager 页面，集群服务–> Kafka 集群–> 配置管理–> default 配置组–> 配置文件 server.properties –> 配置项 zookeeper.connect。
2. 将新的 ZK 节点，通过 vs 统一修改 3 个配置，添加到配置项 zookeeper.connect 中。

02 配置同步

点击配置同步，推送配置。

03 重启服务

需要重启 Kafka 所有服务。请谨慎进行 Kafka 集群的重启操作。

4.9.3.2 HDFS 扩容

前置条件

1. HDFS NameNode 完成扩容与配置生效操作。
2. 配置生效后重启 HDFS 服务。
3. 集群状态正常。

相关服务修改建议

01 YARN

前往翼 MR Manager，重启 ResourceManager，NodeManager，JobHistory，TimelineServer 实例。

02 Spark

前往翼 MR Manager，重启 SparkHistoryServer 实例。

03 Hive

前往翼 MR Manager，重启 HiveServer2、MetaStore 实例。

04 Kyuubi

前往翼 MR Manager，重启 Kyuubi 服务。

05 Flink

前往翼 MR Manager，重启 FlinkHistoryServer 实例。

06 HBase

前往翼 MR Manager，滚动重启 HBase 服务。

07 Trino

前往翼 MR Manager，重启 TrinoCoordinator 和 TrinoWorker 实例。

4.9.3.3 YARN 扩容

前置条件

1. YARN—ResourceManager 完成扩容与配置生效操作。
2. 集群状态正常。

相关服务修改建议

01 Spark

- 请前往翼 MR Manager 重启 SparkHistoryServer 所有的实例。

02 Flink

1. 如$FLINK_HOME/conf 目录下不存在 yarn-site.xml 文件，无需对 Flink 进行操作。
2. 如$FLINK_HOME/conf 目录下存在指向 yarn-site.xml 文件的软链，需确认该软链指向更新后的 yarn-site.xml 文件并且内容一致。
3. 如$FLINK_HOME/conf 目录下存在 yarn-site.xml 文件，需使用更新后的 yarn-site.xml 文件替换此文件。
4. 请前往翼 MR Manager 页面重启 FlinkHistoryServer 所有实例（需使用该文件作为依赖项或资源的客户 Flink 作业）。

4.9.3.4 Hive 扩容

前置条件

1. Hive MetaStore 完成扩容与配置生效操作。
2. 集群状态正常。

相关服务修改建议

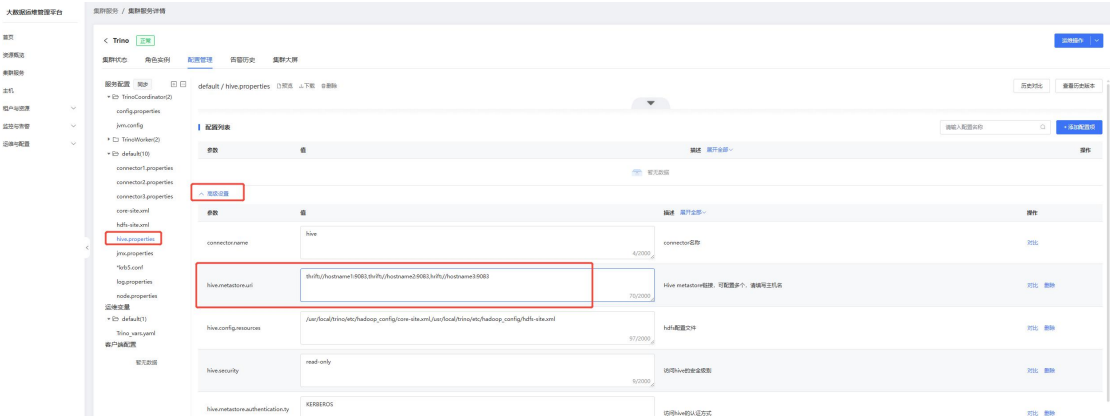
01 Flink

- 如果用户配置了数据湖作业，那么就需要登录到所有的 Flink client 主机，在 flink 配置目录/user/local/flink/conf/中建立对/user/local/hive/conf/hive-site.xml 的软链。
- 如果 Hive MetaStore 进行了节点扩容，需要更新该软链或文件，保证 Flink 使用到的是扩容后的 hive-site.xml 文件。
- 最后，根据扩容后的 Hive MetaStore 节点，更新数据湖作业中配置的 hive.uri，并重启作业。
- 如果用户没有配置数据湖作业，则 Flink 不需要做任何操作。

02 Trino

- 通过翼 MR Manager 进入 Trino 集群的“配置管理”页面，修改 hive.properties 文件。
- 在参数 hive.metastre.uri 中按照实际情况填写 metastore 的地址；例如，三台 Hive MetaStore 的主机名为 hostname1，hostname2，hostnam3，那么该配置的值应为 thrift://hostname1:9083，thrift://hostname2:9083，hrift://hostname3:9083。
- 保存更改后，需要进行配置“同步”操作。

- 最后，重启 Trino 集群服务。



5 最佳实践

5.1 Kibana 用户体验、性能优化

Kibana 是一个用于可视化和分析 ElasticSearch 中数据的强大工具。为了最大化利用 Kibana 的功能，提高用户体验和数据洞察力，以下是 Kibana 的一些最佳实践。

1. 仪表盘设计

明确的目标

在创建仪表盘之前，明确您的目标和需求。确定您希望从数据中获得什么样的见解，以便针对性地设计可视化组件。比如假设您是一个电子商务平台的数据分析师，您想要创建一个仪表盘来监控商品销售情况。您的目标是找出最畅销的商品类别和购买趋势。

简洁和重点突出

避免过于拥挤的仪表盘，尽量保持简洁。突出显示关键指标和信息，避免分散用户的注意力。避免在同一个仪表盘中混杂过多的信息，以确保用户专注于关键指标。

2. 可视化

合适的图表类型

根据数据的性质选择合适的图表类型，例如线形图、柱状图、饼图等。确保图表类型能够最有效地呈现数据。比如为了展示商品销售趋势，您可以选择使用折线图。这可以清晰地显示出不同类别的销售额随时间的变化。

轴标签和标题

为图表添加清晰的轴标签和标题，以便用户理解数据的含义。标签应该简洁明了，避免过于技术性的术语。比如在折线图上添加横轴和纵轴标签，分别标明时间和销售额。为图表添加标题，比如“商品销售趋势”。

3. 搜索和过滤

利用过滤器

使用 Kibana 的过滤器来快速筛选和过滤数据。为常用的过滤条件设置预定义的过滤器，方便用户快速获取特定数据。比如创建预定义的过滤器，如“电子产品”、“服装”和“家居用品”，以便用户可以轻松筛选特定商品类别的数据。

使用查询语言

熟悉 Kibana 查询语言，以便在搜索中精确获取所需数据。使用查询语言来构建复杂的搜索条件。比如如果您想进一步筛选销售额大于一定值的商品，您可以使用查询语言，如 sales_amount：＞1000。

4. 仪表盘布局

自由布局

充分利用 Kibana 的仪表盘自由布局功能，将可视化组件灵活地排列在仪表盘上，以满足特定的需求。

仪表盘间距

注意仪表盘组件之间的间距，以确保组件之间不会重叠或过于拥挤。

5. 安全性

访问控制

根据用户角色和权限设置访问控制，确保只有有权用户可以访问特定的仪表盘和数据。比如设置只有数据分析师可以访问销售数据仪表盘，而其他人员无法访问。

敏感数据处理

对于包含敏感数据的仪表盘，考虑将数据脱敏或使用字段级别的权限来限制敏感信息的访问。

6. 性能和优化

数据缓存

利用 Kibana 的数据缓存功能，减少重复查询对 ElasticSearch 的压力，提高仪表盘加载速度。

数据更新频率

根据数据的更新频率，设置仪表盘的刷新间隔，避免过于频繁的数据刷新。

这些是 Kibana 最佳实践的一部分，能够帮助您更好地设计、创建和使用仪表盘。但请注意，实际实践可能因特定需求和场景而有所不同。

6 API 参考

6.1 API 概览

翼 MapReduce 提供以下相关 API 接口。

集群

API 标题	API 概述
根据 id 查询集群信息	提供查询用户根据 id 查询集群信息功能，系统接收用户输入的查询条件，返回符合条件的集群详细信息。
集群信息分页查询	供查询用户集群信息功能，系统接收用户输入的查询条件，返回用户输入的资源池下符合条件的集群详细信息，不会返回已删除的集群信息。

节点组

API 标题	API 概述
查询集群节点组详情	提供查询用户集群节点组详情功能，系统接收用户输入的查询条件，返回符合条件的集群节点组信息。
获取节点组信息	提供查询用户集群节点组信息功能，系统接收用户输入的查询条件，返回符合条件的集群节点组详细信息。

6.2 如何调用 API

6.2.1 查询集群节点组详情

接口功能介绍

此接口提供查询用户集群节点组详情功能，系统接收用户输入的查询条件，返回符合条件的集群节点组信息。

接口约束

只允许查询处于运行中状态的集群的节点组详情。

接口是否审批

否

URI

POST /v1/emr/openapi/cluster/clusterNodeGroup/getGroupAndHostByCondition

Content-Type

application/json

路径参数

无

Query 参数

无

请求头 header 参数

无

请求体 body 参数

请求体 body 参数

参数	是否必填	参数类型	说明	示例	下级对象
clusterId	是	String	集群 id	6695c8e7c34af927e860ba1f043c8c91	/
nodeState	否	Integer	主机状态（1:启动中,2:正在关机,4:已删除,8:运行中,9:已关机）	1	/
selectKey	否	String	模糊查询参数，可以通过节点名称、内网 IP、外网 IP 进行模糊查询	10.0.02	/

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码	200	/
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	/
error	String	错误码，请求成功时，不返回该字段	EMR_400001	/
returnObj	Object	返回结果	/	returnObj

表 returnObj

参数	参数类型	说明	示例	下级对象
id	String	节点组 id	175187849787779	/
clusterId	String	集群 id	da595eb1d81503b323fdc01d9bf786b7	/
payType	Integer	付费类型	1	/
nodeGroupType	String	节点类型 code 码	MASTER	/
nodeGroupName	String	节点类型 code 值	master	/
imageId	String	镜像 id	f8d28855-ffda-426d-96e2-0f6bf46877e0	/
hostNum	Integer	主机数量	3	/
computeSpecificationId	Integer	主机规格 id	101	/
iaasVmSpecCode	String	IAAS 虚机规	s7.2xlarge.4	/

参数	参数类型	说明	示例	下级对象
		格编码		
cpuNum	Integer	cpu 核数集群列表	8	/
memory	Integer	内存大小	32	/
diskSpecificationList	String	磁盘规格列表	[{diskFunctionType:1,diskFunctionName:系统盘,diskType:cloud,diskTypeName:云硬盘,ioType:SSD-genric,ioTypeName:通用型SSD,volume:200,diskNum:1}]	/
createTime	Integer	创建时间	1706515356000	/
updateTime	Integer	更新时间	1706515356000	/
mountPublicIp	Boolean	是否挂载公网	false	/
iaasVmSpecId	String	IAAS 虚机规格 id	b307034d-cbc3-27bb-24f7-a97565814236	/
highest	Boolean	节点组规格是否最高的	false	/
regionId	String	资源池 id	bb9fdb42056f11eda1610242ac110002	/
availableZoneId	String	可用区 id	cn-huadong1-jsnj1A-public-ctcloud	/
vpcId	String	vpc id	vpc-fr2xjo1gj0	/
clusterHostDtoList	Array of Objects	集群节点组列表	/	clusterHostDtoList

表 clusterHostDtoList

参数	参数类型	说明	示例	下级对象
id	String	主键 id	0088babec7bab47b20409df6646db856	/
nodeGroupId	String	节点组 id	1753298494213554178	/
iaasHostId	String	平台主机 id	b2f2377d-3f0b-bc74-b9c5-67bb557b0e63	/
hostName	String	主机名称	bigdata-emr-vm-l7zb3wnk	/
manageIp	String	管理 IP	10.2.3.0	/
serviceIp	String	内网 IP	192.168.0.149	/
publicIp	String	公网 IP	212.168.0.150	/
ipv6ServiceIp	String	ipv6 内网 ip	240e:982:db0f:6c00:d1ea:8588:69fa:96	/
ipv6PublicIp	String	ipv6 内网 ip	121.237.177.234	/
state	Integer	机器状态枚举创建主机时机器状态非即时状态	8	/
vpcId	String	vpc id	vpc-fr2xjo1gj0	/
subnetId	String	子网 id	subnet-vkixrw8xw7	/
regionId	String	资源池 id	bb9fdb42056f11eda1610242ac110002	/

参数	参数类型	说明	示例	下级对象
availableZoneId	String	可用区id	cn-huadong1-jsnj1A-public-ctcloud	/
eipId	String	弹性IP id	eip_sssww	/
bandwidthId	String	绑定ipv6的带宽id	bandwidth-fnzgeac441	
jobId	String	开机jobId	compute_8d9eabb5-89a5-44de-beae-583bf624f6d5	
masterOrderId	String	IT的主订单id	b8d68f4c36734227ac4c8fb7bba58ae9	
iaasMasterOrderId	String	IAAS的主订单id	INT0220240912185928923050 18	
paasResourceId	String	订单来源	6276b490c2504f4ba90e90e857b043e4	
isDeleted	Integer	判断是否删除了当前节点（发生撤单时会删除该节点）（0：未删除，1：已删除）	0	
createTime	Integer	创建时间	1706515356000	/
updateTime	Integer	更新时间	1706515356000	/

参数	参数类型	说明	示例	下级对象
hostStateValue	String	机器状态	运行中	/
deployRoleInstance	Array of Strings	部署角色实例	["NodeMaster"]	/

枚举参数

无

请求头 header 示例

```
{
    "Cluster-Id":"bdbbc8cc1288a78e5851839dc26f58b9"
}
```

请求 url 示例

```
https://emr-global.ctapi.ctyun.cn/v1/emr/cluster/openapi/clusterNodeGroup/getGroupAndHostByCondition
```

请求体 body 示例

```
{
    "clusterId":"bdbbc8cc1288a78e5851839dc26f58b9",
    "selectKey":"192.168.0.208"
}
```

响应示例

请求成功返回值示例


```
{
    "statusCode": 200,
    "message": "success",
    "returnObj": [
        {
            "id": "1759528882838106113",
            "clusterId": "bdbbc8cc1288a78e5851839dc26f58b9",
            "payType": 1,
            "nodeGroupType": "CORE",
            "nodeGroupName": "core",
            "imageId": "f8d28855-ffda-426d-96e2-0f6bf46877e0",
            "hostNum": 3,
            "computeSpecificationId": 101,
            "iaasVmSpecCode": "s7.2xlarge.4",
            "cpuNum": 8,
            "memory": 32,
            "diskSpecificationList": "[{\"diskFunctionType\":1,\"diskFunctionNa
me\":\"系统盘\",\"diskType\":\"cloud\",\"diskTypeName\":\"云硬盘\",\"ioType\":
\"SATA\",\"ioTypeName\":\"普通 IO\",\"volume\":200,\"diskNum\":1},{\"diskFun
ctionType\":2,\"diskFunctionName\":\"数据盘\",\"diskType\":\"cloud\",\"diskTy
```

peName\":\"云硬盘\",\"ioType\":\"SATA\",\"ioTypeName\":\"普通 IO\",\"volume\":200,\"diskNum\":2}]",
"createTime": 1708339350000,
"updateTime": 1708339350000,
"mountPublicIp": false,
"iaasVmSpecId": "b307034d-cbc3-27bb-24f7-a97565814236",
"highest": null,
"regionId": "bb9fdb42056f11eda1610242ac110002",
"availableZoneId": "cn-huadong1-jsnj1A-public-ctcloud",
"vpcId": "vpc-fr2xjo1gj0",
"clusterHostDtoList": [
{
"id": "af84f22fdefd94bb30f7af4c24368337",
"nodeGroupId": "1759528882838106113",
"iaasHostId": "3d328644-8521-89f7-a114-74b4de602c50",

"hostName": "bigdata-emr-vm-iwxkgpil",
"manageIp": "-",
"serviceIp": "192.168.0.208",
"publicIp": "-",
"ipv6ServiceIp": "240e:982:db0f:6c00:d1ea:8588:69fa:96",
"ipv6PublicIp": "121.237.177.234",
"state": 8,

"vpcId": "vpc-fr2xjo1gj0",

"subnetId": "subnet-vkixrw8xw7",

"regionId": "bb9fdb42056f11eda1610242ac110002",

"availableZoneId": "cn-huadong1-jsnj1A-public-ctcloud",

"eipId": "",

"bandwidthId": "bandwidth-fnzgeac441",

"jobId":"compute_8d9eabb5-89a5-44de-beae-583bf624f6d5",

"masterOrderId": "b8d68f4c36734227ac4c8fb7bba58ae9",

"iaasMasterOrderId": "INT0220240912185928923050 18",

"paasResourceId": "6276b490c2504f4ba90e90e857b043e4",

"isDeleted": 0,

"createTime": 1708339353000,

"updateTime": 1708339353000,

"hostStateValue": "运行中",

"deployRoleInstance": ["DataNode", "NodeManager", "HRegionServer"]

}

]

}

]

}

请求失败返回值示例

```
{
    "statusCode": 200,
    "error": "EMR_400010",
    "message": "clusterId 参数类型错误",
    "returnObj": []
}
```

状态码

请参考状态码

错误码

请参考错误码

6.2.2 根据 id 查询集群信息

接口功能介绍

此接口提供查询用户根据 id 查询集群信息功能，系统接收用户输入的查询条件，返回符合条件的集群详细信息。

接口约束

只允许查询处于运行中状态的集群信息。

接口是否审批

否

URI

GET /v1/emr/openapi/cluster/clusterDetail/getById

Content-Type

application/json

路径参数

无

Query 参数

Query 参数

参数	是否必填	参数类型	说明	示例
id	是	String	集群 id	00c3a04292996955752f073c995a1cc6

请求头 header 参数

无

请求体 body 参数

无

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码	200	/
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	/
error	String	错误码，请求成功时，不返回该字段	EMR_400001	/
returnObj	Object	返回结果	/	returnObj

表 returnObj

参数	参数类型	说明	示例	下级对象
id	String	集群 id	00c3a0429299695 5752f073c995a 1cc6	/
managerClusterId	Integer	manager 定义的集群 id	1	/
iaasType	String	平台	公有云	/
regionId	String	资源池 id	bb9fdb42056f11e da1610242ac110 002	/
regionName	String	资源池名称	华东 1	/
availableZoneId	String	可用区 id	cn-huadong1-js nj1A-public-ctc loud	/
availableZoneNa me	String	可用区名称	可用区 1	/
clusterName	String	集群名称	test_vpc_24	/
payType	String	付费类型	包年包月	/
clusterType	String	集群类型	云搜索	/
clusterTypeVersio n	String	产品版本	翼 MR-2.12.0	/
clusterPlanCode	String	集群规划编码	cloud-search	/
componentNameL ist	String	组件名称列表	[{componentTitle :ElasticSearch,v ersion:7.10.2}, {componentTitle: Kibana,version:7 .10.2}]]	/

参数	参数类型	说明	示例	下级对象
datasourceConfigs	String	数据源信息	[]	/
vpcId	String	vpc id	vpc-0k5xl6w5	/
subnetId	String	子网 id	subnet-i2ys8sp	/
securityGroupId	String	安全组 id	[sg-4h7w9cl1]	/
loginType	String	登录方式	PASSWORD	/
clusterDueTime	Integer	集群到期时间	1709193751000	/
userId	String	用户 id	ba14c8e729e447d69698f81ac7d55555	/
accountId	String	账号 id	ed24e4b414a048b0a9cb995f59cc85jj	/
clusterCreateTime	Integer	集群创建时间	1706515357000	/
clusterState	String	集群状态	运行中	/
managerVersion	String	manager 版本号	2.15.1	/
enableIpv6	String	Ipv6 是否开启(OPEN:打开,CLOSE:关闭,NOT_DISPLAY:不展示,null:不展示)	OPEN	/
createTime	Integer	创建时间	1706515356000	/
updateTime	Integer	更新时间	1706515356000	/
autoRenewStatus	Integer	是否开启自动续订的状态(0:不自动续订,1:自动续订)	1	/

参数	参数类型	说明	示例	下级对象
clusterRunningSeconds	Long	集群运行时间	0	/
pathMap	Map of String	组件名称:组件链接	{ "Kibana": "1.1", "ElasticSearch": "1.1" }	/

枚举参数

无

请求头 header 示例

```
{
    "Cluster-Id":"00c3a04292996955752f073c995a1cc6"
}
```

请求 url 示例

```
https://emr-global.ctapi.ctyun.cn/v1/emr/openapi/cluster/clusterDetail/getById?
id=00c3a04292996955752f073c995a1cc6
```

请求体 body 示例

无

响应示例

```
请求成功返回值示例
{
    "statusCode": 200,
    "message": "success",
```

"returnObj": {

"id": "00c3a04292996955752f073c995a1cc6",

"managerClusterId": 1,

"iaasType": "公有云",

"regionId": "bb9fdb42056f11eda1610242ac110002",

"regionName": "华东 1",

"availableZoneId": "cn-huadong1-jsnj1A-public-ctcloud",

"availableZoneName": "可用区 1",

"clusterName": "test1218",

"payType": "包年包月",

"clusterType": "云搜索",

"clusterTypeVersion": "翼 MR-2.12.0",

"clusterPlanCode": "cloud-search",

"componentNameList": "[{\"componentTitle\":\"ElasticSearch\",\"version\":\"7.10.2\"},{\"componentTitle\":\"Kibana\",\"version\":\"7.10.2\"}]",

"datasourceConfigs": "[]",

"vpcId": "vpc-vgo9wazk0m",

"subnetId": "subnet-6hd5rpedsp",

"securityGroupId": "[\"sg-mz2suubydc\"]",

"loginType": "PASSWORD",

"clusterDueTime": 1705543555000,

"userId": "ba14c8e729e447d69698f81ac7d506b5",

"accountId": "ed24e4b414a048b0a9cb995f59cc85dd",

"clusterCreateTime": 1702865155000,

"clusterState": "运行中",

"managerVersion": "2.15.1",

"enableIpv6": "NOT_DISPLAY",

"createTime": 1702865154000,

"updateTime": 1702979599000,

"autoRenewStatus": 0,

"clusterRunningSeconds": 0,

"pathMap": {

"Kibana": "1,1",

"ElasticSearch": "1,1"

}

}

}

请求失败返回值示例

{

"statusCode": 200,

"error": "EMR_400019",

"message": "集群 id 必传",

"returnObj": []

}

状态码

请参考状态码

错误码

请参考错误码

6.2.3 获取节点组信息

接口功能介绍

此接口提供查询用户集群节点组信息功能，系统接收用户输入的查询条件，返回符合条件的集群节点组详细信息。

接口约束

只允许查询处于运行中状态的集群的节点组信息。

接口是否审批

否

URI

GET /v1/emr/openapi/cluster/clusterNodeGroup/getByClusterId

Content-Type

application/json

路径参数

无

Query 参数

Query 参数

参数	是否必填	参数类型	说明	示例
clusterId	是	String	集群 id	da595eb1d81503b323fdc01d9bf786b7

请求头 header 参数

无

请求体 body 参数

无

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码	200	/
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	/
error	String	错误码，请求成功时，不返回该字段	EMR_400001	/
returnObj	Array of Objects	返回结果	/	returnObj

表 returnObj

参数	参数类型	说明	示例	下级对象
id	String	节点组 id	175187849787779	/
clusterId	String	集群 id	da595eb1d81503b323fdc01d9bf786b7	/

参数	参数类型	说明	示例	下级对象
payType	Integer	付费类型	1	/
nodeGroupType	String	节点类型code码	MASTER	/
nodeGroupName	String	节点类型code值	master	/
imageId	String	镜像id	f8d28855-ffda-426d-96e2-0f6bf46877e0	/
hostNum	Integer	主机数量	3	/
computeSpecificationId	Integer	主机规格id	101	/
iaasVmSpecCode	String	IAAS虚机规格编码	s7.2xlarge.4	/
cpuNum	Integer	cpu核数	8	/
memory	Integer	内存大小	32	/
diskSpecificationList	String	磁盘规格列表	[{diskFunctionType:1,diskFunctionName:系统盘,diskType:cloud,diskTypeName:云硬盘,ioType:SSD-genric,ioTypeName:通用型SSD,volume:200,diskNum:1}]	/
createTime	Integer	创建	1706515356000	/

参数	参数类型	说明	示例	下级对象
		时间		
updateTime	Integer	更新时间	1706515356000	/
mountPublicIp	Boolean	是否挂载公网	false	/
iaasVmSpecId	String	IAAS虚机规格id	b307034d-cbc3-27bb-24f7-a97565814236	/
highest	Boolean	节点组规格是否最高的	false	/

枚举参数

无

请求头 header 示例

```
{
    "Cluster-Id":"da595eb1d81503b323fdc01d9bf786b7"
}
```

请求 url 示例

```
https://emr-global.ctapi.ctyun.cn/v1/emr/openapi/cluster/clusterNodeGroup/getByClusterId?clusterId=da595eb1d81503b323fdc01d9bf786b7
```

请求体 body 示例

无

响应示例

请求成功返回值示例


```
{
    "statusCode": 200,
    "message": "success",
    "returnObj": [
        {
            "id": "1751878497877790722",
            "clusterId": "da595eb1d81503b323fdc01d9bf786b7",
            "payType": 1,
            "nodeGroupType": "MASTER",
            "nodeGroupName": "master",
            "imageId": "f8d28855-ffda-426d-96e2-0f6bf46877e0",
            "hostNum": 3,
            "computeSpecificationId": 101,
            "iaasVmSpecCode": "s7.2xlarge.4",
            "cpuNum": 8,
            "memory": 32,
            "diskSpecificationList": "[{\"diskFunctionType\":1,\"diskFunctionNa
me\":\"系统盘\",\"diskType\":\"cloud\",\"diskTypeName\":\"云硬盘\",\"ioType\":
```

\"SSD-genric\",\"ioTypeName\":\"通用型 SSD\",\"volume\":200,\"diskNum\":1},{\"diskFunctionType\":2,\"diskFunctionName\":\"数据盘\",\"diskType\":\"cloud\",\"diskTypeName\":\"云硬盘\",\"ioType\":\"SSD-genric\",\"ioTypeName\":\"通用型 SSD\",\"volume\":200,\"diskNum\":2}]",
"createTime": 1706515356000,
"updateTime": 1706515356000,
"mountPublicIp": false,
"iaasVmSpecId": "b307034d-cbc3-27bb-24f7-a97565814236",
"highest": null
},
{
"id": "1751878497890373634",
"clusterId": "da595eb1d81503b323fdc01d9bf786b7",
"payType": 1,
"nodeGroupType": "CORE",
"nodeGroupName": "core",
"imageId": "f8d28855-ffda-426d-96e2-0f6bf46877e0",
"hostNum": 1,
"computeSpecificationId": 101,
"iaasVmSpecCode": "s7.2xlarge.4",
"cpuNum": 8,
"memory": 32,

"diskSpecificationList": "[{\"diskFunctionType\":1,\"diskFunctionName\":\"系统盘\",\"diskType\":\"cloud\",\"diskTypeName\":\"云硬盘\",\"ioType\":\"SSD-genric\",\"ioTypeName\":\"通用型 SSD\",\"volume\":200,\"diskNum\":1},{\"diskFunctionType\":2,\"diskFunctionName\":\"数据盘\",\"diskType\":\"cloud\",\"diskTypeName\":\"云硬盘\",\"ioType\":\"SSD-genric\",\"ioTypeName\":\"通用型 SSD\",\"volume\":200,\"diskNum\":2}]",
"createTime": 1706515356000,
"updateTime": 1706515356000,
"mountPublicIp": false,
"iaasVmSpecId": "b307034d-cbc3-27bb-24f7-a97565814236",
"highest": null
}
]
}

请求失败返回值示例

{
"statusCode": 200,
"error": "EMR_400021",
"message": "不是合法的集群 id",
"returnObj": []
}

状态码

请参考状态码

错误码

请参考错误码

6.2.4 集群信息分页查询

接口功能介绍

此接口提供查询用户集群信息功能，系统接收用户输入的查询条件，返回用户输入的资源池下符合条件的集群详细信息，不会返回已删除的集群信息。

接口约束

无

接口是否审批

否

URI

POST /v1/emr/openapi/cluster/clusterDetail/selectPage

Content-Type

application/json

路径参数

无

Query 参数

无

请求头 header 参数

无

请求体 body 参数

请求体 body 参数

参数	是否必填	参数类型	说明	示例	下级对象
regionId	是	String	资源池 id	bb9fdb42056f11eda1610242ac110002	/
pageIndex	是	Integer	当前页，参数值需为正整数	1	/
pageSize	是	Integer	每页大小，参数值需为正整数	10	/
clusterName	否	String	集群名称	cluster_name	/
clusterStateCode	否	Integer	集群状态码(1 启动中，2 运行中，3 异常终止，4 释放中，5 已终止，10 已冻结)	1	/
clusterTypeCode	否	Integer	集群类型码(1：数据湖，2：数据分析，3：数据服务，4：云搜索，6：实时数据流)	1	/

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码	200	/

参数	参数类型	说明	示例	下级对象
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	/
error	String	错误码，请求成功时，不返回该字段	EMR_400001	/
returnObj	Object	返回结果	/	returnObj

表 returnObj

参数	参数类型	说明	示例	下级对象
total	Integer	总集群数量	23	/
size	Integer	当前查询返回的集群数量	1	/
current	Integer	当前页	1	/
orders	Array of Objects	排序字段和排序方式	/	orderItem
optimizeCountSql	Boolean	自动优化 count SQL 查询	true	/
searchCount	Boolean	是否进行 count 查询，默认 true	true	/
countId	String	进行 count 查询的 id	1	/
maxLimit	Integer	单页分页条数限制	20	/
pages	Integer	总页数	5	/
records	Array of Objects	集群列表	/	records

表 orderItem

参数	参数类型	说明	示例	下级对象
column	String	需要进行排序的字段	regionId	/
acs	Boolean	是否正序排列，默认 true	true	/

表 records

参数	参数类型	说明	示例	下级对象
id	String	集群 id	8571aa313aaf77d45bf6765805a9ce55	/
managerClusterId	Integer	manager 定义的集群 id	1	/
iaasType	String	平台	公有云	/
regionId	String	资源池 id	bb9fdb42056f11eda1610242ac110002	/
regionName	String	资源池名称	华东 1	/
availableZoneId	String	可用区 id	cn-huadong1-jsnj1A-public-ctcloud	/
availableZoneName	String	可用区名称	可用区 1	/
clusterName	String	集群名称	test_vpc_24	/
payType	String	付费类型	包年包月	/
clusterType	String	集群类型	云搜索	/
clusterTypeVersion	String	产品版本	翼 MR-2.12.0	/
clusterPlanCode	String	集群规划编码	cloud-search	/
componentNameList	String	组件名称列表	[{componentTitle:ElasticSearch,version:7.10.2},{componentTitle:Kibana,version:7.10.2}]]	/

参数	参数类型	说明	示例	下级对象
datasourceConfigs	String	数据源信息	[]	/
vpcId	String	vpc id	vpc-0k5xl6w5	/
subnetId	String	子网 id	subnet-i2ys8sp	/
securityGroupId	String	安全组 id	[sg-4h7w9cl1]	/
loginType	String	登录方式	PASSWORD	/
clusterDueTime	Integer	集群到期时间	1709193751000	/
userId	String	用户 id	ba14c8e729e447d69698f81ac7d55555	/
accountId	String	账号 id	ed24e4b414a048b0a9cb995f59cc85jj	/
clusterCreateTime	Integer	集群创建时间	1706515357000	/
clusterState	String	集群状态	运行中	/
managerVersion	String	manager 版本号	2.15.1	/
enableIpv6	String	Ipv6 是否开启(OPEN:打开,CLOSE:关闭,NOT_DISPLAY:不展示,null:不展示)		/
createTime	Integer	创建时间	1706515356000	/
updateTime	Integer	更新时间	1706515356000	/
autoRenewStatus	Integer	是否开启自动续订的状态(0:不自动续订,1:自动续订)	0	/

枚举参数

无

请求头 header 示例

无

请求 url 示例

```
https://emr-global.ctapi.ctyun.cn/v1/emr/openapi/cluster/clusterDetail/selectPage
```

请求体 body 示例

```
{
    "clusterName":"",
    "clusterStateCode": 2,
    "clusterTypeCode": 2,
    "pageIndex": 1,
    "pageSize": 10,
    "regionId":"bb9fdb42056f11eda1619242ac110002"
}
```

响应示例

```
请求成功返回值示例
{
    "statusCode": 200,
    "message": "success",
```

"returnObj": {
"records": [
{
"id": "2cd01a834b8e0c14d4a03c01d6052924",
"managerClusterId": 1,
"iaasType": "公有云",
"regionId": "bb9fdb42056f11eda1610242ac110002",
"regionName": "华东 1",
"availableZoneId": "cn-huadong1-jsnj1A-public-ctcloud",
"availableZoneName": "可用区 1",
"clusterName": "sjfw-sxy-0207-3",
"payType": "包年包月",
"clusterType": "数据服务",
"clusterTypeVersion": "翼 MR-2.12.1",
"clusterPlanCode": "data-service",
"componentNameList": "[{\"componentTitle\":\"OpenLDAP\",\"version\":\"2.4.50\"},{\"componentTitle\":\"Kerberos\",\"version\":\"1.18.2\"},{\"componentTitle\":\"ZooKeeper\",\"version\":\"3.7.1\"},{\"componentTitle\":\"HDFS\",\"version\":\"3.3.3\"},{\"componentTitle\":\"YARN\",\"version\":\"3.3.3\"},{\"componentTitle\":\"HBase\",\"version\":\"2.4.12\"},{\"componentTitle\":\"Ranger\",\"version\":\"2.2.0\"}]",
"datasourceConfigs": "[{\"compType\":\"Ranger\",\"properties\":[{\"propName\":\"DB_Host\",\"propValue\":\"192.168.0.156\"},{\"propName\":

\"DB_Port\",\"propValue\":\"13049\"},{\"propName\":\"DB_Name\",\"propValue\":\"ranger0207_4_sxy\"},{\"propName\":\"DB_User\",\"propValue\":\"root\"},{\"propName\":\"DB_Password\",\"propValue\":\"t3O15dEAnNxQbKytUnCpi+ir5lbdkt45Zxm9BSO1YollowerqD0dCA22hMVLDwhff5na3fz4KAnNW7fQpP84ZaKgX6XRovjEsWJ0Gyn9ss0NZ3nRI7jjOW08CnNOmGRirY3OebmYO9PsSucjVMGCPAplyhnjMefhGJHOFnvw4xU=\"}]}]",

"vpcId": "vpc-fr2xjo1gj0",

"subnetId": "subnet-vkixrw8xw7",

"securityGroupId": "[\"sg-72kksosxfn\"]",

"loginType": "PASSWORD",

"clusterDueTime": 1709795918000,

"userId": "ba14c8e729e447d69698f81ac7d506b5",

"accountId": "ed24e4b414a048b0a9cb995f59cc85dd",

"clusterCreateTime": 1707290310000,

"clusterState": "运行中",

"managerVersion": "2.15.1",

"enableIpv6": "NOT_DISPLAY",

"createTime": 1707290309000,

"updateTime": 1707291860000,

"autoRenewStatus": 0

}

],

"total": 1,

```
            "size": 10,
            "current": 1,
            "orders": [],
            "optimizeCountSql": true,
            "searchCount": true,
            "countId": "",
            "maxLimit": null,
            "pages": 1
        }
}
```


请求失败返回值示例


```
{
    "statusCode": 200,
    "error": "EMR_400010",
    "message": "pageIndex,regionId 为必填项，未传",
    "returnObj": []
}
```


状态码

请参考状态码

错误码

请参考错误码

6.2.5 创建集群

接口功能介绍

创建集群（包周期，即包年/包月）。

接口约束

参数校验必须通过。

注意事项：

- 云搜索和数据分析类型集群不支持 task 节点组。
- 云搜索类型集群的 core 类型节点组机器数量为 1—15，且可以没有 core 节点组。
- 非云搜索类型的集群，都需要存在 master 和 core—1 节点组，且 core—1 节点组机器数量为 3—15。
- 自定义类型集群需参考组件依赖关系。

components 组件信息如下：

- data-lake 数据湖

  必选组件：OpenLDAP(2.4.50)、Kerberos(1.18.2)、ZooKeeper(3.7.1)、HDFS(3.3.3)、YARN(3.3.3)、Hive(3.1.2)、Tez(0.10.1)、Spark(3.4.1)、Hudi(0.14.0)、Iceberg(1.4.3)

  可选组件：HBase(2.4.12)、Trino(406)、Kyuubi(1.8.2)、Doris(2.1.5)、Ranger(2.2.0)、Kafka(2.8.1)、KafkaUI(1.0.0)、Knox(1.6.1)、Flink(1.16.2)、JeekeFS(1.1.1)、Flume(1.10.0)、SeaTunnel(2.3.7)、Pushgateway(1.6.2)

- data-analysis 数据分析

  必选组件：Doris(2.1.5)

- data-service 数据服务

  必选组件：OpenLDAP(2.4.50)、Kerberos(1.18.2)、ZooKeeper(3.7.1)、HDFS(3.3.3)、YARN(3.3.3)、HBase(2.4.12)

  可选组件：Ranger(2.2.0)、Knox(1.6.1)、JeekeFS(1.1.1)

- cloud-search 云搜索

  必选组件：Elasticsearch(7.10.2)、Kibana(7.10.2)

可选组件：Logstash(7.10.2)

- real-time-data-processing 实时数据流

必选组件：HDFS(3.3.3)、YARN(3.3.3)、Flink(1.16.2)、ZooKeeper(3.7.1)、Kerberos(1.18.2)、OpenLDAP(2.4.50)、Hudi(0.14.0)、Iceberg(1.4.3)、SeaTunnel(2.3.7)

可选组件：Kafka(2.8.1)、KafkaUI(1.0.0)、Knox(1.6.1)、JeekeFS(1.1.1)、Pushgateway(1.6.2)

- customize 自定义

可选组件：OpenLDAP(2.4.50)、Kerberos(1.18.2)、ZooKeeper(3.7.1)、HDFS(3.3.3)、YARN(3.3.3)、Hive(3.1.2)、Spark(3.4.1)、Hudi(0.14.0)、Iceberg(1.4.3)、HBase(2.4.12)、Trino(406)、Kyuubi(1.8.2)、Doris(2.1.5)、Ranger(2.2.0)、Kafka(2.8.1)、KafkaUI(1.0.0)、Knox(1.6.1)、Flink(1.16.2)、JeekeFS(1.1.1)、Flume(1.10.0)、Tez(0.10.1)、SeaTunnel(2.3.7)、Pushgateway(1.6.2)

自定义集群组件依赖关系如下：

Flink 依赖（全部）HDFS、Hudi、Iceberg、Kerberos、OpenLDAP、SeaTunnel、YARN、ZooKeeper

HBase 依赖（全部）HDFS、Kerberos、OpenLDAP、ZooKeeper

HDFS 依赖（全部）Kerberos、OpenLDAP、ZooKeeper

Hive 依赖（全部）HDFS、Hudi、Iceberg、Kerberos、OpenLDAP、Tez、YARN、ZooKeeper

Hudi 依赖（任意）Flink、Hive、Spark

Iceberg 依赖（任意）Flink、Hive、Spark

JeekeFS 依赖（全部）Kerberos

Kafka 依赖（全部）Kerberos、ZooKeeper

KafkaUI 依赖（全部）Kafka、Kerberos、OpenLDAP

Kerberos 依赖（全部）OpenLDAP

Knox 依赖（全部）Kerberos、OpenLDAP

Kyuubi 依赖(全部)HDFS、Hive、Kerberos、OpenLDAP、Spark、YARN、ZooKeeper

Ranger 依赖（全部）HDFS、Kerberos、OpenLDAP、ZooKeeper

Spark 依赖（全部）HDFS、Hive、Hudi、Iceberg、Kerberos、OpenLDAP、YARN、ZooKeeper

Trino 依赖（全部）Kerberos、OpenLDAP

YARN 依赖（全部）HDFS、Kerberos、OpenLDAP、ZooKeeper

ZooKeeper 依赖（全部）Kerberos、OpenLDAP

接口是否审批

否

URI

POST /v2/emr/openapi/order/new

Content-Type

application/json

路径参数

无

Query 参数

无

请求头 header 参数

无

请求体 body 参数

请求体 body 参数

参数	是否必填	参数类型	说明	示例	下级对象
cluster_name	是	String	集群名称：大写字母、小写字母、数字和特殊符号组成，最大支持 28 个字符	sjfw-test-1	
cluster_type	是	String	集群类型： data-lake：数据湖 data-analysis：数据分析 data-service：数据服务 cloud-search：云搜索 real-time-data-processing：实时数据流 customize：自定义	data-lake	
components	是	String	组件信息：多个组件“,”隔开	OpenLDAP,Kerberos,ZooKeeper,HDFS,YARN,Hive,Tez,Spark,Hudi,Iceberg,Ranger	
region_id	是	String	资源池 ID	bb9fdb42056f11eda1610242ac110002	
available_zone_id	是	String	可用区 ID	cn-huadong1-jsnj3A-public-ctcloud	
vpc_Id	是	String	VPC ID	vpc-n83zi9vuo0	
subnet_id	是	String	子网 ID	subnet-8hzbyype9r	
security_	是	String	安全组 ID	sg-zfh4pgydhe	

参数	是否必填	参数类型	说明	示例	下级对象
groups_id					
enable_ipv6	否	Boolean	开启 IPv6： true：开启 false：关闭 VPC 支持 IPv6 时默认 true	false	
node_root_password	是	String	节点的 root 密码：长度为 12-26 字符，需包含大写字母、小写字母、数字和特殊符号（仅包括：~!@#$%^*_-+{[]}:,.?），不能包含 root、toor 相关大小写变形字符串和 3 位及以上连续数字或字符	Juhwqe7h!21Da.	
charge_info	是	Object	计费信息		表 charge_info
datasource_configs	否	Array of Objects	关联数据连接信息：部署 Hive 和 Ranger 等组件时，可以关联数据连接，将元数据存储于关联的数据库		表 datasource_config
node_details	是	Array of Objects	节点组详情		表 node_

参数	是否必填	参数类型	说明	示例	下级对象
					detail

表 charge_info

参数	是否必填	参数类型	说明	示例	下级对象
charge_mode	是	String	计费模式： prePaid：包周期，即包年/包月	prePaid	
period_type	是	String	周期类型： month：包月 year：包年	month	
period_num	是	Integer	周期数： 当“period_type”为“month”时，取值为 1-11 当“period_type”为“year”时，取值为 1-5	1	
is_auto_renew	否	Boolean	是否自动续费： true：是 false：否 默认 false	false	

表 datasource_config

参数	是否必填	参数类型	说明	示例	下级对象
type	是	String	组件类型：Hive、Ranger	Hive	

参数	是否必填	参数类型	说明	示例	下级对象
properties	是	Array of Objects	配置信息		表 property

表 property

参数	是否必填	参数类型	说明	示例	下级对象
name	是	String	名称：DB_Host、DB_Port、DB_Name、DB_User、DB_Password		
value	是	String	值		

表 node_detail

参数	是否必填	参数类型	说明	示例	下级对象
group_name	是	String	节点组名称：master，core1-11，task1-10，节点组总数不超过12	master	
node_num	是	Integer	节点数量：master 节点组固定 3，core 和 task 节点组 1-15 (非云搜索	3	

参数	是否必填	参数类型	说明	示例	下级对象
			类型的集群 core-1 节点数量最小为 3)		
node_size	是	String	实例规格	s7.2xlarge.4	
affinity_group	是	Boolean	反亲和性云主机组： true：反亲和 false：亲和 默认 true	true	
disk_specifications	是	Array of Objects	磁盘规格		表 disk_specification

表 disk_specification

参数	是否必填	参数类型	说明	示例	下级对象
disk_function	是	String	磁盘分类： SYSTEM：系统盘 DATA：数据盘	SYSTEM	
disk_type	是	String	磁盘类型： cloud：云盘	cloud	
io_type	是	String	磁盘 IO 类型： SATA：普通 IO SAS：高 IO	SATA	

参数	是否必填	参数类型	说明	示例	下级对象
			SSD：超高 IO SSD-genric：通用型 SSD		
disk_size	是	Integer	磁盘大小：80-32768，单位 GB	80	
disk_num	是	Integer	磁盘数量：系统盘固定 1，数据盘 1-6	1	

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码： 成功：200 失败：500	200	
error	String	错误码，请求成功时，不返回该字段	EMR_400000	
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	
returnObj	Object	返回结果		表 returnObj

表 returnObj

参数	参数类型	说明	示例	下级对象

参数	参数类型	说明	示例	下级对象
orderNo	String	订单号	20221018153152727265	

枚举参数

无

请求头 header 示例

无

请求 url 示例

https://emr-global.ctapi.ctyun.cn/v2/emr/openapi/order/new

请求体 body 示例

数据湖请求示例：


```
{
    "cluster_name": "sjfw-test-1",
    "cluster_type": "data-lake",
    "components":
"OpenLDAP,Kerberos,ZooKeeper,HDFS,YARN,Hive,Tez,Spark,Hudi,Iceberg,HBase,Trino,Kyuubi,Doris,Ranger,Kafka,Knox,Flink,JeekeFS,Flume",
    "region_id": "bb9fdb42056f11eda1610242ac110002",
    "available_zone_id": "cn-huadong1-jsnj3A-public-ctcloud",
    "vpc_Id": "vpc-n83zi9vuo0",
    "subnet_id": "subnet-8hzbyype9r",
    "security_groups_id": "sg-zfh4pgydhe",
    "enable_ipv6": false,
    "node_root_password": "Juhwqe7h!21Da.",
```

```
    "charge_info": {
        "charge_mode": "prePaid",
        "period_type": "month",
        "period_num": 1,
        "is_auto_renew": false
    },
    "datasource_configs": [{
            "type": "Hive",
            "properties": [{
                    "name": "DB_Host",
                    "value": "192.168.1.100"
                }, {
                    "name": "DB_Port",
                    "value": "3306"
                }, {
                    "name": "DB_Name",
                    "value": "hive-1"
                }, {
                    "name": "DB_User",
                    "value": "hive"
                }, {
                    "name": "DB_Password",
                    "value": "A121daDAqw2A."
                }
```

```
            ]
        }, {
            "type": "Ranger",
            "properties": [{
                "name": "DB_Host",
                "value": "192.168.1.100"
            }, {
                "name": "DB_Port",
                "value": "3306"
            }, {
                "name": "DB_Name",
                "value": "ranger-1"
            }, {
                "name": "DB_User",
                "value": "ranger"
            }, {
                "name": "DB_Password",
                "value": "A121daDAqw2A."
            }
            ]
        }
    ],
    "node_details": [{
            "group_name": "master",
```

```
            "node_num": 3,
            "node_size": "s7.2xlarge.4",
            "affinity_group": true,
            "disk_specifications": [{
                "disk_function": "SYSTEM",
                "disk_type": "cloud",
                "io_type": "SATA",
                "disk_size": 80,
                "disk_num": 1
            }, {
                "disk_function": "DATA",
                "disk_type": "cloud",
                "io_type": "SATA",
                "disk_size": 80,
                "disk_num": 2
            }
            ]
        }, {
            "group_name": "core-1",
            "node_num": 3,
            "node_size": "s7.2xlarge.4",
            "affinity_group": true,
            "disk_specifications": [{
                "disk_function": "SYSTEM",
```

```
                "disk_type": "cloud",
                "io_type": "SATA",
                "disk_size": 80,
                "disk_num": 1
            }, {
                "disk_function": "DATA",
                "disk_type": "cloud",
                "io_type": "SATA",
                "disk_size": 80,
                "disk_num": 2
            }
        ]
    }, {
        "group_name": "task-1",
        "node_num": 1,
        "node_size": "s7.2xlarge.4",
        "affinity_group": true,
        "disk_specifications": [{
                "disk_function": "SYSTEM",
                "disk_type": "cloud",
                "io_type": "SATA",
                "disk_size": 80,
                "disk_num": 1
            }, {
```

```
                    "disk_function": "DATA",
                    "disk_type": "cloud",
                    "io_type": "SATA",
                    "disk_size": 80,
                    "disk_num": 2
                }
            ]
        }
    ]
}
```


响应示例

请求成功示例：


```
{
    "statusCode": 200,
    "message": "success",
    "returnObj": {
        "orderNo": "20221018153152727265"
    }
}
```


验签失败示例：


```
{
    "statusCode": 500,
    "error": "EMR_401000",
    "message": "OpenAPI 认证失败",
```

```
    "returnObj": {}
}
```


实名认证失败示例：


```
{
    "statusCode": 500,
    "error": "EMR_401009",
    "message": "账号未进行实名认证",
    "returnObj": {}
}
```


非法操作示例：


```
{
    "statusCode": 500,
    "error": "EMR_401004",
    "message": "非法操作",
    "returnObj": {}
}
```


请求参数类型错误示例：


```
{
    "statusCode": 500,
    "error": "EMR_401001",
    "message": "请求参数类型错误",
    "returnObj": {
        "cluster_Name": "应为字符串"
    }
```

```
}
```

请求参数值无效示例 1：

```
{
    "statusCode": 500,
    "error": "EMR_401002",
    "message": "请求参数值无效",
    "returnObj": {
        "cluster_Name": "不能为空"
    }
}
```

请求参数值无效示例 2：

```
{
    "statusCode": 500,
    "error": "EMR_401002",
    "message": "请求参数值无效",
    "returnObj": {
        "cluster_type": "集群类型不存在：xxx"
    }
}
```

规格不足示例：

```
{
    "statusCode": 500,
    "error": "EMR_401012",
    "message": "规格已售罄",
```

```
    "returnObj": {}
}
```

IP 不足示例：

```
{
    "statusCode": 500,
    "error": "EMR_401013",
    "message": "子网下 IP 余量不足",
    "returnObj": {}
}
```

状态码

请参考状态码

错误码

请参考错误码

6.2.6 扩容集群

接口功能介绍

扩容集群。

接口约束

仅支持对 V2.15 及以上版本的集群节点进行扩容。

当前仅支持对数据湖、数据分析、数据服务、实时数据流和自定义业务场景的 core 与 task 节点进行扩容。

被扩容的集群必须处于运行状态中，且当前不存在运维操作。

接口是否审批

否

URI

POST /v2/emr/openapi/order/increase

Content-Type

application/json

路径参数

无

Query 参数

无

请求头 header 参数

无

请求体 body 参数

请求体 body 参数

参数	是否必填	参数类型	说明	示例	下级对象
cluster_id	是	String	集群 ID：云搜索类型的集群不支持扩容	22cbc1c6c69e5abe863e43a091cd02b7	
node_group_name	是	String	节点组名称	core-1	
count	是	Integer	扩容节点数量：MASTER 类型节点组不允许扩容，CORE 和 TASK 类型节点组节点总数最多 15 台	1	
scale_without_start	否	Boolean	扩容后是否选择不启动扩容节点上的组件： true：扩容后不启动组件 false：扩容后启动组件 默认 true	true	

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码： 成功：200 失败：500	200	
error	String	错误码，请求成功时，不返回该字段	EMR_400000	
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	
returnObj	Object	返回结果		表 returnObj

表 returnObj

参数	参数类型	说明	示例	下级对象
orderNo	String	订单号	20221018153152727265	

枚举参数

无

请求头 header 示例

无

请求 url 示例

https://emr-global.ctapi.ctyun.cn/v2/emr/openapi/order/increase

请求体 body 示例

```
{
    "cluster_id":"22cbc1c6c69e5abe863e43a091cd02b7",
    "node_group_name":"core-1",
    "count":1,
    "scale_without_start":true}
```

响应示例

请求成功示例：

```
{
```

```
    "statusCode": 200,
    "message": "success",
    "returnObj": {
        "orderNo": "20221018153152727265"
    }
}
```


请求失败示例 1：


```
{
    "statusCode": 500,
    "error": "EMR_400000",
    "message": "请求失败",
    "returnObj": {}
}
```


请求失败示例 2：


```
{
    "statusCode": 500,
    "error": "EMR_400000",
    "message": "当前资源存在在途单 不允许重复提交，产品规格实例 ID：0cfc7bb1e6ae48649a1749d0b5b3ad11，在途单的订单号：20241206151843080993 ",
    "returnObj": {}
}
```


验签失败示例：


```
{
    "statusCode": 500,
    "error": "EMR_401000",
    "message": "OpenAPI 认证失败",
    "returnObj": {}
}
```


实名认证失败示例：


```
{
    "statusCode": 500,
    "error": "EMR_401009",
    "message": "账号未进行实名认证",
    "returnObj": {}
}
```


非法操作示例：

```
{
    "statusCode": 500,
    "error": "EMR_401004",
    "message": "非法操作",
    "returnObj": {}
}
```


请求参数类型错误示例：


```
{
    "statusCode": 500,
    "error": "EMR_401001",
    "message": "请求参数类型错误",
    "returnObj": {
        "count": "应为 int 范围内整数"
    }
}
```


请求参数值无效示例 1：


```
{
    "statusCode": 500,
    "error": "EMR_401002",
    "message": "请求参数值无效",
    "returnObj": {
        "count": "不能为空"
    }
}
```


请求参数值无效示例 2：


```
{
    "statusCode": 500,
    "error": "EMR_401002",
    "message": "请求参数值无效",
    "returnObj": {
        "cluster_id": "cloud_search 类型的集群无法扩容"
    }
}
```


运维约束示例：


```
{
    "statusCode": 500,
```

```
    "error": "EMR_401010",
    "message": "存在与之约束的运维操作",
    "returnObj": {}
}
```

服务异常示例：

```
{
    "statusCode": 500,
    "error": "EMR_401011",
    "message": "存在服务异常，正在为您加速处理中",
    "returnObj": {}
}
```

规格不足示例：

```
{
    "statusCode": 500,
    "error": "EMR_401012",
    "message": "规格已售罄",
    "returnObj": {}
}
```

IP 不足示例：

```
{
    "statusCode": 500,
    "error": "EMR_401013",
    "message": "子网下 IP 余量不足",
    "returnObj": {}
}
```

状态码

请参考状态码

错误码

请参考错误码

6.2.7 退订集群

接口功能介绍

根据集群 id 退订集群。

接口约束

退订集群必须处于运行状态或异常终止状态。

接口是否审批

否

URI

GET /v2/emr/openapi/order/refund

Content-Type

application/json

路径参数

无

Query 参数

Query 参数

参数	是否必填	参数类型	说明	示例
id	是	String	集群	00c3a04292996955752f073c995a1cc6

请求头 header 参数

无

请求体 body 参数

无

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码： 成功：200 失败：500	200	
message	String	用来简述当前接口调用状态以及必要提示信息	请求成功	
error	String	错误码，请求成功时，不返回该字段	EMR_400000	
returnObj	Object	返回结果		

枚举参数

无

请求头 header 示例

无

请求 url 示例

https://emr-global.ctapi.ctyun.cn/v2/emr/openapi/order/refund?id=00c3a04292996955752f073c995a1cc6

请求体 body 示例

无

响应示例

退订成功示例：


```
{
  "statusCode": 200,
  "message": "success",
  "returnObj": {
    "errorMessage": "",
    "submitted": true,
    "orderPlacedEvents": [
      {
        "errorMessage": "",
        "submitted": true,
        "newOrderId": "1825ded866c7485fb3bb4918e4e3b805",
        "newOrderNo": "20241202110327910353",
        "totalPrice": 7223.04
      }
    ]
  }
}
```


验签失败示例：


```
{
  "statusCode": 500,
  "error": "EMR_401000",
  "message": "OpenAPI 认证失败",
  "returnObj": {}
```

```
}
```

请求参数值无效示例：

```
{
  "statusCode": 500,
  "error": "EMR_401002",
  "message": "请求参数值无效",
  "returnObj": {
    "id": "不能为空"
  }
}
```

集群不存在/集群无退订权限示例：

```
{
  "statusCode": 500,
  "error": "EMR_401004",
  "message": "非法操作",
  "returnObj": {}
}
```

无法退订的集群示例：

```
{
  "statusCode": 500,
  "error": "EMR_401004",
  "message": "非法操作",
  "returnObj": "clusterState: 已终止"
}
```

状态码

请参考状态码

错误码

请参考错误码

6.2.8 元数据概览

接口功能介绍

此接口提供获取集群元数据概览功能，系统接收用户输入的查询条件，返回集群的库总数、表总数、总存储量、总文件数等概览信息。

接口约束

集群必须处于运行状态。

接口是否审批

否

URI

GET /v1/emr/doctor/openapi/meta/hive/overview

Content-Type

application/json

路径参数

无

Query 参数

Query 参数

参数	是否必填	参数类型	说明	示例
clusterId	是	String	集群 id	da595eb1d81503b323fdc01d9bf786b7
timestamp	是	Long	要查询的时间点，以 Unix 时间戳表示（单位：秒）	1700000000

请求头 header 参数

无

请求体 body 参数

无

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码	200	
message	String	用来简述当前接口调用状态以及必要提示信息	success	
error	String	错误码，请求成功时，不返回该字段	EMR_400001	
returnObj	Object	返回结果		returnObj

枚举参数

无

请求头 header 示例

无

请求 url 示例

https://emr-global.ctapi.ctyun.cn/v1/emr/doctor/openapi/meta/hive/overview?clusterId=da595eb1d81503b323fdc01d9bf786b7&timestamp=1700000000

请求体 body 示例

无

响应示例

请求成功返回值示例

```
{
  "statusCode": 200,
  "message": "success",
  "returnObj": {
    "databaseCnt": 5,
    "tableCnt": 237,
    "storageCnt": 10240000,
    "fileCnt": 12223
  }
}
```

请求失败返回值示例

```
{
    "statusCode": 500,
    "error": "EMR_401005",
    "message": "该集群不存在",
    "returnObj": {}
}
```

状态码

请参考状态码

错误码

请参考错误码

6.2.9 元数据信息查询

接口功能介绍

此接口提供查询集群下指定表的元数据的功能，用户输入集群 id、时间戳、Hive 库名、Hive 表名，返回集群的该表的文件数、存储大小、分区数等元数据。

接口约束

集群必须处于运行状态。

接口是否审批

否

URI

GET /v1/emr/doctor/openapi/meta/hive/tableInfo

Content-Type

application/json

路径参数

无

Query 参数

Query 参数

参数	是否必填	参数类型	说明	示例
clusterId	是	String	集群 id	da595eb1d81503b323fdc01d9bf786b7
timestamp	是	Long	要查询的时间点，以 Unix 时间戳表示（单位：秒）	1700000000
databaseName	是	String	Hive 库名	test_db1
tableName	是	String	Hive 表名	test_table1

请求头 header 参数

无

请求体 body 参数

无

响应参数

响应参数

参数	参数类型	说明	示例	下级对象
statusCode	Integer	状态码	200	
message	String	用来简述当前接口调用状态以及必要提示信息	success	

参数	参数类型	说明	示例	下级对象
error	String	错误码，请求成功时，不返回该字段	EMR_400001	
returnObj	Object	返回结果		returnObj

枚举参数

无

请求头 header 示例

无

请求 url 示例

https://emr-global.ctapi.ctyun.cn/v1/emr/doctor/openapi/meta/hive/tableInfo?clusterId=cluster-12345&timestamp=1700000000&databaseName=test_db1&tableName=test_table1

请求体 body 示例

无

响应示例

请求成功返回值示例


```
{
  "statusCode": 200,
  "message": "success",
  "returnObj": {
    "clusterId": "da595eb1d81503b323fdc01d9bf786b7",
    "databaseName": "test_db1",
    "tableName": "test_table1",
    "hivePath": "/hive/xxx/xxx/xx",
    "fileCnt": 100,
    "storageCnt": 204800,
    "avgStorageCnt": 2048,
    "lastAccessTime": "2024-11-06 01:09:37",
    "isExternalTable": false,
    "isMisDelete": false,
    "isPartitionTable": true,
    "partitionCnt": 20,
    "iceFileCnt": 1,
    "iceStorageCnt": 20,
```

```
    "coldFileCnt": 1,
    "coldStorageCnt": 1024,
    "emptyFilePtCnt": 1,
    "emptyFilePtFileCnt": 2,
    "smallFilePtCnt": 5,
    "smallFilePtFileCnt": 5
  }
}
```


请求失败返回值示例


```
{
    "statusCode": 500,
    "error": "EMR_401005",
    "message": "该集群不存在",
    "returnObj": {}
}
```


状态码

请参考状态码

错误码

请参考错误码

7 常见问题

7.1 产品咨询类

7.1.1 翼 MR 支持什么类型的分布式存储?

翼 MR 集群内当前使用主流的大数据 Hadoop，目前支持 Hadoop 3.3.3 版本，并且会随着集群的演进逐步更新相关稳定版本。

同时客户可以在开通翼 MR 产品中选择使用 Doris 组件，Doris 同样提供分布式数据存储能力，当前默认提供的 Doris 为 2.1.8 版本。

更多的翼 MR 产品中组件信息请参见产品文档中的版本概述。

7.1.3 翼 MR 是否支持更换网段?

当前，翼 MR 集群内节点暂不支持更换网段。因此，客户在选择开通翼 MR 集群前，建议基于业务需要，做好相关网段划分并提前在天翼云其它产品，如：虚拟私有云（CT-VPC）中开通好合适的资源。

7.1.4 翼 MR 集群内节点是否支持降配操作?

当前，翼 MR 集群内节点暂不支持降级配置规格。客户可以基于业务的需要，基于翼 MR 集群中的 Master、Core、Task 节点组按需进行升级配置规格操作。

7.1.5 翼 MR 集群是否支持 Hive on Spark?

当前，翼 MR 集群中的 Hive 服务，默认提供的是 Hive on MR 的计算能力。

用户可以在开通好翼 MR 集群后，进入翼 MR Manager 的管控平台中，并在 Hive 服务的配置管理页中，修改相关配置参数，重启 Hive 服务，完成 Hive on Spark 的切换并生效。

7.1.6 不同版本的 Hive 之间是否可以兼容?

Hive 3.1 版本与 Hive 1.2 版本相比不兼容内容主要如下：

- 字段类型约束：Hive 3.1 不支持 String 转成 int。
- UDF 不兼容：Hive 3.1 版本 UDF 内的 Date 类型改为 Hive 内置。
- 索引功能废弃。

- 时间函数问题：Hive 3.1 版本为 UTC 时间，Hive 1.2 版本为当地时区时间。
- 驱动不兼容：Hive 3.1 和 Hive 1.2 版本的 JDBC 驱动不兼容。
- Hive 3.1 对 ORC 文件列名大小写，下划线敏感。
- Hive 3.1 版本列中不能有名为 time 的列。

当前，翼 MR 集群提供的是高稳定、高性能的 3.1.2 版本 Hive 组件能力，建议客户在开通翼 MR 集群后，优先考虑完成业务上 Hive 使用版本的升级适配。

7.1.7 翼 MR 是否支持 Hive on Tez 和 Hive on Kudu?

当前，翼 MR 集群中的 Hive 组件，支持 Hive on Tez，暂不支持 Hive on Kudu 模式。

翼 MR 集群开通后，默认支持 Hive on MR 模式。客户可以在开通翼 MR 集群后，通过进入翼 MR Manager 管控平台，在 Hive 服务的配置管理中，修改相关配置参数，并重启 Hive 服务，实现 Hive on Spark 切换和生效。

7.1.8 翼 MR 集群内节点上/var/log 目录中的日志文件能否清理?

翼 MR 集群内节点上的/var/log 目录中，存放了操作系统的相关操作、安全等关键性日志数据以及翼 MR 集群各个部署组件实例的应用和审计日志，以方便客户对云主机的系统问题以及翼 MR 组件集群进行定位和安全审计。

- 针对上述的系统日志文件，不建议客户进行清理，该操作属于高危操作，易引发弹性云主机的故障不可用问题。
- 针对上述的翼 MR 集群应用和审计日志，建议客户基于必要性考虑，确认是否需要清理。如果需要清理，优先考虑将相关历史久远的日志执行删除操作。

7.1.9 Kafka 支持的访问协议类型有哪些?

Kafka 支持四种协议类型的访问，分别为：PLAINTEXT、SSL、SASL_PLAINTEXT、SASL_SSL。当前，翼 MR 集群默认采用 Kerberos 安全验证服务，Kafka 协议类型建议使用 SASL_PLAINTEXT、SASL_SSL 这两种。

7.1.10 翼 MR 产品支持 Hive 元数据库使用内置 MySQL 类型吗?

当前，翼 MR 产品不支持 Hive 元数据库使用内置 MySQL 类型。内置 MySQL 不具备生产运行所需要的高可用、高安全性，且该方式缺乏有效的技术保障能力，客户需要为此承担不必要的业务风险。

在翼 MR 集群开通的流程中，基于上述产品考量，不开放相关选配功能，默认推荐客户预先开通好天翼云成熟的数据库产品—关系数据库 MySQL 版（CT—RDS MySQL）作为 Hive 服务的元数据库。

7.1.11 是否支持修改翼 MR 集群中节点的 IP 地址?

当前，翼 MR 集群中的节点不支持修改 IP 地址，强行修改 IP 会带来翼 MR 集群服务间通信异常，导致整个翼 MR 集群业务不可用故障。建议客户在开通翼 MR 集群前，考虑好可用区和 VPC 的规划。

7.1.13 翼 MR 管理控制台和翼 MR Manager 页面区别与联系?

用户可以通过翼 MR 管理控制台的“翼 MR Manager”页面进入到翼 MR 的 Manager 页面。管理控制台和 Manager 页面的区别和联系请参考下表：

常用操作	翼 MR 控制台	翼 MR Manager
查看集群基础信息、IAM 同步	支持	不支持
配置升级、远程连接、绑定/解绑弹性 IP	支持	不支持
操作日志	支持	支持
用户管理	支持	不支持
访问链接与端口	支持	不支持
启动服务、停止服务、滚动重启服务	不支持	支持
启动、停止、重启、滚动停止、滚动重启实例	不支持	支持
查看监控、告警信息	不支持	支持

常用操作	翼 MR 控制台	翼 MR Manager
LDAP 租户管理	不支持	支持
运维与配置管理	不支持	支持

7.2 计费类

翼 MapReduce 支持哪些计费方式?

目前翼 MapReduce 支持包周期方式购买。

翼 MapReduce 账户余额不足是否有提醒?

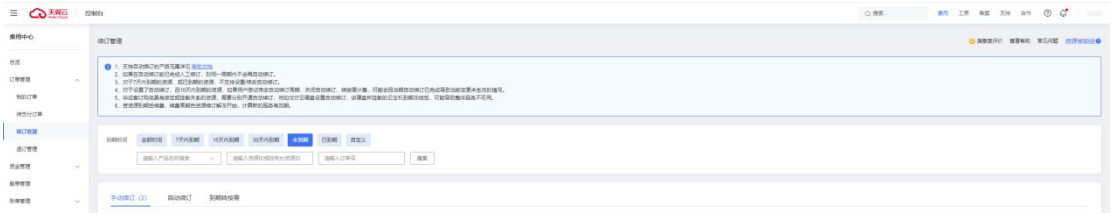
用户可在用户中心自助设置余额阈值，低于阈值即会短信提醒。

购买翼 MapReduce 还需要支付其他费用吗?

在您使用翼 MapReduce 集群时，可按需购买弹性 IP 等产品用于满足公网访问等需求。

翼 MR 如何续费?

目前翼 MR 支持手动续费与自动续费两种方式，若订购时未开启自动续费，可以通过官网左侧导航栏点击“订单管理”–“续订管理”，输入筛选条件，找到需要续订的产品，按提示完成手动续订操作或开启自动续费功能。



7.3 购买类

天翼云都有哪些资源池可订购翼 MapReduce?

目前翼 MapReduce 已在华东 1、西南 1、华北 2、上海 36、华南 2、武汉 41、呼和浩特 3、南昌 5、杭州 7 等一类节点资源池上线，后续会持续扩展更多资源池资源。

翼 MapReduce 支持什么类型的分布式存储?

翼 MapReduce 提供目前主流的 Hadoop，目前支持 Hadoop 3.3.3 版本，并且随社区更新版本。

翼 MapReduce 支持购买哪些业务场景的大数据集群?

翼 MapReduce 业务场景主要包括 6 个：数据湖场景、数据分析场景、云搜索场景、数据服务、实时数据流和自定义场景。

场景	描述
数据湖	数据湖分析 提供更灵活、可靠、高效的管理集群，更快的运行大数据计算引擎并提供出色的数据分析能力。
数据分析	数据分析 Apache Doris：开源的 MPP 架构的 OLAP 分析引擎，支持亚秒级的数据查询和多表 join。
云搜索	云搜索 为结构化/非结构化数据提供低成本、高性能及可靠性的检索、分析服务能力。
数据服务	数据服务 提供更灵活、可靠、高效的数据服务集群。
实时数据流	实时数据流 提供流式计算、消息队列等服务，主要用于实时数据 ETL 和日志采集分析等场景。
自定义	自定义 提供丰富灵活的服务搭配，支持自行选择业务所需服务。

翼 MapReduce 产品购买包括哪些计费项?

翼 MR 集群费用由翼 MR 服务管理费用和硬件产品费用组成，硬件产品费用由主机费用和硬盘费用组成。

7.4 操作类

如何使用组件客户端?

1. 以 root 用户登录任意一个 Master 节点。
2. 执行 su - omm 命令，切换到 omm 用户。
3. 执行 cd 客户端安装目录 ，切换到客户端。
4. 执行 source bigdata_env 命令，配置环境变量。如果当前集群已启用 Kerberos 认证，执行 kinit 组件业务用户认证当前用户。如果当前集群未启用 Kerberos 认证，则无需执行此命令。
5. 环境变量配置成功后，即可执行组件的客户端命令。例如查看组件的相关信息，可执行 HDFS 客户端命令 hdfs dfs -ls /查看 HDFS 根目录文件。

集群支持提交哪些形式的 Spark 作业?

在翼 MR 后台中，集群支持提交 Spark、Spark Script 和 Spark SQL 形式的 Spark 作业。

翼 MR 集群的计算资源最大值为 0 后，还可以跑 Spark 任务吗?

翼 MR 集群的租户计算资源最大值改为 0 后，不可以跑 Spark 任务。

Spark 作业的 Client 模式和 Cluster 模式有什么区别?

理解 YARN-Client 和 YARN-Cluster 深层次的区别之前先清楚一个概念：Application Master。

在 YARN 中，每个 Application 实例都有一个 ApplicationMaster 进程，它是 Application 启动的第一个容器。它负责和 ResourceManager 打交道并请求资源，获取资源之后告诉 NodeManager 为其启动 Container。从深层次的含义讲 YARN-Cluster 和 YARN-Client 模式的区别其实就是 ApplicationMaster 进程的区别。

YARN-Cluster 模式下，Driver 运行在 AM(Application Master)中，它负责向 YARN 申请资源，并监督作业的运行状况。当用户提交了作业之后，就可以关掉 Client，作业会继续在 YARN 上运行，因而 YARN-Cluster 模式不适合运行交互类型的作业。

YARN-Client 模式下，Application Master 仅仅向 YARN 请求 Executor，Client 会和请求的 Container 通信来调度工作，也就是说 Client 不能离开。

如何关闭 ZooKeeper 的 SASL 认证?

1. 登录翼 MR Manager。
2. 选择“集群服务 > ZooKeeper > 配置管理 ”。
3. 在左侧导航栏选择“配置组 > zoo.cfg”，然后，选择该配置文件右上角的“+”号，添加参数名称：zookeeper.sasl.disable，再添加参数值：false。
4. 在左侧导航栏选择“配置组 > 同步”，执行配置同步工作。
5. 重启 ZooKeeper 服务。

在翼 MR 集群外客户端中执行 kinit 报错“Permission denied”如何处理?

问题现象

在翼 MR 集群外节点上安装了客户端后并执行 kinit 命令报错如下：

```
-bash kinit Permission denied
```

执行 java 命令正常如下：

```
-bash: /xxx/java: Permission denied
```

执行 ll / *java 安装路径* /JDK/jdk/bin/java 命令查看该文件执行权限信息正常。

原因分析

执行 mount | column -t 查看挂接的分区状态，发现 java 执行文件所在的挂载点的分区状态是“noexec”。当前环境中将安装翼 MR 客户端所在的数据盘配置成“noexec”，即禁止二进制文件执行，从而无法使用 java 命令。

解决方法

1. 以 root 用户登录翼 MR 客户端所在节点。
2. 移除“/etc/fstab”文件中翼 MR 客户端所在的数据盘的配置项“noexec”。
3. 执行 umount 命令卸载数据盘，然后再执行 mount -a 重新挂载数据盘。

7.5 账号密码类

翼 MR 集群开通后，登录翼 MR Manager 的用户是什么?

系统默认登录翼 MR Manager 的帐号为 op_admin，用户可以直接通过翼 MR 控制台中的免密登录跳转访问到翼 MR Manager。

集群的主机登录密码是什么?

主机登录密码是在翼 MR 集群开通时所设置的，如忘记密码，可前往集群的基础信息页进行密码重置。当前仅支持对 2.16 及以上版本进行密码重置。

组件密码在哪里查询与重置?

组件密码可前往翼 MR Manager 配置管理页面，在 vars.yaml 的高级配置中查询。当前暂不支持在该文件中直接修改组件密码，文件仅用于记录配置信息。2.19 版本起，云搜索场景支持在控制台-集群服务管理页对 Elasticsearch 组件密码进行修改。

LDAP 密码在哪里重置?

LDAP 密码可前往翼 MR Manager LDAP 租户管理页面，选择要修改密码的 LDAP 用户，进入该用户的基础信息页进行密码重置。

Manager 主密码是否支持重置?

当前主密码暂不支持重置，请妥善保存密码。

7.6 账号权限类

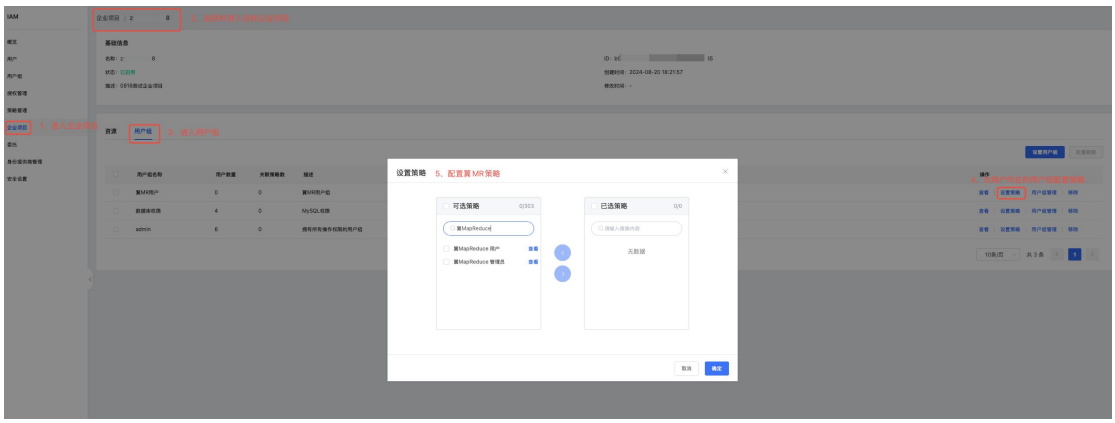
为何在控制台集群列表页看不到已订购的集群?

子用户登录翼 MR 控制台后，可能在集群列表页看不到已订购的集群，此时可按下述步骤进行问题定位与处理。

1、检查资源池是否正确：可与管理员确认已订购集群所在的资源池，并查看页面左上角所选择的资源池是否正确。

2、检查是否具备集群所在企业项目的权限：联系管理员在“账号中心”—“统一身份认证服务”功能中，检查是否有为子用户配置企业项目权限。

管理员可通过用户组授权或基于企业项目完成授权两种方式，为子用户授予在目标企业项目下的翼 MR 使用权限。管理员授权后，权限预计在 1 分钟内生效，子用户可在翼 MR 控制台看到该企业项目下的集群。



如果不开启 Kerberos 认证，翼 MR 集群能否支持访问权限细分?

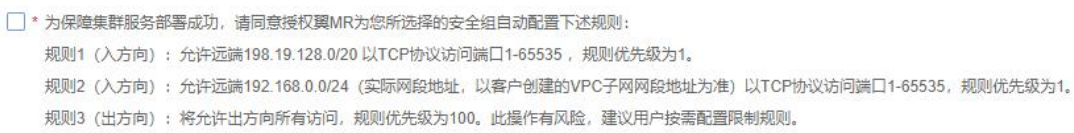
翼 MR 集群在开通过程中，数据湖、数据服务、实时数据流的业务场景是默认开启 Kerberos 认证。基于 Kerberos 认证，才能实现访问权限的细分设置。

翼 MR 集群在开通过程中，云搜索和数据分析的业务场景是默认不使用 Kerberos 认证。其中，数据分析的业务场景采用自带访问权限细分功能，云搜索暂不支持。

7.7 集群创建类

如何使用自定义安全组创建翼 MR 集群?

1. 当前翼 MR 产品“一类节点”仅支持华东 1、西南 1、华北 2、上海 36、华南 2、武汉 41、呼和浩特 3、南昌 5、杭州 7 资源池，后续以实际上线为准。
2. 创建翼 MR 集群前，用户需要在华东 1、西南 1、华北 2、上海 36、华南 2、武汉 41、呼和浩特 3、南昌 5 或杭州 7“一类节点”资源池完成虚拟私有云 VPC 的创建及相关配套安全组规则的添加工作。
3. 参照天翼云-虚拟私有云产品文档中“创建虚拟私有云 VPC”章节，需完成在“一类节点”资源池创建相关 VPC 网段，默认为“192.168.0.0/16”。
4. 完成上述步骤后，在控制中心-网络控制台-访问控制-安全组菜单中，可以发现为创建的 VPC 自动添加了默认安全组。另外，客户可以额外创建自定义的安全组。
5. 在创建翼 MR 集群时，需要授权翼 MR 完成以下规则的配置添加。



购买翼 MR 集群时，找不到 HDFS、YARN 和 ZooKeeper 组件如何处理?

HDFS、YARN 和 ZooKeeper 组件包含在翼 MR 集群的数据湖、数据服务、实时数据流和自定义业务场景中，当购买翼 MR 集群时，如果无法看到 HDFS、YARN 和 ZooKeeper 组件，请确认业务场景选择是否正确，另外确定在开通订购时，已经完成相关组件的勾选。等翼 MR 集群正式开通部署完成后，客户即可在翼 MR Manager 管控平台中查看到 HDFS、YARN 和 ZooKeeper 组件。

翼 MR 集群内节点的登录方式如何切换?

不可以切换。创建集群时选择了集群登录方式后不能更改登录方式。

刚创建的集群为何 Hive 或 Ranger 出现故障?

Hive 与 Ranger 的使用依赖元数据库配置，若创建集群时配置错误，将导致服务异常。此时，您可前往翼 MR Manager 进入 Hive 或 Ranger 的集群服务详情页，在运维操作中选择“元数据库配置”进行信息重置。

7.8 集群管理类

7.8.1 使用类 FAQ

如何查看集群配置信息?

- 集群创建完成后在翼 MR 控制台单击集群名称进入基础信息页面，可以查看集群的基本配置信息，包括集群名称、创建时间、付费类型、集群版本、组件信息、VPC 和安全组、节点的实例规格和存储等。其中，节点的实例规格和存储决定了该集群对数据的分析处理能力。节点实例规格越高，存储越大，集群运行速度越快，分析处理能力越强，相应的成本也越高。
- 点击前往翼 MR Manager，跳转至“运维与配置”中的“配置管理”页面，通过选择不同的“所属集群服务”，可以查看对应集群服务完整的配置列表信息。

升级 Master 节点规格需要关机吗?

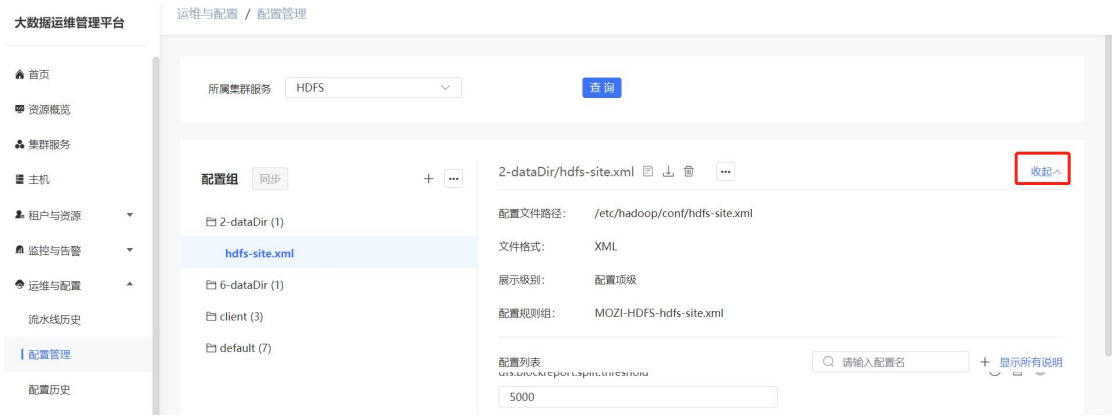
翼 MR 服务集群的 MASTER/CORE/TASK 节点组在进行配置升级时，会自动触发关机操作，当升级完成后会自动触发开机操作，无需人工干预。但是需要注意的是，在配置升级之前需要先停止所有的集群服务，升级完成后再将所有集群服务手动开启。

翼 MR 集群中安装的组件能否删除?

已经创建的集群中的组件不可以删除，如果不使用的话可以登录到 Manager 页面的集群服务页面找到对应的组件将其停止。

如何查看各组件配置文件路径?

集群创建完成后在翼 MR 控制台单击集群名称进入详情后，点击“翼 MR Manager”进入大数据运维管理平台。点击“运维与配置”，进入“配置管理”页面，点击页面右上角的“展开”按钮，即可以查看相应集群服务的配置文件路径。



如何下载客户端?

- 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
- 单击“基础信息”tab，在“软件信息”中单击“客户端”右侧的“前往下载客户端”按钮，即可下载集群服务的客户端信息。

如何启动/停止所有集群服务?

- 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
- 单击“翼 MR Manager”tab，单击菜单“集群服务”，进入集群服务列表页面。单击“更多操作>启动所有集群服务”或“停止所有集群服务”，确认操作后开始启动/停止所有集群服务，等待完成即可。

如何滚动重启集群服务?

- 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
- 单击“翼 MR Manager”tab，单击菜单“集群服务”，进入集群服务列表页面。选择要滚动重启的集群服务，单击集群服务名称进入集群服务详情页面。单击“运维操作>滚动重启集群服务”，进入到滚动重启集群服务页面，按照引导进行下一步操作，等待滚动重启集群服务完成后即可。

如何配置 YARN 队列?

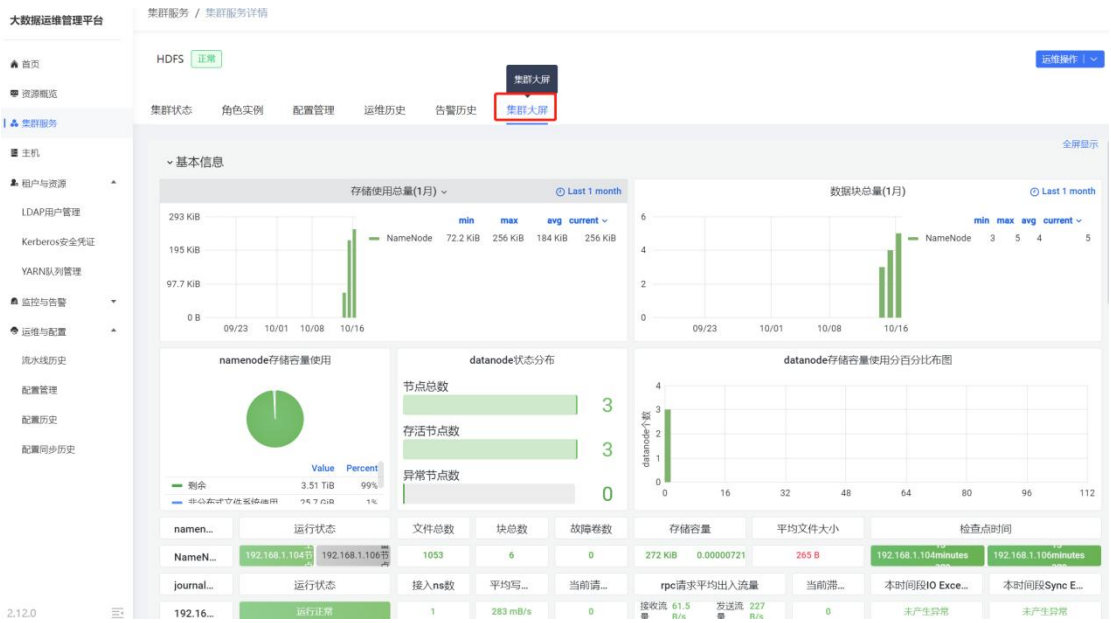
- 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
- 单击“翼 MR Manager”tab，点击菜单“租户与资源 > YARN 队列管理”，即可对集群进行 YARN 队列配置，包括：新建队列、启动队列、停止队列、设置队列容量、队列授权、修改全局属性等操作。

如何分发 Keytab?

- 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
- 单击“翼 MR Manager”tab，点击菜单“租户与资源 > Kerberos 安全凭证”，选择要下发的 Keytab，单击“分发”按钮，出现分发 Keytab 弹框，选择目标主机，依次输入相关字段单击确定等待下发结果即可。

如何查看集群服务监控大屏?

- 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
- 单击“翼 MR Manager”tab，单击菜单“集群服务”，单击指定集群服务，进入集群服务详情页面，单击“集群大屏”tab 即可查看该集群服务的监控大屏。



7.8.2 如何查看所有集群?

翼 MR 所有的集群都展示在“我的集群”页面中，进入“我的集群”页面，可以查看所有集群信息。集群数量较多时，可采用翻页显示，您可以查看任何状态下的集群。集群名称是高亮显示的表示是正常运行的集群，非高亮显示的为非正常运行状态。集群状态包括以下：

- 启动中：代表集群还在部署中，可以查看具体进程。
- 运行中：集群已完成部署/或者续订成功，正在正常运行。
- 异常终止：启动失败/异常原因的终止，可以查看失败原因。
- 释放中：针对退订中的集群，会显示为释放中。
- 已终止：针对已经到期 15 天后或者用户主打退订完成的集群，会显示为已终止，集群会被铲掉。
- 续订中：点击集群进行续订，显示续订中。
- 已冻结：针对已经到期且未超过 15 天的集群，资源会被冻结保留，续费后自动解冻。

7.8.3 如何查看主机和组件日志?

操作步骤

1. 登录翼 MR 管理控制台。
2. 选择“我的集群”，选中一个运行中的集群并单击集群名称，进入集群信息页面。
3. 在“节点管理”页签中单击 Master 节点，选择要进入的 Master 节点。
4. 单击该节点右侧的“远程连接”。

5. 根据界面提示，输入 Master 节点的用户名和密码，用户名、密码分别为 root 和创建集群时设置的密码。
6. 进入对应组件的日志目录，查看相关日志。

7.8.4 如何进行集群服务配置调优?

操作步骤

步骤一 修改配置

1. 登录翼 MR 管理控制台，单击“我的集群”，单击指定的集群名称，进入集群信息页面。
2. 单击“翼 MR Manager”tab，单击“前往翼 MR Manager”。
3. 进入到翼 MR Manager 以后，单击菜单“集群服务”，进入集群服务列表页面。
4. 选择指定集群服务，单击集群服务名称进入集群服务详情页面。
5. 单击“配置管理”tab，单击配置组，单击要修改的配置名称，页面右侧展示该配置的详细信息，对配置进行修改，修改完成后单击“保存更改”按钮。

步骤二 配置同步

1. 单击“同步”按钮，进入待同步配置页面。
2. 单击“配置同步”按钮执行配置同步并进入到配置同步详情页面，查看配置同步进度，等待同步完成。

步骤三 配置生效

1. 配置同步完成后，返回到集群服务详情页面。
2. 单击“运维操作>滚动重启集群服务”，对集群服务滚动重启后即可生效修改后的配置。

7.9 大数据业务开发

7.9.1 执行 Spark 任务报内存不足告警如何处理?

Spark 任务内存不足情况需要根据具体问题进行具体分析。

首先 spark 统一内存模型将内存分为 off-heap 和 heap 两部分内存，每一个部分都可能发生内存不足的情况。

off-heap 内存不足的情况

off-heap 内存作用

为了进一步优化内存的使用以及提高 Shuffle 时排序的效率，Spark 引入了堆外（Off-heap）内存，使之可以直接在工作节点的系统内存中开辟空间，存储经过序列化的二进制数据。

利用 JDK Unsafe API（从 Spark 2.0 开始，在管理堆外的存储内存时不再基于 Tachyon，而是与堆外的执行内存一样，基于 JDK Unsafe API 实现），Spark 可以直接操作系统堆外内存，减少了不必要的内存开销，以及频繁的 GC 扫描和回收，提升了处理性能。堆外内存可以被精确地申请和释放，而且序列化的数据占用的空间可以被精确计算，所以相比堆内内存来说降低了管理的难度，也降低了误差。

off-heap 参数和动态扩展机制

Spark 堆外内存的大小可以由 spark.memory.offHeap.size 控制，spark off-heap 空间只分为 execution 和 storage 1:1 两部分，两部分可以动态扩展。

off-heap 不足

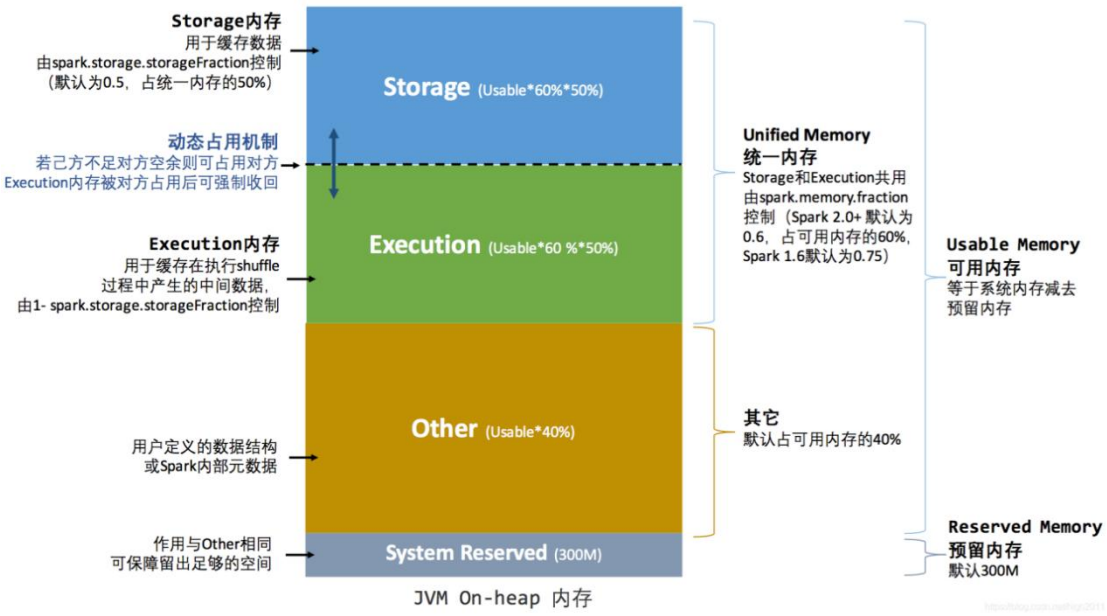
目前常用的 off-heap 常见是 executor 端，map task 侧 shuffle 时候，由于 ShuffleExternalSorter 占用内存过大，导致内存不足。

此外，在 spark native 场景中，spark 将更多的使用 spark off-heap 内存取代 heap 内存，因此 spark off-heap 在非 shuffle 场景下也会占用很多 off-heap 内存。

如果存在 off-heap 内存不足警告，可以酌情添加 spark.memory.offHeap.size 或者 降低 shuffle 内存缓冲区大小。

heap 内存不足的情况

spark 将 heap 内存做如下分割



每一个部分超出使用上线都可能产生内存不足的情况。

Spark 整体 heap 内存是由 spark.executor.memory 和 spark.driver.memory 控制的

Spark 将 spark.xxxx.memory 内存分为 user heap（other）、execution heap、storage heap、system heap（300MB）。

spark.memory.fraction 可以将** **spark.xxxx.memory 分割出 heap storage 和 heap execution 两部分。

Heap storage

Heap storage 作用

Spark Heap Storage 内存用于存储 spark Rdd cache。

在 spark SQL 中也可以用于存储 view、或存储 cache table 的表。

Heap storage 不足

Heap storage 不足一般发生在 executor 端。主要是因为 cache RDD 太多了，因为动态占用机制，占用 heap execution 空间，导致分配到当前 executor 的 heap execution 不够用，从而告警。

我们可以通过 spark UI 查看 作业的 storage 选项卡，定位告警的 executor Id，查看 cache 的大小，与 environment 选项卡中的 spark.executor.memory 做比对。

这种情况下，我们可以酌情，增加 spark.executor.memory，或者 uncache 或者 drop view 不需要的 RDD 或者 cached table。

Heap execution

Heap execution 作用

Heap execution 可以用于存储输入数据。

Spark 是内存计算，需要将父 RDD 的 partition 读入内存中，这个数据将会被 blockManager 读入。

在 spark shuffle map stage 或者 reduce stage 都会发生读入操作。

Heap execution 可以用于 shuffle 缓冲区 和 序列化反序列化。

Spark 的一些使用 heap 缓冲区进行 shuffle 的场景需要分配 heap space。

Heap execution 空间不足

一般是读入的数据太多**或者**发生了数据倾斜。

spark shuffle map task 读入 hdfs 文件块数据。

发生在 shuffle map stage，spark 需要读取 hdfs 文件块。

这个文件快太大且不可分割，或者数据压缩密度太大，会导致 task memory 急速膨胀，超过 1/n 的 executor heap space 和 executor storage space 总和。（此处的 n 是 spark executor 的 running task 数目）。

如果 hdfs 文件块可以分割读入 task，尝试使用 spark.files.maxPartitionBytes 去分割读取的 hdfs 文件 或者 扩大 spark.executor.memory。但是后者不适用于个别 hdfs 大文件块情况。

如果 hdfs 文件块是 gzip 这种的不可压缩，可能只能扩大 spark.files.maxPartitionBytes，或者 在业务上有控制写入 hdfs 的文件格式是可分割的。

spark shuffle reduce task 读入 shuffle block 数据太多了，也就是发生数据倾斜。

一般 shuffle block 都会连续小幅度的拉入 reduce task 侧，此时如果溢出，spark 会将这些数据 spill disk。

如果出现 reduce task 读取超大 shuffle block 或者 并发读取太快，会导致数据内存膨胀太快，而直接 OOM。

此时 spark 会有 task 重启机制，会过滤并发过多情况。如果是读入 shuffle block 太大，则会反复的发生 task failed 情况。

此时可以适当降低 map 侧 shuffle write 的 shuffle block 大小，可以尝试增加 shuffle reduce 的并行度 或者 尽量使用 sortBaseShuffleWriter，使用 sortBaseShuffleWriter 可以实现 map combiner。

如果发生了数据倾斜

此时倾斜 executor 不会出现 Exception，但是内存 heap 使用量偏高，会发出内存告警。

这一点可以在 web UI 的 task 列表中清晰的看到 stage 的某一个 task 相对于其他 task，其 computing time 明显很长。

这种情况下一般是 shuffle 发生了数据倾斜，首先可以使用 map combiner，降低 map 侧的倾斜的 shuffle reduce 分区的数据量。

此外，可以通过编程，提取出 topk 的 shuffle reduce 分区的 key 值，然后将对应 shuffle Key 的 rdd partition 打上随机数，做一个均匀分割，然后执行 reduceByKey 操作，最后去掉随机数，再执行一次 reduceByKey。

如果不想这样，可以使用 hive。

Heap User

Heap User 作用

用户自定义数据结构，这个就是用户使用 spark core api 或者 spark sql api 过程中使用了其他数据结构进行编程，因而产生了一些 heap 消耗。

Heap User 不足

存在 heap 不足的情况，一般是用户将 task 的计算结果，RDD.collect 到 driver 中的数据结构中，这种情况，需要去使用专门的 connector，实现分布式计算和读写。

还有就是用户执行 sql 的时候，会进行非 insert 操作，即直接 select 读取大表，导致 driver 内存不足够，这种情况下建议使用 limit 操作，或者 使用 spark 的流式返回功能。

7.9.2 HBase 表如何设置和修改数据保留期?

- 保证用户具有表的 admin 权限和 create 权限。
- 创建表时指定，创建 t_task_log 表，列族 f，TTL 设置 86400 秒过期。

```
create 't_task_log',{NAME => 'f', TTL=>'86400'}
```

- 在已有表的基础上指定：

```
alter "t_task_log",NAME=>'data',TTL=>'86400' #设置 TTL 值，作用于列族
data
```

7.9.3 如何查看 HBase 日志?

1. 使用 root 用户登录集群的 Master 或 rs 节点。
2. 执行 cd /var/log/Bigdata/hbase/ 命令，进入到"/var/log/Bigdata/hbase/"目录，即可查看 HBase 日志信息。

7.9.4 HBase 支持的压缩算法有哪些?

HBase 目前默认支持的压缩算法有 snappy 和 gz，根据需要可以增加 lzo、lz4、bzip2 等算法支持。

7.9.5 Kafka 目前支持的访问协议类型有哪些?

Kafka 目前支持 4 种协议类型的访问：PLAINTEXT、SSL、SASL_PLAINTEXT、SASL_SSL。

- PLAINTEXT 是最简单的协议，它以明文形式传输数据，不提供任何加密或身份验证机制。
- SSL（Secure Sockets Layer） 是一种安全的协议，用于在客户端和服务器之间加密通信。通过使用 SSL，可以确保数据在传输过程中是加密的，从而提高了安全性。SSL 协议通常需要在 Kafka 服务器上配置 SSL 证书以进行安全通信。
- SASL_PLAINTEXT （Simple Authentication and Security Layer with Plain Text）使用简单身份验证机制，通常用于在不使用加密的情况下进行用户身份验证，但是用户名和密码以明文形式传输。
- SASL_SSL （Simple Authentication and Security Layer with SSL）结合了 SSL 和 SASL，提供了更强大的安全性。它通过 SSL 加密通信，并使用 SASL 进行身份验证。这是 Kafka 中最安全的选项之一，适用于在不同网络环境中进行安全通信。

在 Kafka 中，选择适当的协议取决于集群的安全需求和网络环境。

7.9.6 消费 Kafka Topic 时报错"Not Authorized to access group XXX"如何处理?

该报错是由于 kafka 消费者组缺少 ACL 授权导致的。

可以通过如下命令赋予消费者组 XXX 读取权限：

```
./kafka-acls.sh --authorizer-properties zookeeper.connect=localhost:2181/kafka --add --allow-principal User:UserName --operation Read --group XXX
```

7.9.7 翼 MR 如何连接 spark-shell?

1. 任一用户登录集群客户端节点。
2. 配置/usr/local/ 下 spark 客户端目录 conf 下的配置，主要是 Spark-env.sh 需要配置 SPARK_HOME、HADOOP_HOME、HADOOP_CLASSPATH 等。
3. 如果当前集群已启用 Kerberos 认证，执行以下命令认证当前用户。如果当前集群未启用 Kerberos 认证，则无需执行此命令。
   a. 首先 klist -kt <keytab 文件路径>,获取 keytab 文件的 principal,例如 klist -kt user.keytab 获得 user/hostname@realm。
   b. 然后 kinit -kt <keytab 文件路径> <获取到的 principal>，例如 kinit -kt user.keytab user/hostname@realm。
   c. Kinit 认证完成登录后，可以 klist -l 查看。
4. 如果 SPARK_HOME/bin 已经配置到系统环境变量中，可以直接 spark-shell 进入。否则需要去到 SPARK_HOME/bin 下执行。

7.9.8 翼 MR 如何连接 spark-beeline?

Spark-beeline 连接的目标可以是 spark thrift server、也可以是 kyuubi server。

我们以 spark thrift server 为例，提供链接流程如下：

1. 首先配置 spark thrift 相关配置。
   a. Spark-env.sh 中配置需要配置 JAVA_HOME、HADOOP_HOME、HADOOP_CLASSPATH、HADOOP_CONF_DIR 等环境变量。
   b.

```
export JAVA_HOME=/usr/jdk64/current
export HADOOP_HOME=${HADOOP_HOME:-/usr/local/hadoop3}
export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-/usr/local/hadoop3/etc/hadoop}
SPARK_DAEMON_MEMORY="8g"
```

   c. SPARK_HOME/conf 路径下的 hive-site.xml 文件中，需要配置如下：

```
hive.server2.transport.mode - Set this to value: http
hive.server2.thrift.http.port - HTTP port number to listen on; default is 10001
```

如果 hive-site.xml 是软连接，且 spark 和 hive 混布，那么需要直接 cp 该文件到 SPARK_HOME/conf 下面。

d. Spark-defaults.conf 中如果 hadoop 集群开启了 Kerberos 认证，那么需要配置 spark.sql。

2. 启动 spark thrift server。

```
SPARK_HOME/sbin/start-thriftserver.sh
```

3. 如果当前集群已启用 Kerberos 认证，执行以下命令认证当前用户。如果当前集群未启用 Kerberos 认证，则无需执行此命令。

   a. 首先 klist -kt <keytab 文件路径>,获取 keytab 文件的 principal,例如 klist -kt user.keytab 获得 user/hostname@realm。

   b. 然后 kinit -kt <keytab 文件路径> <获取到的 principal>，例如 kinit -kt user.keytab user/hostname@realm。

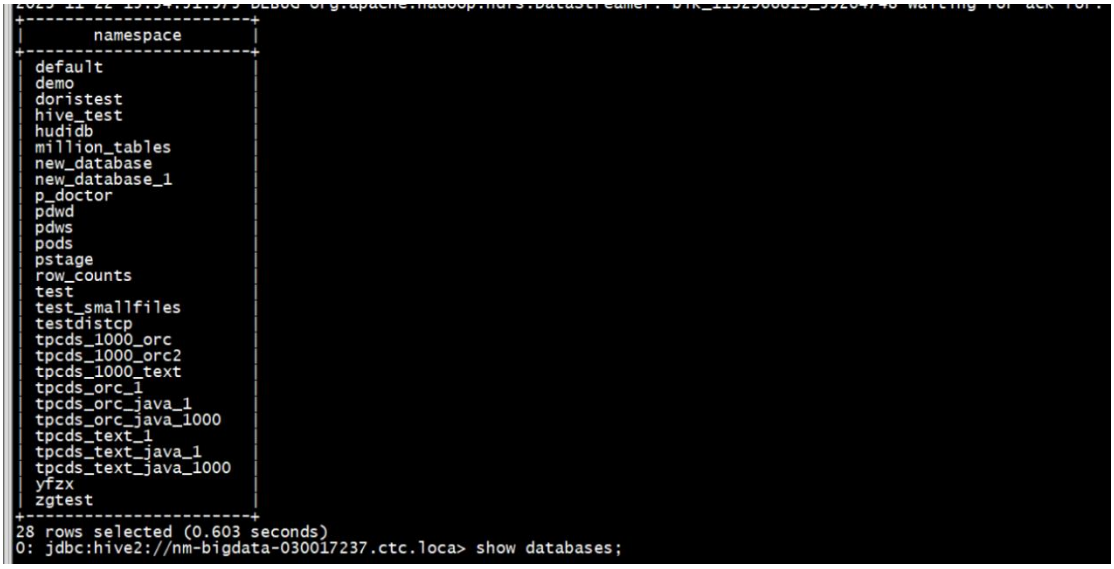
   c. Kinit 认证完成登录后，可以 klist -l 查看。

4. Beeline 链接 spark-thrift-server。

```
SPARK_HOME/bin -u 'jdbc:hive2://<host>:<port>/<database>;principal=user/hostname@realm?spark.yarn.queue=root.default' -n user
```

5. 执行如下命令查询所有表，返回结果中存在表 test，即表示访问 OBS 成功。

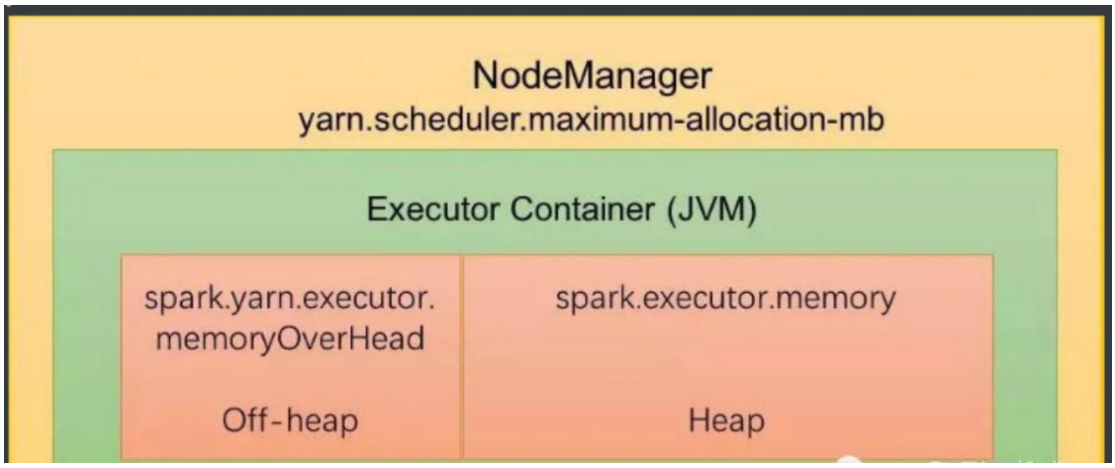
```
show databases
```



7.9.9 Spark.yarn.executor.memoryOverhead 设置不生效如何处理?

参数作用

Spark 是运行在 executor container 进程的 JVM 上的，spark.yarn.executor.memoryOverHead 指的是 executor jvm 的堆外内存。



Executor container 消耗的总内存 = spark.yarn.executor.memoryOverhead + spark.executor.memory + spark.memory.offHeap.size+spark.executor.pyspark.memory

其中 **spark.executor.memory** ，**spark.memory.offHeap.size** 是 spark 运行时消耗的对象的存储内存。

spark.executor.pyspark.memory 是 pyspark 运行时，在 spark.executor.memory + spark.memory.offHeap.size 之外的开销。

spark.yarn.executor.memoryOverhead 指的是 executor container 进程，运行时数据区非堆部分的开销，即执行开销。

它包括 JVM 执行时自身所需要的内容，包括线程堆栈、IO、编译缓存等所使用的内存。

失效原因

在 spark 源码中，该参数出现在如下位置：

```
 * TODO: consolidate it with `ConfigBuilder.withAlternative`.
 */
private val configsWithAlternatives = Map[String, Seq[AlternateConfig]](
  EXECUTOR_USER_CLASS_PATH_FIRST.key -> Seq(
    AlternateConfig("spark.files.userClassPathFirst", "1.3")),
  UPDATE_INTERVAL_S.key -> Seq(
```

```
EXECUTOR_MEMORY_OVERHEAD.key -> Seq(
  AlternateConfig("spark.yarn.executor.memoryOverhead", "2.3")),
```

AlternateConfig 的 key 会被 desprecated。

```
*/
def getDeprecatedConfig(key: String, conf: JMap[String, String]): Option[String] = {
  configsWithAlternatives.get(key).flatMap { alts =>
    alts.collectFirst { case alt if conf.containsKey(alt.key) =>
      val value = conf.get(alt.key)
      if (alt.translation != null) alt.translation(value) else value
    }
  }
}
```

所以，自 spark 2.3 之后，将使用 spark.executor.memoryOverhead，废弃 spark.yarn.executor.memoryOverhead。

```
private[spark] val EXECUTOR_MEMORY_OVERHEAD = ConfigBuilder("spark.executor.memoryOverhead")
  .doc("The amount of non-heap memory to be allocated per executor, in MiB unless otherwise" +
    " specified.")
  .version("2.3.0")
  .bytesConf(ByteUnit.MiB)
  .createOptional
```

7.9.10 SparkSQL 访问 Hive 分区表启动 Job 前耗时较长如何处理?

问题说明

使用 Spark SQL 访问 Hive 表的一个表分区，但是运行速度却很慢。

分析样例：

select** x,y from test where **x=1 （其中 x 是 test 表的 Partition 字段）

原因分析

按照 spark 源码逻辑，在解析逻辑计划时候回去调用 getPartitionsByFilter 方法 去 hive 中只提取 x=1 分区信息。

但是由于一些原因，导致 getPartitionsByFilter 的谓词下推失败，从而去全表扫描所有 test 的分区信息，并返回。

例如，我们 x 字段是 String 类型，但是我们的 SQL 中不是 where **x='1'，而是 where** x=1 ，这就导致了谓词下推失败。

出现 hive 分区表谓词下推失败的情况，我们可以做如下处理：

- 我们需要去检查 sql 中的写法是否正确。
- 可以关闭 SQL 逻辑计划解析过程中的谓词下推逻辑。

处理步骤

关闭 SQL 逻辑计划解析过程中的谓词下推逻辑，具体是 Spark SQL 默认开启基于分区统计信息的执行计划优化，相当于自动执行 Analyze Table（默认开启的设置方法为 spark.sql.statistics.fallBackToHdfs=true，可通过配置为 false 关闭）。

开启后，SQL 执行过程中会扫描表的分区统计信息，并作为执行计划中的代价估算，例如对于代价评估中识别的小表，会广播小表放在内存中广播到各个节点上，进行 join 操作，大大节省 shuffle 时间。

此开关对于 Join 场景有较大的性能优化，但是会带来 获取分区表信息 RPC 请求 的增加。

在 SparkSQL 中设置以下参数后再运行：

```
set spark.sql.statistics.fallBackToHdfs=false;
```

或者在启动之前使用--conf 设置这个值为 false：

```
--conf spark.sql.statistics.fallBackToHdfs=false
```

7.9.11 如何获取 Spark Jar 包?

1、获取非 Spark 依赖

从 maven 中央仓库/华为云仓库等 获取 Spark 运行所需要的依赖 ，可以从 Spark Apache 官网获取 Spark 相关的客户端。

Spark 客户端中 jars 目录包含 Spark 运行依赖。

Spark 客户端下载地址参考 Spark Apache 官网。

2、获取 Spark-core 等 Spark 依赖

除了通过 Spark Apache 官网下载客户端获取和 maven 中央仓库获取有时也可以通过源码打包以获取。

可根据如下命令进行 Spark 源码打包：

整体打包命令：

```
./build/mvn -Pyarn -Phive -Phive-thriftserver -DskipTests clean package
```

单模块打包命令：

```
./build/mvn -pl :spark-streaming_2.12 clean install
```

7.9.12 Trino 如何配置其他数据源?

增加数据源，在 Trino 安装目录配置对应的 catalog 文件即可，以 MySQL 数据源为例：

1. 创建${trino_home}/etc/catalog/mysql.properties。
2. 编辑 mysql.properties：

```
connector.name=mysql

connection-url=jdbc:mysql://127.0.0.1:3306

connection-user=test

connection-password=test
```

✧ **说明**

具体的数据源配置可参考 Trino 相应官方文档。

7.10 性能优化类

翼 MapReduce 集群是否支持重装系统?

翼 MR 集群不支持重装系统。另外，不建议客户通过远程连接登录 CT-ECS，尝试自行执行重装系统等高危操作，由此造成翼 MR 集群故障，从而导致业务不可用。

翼 MapReduce 集群是否支持切换操作系统?

翼 MR 集群节点不支持切换操作系统。建议客户在使用翼 MR 集群过程中，有相关使用问题，可以优先提交天翼云工单，向天翼云专业客服进行咨询和沟通，感谢理解。

如何提高集群 Core 节点的资源使用率?

1. 进入翼 MR Manager 管控平台的 YARN 服务的详情页，切换到“集群大屏”页面，通过查看 YARN 集群的历史资源使用趋势图，确认是否需要提高资源使用率。

2. 搜索并修改“yarn.nodemanager.resource.memory-mb”或者“yarn.nodemanager.resource.cpu-vcores”的值，请根据集群的节点内存和 CPU 核数的实际情况调大该值。
3. 保存配置并重启受影响的服务或实例。

如何关闭防火墙服务?

1. 以 root 用户登录集群的各个节点。
2. 检查防火墙服务是否启动。
3. 在 ECS 节点中，执行 systemctl status firewalld.service 命令。
4. 关闭防火墙服务。
5. 在 ECS 节点中，执行 systemctl stop firewalld.service 命令。

8 相关协议

8.1 翼 MapReduce 产品服务协议

点此预览：《翼 MapReduce 产品服务协议》

8.2 翼 MapReduce 产品服务等级协议

点此预览：《翼 MapReduce 产品服务等级协议》