

天翼云 • 弹性负载均衡

用户使用指南

天翼云科技有限公司

目 录

1	产品概述.....	3
1.1	概念.....	3
1.1.1	弹性负载均衡.....	3
1.1.2	监听器.....	4
1.1.3	健康检查.....	5
2	快速入门.....	6
2.1	典型场景说明.....	6
2.2	创建负载均衡器.....	8
2.2.1	创建增强型负载均衡器.....	8
2.2.2	添加监听器.....	10
2.2.3	添加后端云主机.....	13
3	负载均衡管理.....	15
3.1	负载均衡器管理.....	15
3.1.1	查询负载均衡器.....	15
3.1.2	删除负载均衡器.....	15
3.2	监听器管理.....	16
3.2.1	修改监听器.....	16
3.2.2	删除监听器.....	17
3.3	后端云主机管理.....	17

3.3.1	添加后端云主机.....	17
3.3.2	移除后端云主机.....	18
4	常见问题.....	19
4.1	弹性负载均衡是什么？	19
4.2	弹性负载均衡服务是否收费？	19
4.3	弹性负载均衡支持哪些转发方式？	19
4.4	弹性负载均衡（增强型）是否可以添加不同操作系统的云主机？	19
4.5	单个用户支持保有多个弹性负载均衡？	20
4.6	什么是配额？	20
4.7	如何配置内网或公网负载均衡？	20
4.8	监听器是什么？	20
4.9	什么是负载均衡协议（端口）？	20
4.10	什么是云主机协议（端口）？	20
4.11	弹性负载均衡分配的弹性 IP 是否为独占？	21
4.12	删除弹性负载均衡有什么影响？	21
4.13	ELB 支持什么类型的会话保持？	21

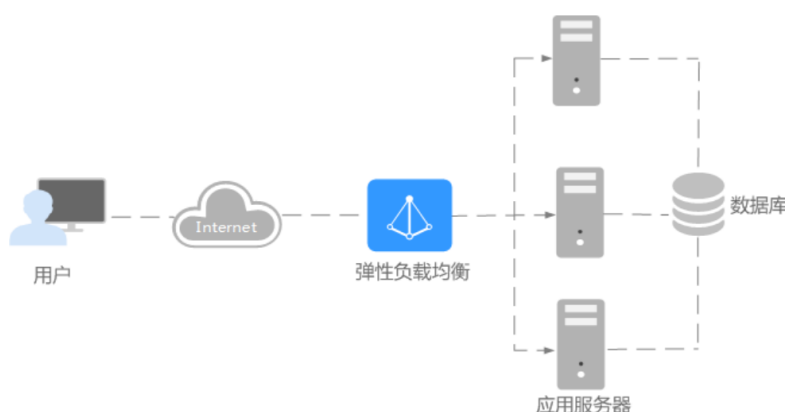
1 产品概述

1.1 概念

1.1.1 弹性负载均衡

弹性负载均衡（Elastic Load Balancing，简称 ELB）是将访问流量根据转发策略分发到后端多台弹性云主机的流量分发控制服务。弹性负载均衡可以通过流量分发扩展应用系统对外的服务能力，实现更高水平的应用程序容错性能。

如下图所示，弹性负载均衡将访问流量分发到后端三台应用服务器，每个应用服务器只需分担三分之一的访问请求。同时，结合健康检查功能，流量只分发到后端正常工作的服务器，从而提升了应用系统的可用性。

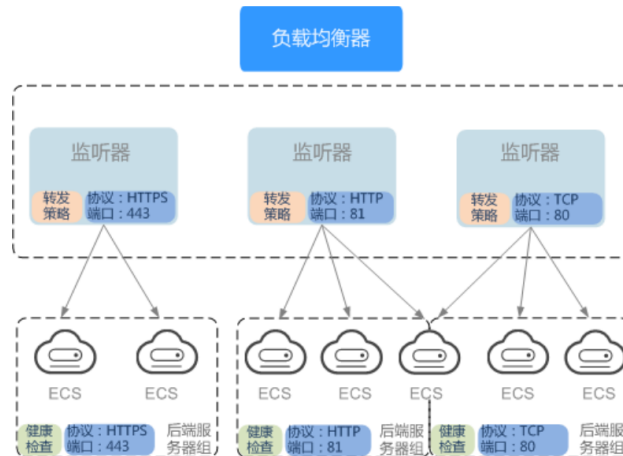


弹性负载均衡器接受来自客户端的传入流量并将请求转发到一个或多个可用区中的后端云主机。

监听器：您可以向您的弹性负载均衡器添加一个或多个监听器。监听器使用您配置的协议和端口检查来自客户端的连接请求，并根据您定义的转发策略将请求转发到一个后端云主机组里的后端云主机。

后端主机组：每个后端云主机组使用您指定的协议和端口号将请求转发到一个或多个后端云主机。

您可以开启健康检查功能，对每个后端云主机配置运行状况检查。当后端某台服务器健康检查出现异常时，弹性负载均衡会自动将新的请求分发到其它健康检查正常的后端云主机上；而当该后端云主机恢复正常运行时，弹性负载均衡会将其自动恢复到弹性负载均衡服务中。



1.1.2 监听器

用户定制的监听器，定义了负载均衡策略和转发规则。负载均衡策略和转发规则的相关概念如下：

监听器使用前端（客户端到负载均衡器）连接的协议以及端口和后端（负载均衡器到后端弹性云主机）连接的协议以及端口配置负载均衡策略。负载均衡器支持协议 HTTP、TCP。负载均衡器可以监听端口 1-65535。

ELB（增强型）支持三种转发规则，用户可以根据自身需求选择相应的算法来分配用户访问流量，提升负载均衡能力。增强型负载均衡的分配策略，支持以下三种调度算法：

- 加权轮询算法：按顺序依次将请求分发给不同的云主机。它用相应的权重表示云主机的处理性能，按照权重的高低以及轮询方式将请求分配给各云主机，相同权重的云主机处理相同数目的连接数。
- 最少连接算法：通过当前活跃的连接数来估计云主机负载情况的一种动态调度算法。
- 源算法：将请求的源 IP 地址作为散列键（HashKey），从静态分配的散列表找出对应的云主机；

ELB（增强型）支持如下三种会话保持方式：

- SOURCE_IP: 将请求的源 IP 地址作为散列键 (HashKey), 从静态分配的散列表找出对应的云主机;
- HTTP_cookie: 负载均衡器会根据客户端第一个请求生成一个 cookie, 后续所有包含这个 cookie 值的请求都会由同一个后端云主机处理;
- APP_cookie: 该选项依赖于后端应用, 后端应用生成一个 cookie 值, 后续所有包含这个 cookie 值的请求都会由同一个后端云主机处理。

1.1.3 健康检查

用户可以配置运行状况检查, 这些检查可用来监控后端云主机的运行状况, 以便负载均衡器只将请求发送到正常运行的后端云主机。而当该故障云主机恢复正常运行时, 负载均衡会将其自动恢复到对外或对内的服务中。健康检查支持协议 TCP 和 HTTP。

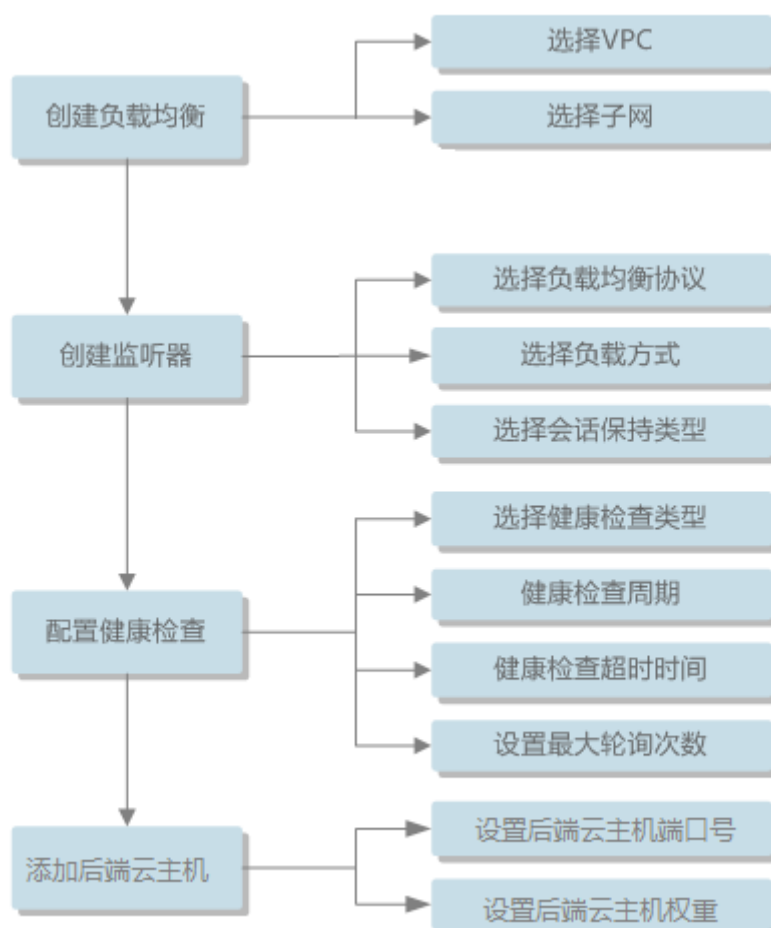
2 快速入门

2.1 典型场景说明

使用负载均衡对后端多台云主机进行流量分发时，需要创建负载均衡，在负载均衡下添加监听器并设置健康检查，最后将后端云主机添加至监听器。

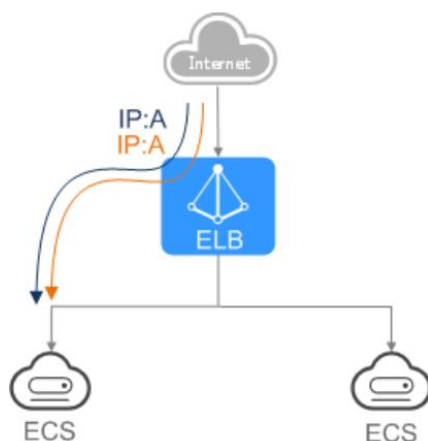
当为用户提供增强型负载均衡服务时，将来自同一个 VPC 下的访问流量自动分发到多台云主机。

配置流程如图所示：



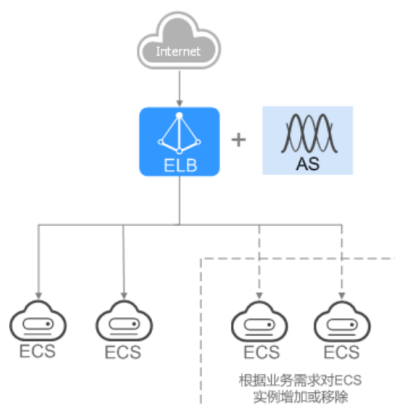
场景一：使用 ELB 为高访问量业务进行流量分发

对于业务量访问较大的业务，可以通过 ELB 设置相应的转发策略，将访问量均匀的分到多个后端处理。例如大型门户网站，移动应用市场等。同时您还可以开启会话保持功能，保证同一个客户请求转发到同一个后端。从而提升访问效率，如图所示。



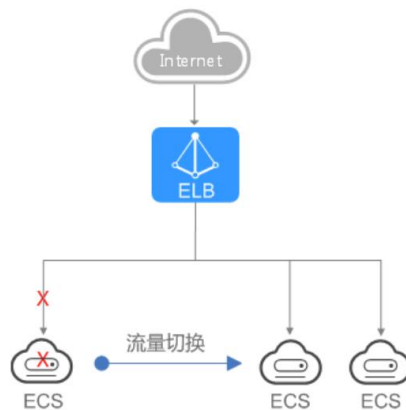
场景二：使用 ELB 和 AS 为潮汐业务弹性分发流量

对于存在潮汐效应的业务，结合弹性伸缩服务，随着业务量的增长和收缩，弹性伸缩服务自动增加或者减少的 ECS 实例，可以自动添加到 ELB 的后端云主机组或者从 ELB 的后端云主机组移除。负载均衡实例会根据流量分发、健康检查等策略灵活使用 ECS 实例资源，在资源弹性的基础上大大提高资源可用性，如图所示。例如电商的“双 11”、“双 12”、“618”等大型促销活动，业务的访问量短时间迅速增长，且只持续短暂的几天甚至几小时。使用负载均衡及弹性伸缩能最大限度的节省 IT 成本。



场景三：使用 ELB 消除单点故障

对于可靠性有较高要求的业务，可以在负载均衡器上添加多个后端云主机。负载均衡器会通过健康检查及时发现并屏蔽有故障的云主机，并将流量转发到其他正常运行的后端云主机，确保业务不中断，如图所示。例如官网，计费业务，Web 业务等。



2.2 创建负载均衡器

本节将说明通过控制中心创建增强型负载均衡器的方法。在这一章节，您将创建一个增强型负载均衡器。

在创建前，请启动您计划添加到负载均衡器的后端云主机，并确保这些云主机的安全组允许端口 22 上的 TCP 访问。

操作步骤如下。

2.2.1 创建增强型负载均衡器

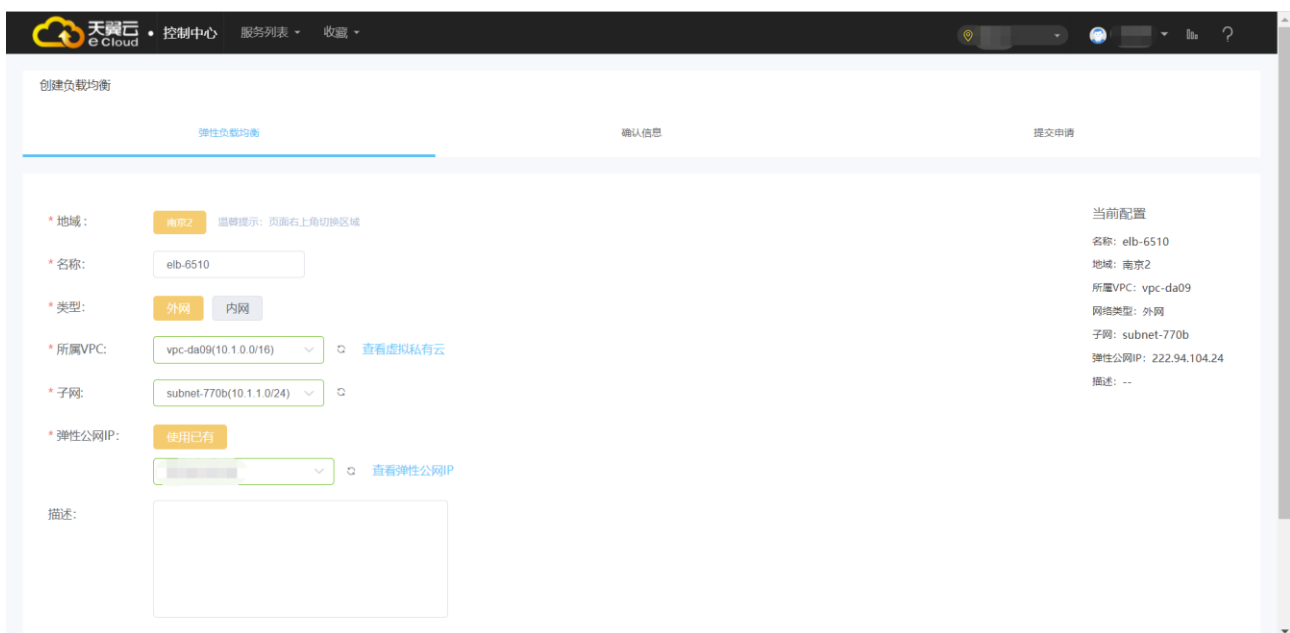
1. 登录天翼云控制中心；
2. 在【所有服务】标签下，选择【网络】【弹性负载均衡】；



3. 在【负载均衡器】界面单击【创建负载均衡】;



4. 在【创建负载均衡】界面，根据界面提示配置参数;

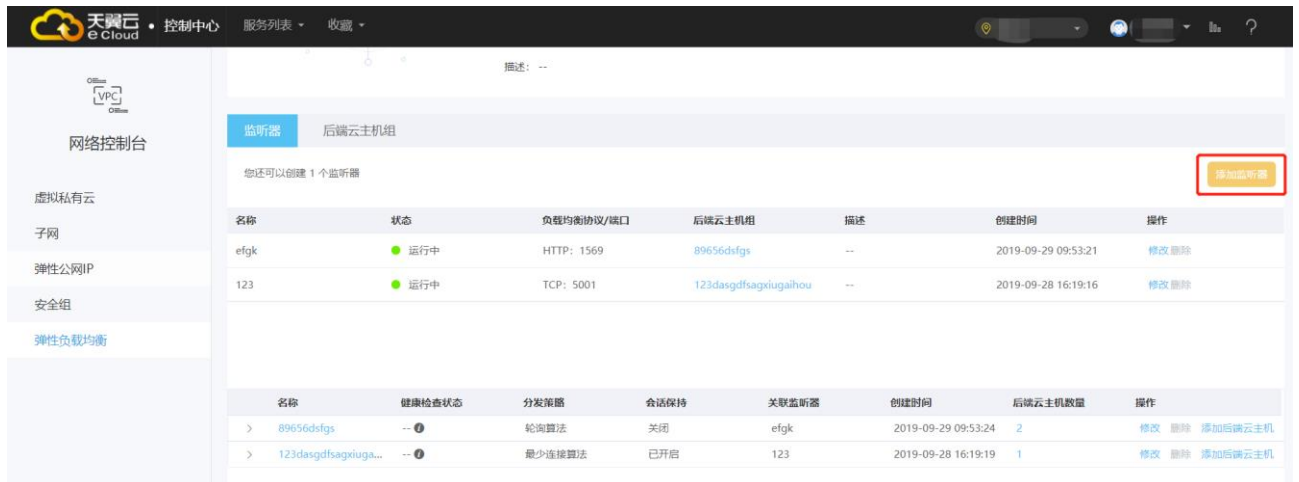


参数	说明	取值样例
名称	负载均衡器的名称。	elb-yss0
类型	可选公网或者内网。 公网：公网负载均衡器通过公网 IP 对外提供服务，将来自公网的客户端请求按照指定的负载均衡策略分发到后端行处理。需要绑定一个已有的 EIP。 内网：内网负载均衡器通过内网 IP 对外提供服务，将来自同一个 VPC 的客户端请求按照指定的负载均衡策略分发到后端进行处理。	内网
所属 VPC	所属虚拟私有云。 您可以选择使用已有的虚拟私有云网络，或者创建新的虚拟私有云。 更多关于虚拟私有云的信息，请参见《虚拟私有云用户指南》。	-
子网	所属子网。	-
负载均衡虚拟 IP	自动分配或手动分配。选择手动分配时需要输入相应的 IP。	自动分配
弹性公网 IP	负载均衡器绑定弹性 IP 后可以接收来自公网的访问请求并自动分发到多台弹性云主机。 您可以将弹性负载均衡绑定一个已有弹性 IP 创建公网内网均可用的负载均衡。	新创建
描述	可添加负载均衡器相关描述。	-

- 单击【下一步】按钮；
- 确认配置无误后，单击【确认下单】，任务下发成功后，关闭创建界面；

2.2.2 添加监听器

- 登录天翼云控制中心；
- 选择【弹性负载均衡】【负载均衡器】；
- 单击已创建的负载均衡器实例名称；
- 在该负载均衡界面的【监听器】区域，单击【添加监听器】按钮；



网络控制台

虚拟私有云

子网

弹性公网IP

安全组

弹性负载均衡

您还可以创建 1 个监听器

[添加监听器](#)

名称	状态	负载均衡协议/端口	后端云主机组	描述	创建时间	操作
efgk	运行中	HTTP: 1569	89656dsfgs	--	2019-09-29 09:53:21	修改 删除
123	运行中	TCP: 5001	123dasgdfsagxiugaihou	--	2019-09-28 16:19:16	修改 删除

名称	健康检查状态	分发策略	会话保持	关联监听器	创建时间	后端云主机数量	操作
> 89656dsfgs	--	轮询算法	关闭	efgk	2019-09-29 09:53:24	2	修改 删除 添加后端云主机
> 123dasgdfsagxiuga...	--	最少连接算法	已开启	123	2019-09-28 16:19:19	1	修改 删除 添加后端云主机

5. 在【添加监听器】界面，根据提示配置参数；



负载均衡业务配置

协议及监听器

新建后端云主机组

负载方式及健康检查

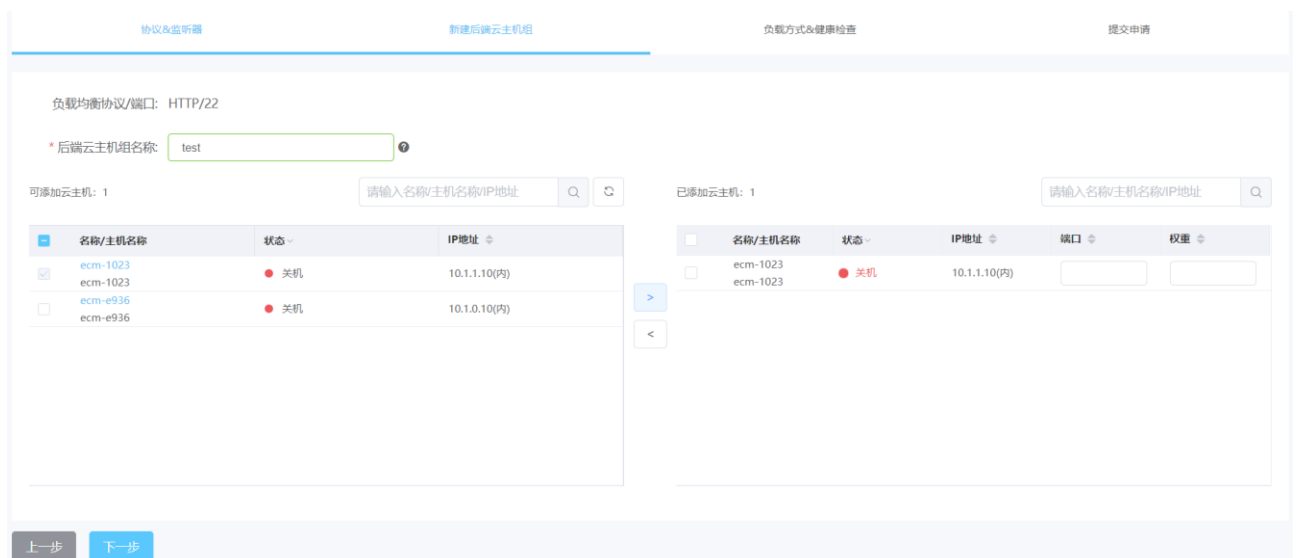
提交申请

* 名称: test

* 负载均衡协议/端口: HTTP 22 取值范围1~65535

描述:

[取消](#) [下一步](#)



协议及监听器

新建后端云主机组

负载方式及健康检查

提交申请

负载均衡协议/端口: HTTP/22

* 后端云主机组名称: test

可添加云主机: 1

请输入名称/主机名称/IP地址

名称/主机名称	状态	IP地址
☒ ecm-1023	关机	10.1.1.10(内)
☒ ecm-e936	关机	10.1.0.10(内)

已添加云主机: 1

请输入名称/主机名称/IP地址

名称/主机名称	状态	IP地址	端口	权重
☐ ecm-1023	关机	10.1.1.10(内)		
☐ ecm-1023	关机	10.1.1.10(内)		

[上一步](#) [下一步](#)

负载均衡业务配置

协议&监听器 新建后端云主机组 负载均衡方式&健康检查 提交申请

* 后端云主机组名称: test

* 负载均衡方式: 轮询算法

* 会话保持: ☒

* 会话保持方式: APP_COOKIE

* APP_Cookie名称:

* 健康检查: ☒

* 健康检查 类型: HTTP

* 间隔时间 (秒): 5 取值范围1~20940 * 超时时间 (秒): 2 取值范围2~60

* 最大重试次数: 2 取值范围1~10

* 检查路径: / 长度范围1~80

* HTTP方法: GET

* HTTP状态码:

[上一步](#)
[立即创建](#)

各参数说明如下:

参数	说明	取值样例
名称	监听器名称。	listener01
负载均衡器协议/端口	负载均衡的协议和端口。 支持以下协议，端口取值范围[1-65535]。 • HTTP • TCP	HTTP/80
分配策略类型	负载均衡采用的算法，用户可以根据自身需求选择相应的算法来分配用户访问流量： • 加权轮询算法：按顺序依次将请求分发给不同的云主机。它用相应的权重表示云主机的处理性能，按照权重的高低以及轮询方式将请求分配给各云主机，相同权重的云主机处理相同数目的连接数； • 加权最少连接：通过当前活跃的连接数来估计云主机负载情况的一种动态调度算法；	加权轮询算法

参数	说明	取值样例
	源 IP 算法：将请求的源 IP 地址作为散列键（HashKey），从静态分配的散列表找出对应的云主机。	
会话保持类型	会话保持的方式，用户可以根据自身需求选择相应的会话保持方式来分配用户访问流量： • SOURCE_IP：将请求的源 IP 地址作为散列键（HashKey），从静态分配的散列表找出对应的云主机； • HTTP_COOKIE：负载均衡器会根据客户端第一个请求生成一个 cookie，后续所有包含这个 cookie 值的请求都会由同一个后端云主机处理； • APP_COOKIE：该选项依赖于后端应用。后端应用生成一个 cookie 值，后续所有包含这个 cookie 值的请求都会由同一个后端云主机处理；	HTTP_COOKIE
健康检查协议	当负载分发协议选择“HTTP”时，健康检查支持的两种类型，设置后不可修改： • TCP • HTTP 当负载分发协议选择“TCP”时，健康检查支持的一种类型，设置后不可修改： • TCP	HTTP
间隔周期（秒）	每次健康检查响应的最大间隔时间；	5
超时时间（秒）	每次健康检查响应的最大超时时间；	10
检查路径	当健康检查方式为 HTTP 时需要配置的选项。为需要被请求的 URL 地址；	/index.html
最大重试次数	健康检查最大的重试次数，范围[1-10]；	3
HTTP 方法	当健康检查方式为 HTTP 时需要配置的选项，为 HTTP 请求的方法；	GET
HTTP 状态码	当健康检查方式为 HTTP 或 HTTPS 时需要配置的选项，为 HTTP 或 HTTPS 请求后表示请求返回的状态码；	201

6. 单击【立即创建】按钮完成监听器创建；

2.2.3 添加后端云主机

您必须将在运行中的云主机添加至您的的负载均衡器中，才能实现负载均衡器对云主机流量分

发的功能。

1. 登录天翼云控制中心；
2. 选择【弹性负载均衡】【负载均衡器】；
3. 单击已创建的负载均衡器实例名称，进入后点击添加后端云主机；

监听器 后端云主机组						
您还可以创建 2 个监听器						
名称	状态	负载均衡协议/端口	后端云主机组	描述	创建时间	操作
test	运行中	HTTP: 22	test	--	2019-10-11 00:25:09	修改 删除

名称	健康检查状态	会话保持	分发策略	关联监听器	创建时间	后端云主机数量	操作
test	运行中	已开启	轮询算法	test	2019-10-11 00:25:11	1	修改 删除 添加后端云主机

名称	主机名称	状态	健康检查状态	IP地址	端口	权重	操作
win888	win888	关机	离线	10.68.0.7(内)	22	1	移除

在页面左侧勾选相应的云主机后，点击“>”按钮将勾选的云主机选入“已添加云主机列表”输入相应的端口和权重后单击“确定”

* 后端云主机组名称: test			请输入名称/主机名称/IP地址		
可添加云主机: 2			已添加云主机: 1		
名称/主机名称	状态	IP地址	名称/主机名称	状态	IP地址
vpc-hah-sub1-vm	关机	10.68.0.15(内)	win888	关机	10.68.0.7(内)
vpc-hah-sub1-vm	关机	10.68.0.7(内)	win888	关机	10.68.0.9(内)
win888	关机	10.68.0.7(内)			
win222	关机	10.68.0.9(内)			
win222	关机	10.68.0.9(内)			

各参数说明如下：

参数	说明	取值样例
端口	后端云主机的服务监听端口，取值范围[1-65535]；	123
权重	后端虚拟机权重。权重值决定了后端云主机处理的请求的比例。例如，一个权重为 2 的云主机处理的请求数是权重为 1 的两倍。默认情况下，权重为 1；	10

3 负载均衡配置

3.1 负载均衡器管理

本章节提供查询和删除负载均衡器的操作步骤。当您需要查看某负载均衡器详情或不再使用该负载均衡器时可参考本章节。

3.1.1 查询负载均衡器

在控制中心的【弹性负载均衡】界面的信息列表，可以查看已创建负载均衡器的状态、子网等详细信息。

1. 登录天翼云控制中心；
2. 选择【网络】【弹性负载均衡】【负载均衡器】；
3. 在负载均衡器信息列表右上角的下拉框中，可设置通过名称、服务地址等参数搜索负载均衡器；



4. 单击负载均衡器名称，进入负载均衡器详情页面，查看负载均衡器的详细信息；

3.1.2 删除负载均衡器

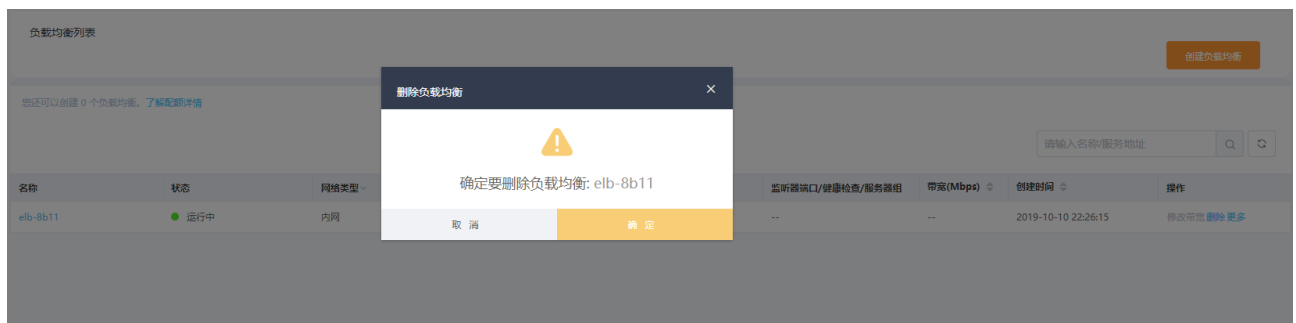
当您不需要再使用某个负载均衡器时，可删除该负载均衡器。

1. 登录天翼云控制中心；
2. 选择【网络】【弹性负载均衡】【负载均衡器】；

3. 在【负载均衡器】界面，单击负载均衡器所在行的【删除】按钮；
4. 在确认对话框单击【确定】；

【注 1】如果该负载均衡器下有监听器，不能删除，需先删除监听器后才可删除负载均衡器。

- 1、进行负载均衡删除操作。
- 2、弹出删除提醒框如下图，



3.2 监听器管理

本章节提供修改监听器和删除监听器的操作步骤。

3.2.1 修改监听器

1. 登录天翼云控制中心；
2. 选择【弹性负载均衡】【负载均衡器】；
3. 单击已创建的负载均衡器实例名称；
4. 在该负载均衡界面的【监听器】区域，单击监听器所在行的【修改】选项；



5. 在【修改监听器】界面，根据页面提示配置参数，参数说明请见“3.2.1 添加监听器”；

- 单击【确定】按钮；

3.2.2 删除监听器

- 登录天翼云控制中心；
- 选择【弹性负载均衡】【负载均衡器】；
- 单击已创建的负载均衡器实例名称；
- 在该负载均衡界面的【监听器】区域，单击监听器所在行的【删除】选项；

说明：如果该负载均衡器下有监听器，不能删除，需先删除监听器后才可删除负载均衡器。

3.3 后端云主机管理

本章节提供添加、修改监听器和删除监听器的操作步骤。当您需要向负载均衡器添加监听器，修改监听器的负载均衡模式、会话保持类型、健康检查配置或不再使用该监听器时可参考本章节。

3.3.1 添加后端云主机

本章节提供添加和移除后端云主机的操作步骤。当您需要将云主机添加至负载均衡器或将云主机从负载均衡器下移除时可参考本章节。

- 登录天翼云控制中心；
- 选择【弹性负载均衡】【负载均衡器】；
- 单击已创建的负载均衡器实例名称；
- 单击【添加后端云主机】选项；
- 选择需要和负载均衡关联的云主机，并配置参数；

各参数说明如下：

参数	说明	取值样例
端口	后端云主机的服务监听端口，取值范围[1-65535]；	123

参数	说明	取值样例
权重	后端虚拟机权重。权重值决定了后端云主机处理的请求的比例。例如，一个权重为 2 的云主机处理的请求数是权重为 1 的两倍。默认情况下，权重为 1；	10

- 单击【确定】按钮完成后端云主机添加；

3.3.2 移除后端云主机

- 登录天翼云控制中心；
- 选择【弹性负载均衡】【负载均衡器】；
- 单击已创建的负载均衡器实例名称；
- 在该负载均衡详情界面，选择【后端云主机】标签；
- 需要移除多个后端云主机时，可勾选云主机并单击列表上方的【移除】按钮；需要移除单个后端云主机，可单击列表中云主机所在行的【移除】按钮或勾选云主机并单击列表上方的【移除】按钮；
- 单击【确定】按钮；

4 常见问题

4.1 弹性负载均衡是什么？

弹性负载均衡（Elastic Load Balancing，简称 ELB）是将访问流量根据转发策略分发到后端多台弹性云主机的流量分发控制服务。弹性负载均衡可以通过流量分发扩展应用系统对外的服务能力，实现更高水平的应用程序容错性能。

用户通过基于浏览器、统一化视图的云计算管理图形化界面，可以创建 ELB，为服务配置需要监听的端口，配置云主机。消除单点故障，提高整个系统的可用性。

4.2 弹性负载均衡服务是否收费？

负载均衡服务本身不收费，但绑定弹性 IP 时会收取弹性 IP 及带宽的费用。

4.3 弹性负载均衡支持哪些转发方式？

当前 ELB（增强型）支持加权轮询、加权最少连接和源 IP 三种模式的转发规则。

- 轮询算法：按顺序依次将请求分发给不同的云主机。它用相应的权重表示云主机的处理性能，按照权重的高低以及轮询方式将请求分配给各云主机，相同权重的云主机处理相同数目的连接数。
- 最少连接：通过当前活跃的连接数来估计云主机负载情况的一种动态调度算法。
- 源算法：将请求的源 IP 地址作为散列键（HashKey），从静态分配的散列表找出对应的云主机。

4.4 弹性负载均衡（增强型）是否可以添加不同操作系统的云主机？

可以。ELB 本身不会限制后端的云主机使用哪种操作系统，只要您的 2 台云主机中的应用服务部

署是相同且保证数据的一致性即可。但是，我们建议您选择 2 台相同操作系统的云主机进行配置，以便您日后的管理维护。

4.5 单个用户支持保有多少个弹性负载均衡？

单个用户默认可创建 1 个增强型负载均衡器。如果需要创建更多弹性负载均衡器，请申请更高配额。

4.6 什么是配额？

为防止资源滥用，平台限定了各服务资源的配额，对用户的资源数量和容量做了限制。如果当前资源配额限制无法满足使用需要，您可以申请扩大配额。

4.7 如何配置内网或公网负载均衡？

创建一个增强型负载均衡，系统分配一个内网 IP，默认是内网负载均衡，如果为这个内网 IP 绑定一个公网的 IP，则可作为公网负载均衡。增强型负载均衡同时可支持内网、公网访问。

4.8 监听器是什么？

承担 ELB 具体的协议和端口配置，云主机协议和端口配置，监听策略配置。

4.9 什么是负载均衡协议（端口）？

增强型负载均衡系统支持 4 层（TCP）和 7 层（HTTP）协议的负载均衡，可通过具体提供的服务能力选择对应的协议以及该协议对外呈现的端口。

监听器协议	用途
TCP	TCP 的应用部署
HTTP	Web 应用

4.10 什么是云主机协议（端口）？

后端云主机自身提供的网络服务的协议以及协议的端口，如使用 windows 操作系统上安装的 IIS (webservice)，该服务默认的协议为 HTTP，端口为 80。

4.11 弹性负载均衡分配的弹性 IP 是否为独占？

负载均衡：分配的弹性公网 IP 支持解绑，解绑后的增强型负载均衡变成内网型负载均衡，解绑后的公网 IP 可被其他资源绑定。

4.12 删除弹性负载均衡有什么影响？

如果您的 ELB 服务地址（IP）已经正常解析到域名且对外提供服务，除非必要请不要删除您创建的 ELB 服务，删除了 ELB 服务以后相应的服务配置和服务地址（IP）将会被释放掉，数据一旦删除，不可恢复。如果您重新创建 ELB 服务，可以重新由系统重新给您分配一个新的服务地址（IP）。

4.13 ELB 支持什么类型的会话保持？

增强型负载均衡器支持 SOURCE_IP、HTTP_cookie、APP_COOKIE 三种会话保持类型。

4.14 如何获取来访者的真实 IP？

（1）七层服务

针对七层（HTTP 协议）服务，需要对应用服务器进行配置，然后使用 X-Forwarded-For 的方式获取来访者的真实 IP 地址。真实的来访者 IP 会被负载均衡放在 HTTP 头部的 X-Forwarded-For 字段，格式如下：

X-Forwarded-For：来访者真实 IP，代理服务器 1-IP，代理服务器 2-IP，...

当使用此方式获取来访者真实 IP 时，获取的第一个地址就是来访者真实 IP。

（2）四层服务

针对四层（TCP 协议）服务，需要配置 TOA 插件获取。需要配置 TOA 插件获取。

4.15 健康检查为什么会导致 ELB 会频繁向后端云主机发送探测请求？

ELB 是高可用集群部署的，集群内的所有的转发节点会同时向后端云主机发送探测请求，检查周期用户可配，健康检查会根据检查周期一直探测，所以每隔几秒会有访问。您可以通过配置健康检

查的周期来控制访问后端云主机的频率。。

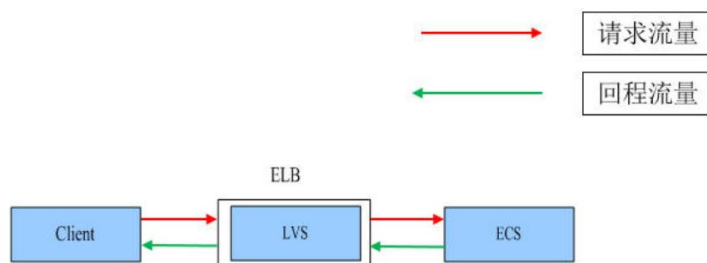
4.16 弹性负载均衡是否支持后端 FTP 服务？

弹性负载均衡不支持后端 FTP 服务。

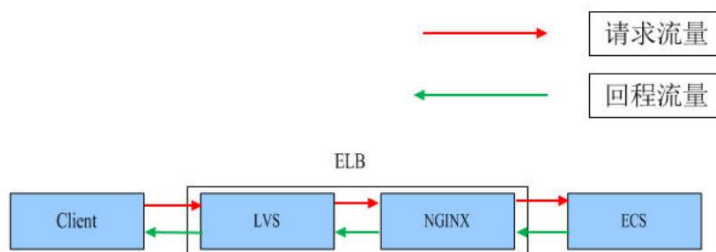
4.17 ELB 如何根据不同的协议来分发流量？

公网 ELB 采用 FullNAT 模式转发。如下图所示，四层转发转发经过 LVS，七层转发协议，经过 LVS 后再到 NGINX。

四层转发协议：



七层转发协议：



4.18 ELB 是否支持 TCP 长连接？

支持，客户端到 ELB 之间支持长连接，TCP 长连接是指客户端和 ELB 之间建立 TCP 连接之后，可以持续发送业务请求（HTTP 请求），可提高 TCP 连接复用率，降低 TCP 频繁建连的开销。

4.19 ELB 支持跨用户、跨 VPC 使用么？

当前 ELB 不支持跨用户使用，也不支持添加跨 VPC 的后端云主机。

4.20 如何检测压测性能上不去？

- 1、检查后端云主机的负载状态，如果 CPU 达到 100%，可能是后端应用达到性能瓶颈。
- 2、查看流量是否超过绑定到弹性负载均衡的 EIP 的带宽，带宽超限后，会有大量丢包和请求失败，影响压测性能。
- 3、如果是短连接测试，可能是客户端端口不足导致建立连接失败，可以通过客户端处于 `time_wait` 状态的连接数量来判断。可通过增加客户端 IP 来解决。
- 4、后端云主机的监听队列 `backlog` 满了，导致后端云主机不回复 `syn_ack` 报文，使得客户端连接超时。可以通过调整 `net.core.somaxconn` 参数来调大 `backlog` 的上限值。

4.21 如何检查请求不均衡？

- 1、检查是否开启了会话保持。如果配置了会话保持，而客户端的个数又比较少时，很容易导致不均衡。
- 2、检查后端云主机的健康检查状态是否正常，特别要关注下是否有健康检查状态一会正常一会异常的情况。健康检查异常或者状态切换都会导致流量不均衡。
- 3、检查负载均衡算法是否是源 IP 算法。此时同一个 IP 发过来的请求都会分发到同一个后端，导致流量不均衡。
- 4、后端服务是否开启了 TCP `keepalive` 保持长连接。如果开启，则有可能因为长连接上的请求数不同导致流量不均衡。
- 5、将云主机添加到 ELB 后端时是否设置了权重，权重不同，分发的流量也不同。

4.22 后端云主机什么时候被认为是健康的？（TCP、HTTP 健康检查）

首次添加的服务器健康检查成功一次就上线，后续按照配置的“最大重试次数”上线。

负载均衡器会定期向后端云主机发送请求以测试其运行状态，这些测试称为健康检查。通过健康检查来判断后端云主机是否可用。负载均衡器如果判断后端云主机健康检查异常，就不会将流量

分发到异常后端云主机，而是分发到健康检查正常的后端云主机，从而提高了业务的可靠性。当异常的后端云主机恢复正常运行后，负载均衡器会将其自动恢复到负载均衡服务中，承载业务流量。

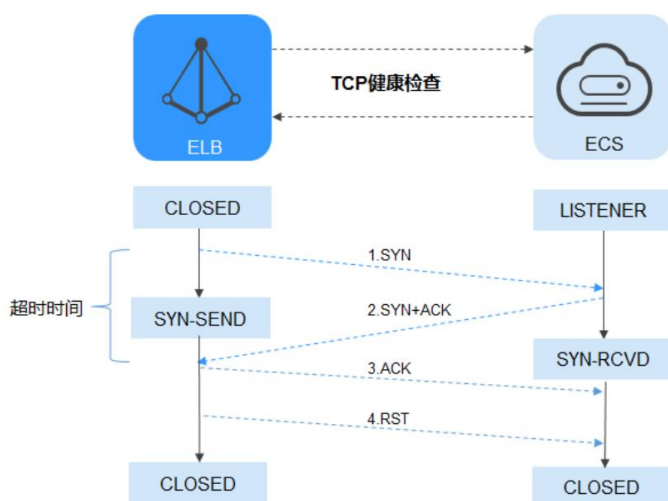
如果您的业务对负载比较敏感，过于频繁的健康检查报文可能会对您的正常业务产生影响。您可以根据实际的业务情况，通过增大健康检查周期，或者将七层健康检查改为四层健康检查等方式来降低对业务的影响。如果您的业务系统自身有健康检查机制，也可以关闭负载均衡器的健康检查，但是为了保障业务的持续可用，不建议这样做。

前提条件：

对于四层监听器，健康检查适用的版本是 HTTP 1.1；对于七层监听器，共享型负载均衡健康检查适用的版本是 HTTP 1.1，独享型适用 HTTP 1.0。

TCP 健康检查：

对于四层（TCP）和七层（HTTP/HTTPS）监听器，您可以配置 TCP 健康检查，通过发起 TCP 三次握手来获取后端云主机的状态信息，如图 1 所示。



TCP 健康检查的机制如下：

- 1、ELB 节点根据健康检查配置，向后端服务器（IP+健康检查端口）发送 TCP SYN 报文。
- 2、后端服务器收到请求报文后，如果相应的端口已经被正常监听，则会返回 SYN+ACK 报文。

3、如果在超时时间内没有收到后端服务器的 SYN+ACK 报文，则判定健康检查失败。然后发送 RST 报文给后端服务器中断 TCP 连接。

4、如果在超时时间内收到了 SYN+ACK 报文，则发送 ACK 给后端服务器，判定健康检查成功，并发送 RST 报文给后端服务器中断 TCP 连接。

须知：

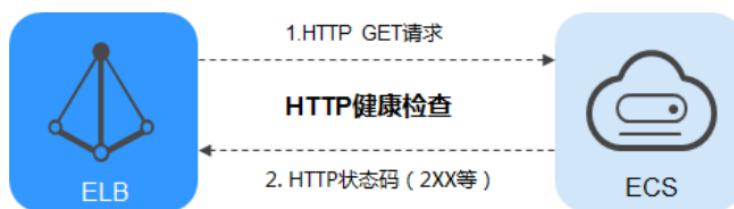
正常的 TCP 三次握手后，会进行数据传输，但是在健康检查时会发送 RST 中断建立的 TCP 连接。该实现方式可能会导致后端服务器中的应用认为 TCP 连接异常退出，并打印错误信息，如“Connection reset by peer”。解决方案如下：

- 1、采用 HTTP 方式进行健康检查。
- 2、后端服务器忽略健康检查的连接错误。

HTTP 健康检查：

对于四层（TCP）和七层（HTTP/HTTPS）监听器，您可以配置 HTTP 健康检查，通过 HTTP GET 请求来获取状态信息。对于 HTTPS 监听器，由于负载均衡器对 TLS 协议进行了卸载，负载均衡器与后端服务器之间使用 HTTP 传输，健康检查也采用 HTTP 方式，以提高系统的性能。健康原理如图 3 所示。

图 3 HTTP 健康检查



具体机制如下：

- 1、七层 ELB 节点根据健康检查配置，向后端服务器（IP+端口+检查路径）发出 HTTP GET 请求（可以选择设置域名）。

2、后端服务器收到请求后，根据服务的情况返回相应的 HTTP 状态码。

3、如果七层 ELB 节点在响应超时时间内收到了后端服务器的响应，将 HTTP 状态码与预置的状态码进行对比，如果匹配则认为健康检查成功，后端服务器运行正常。

4、如果七层 ELB 节点在响应超时时间内没有收到后端服务器的响应，则判定健康检查失败。

健康检查时间窗：

健康检查机制的引入，有效提高了业务服务的可用性。但是，为了避免频繁的健康检查失败引起的切换对系统可用性的冲击，健康检查只有连续多次检查成功或失败后，才会进行状态切换。

健康检查时间窗由以下三个因素决定：

- 1、间隔时间（每隔多久进行一次健康检查）
- 2、超时时间（等待服务器返回健康检查的时间）
- 3、最大重试次数（健康检查连续成功、失败的次数）。