弹性负载均衡

用户指南

文档版本 01

发布日期 2024-04-16

华为云计算技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目 录

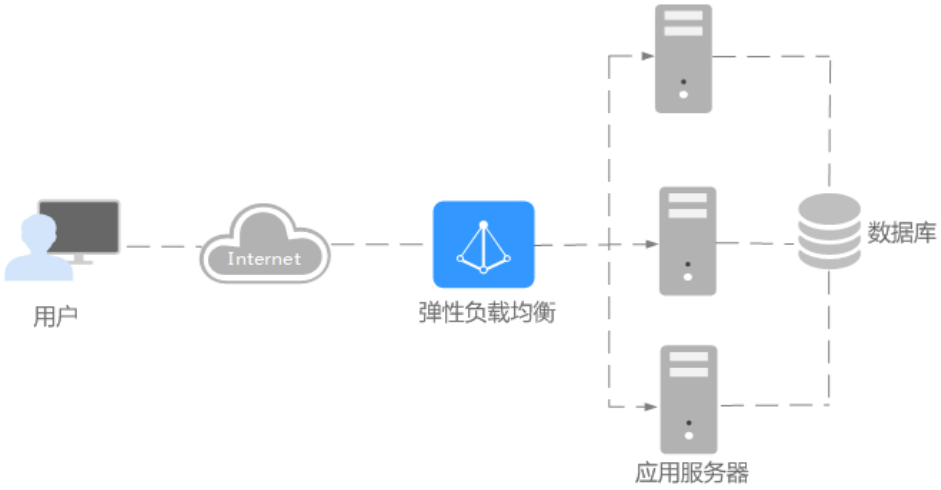
1 产品介绍

1.1 什么是弹性负载均衡

弹性负载均衡（Elastic Load Balance，简称ELB）是将访问流量根据分配策略分发到后端多台服务器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，同时通过消除单点故障提升应用系统的可用性。

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。

图 1-1 使用弹性负载均衡实例



弹性负载均衡的组件

弹性负载均衡由以下3部分组成：

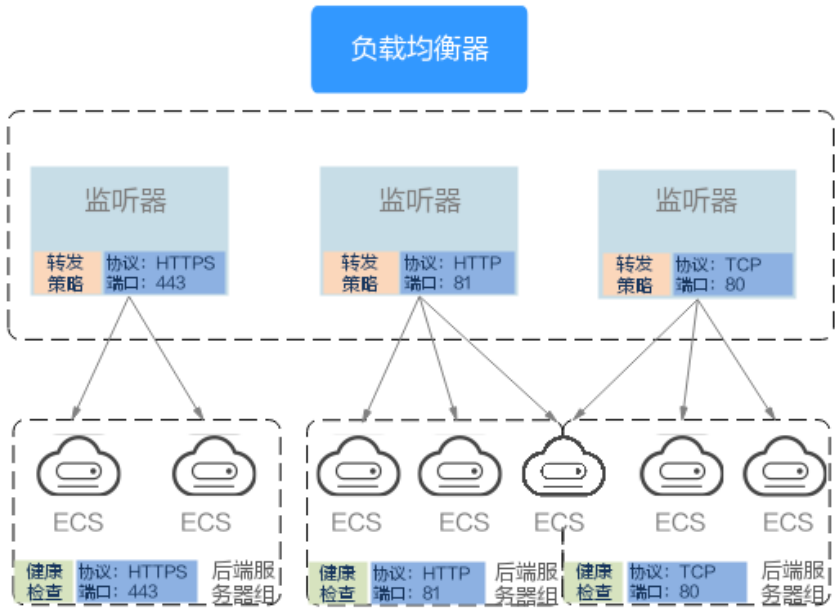
- **负载均衡器**：接受来自客户端的传入流量并将请求转发到一个或多个可用区中的后端服务器。
- **监听器**：您可以向您的弹性负载均衡器添加一个或多个监听器。监听器使用您配置的协议和端口检查来自客户端的连接请求，并根据您定义的分配策略和转发策略将请求转发到一个后端服务器组里的后端服务器。

- **后端服务器**：每个监听器会绑定一个后端服务器组，后端服务器组中可以添加一个或多个后端服务器。后端服务器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服务器。

  可以为后端服务器配置流量转发权重，不能为后端服务器组配置权重。

  您可以开启健康检查功能，对每个后端服务器组配置运行状况检查。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服务器上；而当该后端服务器恢复正常运行时，弹性负载均衡会将其自动恢复到弹性负载均衡服务中。

图 1-2 弹性负载均衡组件图



如何访问弹性负载均衡

可以使用以下方式访问和管理弹性负载均衡：

- 管理控制台

  请使用管理控制台方式访问弹性负载均衡。可直接登录管理控制台，从主页选择“弹性负载均衡”。

- 查询API

  通过调用API的方式访问弹性负载均衡，具体操作请参见《弹性负载均衡API参考》。

1.2 产品优势

独享型负载均衡的优势

- **超高性能**

  可实现性能独享，资源隔离，单实例单AZ最高支持2千万并发连接，满足用户的海量业务访问需求。

  选择多个可用区之后，对应的最高性能规格（新建连接数/并发连接数等）会加倍。例如：单实例单AZ最高支持2千万并发连接，那么单实例双AZ最高支持4千万并发连接。

说明

- 对于公网访问，会根据源IP的不同将流量分配到创建的多个AZ中的ELB上，多个AZ的ELB性能加倍。
- 对于内网访问：
  - 当从创建ELB的AZ发起访问时，流量将被分配到本AZ中的ELB上，当本AZ的ELB不可用时，容灾到创建的其他AZ的ELB上。

    如果本AZ的ELB正常，但是本AZ的流量超过规格，业务也会受影响，内网场景要考虑客户端访问的均衡性。内网流量使用率建议通过AZ粒度监控观察是否超限。
  - 当从未创建ELB的AZ访问时，会根据源IP的不同将流量分配到创建的多个AZ中的ELB上。
- 对于客户端跨VPC访问，流量优先分配至源VPC子网所在AZ部署的ELB，否则分配到其他AZ下的ELB。

- **超安全**

  支持TLS 1.3，提供全链路HTTPS数据传输，支持多种安全策略并支持创建自定义安全策略，根据业务不同安全要求灵活选择安全策略。
- **多协议**

  支持TCP/UDP/HTTP/HTTPS/QUIC协议，满足不同协议接入需求。
- **更灵活**

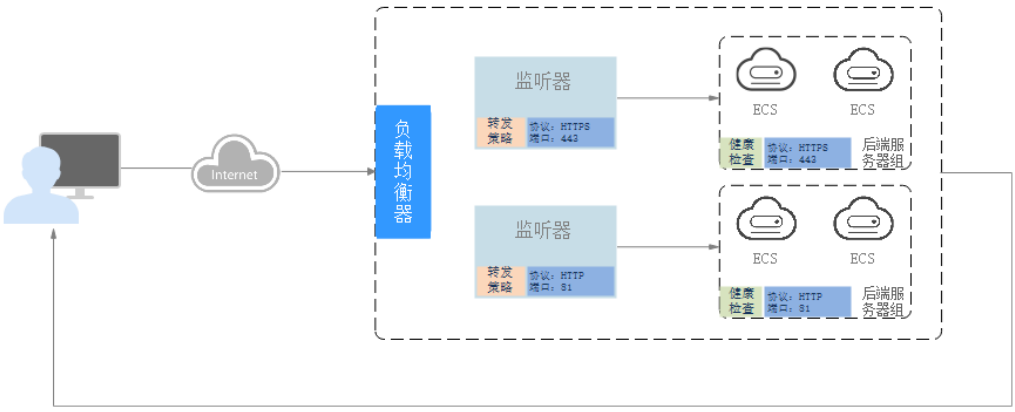
  支持请求方法、HEADER、URL、PATH、源IP等不同应用特征，并可对流量进行转发、重定向、固定返回码等操作。
- **无边界**

  提供混合负载均衡能力（跨VPC后端），可以将云上的资源和云下、多云之间的资源进行统一负载。
- **简单易用**

  快速部署ELB，实时生效，支持多种协议、多种调度算法可选，用户可以高效地管理和调整分发策略。

1.3 弹性负载均衡是如何工作的

图 1-3 ELB 工作原理图



弹性负载均衡的工作原理如下：

1. 客户端向您的应用程序发出请求。
2. 负载均衡器中的监听器接收与您配置的协议和端口匹配的请求。
3. 监听器再根据您的配置将请求转发至相应的后端服务器组。如果配置了转发策略，监听器会根据您配置的转发策略评估传入的请求，如果匹配，请求将被转发至相应的后端服务器组。
4. 后端服务器组中健康检查正常的后端服务器将根据分配策略和您在监听器中配置的转发策略的路由规则接收流量，处理流量并返回客户端。

请求的流量分发与负载均衡器所绑定的监听器配置的转发策略和后端服务器组配置的**分配策略**类型相关。

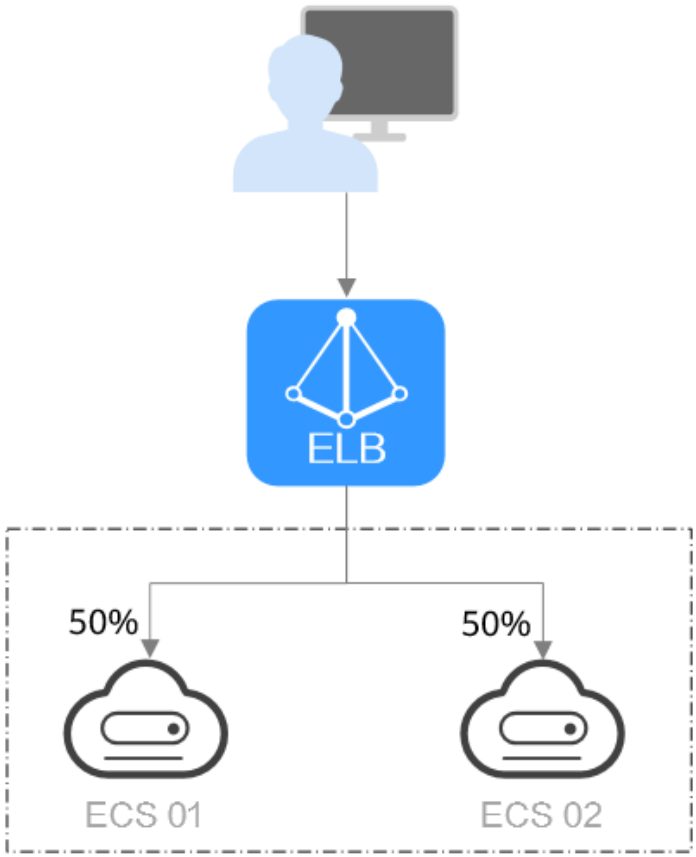
分配策略类型

负载均衡支持以下三种分配策略：

- 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器。它用相应的权重表示服务器的处理性能，按照权重的高低以及轮询方式将请求分配给各服务器，权重大的后端服务器被分配的概率高。相同权重的服务器处理相同数目的连接数。常用于短连接服务，例如HTTP等服务。

  图1-4展示弹性负载均衡器使用加权轮询算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，负载均衡器节点会将50%的客户端流量分发到其可用区中的每一台后端服务器。

  图 1-4 加权轮询算法流量分发

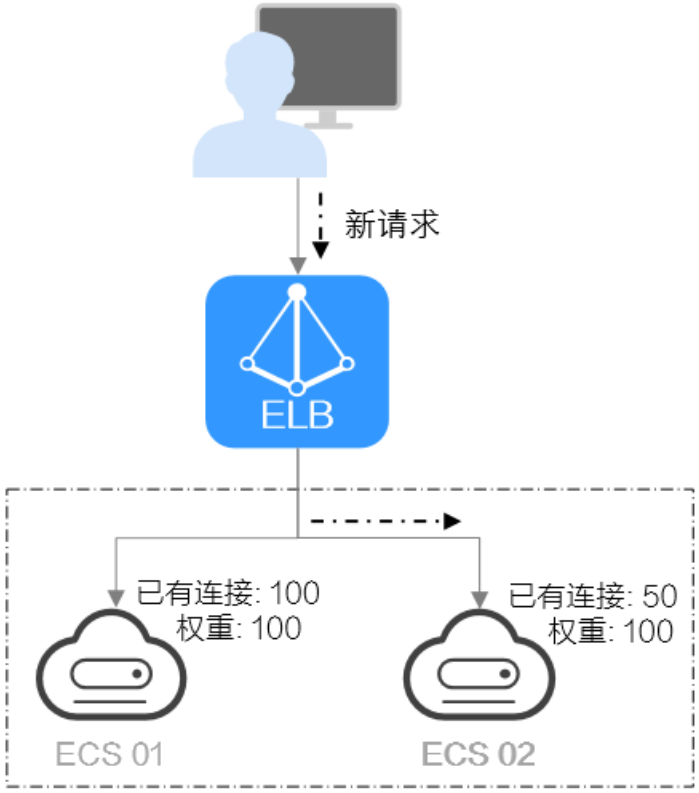


- 加权最少连接：最少连接是通过当前活跃的连接数来估计服务器负载情况的一种动态调度算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处

理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。常用于长连接服务，例如数据库连接等服务。

图1-5展示弹性负载均衡器使用加权最少连接算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，ECS 01已有100个连接，ECS 02已有50个连接，则新的连接会优先分配到ECS 02上。

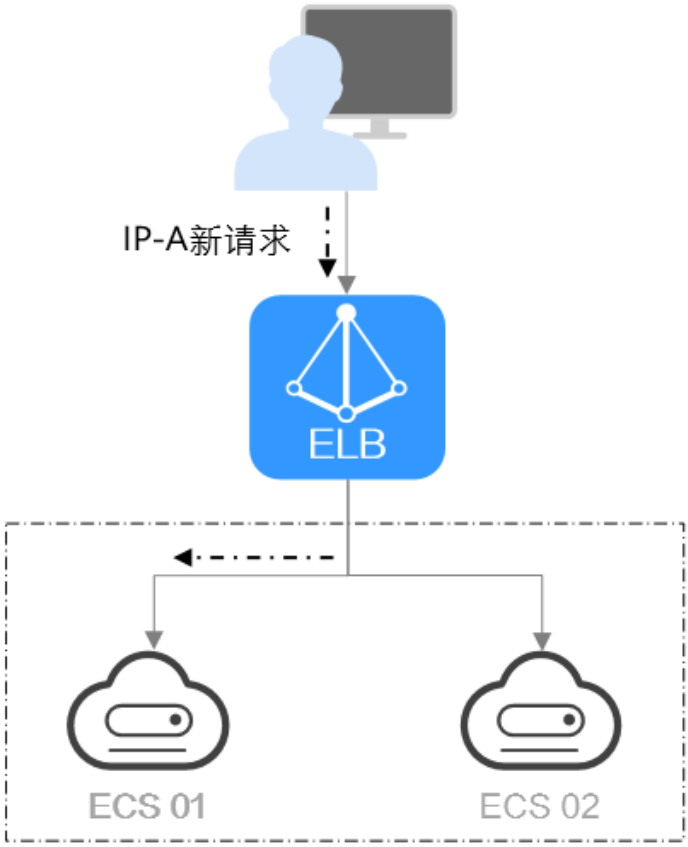
图 1-5 加权最少连接算法流量分发



- 源IP算法：将请求的源IP地址进行一致性Hash运算，得到一个具体的数值，同时对后端服务器进行编号，按照运算结果将请求分发到对应编号的服务器上。这可以使得对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某特定的服务器。该方式适合负载均衡无cookie功能的TCP协议。

  图1-6展示弹性负载均衡器使用源IP算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，ECS 01已经处理了一个IP-A的请求，则IP-A新发起的请求会自动分配到ECS 01上。

图 1-6 源 IP 算法流量分发



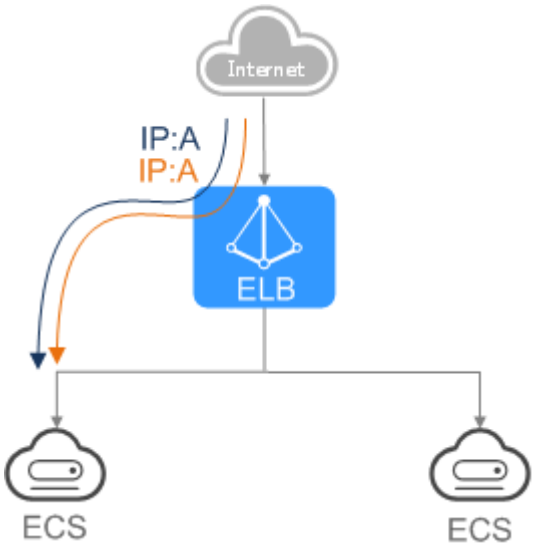
1.4 应用场景

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设置相应的分配策略，将访问量均匀的分到多个后端服务器处理。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服务器。从而提升访问效率，如图1-7所示。

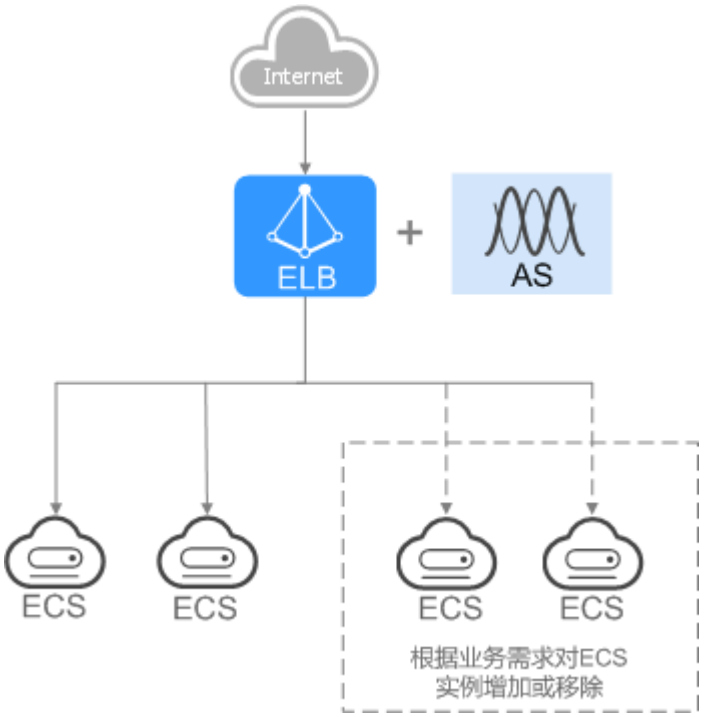
图 1-7 会话保持流量分发



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的后端服务器组或者从ELB的后端服务器组移除。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1-8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。

图 1-8 灵活扩展

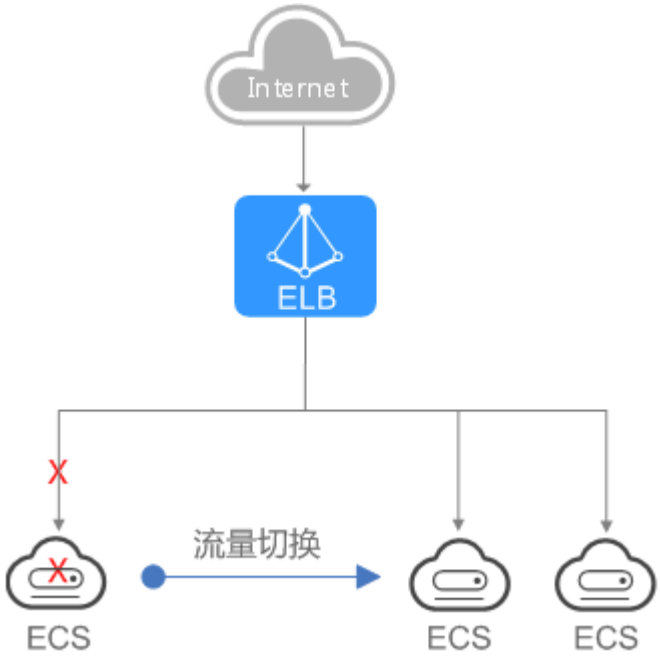


使用 ELB 消除单点故障

对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服务器。负载均衡器会通过健康检查及时发现并屏蔽有故障的服务器，并将流量转发到其他正常运行的后端服务器，确保业务不中断，如图1-9所示。

例如官网，计费业务，Web业务等。

图 1-9 消除单点故障

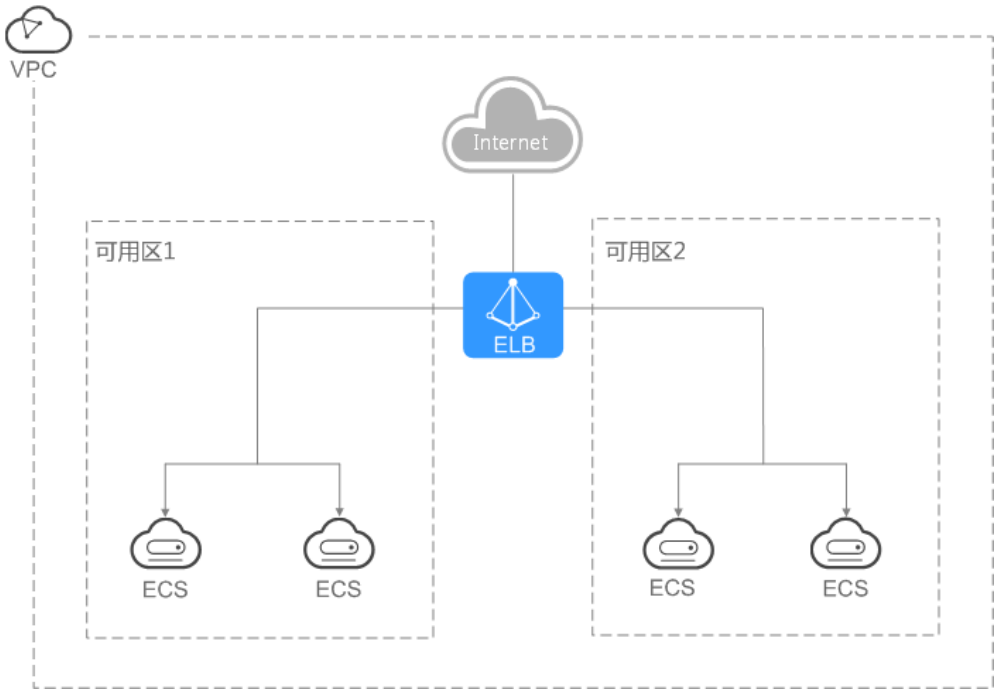


使用 ELB 跨可用区特性实现业务容灾部署

对可靠性和容灾有很高要求的业务，弹性负载均衡可将流量跨可用区进行分发，建立实时的业务容灾部署。即使出现某个可用区网络故障，负载均衡器仍可将流量转发到其他可用区的后端服务器进行处理，如图1-10所示。

例如银行业务，警务业务，大型应用系统等。

图 1-10 多可用区部署



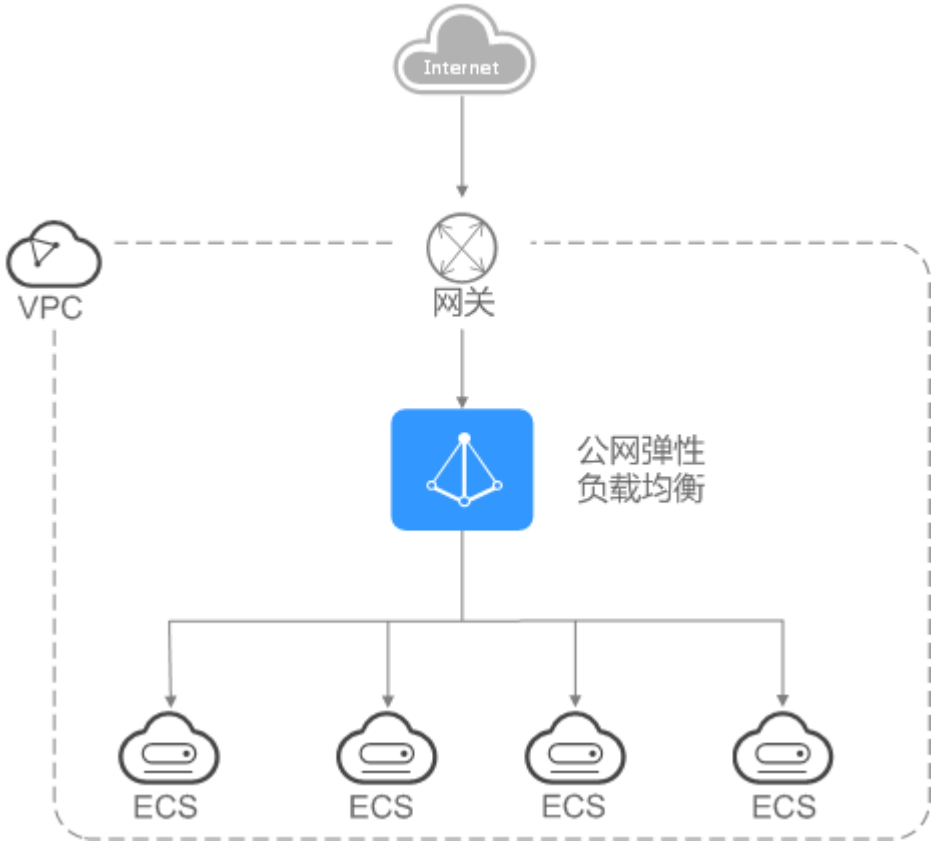
1.5 公网和私网负载均衡器

负载均衡按照**支持的网络类型的**不同分为**公网负载均衡器**和**私网负载均衡器**。

公网负载均衡器

通过给负载均衡器绑定弹性公网IP，使其支持转发公网流量请求，称为公网负载均衡器。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服务器进行处理。

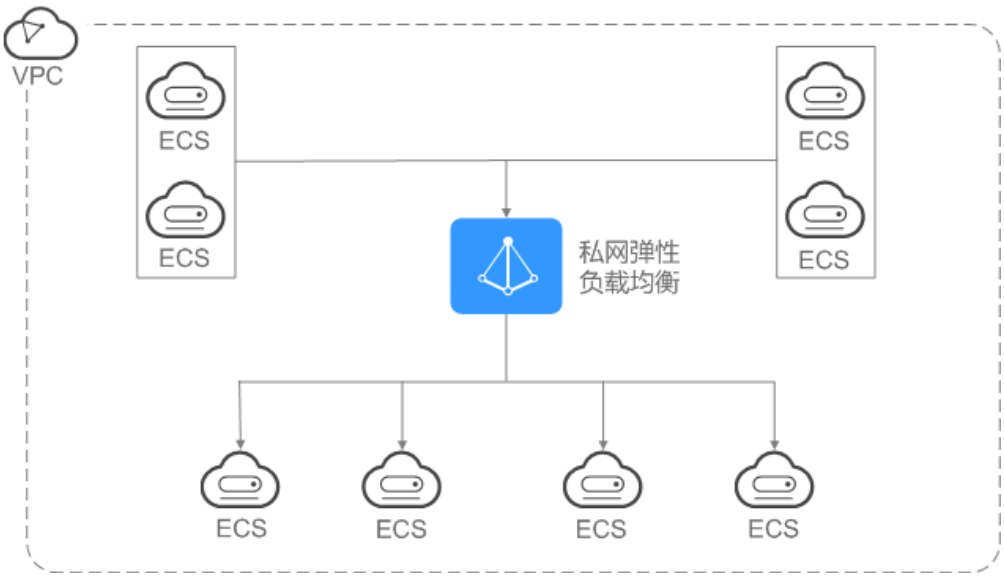
图 1-11 公网负载均衡器



私网负载均衡器

通过给负载均衡器绑定弹性私网IP，使其支持转发私网流量请求，称为私网负载均衡器。通过私网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服务器进行处理。

图 1-12 私网负载均衡器



1.6 ELB 网络流量路径说明

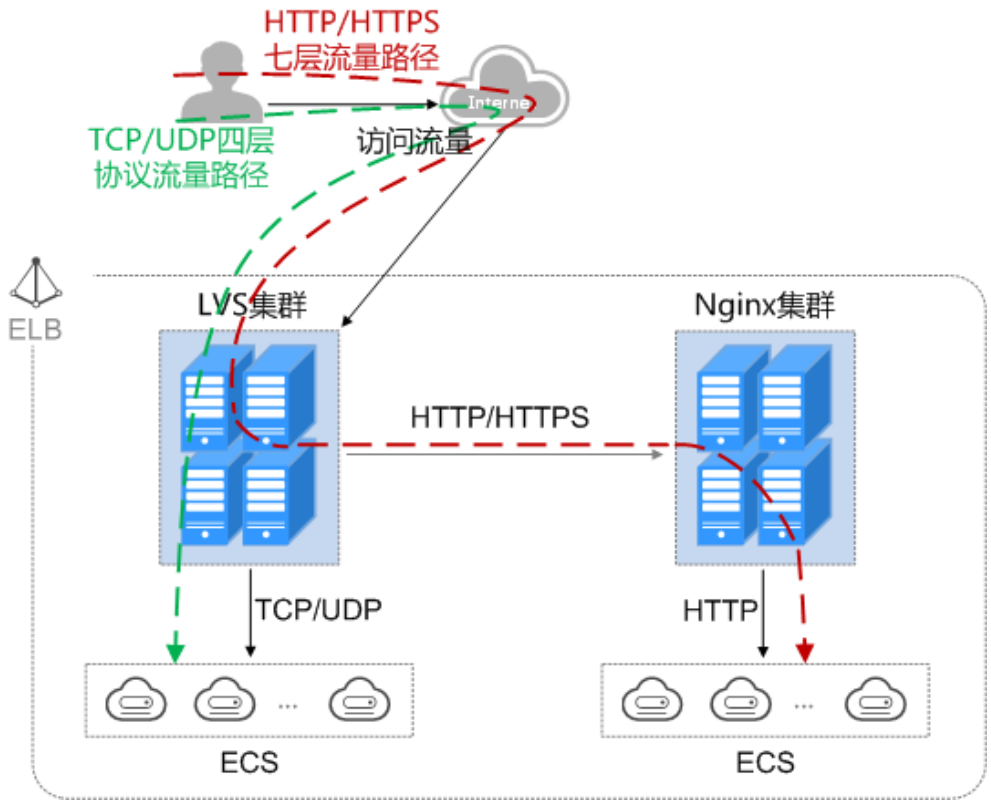
负载均衡将来自客户端的请求通过负载均衡器分发至后端服务器，后端服务器再将响应通过内网返回给负载均衡。负载均衡器和后端服务器之间是通过内网进行通信的。

- 如果负载均衡器后端服务器仅处理来自负载均衡的访问请求，服务器可以不购买EIP或者NAT网关等服务，仅有私网IP即可。
- 如果负载均衡器后端服务器还需要直接对公网提供服务，或者需要访问公网资源，则服务器需要购买EIP或者NAT网关等服务。

入网流量路径

对于入网流量，负载均衡会根据用户配置的流量分配策略，对来自公网或者私网的访问请求进行转发和处理。如图1-13所示。

图 1-13 入网流量路径



当负载均衡器使用四层协议TCP/UDP时：

- 四层协议TCP/UDP的流量只经过LVS集群进行转发。
- LVS集群的所有节点会根据负载均衡器的流量分配策略，将接收到的访问请求直接分发到后端服务器。

当负载均衡器使用七层协议HTTP/HTTPS时：

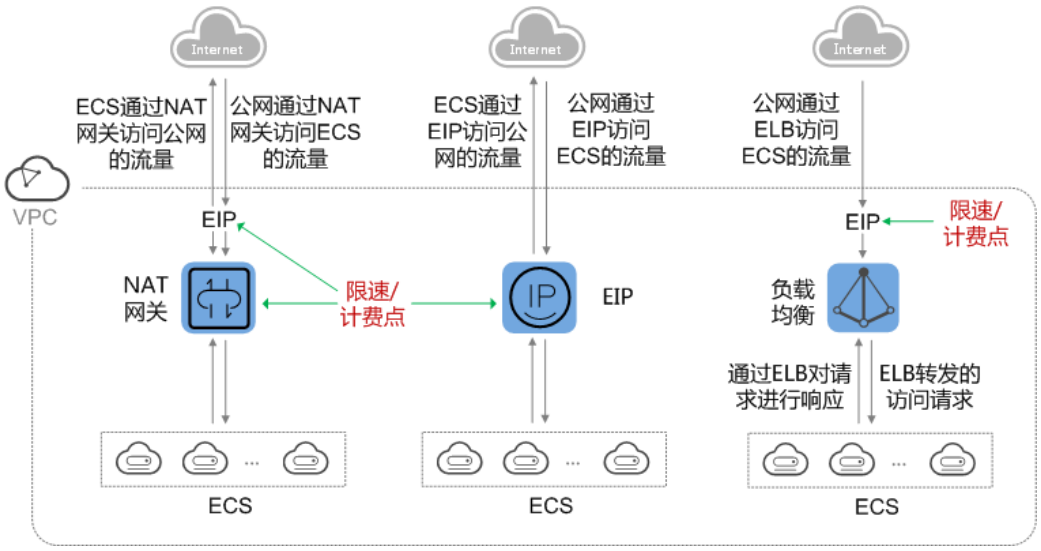
- 七层协议HTTP/HTTPS的流量，需要经过LVS集群先将访问请求平均分发到Nginx集群的所有节点，然后Nginx集群的节点再根据负载均衡器的转发策略，将接收到的请求最终分发到后端服务器。

- 七层协议HTTPS的流量，在最终分发到服务器前，还需要在Nginx集群内进行证书验证以及数据包的解密操作。然后通过HTTP协议将请求分发到后端服务器。

出网流量路径

出网流量遵循请求从哪进来，响应从哪出去的原则。如**出网流量路径**所示。

图 1-14 出网流量路径



- 通过负载均衡器进入的访问流量，对应的响应流量通过负载均衡器返回。

  由于负载均衡器实际是通过绑定的EIP接收来自公网的流量和响应请求，所以负载均衡器的限制实际是在负载均衡器绑定的EIP 上，并在EIP上进行计费。从负载均衡到后端云服务器之间通过VPC内网进行通信，不收取费用。
- 通过NAT网关进入的访问流量，对应的响应流量通过NAT网关返回。在NAT网关上限速和计费。

  由于NAT网关实际是通过绑定的EIP接收来自公网的流量和访问公网，所以NAT网关上进行的是连接数的限制，带宽或者流量的限制是在NAT网关绑定的EIP上，并分别在NAT网关和弹性公网IP上进行计费。
- 通过EIP进入的访问流量，对应的响应流量通过EIP返回，在EIP上限速和计费。

1.7 约束与限制

弹性负载均衡包含独享型负载均衡和共享型负载均衡，本文为您介绍弹性负载均衡资源的使用限制。

弹性负载均衡的服务配额

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。如您最多可以创建多少台弹性云服务器、多少块云硬盘。

默认资源配额如**表1-1**，不同用户拥有的实际资源配额略有差异。

表 1-1 弹性负载均衡的服务配额

资源名称	资源说明	默认配额
弹性负载均衡	一个用户创建弹性负载均衡的数量	50个
弹性负载均衡监听器	一个用户创建监听器的数量	100个
弹性负载均衡转发策略	一个用户创建转发策略的数量	500条
弹性负载均衡后端主机组	一个用户创建转发后端服务器组的数量	500个
弹性负载均衡证书	一个用户拥有弹性负载均衡证书的数量	120个
弹性负载均衡后端服务器	一个用户拥有后端服务器的数量	500个
单负载均衡器可添加监听器数量	一个负载均衡器支持添加监听器的数量	50个

说明

以上配额说明针对单租户情况。

弹性负载均衡的资源配额

除**弹性负载均衡的服务配额**外，弹性负载均衡的使用中还存在部分资源配额限制。

表 1-2 弹性负载均衡的资源配额

资源名称	资源说明	默认配额
单转发策略可添加的转发条件数量	一条转发策略支持的转发条件数量	10个
单后端服务器组可添加的后端服务器数量	一个后端服务器组支持添加后端服务器的数量	500个
IP地址组		
弹性负载均衡IP地址组	一个用户创建IP地址组的数量	50个
单IP地址组可关联监听器	一个IP地址组可关联监听器的数量	50个
单IP地址组可添加IP地址数量	一个IP地址组支持添加IP地址的数量	300个

负载均衡器

- 负载均衡器对转发数据的限制：

  - 四层监听器：无限制。
  - 七层监听器：
    - 上传文件大小限制为10GB。
    - HTTP请求行加HTTP请求头之和限制为32KB。

监听器

- 独享型负载均衡的监听器最多与50个后端服务器组关联使用。
- 一个HTTPS监听器最多支持配置30个SNI证书。
- 监听器的前端协议和端口设置后不允许修改。

转发策略

- 仅HTTP和HTTPS协议的监听器支持配置转发策略。
- 不支持创建相同的转发策略。
- 单个监听器最多支持配置100条转发策略，超过配额的转发策略不生效。
- 转发条件数量：
  - 未开启高级转发策略：一种转发规则仅支持一个转发条件
  - 高级转发策略：一种转发规则支持多个转发条件，一条转发策略最多支持10个转发条件。

表 1-3 弹性负载均衡的转发策略限制

实例类型	高级转发策略	转发规则	转发动作
独享型负载均衡	未开启高级转发策略	域名、URL	转发至后端服务器组、重定向至监听器
	开启高级转发策略	域名、URL、HTTP请求方法、HTTP请求头、查询字符串、网段	转发至后端服务器组、重定向至监听器、重定向至URL、返回固定响应

后端服务器组

仅前端协议与后端协议匹配的监听器和后端服务器组才可关联使用，协议匹配关系详见**表4 前端/后端协议匹配关系**。

表 1-4 前端/后端协议匹配关系

监听器的前端协议	后端服务器组的后端协议
TCP	TCP
UDP	• UDP • QUIC

监听器的前端协议	后端服务器组的后端协议
HTTP	HTTP
HTTPS	• HTTP • HTTPS

后端服务器

开启**获取客户端IP**功能后，不支持同一台服务器既作为后端服务器又作为客户端的场景。

TLS 安全策略

一个用户可以创建50个自定义TLS安全策略。

1.8 权限管理

如果您需要对云上创建的弹性负载均衡（Elastic Load Balance，简称ELB）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，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制云上资源的访问。

通过IAM，您可以在云账号中给员工创建IAM用户，并使用策略来控制他们对云上资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有ELB的使用权限，但是不希望他们拥有删除负载均衡器等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用负载均衡器，但是不允许删除负载均衡器的权限策略，控制他们对ELB资源的使用范围。

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ELB服务的其它功能。

IAM是提供权限管理的基础服务。关于IAM的详细介绍，请参见《IAM产品介绍》。

ELB 权限

默认情况下，账号管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的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。

ELB部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域对应的项目中设置相关权限，并且该策略仅对此项目生效，如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问ELB时，需要先切换至授权区域。

根据授权精细程度分为角色和策略。

- 角色：IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于云上各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。

- 策略：IAM最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。例如：针对ELB服务，账号管理员能够控制IAM用户仅能对某一类云服务器资源进行指定的管理操作。多数细粒度策略以API接口为粒度进行权限拆分，ELB支持的API授权项请参见《弹性负载均衡接口参考》。

如表1-5所示，包括了ELB的所有系统权限。

表 1-5 ELB 系统权限

系统角色/策略名称	描述	类型
ELB FullAccess	操作权限：对弹性负载均衡服务的所有执行权限。 作用范围：项目级服务。	系统策略
ELB ReadOnly Access	操作权限：对弹性负载均衡服务的只读权限。 作用范围：项目级服务。	系统策略
ELB Administrator	操作权限：对弹性负载均衡服务的所有执行权限。拥有该权限的用户必须同时拥有Tenant Administrator、VPC Administrator、CES Administrator、Server Administrator、Tenant Guest权限。 作用范围：项目级服务。 **说明** 如果账号已经申请开通细粒度权限，设置ELB系统权限时请配置细粒度策略，不要配置ELB Administrator策略。	系统角色

表1-6列出了ELB常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 1-6 常用操作与系统策略的关系

操作	ELB FullAccess	ELB ReadOnlyAccess	ELB Administrator
创建负载均衡器	√	×	√
查询负载均衡器	√	√	√
查询负载均衡器状态树	√	√	√
查询负载均衡器列表	√	√	√
更新负载均衡器	√	×	√
删除负载均衡器	√	×	√
创建监听器	√	×	√

操作	ELB FullAccess	ELB ReadOnlyAccess	ELB Administrator
查询监听器	√	√	√
修改监听器	√	×	√
删除监听器	√	×	√
创建后端服务器组	√	×	√
查询后端服务器组	√	√	√
修改后端服务器组	√	×	√
删除后端服务器组	√	×	√
创建后端服务器	√	×	√
查询后端服务器	√	√	√
修改后端服务器	√	×	√
删除后端服务器	√	×	√
创建健康检查	√	×	√
查询健康检查	√	√	√
修改健康检查	√	×	√
关闭健康检查	√	×	√
创建弹性公网IP	×	×	√
绑定弹性公网IP	×	×	√
查询弹性公网IP	√	√	√
解绑弹性公网IP	×	×	√
查看监控指标	×	×	√
查看访问日志	×	×	√

说明

- 解绑弹性公网IP，还需要配置VPC服务的vpc:bandwidths:update和vpc:publicIps:update细粒度权限，具体详见《虚拟私有云API参考》。
- 查看监控指标，还需要配置CES服务的CES ReadOnlyAccess权限，具体详见《云监控服务API参考》。
- 查看访问日志，还需要配置LTS服务的LTS ReadOnlyAccess权限，具体详见《云日志服务API参考》。

1.9 基本概念

1.9.1 产品基本概念

表 1-7 弹性负载均衡基本概念

名词	说明
负载均衡器	负载均衡器是指您创建的承载业务的负载均衡服务实体。
监听器	监听器负责监听负载均衡器上的请求，根据配置的流量分配策略，分发流量到后端云服务器处理。
后端服务器	负载均衡器会将客户端的请求转发给后端服务器处理。例如，您可以添加ECS实例作为负载均衡器的后端服务器，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，并根据您定义的分配策略将请求转发到后端服务器组里的后端服务器。
后端服务器组	把具有相同特性的后端服务器放在一个组，负载均衡实例进行流量分发时，流量分配策略以后端服务器组为单位生效。
健康检查	负载均衡器会定期向后端服务器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服务器是否可用。负载均衡器如果判断后端服务器健康检查异常，就不会将流量分发到异常后端服务器，而是分发到健康检查正常的后端服务器，从而提高了业务的可靠性。当异常的后端服务器恢复正常运行后，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。
重定向	HTTPS是加密数据传输协议，安全性高，如果您需要保证业务建立安全连接，可以通过负载均衡的HTTP重定向功能，将HTTP访问重定向至HTTPS。
会话保持	会话保持，指负载均衡器可以识别客户与服务器之间交互过程的关联性，在实现负载均衡的同时，保持将其他相关联的访问请求分配到同一台后端服务器上。
WebSocket	WebSocket (WS)是HTML5一种新的协议。它实现了浏览器与服务器全双工通信，能更好地节省服务器资源和带宽并达到实时通讯。WebSocket建立在TCP之上，同HTTP一样通过TCP来传输数据，但是它和HTTP最大不同在于，WebSocket是一种双向通信协议，在建立连接后，WebSocket服务器和Browser/Client Agent都能主动的向对方发送或接收数据，就像Socket一样；WebSocket需要类似TCP的客户端和服务器端通过握手连接，连接成功后才能相互通信。
SNI	您需要在创建HTTPS监听器时开启SNI功能。SNI（Server Name Indication）是为了解决一个服务器使用域名证书的TLS扩展，开启SNI之后，用户需要添加域名对应的证书。开启SNI后，允许客户端在发起SSL握手请求时就提交请求的域名信息，负载均衡收到SSL请求后，会根据域名去查找证书，如果找到域名对应的证书，则返回该证书；如果没有找到域名对应的证书，则返回缺省证书。

名词	说明
长连接	长连接是指在一个连接上可以连续发送多个数据包，在连接保持期间，如果没有数据包发送，需要双方发链路检测包。
短连接	短连接是指通讯双方有数据交互时，就建立一个连接，数据发送完成后，则断开此连接，即每次连接只完成一项业务的发送。
并发连接	并发连接指客户端向服务器发起请求并建立了TCP连接的总和，负载均衡的并发连接是指每秒钟所能接收并处理的TCP连接总和。

1.9.2 区域和可用区

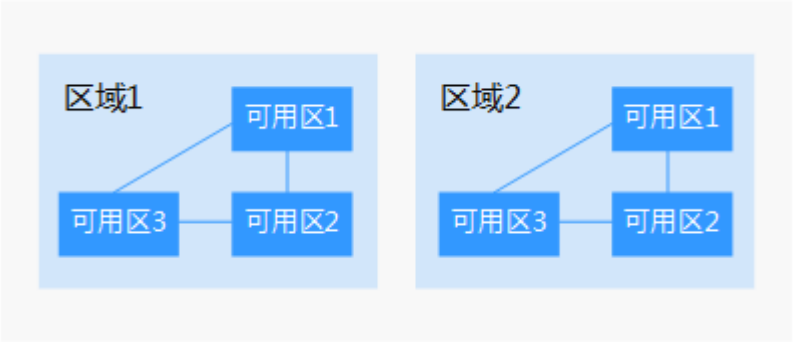
什么是区域、可用区?

区域和可用区用来描述数据中心的位置，您可以在特定的区域、可用区创建资源。

- 区域（Region）指物理的数据中心。每个区域完全独立，这样可以实现最大程度的容错能力和稳定性。资源创建成功后不能更换区域。
- 可用区（AZ，Availability Zone）是同一区域内，电力和网络互相隔离的物理区域，一个可用区不受其他可用区故障的影响。一个区域内可以有多个可用区，不同可用区之间物理隔离，但内网互通，既保障了可用区的独立性，又提供了低价、低时延的网络连接。

图1-15阐明了区域和可用区之间的关系。

图 1-15 区域和可用区



如何选择区域?

建议就近选择靠近您或者您的目标用户的区域，这样可以减少网络时延，提高访问速度。

如何选择可用区?

是否将资源放在同一可用区内，主要取决于您对容灾能力和网络时延的要求。

- 如果您的应用需要较高的容灾能力，建议您将资源部署在同一区域的不同可用区内。
- 如果您的应用要求实例之间的网络延时较低，则建议您将资源创建在同一可用区内。

区域和终端节点

当您通过API使用资源时，您必须指定其区域终端节点。有关的区域和终端节点的更多信息，请参阅**地区和终端节点**。

1.10 与其他服务的关系

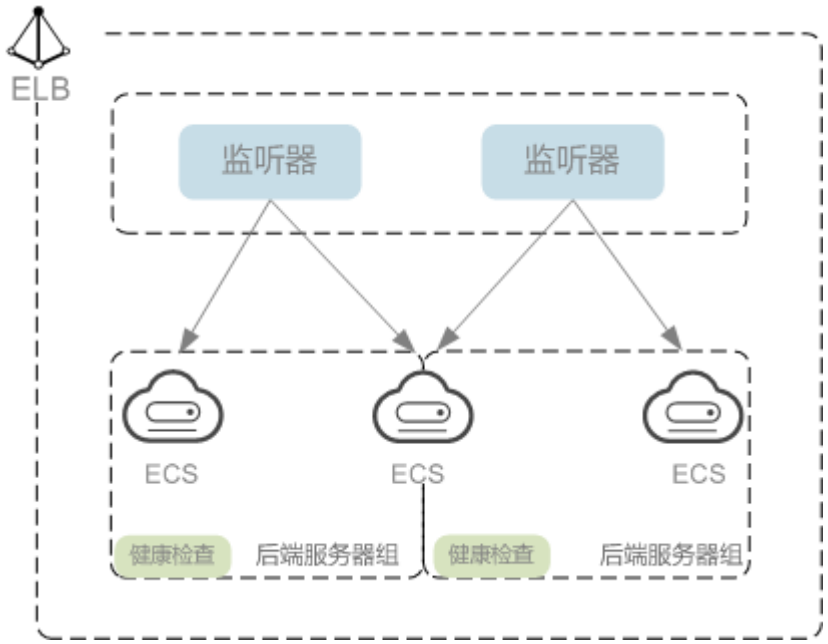
- 虚拟私有云（VPC）
  创建ELB时需要使用虚拟私有云服务创建的弹性IP、带宽。
- 弹性伸缩（AS）
  当配置了负载均衡服务后，弹性伸缩在添加和移除云服务器时，自动在负载均衡服务中添加和移除云服务器。
- 统一身份认证服务（IAM）
  需要统一身份认证提供鉴权。
- 弹性云服务器（Elastic Cloud Server，ECS）
  通过相关服务部署用户业务，并接收ELB分发的访问流量。
- 云日志服务（Log Tank Service，LTS）
  配置访问日志时需要您对接云日志服务，查看和分析对七层负载均衡HTTP和HTTPS进行请求的详细访问日志记录。
- 云监控（Cloud Eye）
  当用户开通了弹性负载均衡服务后，无需额外安装其他插件，即可在云监控查看对应服务的实例状态。

2 负载均衡器

2.1 什么是负载均衡器

负载均衡器是指您创建的承载业务的负载均衡服务实体。创建负载均衡器后，您还需要在负载均衡器中添加监听器和后端服务器，然后才能使用负载均衡服务提供的功能。

图 2-1 负载均衡器结构图



网络类型

按照网络类型分类，负载均衡器分为**公网负载均衡器**和**私网负载均衡器**。

- **公网负载均衡器：**负责处理来自公网访问请求分发的负载均衡实体。
  公网负载均衡器接收公网的访问请求，然后向绑定了监听器的后端服务器分发这些请求。创建公网负载均衡器时，需要为负载均衡器创建EIP或者绑定已有的EIP 。

使用场景

- 需要通过服务器集群对公网提供服务，且需要统一的入口，并将公网用户请求合理地分配到服务器集群时。
- 需要对服务器集群做故障容错和故障恢复时。

- **私网负载均衡器：**负责处理来自弹性负载均衡同一个VPC内访问请求的负载均衡实体。

  私网负载均衡器由于没有公网域名和EIP，所以只能在VPC内部被访问，不能被Internet的公网用户访问。私网负载均衡通过使用私有IP将来自同一个VPC内的访问请求分发到后端服务器上，通常用于内部服务集群。

  使用场景

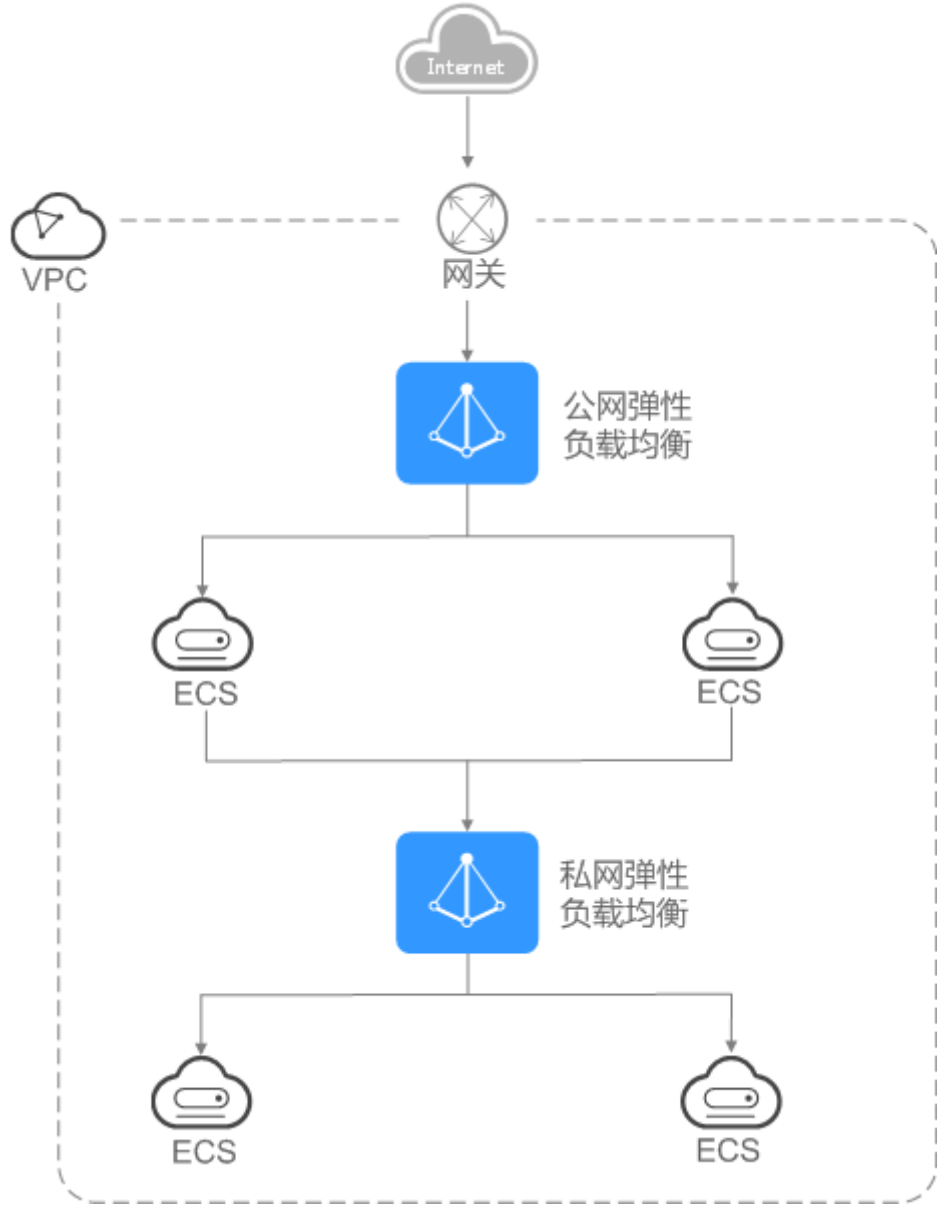
  私网负载均衡的客户端和服务器端均在云平台内部，通过VPC内网访问，主要场景如下：

  - 当内部服务器端有多台，需要将客户端请求合理地分发到各台服务器时；
  - 当需要对内部服务器集群做故障容错和故障恢复时；
  - 当用户想对外屏蔽自己的物理IP地址，对客户端提供透明化的服务时；

  同时使用公网负载均衡和私网负载均衡：

  例如，某业务Web服务器和数据库服务器分开部署，Web服务器需要对公网用户提供访问，后端的数据库服务器只能通过内网进行访问。该场景可以同时使用公网负载均衡器和私网负载均衡器，将Web服务器连接至公网负载均衡器，将相应的数据库服务器连接至私网负载均衡器。公网负载均衡器接收来自公网的请求并分发至后端Web服务器，处理后将对数据库的请求发送到私网负载均衡，再由私网负载均衡转发请求至数据库服务器。

图 2-2 同时使用公网负载均衡和私网负载均衡



2.2 规划和准备

在使用负载均衡前，需要根据业务规划待创建负载均衡器的区域、类型、协议以及后端服务器等。

规划实例区域

负载均衡器选择区域时需要注意以下事项：

- 选择距离业务目标客户距离最近的区域，可以减少网络时延以及提高下载速度。
- 独享型负载均衡可通过以下方式实现跨区域、跨VPC添加后端服务器：
  - 通过ELB的跨VPC后端功能实现跨VPC添加后端服务器，详见《 弹性负载均衡 用户指南 》。

规划实例可用区

独享型负载均衡支持多可用区，选择的每个可用区都会创建相应的负载均衡实例。

这些可用区的负载均衡实例间采用双活或者多活模式，客户端访问的请求就近分配到同可用区的实例。

弹性负载均衡可将客户端请求跨可用区分发，选择与后端服务器相同的可用区，可以减少网络时延以及提高访问速度。

如果**业务需要考虑容灾能力**，建议采取以下两种方式创建负载均衡实例：

- **单实例多可用区（可用区容灾）**

  对于业务量没有超过独享型负载均衡最大规格（大型 II）限制的，可以创建一个负载均衡实例，并选择多个可用区，这样单个可用区的负载均衡实例故障不会影响所有业务，多个可用区之间可以实现业务容灾。

- **多实例多可用区（实例容灾+可用区容灾）**

  对于超高业务量，超过独享型负载均衡最大规格（大型 II）限制的，可以创建多个负载均衡实例，并且每个负载均衡实例选择多个可用区，这样单个负载均衡实例故障不会影响所有业务，多个负载均衡实例和多个可用区之间均可以实现业务容灾。

说明

- 对于公网访问，会根据源IP的不同将流量分配到创建的多个AZ中的ELB上，多个AZ的ELB性能加倍。
- 对于内网访问：
  - 当从创建ELB的AZ发起访问时，流量将被分配到本AZ中的ELB上，当本AZ的ELB不可用时，容灾到创建的其他AZ的ELB上。

    如果本AZ的ELB正常，但是本AZ的流量超过规格，业务也会受影响，内网场景要考虑客户端访问的均衡性。内网流量使用率建议通过AZ粒度监控观察是否超限。
  - 当从未创建ELB的AZ访问时，会根据源IP的不同将流量分配到创建的多个AZ中的ELB上。
- 对于客户端跨VPC访问，流量优先分配至源VPC子网所在AZ部署的ELB，否则分配到其他AZ下的ELB。

选择网络类型

独享型负载均衡网络类型可以选择IPv4公网、IPv4私网和IPv6。

- 如果选择了IPv4公网，负载均衡实例会分配到一个IPv4的公网IP地址，可以处理来自Internet上IPv4公网的访问请求。
- 如果选择了IPv4私网，负载均衡实例会分配到一个IPv4的私网IP地址，可以处理来自VPC内部IPv4私网的访问请求。
- 如果选择了IPv6，负载均衡实例就会分配到一个IPv6的IP地址，可以处理来自VPC内部IPv6私网的访问请求，如果同时购买了公网带宽，则可以同时来自VPC内部IPv6私网的访问请求和来自Internet上IPv6公网的访问请求。

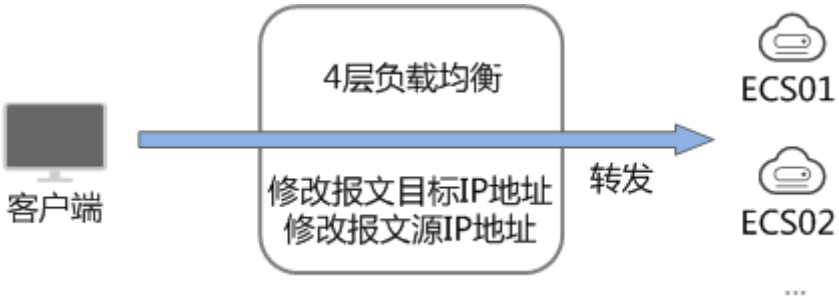
选择协议类型

提供基于四层协议和七层协议的负载均衡，在负载均衡器中通过加监听器选择相应的协议。

- 使用四层协议的负载均衡，监听器收到访问请求后，将请求直接转发给后端服务器。转发过程仅修改报文中目标IP地址和源IP地址，将目标地址改为后端云服务器

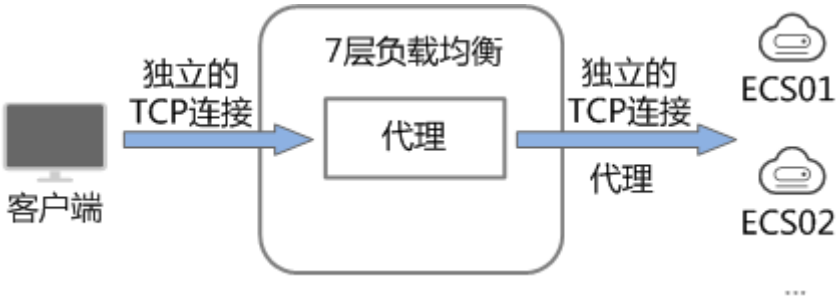
的IP地址，源地址改为负载均衡器的IP地址。四层协议连接的建立，即三次握手是客户端和后端服务器直接建立的，负载均衡只是进行了数据的转发。

图 2-3 四层负载均衡



- 使用七层协议的负载均衡，也称为“内容交换”。监听器收到访问请求后，需要识别并通过HTTP/HTTPS协议报文头中的相关字段，进行数据的转发。监听器收到访问请求后，先代理后端服务器和客户端建立连接（三次握手），接收客户端发送的包含应用层内容的报文，然后根据报文中的特定字段和流量分配策略判断需要转发的后端服务器。此场景中，负载均衡类似一个代理服务器，分别和客户端以及后端服务器建立连接。

图 2-4 七层负载均衡



后端服务器

在使用负载均衡器前，需要先创建ECS实例或者BMS实例并部署相关业务应用，然后将ECS实例或者BMS实例添加到负载均衡器的后端服务器组来处理转发的客户端访问请求。创建后端服务器时，请注意以下事项：

- 确保后端服务器实例的所属地域和负载均衡器的所属地域相同。
- 建议您选择相同操作系统的后端服务器实例作为后端服务器，以便后续管理和维护。

2.3 创建独享型负载均衡器

操作场景

在您创建独享型负载均衡器前，确保您已经做好了相关规划，详情参考**规划和准备**。

约束与限制

- 负载均衡器创建后，不支持修改VPC。如果要修改VPC，请重新创建负载均衡器，并选择对应的VPC。
- 独享型负载均衡实例创建完成后，您还需要创建监听器，才可以对负载均衡实例地址进行**ping**验证。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面单击“创建弹性负载均衡器”。根据界面提示选择负载均衡器的基础配置，配置参数如**表2-1**所示。

   表 2-1 负载均衡器的基础配置

参数	说明	取值样例
区域	不同区域的资源之间内网不互通。请选择靠近业务的区域，可以降低网络时延、提高访问速度。	-
可用区	可以选择在多个可用区创建负载均衡实例，提高服务的可用性。如果业务需要考虑容灾能力，建议选择多个可用区。当一个可用区出现故障或不可用时，业务可以快速切换到另一个可用区的负载均衡继续提供服务。 **说明** 针对已有实例，如果进行可用区配置修改，可能会导致该实例的业务闪断数秒，请做好规划，确实要修改的话建议选择闲时操作。	-
名称	负载均衡器的名称。	elb-test
描述	可添加负载均衡器相关描述。	-

5. 选定负载均衡器的规格后，请根据界面提示选择负载均衡器的网络配置，配置参数如**表2-2**所示。

表 2-2 负载均衡器的网络配置

参数	说明	取值样例
网络类型	可以单独选择一个网络类型，也可以同时选择多个。 ● IPv4公网：负载均衡器通过IPv4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服务器进行处理。 ● IPv4私网：负载均衡器通过IPv4私网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服务器进行处理。 ● IPv6公网私网：系统会为实例分配一个IPv6地址，转发来自IPv6客户端的请求。 **说明** 如果公网或私网IP均未选择，则ELB实例创建完成后无法与客户端通信。请在使用ELB或测试业务连通性时，务必确保该ELB绑定了公网或私网IP。	IPv4公网
所属VPC	负载均衡器所属虚拟私有云。 您可以选择使用已有的虚拟私有云网络，或者单击“查看虚拟私有云”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》。	vpc-test
前端子网	独享型负载均衡所在的子网，从该子网中分配ELB实例对外服务的IP地址。 根据配置的网络类型为ELB实例分配对应的IP地址： ● IPv4私网：前端子网作为IPv4子网为ELB实例下发IPv4私有地址。 ● IPv6公网私网：前端子网作为IPv6子网为ELB实例下发IPv6地址。 **说明** 当网络类型选择“IPv6公网私网”，且所选的VPC下无支持IPv6的子网时，请为已有子网开启IPv6或创建支持IPv6的子网。详见《虚拟私有云用户指南》。	subnet-test

参数	说明	取值样例
后端子网	负载均衡实例将使用后端子网中的IP地址与后端服务器建立连接。 ● 默认与前端子网保持一致 ● 可选择负载均衡器所属VPC下的其他子网 ● 添加子网 **说明** ● 负载均衡实例会占用后端子网中的部分IP地址，负载均衡实例的规格，可用区的数量和跨VPC功能的使用会影响占用IP地址的数量。实际占用IP地址的数量以您在控制台创建的负载均衡实例所占用的IP地址个数为准。 ● 应用型负载均衡器需要额外占用8-30个后端子网中的IP地址进行流量转发，具体占用地址数量与ELB集群规模有关，请以最终结果为准。如果多个ELB实例在同一集群且实例的后端子网相同，会复用占用的IP地址，以节省占用地址数量。	与前端子网保持一致
IPv4私网配置		
IPv4地址	选择IPv4地址的分配方式。 ● 自动分配IP地址：由系统自动分配IPv4地址。 ● 手动指定IP地址：手动指定IPv4地址。 **说明** 负载均衡器的IP地址不受所在后端子网ACL配置的限制，所以能够被客户端直接访问。建议您使用监听器的访问控制功能限制客户端访问负载均衡器。 详细请参考**访问控制策略**。	自动分配IP地址
IPv6网络配置		
IPv6地址	选择IPv6的IP地址的分配方式。 **说明** 负载均衡器的IP地址不受所在后端子网ACL配置的限制，所以能够被客户端直接访问。建议您使用监听器的访问控制功能限制客户端访问负载均衡器。 详细请参考**访问控制策略**。	自动分配IP地址
共享带宽	选择IPv6的共享带宽。 可以选择暂不设置共享带宽或选择已有的共享带宽或新建共享带宽。	暂不设置
IPv4公网配置		

参数	说明	取值样例
弹性公网IP	当网络类型勾选“IPv4公网”时，需要指定弹性公网IP。 ● 新创建：系统为弹性负载均衡实例新创建一个弹性公网IP。 ● 使用已有：为弹性负载均衡实例选择一个已有的弹性公网IP地址。	-
弹性公网IP类型	使用新创建弹性公网IP时，选择的弹性公网IP的链路类型。 全动态BGP：可以根据设定的寻路协议实时自动优化网络结构，以保证客户使用的网络持续稳定、高效。	全动态BGP
公网带宽	弹性公网IP使用的带宽类型。 可选“按带宽计费”或“按流量计费”或“加入共享带宽”。 ● 按带宽计费：指定带宽上限，按使用时间计费，与使用的流量无关。 ● 按流量计费：指定带宽上限，按实际使用的上行流量计费，与使用时间无关。 ● 加入共享带宽	加入共享带宽
带宽	指定具体的带宽上限。	100 Mbit/s

6. 确认配置信息，单击“提交”。

2.4 修改公网带宽

操作场景

当负载均衡器支持公网流量请求时（IPv4公网或IPv6），公网与负载均衡器之间的流量通过公网带宽进行访问，用户可以按照实际需求更改负载均衡实例关联的公网带宽。

说明

- 变更独享型负载均衡实例的公网带宽时，需考虑变更独享型负载均衡实例的规格，避免因负载均衡实例的带宽不足造成流量通过ELB时限速。
- 公网带宽为ELB绑定的EIP带宽，是客户端访问ELB时的最高流量限制。

修改公网带宽

弹性负载均衡在变更带宽的时候，访问流量不会中断。

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。

3. 单击页面左上角的 ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，待修改带宽的负载均衡器所在行的“操作”列，单击“更多”。
5. 单击“修改IPv4带宽”或“修改IPv6带宽”。
6. 在“修改带宽”区域，设置新的带宽大小，单击“下一步”。

   可以选择系统定义好的带宽也可以自定义带宽大小。自定义修改带宽的范围为1-2,000 Mbit/s。
7. 确认修改后的带宽大小，单击“提交”。

2.5 修改 IP 地址

操作场景

弹性负载均衡支持修改IPv4私有IP，可以将负载均衡当前使用IPv4私有IP修改为当前子网或者其他子网的目标IP地址。

说明

仅独享型负载均衡支持修改IP地址。

修改 IPv4 私有 IP

1. 登录管理控制台。
2. 在管理控制台左上角单击 图标，选择区域和项目。
3. 单击页面左上角的 ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”页面，需修改负载均衡器所在行，单击“更多 > 修改IPv4私有IP”。
5. 在“修改IPv4私有IP”对话框中，选择需要修改的目标IP所在子网，并设置目标IP地址。
   - 不同子网下修改IPv4地址，可以勾选“自动分配IPv4地址”，勾选后，系统会自动分配一个所选择子网的IPv4地址。
   - 同一子网下修改IPv4地址，必须指定IP，不支持自动分配。
6. 单击“确定”。

2.6 为实例绑定/解绑 IP

操作场景

可以根据业务需要为负载均衡实例绑定IP地址，或者将负载均衡实例已经绑定的IP地址进行解绑。

说明

- 解绑IPv4公网IP后，对应的弹性负载均衡器将无法进行IPv4公网流量转发。
- 解绑IPv4私有IP后，对应的弹性负载均衡器将无法基于IPv4私有IP进行私网流量转发。
- 解绑IPv6地址后，对应的弹性负载均衡器将无法基于IPv6地址进行流量转发，请谨慎操作。

绑定 IPv4 公网 IP

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，待修改的负载均衡器所在行，选择“更多 > 绑定IPv4公网IP”。
5. 在“绑定IPv4公网IP”对话框中，选择需要绑定的公网IP。
6. 单击“确定”。

绑定 IPv4 私有 IP

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，待修改的负载均衡器所在行，选择“更多 > 绑定IPv4私有IP”。
5. 在“绑定IPv4私有IP”对话框中，选择待绑定的IPv4地址所在子网，并设置目标IP地址。
   - 系统默认自动分配IP地址，如果需要手动指定IP地址，请去勾选“自动分配IPv4地址”，并在参数“IPv4地址”行输入目标IP地址。
   - 输入的IP地址必须属于所选择的子网且未被使用。
6. 单击“确定”。

解绑 IPv4 公网 IP

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，待修改的负载均衡器所在行，选择“更多 > 解绑IPv4公网IP”。
5. 在“解绑IPv4公网IP”对话框中，确认需要释放的IPv4公网IP地址，单击“是”。

说明

解绑IPv4公网IP后，对应弹性负载均衡器将无法进行IPv4公网流量转发，请谨慎操作。

解绑 IPv4 私有 IP

当前仅独享型负载均衡支持此功能。

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，待设置的负载均衡器所在行，选择“更多 > 解绑IPv4私有IP”。
5. 在“解绑IPv4私有IP”对话框中，确认需要释放的IPv4私有IP地址，单击“是”。

说明

解绑IPv4私有IP后，对应的弹性负载均衡器将无法基于IPv4私有IP进行私网流量转发，请谨慎操作。

解绑 IPv6 地址

当前仅独享型负载均衡支持此功能。

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”页面，待设置的负载均衡器所在行单击“更多 > 解绑IPv6地址”。
5. 在“解绑IPv6地址”对话框中，确认需要释放的IPv6地址，单击“是”。

说明

解绑IPv6地址后，对应的弹性负载均衡器将无法基于IPv6地址进行流量转发，请谨慎操作。

2.7 导出负载均衡器列表

操作场景

您可以选择导出弹性负载均衡器列表，作为本地备份数据查看。

导出弹性负载均衡器列表

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在负载均衡器列表左上方，单击“导出”，导出弹性负载均衡器列表。

2.8 删除负载均衡器

操作场景

当您确认负载均衡不需要继续使用时，您可以根据需求随时删除负载均衡器。

⚠ 注意

删除弹性负载均衡后无法恢复，请谨慎操作。

删除公网类型负载均衡器时，绑定的EIP不会被默认自动删除，不会影响EIP的正常使用。

前提条件

请先按以下顺序删除该负载均衡器配置的资源：

1. **转发策略：**如果监听器配置了转发策略，请删除所有转发策略。
2. **重定向：**如果配置了HTTP监听器重定向至HTTPS监听器，请删除所有重定向。
3. **后端服务器：**如果监听器对应的后端服务器组添加了后端服务器，请删除所有后端服务器。
4. **监听器：**请删除ELB下的所有监听器。
5. **后端服务器组：**请删除监听器对应的所有后端服务器组。

删除负载均衡器

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，选择目标负载均衡器所在行的操作列下的“更多 > 删除”。

   弹出删除确认对话框。根据实际需要勾选“释放该负载均衡绑定的弹性公网IP”。
5. 单击“是”。

3 监听器

3.1 什么是监听器

创建负载均衡器后，需要为负载均衡器配置监听器。监听器负责监听负载均衡器上的请求，根据配置流量分配策略，分发流量到后端服务器处理。

支持的协议类型

负载均衡提供四层协议和七层协议监听，您可根据从客户端到负载均衡器的应用场景选择监听协议，详细说明可参见**表3-1**。

对于支持**四层能力**的负载均衡器，在创建监听器时，支持选择**TCP**或者**UDP**。

对于支持**七层能力**的负载均衡器，在创建监听器时，支持选择**HTTP**或者**HTTPS**。

表 3-1 监听协议类型说明

协议类型		说明	适用场景
四层协议	TCP	• 基于源地址的会话保持。 • 数据传输快。	• 适用于注重可靠性，对数据准确性要求高的场景，如文件传输、发送或接收邮件、远程登录。 • 对性能和并发规模有要求的Web应用。
四层协议	UDP	• 可靠性相对低 • 数据传输快	适用于关注实时性而相对不注重可靠性的场景，如视频聊天、游戏、金融实时行情推送。
七层协议	HTTP	• 基于Cookie的会话保持。 • 使用X-Forward-For获取源地址。	需要对数据内容进行识别的应用，如Web应用、移动游戏等。

协议类型		说明	适用场景
七层协议	HTTPS	• 加密传输数据，可以阻止未经授权的访问。 • 加解密操作在负载均衡器上完成，可减少后端服务器的处理负载。 • 多种加密协议和加密套件可选。	需要加密传输的应用。

3.2 添加 TCP 监听器

操作场景

TCP协议适用于注重可靠性，对数据准确性要求高，速度可以相对较慢的场景，如文件传输、发送或接收邮件、远程登录等。您可以添加一个TCP监听器转发来自TCP协议的请求。

约束与限制

- 前端协议为“TCP”时，后端协议默认为“TCP”，且不支持修改。
- 如果您的独享型负载均衡实例类型为应用型（HTTP/HTTPS），则无法创建TCP监听器。

添加独享型负载均衡 TCP 监听器

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要添加监听器的负载均衡名称。
5. 切换到“监听器”页签，单击“添加监听器”，配置监听器。配置监听器参数参见表3-2。

表 3-2 独享型负载均衡配置监听器参数说明

参数	说明	示例
名称	监听器名称。	listener-pnqy
前端协议	客户端与负载均衡监听器建立流量分发连接的协议。 协议选择TCP。	TCP

参数	说明	示例
前端端口	客户端与负载均衡监听器建立流量分发连接的端口。 取值范围为：[1-65535]。	80
访问控制	支持通过白名单和黑名单进行访问控制，更多信息请参见**访问控制策略**： ● 允许所有IP访问 ● 黑名单 ● 白名单	黑名单
IP地址组	设置白名单或者黑名单时，必须选择一个IP地址组。如果还未创建IP地址组，需要先创建IP地址组，更多关于IP地址组的信息请参见**创建IP地址组**。	ipGroup-b2
获取客户端IP	开启此开关，后端服务器可以获取到客户端的真实IP地址。 独享型负载均衡默认开启，且不可关闭。	开启
高级配置		
空闲超时时间	如果在超时时间内一直没有访问请求，负载均衡会中断当前连接，直到下一次请求到来时再重新建立新的连接。 取值范围：10~4000s	300
描述	对于监听器描述。 字数范围：0/255。	-

6. 单击“下一步：配置后端分配策略”，配置后端服务器组及健康检查。配置后端服务器组参数请参见**表3-3**。

表 3-3 独享型负载均衡配置后端服务器组参数说明

参数	说明	示例
后端服务器组	把具有相同特性的后端服务器放在一个组。 ● 新创建 ● 使用已有 **说明** 只能选择与前端协议匹配的后端服务器组。例如前端协议是TCP协议，后端协议应为TCP协议。	新创建

参数	说明	示例
名称	后端服务器组名称。	server_group
后端协议	云服务器开通的协议。 前端协议为TCP时，后端协议默认为TCP，不支持修改。	TCP
分配策略类型	负载均衡采用的算法。 ● 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器。它用相应的权重表示服务器的处理性能，按照权重的高低以及轮询方式将请求分配给各服务器，相同权重的服务器处理相同数目的连接数。 ● 加权最少连接：最少连接是通过当前活跃的连接数来估计服务器负载情况的一种动态调度算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。 ● 源IP算法：将请求的源IP地址进行一致性Hash运算，得到一个具体的数值，同时对后端服务器进行编号，按照运算结果将请求分发到对应编号的服务器上。这可以使得对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某特定的服务器。 **说明** ● 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。 ● 对于加权轮询算法和加权最少连接，当服务器的权重为“0”时，将不会被分发访问请求。	加权轮询算法
会话保持	如果“分配策略类型”选择“加权轮询算法”时，则该项是可选参数。 开启会话保持后，弹性负载均衡将属于同一个会话的请求都转发到同一个服务器进行处理。	-

参数	说明	示例
会话保持类型	开启会话保持功能，支持配置会话保持类型。 TCP和UDP协议仅支持源IP地址类型。 **源IP地址：**基于源IP地址的简单会话保持，将请求的源IP地址作为散列键（HashKey），从静态分配的散列表中找出对应的服务器。即来自同一IP地址的访问请求会被转发到同一台后端服务器上进行处理。	源IP地址
描述	后端服务器组的描述。 字数范围：0/255。	-

7. 单击“下一步：添加后端服务器”，添加后端服务器并配置健康检查。添加后端服务器详见**后端服务器概述**，配置健康检查参数请参见**表3-4**。

表 3-4 独享型负载均衡配置健康检查参数说明

参数设置	说明	示例
是否开启	开启或者关闭健康检查。 如果开启健康检查，您可单击“参数设置”设置健康检查的参数。	-
健康检查协议	健康检查请求的协议类型。 后端协议为TCP协议时，健康检查仅支持TCP、HTTP、HTTPS协议。	HTTP
健康检查域名	如果健康检查协议选择HTTP/HTTPS协议，则该项是必选参数。 健康检查的请求域名。 • 默认使用后端服务器的内网IP为域名。 • 您也可选择指定特定域名，特定域名只能由字母，数字，中划线组成，中划线不能在开头或末尾，至少包含两个字符串，单个字符串不能超过63个字符，字符串间以点分割，且总长度不超过100个字符。	www.elb.com
健康检查端口	健康检查端口号，取值范围[1，65535]，为可选参数。 **说明** 默认使用后端云服务器的业务端口进行健康检查。指定特定端口后，使用指定的端口进行健康检查。	80

参数设置	说明	示例
健康检查路径	如果健康检查协议选择HTTP/HTTPS协议，则该项是必填参数。 指定健康检查的URL地址。检查路径只能以/开头，长度范围[1-80]。 支持使用英文字母、数字和字符-/.%?&_。	/index.html
检查间隔（秒）	发送健康检查请求的时间间隔。 取值范围[1-50]。	5
超时时间（秒）	每次健康检查响应的最大超时时间。取值范围[1-50]。	3
最大重试次数	健康检查最大的重试次数，取值范围[1-10]。	3

8. 单击“下一步：确认配置”。
9. 确认配置无误后，单击“提交”。

3.3 添加 UDP 监听器

操作场景

UDP协议适用于关注实时性而相对不注重可靠性的场景，如视频聊天、游戏、金融实时行情推送。您可以添加一个UDP监听器转发来自UDP协议的请求。

约束与限制

- UDP监听器不支持分片包。
- UDP监听器的前端端口当前不支持4789。
- UDP监听器支持的最大MTU为1500，请确保与ELB通信的网卡的MTU不大于1500（有些应用程序需要根据此MTU值同步修改其配置文件），否则数据包可能会因过大被丢弃。
- 如果您的独享型负载均衡实例类型为应用型（HTTP/HTTPS），则无法创建UDP监听器。

添加独享型负载均衡 UDP 监听器

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要添加监听器的负载均衡名称。
5. 切换到“监听器”页签，单击“添加监听器”，配置监听器。配置监听器参数参见**表3-5**。

表 3-5 独享型负载均衡配置监听器参数说明

参数	说明	示例
名称	监听器名称。	listener
前端协议	客户端与负载均衡监听器建立流量分发连接的协议。 协议选择UDP。	UDP
前端端口	客户端与负载均衡监听器建立流量分发连接的端口。 取值范围为：[1-65535]。	80
访问控制	支持通过白名单和黑名单进行访问控制，更多信息请参见**访问控制策略**： • 允许所有IP访问 • 黑名单 • 白名单	黑名单
IP地址组	设置白名单或者黑名单时，必须选择一个IP地址组。如果还未创建IP地址组，需要先创建IP地址组，更多关于IP地址组的信息请参见**创建IP地址组**。	ipGroup
获取客户端IP	开启此开关，后端服务器可以获取到客户端的真实IP地址。 独享型负载均衡默认开启，且不可关闭。	开启
高级配置		
空闲超时时间	如果在超时时间内一直没有访问请求，负载均衡会中断当前连接，直到下一次请求到来时再重新建立新的连接。 取值范围：10~4000s	300
描述	对于监听器描述。 字数范围：0/255。	-

6. 单击“下一步：配置后端分配策略”，配置后端服务器组及健康检查。配置后端服务器组参数请参见**表3-6**。

表 3-6 独享型负载均衡配置后端服务器组参数说明

参数	说明	示例
后端服务器组	把具有相同特性的后端服务器放在一个组。 ● 新创建 ● 使用已有 **说明** 只能选择与前端协议匹配的后端服务器组。例如前端协议是TCP协议，后端协议应为TCP协议。	新创建
名称	后端服务器组名称。	server_group
后端协议	云服务器开通的协议。 前端协议为UDP时，后端协议支持修改，可以选择为UDP或QUIC。	UDP
分配策略类型	负载均衡采用的算法。 ● 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器。它用相应的权重表示服务器的处理性能，按照权重的高低以及轮询方式将请求分配给各服务器，相同权重的服务器处理相同数目的连接数。 ● 加权最少连接：最少连接是通过当前活跃的连接数来估计服务器负载情况的一种动态调度算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。 ● 源IP算法：将请求的源IP地址进行一致性Hash运算，得到一个具体的数值，同时对后端服务器进行编号，按照运算结果将请求分发到对应编号的服务器上。这可以使得对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某特定的服务器。 **说明** ● 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。 ● 对于加权轮询算法和加权最少连接，当服务器的权重为“0”时，将不会被分发访问请求。	加权轮询算法

参数	说明	示例
会话保持	如果“分配策略类型”选择“加权轮询算法”时，则该项是可选参数。 开启会话保持后，弹性负载均衡将属于同一个会话的请求都转发到同一个后端服务器进行处理。	-
会话保持类型	TCP和UDP协议仅支持源IP地址类型。 **源IP地址：**基于源IP地址的简单会话保持，将请求的源IP地址作为散列键（HashKey），从静态分配的散列表中找出对应的服务器。即来自同一IP地址的访问请求会被转发到同一台后端服务器上进行处理。	源IP地址
描述	后端服务器组的描述。 字数范围：0/255。	-

7. 单击“下一步：添加后端服务器”，添加后端服务器并配置健康检查。添加后端服务器详见**后端服务器概述**，配置健康检查参数请参见**表3-7**。

表 3-7 独享型负载均衡配置健康检查参数说明

参数	说明	示例
是否开启	开启或者关闭健康检查。 如果开启健康检查，您可单击“参数设置”设置健康检查的参数。	-
健康检查协议	健康检查请求的协议类型。 后端协议为UDP协议时，健康检查仅支持UDP协议，且不支持修改。	UDP
健康检查端口	健康检查端口号，取值范围[1，65535]，为可选参数。 **说明** 默认使用后端云服务器的业务端口进行健康检查。指定特定端口后，使用指定的端口进行健康检查。	80
检查间隔（秒）	发送健康检查请求的时间间隔。 取值范围[1-50]。	5
超时时间（秒）	每次健康检查响应的最大超时时间。取值范围[1-50]。	3
最大重试次数	健康检查最大的重试次数，取值范围[1-10]。	3

8. 单击“下一步：确认配置”。
9. 确认配置无误后，单击“提交”。

3.4 添加 HTTP 监听器

操作场景

HTTP协议适用于需要对数据内容进行识别的应用，如Web应用、小的手机游戏等。您可以添加一个HTTP监听器转发来自HTTP协议的请求。

约束与限制

- 前端协议为“HTTP”时，后端协议默认为“HTTP”，且不支持修改。
- 如果您的独享型负载均衡实例类型为网络型（TCP/UDP），则无法创建HTTP监听器。

添加独享型负载均衡 HTTP 监听器

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要添加监听器的负载均衡名称。
5. 切换到“监听器”页签，单击“添加监听器”，配置监听器。配置监听器参数参见表3-8。

表 3-8 独享型负载均衡配置监听器参数说明

参数	说明	示例
名称	监听器名称。	listener
前端协议	客户端与负载均衡监听器建立流量分发连接的协议。 协议选择HTTP。	HTTP
前端端口	客户端与负载均衡监听器建立流量分发连接的端口。 取值范围为：[1-65535]。	80
重定向	重定向开关是否开启。 协议类型为HTTP时，可根据需要设置该项。需要保证业务建立安全连接时，若同时创建了HTTPS监听器和HTTP监听器，可以通过重定向功能，将HTTP监听器访问重定向至HTTPS监听器。	-
重定向至	重定向开关开启，需要选择重定向至的HTTPS监听器的名称。	listener_HTTPS_443

参数	说明	示例
访问控制	支持通过白名单和黑名单进行访问控制，更多信息请参见**访问控制策略**： • 允许所有IP访问 • 黑名单 • 白名单	黑名单
IP地址组	设置白名单或者黑名单时，必须选择一个IP地址组。如果还未创建IP地址组，需要先创建IP地址组，更多关于IP地址组的信息请参见**创建IP地址组**。	ipGroup
获取客户端IP	开启此开关，后端服务器可以获取到客户端的真实IP地址。 独享型负载均衡默认开启，且不可关闭。	开启
高级配置		
获取弹性公网IP	通过X-Forwarded-ELB-IP头字段获取ELB实例公网IP地址。 若您需要将ELB公网IP透传到后端，只需在创建HTTP监听器时，打开该开关。	-
获取监听器端口号	通过X-Forwarded-Port头字段获取ELB实例监听器端口号。 若您需要将ELB实例监听器的端口号透传到后端，只需在创建HTTP监听器时，打开该开关。	-
获取客户端请求端口号	通过X-Forwarded-For-Port头字段获取客户端请求端口号。 若您需要将客户端请求的端口号透传到后端，只需在创建HTTP监听器时，打开该开关。	-
重写X-Forwarded-Host	• 开关关闭：ELB透传客户端的X-Forwarded-Host。 • 开关开启：ELB以客户端请求头的Host重写X-Forwarded-Host向后端传输。	-
空闲超时时间（秒）	如果在超时时间内一直没有访问请求，负载均衡会中断当前连接，直到下一次请求到来时再重新建立新的连接。 时间取值范围[0-4000]。	60

参数	说明	示例
请求超时时间（秒）	客户端向负载均衡发起请求，如果在超时时间内客户端没有完成整个请求的传输，负载均衡将放弃等待关闭连接。 时间取值范围[1-300]。	60
响应超时时间（秒）	负载均衡向后端服务器发起请求，如果超时时间内接收请求的后端服务器无响应，负载均衡会向其他后端服务器重试请求。如果重试期间后端服务器一直没有响应，则负载均衡会给客户端返回HTTP 504错误码。 时间取值范围[1-300]。 **说明** 当开启了会话保持功能时，响应超时时间内如果对应的后端服务器无响应，则直接会返回HTTP 504错误码。	60
描述	对于监听器描述。 字数范围：0/255。	-

6. 单击“下一步：配置后端分配策略”，配置监听器的默认后端服务器组。
   a. 推荐选择“使用已有”后端服务器组。
   b. 您也可选择“新创建”后端服务器组，配置后端服务器组参数请参见**表3-9**。

表 3-9 独享型负载均衡配置后端服务器组参数说明

参数	说明	示例
后端服务器组	把具有相同特性的后端服务器放在一个组。 • 新创建 • 使用已有 **说明** 只能选择与前端协议匹配的后端服务器组。例如前端协议是TCP协议，后端协议应为TCP协议。	新创建
名称	后端服务器组名称。	server_group
后端协议	云服务器开通的协议。 前端协议为HTTP时，后端协议默认为HTTP，不支持修改。	HTTP

参数	说明	示例
分配策略类型	负载均衡采用的算法。 ● 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器。它用相应的权重表示服务器的处理性能，按照权重的高低以及轮询方式将请求分配给各服务器，相同权重的服务器处理相同数目的连接数。 ● 加权最少连接：最少连接是通过当前活跃的连接数来估计服务器负载情况的一种动态调度算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。 ● 源IP算法：将请求的源IP地址进行一致性Hash运算，得到一个具体的数值，同时对后端服务器进行编号，按照运算结果将请求分发到对应编号的服务器上。这可以使得对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某特定的服务器。 **说明** ● 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。 ● 对于加权轮询算法和加权最少连接，当服务器的权重为“0”时，将不会被分发访问请求。	加权轮询算法
会话保持	如果“分配策略类型”选择“加权轮询算法”时，则该项是可选参数。 开启会话保持后，弹性负载均衡将属于同一个会话的请求都转发到同一个服务器进行处理。	-
会话保持类型	前端协议为HTTP或HTTPS时，支持会话保持。 ● **负载均衡器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服务器处理。 **说明**	负载均衡器cookie

参数	说明	示例
慢启动	慢启动默认关闭。 当开启慢启动时，负载均衡器向该模式下的后端服务器线性增加请求分配权重，当配置的慢启动持续时间期限结束后，负载均衡器向后端服务器发送完整的请求比例，此后本次添加的后端服务器退出慢启动模式。 详情见**慢启动介绍（独享型）**。	-
慢启动时间（秒）	开启慢启动功能时，需配置慢启动的时间。 取值范围为30~1200，默认为30秒。	30
描述	后端服务器组的描述。 字数范围：0/255。	-

7. 单击“下一步：添加后端服务器”。添加后端服务器并配置健康检查。添加后端服务器详见**后端服务器概述**，配置健康检查参数请参见**表3-10**。

表 3-10 独享型负载均衡配置健康检查参数说明

参数设置	说明	示例
是否开启	开启或者关闭健康检查。 如果开启健康检查，您可单击“参数设置”设置健康检查的参数。	-
健康检查协议	健康检查请求的协议类型。 后端协议为HTTP/HTTPS协议时，健康检查支持TCP、HTTP、HTTPS协议。	HTTP
健康检查域名	如果健康检查协议选择HTTP/HTTPS协议，则该项是必选参数。 健康检查的请求域名。 ● 默认使用后端服务器的内网IP为域名。 ● 您也可选择指定特定域名，特定域名只能由字母，数字，中划线组成，中划线不能在开头或末尾，至少包含两个字符串，单个字符串不能超过63个字符，字符串间以点分割，且总长度不超过100个字符。	www.elb.com

参数设置	说明	示例
健康检查端口	健康检查端口号，取值范围[1，65535]，为可选参数。 **说明** 默认使用后端云服务器的业务端口进行健康检查。指定特定端口后，使用指定的端口进行健康检查。	80
健康检查路径	如果健康检查协议选择HTTP/HTTPS协议，则该项是必填参数。 指定健康检查的URL地址。检查路径只能以/开头，长度范围[1-80]。	/index.html
检查间隔（秒）	发送健康检查请求的时间间隔。 取值范围[1-50]。	5
超时时间（秒）	每次健康检查响应的最大超时时间。取值范围[1-50]。	3
最大重试次数	健康检查最大的重试次数，取值范围[1-10]。	3

8. 单击“下一步：确认配置”。
9. 确认配置无误后，单击“提交”。

3.5 添加 HTTPS 监听器

操作场景

HTTPS协议适用于需要加密传输的应用。您可以添加一个HTTPS监听转发来自HTTPS协议的请求。ELB对于用户的HTTPS的请求进行解密，然后发送至后端服务器；后端服务器处理完请求后的返回包首先发送至ELB，由ELB进行加密后，再传回用户侧。

约束与限制

- 独享型负载均衡前端协议为“HTTPS”时，后端协议可以选择“HTTP”或“HTTPS”。
- 如果您的独享型负载均衡实例类型为网络型（TCP/UDP），则无法创建HTTPS监听器。

添加独享型负载均衡 HTTPS 监听器

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要添加监听器的负载均衡名称。
5. 切换到“监听器”页签，单击“添加监听器”，配置监听器。配置监听器参数参见**表3-11**。

表 3-11 独享型负载均衡配置监听器参数说明

参数	说明	示例
名称	监听器名称。	listener-pnqy
前端协议	客户端与负载均衡监听器建立流量分发连接的协议。 协议选择HTTPS。	HTTPS
前端端口	客户端与负载均衡监听器建立流量分发连接的端口。 取值范围为：[1-65535]。	80
SSL解析方式	确保服务安全，请选择客户端到服务器端认证方式。 可选择“单向认证”或“双向认证”。 • 如仅进行服务器端认证，请选择单向认证。 • 双向认证需要负载均衡实例与访问用户互相提供身份认证，从而允许通过认证的用户访问负载均衡实例，后端服务器无需额外配置双向认证。	单向认证
服务器证书	协议类型为HTTPS时，需绑定服务器证书。 服务器证书用于SSL握手协商，需提供证书内容和私钥。	-
CA证书	协议类型为HTTPS且SSL解析方式为“双向认证”时，需绑定CA证书。 CA证书又称客户端CA公钥证书，用于验证客户端证书的签发者；在开启HTTPS双向认证功能时，只有当客户端能够出具指定CA签发的证书时，HTTPS连接才能成功。	-
开启SNI	HTTPS协议的负载均衡可以选择是否开启SNI。 SNI是为了解决一个服务器使用多个域名和证书的TLS扩展。 开启SNI后，允许客户端在发起SSL握手请求时就提交请求的域名信息，ELB收到SSL请求后，会根据域名去查找证书，如果找到域名对应的证书，则返回该证书；如果没有找到域名对应的证书，则返回缺省证书。详见**SNI证书-多域名访问**。	-

参数	说明	示例
SNI证书	HTTPS协议的负载均衡设置开启SNI后需要选择域名对应的证书。 可选择已创建或者创建新的SNI证书。	-
访问控制	支持通过白名单和黑名单进行访问控制，更多信息请参见**访问控制策略**： • 允许所有IP访问 • 黑名单 • 白名单	白名单
IP地址组	设置白名单或者黑名单时，必须选择一个IP地址组。如果还未创建IP地址组，需要先创建IP地址组，更多关于IP地址组的信息请参见**创建IP地址组**。	ipGroup-b2
获取客户端IP	开启此开关，后端服务器可以获取到客户端的真实IP地址。 独享型负载均衡默认开启，且不可关闭。	开启
高级设置		
HTTP/2	协议类型为HTTPS时，可选择是否支持该协议类型。详见**HTTP/2**。	-
获取弹性公网IP	通过X-Forwarded-ELB-IP头字段获取ELB实例公网IP地址。 若您需要将ELB公网IP透传到后端，只需在创建HTTPS监听器时，打开该开关。	-
获取监听器端口号	通过X-Forwarded-Port头字段获取ELB实例监听器端口号。 若您需要将ELB实例监听器的端口号透传到后端，只需在创建HTTP监听器时，打开该开关。	-
获取客户端请求端口号	通过X-Forwarded-For-Port头字段获取客户端请求端口号。 若您需要将客户端请求的端口号透传到后端，只需在创建HTTP监听器时，打开该开关。	-
重写X-Forwarded-Host	• 开关关闭：ELB透传客户端的X-Forwarded-Host。 • 开关开启：ELB以客户端请求头的Host重写X-Forwarded-Host向后端传输。	-

参数	说明	示例
空闲超时时间（秒）	如果在超时时间内一直没有访问请求，负载均衡会中断当前连接，直到下一次请求到来时再重新建立新的连接。 时间取值范围[0-4000]。	60
请求超时时间（秒）	客户端向负载均衡发起请求，如果在超时时间内客户端没有完成整个请求的传输，负载均衡将放弃等待关闭连接。 时间取值范围[1-300]。	60
响应超时时间（秒）	负载均衡向后端服务器发起请求，如果超时时间内接收请求的后端服务器无响应，负载均衡会向其他后端服务器重试请求。如果重试期间后端服务器一直没有响应，则负载均衡会给客户端返回HTTP 504错误码。 时间取值范围[1-300]。 **说明** 当开启了会话保持功能时，响应超时时间内如果对应的后端服务器无响应，则直接会返回HTTP 504错误码。	60
描述	对于监听器描述。 字数范围：0/255。	-

6. 单击“下一步：配置后端分配策略”，配置监听器的默认后端服务器组。
   a. 推荐选择“使用已有”后端服务器组。
   b. 您也可选择“新创建”后端服务器组，配置后端服务器组参数请参见**表3-12**。

表 3-12 独享型负载均衡配置后端服务器组参数说明

参数	说明	示例
后端服务器组	把具有相同特性的后端服务器放在一个组。 • 新创建 • 使用已有 **说明** 只能选择与前端协议匹配的后端服务器组。例如前端协议是TCP协议，后端协议应为TCP协议。	新创建
名称	后端服务器组名称。	server_group-sq4v

参数	说明	示例
后端协议	云服务器开通的协议。 前端协议为HTTPS时，后端协议支持HTTP、HTTPS。	HTTP
分配策略类型	负载均衡采用的算法。 ● 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器。它用相应的权重表示服务器的处理性能，按照权重的高低以及轮询方式将请求分配给各服务器，相同权重的服务器处理相同数目的连接数。 ● 加权最少连接：最少连接是通过当前活跃的连接数来估计服务器负载情况的一种动态调度算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。 ● 源IP算法：将请求的源IP地址进行一致性Hash运算，得到一个具体的数值，同时对后端服务器进行编号，按照运算结果将请求分发到对应编号的服务器上。这可以使得对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某特定的服务器。 **说明** ● 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。 ● 对于加权轮询算法和加权最少连接，当服务器的权重为“0”时，将不会被分发访问请求。	加权轮询算法
会话保持	如果“分配策略类型”选择“加权轮询算法”时，则该项是可选参数。 开启会话保持后，弹性负载均衡将属于同一个会话的请求都转发到同一个服务器进行处理。	-

参数	说明	示例
会话保持类型	前端协议为HTTP或HTTPS时，支持以下两种类型的会话保持。 ● **负载均衡器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服务器处理。 **说明**	负载均衡器cookie
慢启动	慢启动默认关闭。 当开启慢启动时，负载均衡器向该模式下的后端服务器线性增加请求分配权重，当配置的慢启动持续时间期限结束后，负载均衡器向后端服务器发送完整的请求比例，此后本次添加的后端服务器退出慢启动模式。 详情见**慢启动介绍（独享型）**。	-
慢启动时间（秒）	配置慢启动的时间。 取值范围为30~1200，默认为30秒。	30
描述	后端组的描述。 字数范围：0/255。	-

7. 单击“下一步：添加后端服务器”。添加后端服务器并配置健康检查。添加后端服务器详见**后端服务器概述**，配置健康检查参数请参见**表3-13**。

表 3-13 独享型负载均衡配置健康检查参数说明

参数设置	说明	示例
是否开启	开启或者关闭健康检查。 如果开启健康检查，您可单击“参数设置”设置健康检查的参数。	-
健康检查协议	健康检查请求的协议类型。 后端协议为HTTP/HTTPS协议时，健康检查支持TCP、HTTP、HTTPS协议。	HTTP

参数设置	说明	示例
健康检查域名	如果健康检查协议选择HTTP/HTTPS协议，则该项是必选参数。 健康检查的请求域名。 ● 默认使用后端服务器的内网IP为域名。 ● 您也可选择指定特定域名，特定域名只能由字母，数字，中划线组成，中划线不能在开头或末尾，至少包含两个字符串，单个字符串不能超过63个字符，字符串间以点分割，且总长度不超过100个字符。	www.elb.com
健康检查端口	健康检查端口号，取值范围[1，65535]，为可选参数。 **说明** 默认使用后端云服务器的业务端口进行健康检查。指定特定端口后，使用指定的端口进行健康检查。	80
健康检查路径	如果健康检查协议选择HTTP/HTTPS协议，则该项是必填参数。 指定健康检查的URL地址。检查路径只能以/开头，长度范围[1-80]。	/index.html
检查间隔（秒）	发送健康检查请求的时间间隔。 取值范围[1-50]。	5
超时时间（秒）	每次健康检查响应的最大超时时间。取值范围[1-50]。	3
最大重试次数	健康检查最大的重试次数，取值范围[1-10]。	3

8. 单击“下一步：确认配置”。
9. 确认配置无误后，单击“提交”。

3.6 配置监听器的超时时间

操作场景

弹性负载均衡支持配置监听器的超时时间（**空闲超时时间**、**请求超时时间**、**响应超时时间**），方便用户根据自身业务情况，自定义调整超时时间。例如，HTTP/HTTPS协议客户端的请求文件比较大，可以增加请求超时时间，以便能够顺利完成文件的传输。

独享型负载均衡支持修改TCP/UDP/HTTP/HTTPS协议的超时时间。

图 3-1 七层监听器超时时间示意图

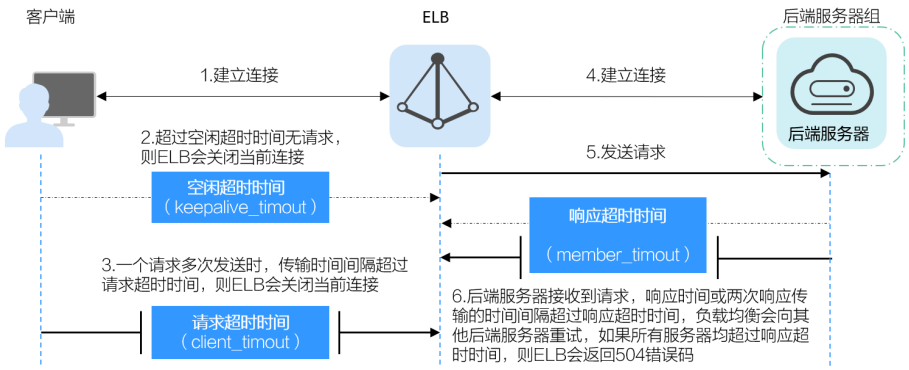


图 3-2 四层监听器超时时间示意图

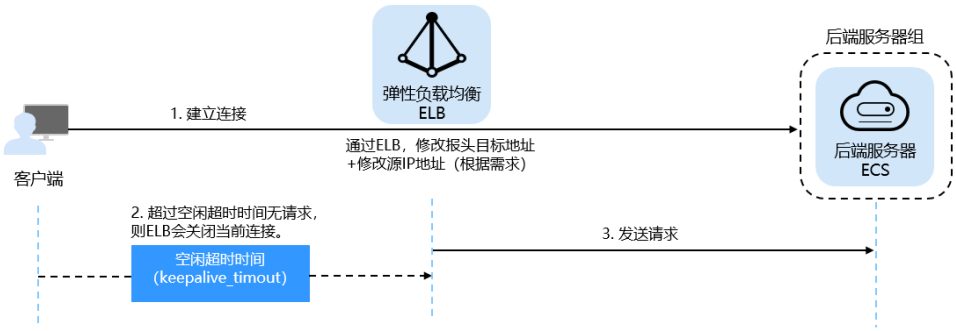


表 3-14 超时时间

协议	类别	描述	取值范围	默认超时时间
TCP	空闲超时时间	如果在超时时间内一直没有访问请求，负载均衡会关闭当前连接，直到下一次请求到来时再重新建立新的连接。	10~4000s	300s
UDP	空闲超时时间		10~4000s	独享型负载均衡：300s
HTTP/HTTPS	空闲超时时间		0~4000s	60s
	请求超时时间	客户端向负载均衡发起请求，一个请求多次发送时，传输时间间隔超过请求超时时间，则负载均衡将放弃等待关闭连接。	1~300s	60s

协议	类别	描述	取值范围	默认超时时间
	响应超时时间	负载均衡向后端服务器发起请求，如果超时时间内接收请求的后端服务器无响应或两次响应传输的时间间隔超过响应超时时间，负载均衡会向其他后端服务器重试请求。如果重试期间后端服务器一直没有响应，则负载均衡会给客户端返回HTTP 504错误码。 **说明** 当开启了会话保持功能时，响应超时时间内如果对应的后端服务器无响应，则直接会返回HTTP 504错误码。	1~300s	60s

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要修改监听器的负载均衡名称。
5. 切换到“监听器”页签，单击需要配置超时时间的目标监听器名称。
6. 在监听器的“基本信息”页面，单击“编辑监听器”。
7. 在“编辑监听器”页面，单击“高级配置”。
8. 根据需要配置“空闲超时时间”或“请求超时时间”或“响应超时时间”。
9. 单击“确定”。

3.7 修改/删除监听器

操作场景

如果您已创建监听器，您可以根据实际业务需求，可以修改或者删除监听器。

监听器被删除后无法恢复，请谨慎操作。

说明

目前暂不支持修改“前端协议/端口”和“后端协议”，如果要修改监听器的协议或端口，请重新创建监听器。

修改监听器

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要修改监听器的负载均衡名称。
5. 您可以通过以下两种操作入口，修改监听器。
   - 在目标监听器所在行的“操作”列，单击“编辑”。
   - 单击目前监听器的名称，进入监听器的“基本信息”页面，单击“编辑监听器”。
6. 在“编辑监听器”页面修改参数，单击“确定”。

删除监听器

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要删除监听器的负载均衡名称。
5. 在“监听器”页签，需要删除监听器所在行的“操作”列，单击“删除 ”。

   说明

   - 如果该监听器下有后端服务器，删除监听器之前需移除后端服务器。
   - 如果HTTP设置了重定向至HTTPS，删除HTTPS监听器之前需删除HTTPS重定向规则。
   - 如果监听器包含了转发策略，删除监听器之前需先删除转发策略。
   - 删除监听器后会同时删除所绑定的后端服务器组。
6. 在删除监听器的弹窗页面，输入“DELETE”。
7. 单击“确定”。

3.8 获取客户端 IP

获取客户端 IP 功能

监听器开启“获取客户端IP”功能后，负载均衡器和后端服务器之间直接使用真实的IP进行通信。

目前，独享型负载均衡对“获取客户端IP”功能的支持情况如表3-15。

表 3-15 独享型负载均衡“获取客户端 IP”功能说明

监听器类型	开启“获取客户端IP”	关闭“获取客户端IP”
四层（TCP/UDP）监听器	默认开启	×
七层（HTTP/HTTPS）监听器	默认开启	×

约束与限制

- 开启“获取客户端IP”之后，不支持同一台服务器既作为后端服务器又作为客户端的场景。

  如果后端服务器和客户端使用同一台服务器，且开启“获取客户端IP”，则后端服务器会根据报文源IP为本地IP判定该报文为本机发出的报文，无法将应答报文返回给ELB，最终导致回程流量不通。
- 开启此功能后，执行后端服务器迁移任务时，可能出现流量中断（例如单向下载、推送类型的流量）。所以后端服务器迁移完成后，需要通过报文重传来恢复流量。
- 通过跨VPC后端功能添加的后端服务器，默认开启的获取客户端IP功能会失效。

其他获取客户端真实 IP 方法

负载均衡的监听器还可通过如下补充方法获取客户端的真实IP，详情见**表3-16**。

表 3-16 独享型负载均衡获取客户端真实 IP 补充方法

监听器类型	其他获取客户端真实IP方法
四层（TCP/UDP）监听器	-
七层（HTTP/HTTPS）监听器	**七层服务获取客户端IP**。

4 HTTP/HTTPS 监听器高级配置

4.1 转发策略（独享型）

转发策略概述

您可以通过给独享型负载均衡添加转发策略，将来自不同域名或者不同URL的请求转发到不同的后端服务器组处理，便于灵活的分流业务，合理的分配资源。

转发策略由**转发规则**和**转发动作**两部分组成，参见表4-1。

表 4-1 转发策略支持的规则与动作

策略分类	转发规则	动作
转发策略	域名、URL。	转发至后端服务器组、重定向至监听器（仅HTTP监听器支持）。
高级转发策略	域名、URL、HTTP请求方法、HTTP请求头、查询字符串、网段。	转发至后端服务器组、重定向至监听器、重定向至URL、返回固定响应。

匹配原理

- 在添加了转发策略后，负载均衡器将按以下规则转发前端请求：
  - 如果能匹配到监听器的转发策略，则按该转发策略将请求转发到对应的后端服务器组。
  - 如果不能匹配到监听器的转发策略，则按照默认转发策略将请求转发到监听器默认的后端服务器组（创建监听器时配置的后端服务器组）。
- 匹配优先级：
  - 不同域名间优先级互相独立，转发规则域名与URL同时存在时，优先按照域名进行匹配。
  - 转发规则为URL时，匹配优先级如下：精确匹配 > 前缀匹配 > 正则匹配，匹配类型相同时URL长度越长，优先级越高。

表 4-2 转发策略示例

访问请求	转发策略	转发规则	设定值
www.elb.com/test	1	URL	/test
	2	域名	www.elb.com

说明

如表4-2中，访问请求www.elb.com/test同时满足转发策略1和转发策略2，优先按照域名进行匹配，则请求将按照转发策略2进行转发。

约束与限制

- 此功能目前仅支持协议类型为HTTP、HTTPS的监听器。
- 负载均衡控制台不支持创建相同的转发策略。
- 一个监听器最多支持配置100条转发策略，超过配额的转发策略不生效。
- 配置转发策略时，请注意以下事项：
  - 每个URL路径需要存于后端服务器（即必须是后端服务器上真实存在的路径），否则访问后端服务器时，后端服务器会返回404。
  - 因为正则匹配采用顺序匹配的方式，只要任意规则匹配成功就结束匹配。所以配置“URL匹配规则”为“正则匹配”的多个匹配规则时，规则之间不能重叠。
  - 不能配置URL路径完全相同的转发策略。
  - 输入的域名总长度不能超过100个字符。

添加转发策略

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要添加转发策略的负载均衡器名称。
5. 在“监听器”页签，您可以通过以下两种操作入口，进入监听器的“转发策略”页签。
   - 在目标监听器所在行的“转发策略”列，单击“添加/编辑转发策略”。
   - 单击目标监听器的名称，并切换到“转发策略”页签。
6. 单击“添加转发策略”按钮。参考表4-3配置参数。

表 4-3 添加转发策略的参数

参数	类型	说明	样例
如果**转发规则**	域名	触发转发的域名，仅支持精确域名。 域名或者URL至少要指定一个。	www.test.com

参数	类型	说明	样例
	URL	触发转发的URL。URL的匹配规则有如下三种： ● 精确匹配：请求的URL和设定URL完全一致。 ● 前缀匹配：请求的URL匹配已设定URL开头的URL。 ● 正则匹配：请求的URL和设定的URL正则表达式匹配。	/login.php
然后**转发动作**	转发至后端服务器组	如果请求与配置的转发规则匹配，则将请求转发至配置的后端服务器组。	-
	重定向至监听器	如果请求与配置的转发规则匹配，则将请求重定向至配置的监听器。 仅HTTP监听器支持配置该动作类型。 **说明** 选择“重定向至监听器”后，除访问控制以外原HTTP监听器的配置会失效，将以重定向至的HTTPS监听器的配置进行转发。	-

7. 配置完成，单击“保存”。

4.2 HTTPS 双向认证

使用场景

一般的HTTPS业务场景只对服务器做认证，因此只需要配置服务器的证书即可。某些关键业务（如银行支付），需要对通信双方的身份都要做认证，即双向认证，以确保业务的安全性。

此时，除了配置服务器的证书之外，还需要配置客户端的证书，以实现通信双方的双向认证功能。

本章节以自签名证书为例，介绍如何配置HTTPS双向认证。但是自签名证书存在安全隐患，建议购买权威机构颁发的证书。

使用 OpenSSL 制作 CA 证书

1. 登录到任意一台安装有openssl工具的Linux机器。
2. 创建工作目录并进入该目录。

   mkdir ca

   cd ca
3. 创建CA证书的openssl配置文件ca_cert.conf，内容如下：

```
[ req ]
distinguished_name     = req_distinguished_name
```



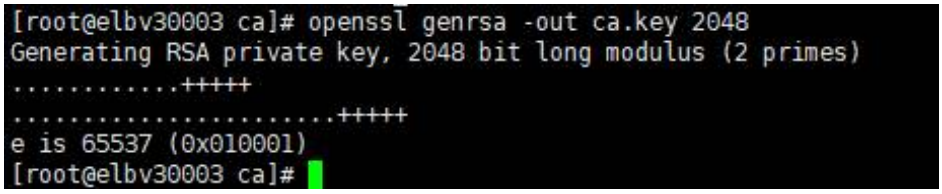
```
prompt              = no

[ req_distinguished_name ]
 O                  = ELB
```

4. 创建CA证书私钥文件ca.key。

   openssl genrsa -out ca.key 2048

   图 4-1 生成 CA 证书私钥文件



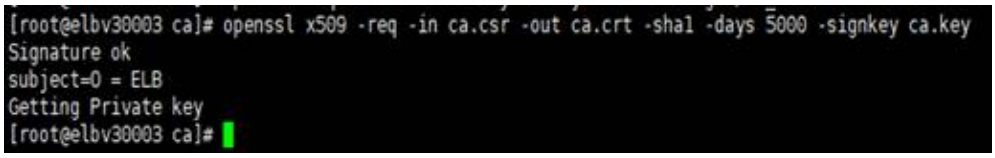
5. 创建CA证书的csr请求文件ca.csr。

   openssl req -out ca.csr -key ca.key -new -config ./ca_cert.conf

6. 创建自签名的CA证书ca.crt。

   openssl x509 -req -in ca.csr -out ca.crt -sha1 -days 5000 -signkey ca.key

   图 4-2 创建自签名 CA 证书



使用 CA 证书签发服务器证书

用户可以用权威CA签发的证书或者自签名的证书，这里以自签名证书为例说明如何创建服务器证书。

1. 登录到生成CA证书的服务器。
2. 创建与CA平级的目录，并进入该目录。

   mkdir server

   cd server

3. 创建服务器证书的openssl配置文件server_cert.conf，内容如下：

```
[ req ]
distinguished_name     = req_distinguished_name
prompt                 = no

[ req_distinguished_name ]
 O                     = ELB
 CN                    = www.test.com
```

   说明

   CN字段可以根据需求改为服务器对应的域名、IP地址。

4. 创建服务器证书私钥文件server.key。

   openssl genrsa -out server.key 2048

5. 创建服务器证书的csr请求文件server.csr。

   openssl req -out server.csr -key server.key -new -config ./server_cert.conf

6. 使用CA证书签发服务器证书server.crt。

   openssl x509 -req -in server.csr -out server.crt -sha1 -CAcreateserial -days 5000 -CA ../ca/ca.crt -CAkey ../ca/ca.key

   图 4-3 签发服务器证书

   ```
   [root@elbv30003 server]# openssl x509 -req -in server.csr -out server.crt -sha1 -CAcreateserial -days 5000 -CA ../ca/ca.crt -CAkey ../ca/ca.key
   Signature ok
   subject=O = ELB, CN = www.test.com
   Getting CA Private Key
   [root@elbv30003 server]#
   ```

使用 CA 证书签发客户端证书

1. 登录到生成CA证书的服务器。
2. 创建与CA平级的目录，并进入该目录。

   mkdir client

   cd client

3. 创建客户端证书的openssl配置文件client_cert.conf，内容如下：

   ```
   [ req ]
   distinguished_name     = req_distinguished_name
   prompt                 = no

   [ req_distinguished_name ]
    O                     = ELB
    CN                    = www.test.com
   ```

   说明

   CN字段可以根据需求改为对应的域名、IP地址。

4. 创建客户端证书私钥文件client.key。

   openssl genrsa -out client.key 2048

   图 4-4 创建客户端证书私钥文件

   ```
   [root@elbv30003 client]# openssl genrsa -out client.key 2048
   Generating RSA private key, 2048 bit long modulus (2 primes)
   ..............................................................................+++++
   ............+++++
   e is 65537 (0x010001)
   [root@elbv30003 client]#
   ```

5. 创建客户端证书的csr请求文件client.csr。

   openssl req -out client.csr -key client.key -new -config ./client_cert.conf

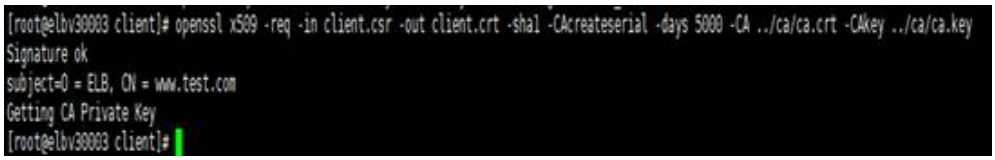
   图 4-5 创建客户端证书 csr 文件

   ```
   [root@elbv30003 client]# openssl req -out client.csr -key client.key -new -config ./client_cert.conf
   ```

6. 使用CA证书签发客户端证书client.crt。

   openssl x509 -req -in client.csr -out client.crt -sha1 -CAcreateserial -days 5000 -CA ../ca/ca.crt -CAkey ../ca/ca.key

图 4-6 签发客户端证书



7. 把客户端证书格式转为浏览器可识别的p12格式。

   openssl pkcs12 -export -clcerts -in client.crt -inkey client.key -out client.p12

   说明

   该命令执行时需要输入导出密码，请输入并记住该密码，在证书导入浏览器时需要使用。

配置服务器证书和私钥

1. 登录负载均衡控制台页面。
2. 单击“证书管理 > 创建证书”。
3. 在创建证书页面，证书类型选择“服务器证书”，同时把前面生成的服务器证书server.crt以及私钥server.key的内容复制到对应的区域，单击“确定”按钮。

   说明

   复制内容时请将最后的换行符删除，避免保存时报错。

   说明

   服务器证书和私钥内容只支持上传pem格式。

配置 CA 证书

步骤1 登录负载均衡控制台页面。

步骤2 单击“证书管理 > 创建证书”。

步骤3 在创建证书页面，证书类型选择“CA证书”，同时把**使用OpenSSL制作CA证书**创建的客户端CA证书ca.crt的内容复制到证书内容区域，单击“确定”按钮。

说明

复制内容时请将最后的换行符删除，避免保存时报错。

说明

CA证书内容只支持上传pem格式。

----结束

配置 HTTPS 双向认证

1. 登录负载均衡控制台页面。
2. 在添加监听器页面，协议类型选择“HTTPS”，“SSL解析方式”选择“双向认证”，并且在服务器证书和CA证书两个配置项中选择所添加的服务器证书和CA证书对应的名称。

添加后端服务器

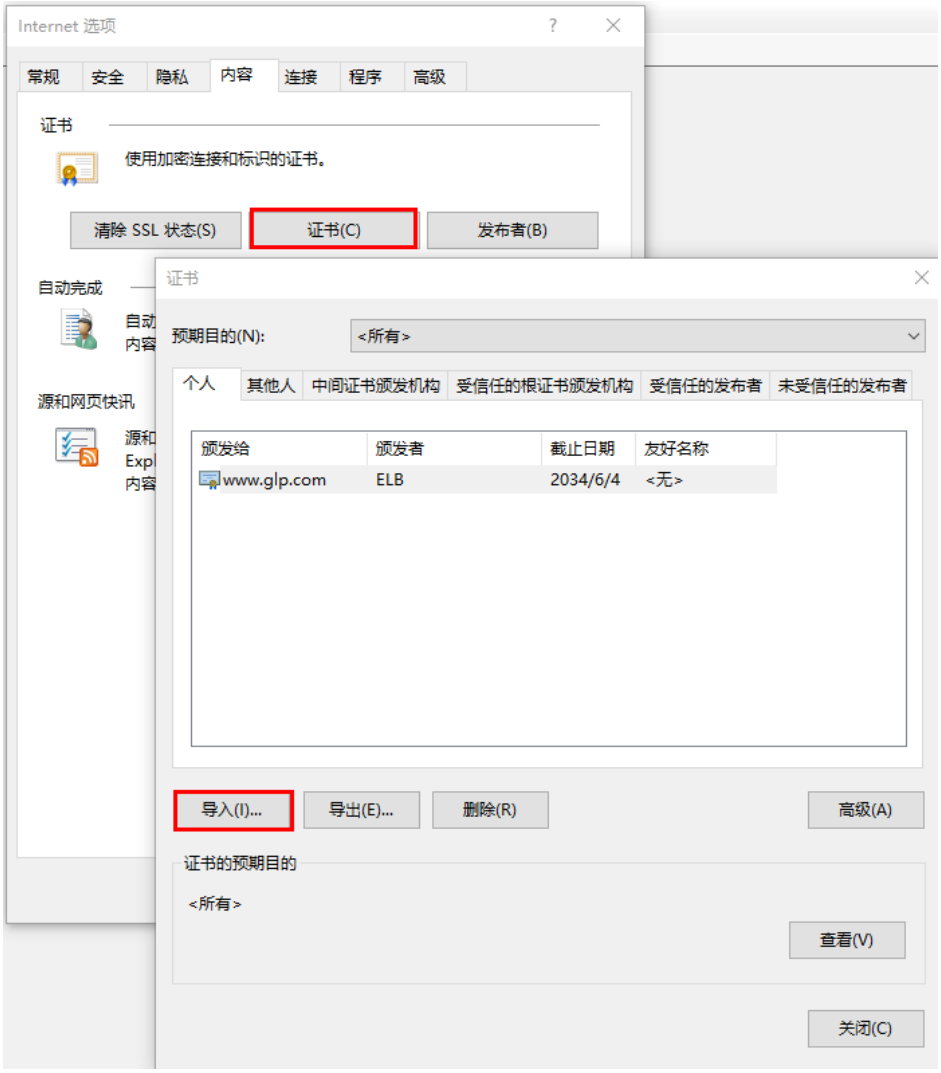
请参考后端服务器相关操作指导，此处不展开描述。

导入客户端证书并测试

浏览器方式功能测试

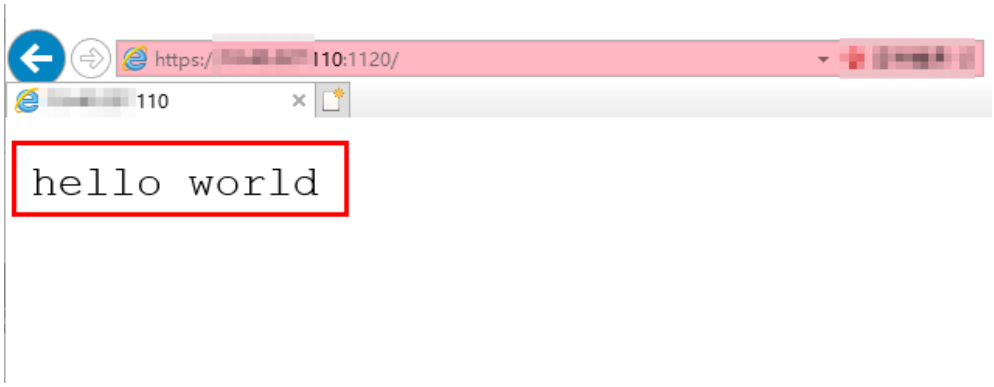
1. 浏览器导入客户端证书（以Internet Explorer 11为例说明）
   a. 把客户端证书从Linux机器导出来，即前面签发的client.p12证书文件。
   b. 单击“设置 > Internet选项”，切换到“内容”页签。
   c. 单击“证书”，然后单击“导入”，导入client.p12证书文件。

   图 4-7 安装 client.p12 证书



2. 测试验证
   在浏览器中输入地址，浏览器会弹出证书选择窗口，如下，选择客户端证书，然后点确定按钮，可以正常访问网站，如图12 正常访问网站。

图 4-8 正常访问网站



Curl工具方式功能测试

1. 导入客户端证书

   把客户端证书client.crt和客户端私钥文件client.key拷贝到新目录，如目录/home/client_cert。

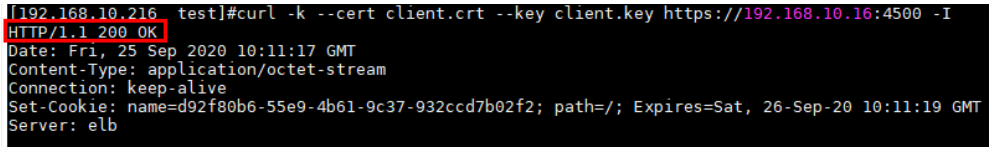
2. 测试验证

   在shell界面，输入以下命令，请输入正确的证书地址和密钥文件地址，以及负载均衡器的IP地址和监听器端口(以下用https://XXX.XXX.XXX.XXX:XXX 表示，以实际IP地址和端口为准)。

   ```
   curl -k --cert /home/client_cert/client.crt --key /home/client_cert/client.key https://XXX.XXX.XXX.XXX:XXX/ -I
   ```

   如果可以正确获得响应码，如图4-9说明验证成功。

图 4-9 正确响应码示例



4.3 HTTP/2

操作场景

HTTP/2，即超文本传输协议 2.0，是下一代HTTP协议。如果您需要保证HTTPS业务更加安全，可以在配置HTTPS监听器时，开启HTTP/2功能。如果您已创建了HTTPS监听器，可以在已创建的HTTPS监听器中开启或者关闭支持HTTP/2功能。

约束与限制

仅HTTPS监听器支持HTTP/2功能。

开启 HTTPS 监听器的 HTTP/2 功能

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。

3. 单击页面左上角的 ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要开启HTTP/2功能的监听器的负载均衡器名称。
5. 在该负载均衡界面的“监听器”页签，单击“添加监听器”。
6. 在“添加监听器”界面，前端协议选择“HTTPS”。
7. 在“添加监听器”界面，展开高级配置，打开HTTP/2功能。
8. 确认配置，单击“提交”。

修改 HTTPS 监听器的 HTTP/2 功能

1. 登录管理控制台。
2. 在管理控制台左上角单击 图标，选择区域和项目。
3. 单击页面左上角的 ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要修改HTTP/2功能的负载均衡器名称。
5. 在“监听器”页签，单击需要修改HTTP/2功能开关的监听器名称。
6. 在监听器的“基本信息”页面，单击“编辑监听器”。
7. 在“编辑监听器”界面，展开高级配置，开启或者关闭HTTP/2功能。
8. 单击“确定”。

4.4 HTTP 重定向至 HTTPS

操作场景

HTTPS是加密数据传输协议，安全性高，如果您需要保证业务建立安全连接，可以通过负载均衡的HTTP重定向功能，将HTTP访问重定向至HTTPS。

该功能可以满足您如下需求，PC、手机浏览器等以HTTP请求访问Web服务，配置了HTTP访问重定向至HTTPS后，后端服务器返回HTTPS的响应。默认强制以HTTPS访问网页。

⚠ 注意

- 因为HTTP标准协议只支持GET和HEAD方法的重定向，所以设置了HTTP重定向至HTTPS后，POST和其他方法会被改为GET方法，这是客户端浏览器的行为，而非ELB修改的。如果您需要实现除GET和HEAD方法以外的访问方式，建议直接使用HTTPS方式进行访问。
- HTTP重定向至HTTPS是指所有的HTTP请求都将转给HTTPS监听器处理为HTTPS请求，但HTTPS请求是通过HTTP被发送给后端服务器的。
- HTTP监听器重定向至HTTPS监听器，HTTPS监听器所关联的后端服务器上不能再安装证书，否则会引起HTTPS请求不生效。

前提条件

- 已经创建HTTPS监听器

- 已经创建HTTP监听器

添加重定向至 HTTPS

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要重定向的HTTP监听器的负载均衡名称。
5. 在该负载均衡界面的“监听器”页签，单击需要重定向的HTTP监听器名称。
6. 切换到“转发策略”页签，单击“添加转发策略”进行添加。

   表 4-4 重定向至 HTTPS 配置

参数	配置说明
动作	选择“重定向至监听器”。
监听器	选择需要重定向至的HTTPS监听器的名称。

7. 转发策略添加完成后，单击“保存”。

说明

- HTTP监听器被重定向，除访问控制以外原有监听器配置会失效。
- HTTP监听器被重定向后，会返回301返回码。

修改重定向至 HTTPS

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要重定向的HTTP监听器的负载均衡名称。
5. 在该负载均衡界面的“监听器”页签，单击需要修改重定向的HTTP监听器名称。
6. 切换到“转发策略”页签下，单击目标转发策略右侧的“编辑”。
7. 用户可根据使用需求，更换重定向至的HTTPS监听器。
8. 单击“保存”。

删除重定向至 HTTPS

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击已经重定向的HTTP监听器的负载均衡名称。

5. 在该负载均衡界面的“监听器”页签，单击已经重定向的HTTP监听器名称。
6. 切换到“转发策略”页签，在重定向至监听器的转发策略右侧单击“删除”。
7. 在确认对话框单击“是”。

4.5 获取 ELB 实例弹性公网 IP

操作场景

对于需要将ELB的弹性公网IP透传到后端服务器的用户，在创建HTTPS监听器和HTTP监听器时，可以开启获取弹性公网IP开关，传输到后端服务器的报文中，HTTPS或HTTP报文头会包含ELB的弹性公网IP。

ELB的弹性公网IP会被放在HTTPS或HTTP报文头的X-Forwarded-ELB-IP字段，格式如下(XX.XXX.XX.XXX代表ELB的弹性公网IP)：

```
X-Forwarded-ELB-IP：XX.XXX.XX.XXX
```

添加获取弹性公网 IP

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要创建ELB公网IP透传到后端服务器的监听器的负载均衡器名称。
5. 在该负载均衡界面的“监听器”页签，单击“添加监听器”。
6. 在“添加监听器”界面，展开高级配置，打开获取弹性公网IP开关。
7. 确认配置，单击“提交”。

说明

“获取弹性公网IP”仅在创建HTTPS监听器和HTTP监听器时可配置。

修改获取弹性公网 IP

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击含有需要修改透传ELB公网IP到后端服务器的监听器的负载均衡器名称。
5. 在“监听器”页签，单击需要修改透传ELB公网IP的监听器名称。
6. 在基本信息页面，单击“编辑监听器”。
7. 在“编辑监听器”界面，展开高级配置，打开或关闭获取弹性公网IP开关。
8. 单击“确定”。

4.6 SNI 证书-多域名访问

操作场景

本章节指导用户通过配置HTTPS监听器绑定多个证书，实现同一个监听器根据多个域名自动选择证书来完成HTTPS认证和访问后端的诉求。

您需要在创建HTTPS监听器时开启SNI功能。SNI（Server Name Indication）是为了解决一个服务器使用域名证书的TLS扩展，开启SNI之后，用户需要添加域名对应的证书。开启SNI后，允许客户端在发起SSL握手请求时就提交请求的域名信息，负载均衡收到SSL请求后，会根据域名去查找证书，如果找到域名对应的证书，则返回该证书；如果没有找到域名对应的证书，则返回缺省证书。

负载均衡在配置HTTPS 监听器时，支持此功能，支持绑定多个证书。

约束与限制

一个HTTPS监听器最多支持配置30个SNI证书。

前提条件

- 已经创建用于SNI证书，具体步骤可参照**创建/修改/删除证书**。
- 已经创建HTTPS监听器，具体步骤可参照**添加HTTPS监听器**。

说明

- 用于SNI的证书，需要指定域名，指定的域名必须与证书中的域名保持一致。
- 目前支持一个域名可以同时绑定ECC类型的证书和RSA类型的证书，在选择SNI证书时，支持选择同域名绑定的两个证书，在使用时，会优先选择ECC类型的证书。
- ELB不会自动选择未过期的证书，如果您有证书过期了，需要手动更换或者删除证书，详见**创建/修改/删除证书**。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击负载均衡名称。
5. 在“监听器”页签，单击需要添加SNI的监听器名称。
6. 在监听器基本信息页面，单击SNI右侧“配置”。
7. 开启SNI的开关，选择需要配置的SNI证书。
8. 单击“确定”。

5 后端服务器组

5.1 后端服务器组概述

后端服务器组简介

后端服务器组是一个或多个后端服务器的逻辑集合，用于将客户端的流量转发到一个或多个后端服务器，满足用户同时处理海量并发业务的需求。后端服务器可以是云服务器实例或IP地址。

后端服务器组参与流量转发过程如下：

1. 来自客户端的请求先传入负载均衡器，再经由负载均衡器上的监听器转发到后端服务器组。
2. 后端服务器组中健康检查正常的后端服务器处理转发的业务请求。
3. 实现同时对用户的海量并发业务进行处理，从而提升用户应用系统的可用性。

后端服务器组优势

在负载均衡器的使用中引入后端服务器组有如下优势：

- 通过后端服务器组可以对后端服务器进行统一管理，灵活地添加或者移除后端服务器，降低用户的管理和使用成本。
- 后端服务器组支持**健康检查功能**，可保证流量转发到正常的后端服务器，提升用户业务的可靠性。

后端服务器组关键功能

为保证用户业务的稳定和多样化的流量转发需求，后端服务器组提供了如**表5-1**所示的关键功能可供用户配置。

表 5-1 后端服务器组关键功能

关键功能	功能说明	功能详情
健康检查	负载均衡器通过健康检查来判断后端服务器是否可用。 如果某个后端服务器健康检查异常，负载均衡器将不会把流量转发给异常后端服务器，从而提升了业务的可靠性。	健康检查介绍。
流量分配策略	负载均衡器按照后端服务器组配置的流量分配策略对请求的流量进行分发。	流量分配策略介绍。
会话保持	开启会话保持后，负载均衡器将属于同一个会话的请求都转发到固定的后端服务器进行处理，避免了客户端重复登录后端服务器。	会话保持介绍。
慢启动	慢启动指负载均衡器向组内新增的后端服务器线性增加请求分配权重的启动模式。 当配置慢启动时间结束，负载均衡向后端服务器发送完整比例的流量请求，实现业务的平滑启动。 **说明** 仅独享型负载均衡支持HTTP和HTTPS类型的后端服务器组开启慢启动功能。	慢启动介绍（独享型）。

后端服务器组创建指引

后端服务器组独立创建后仅可关联至前端协议与后端协议匹配的监听器使用，监听器与后端服务器组的前端/后端协议匹配关系详见表5-2。

您有多种方式创建后端服务器组，详见表5-3。

表 5-2 前端/后端协议匹配关系

监听器的前端协议	后端服务器组的后端协议
TCP	TCP
UDP	• UDP • QUIC
HTTP	HTTP
HTTPS	• HTTP • HTTPS

表 5-3 后端服务器组创建指引

弹性负载均衡类型	后端服务器组创建方法
独享型负载均衡	创建后端服务器组。

5.2 后端服务器组关键功能

5.2.1 健康检查介绍

负载均衡器会定期向后端服务器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服务器是否可用。

负载均衡器如果判断后端服务器健康检查异常，就不会将流量分发到异常后端服务器，而是分发到健康检查正常的后端服务器，从而提高了业务的可靠性。当异常的后端服务器恢复正常运行后，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，通过增大健康检查间隔，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。

健康检查协议

您可以在创建后端服务器组和创建监听器时为后端服务器组配置健康检查，通常，使用默认的健康检查配置即可，也根据业务需要选择不同的健康检查协议。

您也可以在后端服务器组创建后修改健康检查，详情可见**修改健康检查配置**。

后端服务器组的后端协议与支持的健康检查协议存在匹配关系，详情请参见**表5-4**。

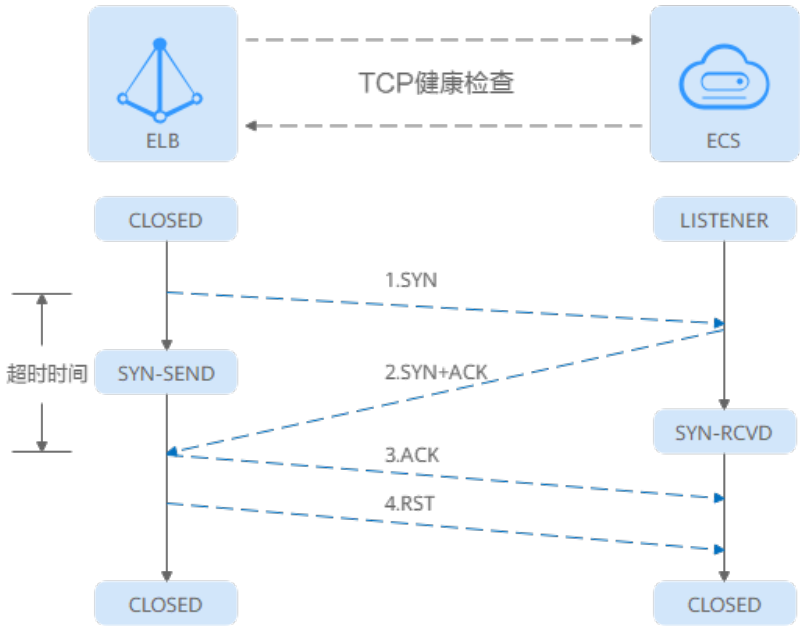
表 5-4 后端服务器组支持的健康检查协议（独享型）

后端服务器组的后端协议	健康检查协议
TCP	TCP、HTTP、HTTPS
UDP	UDP
QUIC	UDP
HTTP	TCP、HTTP、HTTPS
HTTPS	TCP、HTTP、HTTPS

TCP 健康检查

对于四层（TCP）和七层（HTTP/HTTPS）后端协议，您可以配置TCP健康检查，通过发起TCP三次握手来获取后端服务器的状态信息，如**图5-1**所示。

图 5-1 TCP 健康检查



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。
   - 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。随后发送RST报文给后端服务器中断TCP连接。
   - 如果在超时时间内收到了SYN+ACK报文，则判定健康检查成功，并进一步发送ACK报文给后端服务器。随后发送RST报文给后端服务器中断TCP连接。

须知

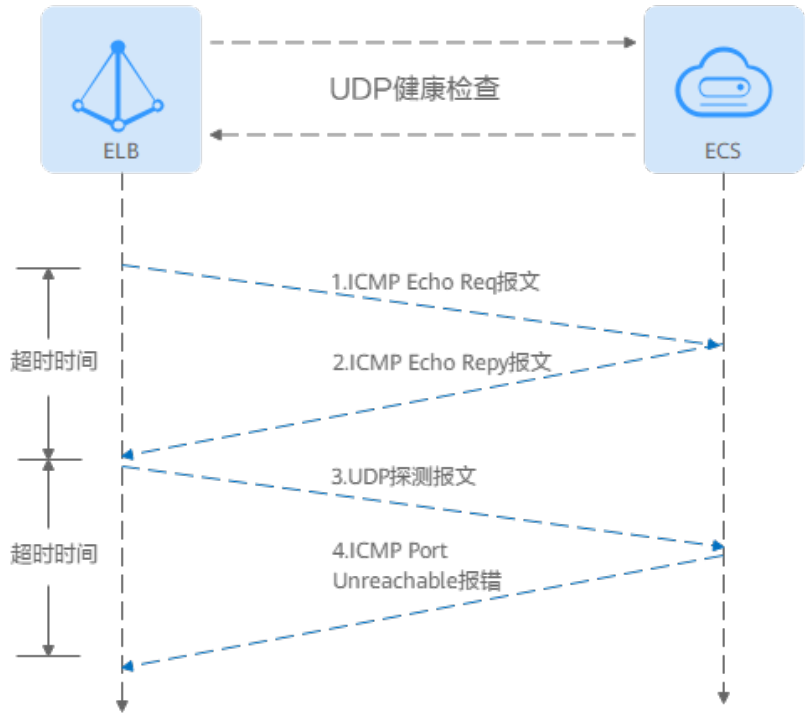
正常的TCP三次握手后，会进行数据传输，但是在健康检查时会发送RST中断建立的TCP连接。该实现方式可能会导致后端服务器中的应用认为TCP连接异常退出，并打印错误信息，如“Connection reset by peer”。解决方案如下：

- 采用**HTTP健康检查**。
- 后端服务器忽略健康检查的连接错误。

UDP 健康检查

对于四层（UDP）后端协议，默认配置UDP健康检查，通过发送UDP探测报文获取后端服务器的状态信息，如**图5-2**所示。

图 5-2 UDP 健康检查



UDP健康检查机制如下：

1. 四层ELB节点根据健康检查配置，向后端服务器发送ICMP Echo Request报文。
   - 如果在超时时间内没有收到ICMP Echo Reply报文，则判定健康检查失败。
   - 如果在超时时间内收到了ICMP Echo Reply报文，则向后端服务器发送UDP探测报文。
2. 如果在超时时间内没有收到后端服务器返回的ICMP Port Unreachable报文，则判定健康检查成功。否则，判定健康检查失败。

HTTP 健康检查

对于四层（TCP）和七层（HTTP/HTTPS）后端协议，您可以配置HTTP健康检查，通过HTTP GET请求来获取状态信息。检查原理如图5-3所示。

图 5-3 HTTP 健康检查

HTTP健康检查机制如下：

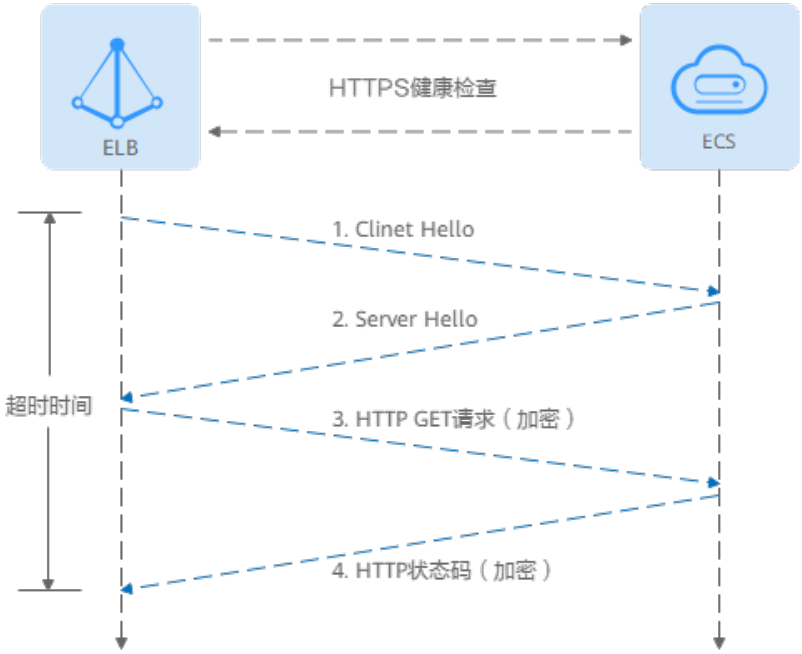
1. ELB节点根据健康检查配置，向后端服务器（IP+端口+检查路径）发出HTTP GET请求（可以选择设置域名）。

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。
   - 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。
   - 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。

HTTPS 健康检查

对于四层（TCP）和七层（HTTP/HTTPS）后端协议，您也可以配置HTTPS健康检查。HTTPS健康检查首先通过TLS握手建立SSL连接，再通过发送加密的HTTP GET请求来获取后端服务器的状态信息。检查原理如图5-4所示。

图 5-4 HTTPS 健康检查



HTTPS健康检查机制如下：

1. ELB节点向后端服务器发送Client Hello请求，与后端服务器建立SSL连接。
2. ELB节点收到后端服务器返回Server Hello报文后，根据健康检查配置，向后端服务器（IP+端口+检查路径）发出加密的HTTP GET请求（可以选择设置域名）。
3. 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。
   - 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。
   - 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。

健康检查时间窗由表5-5中的三个因素决定：

表 5-5 健康检查时间窗的影响因素

影响因素	说明
检查间隔	每隔多久进行一次健康检查。
超时时间	等待服务器返回健康检查的时间。
健康检查阈值	判定健康检查结果正常或异常时，所需的健康检查连续成功或失败的次数。

健康检查时间窗的计算方法如下：

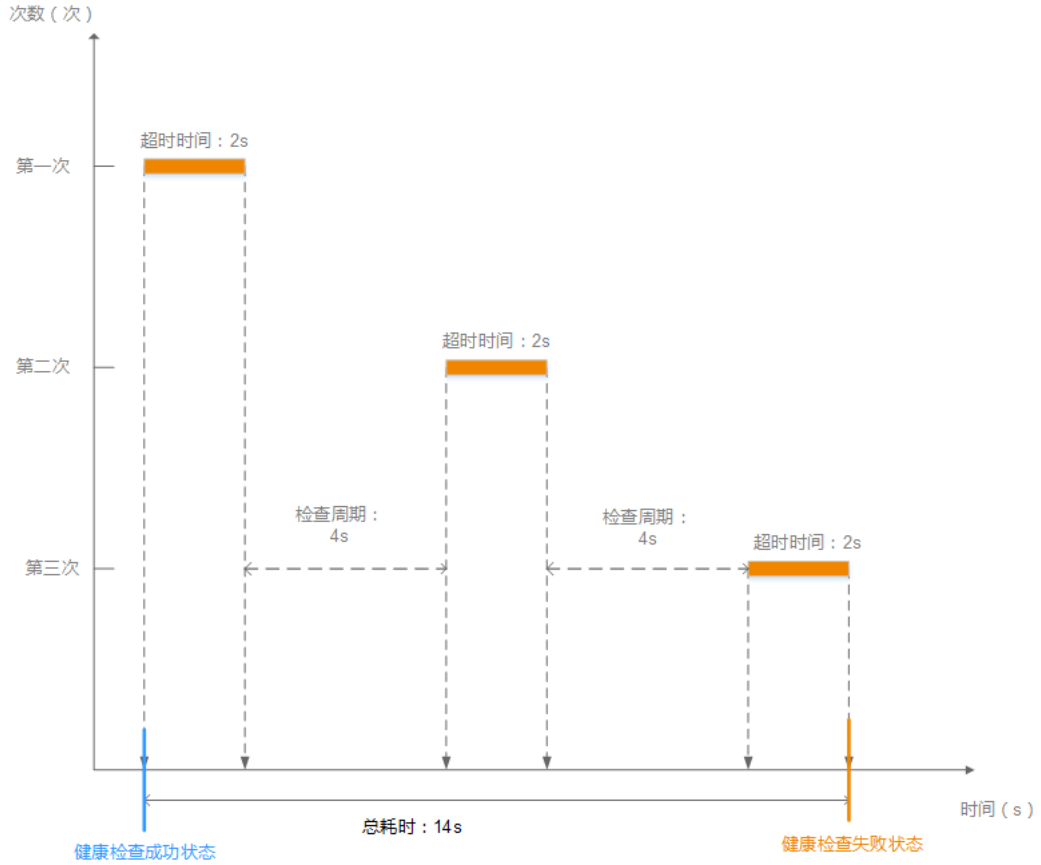
- 健康检查成功时间窗 = 超时时间×健康检查正常阈值 + 检查间隔×(健康检查正常阈值-1)
- 健康检查失败时间窗 = 超时时间×健康检查异常阈值 + 检查间隔×(健康检查异常阈值-1)

如图5-5所示：

- 检查间隔：4s
- 超时时间：2s
- 健康检查异常阈值：3次

健康检查检测到后端服务器从正常到失败状态，健康检查失败时间窗 = 超时时间×健康检查异常阈值+检查间隔×(健康检查异常阈值-1) = 2 x 3+4 x (3-1) = 14s。

图 5-5 健康检查失败时间窗



健康检查异常排查

如果您的健康检查异常，排查方法请参考**健康检查异常如何排查**。

5.2.2 流量分配策略介绍

分配策略类型总览

负载均衡会根据配置的流量分配策略，将来自客户端的请求按照对应的流量分配策略转发至相应的后端服务器。

弹性负载均衡支持加权轮询算法、加权最小连接、源IP算法、连接ID算法等多种分配策略，用于支持不同的业务场景。

本文列出弹性负载均衡支持的所有分配策略，不同类型的负载均衡器和后端服务器组支持的流量分配策略不同。

表 5-6 流量分配策略对比

分配策略类型	描述
加权轮询算法	根据组内后端服务器设置的权重，按照访问顺序依次将请求分发给不同的服务器。

分配策略类型	描述
加权最少连接	将请求分发给（当前连接/权重）比值最小的后端服务器进行处理。
一致性哈希算法 • 源IP算法 • 连接ID算法	对请求的特定字段进行一致性哈希计算，并根据计算的哈希值将请求均匀地分配到后端服务器中。相同哈希值的请求，将会被分配到相同的后端服务器，即使后端服务器组中的后端服务器个数在发生变化。 • 源IP算法：根据请求的源IP地址进行哈希计算，源IP相同的请求会被分配到同一台后端服务器。 • 连接ID算法：根据QUIC协议请求的ID进行哈希计算，相同QUIC ID连接上的请求会被分配到同一台后端服务器。

加权轮询算法

图5-6展示弹性负载均衡器使用加权轮询算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，负载均衡器节点会将50%的客户端流量分发到其可用区中的每一台后端服务器。

图 5-6 加权轮询算法流量分发

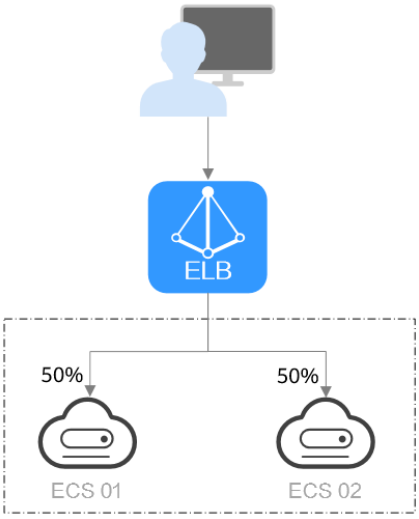


表 5-7 加权轮询算法说明

概述	加权轮询算法根据组内后端服务器设置的权重，将请求分发给不同的服务器。权重大的后端服务器被分配的概率高，相同权重的服务器处理相同数目的连接数。

推荐场景	加权轮询算法常用于短连接服务，例如HTTP等服务。 • 灵活负载：当对后端服务器的负载分配有更精细的要求时，可以通过设置不同的权重来实现对服务器的灵活调度，使得性能较好的服务器能够处理更多的请求。 • 动态负载：当后端服务器的性能和负载情况经常发生变化时，可以通过动态调整权重来适应不同的场景，实现负载均衡。
缺点	• 加权轮询算法需要配置每个后端服务器的权重，对于有大量后端服务器或频繁变动的场景，运维工作量较大。 • 权重设置不准确可能会导致负载不均衡的情况，需要根据后端服务器的实际性能进行调整。

加权最少连接

图5-7展示弹性负载均衡器使用加权最少连接算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，ECS 01已有100个连接，ECS 02已有50个连接，则新的连接会优先分配到ECS 02上。

图 5-7 加权最少连接算法流量分发

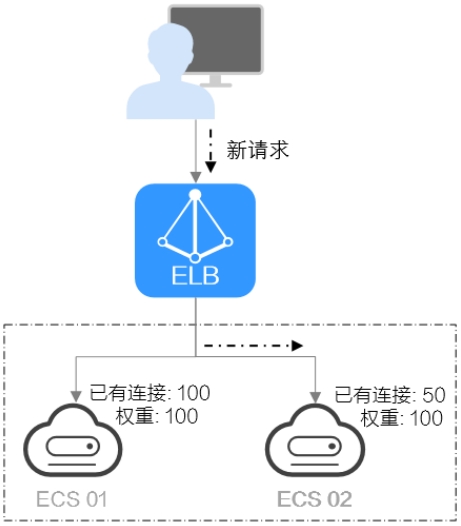


表 5-8 加权最少连接说明

概述	最少连接是通过当前活跃的连接数来评估服务器负载情况的一种动态负载均衡算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。

推荐场景	加权最少连接常用于长连接服务，例如数据库连接等服务。 • 灵活负载：当后端服务器的性能差异较大时，同时考虑后端服务器的连接数和权重来进行负载，可以更精确地将请求分配到后端服务器上，避免出现过载或空闲的情况。 • 动态负载：当后端服务器的连接数和负载情况经常发生变化时，可以通过实时监控连接数变化进行动态的负载调整。 • 更高稳定负载：对于需要高稳定性的业务场景，加权最小连接算法可以降低后端服务器的峰值负载，提高业务的稳定性和可靠性。
缺点	• 加权最小连接算法的实现更复杂：需要实时监控负载均衡器与后端服务器之间的连接数变化。 • 对后端服务器的连接数存在依赖：算法依赖于准确获取负载均衡服务和后端服务器的连接数，如果获取不准确或监控不及时，可能导致负载分配不均衡。同时由于算法只能统计到负载均衡器与后端服务器之间的连接，后端服务器整体连接数无法获取，因此对于后端服务器挂载到多个弹性负载均衡的场景，也可能导致负载分配不均衡。 • 新增后端服务器时可能导致过载：如果已有的连接数过大，大量的新建连接会被分配到新加入的后端服务器上，可能会导致新加入的后端服务器瞬间过载影响系统稳定性。

源 IP 算法

图5-8展示弹性负载均衡器使用源IP算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，ECS 01已经处理了一个IP-A的请求，则IP-A新发起的请求会自动分配到ECS 01上。

图 5-8 源 IP 算法流量分发

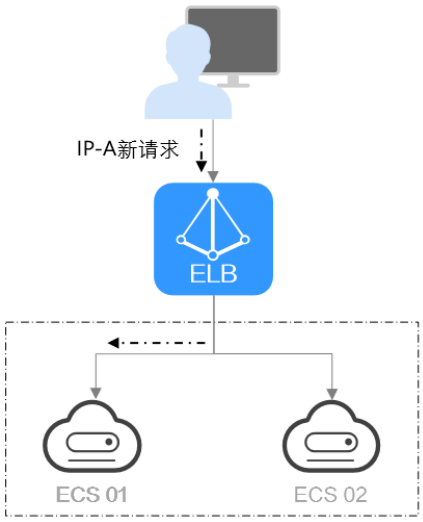


表 5-9 源 IP 算法说明

概述	根据请求的源IP地址进行一致性哈希计算，源IP地址相同的请求会被分配到同一台后端服务器。
推荐场景	源IP算法常用于需要保持用户状态或会话的应用。 • 基于源IP的会话保持：源IP算法可以确保源IP相同的请求具有相当的哈希值并被分配到同一台后端服务器上，从而实现会话保持。 • 保持数据一致：一致性哈希算法将相同哈希值的请求调度到相同后端服务器上，保证多次请求数据的一致性。 • 均衡性要求较高：一致性哈希算法能够提供相对均衡的负载分配效果，减少后端服务器的负载差异。
缺点	• 后端服务器数量变动可能导致不均衡：一致性哈希算法在后端服务器数量变动时会尽力保障请求的一致性，部分请求会重新分配。当后端服务器数量较少时，重新分配过程中有可能导致负载不均衡的情况发生。 • 扩展复杂性增加：由于一致性哈希算法将请求根据哈希因子进行哈希计算，当后端服务器数量变化时，会导致一部分请求需要重新分配，这会引入一定的复杂性。

连接 ID 算法

图5-9展示弹性负载均衡器使用连接ID算法的流量分发流程。假设可用区内有2台权重相同的后端服务器，ECS 01已经处理了一个客户端A的请求，则客户端A上新发起的请求会自动分配到ECS 01。

图 5-9 连接 ID 算法流量分发

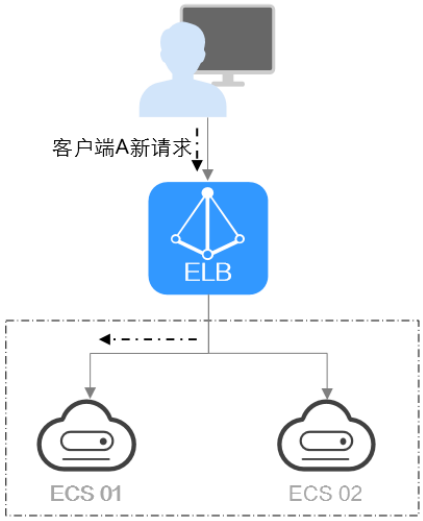


表 5-10 连接 ID 算法说明

概述	根据QUIC 协议请求的QUIC ID进行哈希计算，相同QUIC连接上的请求会被分配到同一台后端服务器。QUIC ID是QUIC连接的唯一标识符，连接ID算法可以实现基于连接级别的负载均衡。 仅QUIC协议的后端服务器组支持连接ID算法。
推荐场景	连接ID算法常用于实现连接级别负载均衡的应用。 ● 基于QUIC连接的会话保持：连接IP算法可以确保源相同QUIC连接上的请求具有相当的哈希值并被分配到同一台后端服务器上，从而实现会话保持。 ● 保持数据一致：一致性哈希算法将相同哈希值的请求调度到相同后端服务器上，保证多次请求数据的一致性。 ● 均衡性要求较高：一致性哈希算法能够提供相对均衡的负载分配效果，减少后端服务器的负载差异。
缺点	● 后端服务器数量变动可能导致不均衡：一致性哈希算法在后端服务器数量变动时会尽力保障请求的一致性，部分请求会重新分配。当后端服务器数量较少时，重新分配过程中有可能导致负载不均衡的情况发生。 ● 扩展复杂性增加：由于一致性哈希算法将请求根据哈希因子进行哈希计算，当后端服务器数量变化时，会导致一部分请求需要重新分配，这会引入一定的复杂性。

5.2.3 会话保持介绍

会话保持，指负载均衡器可以识别客户与服务器之间交互过程的关联性，在实现负载均衡的同时，保持将其他相关联的访问请求分配到同一台服务器上。

会话保持有什么作用呢，举例说明如下：如果有一个用户在服务器甲登录了，访问请求被分配到服务器甲，在很短的时间，这个用户又发出了一个请求，如果没有会话保持功能的话，这个用户的请求很有可能会被分配到服务器乙去，这个时候在服务器乙上是没有登录的，所以需要重新登录。如果配置了会话保持功能，上述一系列的操作过程将由同一台服务器完成，避免被负载均衡器分配到不同的服务器上，提供访问效率。

四层会话保持和七层会话保持的区别

按照所使用的协议的不同，会话保持可以分为**四层会话保持**和**七层会话保持**。

表 5-11 四层会话保持和七层会话保持的区别

类型	说明	支持的会话保持类型	会话保持失效的场景
四层会话保持	当使用的协议为TCP或UDP时，即为四层会话保持。	**源IP地址**：基于源IP地址的简单会话保持，将请求的源IP地址作为散列键（HashKey），从静态分配的散列表中找出对应的服务器。即来自同一IP地址的访问请求会被转发到同一台后端服务器上进行处理。	• 客户端的源IP地址发生变化。 • 客户端访问请求超过会话保持时间。
七层会话保持	当使用的协议为HTTP或HTTPS时，即为七层会话保持。	• **负载均衡器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服务器处理。 负载均衡器本身不会增加额外的cookie。 • **应用程序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服务器处理。	• 如果客户端发送请求未附带cookie，则会话保持无法生效。 • 客户端访问请求超过会话保持时间。

说明

- 当**分配策略类型**选择“源IP算法”时，四层和七层会话已支持基于源IP地址的会话保持。
- 当**分配策略类型**选择“加权轮询算法”或“加权最少连接”时，才可配置会话保持。

约束与限制

- 如果您需要从**云专线**、**VPN**访问ELB，请您使用源IP负载均衡算法代替会话保持功能。
- 独享型负载均衡器支持源IP地址、负载均衡器cookie的会话保持类型。

说明

- 对于HTTP、HTTPS类型的后端服务器，变更会话保持的状态可能会导致监听器与后端服务器组的访问出现秒级中断。
- 如果您开启了会话保持功能，那么有可能会造成后端服务器的访问量不均衡。如果出现了访问不均衡的情况，建议您暂时关闭会话保持功能，再观察一下是否依然存在访问不均衡的情况。

5.2.4 慢启动介绍（独享型）

慢启动指负载均衡器向组内新增的后端服务器线性增加请求分配权重，直到配置的慢启动时间结束，负载均衡器向后端服务器正常发送完请求的启动模式。更多后端服务器分配权重设置，请见**后端服务器的权重**。

慢启动能够实现业务的平滑启动，完美避免业务抖动问题。

说明

仅独享型负载均衡支持HTTP和HTTPS类型的后端服务器组开启慢启动功能。

后端服务器在以下两种状态会退出慢启动状态。

- 到达已设定的慢启动时间。
- 慢启动时间内后端服务器变为异常。

约束与限制

- 仅在流量分配策略使用加权轮询算法时生效。
- 慢启动仅对新增后端服务器生效，后端服务器组首次添加后端服务器时慢启动不生效。
- 后端服务器的慢启动结束之后，不会再次进入慢启动模式。
- 在健康检查开启时，后端服务器健康检查结果正常后慢启动生效。
- 在健康检查关闭时，慢启动立即生效。

5.3 创建后端服务器组

操作场景

负载均衡实例的监听器绑定后端服务器组后，才能正常转发访问请求。

说明

本章节指导用户创建可关联至独享型负载均衡使用的后端服务器组。

您可通过三种方式为负载均衡实例创建后端服务器组，详见**表5-12**。

表 5-12 创建后端服务器组（独享型）指引

创建场景	创建步骤
独立创建后端服务器组后关联至负载均衡实例使用	**操作步骤**。
添加监听器时，选择“新创建”后端服务器组。	您可根据使用需求添加不同协议的监听器，详情见**什么是监听器**。 具体添加步骤如下： • **添加TCP监听器**。 • **添加UDP监听器**。 • **添加HTTP监听器**。 • **添加HTTPS监听器**。
更换监听器的后端服务器组时，选择“创建后端服务器组”。	**更换后端服务器组**。

约束与限制

后端服务器组独立创建后仅可关联至前端协议与后端协议匹配的监听器使用，协议匹配关系详见表5-13。

表 5-13 前端/后端协议匹配关系

监听器的前端协议	后端服务器组的后端协议
TCP	TCP
UDP	• UDP • QUIC
HTTP	HTTP
HTTPS	• HTTP • HTTPS

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击页面右上角“创建后端服务器组”按钮。
6. 配置后端分配策略，参数详情请参见表5-14。

表 5-14 配置后端分配策略参数说明

参数	说明	示例
负载均衡类型	可使用该后端服务器组的负载均衡实例类型，请选择独享型。 以下参数均针对独享型负载均衡。	-
所属负载均衡器	使用该后端服务器组的负载均衡实例。	-
名称	待创建的后端服务器组的名称。	server_group
后端协议	后端云服务器自身提供的网络服务的协议。	HTTP

参数	说明	示例
分配策略类型	负载均衡采用的算法。 ● 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器，权重大的后端服务器被分配的概率高。 ● 加权最少连接：加权最少连接是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。 ● 源IP算法：对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某特定的服务器。 更多关于分配策略的信息，请参见**流量分配策略介绍**。	加权轮询算法
会话保持	仅分配策略类型选择加权轮询算法或加权最少连接时支持开启会话保持。 开启会话保持后，弹性负载均衡将属于同一个会话的请求都转发到同一个后端服务器进行处理。 更多关于会话保持的信息，请参见**会话保持介绍**。	-
会话保持类型	如果“会话保持”功能开启，则该项是必选参数。 选择会话保持的类型： ● **源IP地址：**基于源IP地址的简单会话保持，将请求的源IP地址作为散列键（HashKey），从静态分配的散列表中找出对应的服务器。即来自同一IP地址的访问请求会转发到同一台后端服务器上进行处理。 ● **负载均衡器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服务器处理。 **说明** ● 当后端协议选择TCP/UDP时，支持源IP地址类型。 ● 当后端协议选择HTTP/HTTPS时，支持负载均衡器cookie。	源IP地址

参数	说明	示例
慢启动	如果“分配策略类型”选择“加权轮询算法”，则该项是可选参数。 慢启动指负载均衡器向组内新增的后端服务器线性增加请求分配权重的启动模式。 当配置慢启动时间结束，负载均衡向后端服务器发送完整比例的流量请求，实现业务的平滑启动。 **说明** 仅独享型负载均衡支持HTTP和HTTPS类型的后端服务器组开启慢启动功能。 更多关于慢启动的信息，请参见**慢启动介绍（独享型）**。	-
慢启动时间（秒）	如果“慢启动”功能开启，则该项是必填参数。 慢启动开启后需添加的慢启动时间。	30
描述	后端服务器组的描述	-

7. 单击“下一步”，添加后端服务器并配置健康检查。

   独享型后端服务器组支持添加云服务器、跨VPC后端作为后端服务器，详情可参见**后端服务器概述**。

   配置健康检查参数请参见**表5-15**。更多关于健康检查的信息，请参见**健康检查介绍**。

   表 5-15 配置健康检查参数说明

参数	说明	示例
是否开启	开启或者关闭健康检查。 如果开启健康检查，您可单击“参数设置”设置健康检查的参数。	-
健康检查协议	健康检查请求的协议类型。 • 支持选择TCP、HTTP、HTTPS协议。 • 当后端协议选择UDP，健康检查协议默认为UDP且不可修改。	HTTP

参数	说明	示例
健康检查域名	如果健康检查协议选择HTTP/HTTPS协议，则该项是必选参数。 健康检查的请求域名。 ● 默认使用后端服务器的内网IP为域名。 ● 您也可选择指定特定域名，特定域名只能由字母，数字，中划线组成，中划线不能在开头或末尾，至少包含两个字符串，单个字符串不能超过63个字符，字符串间以点分割，且总长度不超过100个字符。	www.elb.com
健康检查端口	健康检查端口号，取值范围[1，65535]，为可选参数。 **说明** 默认使用后端云服务器的业务端口进行健康检查。指定特定端口后，使用指定的端口进行健康检查。	80
健康检查路径	如果健康检查协议选择HTTP/HTTPS协议，则该项是必填参数。 指定健康检查的URL地址。检查路径只能以/开头，长度范围[1-80]。 支持使用英文字母、数字和字符-/.%?&_。	/index.html
检查间隔（秒）	每次健康检查响应的最大间隔时间。 取值范围[1-50]。	5
超时时间（秒）	每次健康检查响应的最大超时时间。取值范围[1-50]。	3
最大重试次数	健康检查最大的重试次数，取值范围[1-10]。	3

8. 单击“下一步”。
9. 确认配置无误后，单击“立即创建”。

后续操作

创建后端服务器组后，您可通过两种方式将后端服务器组关联到独享型负载均衡实例的监听器上使用，详见表5-12。

5.4 修改后端服务器组配置

5.4.1 修改后端服务器组配置场景说明

后端服务器组创建后，用户可根据使用需求修改后端服务器组的健康检查配置和基本信息。

健康检查

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，通过增大健康检查间隔，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。

健康检查功能详情参见**健康检查介绍**。

修改健康检查步骤详情见**修改健康检查配置**。

后端服务器组的基本信息

选定目标后端服务器后，可对以下基本信息进行修改，详情见**表5-16**。

表 5-16 支持修改的后端服务器组信息

参数	修改场景说明
名称	用户可自定义后端服务器组的名称。 修改名称步骤详情见**修改流量分配策略配置**。
分配策略类型	用户可根据使用需求修改后端服务器组的流量分配策略。 后端服务器组根据配置的流量分配策略转发流量到不同的后端服务器。 流量分配策略详情参见**流量分配策略介绍**。 修改流量分配策略步骤详情见**修改流量分配策略配置**。
会话保持	用户可根据使用需求开启或关闭会话保持。 当用户开启了会话保持功能后，会话保持可以使来自同一客户端的请求被转发到同一台后端服务器上，客户端的请求将无需重复登录后端服务器。 开启了会话保持功能，也可能会造成后端服务器的访问量不均衡，此时建议您暂时关闭会话保持功能，再观察是否依然存在访问不均衡的情况。 会话保持功能详情参见**会话保持介绍**。 修改会话保持步骤详情见**修改会话保持配置**。
慢启动	用户可根据使用需求开启或关闭慢启动。 慢启动能够实现业务的平滑启动，完美避免业务抖动问题。建议用户在添加后端服务器前开启慢启动。 慢启动功能详情参见**慢启动介绍（独享型）**。 修改慢启动步骤详情见**修改慢启动配置（独享型）**。

参数	修改场景说明
描述	用户可自定义对目标后端服务器组的描述。 修改描述步骤详情见**修改流量分配策略配置**。

5.4.2 修改健康检查配置

操作场景

本章节指导用户在后端服务器组创建后修改健康检查配置。

若切换健康检查协议，负载均衡会根据新的健康检查协议重新检查后端服务器。健康检查通过后，负载均衡向后端服务器继续转发流量。

健康检查切换周期内，客户端可能收到503错误码。

约束与限制

- 健康检查协议与服务器组的后端协议是两个相互独立的能力，所以健康检查协议可以与后端协议不同。
- 为了减少后端服务器的CPU占用，建议您使用TCP协议做健康检查。如果您希望使用HTTP健康检查协议，建议使用HTTP+静态文件的方式。
- 为保证健康检查功能正常，配置健康检查后必须放通对应的安全组规则：
  - 独享型负载均衡：安全组配置详情请参考**配置后端服务器的安全组**。

说明

开启健康检查后不会影响已建立连接的流量转发，负载均衡会立即对后端服务器执行健康检查。

- 如果健康检查正常，则新建连接的流量会根据分配策略和权重向该服务器转发流量。
- 如果健康异常，则系统会设置该服务器状态为异常，不转发新的流量到该服务器。

开启健康检查

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组界面，单击需要修改健康检查的后端服务器组名称。
6. 在后端服务器组的“基本信息”页签下，单击健康检查区域右侧的“配置健康检查”。
7. 在“配置健康检查”弹窗，可根据需要参考**表5-17**进行配置。

表 5-17 配置健康检查参数说明

参数	说明	示例
是否开启	开启或者关闭健康检查。	-

参数	说明	示例
健康检查协议	• 健康检查支持选择TCP、HTTP、HTTPS方式。 • 当后端协议选择UDP，健康检查协议默认为UDP且不可修改。	HTTP
健康检查域名	如果健康检查协议选择HTTP/HTTPS协议，则该项是必选参数。 健康检查的请求域名。 • 默认使用后端服务器的内网IP为域名。 • 您也可选择指定特定域名，特定域名只能由字母，数字，中划线组成，中划线不能在开头或末尾，至少包含两个字符串，单个字符串不能超过63个字符，字符串间以点分割，且总长度不超过100个字符。	www.elb.com
健康检查端口	健康检查端口号，取值范围[1，65535]，为可选参数。 **说明** 默认使用后端云服务器的业务端口进行健康检查。指定特定端口后，使用指定的端口进行健康检查。	80
检查间隔（秒）	每次健康检查响应的最大间隔时间。 取值范围[1-50]。	5
超时时间（秒）	每次健康检查响应的最大超时时间。取值范围[1-50]。	3
最大重试次数	健康检查最大的重试次数，取值范围[1-10]。	3

8. 单击“确定”。

关闭健康检查

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组界面，单击需要关闭健康检查的后端服务器组名称。
6. 在后端服务器组的“基本信息”页签下，单击健康检查区域右侧的“配置健康检查”。
7. 在“配置健康检查”界面，可根据需要关闭健康检查。

8. 单击“确定”。

5.4.3 修改流量分配策略配置

操作场景

本章节指导用户在后端服务器组中的修改流量分配策略。

流量分配策略详情参见**流量分配策略介绍**。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组列表，在目标后端服务器组所在行的操作列单击“编辑”。
6. 在“修改后端服务器组”弹窗中进行修改，选择“分配策略类型”。
7. 单击“确定”。

说明

修改分配策略立即生效，不影响已经建立连接的流量转发，只影响新建连接的流量分配。

5.4.4 修改会话保持配置

操作场景

本章节指导用户在后端服务器组中修改会话保持功能。

说明

- 本章节适用于独享型弹性负载均衡和共享型弹性负载均衡。
- 您还可以在进行“添加监听器”或“创建后端服务器组”操作时，配置会话保持功能。

开启会话保持

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组界面，在目标后端服务器组所在行的操作列单击“编辑”。
6. 在“修改后端服务器组”弹窗中，开启会话保持功能，配置会话保持类型以及会话保持时间参数。
7. 单击“确定”。

关闭会话保持

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组界面，在目标后端服务器组所在行的操作列单击“编辑”。
6. 在“修改后端服务器组”弹窗中，关闭会话保持功能。
7. 单击“确定”。

5.4.5 修改慢启动配置（独享型）

操作场景

本章节指导用户在后端服务器组中修改慢启动功能。

慢启动详情参见**慢启动介绍（独享型）**。

说明

- 本章节仅适用于独享型弹性负载均衡。
- 您还可以在进行“添加监听器”或“创建后端服务器组”操作时，配置慢启动。

开启慢启动

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组界面，在目标后端服务器组所在行的操作列单击“编辑”。
6. 在“修改后端服务器组”弹窗中，开启慢启动功能并配置慢启动时间。

   慢启动时间（秒）：取值范围为30~1200，当慢启动时间结束，负载均衡向后端服务器发送完整比例的流量请求。
7. 单击“确定”。

关闭慢启动

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组界面，在目标后端服务器组所在行的操作列单击“编辑”。
6. 在“修改后端服务器组”弹窗中，关闭慢启动功能。

7. 单击“确定”。

5.5 更换后端服务器组

操作场景

本章节指导用户更换在监听器下配置的默认转发后端服务器组。

ELB四层监听器（TCP/UDP）将客户端请求转发到默认后端服务器组。

ELB七层监听器（HTTP/HTTPS）将客户端的请求按转发策略的优先级进行转发。若用户未自定义转发策略，客户端请求将被转发至默认后端服务器组。

约束与限制

- 监听器开启重定向，不支持更换后端服务器组。
- 后端服务器组的后端协议应与监听器的前端协议匹配，匹配关系详见表5-2。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”列表，单击目标监听器所在的负载均衡名称。
5. 选择“监听器”页签，在监听器列表中，单击目标监听器的名称。
6. 在监听器的“基本信息”页签，单击“后端服务器组”区域右侧“更换后端服务器组”。
7. 在弹出的对话框中，单击服务器组名称方框。

   将显示搜索框、所有可选服务器组和“创建后端服务器组”。

   a. 选择已有服务器组，可直接单击目标服务器组名称，也可在搜索框中按名称搜索。
   b. 您也可单击“创建后端服务器组”创建新的后端服务器组。创建完成后单击刷新按钮，在已有服务器组中进行选择。

   说明

   若创建新的服务器组，后端协议应与监听器的前端协议匹配才可被当前监听器使用。
8. 单击“确定”。

5.6 查看后端服务器组

操作场景

本章节指导用户查看后端服务器组的的详细信息，主要信息如下：

- 基本信息：后端服务器组的基本信息，包括名称、ID和后端协议等信息。
- 健康检查：后端服务器组是否开启健康检查以及健康检查的详细配置信息。

- 后端服务器：后端服务器组中已添加的后端服务器资源。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组列表，单击待查看的后端服务器组名称。
6. 在后端服务器组“基本信息”页签下，可查看服务器组基本信息和健康检查配置。

5.7 删除后端服务器组

操作场景

本章节指导用户删除已创建的后端服务器组。

约束与限制

- 如果后端服务器组已被监听器使用，无法执行删除，需先将目标后端服务器组从监听器下释放。
  - 在监听器下释放默认转发后端服务器组，详情请参见**更换后端服务器组**。
  - 七层监听器还需保证自定义的转发策略不使用该后端服务器组。
- 如果后端服务器组中包含后端服务器，不能执行删除操作，需先移除已添加的后端服务器。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在后端服务器组列表，在目标后端服务器组所在行的操作列单击“删除”。
6. 在“确认删除后端服务器组”对话框中，单击“是”。

6 后端服务器

6.1 后端服务器概述

负载均衡器会将客户端的请求转发给后端服务器处理。

负载均衡器支持随时增加或减少后端服务器数量，保证应用业务的稳定和可靠，屏蔽单点故障。

如果负载均衡器与某个弹性伸缩组关联，则该弹性伸缩组中的实例会自动添加至负载均衡后端实例，从弹性伸缩组移除的服务器实例会自动从负载均衡后端服务器中删除。

注意事项

- 建议您选择相同操作系统的后端服务器，以便日后管理和维护。
- 新添加后端服务器后，若健康检查开启，负载均衡器会向后端服务器发送请求以检测其运行状态，响应正常则直接上线，响应异常则开始健康检查机制定期检查，检查正常后上线。
- 关机或重启已有业务的后端服务器，会断开已经建立的连接，正在传输的流量会丢失。建议在客户端上面配置重试功能，避免业务数据丢失。
- 如果您开启了会话保持功能，那么有可能会造成后端服务器的访问量不均衡。如果出现了访问不均衡的情况，建议您暂时关闭会话保持功能，观察一下是否依然存在这种情况。

约束与限制

- 一个后端服务器组最多支持添加500个后端服务器。
- 确保后端服务器的安全组已针对后端服务器端口和健康检查端口配置了相应的入方向规则，详情请参见**配置后端服务器的安全组**。
- 独享型负载均衡的网络型(TCP/UDP)实例不支持同一台服务器既作为后端服务器又作为客户端的场景。

后端服务器的权重

在后端服务器组内添加后端服务器后，需设置后端服务服务器的转发权重。权重越高的后端服务器将被分配到越多的访问请求。

每台后端服务器的权重取值范围为[0, 100]，新的请求不会转发到权重为0的后端服务器上。

以下三种流量分配策略支持权重设置，详情见**表6-1**，更多流量策略分配策略详情见**流量分配策略介绍**。

表 6-1 流量分配策略的权重设置说明

流量分配策略类型	权重设置说明
加权轮询算法	• 在非0的权重下，负载均衡器会将请求按权重值的大小分配给所有的后端服务器，且在轮询时，权重大的后端服务器被分配的概率高。 • 当后端服务器的权重都设置为相等时，负载均衡器将按照简单的轮询策略分发请求。
加权最少连接	• 在非0的权重下，负载均衡器会通过 overhead=当前连接数/权重 来计算每个服务器负载。 • 每次调度会选择overhead最小的后端服务器。
源IP算法	• 在非0的权重下，在一段时间内，同一个客户端的IP地址的请求会被调度至同一个后端服务器上。 • 每台后端服务器的权重取只做0和非0的区分。

6.2 配置后端服务器的安全组

操作场景

为了确保负载均衡器与后端服务器进行正常通信和健康检查正常，添加后端服务器后必须检查后端服务器所在的安全组规则和网络ACL规则。

- 后端服务器的安全组规则必须放通源地址为ELB后端子网所属网段。默认情况下，ELB后端子网与ELB所在子网一致。查看如何**配置安全组规则**。
- 网络ACL为子网级别的可选安全层，若ELB的后端子网关联了网络ACL规则，网络ACL规则必须配置允许源地址为ELB后端子网所属网段。查看如何**配置网络ACL规则**。

说明

若独享型ELB实例未开启“跨VPC后端”功能，ELB四层监听器转发的流量将不受安全组规则和网络ACL规则限制，安全组规则和网络ACL规则无需额外放通。

建议您使用监听器的访问控制功能对访问IP进行限制，详情请参考**访问控制策略**。

约束与限制

- 后端服务器组开启健康检查，后端服务器的安全组规则必须配置放通ELB用于健康检查的协议和端口。
- 如果健康检查使用UDP协议，则还必须配置安全组规则放行ICMP协议，否则无法对已添加的后端服务器执行健康检查。

配置安全组规则

首次创建后端服务器时，如果用户未配置过VPC，系统将会创建默认VPC。由于默认VPC的安全组策略为组内互通、禁止外部访问，即外部网络无法访问后端服务器，为了确保负载均衡器可同时在监听器端口和健康检查端口上与已创建后端服务器的进行通信，就需要配置安全组入方向的访问规则。

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 选择“计算 > 弹性云服务器”。
4. 在弹性云服务器列表，单击待变更安全组规则的弹性云服务器名称。
   系统跳转至该弹性云服务器详情页面。
5. 选择“安全组”页签，单击安全组名称，查看安全组规则。
6. 单击“ID”或者“更改安全组规则”，系统自动跳转至安全组界面。
7. 在入方向规则页签，单击“添加规则”，根据所在后端服务器组的后端协议类型按表6-2配置安全组入方向的访问规则。

   表 6-2 放通安全组规则（独享型）

后端协议	协议端口	源地址
HTTP或者HTTPS	协议：TCP 端口：后端服务器端口和健康检查端口	**ELB后端子网所属网段**
TCP	协议：TCP 端口：健康检查端口	
UDP	协议：UDP、ICMP 端口：健康检查端口	

   说明

   - 创建负载均衡实例后，不建议变更后端子网。若更换后端子网，负载均衡器已占用的后端子网IP地址不会释放，原后端子网所属网段仍需保持放通状态。
   - 为负载均衡实例新增后端子网，新增后端子网所属网段也需全部放通。
8. 单击“确定”，完成安全组规则配置。

配置网络 ACL 规则

网络ACL是一个子网级别的可选安全层，通过与子网关联的出方向/入方向规则控制出入子网的数据流。网络ACL与安全组类似，都是安全防护策略，当您想增加额外的安全防护层时，就可以启用网络ACL。

网络ACL默认规则会拒绝所有入站和出站流量，启用网络ACL后，您可以通过配置网络ACL入方向规则，放行源网段为ELB后端子网所在网段，目的端口为后端服务器端口。

1. 登录管理控制台。

2. 在管理控制台左上角单击图标，选择区域和项目。
3. 在系统首页，选择“网络 > 虚拟私有云”。
4. 在左侧导航栏选择“访问控制 > 网络ACL”。
5. 在“网络ACL”列表区域，选择网络ACL的名称列，单击您需要修改的“网络ACL名称”进入网络ACL详情页面。
6. 在入方向规则或出方向规则页签，单击“添加规则”，添加入方向或出方向规则。
   - 类型：与后端服务器的IP类型保持一致。
   - 协议：和后端协议一致。
   - 源地址：此方向允许的源地址，填写ELB后端子网网段。
   - 源端口范围：选择业务所在端口范围。
   - 目的地址：此方向允许的目的地址。选择默认值为0.0.0.0/0，代表支持所有的IP地址。
   - 目的端口范围：选择业务所在端口范围。
   - 描述：网络ACL规则的描述信息，非必填项。
7. 单击“确定”。

6.3 后端云服务器

6.3.1 添加后端云服务器

操作场景

本章节指导用户在后端服务器组中添加云服务器类型的后端服务器。

在使用负载均衡服务时，确保至少有一台后端服务器在正常运行，可以接收负载均衡转发的客户端请求。

负载均衡器支持随时增加或减少后端服务器数量，保证应用业务的稳定和可靠。

约束与限制

- 仅支持添加与后端服务器组同VPC的云服务器。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要添加后端服务器的后端服务器组名称。
6. 切换到“后端服务器”页签，选择下方“云服务器”页签，并单击“添加”。
7. 支持通过指定关键字搜索后端服务器。
   勾选添加的后端服务器，单击“下一步”。

8. 设置后端端口和服务器的权重，单击“完成”。
   支持批量设置后端端口。

6.3.2 查看后端云服务器

操作场景

本章节指导用户在后端服务期组中查看已经添加的云服务器类型的后端服务器。

支持查看的云服务器信息包括状态、私网IP地址、健康检查结果、权重和业务端口等。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要查看后端服务器的后端服务器组名称。
6. 切换到“后端服务器”页签，选择下方“云服务器”页签。
7. 在“云服务器”列表中，查看已经添加的后端云服务器。

6.3.3 移除后端云服务器

操作场景

本章节指导用户在后端服务器组中移除云服务器类型的后端服务器。

移除负载均衡器绑定的后端服务器，后端服务器将不再收到负载均衡器转发的需求，但不会对服务器本身产生任何影响，只是解除了后端服务器和负载均衡器的关联关系。您可以在业务增长或者需要增强可靠性时再次将它添加至后端服务器组中。

约束与限制

移除后端服务器后，长连接在超时时间内会复用TCP连接，请求会继续转发，仍然会有流量进入后端服务器。

已有连接在请求超时时间后没有数据传输，ELB会将连接断开。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要移除后端服务器的后端服务器组名称。、
6. 切换到“后端服务器”页签，选择下方“云服务器”页签。

7. 勾选需要移除的云服务器，单击服务器列表上方的“移除”。
8. 在移除后端服务器的对话框中单击“是”。

6.3.4 修改后端云服务器权重

操作场景

本章节指导用户在后端服务器组中修改后端云服务器的转发权重。

约束与限制

- 每台后端服务器的权重取值范围为[0, 100]，新的请求不会转发到权重为0的后端服务器上。
- 仅当流量分配策略为加权轮询算法、加权最少连接算法和源IP算法时支持权重设置，更多详情见**后端服务器的权重**。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要修改后端服务器权重的后端服务器组名称。
6. 在该后端服务器组界面，选择“后端服务器”页签，单击下方“云服务器”区域。
7. 勾选需要设置权重的后端服务器，单击服务器列表上方的“修改权重”。
8. 在“修改权重”弹窗页面，根据需要修改权重的后端数量进行相应操作。
   - 修改权重：
     - 修改单个后端服务器权重：在目标服务器所在行，设置“修改后权重”。
     - 批量修改后端服务器权重：在“批量修改权重”后的输入框中设置权重值，单击输入框右侧的“确定”。

     说明

     将后端服务器的权重值批量设置为“0”，可以实现批量屏蔽后端服务器。
9. 单击弹窗下方的“确定”，完成设置。

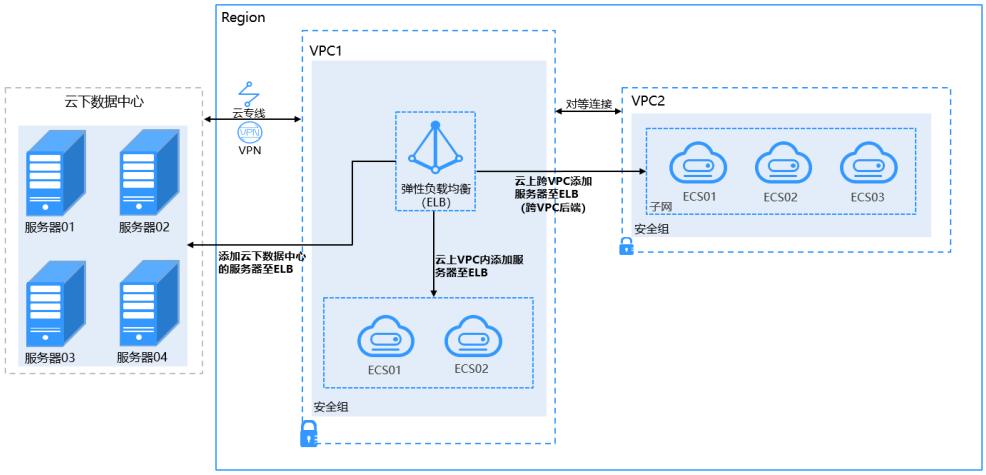
6.4 跨 VPC 后端

6.4.1 跨 VPC 后端概述

独享型负载均衡实例支持混合负载均衡的能力，后端服务器组不仅支持添加云上同VPC内的云服务器和辅助弹性网卡，还支持通过跨VPC后端功能添加云上其他VPC和云下数据中心的IP地址作为后端服务器。

添加跨VPC后端帮助用户根据业务诉求灵活配置后端服务，将流量请求转发到云上、云下的服务器上。

图 6-1 ELB 支持添加云上、云下的服务器



约束与限制

- 跨VPC后端功能开启后无法关闭。
- 只支持添加私网IPv4地址作为后端服务器。
- 跨VPC添加的单个后端服务器最多支持5W并发连接数。
- 通过跨VPC后端功能添加的后端服务器，默认开启的获取客户端IP功能会失效。

添加跨 VPC 后端场景

开启跨VPC后端功能后，独享型负载均衡可添加IP类型的后端服务器。根据添加的IP地址来源不同需做不同的准备，如表6-3。

表 6-3 跨 VPC 后端添加 IP 地址

ELB实例添加跨VPC后端	必做准备
添加云上其他VPC中的IP	需要先在ELB所在的VPC和云上其他VPC之间建立对等连接，然后通过跨VPC功能添加。 建立对等连接详见。
添加云下数据中心的IP	需要先通过云专线或VPN连通云上ELB所在的VPC和云下数据中心，详见或。

6.4.2 开启跨 VPC 后端功能

操作场景

本章节指导用户在独享型负载均衡实例下开启跨VPC后端功能。

约束与限制

- 跨VPC后端功能开启后无法关闭。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要开启跨VPC后端功能的负载均衡名称。
5. 在“基本信息”页面，单击“开启跨VPC后端”。
6. 单击“确定”。

6.4.3 添加跨 VPC 后端

操作场景

本章节指导用户在后端服务器组中通过跨VPC后端添加IP地址作为后端服务器处理来自客户端的请求。

根据添加的IP地址来源不同需做不同的准备，见**表6-4**。

表 6-4 跨 VPC 后端添加 IP 地址

ELB实例添加跨VPC后端	必做准备
添加云上其他VPC中的IP	需要先在ELB所在的VPC和云上其他VPC之间建立对等连接，然后通过跨VPC功能添加。 建立对等连接详见。
添加云下数据中心的IP	需要先通过云专线或VPN连通云上ELB所在的VPC和云下数据中心，详见或。

约束与限制

- 已在负载均衡器基本信息页面开启跨VPC后端功能，否则该功能无法正常使用。
- 跨VPC后端添加的IP地址只支持私网IPv4类型的地址。
- 请确保负载均衡器的后端子网有足够的IP地址，否则该功能无法正常使用。可以通过负载均衡器的“基本信息 > 后端子网”添加多个后端子网来增加后端子网的IP地址。
- 跨VPC后端的安全组规则必须放通负载均衡器的后端子网网段，否则会导后端业务流量与健康检查异常。

操作步骤

1. 登录管理控制台。

2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要添加后端服务器的后端服务器组名称。
6. 切换到“后端服务器”页签，选择下方“跨VPC后端”页签，并单击“添加”。
7. 填写“跨VPC后端IP”、“业务端口”和“权重”。
8. 单击“确定”。

6.4.4 查看跨 VPC 后端

操作场景

本章节指导用户在后端服务期组中查看已经添加的跨VPC后端。

跨VPC后端的信息包括跨VPC后端IP、健康检查结果、权重和业务端口等。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要查看跨VPC后端的后端服务器组名称。
6. 切换到“后端服务器”页签，选择下方“跨VPC后端”页签。
7. 在“跨VPC后端”列表中，查看已经添加的跨VPC后端。

6.4.5 移除跨 VPC 后端

操作场景

本章节指导用户在后端服务器组中移除已添加的跨VPC后端。

约束与限制

移除后端服务器后，长连接在超时时间内会复用TCP连接，请求会继续转发，仍然会有流量进入后端服务器。

已有连接在请求超时时间后没有数据传输，ELB会将连接断开。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。

4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要移除后端服务器的后端服务器组名称。、
6. 切换到“后端服务器”页签，选择下方“跨VPC后端”页签。
7. 勾选需要移除的跨VPC后端，单击服务器列表上方的“移除”。
8. 在移除后端服务器的对话框中单击“是”。

6.4.6 修改跨 VPC 后端的权重

操作场景

本章节指导用户在后端服务器组中修改跨VPC后端服务的转发权重。

约束与限制

- 每台后端服务器的权重取值范围为[0, 100]，新的请求不会转发到权重为0的后端服务器上。
- 仅当流量分配策略为加权轮询算法、加权最少连接算法和源IP算法时支持权重设置，更多详情见**后端服务器的权重**。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > 后端服务器组”。
5. 在“后端服务器组”界面，单击需要修改跨VPC后端权重的后端服务器组名称。
6. 在该后端服务器组界面，切换到“后端服务器”页签，单击下方“跨VPC后端”页签。
7. 勾选需要设置权重的后端服务器，单击服务器列表上方的“修改权重”。
8. 在“修改权重”弹窗页面，根据需要修改权重的后端数量进行相应操作。
   - 修改权重：
     - 修改单个后端服务器权重：在目标服务器所在行，设置“修改后权重”。
     - 批量修改后端服务器权重：在“批量修改权重”后的输入框中设置权重值，单击输入框右侧的“确定”。

     说明

     将后端服务器的权重值批量设置为“0”，可以实现批量屏蔽后端服务器。
9. 单击弹窗下方的“确定”，完成批量设置。

7 证书管理

7.1 证书概述

负载均衡器支持三种类型的证书，服务器证书、CA证书、服务器SM双证书。配置HTTPS监听器时，需要为监听器绑定服务器证书，如果开启双向认证功能，还需要绑定CA证书。

- 服务器SM双证书：在使用HTTPS协议时，若采用商密SSL协议，需提供双证书。双证书包括**签名证书**和**加密证书**，需成套使用。
  - **签名证书：**在签名时使用，仅用于验证身份使用，其公钥和私钥均由服务器自己产生，并由服务器自己保管，证书颁发机构（Certificate Authority，简称“CA”）不负责其保管任务。
  - **加密证书：**在密钥协商时使用，其私钥和公钥均由CA产生，并由CA保管（存根）。

使用证书的注意事项

- 同一个证书在负载均衡器上只需上传一次，可以使用在多个负载均衡器实例中。
- 如果创建的服务器证书用于SNI，则需要指定域名，且指定的域名必须与证书中的域名保持一致。每个证书只能指定一个域名。
- 默认情况下，一个监听器每种类型的证书只能绑定一个，但是一个证书可以被多个监听器绑定。如果监听器开启了SNI功能，则支持绑定多个服务器证书。
- 负载均衡器只支持原始证书，不支持对证书进行加密。
- 可以使用自签名的证书，使用自签名证书和第三方机构颁发的证书对负载均衡器无区别，但是使用自签名证书会存在安全隐患，建议客户使用权威机构颁发的证书。
- 负载均衡器只支持PEM格式的证书，其它格式的证书需要转换成PEM格式后，才能上传到负载均衡。
- ELB不会自动选择未过期的证书，如果您有证书过期了，需要手动更换或者删除证书。

7.2 证书格式

证书格式要求

在创建证书时，您可以直接输入证书内容或上传证书文件。

如果是通过根证书机构颁发的证书，您拿到的证书是唯一的一份，不需要额外的证书，配置的站点即可被浏览器等访问设备认为可信。

服务器证书、CA证书的“证书内容”格式均需按以下要求。

服务器SM双证书中的“SM签名证书内容”和“SM加密证书内容”格式均需按以下要求。

- 以“-----BEGIN CERTIFICATE-----”作为开头，“-----END CERTIFICATE-----”作为结尾。
- 每行64字符，最后一行不超过64字符。
- 证书之间不能有空行。

示例如下：

```
-----BEGIN CERTIFICATE-----
MIIDIjCCAougAwIBAgIJALV96mEtVF4EMA0GCSqGSIb3DQEBBQUAMGoxCzAJBgNV
BAYTAnh4MQswCQYDVQQIEwJ4eDELMAkGA1UEBxMCeHgxCzAJBgNVBAoTAnh4MQsw
CQYDVQQLEwJ4eDELMAkGA1UEAxMCeHgxGjAYBgkqhkiG9w0BCQEWC3h4eEAxNjMu
Y29tMB4XDTE3MTExMzAyMjYxM1oXDTIwMTExMjAyMjYxM1owajELMAkGA1UEBhMC
eHgxCzAJBgNVBAgTAnh4MQswCQYDVQQHEwJ4eDELMAkGA1UEChMCeHgxCzAJBgNV
BAsTAnh4MQswCQYDVQQDEwJ4eDEaMBgGCSqGSIb3DQEJARYLeHh4QDE2My5jb20w
gZ8wDQYJKoZIhvcNAQEBBQADgY0AMIGJAoGBAMU832iM+d3FILgTWmpZBUoYcIWV
cAAYE7FsZ9LNerOyjJpyi256oypdBvGs9JAUBN5WaFk81UQx29wAyNixX+bKa0DB
WpUDqr84V1f9vdQc75v9WoujcnlKszzpV6qePPC7igJJpu4QOI362BrWzJCYQbg4
Uzo1KYBhLFxl0TovAgMBAAGjgc8wgcwwHQYDVR0OBBYEFMbTvDyvE2KsRy9zPq/J
WOjovG+WMIGcBgNVHSMEgZQwgZGAFMbTvDyvE2KsRy9zPq/JWOjovG+WoW6kbDBq
MQswCQYDVQQGEwJ4eDELMAkGA1UECBMCeHgxCzAJBgNVBAcTAnh4MQswCQYDVQQK
EwJ4eDELMAkGA1UECxMCeHgxCzAJBgNVBAMTAnh4MRowGAYJKoZIhvcNAQkBFgt4
eHhAMTYzLmNvbYIJALV96mEtVF4EMAwGA1UdEwQFMAMBAf8wDQYJKoZIhvcNAQEF
BQADgYEAASkC/1iwiALa2RU3YCxqZFEEsZZvQxikrDkDbFeoa6Tk49Fnb1f7FCW6
PTtY3HPWl5ygsMsSy0Fi3xp3jmuIwzJhcQ3tcK5gC99HWp6Kw37RL8WoB8GWFU0Q
4tHLOjBIxkZROPRhH+zMIrqUexv6fsb3NWKhnlfh1Mj5wQE4Ldo=
-----END CERTIFICATE-----
```

私钥格式要求

在创建服务器证书或服务器SM双证书时，您也需要上传证书的私钥。您可直接输入私钥文件内容或上传符合格式的私钥文件。

服务器SM双证书中的“SM签名证书私钥”和“SM加密证书私钥”格式均需按以下要求。

需注意必须是无密码的私钥，私钥内容格式为：

- 符合PEM格式，如下示例：
  - 以“-----BEGIN RSA PRIVATE KEY-----”作为开头，“-----END RSA PRIVATE KEY-----” 作为结尾。
  - 以“-----BEGIN EC PRIVATE KEY-----”作为开头，“-----END EC PRIVATE KEY-----” 作为结尾。

- 私钥之间不能有空行，并且每行64字符，最后一行不超过64字符。

示例如下：

```
-----BEGIN RSA PRIVATE KEY-----
MIICXQIBAAKBgQDFPN9ojPndxSC4E1pqWQVKGHCFlXAAGBOxbGfSzXqzsoyacotu
eqMqXQbxrPSQFATeVmhZPNVEMdvcAMjYsV/mymtAwVqVA6q/OFdX/b3UHO+b/VqL
o3J5SrM86Veqnjzwu4oCSabuEDiN+tga1syQmEG4OFM6NSmAYSxcZdE6LwIDAQAB
AoGBAJvLzJCyIsCJcKHWL6onbSUtDtyFwPViD1QrVAtQYabF14g8CGUZG/9fgheu
TXPtTDcvu7cZdUArvgYW3I9F9IBb2lmF3a44xfiAKdDhzr4DK/vQhvHPuuTeZA41
r2zp8Cu+Bp40pSxmoAOK3B0/peZAka01Ju7c7ZChDWrxleHZAkEA/6dcaWHotfGS
eW5YLbSms3f0m0GH38nRl7oxyCW6yMIDkFHURVMBKW1OhrcuGo8u0nTMi5IH9gRg
5bH8XcujlQJBAMWBQgzCHyoSeryD3TFieXIFzgDBw6Ve5hyMjUtjvgdVKoxRPvpO
kclc39QHP6Dm2wrXXHEej+9RILxBZCVQNbMCQQC42i+Ut0nHvPuXN/UkXzomDHde
h1ySsOAO4H+8Y6OSI87l3HUrByCQ7stX1z3L0HofjHqV9Koy9emGTFLZEzSdAkB7
Ei6cUKKmztkYe3rr+RcATEmwAw3tEJOHmrW5ErApVZKr2TzLMQZ7WZpIPzQRCYnY
2ZZLDuZWFFG3vW+wKKktAkAaQ5GNzbwkRLpXF1FZFuNF7erxypzstbUmU/31b7tS
i5LmxTGKL/xRYtZEHjya4Ikkkgt40q1MrUsgIYbFYMf2
-----END RSA PRIVATE KEY-----
```

7.3 格式转换

操作场景

负载均衡只支持PEM格式的证书，其它格式的证书需要转换成PEM格式后，才能上传到负载均衡。以下是转换成PEM格式的几种常用办法。

DER 转换为 PEM

DER格式通常使用在Java平台。

运行以下命令进行证书转化：

```
openssl x509 -inform der -in certificate.cer -out certificate.pem
```

运行以下命令进行私钥转化：

```
openssl rsa -inform DER -outform PEM -in privatekey.der -out privatekey.pem
```

P7B 转换为 PEM

P7B格式通常使用在Windows Server和Tomcat中。

运行以下命令进行证书转化：

```
openssl pkcs7 -print_certs -in incertificate.p7b -out outcertificate.cer
```

PFX 转换为 PEM

PFX格式通常使用在Windows Server中。

运行以下命令进行证书转化：

```
openssl pkcs12 -in certname.pfx -nokeys -out cert.pem
```

运行以下命令进行私钥转化：

```
openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes
```


7.4 创建/修改/删除证书

操作场景

为了支持HTTPS数据传输加密认证，在创建HTTPS协议监听的时候需绑定证书，负载均衡提供证书管理功能，您可以创建证书、修改证书、删除证书。

说明

- 新建证书只能绑定于所选类型的负载均衡器，请确保负载均衡器类型选择正确。

创建证书

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏单击“证书管理”。
5. 单击“创建证书”，配置证书内容。
   - 证书名称
   - 证书类型：
     - 服务器证书：在使用HTTPS协议时，服务器证书用于SSL握手协商，需提供证书内容和私钥。
     - CA证书： 又称客户端CA公钥证书，用于验证客户端证书的签发者；在开启HTTPS双向认证功能时，只有当客户端能够出具指定CA签发的证书时，HTTPS连接才能成功。
     - 服务器SM双证书：在使用HTTPS协议时，若采用商密SSL协议，需提供双证书。双证书包括**签名证书**和**加密证书**，需成套使用。
       - **签名证书：**在签名时使用，仅用于验证身份使用，其公钥和私钥均由服务器自己产生，并由服务器自己保管，证书颁发机构（Certificate Authority，简称“CA”）不负责其保管任务。
       - **加密证书：**在密钥协商时使用，其私钥和公钥均由CA产生，并由CA保管（存根）。
   - 证书内容：证书内容必须为PEM格式。当证书类型为“**服务器证书**”和“**CA证书**”时，需要填写。

     单击“上传”，选择上传证书文件，请确保您的浏览器是最新版本。

     证书内容格式如下：

     ```
     -----BEGIN CERTIFICATE-----
     Base64-encoded certificate
     -----END CERTIFICATE-----
     ```
   - 私钥：当证书类型为“**服务器证书**”时，需要填写。

     单击“上传”，选择上传私钥文件，请确保您的浏览器是最新版本。

     需注意必须是无密码的私钥。符合PEM格式，私钥格式如下：

```
-----BEGIN PRIVATE KEY-----
[key]
-----END PRIVATE KEY-----
```

- SM签名证书内容：证书内容必须为PEM格式。当证书类型为“**服务器SM双证书**”时，需要填写。

  单击“上传”，选择上传证书文件，请确保您的浏览器是最新版本。

  证书内容格式如下：

  ```
  -----BEGIN CERTIFICATE-----
  Base64–encoded certificate
  -----END CERTIFICATE-----
  ```

- SM签名证书私钥：当证书类型为“**服务器SM双证书**”时，需要填写。

  单击“上传”，选择上传私钥文件，请确保您的浏览器是最新版本。

  需注意必须是无密码的私钥。私钥格式如下：

  ```
  -----BEGIN PRIVATE KEY-----
  [key]
  -----END PRIVATE KEY-----
  ```

- SM加密证书内容：证书内容必须为PEM格式。当证书类型为“**服务器SM双证书**”时，需要填写。

  单击“上传”，选择上传证书文件，请确保您的浏览器是最新版本。

  证书内容格式如下：

  ```
  -----BEGIN CERTIFICATE-----
  Base64–encoded certificate
  -----END CERTIFICATE-----
  ```

- SM加密证书私钥：当证书类型为“**服务器SM双证书**”时，需要填写。

  单击“上传”，选择上传私钥文件，请确保您的浏览器是最新版本。

  需注意必须是无密码的私钥。私钥格式如下：

  ```
  -----BEGIN PRIVATE KEY-----
  [key]
  -----END PRIVATE KEY-----
  ```

  说明

  若是证书链，则需要配置从子证书到根证书的所有证书内容，且证书内容的配置顺序需要为：子证书（服务器证书）> 中间证书 > 根证书。从权威机构颁发的证书，有可能根证书已经预置到服务器内，所以签发证书不包含根证书。此时直接按照“子证书（服务器证书）> 中间证书”完成配置。

  例如，某机构拿到的证书包含2个证书文件：子证书（服务器证书）文件**server.cer**、中间证书文件**mid.crt**和1个私钥文件**private.key**。那么需要在“证书内容”输入框中粘贴**server.cer**内容、然后回车继续粘贴**mid.crt**的内容，并且在“私钥”输入框中粘贴**private.key**的内容，才能使整个证书链生效。证书链内容格式如下：

  证书内容：

  ```
  -----BEGIN CERTIFICATE-----
  子证书（服务器证书）文件server.cer内容
  -----END CERTIFICATE-----
  -----BEGIN CERTIFICATE-----
  中间证书文件mid.crt内容
  -----END CERTIFICATE-----
  ```

  私钥：

  ```
  -----BEGIN PRIVATE KEY-----
  私钥文件private.key内容
  -----END PRIVATE KEY-----
  ```

- 域名

  如果创建的证书用于SNI，则需要指定域名，每个证书只能指定一个域名。且域名必须与证书中的域名一致。

- 描述

6. 填写完成后，单击“确定”。

修改证书

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏单击“证书管理”。
5. 在证书列表中，在需要修改的证书所在行，单击“修改”。
6. 在“修改证书”对话框中，修改证书的相关信息。
7. 确认修改信息，单击“确定”，完成修改。

删除证书

删除证书时，只能删除未使用的证书，在使用中的证书无法删除。

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏单击“证书管理”。
5. 在证书列表中，在需要修改的证书所在行，单击“删除”。
6. 在确认对话框中单击“是”，完成删除。

7.5 绑定/更换证书

操作场景

为了支持HTTPS数据传输加密认证，在创建HTTPS协议监听的时候需绑定证书，您可以参考本章节绑定证书。如果弹性负载均衡实例使用的证书过期或者其它原因需要更换，您可以参考本章节更换证书。

如果还有其他的服务也使用了待更换的证书，例如Web应用防火墙服务。请在所有服务上完成更换证书的操作，以免证书更换不全面而导致业务不可用。

说明

弹性负载均衡的证书和私钥的更换对业务没有影响。

前提条件

已经在弹性负载均衡的“证书管理”页面创建待更换的新证书，如果还未创建，请先**创建证书**。

绑定证书

通过添加HTTPS监听器来绑定证书。详见**添加HTTPS监听器**。

更换证书

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要修改HTTPS监听器的负载均衡名称。
5. 在“监听器”页签下，单击目标监听器所在行操作列的“编辑”。
6. “服务器证书”选择需要更换的证书。
7. 在“编辑监听器”对话框中，单击“确定”。

7.6 批量更换证书

操作场景

如果使用的证书过期或者其它原因需要更换，您可以通过修改证书功能批量更换监听器所绑定的证书。

说明

弹性负载均衡的证书和私钥的更换对业务没有影响。

约束与限制

- 只有HTTPS协议的监听器才支持绑定/更换证书，TCP/UDP/HTTP协议的监听器不支持绑定/更换证书。
- 切换证书后立即生效。

修改证书

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏单击“证书管理”。
5. 在证书列表中，在需要修改的证书所在行，单击“修改”。
6. 在“修改证书”对话框中，修改证书的相关信息。
7. 确认修改信息，单击“确定”，完成修改。

7.7 快速查询证书所关联的监听器

操作场景

您需要快速查询证书所关联的监听器，方便定位相关配置信息。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏单击“证书管理”。
5. 在证书列表中，在“监听器 (前端协议/端口)”所在列，单击监听器名称，即可查看监听器详细信息。

   当关联监听器数量大于5个，在“监听器 (前端协议/端口)”所在列，单击“查看所有”，单击监听器名称，即可查看监听器详细信息。

8 访问控制管理

8.1 访问控制策略

负载均衡器用户可以通过添加白名单和黑名单的方式控制访问负载均衡监听器的IP。通过白名单能够设置允许特定IP访问，而其它IP不许访问。通过黑名单能够设置允许特定的IP不能访问，而其它IP允许访问。

须知

- 设置白名单和黑名单存在一定业务风险。
  - 设置白名单，只有白名单中的IP可以访问负载均衡监听器。
  - 设置黑名单，黑名单中的IP不能访问负载均衡监听器。
- 访问流量的IP先通过白名单或黑名单访问控制，然后负载均衡转发流量，通过安全组安全规则限制，所以安全组的规则设置不会影响负载均衡的白名单或黑名单设置的访问控制。
- 访问控制只限制实际业务的流量转发，不限制ping命令操作，被限制的IP仍可以ping通后端服务器。
- 配置了白名单，但是不在白名单的IP也能访问后端服务器，可能的原因是该连接为长连接。需要客户端或后端服务器断开该长连接。访问控制策略对新建的连接是实时生效的。

设置访问控制策略

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击负载均衡名称，进入监听器管理界面。
5. 您可以通过以下两种操作入口，为监听器设置访问控制策略。
   - 在目标监听器所在行的“访问控制”列，单击“设置”。

- 单击目标监听器名称，进入监听器的基本信息页面，单击访问控制右侧的“设置”。

6. 在“设置访问控制”的弹窗中，如表8-1所示配置访问控制。

表 8-1 访问控制参数说明

参数	说明	样例
访问控制	可以选择允许所有IP访问、白名单和黑名单。 ● 允许所有IP访问：不进行访问控制，允许所有IP访问负载均衡监听器。 ● 白名单：仅允许IP地址组中的IP访问负载均衡监听器。 ● 黑名单：不允许IP地址组中的IP访问负载均衡监听器。	黑名单
IP地址组	设置白名单或者黑名单时，必须选择一个IP地址组。如果还未创建IP地址组，需要先创建IP地址组，更多关于IP地址组的信息请参见IP地址组。	ipGroup-b2
访问控制开关	当访问控制选择白名单或者黑名单时，可以开启或者关闭访问控制开关。 ● 开启：开启访问控制开关，设置的白名单和黑名单才会生效。 ● 关闭：关闭访问控制开关，设置的白名单和黑名单不生效。	-

7. 配置完成，单击“确定”。

8.2 访问控制 IP 地址组

8.2.1 创建 IP 地址组

IP 地址组简介

IP地址组是多个IP地址的集合，用来统一管理具有相同安全要求或需要频繁修改的IP地址。

弹性负载均衡支持对监听器设置访问控制策略。对于需要使用**黑名单**和**白名单**，在监听器上设置**访问控制**的用户，开启白名单或黑名单时必须选择一个IP地址组。

- 白名单：允许IP地址组中的IP访问负载均衡的监听器。如果IP地址组未包含任何IP地址，则对应的负载均衡监听器禁止任何IP地址访问。
- 黑名单：限制IP地址组中的IP访问负载均衡的监听器。如果IP地址组未包含任何IP地址，则对应的负载均衡监听器允许所有IP地址访问。

约束与限制

- 默认情况下，一个用户可以创建50个IP地址组。
- 同一个IP地址组，最多可以关联50个监听器。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > IP地址组”。
5. 在“IP地址组”界面，单击“创建IP地址组”。
6. 配置IP地址组参数，参数说明参见**表8-2**。

表 8-2 IP 地址组参数说明

参数	说明	样例
名称	IP地址组的名称。	ipGroup-01
IP地址	需要通过白名单或黑名单进行访问控制的IP地址，支持IPv4地址和IPv6地址。 ● 每行一个IP地址或一个网段，以回车结束； ● 每个IP地址或者网段都可以用“|”分隔添加备注，如“192.168.10.10 | ECS01”，备注长度范围是0到255字符，不能包含<>； ● 每个IP地址组最多可添加300个IP地址或网段。	10.168.2.24 10.168.16.0/24
描述	IP地址组相关信息的描述说明。	-

7. 确认参数配置，单击“确定”。

8.2.2 查看 IP 地址组详情

操作场景

本章节指导用户查看IP地址组的详情，包括如下信息：

- IP地址组的基本信息，包括IP地址组的名称、ID、创建时间和描述。
- IP地址组内添加的IP地址。
- IP地址组关联的监听器资源。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > IP地址组”。
5. 在“IP地址组”界面，单击需要查看的IP地址组名称。
   进入IP地址组的详情页面。
6. 支持直接查看IP地址组基本信息。
   a. 在“IP地址”页签下，查看IP地址组内的IP地址条目。
   b. 在“关联监听器”页签下，查看IP地址组已关联的监听器。

8.2.3 管理 IP 地址组内的 IP 地址

IP地址组创建后，您可根据使用需求对组内的IP地址进行修改，支持的修改操作如下：

- **添加IP地址**
- **批量修改IP地址**
- **删除IP地址**

约束与限制

输入IP地址支持的格式如下：

- 每行一个IP地址或一个网段，以回车结束。
- 每个IP地址或者网段都可以用“|”分隔添加备注，如“192.168.10.10 | ECS01”，备注长度范围是0到255字符，不能包含<>。
- 每个IP地址组最多可添加300个IP地址或网段。

添加 IP 地址

IP地址组创建后您可向其中添加IP地址，不影响IP地址组中已有的IP地址。

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。

3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > IP地址组”。
5. 在“IP地址组”界面，单击需要添加IP地址的地址组名称。
   进入IP地址组的详情页面。
6. 在IP地址页签下方，单击“添加IP地址”。
7. 在“添加IP地址”页面，添加IP地址。
8. 单击“确定”，完成添加。

批量修改 IP 地址

如果您希望对IP地址组内的所有IP地址进行批量修改，请参考以下操作。

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > IP地址组”。
5. 在“IP地址组”界面，您可以通过以下两种操作入口，批量修改IP地址。
   a. 批量修改IP地址及其基本信息：
      i. 在需要修改IP地址的地址组所在行的操作列，单击“修改”。
      ii. 支持修改IP地址组的名称，组内所有IP地址和描述。
      iii. 单击“确定”，完成修改。
   b. 仅批量修改IP地址：
      i. 单击需要修改IP地址的地址组名称，进入IP地址组的详情页面。
      ii. 在IP地址页签下方，单击“修改IP地址”。
      iii. 支持修改IP地址组的内所有IP地址
      iv. 单击“确定”，完成修改。

删除 IP 地址

如果你希望批量删除IP地址组内的多个IP地址，请参考**批量修改IP地址**。

如果您希望对IP地址组内的单IP地址进行删除，请参考以下操作。

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > IP地址组”。
5. 在“IP地址组”界面，单击需要修改IP地址的地址组名称。
   进入IP地址组的详情页面。
6. 在IP地址列表中，单击目标IP地址所在行的“删除”。
   弹出删除确认对话框。

7. 确认无误后，单击“是”，删除IP地址。

8.2.4 删除 IP 地址组

操作场景

本章节指导用户删除IP地址组。

约束与限制

如果IP地址组已经关联监听器的访问控制策略使用，无法完成删除。您可在IP地址组列表页或通过**查看IP地址组详情**查看IP地址组已关联的监听器资源，解除IP地址组与监听器的关联请参考**设置访问控制策略**。

操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“网络 > 弹性负载均衡”。
4. 在左侧导航栏，选择“ 弹性负载均衡 > IP地址组”。
5. 在“IP地址组”界面，需要删除的IP地址组所在行，单击“删除”。
6. 确认需要删除的IP地址组，单击“是”。

9 TLS 安全策略

操作场景

对于银行，金融类加密传输的应用 ，在创建和配置HTTPS监听器时，您可以选择使用安全策略，可以提高您的业务安全性。安全策略包含TLS协议版本和配套的加密算法套件。

添加安全策略

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要创建安全策略的监听器的负载均衡器名称。
5. 在该负载均衡界面的“监听器”区域，单击“添加监听器”。
6. 在“添加监听器”界面，前端协议选择“HTTPS”。
7. 在“添加监听器”界面，选择“高级配置 > 安全策略”。
   默认策略如**表9-1**所示。

表 9-1 默认安全策略参数说明

名称	说明	支持的TLS版本类型	使用的加密套件列表
TLS-1-0	支持TLS 1.0、TLS 1.1、TLS 1.2版本与相关加密套件，兼容性好，安全性低。	TLS 1.2 TLS 1.1 TLS 1.0	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-GCM-SHA256 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-RSA-AES128-SHA256 • AES128-SHA256 • AES256-SHA256 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES128-SHA • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA • ECDHE-ECDSA-AES256-SHA • AES128-SHA • AES256-SHA
TLS-1-1	支持TLS 1.1、TLS 1.2版本与相关加密套件，兼容性较好，安全性中。	TLS 1.2 TLS 1.1	
TLS-1-2	支持TLS 1.2版本与相关加密套件，兼容性较好，安全性高。	TLS 1.2	

名称	说明	支持的TLS版本类型	使用的加密套件列表
TLS-1-2-Strict	支持TLS 1.2版本与相关加密套件，兼容性一般，安全性高。	TLS 1.2	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-GCM-SHA256 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-RSA-AES128-SHA256 • AES128-SHA256 • AES256-SHA256 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES256-SHA384

名称	说明	支持的TLS版本类型	使用的加密套件列表
TLS-1-0-WITH-1-3	支持TLS 1.0及以上版本与相关加密套件，兼容性最好，安全性低。	TLS 1.3 TLS 1.2 TLS 1.1 TLS 1.0	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-GCM-SHA256 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-RSA-AES128-SHA256 • AES128-SHA256 • AES256-SHA256 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES128-SHA • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA • ECDHE-ECDSA-AES256-SHA • AES128-SHA • AES256-SHA • TLS_AES_128_GCM_SHA256 • TLS_AES_256_GCM_SHA384 • TLS_CHACHA20_POLY1305_SHA256 • TLS_AES_128_CCM_SHA256 • TLS_AES_128_CCM_8_SHA256

名称	说明	支持的TLS版本类型	使用的加密套件列表
TLS-1-2-FS-WITH-1-3	支持TLS 1.2及以上版本与前向安全相关的加密套件，兼容性较好，安全性最高。	TLS 1.3 TLS 1.2	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-RSA-AES128-SHA256 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES256-SHA384 • TLS_AES_128_GCM_SHA256 • TLS_AES_256_GCM_SHA384 • TLS_CHACHA20_POLY1305_SHA256 • TLS_AES_128_CCM_SHA256 • TLS_AES_128_CCM_8_SHA256

名称	说明	支持的TLS版本类型	使用的加密套件列表
hybrid-policy-1-0	支持TLS 1.1、TLS 1.2版本与相关加密套件，兼容性较好，安全性中。	TLS 1.2 TLS 1.1	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-GCM-SHA256 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-RSA-AES128-SHA256 • AES128-SHA256 • AES256-SHA256 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES128-SHA • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA • ECDHE-ECDSA-AES256-SHA • AES128-SHA • AES256-SHA • ECC-SM4-SM3 • ECDHE-SM4-SM3

说明

- 上述列表为ELB支持的加密套件，同时客户端也支持多个加密套件，这样在实际使用时，加密套件的选择范围为：ELB和客户端支持的加密套件的交集，加密套件的选择顺序为：ELB支持的加密套件顺序。

8. 配置完成，单击“确定”。

安全策略差异说明

表 9-2 安全策略差异说明

安全策略	TLS-1-0	TLS-1-1	TLS-1-2	TLS-1-2-Strict
TLS 协议				

安全策略	TLS-1-0	TLS-1-1	TLS-1-2	TLS-1-2-Strict
Protocol-TLS 1.3	-	-	-	-
Protocol-TLS 1.2	√	√	√	√
Protocol-TLS 1.1	√	√	-	-
Protocol-TLS 1.0	√	-	-	-
加密套件				
EDHE-RSA-AES128-GCM-SHA256	√	√	√	√
ECDHE-RSA-AES256-GCM-SHA384	√	√	√	√
ECDHE-RSA-AES128-SHA256	√	√	√	√
ECDHE-RSA-AES256-SHA384	√	√	√	√
AES128-GCM-SHA256	√	√	√	√
AES256-GCM-SHA384	√	√	√	√
AES128-SHA256	√	√	√	√
AES256-SHA256	√	√	√	√
ECDHE-RSA-AES128-SHA	√	√	√	-
ECDHE-RSA-AES256-SHA	√	√	√	-
AES128-SHA	√	√	√	-
AES256-SHA	√	√	√	-
ECDHE-ECDSA-AES128-GCM-SHA256	√	√	√	√
ECDHE-ECDSA-AES128-SHA256	√	√	√	√
ECDHE-ECDSA-AES128-SHA	√	√	√	-
ECDHE-ECDSA-AES256-GCM-SHA384	√	√	√	√
ECDHE-ECDSA-AES256-SHA384	√	√	√	√
ECDHE-ECDSA-AES256-SHA	√	√	√	-

修改安全策略

修改安全策略时，后端需要放通安全组，放开对ELB健康检查的限制（100.125IP的限制，UDP健康检查icmp报文的限制等），否则后端健康检查没上线，会影响业务。

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的 ，选择“网络 > 弹性负载均衡”。
4. 在“负载均衡器”界面，单击需要修改安全策略的监听器的负载均衡器名称。
5. 切换至“监听器”页签，单击需要修改安全策略的监听器名称。
6. 在监听器的基本信息页面，单击“编辑监听器”。
7. 在“编辑监听器”界面，展开高级配置，选择安全策略参数。
8. 单击“确定”。

10 访问日志

操作场景

负载均衡的访问日志功能支持查看和分析对七层负载均衡HTTP和HTTPS进行请求的详细访问日志记录，包括请求时间、客户端IP地址、请求路径和服务器响应等。配置访问日志时需要您对接云日志服务，并且已经创建需要关联的云日志组和日志流。

说明

由于弹性负载均衡会将访问日志等运维数据内容展示到云日志服务控制台，请您在使用过程中，注意您的隐私及敏感信息数据保护，不建议将隐私或敏感数据通过访问日志涉及的的字段传输，必要时请加密保护。

配置云日志服务

为了能够在云日志服务上面看到弹性负载均衡的日志，需要配置云日志服务。关于云日志服务的详细配置和操作方法，请参见《云日志服务用户指南》

1. 在“云日志服务”界面创建日志组。
   a. 登录管理控制台。
   b. 在管理控制台左上角单击图标，选择区域和项目。
   c. 单击页面左上角的，选择“ 管理与部署 > 云日志服务”。
   d. 单击左侧导航栏“日志管理”。
   e. 单击 “创建日志组”，在弹出框内，输入日志组名称。
      根据实际需要设置“日志存储时间（天）”。
   f. 单击“确定”，创建完成。
2. 在“云日志服务”界面创建日志流。
   a. 在云日志服务管理控制台，单击日志组名称对应的 按钮。
   b. 单击“创建日志流”，在弹出框内，输入日志流名称。
   c. 单击“确定”，创建完成。

配置 ELB 访问日志

在“弹性负载均衡”界面配置访问日志。

1. 单击页面左上角的 ，选择“网络 > 弹性负载均衡”。
2. 在“负载均衡器”界面，单击需要配置访问日志的负载均衡器名称。
3. 在该负载均衡器界面的“访问日志”页签，单击“配置访问日志”。
4. 开启日志记录，选择您在云日志服务中创建的日志组和日志流。
5. 单击“确定”，配置完成。

查看访问日志

当您配置了访问日志，可以查看访问日志的详细信息。

查看方式以下两种：

- 通过“弹性负载均衡”控制台，进入访问日志界面，即可查看访问日志。
- （推荐）通过“云日志服务”控制台，进入日志主题界面，选择相应日志主题名称，单击“实时日志”，即可查看访问日志。

日志显示格式如下，日志字段说明如表10-1所示。不支持修改日志格式。

```
$msec $access_log_topic_id [$time_iso8601] $log_ver $remote_addr:$remote_port $status
"$request_method $scheme://$host$router_request_uri $server_protocol" $request_length $bytes_sent
$body_bytes_sent $request_time "$upstream_status" "$upstream_connect_time" "$upstream_header_time"
"$upstream_response_time" "$upstream_addr" "$http_user_agent" "$http_referer" "$http_x_forwarded_for"
$lb_name $listener_name $listener_id
$pool_name "$member_name" $tenant_id $eip_address:$eip_port "$upstream_addr_priv" $certificate_id
$ssl_protocol $ssl_cipher $sni_domain_name $tcpinfo_rtt $self_defined_header
```

表 10-1 字段说明

参数	描述	取值说明	取值示例
msec	以秒为单位的时间，日志写入时的分辨率为毫秒。	浮点型数据	1530153091.868
access_log_topic_id	访问日志流ID。	uuid	04465dfa-640f-4567-8b58-45c9f8bbc23f
time_iso8601	日志写入时的时间，采用ISO 8601标准格式本地时间。	-	2018-06-28T10:31:31+08:00
log_ver	ELB服务日志版本号。	固定值：elb_01	elb_01
remote_addr:remote_port	客户端IP地址：客户端端口。	记录客户端IP地址和客户端端口号。	10.184.30.170:59605
status	ELB响应的状态码。	记录请求状态码。	200

参数	描述	取值说明	取值示例
request_method scheme:// host request_uri server_protocol	请求方法。请求方式：//主机名：请求URI 请求协议。	• request_method：请求方法。 • scheme: http或https。 • host: 主机名，可能为域名或者IP。 • request_uri：浏览器发起的不做任何修改的原生URI。不包括协议及主机名。	POST https://setting1.hicloud.com/AccountServer/IUserInfoMng/stAuth?Version=26400&cVersion=ID_SDK_2.6.4.300
request_length	从客户端收到的请求长度（包括请求header和请求body）。	整型数据	295
bytes_sent	发送到客户端的字节数 。	整型数据	58470080
body_bytes_sent	发送到客户端的字节数（不包括响应头）。	整型数据	58469792
request_time	请求处理时间，即ELB收到第一个客户端请求报文到ELB发送完响应报文的时间间隔（单位：秒）。	浮点型数据	499.769
upstream_status	从上游服务器获得的响应状态码，当ELB代理进行请求重试时会包含多个响应的状态码，当请求未被正确转发到后端服务器时此字段为 -。	后端返回给ELB的状态码	200 或者 "-, 200"，或者"502, 502 : 200"，或者"502 : "
upstream_connect_time	与上游服务器建立连接所花费的时间，时间以秒为单位，分辨率为毫秒。当ELB代理进行请求重试时会包含多个连接的时间，当请求未被正确转发到后端服务器时此字段为 -。	浮点型数据	0.008 或者 "-, 0.008" ，或者 "0.008, 0.005 : 0.004"，或者 "0.008 : "

参数	描述	取值说明	取值示例
upstream_header_time	从上游服务器接收响应头所花费的时间，时间以秒为单位，分辨率为毫秒。当ELB代理进行请求重试时会包含多个响应时间，当请求未被正确转发到后端服务器时此字段为 -。	浮点型数据	0.008 或者 "-, 0.008" ，或者 "0.008, 0.005 : 0.004"，或者 "0.008 : "
upstream_response_time	从上游服务器接收响应所花费的时间，时间以秒为单位，分辨率为毫秒。当ELB代理进行请求重试时会包含多个响应时间，当请求未被正确转发到后端服务器时此字段为 -。	浮点型数据	0.008 或者 "-, 0.008" ，或者 "0.008, 0.005 : 0.004"，或者 "0.008 : "
upstream_addr	后端主机的IP地址和端口号。可能有多个值，每个值都是ip:port或者-，用逗号空格隔开。 （该参数适用于独享型负载均衡）	IP地址+端口号	-，192.168.1.2:8080（可能有多个值，每个值都是ip:port或者-,用逗号空格隔开）
http_user_agent	ELB收到请求头中的http_user_agent内容，表示客户端的系统型号、浏览器信息等。	记录浏览器的相关信息	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.99 Safari/537.36
http_referer	ELB收到请求头中的http_referer内容，表示该请求所在的页面链接。	页面链接请求	http://10.154.197.90/
http_x_forwarded_for	ELB收到请求头中的http_x_forwarded_for内容，表示请求经过的代理服务器IP地址。	IP地址	10.154.197.90

参数	描述	取值说明	取值示例
lb_name	负载均衡器的名称（格式为“loadbalancer_”+“负载均衡器ID”）。	字符串	loadbalancer_789424af-3fd2-4292-8c62-2a2dd7005175
listener_name	监听器的名称（格式为“listener_”+“监听器ID”）。	字符串	listener_fde03b66-f960-440e-954a-0be8b2b75093
listener_id	监听器在ELB服务内部的ID（客户可忽略）。	字符串	-
pool_name	后端服务器组名称（格式为“pool_”+“后端服务器组ID”）。	字符串	pool_066a5dc5-a3e4-4ea1-99f1-2a5716b681f6
member_name	后端服务器的名称（格式为“member_”+“服务器ID”，尚未支持）。可能有多个值，每个值都是member_id或者-，用逗号空格隔开。	字符串	member_47b07465-075a-4d2f-8ce9-0b9f39bff160(可能有多个值，每个值都是member_id或者-，用逗号空格隔开)
tenant_id	租户ID。	字符串	04dd36f921000fe20f95c00bba986340
eip_address:eip_port	弹性IP地址和监听器监听的端口号。	弹性IP地址和监听器监听的端口号。	4.17.12.248:443
upstream_addr_priv	后端主机的IP地址和端口号。可能有多个值，每个值都是ip:port或者-，用逗号空格隔开。 （该参数适用于共享型负载均衡）	IP地址+端口号	-，192.168.1.2:8080（可能有多个值，每个值都是ip:port或者-,用逗号空格隔开）
certificate_id	[HTTPS监听器]SSL连接建立时使用的证书ID（尚未支持）。	字符串	17b03b19-b2cc-454e-921b-4d187cce31dc
ssl_protocol	[HTTPS监听器]SSL连接建立使用的协议，非HTTPS监听器，此字段为 -。	字符串	TLS 1.2

参数	描述	取值说明	取值示例
ssl_cipher	[HTTPS监听器]SSL连接建立使用的加密套件，非HTTPS监听器，此字段为 -。	字符串	ECDHE-RSA-AES256-GCM-SHA384
sni_domain_name	[HTTPS监听器]SSL握手时客户端提供的SNI域名，非HTTPS监听器，此字段为 -。	字符串	www.test.com
tcpinfo_rtt	ELB与客户端之间的tcp rtt时间，单位：微秒。	整型数据	39032
self_defined_header	该字段为保留字段，默认为“-”。	字符串	-

日志示例

```
1644819836.370 eb11c5a9-93a7-4c48-80fc-03f61f638595 [2022-02-14T14:23:56+08:00] elb_01
192.168.1.1:888 200 "POST https://www.test.com/example /HTTP/1.1" 1411 251 3 0.011 "200" "0.000"
"0.011" "0.011" "100.64.0.129:8080" "okhttp/3.13.1" "-" "-"
loadbalancer_295a7eee-9999-46ed-9fad-32a62ff0a687 listener_20679192-8888-4e62-a814-a2f870f62148
3333fd44fe3b42cbaa1dc2c641994d90 pool_89547549-6666-446e-9dbc-e3a551034c46 "-"
f2bc165ad9b4483a9b17762da851bbbb 121.64.212.1:443 "10.1.1.2:8080" - TLSv1.2 ECDHE-RSA-AES256-
GCM-SHA384 www.test.com 56704 -
```

以上日志示例对应的字段如下：

表 10-2 日志示例对应的字段

参数	示例
msec	1644819836.370
access_log_topic_id	eb11c5a9-93a7-4c48-80fc-03f61f638595
time_iso8601	[2022-02-14T14:23:56+08:00]
log_ver	elb_01
remote_addr: remote_port	192.168.1.1:888
status	200
request_method scheme://host request_uri server_protocol	"POST https://www.test.com/example/1 HTTP/1.1"
request_length	1411
bytes_sent	251
body_bytes_sent	3

参数	示例
request_time	0.011
upstream_status	"200"
upstream_connect_time	"0.000"
upstream_header_time	"0.011"
upstream_response_time	"0.011"
upstream_addr	"100.64.0.129:8080"
http_user_agent	"okhttp/3.13.1"
http_referer	"-"
http_x_forwarded_for	"-"
lb_name	loadbalancer_295a7eee-9999-46ed-9fad-32a62ff0a687
listener_name	listener_20679192-8888-4e62-a814-a2f870f62148
listener_id	3333fd44fe3b42cbaa1dc2c641994d90
pool_name	pool_89547549-6666-446e-9dbc-e3a551034c46
member_name	"-"
tenant_id	f2bc165ad9b4483a9b17762da851bbbb
eip_address:eip_port	121.64.212.1:443
upstream_addr_priv	"10.1.1.2:8080"
certificate_id	-
ssl_protocol	TLSv1.2
ssl_cipher	ECDHE-RSA-AES256-GCM-SHA384
sni_domain_name	www.test.com
tcpinfo_rtt	56704
self_defined_header	-

日志分析：

在[2022-02-14T14:23:56+08:00]时，ELB接收到客户端地址和端口（192.168.1.1:888）发起的“POST /HTTP/1.1”请求，ELB将请求转发给后端服务器（100.64.0.129:8080），后端服务器响应状态码200，ELB最终向客户端响应状态码200。

分析结果：

后端服务器正常响应请求。

配置日志转储

如果您希望将日志转储进行二次分析，您可以参考本章设置日志转储。

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“管理与部署 > 云日志服务”。
4. 在左侧导航栏，单击“日志转储”。
5. 在日志转储页面，单击“配置转储”。

1. 根据实际情况设置转储方式和其他配置项，具体操作请参见《云日志服务用户指南》。

11 监控

11.1 监控指标说明

功能说明

本节定义了弹性负载均衡服务上报云监控的监控指标的命名空间，监控指标列表和维度定义。您可以在云监控服务控制台查看弹性负载均衡服务上报的监控指标以及产生告警信息，详见**查看监控指标**。

命名空间

SYS.ELB

监控指标

表 11-1 ELB 支持的监控指标

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
m1_cps	并发连接数	在四层负载均衡器中，指从测量对象到后端服务器建立的所有TCP和UDP连接的数量。 在七层负载均衡器中，指从客户端到ELB建立的所有TCP连接的数量。 单位：个	≥ 0个	• 负载均衡器 • 负载均衡监听器	1分钟

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
m2_act_conn	活跃连接数	从测量对象到后端服务器建立的所有**ESTABLISHED**状态的TCP或UDP连接的数量。 Windows和Linux服务器都可以使用如下命令查看。 `netstat -an` 单位：个	≥ 0个		
m3_inact_conn	非活跃连接数	从测量对象到所有后端服务器建立的所有除**ESTABLISHED**状态之外的TCP连接的数量。 Windows和Linux服务器都可以使用如下命令查看。 `netstat -an` 单位：个	≥ 0个		
m4_ncps	新建连接数	从客户端到测量对象每秒新建立的连接数。 单位：个/秒	≥ 0个/秒		
m5_in_pps	流入数据包数	测量对象每秒接收到的数据包的个数。 单位：个/秒	≥ 0个/秒		
m6_out_pps	流出数据包数	测量对象每秒发出的数据包的个数。 单位：个/秒	≥ 0个/秒		
m7_in_Bps	网络流入速率	从外部访问测量对象所消耗的流量。 单位：字节/秒	≥ 0bytes/s		
m8_out_Bps	网络流出速率	测量对象访问外部所消耗的流量。 单位：字节/秒	≥ 0bytes/s		
m9_abnormal_servers	异常主机数	健康检查统计监控对象后端异常的主机个数。 单位：个	≥ 0个	• 负载均衡器	1分钟

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
ma_normal_servers	正常主机数	健康检查统计监控对象后端正常的主机个数。 单位：个	≥ 0个		
mb_l7_qps	7层查询速率	统计测量对象当前7层查询速率。（HTTP和HTTPS监听器才有此指标） 单位：次/秒。	≥ 0次/秒	• 负载均衡器 • 负载均衡监听器	1分钟
md_l7_http_3xx	7层协议返回码(3XX)	统计测量对象当前7层3XX系列状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒	• 负载均衡器 • 负载均衡监听器	1分钟
mc_l7_http_2xx	7层协议返回码(2XX)	统计测量对象当前7层2XX系列状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒	• 负载均衡器 • 负载均衡监听器	1分钟
me_l7_http_4xx	7层协议返回码(4XX)	统计测量对象当前7层4XX系列状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒		
mf_l7_http_5xx	7层协议返回码(5XX)	统计测量对象当前7层5XX系列状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒		
m10_l7_http_other_status	7层协议返回码(Others)	统计测量对象当前7层非2XX,3XX,4XX,5XX系列状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒		
m11_l7_http_404	7层协议返回码(404)	统计测量对象当前7层404状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒		

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
m12_l7_http_499	7层协议返回码(499)	统计测量对象当前7层499状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒		
m13_l7_http_502	7层协议返回码(502)	统计测量对象当前7层502状态响应码的数量。（HTTP和HTTPS监听器才有此指标） 单位：个/秒。	≥ 0个/秒		
m14_l7_rt	7层协议RT平均值	统计测量对象当前7层平均响应时间。（HTTP和HTTPS监听器才有此指标） 从测量对象收到客户端请求开始，到测量对象将所有响应返回给客户端为止。 单位：毫秒。 **说明** websocket场景下RT平均值可能会非常大，此时该指标无法作为时延指标参考。	≥ 0ms		

维度

Key	Value
lbaas_instance_id	负载均衡器的ID。
lbaas_listener_id	负载均衡监听器的ID。
lbaas_pool_id	后端服务器组的ID

11.2 设置告警规则

本章节主要介绍添加、修改告警规则，删除告警规则详见《云监控服务用户指南》。

11.2.1 添加告警规则

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“管理与部署 > 云监控服务”。
4. 在左侧导航树栏，选择“告警 > 告警规则”。
5. 在“告警规则”界面，单击“创建告警规则”进行添加。

   以创建弹性负载均衡器的告警规则为例：

   a. “资源类型”，选择“弹性负载均衡”。
   b. “维度”，可以选择“负载均衡器”或“监听器”，这里以选择“负载均衡器”为例。
   c. 单击“监控范围”，选择已经创建的负载均衡器。
   d. 按照需要设置其他参数，修改完成后单击“立即创建”。

   弹性负载均衡器告警规则设置完成后，如果通知功能已开启，当符合规则的告警产生时，系统会自动进行通知。

   说明

   更多关于弹性负载均衡器监控规则的信息，请参见《云监控服务用户指南》。

11.2.2 修改告警规则

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 单击页面左上角的，选择“管理与部署 > 云监控服务”。
4. 在左侧导航树栏，选择“告警 > 告警规则”。
5. 在“告警规则”界面，单击需修改的告警规则操作列的“修改”。

   a. 在“修改告警规则”界面，根据界面提示修改配置参数。
   b. 按照需要设置其他参数，修改完成后单击“立即修改”。

   弹性负载均衡器告警规则设置完成后，如果通知功能已开启，当符合规则的告警产生时，系统会自动进行通知。

   说明

   更多关于弹性负载均衡器监控规则的信息，请参见《云监控服务用户指南》。

11.3 查看监控指标

操作场景

云服务平台提供的云监控服务，可以对弹性负载均衡器的运行状态进行日常监控。

您可以通过云监控管理控制台，查看弹性负载均衡器的各项监控指标。

由于监控数据的获取与传输会花费一定时间，因此，云监控显示的是当前时间5～10分钟前的弹性负载均衡状态。如果您的弹性负载均衡器刚刚创建完成，请等待5～10分钟后查看监控数据。

前提条件

- 已经正常运行了一段时间的弹性负载均衡器。

  关机、故障、删除状态的后端服务器，无法在云监控中查看其监控指标。当后端服务器再次启动或恢复后，即可正常查看。

  说明

  关机、故障24小时以上的后端服务器，云监控将默认该负载均衡器不存在，并在监控列表中删除，不再对其进行监控，但告警规则需要用户手动清理。

- 负载均衡器已对接云监控服务，即已在云监控服务页面设置告警规则。

  对接云监控服务之前，用户无法查看到未对接资源的监控数据。具体操作，请参见**设置告警规则**。

- 子账号用户如果需要在云监控页面中查看ELB监控数据，需要为子账号添加“ELB Administrator”权限，否则无法查询到完整的ELB监控数据。

在云监控服务控制台查看监控指标

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击页面左上角的  ，选择“管理与部署 > 云监控服务”。
4. 在左侧导航树选择“云服务监控 > 弹性负载均衡”。
5. 在“云服务监控”页面，单击需要查看监控指标的负载均衡器名称。或单击目标负载均衡器右侧操作列的“查看监控指标”。
6. 选择需要查看监控指标的时间段。支持选择系统定义的时间段（如“近1小时”），或自定义时间段。
7. 单击右上角的“设置监控指标”，设置需要查看的监控指标。

说明

查看云服务监控指标详见《云监控服务用户指南》。

12 关于配额

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。如您最多可以创建多少台弹性云服务器、多少块云硬盘。

如果当前资源配额限制无法满足使用需要，您可以申请扩大配额。

怎样查看我的配额?

1. 登录管理控制台。
2. 单击页面右上角的“My Quota”图标 。
   系统进入“服务配额”页面。
3. 您可以在“服务配额”页面，查看各项资源的总配额及使用情况。
   如果当前配额不能满足业务要求，请参考后续操作，申请扩大配额。

如何申请扩大配额?

目前系统暂不支持在线调整配额大小。如您需要调整配额，请联系运营管理员。

在联系运营管理员之前，请您准备好以下信息：

- 账号名，获取方式如下：
  登录云账户管理控制台，在右上角单击用户名，选择“我的凭证”，在“我的凭证”页面获取“账号名”。
- 配额信息，包括：服务名、配额类别、需要的配额值。

13 常见问题

13.1 高频常见问题

- 如何获取来访者的真实IP?
- 使用UDP协议有什么注意事项?
- ELB支持什么类型的会话保持?
- 如何启用WebSocket支持?
- 如何检查弹性负载均衡会话保持不生效?
- 监听器中分配算法和会话保持算法是什么关系?
- ELB如何根据不同的协议来分发流量?

13.2 功能支持

13.2.1 弹性负载均衡器是否可以单独使用?

不可以。

弹性负载均衡是将访问流量根据分配策略分发到后端多台服务器的流量分发控制服务，通过流量分发扩展应用系统对外的服务能力，同时通过消除单点故障提升应用系统的可用性。因此弹性负载均衡要基于后端实例（如：弹性云服务器）来使用，不可以单独使用。

13.2.2 ELB 是否支持 TCP 长连接?

支持。

客户端到ELB之间支持TCP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。

13.2.3 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。但是可以支持SFTP场景。

13.2.4 弹性负载均衡分配的 EIP 是否为独占?

在您创建使用ELB服务的整个生命周期内：弹性公网IP支持解绑，解绑后的负载均衡变成私网型负载均衡，解绑后的弹性公网IP可被其他资源绑定。

13.2.5 单个用户默认可以创建多少个负载均衡器或监听器?

单个用户默认可创建50个负载均衡器，默认可创建100个监听器。如果需要创建更多弹性负载均衡器或监听器，请申请更高配额。

单个弹性负载均衡器下可创建的监听器个数，与当前用户下的监听器剩余配额相等。

13.2.6 ELB 权限和使用范围是什么?

ELB目前提供的系统策略如下：

表 13-1 策略说明

策略类型	策略名	策略说明
RBAC策略	ELB Administrator	弹性负载均衡的所有执行权限。 给用户组授予RBAC策略时，需要确认是否有依赖的策略，如果有，需要同时设置依赖的权限，授予的RBAC权限才会生效。
细粒度策略	ELB FullAccess	弹性负载均衡的所有执行权限。 如果不开通细粒度策略功能，给用户组授权时，无法使用细粒度策略。
	ELB ReadOnlyAccess	弹性负载均衡的只读权限。

表 13-2 常用操作与系统策略的关系

操作	ELB FullAccess	ELB ReadOnlyAccess	ELB Administrator
创建负载均衡器	√	×	√
查询负载均衡器	√	√	√
查询负载均衡器状态树	√	√	√
查询负载均衡器列表	√	√	√
更新负载均衡器	√	×	√
删除负载均衡器	√	×	√
创建监听器	√	×	√

操作	ELB FullAccess	ELB ReadOnlyAccess	ELB Administrator
查询监听器	√	√	√
修改监听器	√	×	√
删除监听器	√	×	√
创建后端服务器组	√	×	√
查询后端服务器组	√	√	√
修改后端服务器组	√	×	√
删除后端服务器组	√	×	√
创建后端服务器	√	×	√
查询后端服务器	√	√	√
修改后端服务器	√	×	√
删除后端服务器	√	×	√
创建健康检查	√	×	√
查询健康检查	√	√	√
修改健康检查	√	×	√
关闭健康检查	√	×	√
创建弹性公网IP	×	×	√
绑定弹性公网IP	×	×	√
查询弹性公网IP	√	√	√
解绑弹性公网IP	×	×	√
查看监控指标	×	×	√
查看访问日志	×	×	√

关于ELB的细粒度权限详见《弹性负载均衡API参考》。

13.2.7 当负载均衡器正在运行中是否可以调整后端服务器的数量?

我们支持在任意时刻增加或减少负载均衡器的后端服务器的数量，且可以支持不同的后端服务器切换操作。但是，为了保证您对外业务的稳定，请确保在执行上述操作时能够开启负载均衡器的健康检查功能，并同时保证负载均衡后端至少有1台正常运行的服务器。

13.2.8 弹性负载均衡是否可以添加不同操作系统的服务器?

可以。

ELB本身不会限制后端的服务器使用哪种操作系统，只要您的2台服务器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台相同操作系统的服务器进行配置，以便您日后的管理维护。

13.2.9 ELB 添加后端的端口号是否可以不一致?

弹性负载均衡添加后端的端口号可以不同。

13.2.10 ELB 支持跨用户、跨 VPC 使用么?

- 独享型负载均衡实例支持混合负载均衡的能力，后端服务器组不仅支持添加云上VPC内的服务器，还支持添加其他VPC、其他Region、云下数据中心的服务器。详情请参见。

13.2.11 负载均衡器的后端服务器可以反过来访问公网/私网负载均衡器上的端口吗?

不可以。

13.2.12 ELB 能否实现前端是 HTTPS 协议，后端也是 HTTPS 协议?

独享型ELB支持，共享型ELB不支持。

独享型负载均衡支持全链路HTTPS数据传输，即在添加监听器时，前端协议选择“HTTPS”，后端协议也支持选择“HTTPS”。添加监听器请参见《弹性负载均衡用户指南》中的“添加HTTPS监听器”。

13.2.13 ELB 所属的 VPC 和子网支持修改吗?

独享型负载均衡仅支持修改子网，不支持修改VPC。

13.3 负载均衡器

13.3.1 ELB 如何根据不同的协议来分发流量?

ELB采用“FullNAT”模式转发。如下图所示，四层协议转发经过LVS，七层转发协议，经过LVS后再到NGINX。

说明

“FullNAT”是转发模式，是指LVS会转换客户端的源IP和目的IP。

图 13-1 四层转发协议

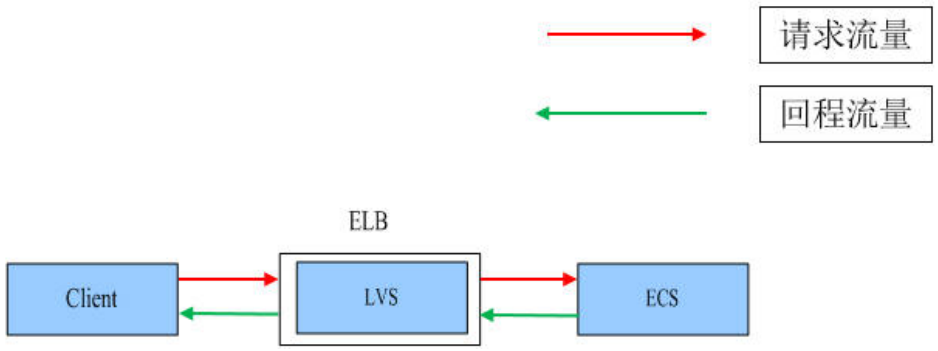
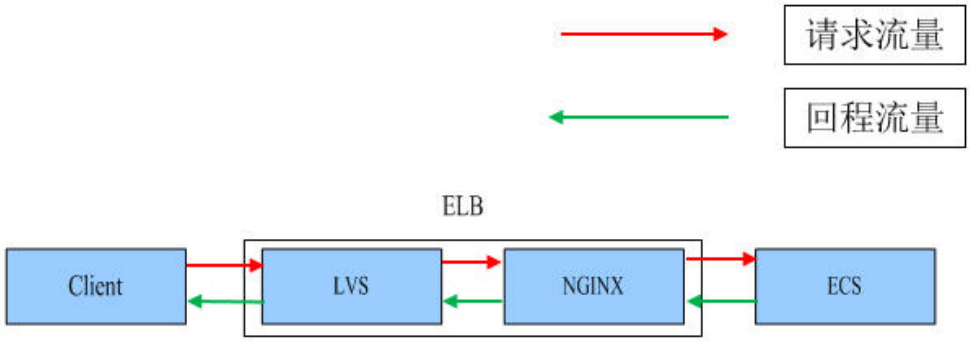


图 13-2 七层转发协议



13.3.2 如何跨 VPC 访问 ELB?

通过VPC对等连接，可以实现跨VPC访问ELB。例如：客户在VPC01中创建了ELB01，位于VPC02的客户端需要去访问ELB01。这时需要为VPC01和VPC02建立对等连接，并添加对等连接路由。

13.3.3 使用负载均衡器必须配置弹性公网带宽吗?

如果您使用的是私网负载均衡，则不需要配置弹性公网带宽；如果您使用的是公网负载均衡，则需要为负载均衡绑定的弹性公网IP配置公网带宽。

13.3.4 一个负载均衡器可以绑定多个 EIP 吗?

不可以。

- 如果您使用的是**公网负载均衡器**，则需要给负载均衡器绑定一个EIP，用来接收来自Internet公网的访问请求。
- 如果您使用的是**私网负载均衡器**，则需要给负载均衡器分配一个私网IP，仅能用来接收来自同一个VPC内的访问请求。如果需要接收来自不同VPC内的访问请求，则需要在ELB所在的VPC和其他VPC之间建立对等连接，打通VPC。建立对等连接请参见《虚拟私有云用户指南》中的“创建同一账户下的对等连接”。

13.3.5 创建/启用独享型负载均衡后为什么会占用多个子网 IP?

这些IP是供ELB内部使用的。

一般情况下，创建单可用区会占用2个IP，开启跨VPC后端会占用6个IP，如果是多可用区，占用的IP数会根据算法增加，目前没有具体的数值，实际使用IP的数量以您创建出来的独享型负载均衡占用的IP个数为准。

13.3.6 分配策略类型选择了“源 IP 算法”，但是同一个 IP 地址同时出现在了两台后端服务器上是什么原因?

选择“源IP算法”，ELB会将请求的源IP地址进行一致性Hash运算，同时对后端服务器进行编号，按照运算结果将请求分发到对应编号的服务器上。这可以使得对不同源IP的访问进行负载分发，同时使得同一个客户端IP的请求始终被派发至某台特定的服务器上。

但是如果后端服务器存在下线后重新上线的情况，ELB会将请求的源IP地址重新进行一致性Hash运算并对后端服务器重新进行编号，因此会出现同一个IP地址同时出现在了两台后端服务器上。

13.3.7 ELB 绑定了 EIP，后端的服务器可以通过 ELB 访问公网吗?

不可以。

弹性负载均衡是将访问流量根据分配策略分发到后端多台服务器的流量分发控制服务。ELB绑定EIP只能做负载，即外部访问后端服务。

ELB绑定EIP后，后端服务器可以绑定EIP，也可以不绑定EIP。如果后端服务器想要访问公网，要么直接绑定EIP，要么做NAT网关。

13.3.8 修改分配策略类型会导致业务中断吗?

修改分配策略类型不会影响现有连接，因此业务不会中断。

13.3.9 独享型负载均衡器的带宽和 EIP 的带宽有什么区别?

独享型负载均衡器的带宽，又称为“每秒带宽（Mbit/s）”，是指入流量或出流量不超过带宽规格的数值。ELB绑定的EIP的带宽是指客户端访问ELB时的最高流量限制。

13.4 监听器

13.4.1 监听器中分配算法和会话保持算法是什么关系?

会话保持功能，目的是将同一个用户的会话分发到相同的后端节点，负载均衡支持情况如**负载均衡会话保持支持情况**所示。

表 13-3 独享型负载均衡会话保持支持情况

分配策略	会话保持类型	L4（TCP、UDP）	L7（HTTP/HTTPS）
加权轮询算法	源IP地址	支持	不支持
	负载均衡器cookie	不涉及	支持
	应用程序cookie	不涉及	不支持

分配策略	会话保持类型	L4（TCP、UDP）	L7（HTTP/HTTPS）
加权最少连接	源IP地址	不支持支持	不支持
	负载均衡器cookie	不涉及	不支持支持
	应用程序cookie	不涉及	不支持支持
源IP地址	源IP地址	不涉及	不支持
	负载均衡器cookie	不涉及	不支持
	应用程序cookie	不涉及	不支持

一般建议：算法可以使用轮询算法，四层会话保持使用源IP地址，七层使用负载均衡器cookie方式。

13.4.2 弹性负载均衡如何支持多证书?

HTTPS监听器支持配置多个证书，您可以通过开启SNI配置单个HTTPS监听器绑定多个证书，实现同一个监听器根据多个域名选择证书来完成HTTPS认证和访问后端的诉求。

详细请参考**SNI证书-多域名访问**。

13.4.3 监听器删除之后，ELB 是否会立即停止转发业务流量?

- 当删除四层监听器时，由于客户端和ELB之间都是短连接，ELB会立即停止转发业务流量；
- 当删除七层监听器时，由于客户端和ELB之间保持长连接，客户端和ELB之间仍然会有部分TCP长连接存在，这些TCP长连接已经建立，不受监听器是否删除的影响，直到客户端在这些TCP连接上停止发送请求时间间隔达到keepalive_timeout超时时间（300s）之后，ELB才会断开这些长连接并停止转发业务流量。

  说明

  keepalive_timeout为空闲超时时间，只有客户端和ELB之间长连接时才会存在keepalive_timeout。

13.4.4 ELB 对上传文件的速度和大小是否有限制?

- ELB 七层监听器与四层监听器对客户端上传文件的速度都没有限制，可能EIP带宽限制会影响上传速度。
- 七层监听器的上传文件大小有限制，最大为10G；四层监听器的上传文件大小没有限制。

13.4.5 支持多个 ELB 转发到同一台后端服务器么?

支持，只要ELB和后端服务器在同一个子网下就可以。

13.4.6 如何启用 WebSocket 支持?

无需配置，当选用HTTP监听时，默认支持无加密版本WebSocket协议（WS协议）；当选择HTTPS监听时，默认支持加密版本的WebSocket协议（WSS协议）。

13.4.7 添加/修改监听器时，选择不到想选择的后端服务器组是什么原因?

这是因为后端服务器组的协议（后端协议）与监听器的协议（前端协议）存在对应关系，在给监听器添加后端服务器组时，只能添加与其协议对应的后端服务器组。如下所示：

表 13-4 独享型负载均衡-前端协议与后端协议对应情况

前端协议	后端协议
TCP	TCP
UDP	UDP/QUIC
HTTP	HTTP
HTTPS	HTTP/HTTPS

表 13-5 共享型负载均衡-前端协议与后端协议对应情况

前端协议	后端协议
TCP	TCP
UDP	UDP
HTTP	HTTP
HTTPS	HTTP

13.4.8 独享型负载均衡器为什么添加不了监听器?

这是因为您在创建独享型负载均衡时，只选择了网络型（TCP/UDP）实例规格或只选择了应用型（HTTP/HTTPS）实例规格，只能添加对应协议的监听器。

独享型ELB实例的类型选定后无法修改，请您合理评估选择。例如：您初始创建了网络型ELB实例，则只能创建TCP/UDP监听器，无法添加或修改为应用型ELB实例，也就无法添加HTTP/HTTPS监听器。

表 13-6 独享型负载均衡类型与监听器的关系

独享型负载均衡的类型	对应协议	可添加的监听器类型
网络型	TCP/UDP	TCP监听器、UDP监听器

独享型负载均衡的类型	对应协议	可添加的监听器类型
应用型	HTTP/HTTPS	HTTP监听器、HTTPS监听器

13.5 后端服务器

13.5.1 为什么后端服务器上收到的健康检查报文间隔和设置的间隔时间不一致?

ELB的每个lvs、nginx节点都会探测后端服务器，每个节点的间隔时间与设置的间隔时间保持一致。

后端服务器收到的是多个节点的探测报文，故在间隔时间内会收到多个检查报文。

13.5.2 使用 ELB 后，后端服务器能否访问公网?

后端服务器能否访问公网和ELB没有关系，如果后端服务器本身可以访问公网，使用了ELB以后仍可以访问，如果服务器本身不可以访问公网，使用ELB之后仍不可以。

13.5.3 ELB 可以跨区域关联后端服务器么?

- 独享型负载均衡支持跨VPC添加后端服务器，详见**跨VPC后端概述**。

13.5.4 公网负载均衡的后端服务器要不要绑定 EIP?

负载均衡实例都是通过私网转发访问请求，不需要后端服务器绑定EIP。

13.5.5 如何检查后端服务器网络状态?

1. 确认虚拟机主网卡已经正确分配到IP地址。
   a. 登录虚拟机内部。
   b. 执行**ifconfig**命令或**ip address**查看网卡的IP信息。

      说明

      Windows虚拟机可以在命令行中执行ipconfig查看。
2. 从虚拟机内部ping所在子网的网关，确认基本通信功能是否正常。
   a. 通常网关地址结尾为.1，可以在VPC详情页面中确认，切换“子网”页签，查看“网关”列，显示网关地址。
   b. 执行**ping**命令，观察能否**ping**通即可。若无法**ping**通网关则需首先排查二三层网络问题。

13.5.6 如何检查后端服务器网络配置?

1. 确认虚拟机使用的网卡安全组配置是否正确。
   a. 在弹性云服务器详情页面查看网卡使用的安全组。
   b. 检查安全组规则是否放通了对应的网段：

- 对于独享型负载均衡，检查后端服务器所在的安全组入方向是否放通ELB所在VPC的网段。如果没有放通，请在安全组入方向规则中添加ELB所在VPC网段。

⚠ 注意

2. 确认虚拟机使用网卡子网的网络ACL不会对流量进行拦截。
   在虚拟私有云页面左侧导航栏，单击“网络ACL”，确认涉及的子网已放通。

13.5.7 如何检查后端服务器服务状态?

1. 确认服务器服务是否开启。
   a. 登录虚拟机内部。
   b. 执行如下命令，查看系统的端口监听状态，如图13-3所示。
      netstat -ntpl

      说明

      Windows虚拟机可以在命令行中执行**netstat -ano**查看系统的端口监听状态，或者查看服务端软件状态。

      图 13-3 系统的端口监听状态

```
[root@ecs-67a0 ~]# netstat -ntpl
Active Internet connections (only servers)
Proto Recv-Q Send-Q Local Address           Foreign Address         State       PID/Program name
tcp        0      0 0.0.0.0:80              0.0.0.0:*               LISTEN      25847/./httpterm-s
tcp        0      0 0.0.0.0:22              0.0.0.0:*               LISTEN      1437/sshd
tcp6       0      0 :::22                   :::*                    LISTEN      1437/sshd
[root@ecs-67a0 ~]#
```

2. 从虚拟机测试服务通信功能是否正常。
   例如：该虚拟机的端口为http 80，使用curl命令，校验服务通信功能是否正常。

```
[root@ecs-67a0 ~]# curl 127.0.0.1:80 -v
* About to connect() to 127.0.0.1 port 80 (#0)
*   Trying 127.0.0.1...
* Connected to 127.0.0.1 (127.0.0.1) port 80 (#0)
> GET / HTTP/1.1
> User-Agent: curl/7.29.0
> Host: 127.0.0.1
> Accept: */*
>
< HTTP/1.1 200
< Connection: close
< Content-length: 14
< Cache-Control: no-cache
< X-req: size=14, time=500 ms
< X-rsp: id=test1, code=200, cache=0, size=14, time=500 ms
<
helloworld@!!
* Closing connection 0
[root@ecs-67a0 ~]#
```

13.5.8 后端服务器什么时候被认为是健康的?

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。

13.5.9 为什么配置了白名单后还能访问后端服务器?

白名单只对经过ELB实例的访问进行控制，通过控制访问负载均衡的监听器的IP地址，能够设置允许特定IP进行访问，而其它IP不许访问。如果需要对后端服务器进行访问控制，可以通过配置网络ACL或者安全组规则实现。

13.5.10 ELB 修改后端服务器权重后多久生效?

ELB修改后端服务器权重后，新的权重5秒内会生效。

- 对于TCP、UDP监听器，新的连接会根据修改后的权重转发，已经建立的连接不受影响。
- 对于HTTP、HTTPS监听器，新的请求会根据修改后的权重转发，已有请求不受影响。

说明

后端服务器的权重修改为0后，不会立即生效，仍然会有流量进入服务器，这是因为长连接在超时时间内会复用TCP连接，请求会继续转发。

- TCP和UDP监听器：长连接在空闲超时时间后断开。
- HTTP和HTTPS监听器：长连接在响应超时时间后断开。

13.5.11 为什么开启跨 VPC 后端需要确保负载均衡所属子网至少拥有 16 个可用 IP 地址?

这些IP是供ELB内部使用的。一般情况下，创建ELB，单可用区会占用2个IP，开启跨VPC后端会占用6个IP，如果是多可用区，占用的IP数会根据算法增加，目前没有具体的数值，实际使用IP的数量以您创建出来的ELB占用的IP个数为准。

13.6 健康检查

13.6.1 健康检查异常如何排查?

问题描述

客户端通过负载均衡器访问后端服务器异常，负载均衡器的“后端服务器组”页签显示后端服务器的健康检查结果为“异常”。

在“负载均衡器”界面，单击后端服务器所在的负载均衡器名称，切换到“后端服务器组”页签，在基本信息页面，查看“健康检查结果”列是否显示“异常”。

背景介绍

ELB的健康检查通过向后端服务器发起心跳检查的方式来实现，在检查过程中使用ELB后端子网所在的VPC地址通信。为确保健康检查正常进行，您需要确保服务器已经放通ELB后端子网所在的VPC网段。

⚠ 注意

- 在给负载均衡添加后端服务器之前首先要检查其所在安全组规则是否放行源网段为后端子网所在的VPC网段地址。
- 四层监听器未开启“跨VPC后端”功能时，后端服务器安全组规则无需放通ELB后端子网所在的VPC网段。

当健康检查探测到您的后端服务器异常时，ELB将不再向异常的后端服务器转发流量。直到健康检查检测到后端服务器恢复正常时，ELB才会向此服务器继续转发流量。

排查思路

以下排查思路根据原因的出现概率进行排序，建议您从高频率原因往低频率原因排查，从而帮助您快速找到问题的原因。

如果解决完某个可能原因仍未解决问题，请继续排查其他可能原因。

说明

相关修改配置的操作，修改完配置后需要等待一定的时间，配置才会生效，因为健康检查包含检查间隔和阈值（根据默认配置为几十秒生效，如果健康检查恢复正常，在ELB关联的后端服务器基本信息界面可以看到健康检查状态是否正常）。

图 13-4 健康检查异常排查

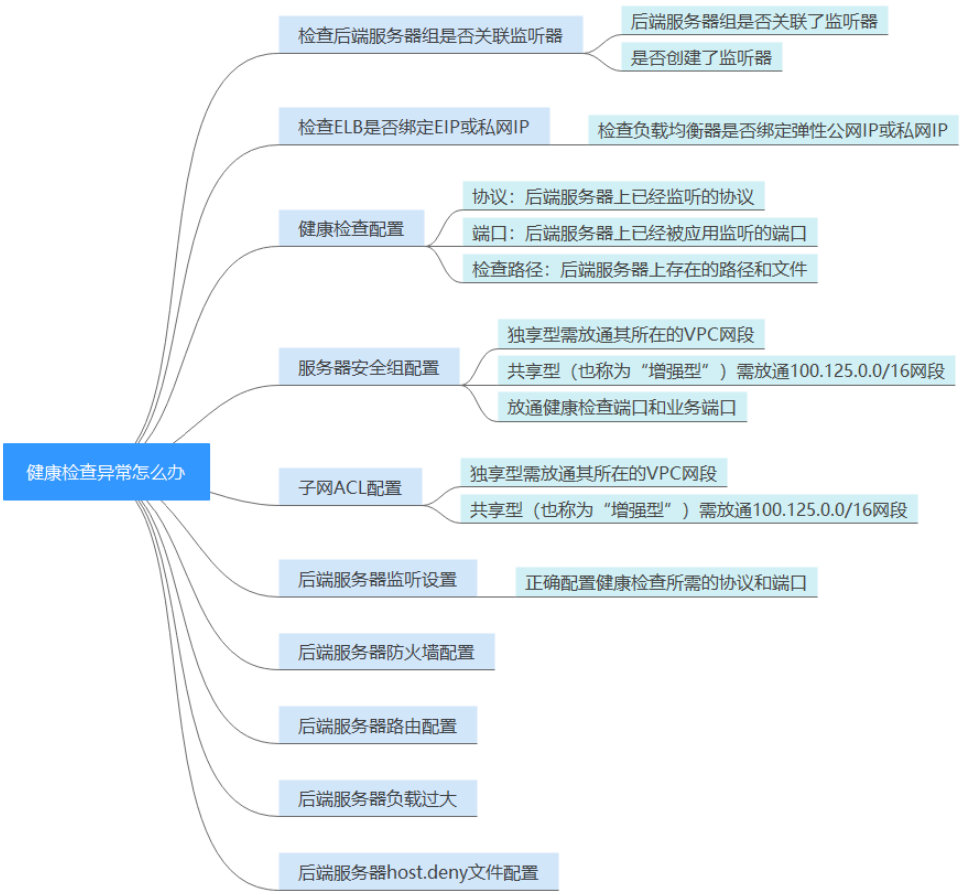


图 13-5 健康检查异常如何排查

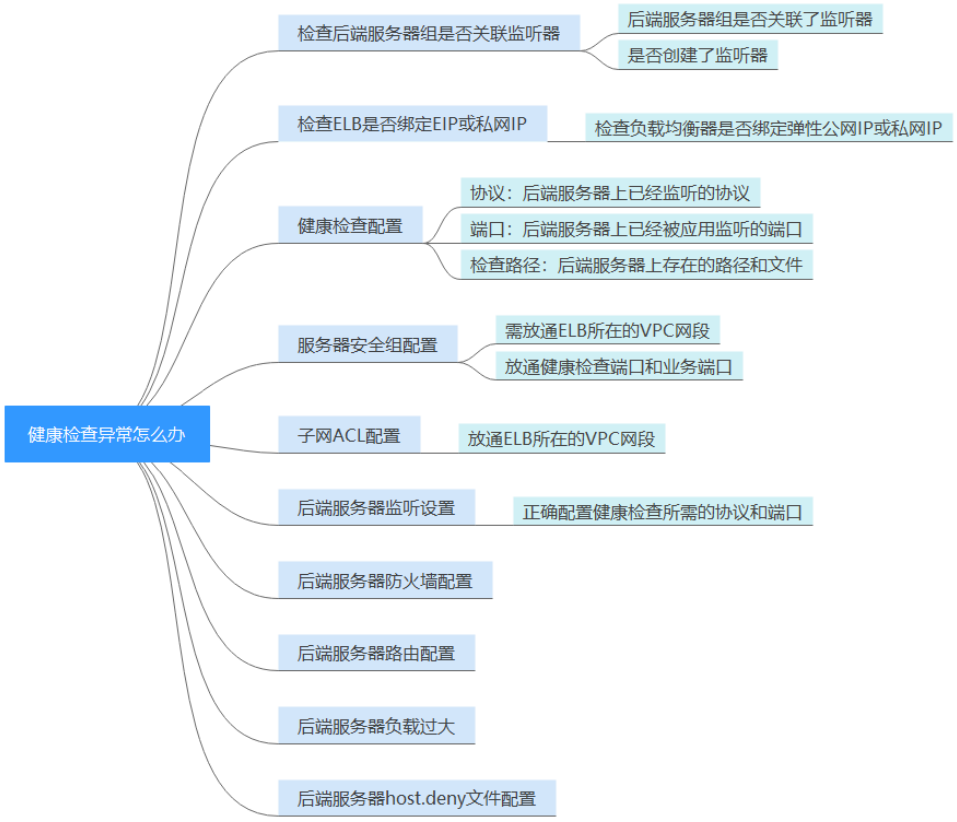


表 13-7 排查思路

可能原因	处理措施
检查后端服务器组是否关联监听器	解决方法请参考**检查后端服务器组是否关联监听器**。
检查ELB是否绑定EIP或私网IP	解决方法请参考**检查ELB是否绑定EIP或私网IP**。
健康检查配置	解决方法请参考**检查健康检查配置**。
服务器安全组配置	解决方法请参考**检查服务器所在安全组**。
子网ACL配置	解决方法请参考**检查网络ACL规则**。
后端服务器监听配置	解决方法请参考**检查后端服务器是否正常**。
后端服务器防火墙配置	解决方法请参考**检查服务器防火墙**。
后端服务器路由配置	解决方法请参考**检查服务器路由**。
后端服务器负载过大	解决方法请参考**检查服务器负载**。
后端服务器host.deny文件配置	解决方法请参考**检查服务器hosts.deny文件**。

检查后端服务器组是否关联监听器

检查健康检查异常的服务器所在的后端服务器组是否关联了监听器。

- 如果后端服务器组未关联监听器，请检查是否已创建了监听器。
  - 已经创建了监听器，请将后端服务器组关联至监听器。
  - 未创建监听器，请先添加监听器，然后为监听器绑定后端服务器组。
- 如果后端服务器组已经关联了监听器，请再按照以下原因排查。

检查 ELB 是否绑定 EIP 或私网 IP

说明

- 该检查项仅适用于四层监听器（TCP/UDP）。
- 对于七层监听器（HTTP/HTTPS），无论ELB是否绑定弹性公网IP或私网IP，均不会影响后端服务器健康检查。

对于四层监听器（TCP/UDP），请检查其关联的负载均衡器是否绑定弹性公网IP或私网IP。

如果ELB未绑定弹性公网IP和私网IP，请绑定弹性公网IP或私网IP。

说明

ELB初次创建时，如果未绑定EIP或私网IP时，四层监听器（TCP/UDP）所关联的后端服务器会显示健康检查异常。当给ELB绑定EIP或私网IP后，健康检查结果显示正常，再解绑EIP或私网IP后，健康检查结果依然会显示正常。

检查健康检查配置

单击对应的负载均衡名称，进入负载均衡基本信息页面。切换到“后端服务器组”页签，单击对应的后端服务器组名称，在其基本信息页面，单击“健康检查”右侧的配置按钮。查看以下参数：

- 域名。健康检查使用HTTP协议时，如果后端服务器设置了校验HOST头能力，需要将后端服务器配置的域名填写到“健康检查配置”页面中的“域名”处。
- 协议。
- 端口。端口必须是后端服务器上真实业务所监听的端口，不是自定义端口。检查您配置的健康检查端口和监听的端口是否一致。不一致则会导致健康检查异常。
- 检查路径。如果是使用HTTP健康检查需要查看此参数，建议配置简单的静态HTML文件。

说明

- 健康检查协议为“HTTP”，则会检查端口和路径。
- 健康检查协议为“TCP”，则只检查端口。
- 您的健康检查协议为“HTTP”，健康检查异常时，如果您已检查端口没有问题，请修改检查路径或者将健康检查协议修改为“TCP”，只检查端口。
- 检查路径需填写绝对路径。

  例如：

  访问链接为：http://www.example.com或http://192.168.63.187:9096，则检查路径填写“/”。

  访问链接为：http://www.example.com/chat/try/，则检查路径填写“/chat/try/”。

  访问链接为：http://192.168.63.187:9096/chat/index.html，则检查路径填写“/chat/index.html”。

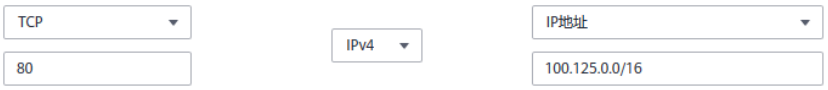
检查服务器所在安全组

- **TCP、HTTP或HTTPS协议监听器**：后端服务器所在的安全组入方向规则无需放通100.125.0.0/16网段，但需放通ELB所在VPC的网段，并在TCP协议中放通健康检查的端口。
  - **健康检查端口与后端服务器业务端口相同**：需要放通后端服务器的业务端口，例如80。
  - **健康检查端口与后端服务器业务端口不同**：需要放通后端服务器的业务端口和健康检查端口，例如80和443。

    说明

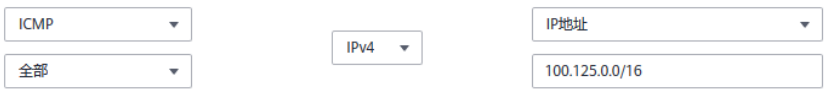
    健康检查的协议和端口在配置的健康检查配置项提示框中获取。

  图 13-6 安全组入方向规则配置示例

- **UDP协议监听器**：不仅需要保证安全组入方向规则放通健康检查的协议、端口和ELB所在VPC的网段。还需要放通后端服务器所在安全组入方向的ICMP协议。

  图 13-7 安全组入方向规则放通 ICMP 协议示例

说明

- 如果不确认是否是安全组问题，可以把安全组入方向规则的“协议”和“端口范围/ICMP类型”均放通Any测试下。
- UDP协议监听器，也可以参考**使用UDP协议有什么注意事项?** 。

检查网络 ACL 规则

网络ACL是一个子网级别的可选安全层，通过与子网关联的出方向/入方向规则控制出入子网的数据流。网络ACL与安全组类似，都是安全防护策略，当您想增加额外的安全防护层时，就可以启用网络ACL。但是网络ACL默认规则会拒绝所有入站和出站流量，

如果此网络ACL和负载均衡所属同一个子网，或者此网络ACL和负载均衡相关联的后端服务器所属同一个子网那么负载均衡的业务也会受到影响，收不到来自于公网或者私网的任何请求流量，或者会导致后端服务器异常。

您可以通过配置网络ACL入方向规则，放行源网段为ELB所在的VPC网段，目的端口为后端服务器端口。

检查后端服务器是否正常

说明

如果后端服务器的操作系统为Windows，请通过浏览器直接访问*https://后端服务器的IP*:*健康检查配置的端口*。如果返回码为2xx或3xx，则表示后端服务器正常。

- 您可以在后端服务器上通过以下命令查看后端服务器的健康检查端口是否被健康检查协议正常监听。

```
netstat -anlp | grep port
```

回显中包含健康检查端口信息并且显示LISTEN，则表示后端服务器的健康检查端口在监听状态，如图13-8中表示880端口被TCP进程所监控。

如果您没有配置健康检查端口信息，默认和后端服务器业务端口一致。

图 13-8 后端服务器正常被监听的回显示例

图 13-9 后端服务器没有被监听的回显示例

如果健康检查端口没有在监听状态（后端服务器没有被监听），您需要先启动后端服务器上的业务，启动业务后再查看健康检查端口是否被正常监听。

- 如果是HTTP健康检查，请您在后端服务器上执行以下命令查看回显中返回的状态码。

```
curl 后端服务器的私有IP:健康检查端口/健康检查路径 -iv
```

HTTP健康检查是ELB向后端服务器发起GET请求，当获取到以下所列的响应状态码，认为服务器是正常状态。

对于TCP的监听器，HTTP健康检查正常返回状态码是200。

对于ELB，HTTP健康检查正常返回状态码是200、202或者401。

图 13-10 后端服务器异常的回显示例

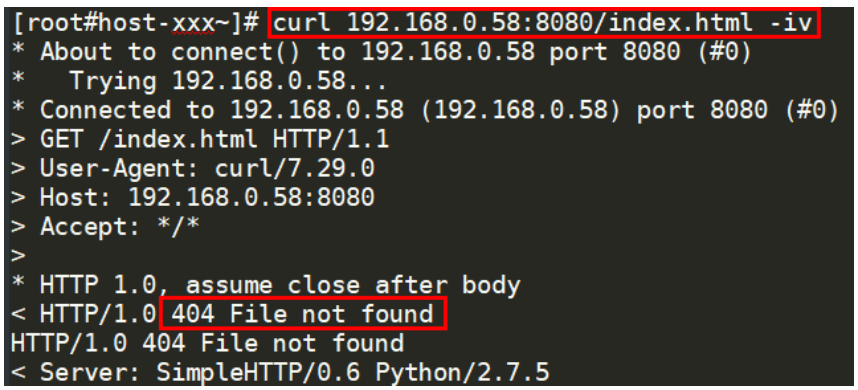


图 13-11 后端服务器正常的回显示例

```
[root#host-xxx~]# curl 192.168.0.58:8080/index.html -iv
* About to connect() to 192.168.0.58 port 8080 (#0)
*   Trying 192.168.0.58...
* Connected to 192.168.0.58 (192.168.0.58) port 8080 (#0)
> GET /index.html HTTP/1.1
> User-Agent: curl/7.29.0
> Host: 192.168.0.58:8080
> Accept: */*
>
192.168.0.58 . . [08/Apr/2019 17:37:34] *GET /index.html HTTP/1.1* 200
* HTTP 1.0, assume close after body
< HTTP/1.0 200 OK
HTTP/1.0 200 OK
< Server: SimpleHTTP/0.6 Python/2.7.5
```

- 如果HTTP健康检查异常，除了检查健康检查路径外，建议您**将配置的HTTP健康检查修改为TCP健康检查**。操作如下：

  在监听器界面，修改目标监听器，在配置参数里选择已有TCP健康检查的后端服务器组，或者选择新创建TCP健康检查的后端服务器组。配置完成之后，几十秒后去查看健康检查状态是否恢复正常。

检查服务器防火墙

如果后端服务器内部开启了防火墙或其他安全类防护软件，这些软件可能会屏蔽ELB后端子网所在的VPC网段或100.125.0.0/16网段的IP。

检查服务器路由

请检查是否手动修改了后端服务器内部的路由，查看主网卡（比如eth0）上是否配置默认路由，默认路由是否修改。如果默认路由更改，可能导致健康检查报文无法到达后端服务器。

您可以在后端服务器上通过以下命令查看您的默认路由是否指向网关（经过ELB转发属于跨网段访问，三层通信需要配置默认路由指向网关）。

```
ip route
```

或

```
route -n
```

正常的回显如图13-12所示。

图 13-12 默认路由指向网关示例

```
root@donatdel.wangfei.iperf ~]# ip route
default via 192.168.2.1 dev eth0 proto dhcp metric 100
169.254.169.254 via 192.168.2.1 dev eth0 proto dhcp metric 100
192.168.2.0/24 dev eth0 proto kernel scope link src 192.168.2.124 metric 100
root@donatdel.wangfei.iperf ~]#
```

图 13-13 默认路由未指向网关示例

```
[root@test ~]# ip route
default via 192.168.0.134 dev eth0
169.254.0.0/16 dev eth0  scope link  metric 1002
169.254.169.254 via 192.168.0.1 dev eth0  proto static
192.168.0.0/24 dev eth0  proto kernel  scope link  src 192.168.0.242
```

如果回显中没有像图13-12中的第一条路由信息，或者路由指向的IP的不是后端服务器所在VPC子网的网关，请您配置默认路由指向网关。

检查服务器负载

通过云监控服务，查看后端服务器的CPU/内存/网络连接数等，来判断后端服务器的负载是否过高。

如果负载很高，可能会导致健康检查的连接或请求超时。

检查服务器 hosts.deny 文件

建议您排查后端服务器的/etc/hosts.deny文件，文件中不能写入ELB后端子网所在的VPC网段。

13.6.2 为什么后端服务器上收到的健康检查报文间隔和设置的间隔时间不一致?

ELB的每个lvs、nginx节点都会探测后端服务器，每个节点的间隔时间与设置的间隔时间保持一致。

后端服务器收到的是多个节点的探测报文，故在间隔时间内会收到多个检查报文。

13.6.3 使用 UDP 协议有什么注意事项?

什么是 UDP 健康检查

UDP是面向非连接的一种协议，在发送数据前不会通过进行三次握手建立连接，UDP健康检查的实现过程如下：

- 健康检查的节点根据健康检查配置，向后端发送ICMP request 消息。
  - 如果健康检查节点收到了后端服务器返回的ICMP reply消息，则认为服务正常，继续进行健康检查。
  - 如果健康检查节点没有收到后端服务器返回的ICMP reply消息，则认为服务异常，判定健康检查失败。
- 健康检查的节点收到ICMP reply消息后，会给后端服务器发送UDP探测报文。
  - 如果在【超时时间】之内，健康检查的节点服务器收到了后端服务器返回的port unreachable的ICMP消息，则认为服务异常，判定健康检查失败。
  - 如果在【超时时间】之内，健康检查的节点服务器没有收到后端服务器返回的ICMP错误信息，则认为服务正常，判定健康检查成功。

当您配置UDP健康检查时，推荐使用配置页面默认的各项数值。

异常排查方法

请您按照以下两种方法排查。

- 检查健康检查超时时间是否过小。

  可能的原因：后端服务器回复的reply或port unreachable类型的ICMP消息未能在超时时间内到达健康检查的节点，导致健康检查结果不准确。

  建议采取的措施：将超时时间调整为更大的值。

  由于UDP健康检查的原理不同于其他健康检查，建议健康检查超时时间不要过小，否则后端服务器可能会反复上线或下线。
- 后端服务器是否限制了ICMP消息产生的速率。

Linux系统下，请用以下命令检查ICMP消息速率的限制。

```
sysctl -q net.ipv4.icmp_ratelimit
```

默认值为：1000

```
sysctl -q net.ipv4.icmp_ratemask
```

默认值为：6168

请确认第一条命令返回值为默认值或0，并用以下命令放开port unreachable消息产生的速率限制。

```
sysctl -w net.ipv4.icmp_ratemask=6160
```

更详细的信息请参考Linux Programmer's Manual相关页面：

```
man 7 icmp
```

或者访问地址：http://man7.org/linux/man-pages/man7/icmp.7.html

说明

放开port unreachable类型ICMP消息的速率限制，会让暴露在公网上的服务器在端口扫描时，不受限制次数地产生port unreachable消息。

注意事项

使用UDP协议注意以下事项：

- 负载均衡健康检查是通过UDP报文和Ping报文探测来获取后端服务器的状态信息。针对此种情况，用户需要确保后端服务器开启ICMP协议，确认方法如下：

  用户登录后端服务器，以root权限执行以下命令：

  cat /proc/sys/net/ipv4/icmp_echo_ignore_all

  若返回值为1，表示ICMP协议关闭；若为0，则表示开启。
- 当前UDP协议服务健康检查可能存在服务真实状态与健康检查不一致的问题：

  如果后端服务器是Linux服务器，在大并发场景下，由于Linux的防ICMP攻击保护机制，会限制服务器发送ICMP的速度。此时，即便服务器已经出现异常，但由于无法向前端返回“port XX unreachable”报错信息，会导致负载均衡由于没收到ICMP 应答进而判定健康检查成功，最终导致服务真实状态与健康检查不一致。

13.6.4 健康检查为什么会导致 ELB 会频繁向后端服务器发送探测请求?

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服务器发送探测请求，检查间隔用户可配，健康检查会根据检查间隔一直探测，所以每隔几秒会有访问。您可以通过**修改健康检查配置**的周期来控制访问后端服务器的频率。

13.6.5 健康检查什么时候启动?

后端服务器新加入后，在第一个周期内随机一个时间开始检测，后续按照“检查间隔”启动。

13.6.6 “最大重试次数”是否包括健康检查失败的场景?

是，最大重试次数既是健康最大重试次数，也是不健康最大重试次数。

13.6.7 如何处理健康检查导致的大量日志?

1. 可以增加健康检查间隔时间，配置方法详见**修改健康检查配置**。

   存在的风险：延长健康检查的间隔时间后，后端ECS实例出现故障时，负载均衡发现故障ECS实例的时间也会增长。

2. 可以关闭健康检查，配置方法详见**修改健康检查配置**。

   存在的风险：关闭健康检查后，负载均衡不再检查后端服务器，一旦某台后端服务器发生故障，则无法实现访问流量自动切换至其它正常的后端服务器。

13.6.8 健康检查正常默认返回的状态码有哪些?

表 13-8 健康检查正常返回的状态码

ELB类型	健康检查协议	健康检查正常返回的状态码
负载均衡	HTTP	200
	HTTPS	200

13.7 获取源 IP

13.7.1 如何获取来访者的真实 IP?

当客户端通过ELB访问后端服务器时，客户端真实的IP地址会被ELB转换，后端服务器获取到的往往是ELB转换后的客户端IP地址。如果需要获取到客户端的真实IP，可以按如下方法操作。

- 七层服务（HTTP/HTTPS协议）：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。

  配置详情见**七层服务**。

约束与限制

- 如果IP经过NAT，则只能获取到NAT转化后的IP地址，无法获取到NAT转化前的IP地址。
- 如果客户端为容器，只能获取到容器所在主机的IP地址，无法获取容器的IP。
- 四层监听器（TCP/UDP）开启“获取客户端IP”功能之后，不支持同一台服务器既作为后端服务器又作为客户端的场景。
- 独享型负载均衡的四层监听器（TCP/UDP）默认开启源地址透传功能，无需手动开启，且不支持关闭。

七层服务

针对七层服务（HTTP/HTTPS协议），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：

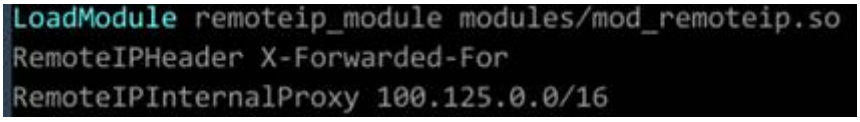
```
X-Forwarded-For: 来访者真实IP, 代理服务器1-IP，代理服务器2-IP, ...
```

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。

配置Apache服务器

1. 安装Apache 2.4。
   例如在CentOS 7.5环境下，可以执行如下命令执行安装：
   ```
   yum install httpd
   ```
2. 修改Apache的配置文件/etc/httpd/conf/httpd.conf，在最末尾添加以下配置信息。
   ```
   LoadModule remoteip_module modules/mod_remoteip.so
   RemoteIPHeader X-Forwarded-For
   RemoteIPInternalProxy 100.125.0.0/16
   ```

   图 13-14 修改 Apache 的配置文件示例图

   说明

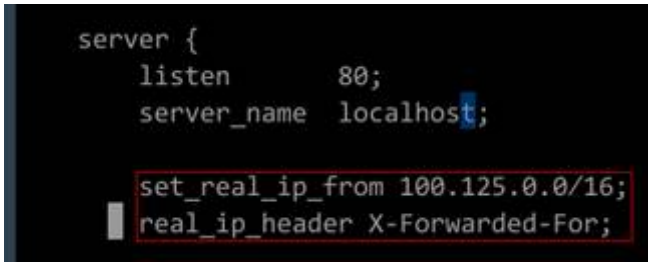
   将代理服务器的网段添加到 RemoteIPInternalProxy <IP_address>。
   负载均衡需要添加ELB实例关联的VPC子网网段。
3. 修改Apache的配置文件/etc/httpd/conf/httpd.conf，将日志输出格式修改为如下所示（%a代表源IP地址）：
   ```
   LogFormat "%a %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\"" combined
   ```
4. 重启Apache。
   ```
   systemctl restart httpd
   ```
5. 查看httpd的访问日志，您可以获取真实的来访者IP。

配置Nginx服务器

例如在CentOS 7.5环境下，可以执行如下命令执行安装：

1. 运行以下命令安装http_realip_module。
   ```
   yum -y install gcc pcre pcre-devel zlib zlib-devel openssl openssl-devel
   wget http://nginx.org/download/nginx-1.17.0.tar.gz
   tar zxvf nginx-1.17.0.tar.gz
   cd nginx-1.17.0
   ./configure --prefix=/path/server/nginx --with-http_stub_status_module --without-http-cache --with-http_ssl_module --with-http_realip_module
   make
   make install
   ```
2. 打开nginx.conf文件。
   ```
   vi /path/server/nginx/conf/nginx.conf
   ```
3. 在以下配置信息后添加新的配置字段和信息。
   在http或者server处，需要添加的配置字段和信息：
   ```
   set_real_ip_from 100.125.0.0/16;
   real_ip_header X-Forwarded-For;
   ```


图 13-15 添加配置字段和信息示例图



说明

将代理服务器的网段添加到 RemoteIPInternalProxy <IP_address>。

负载均衡需要添加ELB实例关联的VPC子网网段。

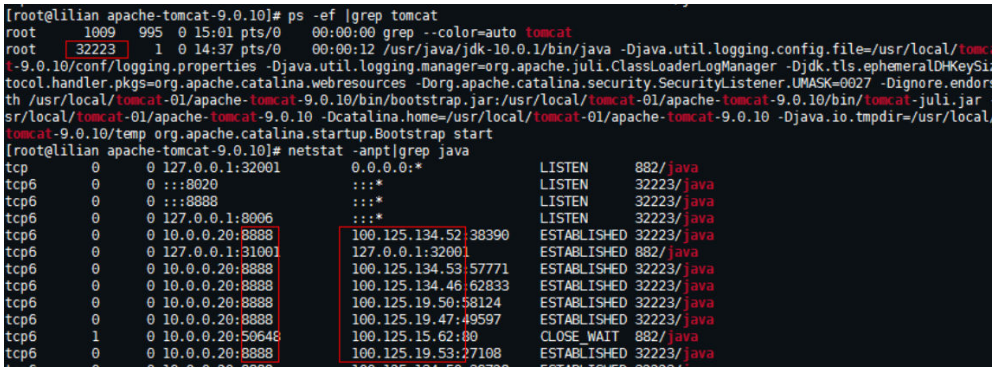
4. 启动Nginx。
   ```
   /path/server/nginx/sbin/nginx
   ```
5. 查看Nginx的访问日志，您可以获取真实的来访者IP。
   ```
   cat /path/server/nginx/logs/access.log
   ```

配置Tomcat服务器

本教程中的Tomcat的安装路径为“/usr/tomcat/tomcat8/”。

1. 登录已安装Tomcat的服务器。
2. 执行如下命令，确定Tomcat已经正常运行。
   ```
   ps -ef|grep tomcat
   netstat -anpt|grep java
   ```

图 13-16 正常运行结果示例



3. 将server.xml文件中的className="org.apache.catalina.valves.AccessLogValve" 模块修改为如下内容。
   ```
   vim /usr/tomcat/tomcat8/conf/server.xml
   <Valve className="org.apache.catalina.valves.AccessLogValve" directory="logs"
   prefix="localhost_access_log." suffix=".txt"
   pattern="%{X-FORWARDED-FOR}i %l %u %t %r %s %b %D %q %{User-Agent}i %T"
   resolveHosts="false" />
   ```


图 13-17 配置示例

```
        <!-- Access log processes all example.
             Documentation at: /docs/config/valve.html
             Note: The pattern used is equivalent to using pattern="common" -->
        <Valve className="org.apache.catalina.valves.AccessLogValve" directory="logs"
               prefix="localhost_access_log." suffix=".txt"
               pattern="%{X-FORWARDED-FOR}i %l %u %t %r %s %b %D %q %{User-Agent}i %T" resolveHosts="false" />

      </Host>
    </Engine>
```

4. 执行如下命令，重启Tomcat服务。

```
cd /usr/tomcat/tomcat8/bin && sh shutdown.sh && sh startup.sh
```

其中“/usr/tomcat/tomcat8/”为Tomcat安装路径，请根据实际情况替换。

图 13-18 重启 Tomcat 服务

```
[root@ecs-ddef bin]# sh startup.sh
Using CATALINA_BASE:   /usr/tomcat/tomcat8
Using CATALINA_HOME:   /usr/tomcat/tomcat8
Using CATALINA_TMPDIR: /usr/tomcat/tomcat8/temp
Using JRE_HOME:        /usr/java/jdk1.8.0_261
Using CLASSPATH:       /usr/tomcat/tomcat8/bin/bootst
Tomcat started.
```

5. 执行如下命令，查看最新的日志。

如图中红框所示获取到的非100.125网段的IP地址，即为获取到的源IP地址。

```
cd /usr/tomcat/tomcat8/logs/
cat localhost_access_log..2021-11-29.txt
```

其中“localhost_access_log..2021-11-29.txt”为当天日志路径，请根据实际情况替换。

图 13-19 查询源 IP 地址

```
100.125.68.197 - - [29/Nov/2021:14:33:27 +0800] "GET /bg-upper.png HTTP/1.1" 200 3103
100.125.68.197 - - [29/Nov/2021:14:33:27 +0800] "GET /bg-middle.png HTTP/1.1" 200 1918
100.125.68.197 - - [29/Nov/2021:14:33:27 +0800] "GET /bg-button.png HTTP/1.1" 200 713
100.125.68.197 - - [29/Nov/2021:14:33:27 +0800] "GET /favicon.ico HTTP/1.1" 200 21630
100.125.68.197 - - [29/Nov/2021:14:33:38 +0800] "GET / HTTP/1.1" 200 11250
100.125.68.197 - - [29/Nov/2021:14:35:09 +0800] "GET / HTTP/1.1" 200 11250
[[redacted]logs]# cat localhost_access_log..2021-11-29.txt
124.7[redacted]6 - - [29/Nov/2021:14:41:09 +0800] GET / HTTP/1.1 200 11250 178  Mozilla/5.0
0.178
124.7[redacted] - - [29/Nov/2021:14:41:47 +0800] GET / HTTP/1.1 200 11250 3  Mozilla/5.0 (
003
124.7[redacted] - - [29/Nov/2021:14:42:10 +0800] GET / HTTP/1.1 200 11250 3  Mozilla/5.0 (
003
```

配置Windows IIS服务器

本教程以Windows Server 2012配置IIS7为例介绍，其他版本操作可能略有不同。

1. 下载并安装IIS。
2. 从第三方网站下载F5XForwardedFor.dll插件，并获取x86和x64目录下的F5XForwardedFor.dll插件拷贝到IIS服务具有访问权限的目录下，例如C:\F5XForwardedFor2008。
3. 打开IIS管理器，选择“模块 > 配置本机模块”注册拷贝的2个插件。

图 13-20 选择模块选项

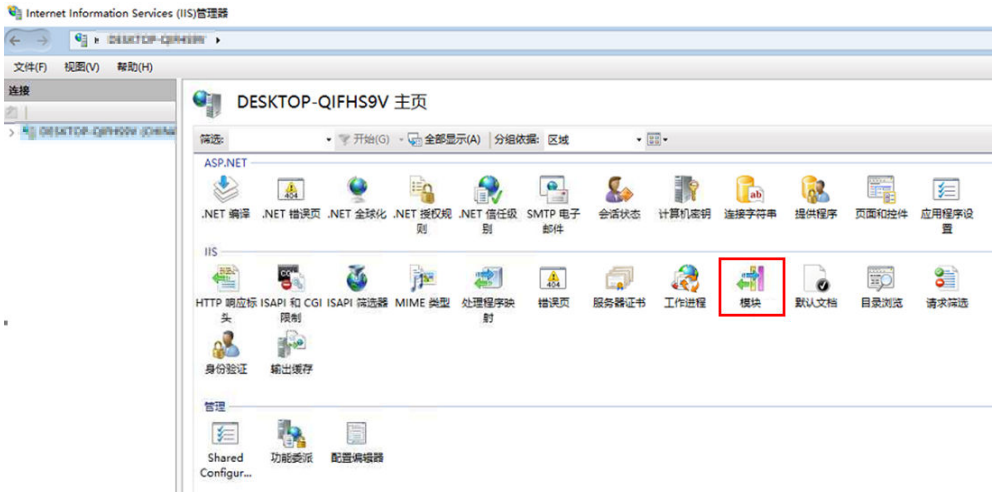
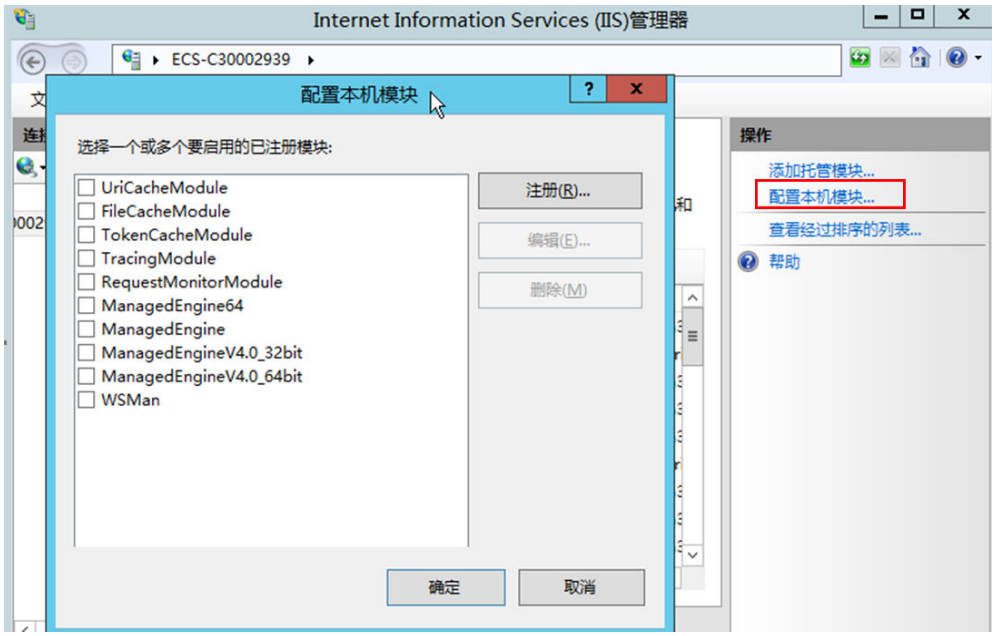
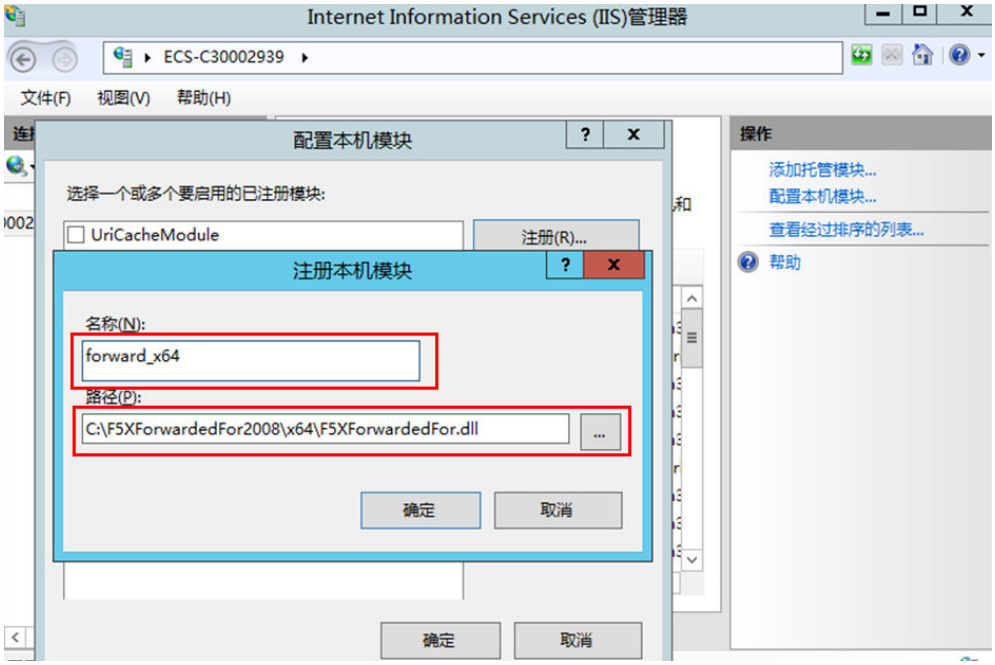


图 13-21 配置本机模块



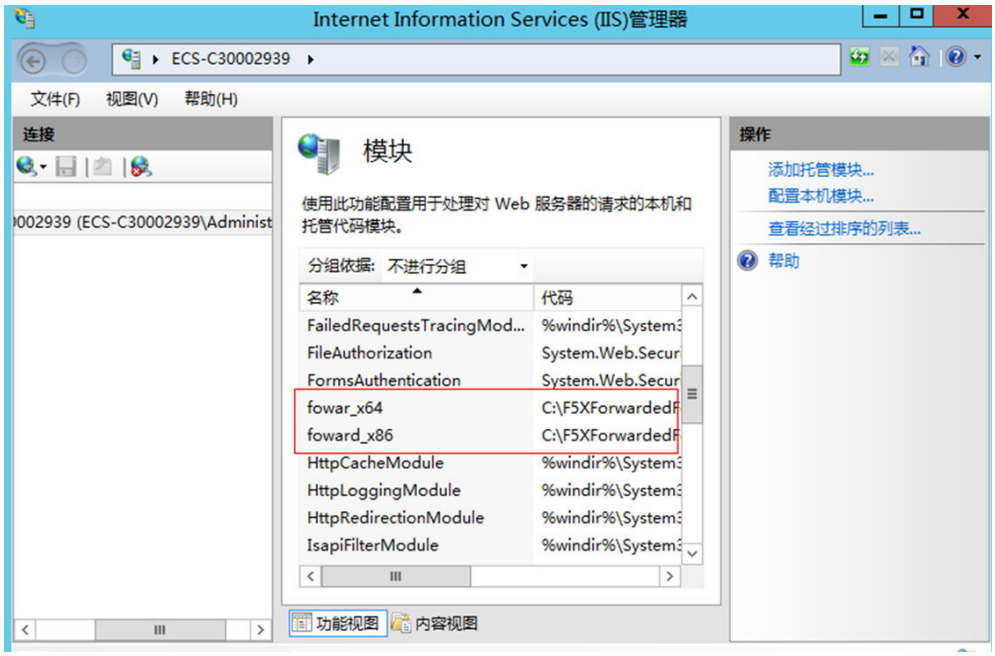
4. 单击“注册”，分别注册x86和x64插件。

图 13-22 注册插件



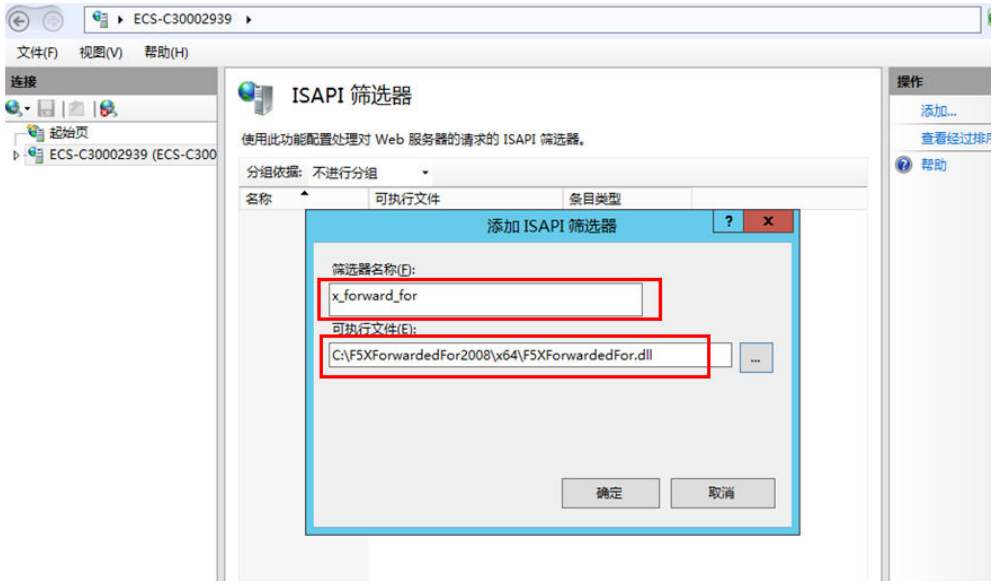
5. 在“模块”页面，确认注册的模块名称出现在列表中。

图 13-23 确认注册成功



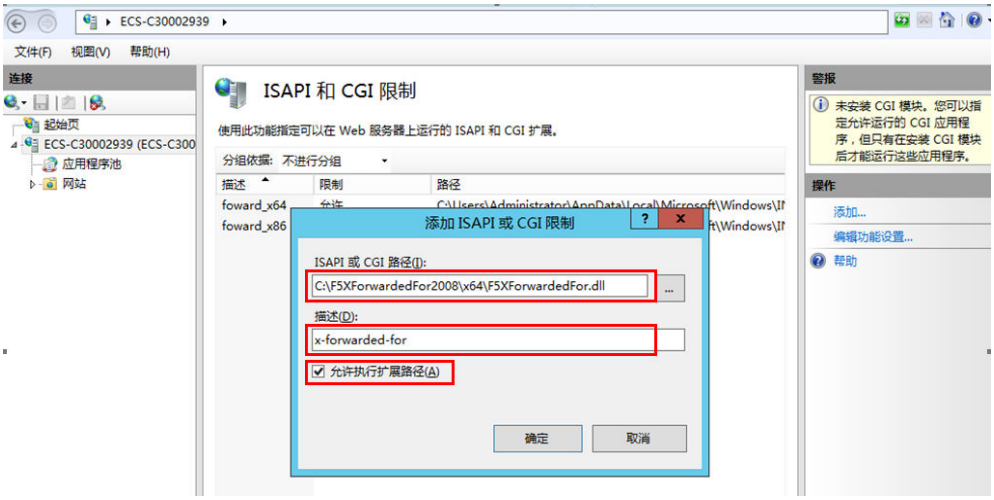
6. 选择IIS管理器主页的“ISAPI筛选器”，为2个插件授权运行ISAPI和CGI扩展。

图 13-24 添加授权



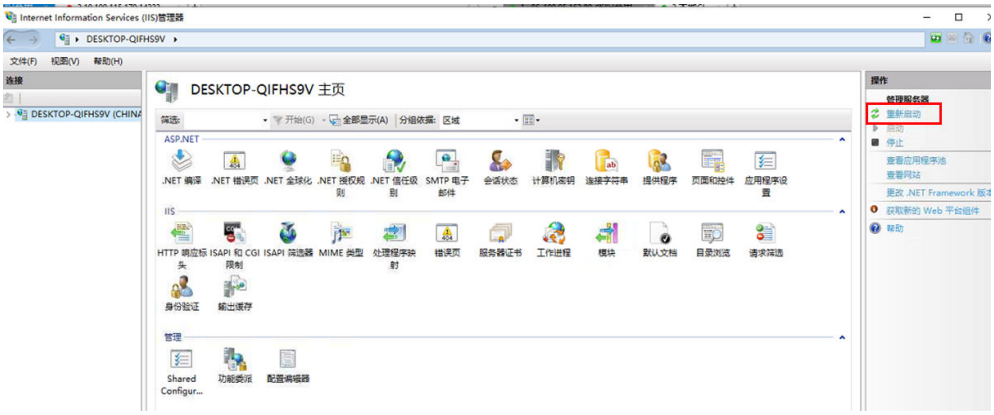
7. 选择“ISAPI和CGI限制”，为2个插件设置执行权限。

图 13-25 允许执行



8. 单击主页的“重新启动”，重启IIS服务，重启后配置生效。

图 13-26 重启 IIS 服务



13.8 HTTP/HTTPS 监听器

13.8.1 HTTPS 监听器的后端协议该如何选择?

如果您想实现全链路HTTPS访问，可以使用独享型负载均衡器，添加HTTPS监听器，后端协议选择HTTPS协议。

如果是非全链路HTTPS，负载均衡支持后端协议选择HTTP协议。

说明

全链路HTTPS仅支持在负载均衡器上做双向验证。

13.8.2 为什么配置证书后仍出现不安全提示?

可能由于以下原因导致配置证书后仍出现不安全提示。

- 证书所记录的域名与用户访问的域名不一致，建议排查证书所记录的域名，或创建自签名证书。
- 配置了SNI，输入的域名与证书所记录的域名不一致。
- 域名级别与证书级别不一致，例如域名为5级而证书为4级。

其他情况您也可以使用 *curl 访问的域名*命令，根据系统返回的错误信息进行排查。

13.8.3 转发策略的状态显示为“故障”的原因是什么?

可能的原因是：如果创建了相同的转发策略（出现转发策略冲突），则会出现转发策略故障，此时即使把前面创建的转发策略删除，后面的转发策略依然会显示故障。

解决办法：将出现冲突的转发策略全部都删除后再重新添加，即可恢复正常。

13.8.4 为什么找不到配置转发策略的入口?

请检查您的监听器协议类型是否为TCP或UDP。

目前转发策略仅支持协议类型为HTTP、HTTPS的监听器，不支持协议类型为TCP、UDP的监听器。

13.8.5 配置转发策略时，为什么无法选择已有的后端服务器组?

后端服务器组只能被一个转发策略所引用，因为该后端服务器组已经被另一个转发策略所引用，所以选不到想选择的后端服务器组。

13.9 会话保持

13.9.1 长连接和会话保持区别是什么?

长连接和会话保持没有必然联系。

长连接是指在一个连接上可以连续发送多个数据包，在连接保持期间，如果没有数据包发送，需要双方发链路检测包。会话保持是指弹性负载均衡将属于同一个会话的请求都转发到同一个服务器进行处理。

13.9.2 如何检查弹性负载均衡会话保持不生效问题?

1. 查看后端服务器组上是否开启了会话保持。
2. 查看后端云服务器的健康检查状态是否正常，如果异常，流量会切换到其他后端云服务器，导致会话保持失效。
3. 如果选择的是源IP算法，需要注意请求到达弹性负载均衡之前IP是否发生变化。
4. 如果是HTTP或HTTPS监听器，配置了会话保持，不用观察session是否丢失，而需要注意发送的请求是否带有cookie，如果带有cookie，则观察该cookie值是否发生了变化（因为7层会话保持基于cookie）。

13.9.3 如何使用 Linux curl 测试负载均衡会话保持?

1. 申请ELB与ECS资源。
   a. 创建3个ECS实例，1个做客户端，2个做服务端。
   b. 创建1个ELB实例与HTTP监听器实例，注意务必开启“会话保持”功能。
2. 启动服务端ECS的HTTP服务。

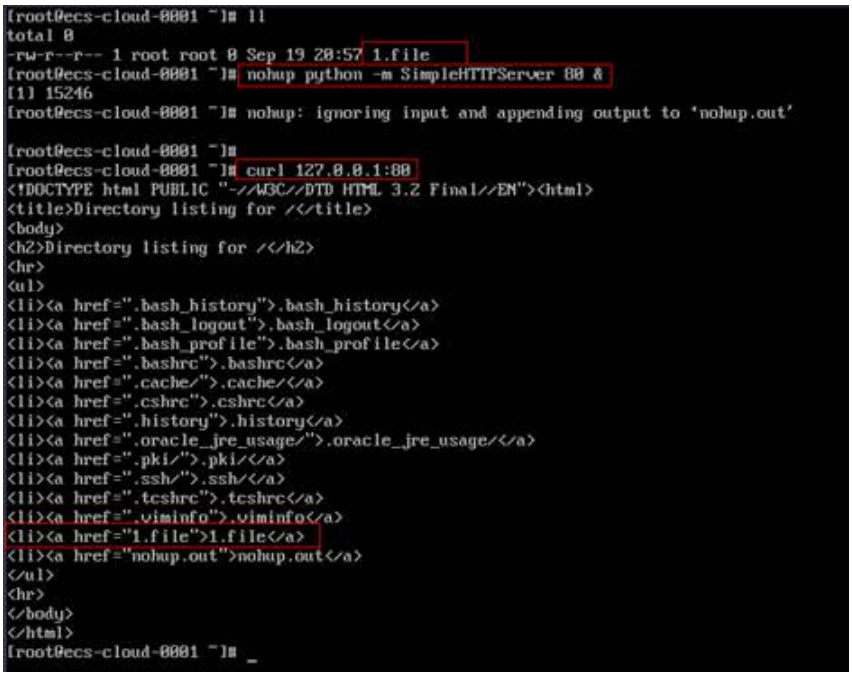
   登录第一个服务端ECS，在当前路径下创建名为“1.file”的文件，以标示第一个节点。

   并在当前路径执行以下命令启动HTTP服务。

   nohup python –m SimpleHTTPServer 80 &

   在第一个部署后端服务的虚拟机执行以下命令，确认HTTP服务正常。

   curl http://127.0.0.1:80



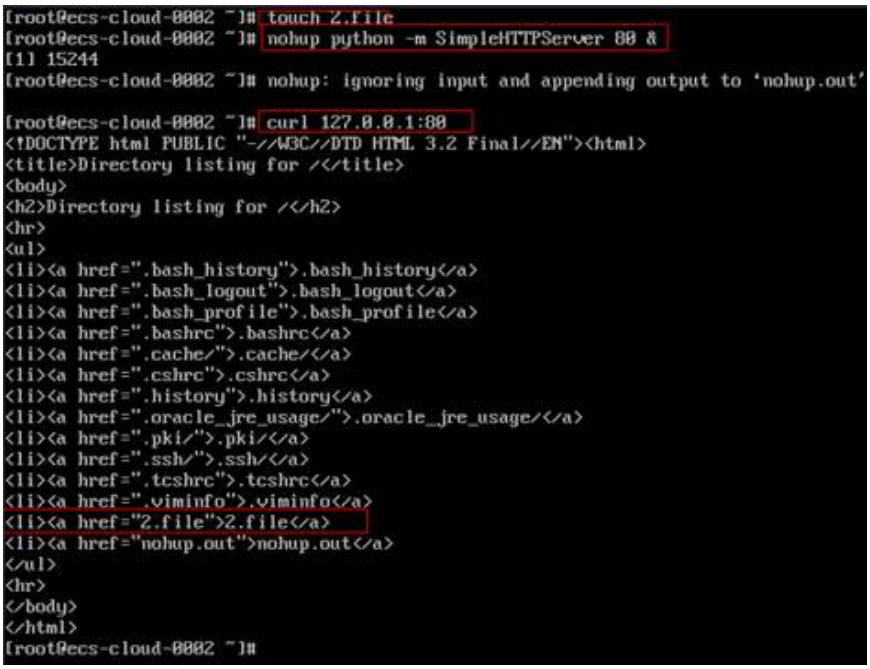
   登录第二个服务端ECS，在当前路径下创建名为“2.file”的文件，以标示第二个节点。

   并在当前路径执行以下命令启动HTTP服务。

nohup python –m SimpleHTTPServer 80 &

在本机执行以下命令，确认HTTP服务正常。

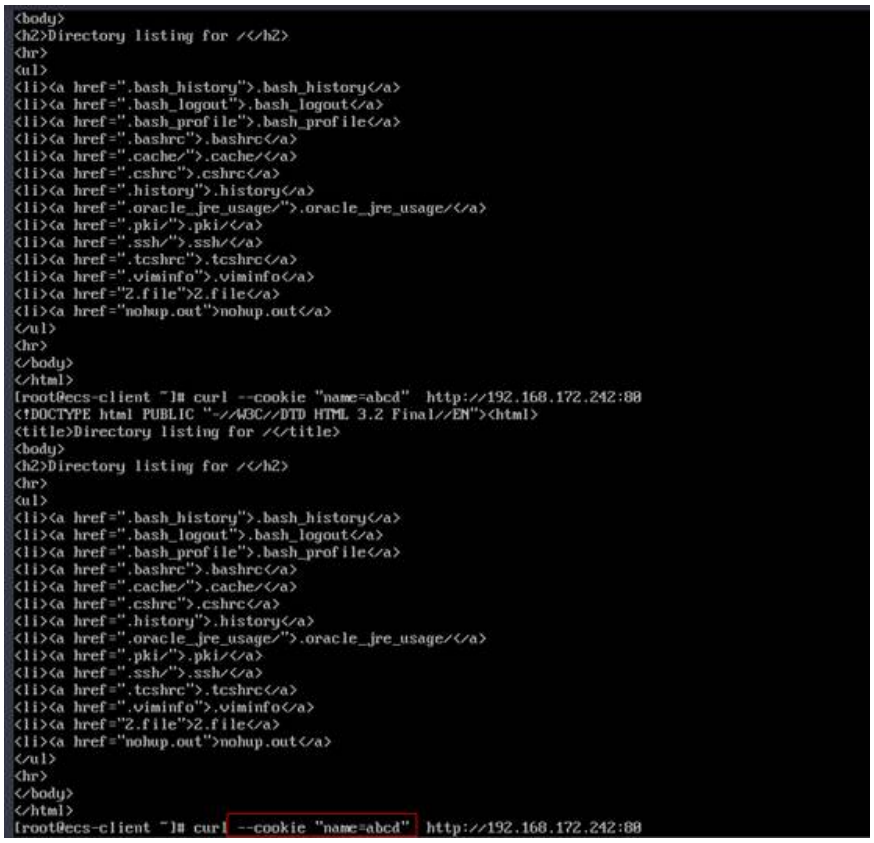
curl http://127.0.0.1:80



3. 从客户端ECS指定cookie值对ELB实例发起访问。

   调整以下命令，从客户端ECS对ELB实例发起访问，确认每次请求返回的file名称一致。

   curl --cookie "name=abcd" http://ELB_IP:Port



13.9.4 ELB 支持什么类型的会话保持?

负载均衡器支持源IP、负载均衡器cookie、应用程序cookie三种会话保持类型。

13.10 证书管理

13.10.1 如何生成服务器证书和 CA 证书?

一般的HTTPS业务场景只对服务器做认证，因此只需要配置服务器的证书即可，关于服务器证书和CA证书的生成，请参考。

13.10.2 ELB 是否支持泛域名证书?

支持，客户上传泛域名证书即可。

独享型负载均衡使用的SNI证书泛域名匹配方式默认为标准域名分级匹配，即只能匹配同级别的子域名。如您希望修改为最长尾缀匹配，请参考《API参考》修改参数sni_match_algo。

表 13-9 泛域名匹配规则示例

域名	标准域名分级匹配	最长尾缀匹配
*.example.com	abc.example.com、sport.example.com、good.example.com等域名	abc.example.com、mycalc.good.example.com等域名

13.10.3 配置了证书，访问异常是什么原因?

可能的原因有：

- 您在证书管理界面创建了证书，但因为您未使用HTTPS监听器，所以无法给监听器绑定证书。

  可以使用以下方法解决：

  - 继续使用现有非HTTPS监听器，并在后端服务器上安装证书。
  - 删除现有非HTTPS监听器，重新创建HTTPS监听器，并绑定证书。
- 您在证书管理界面创建了证书，且使用的是HTTPS监听器，但未将证书绑定至该监听器。
- 您的证书已过期。
- 创建证书时指定了域名，但访问的域名和创建证书时配置的域名不一致。
- 创建的证书为证书链时，没有按照证书链的格式拼接证书。

13.10.4 更换证书会导致网络或者 ELB 连接中断吗?

不会。

更换证书后，新的证书会立即生效，已经建立的连接会继续使用老证书，新建立的连接将会使用新的证书。

说明

证书过期后，用户访问时会提示“不安全的链接”，一般情况下忽略掉安全告警后，还是可以访问的。

13.11 监控

13.11.1 云监控 EIP 带宽使用统计与 ELB 监控的网络流出速率数据为何不一致?

以下两种情况监控EIP带宽使用统计与ELB监控的网络流出速率数据不一致：

- 如果流量没有超过EIP带宽，EIP未被限流，云监控EIP带宽使用统计外网访问数据，而ELB不仅采集外网访问数据，而且采集内网访问的数据。
- 如果流量超过EIP带宽，EIP会被限流，ELB内访问的数据流量跟EIP访问数据流量不是一个路径，ELB内访问数据流量不会被限流。

13.11.2 ELB 监控指标中七层协议返回码和七层后端返回码的区别?

ELB七层监听器会终结TCP连接。即客户端和ELB之间会建立TCP连接，ELB和后端主机之间会建立另外一条TCP连接。客户端把HTTP请求发送给ELB之后，ELB会解析并转发HTTP请求到后端主机，然后后端主机再返回HTTP响应给ELB，ELB再解析和转发HTTP响应到客户端，所以通信过程被分成前后两个阶段。协议返回码是指ELB返回给客户端的状态码，后端返回码是指后端主机返回给ELB的状态码。

协议返回码和后端返回码有如下三种情况：

- 后端主机有返回码，这种情况ELB会透传后端主机返回码到客户端，即协议返回码和后端返回码一致；
- ELB和后端主机连接异常或者超时等，ELB会填充后端返回码为502或者504，然后转发给客户端；
- 监听器配置异常或者客户端请求格式和内容异常时，ELB会直接返回4xx或者502返回码，不继续向后端主机转发请求，即有协议返回码，无后端返回码。

13.11.3 为什么七层监听器的监控中有大量 499 返回码?

HTTP返回码499对应的说明为：client has closed connection，即说明客户端主动断开了连接。

可能的原因：

- 客户端设置的请求超时时间太短，导致客户端未发送完HTTP请求就因为请求超时关闭了连接，建议排查访问日志中的request_time字段，该字段代表客户端请求的总时间，参考该字段的值设置合理的客户端请求超时时间。
- 访问ELB实例的流量太大，触发带宽限速丢包，建议通过云监控排查实例的出带宽使用率指标。更多信息，请参见**监控指标说明**。
- 客户端到ELB的网络链路有问题，存在往返延时比较大或丢包等问题，建议排查访问日志的request_time和tcpinfo_rtt字段或抓包排查客户端网络是否有异常。
- 后端服务器处理请求时间太长，超过了客户端的请求超时时间，建议排查后端服务器的CPU、内存、网络是否存在性能瓶颈。
- 客户端遇到未知问题，在未完成HTTP请求的情况下，提前关闭连接。建议排查客户端是否有提前关闭连接的行为。

14 修订记录

版本日期	变更说明
2024-04-15	第一次正式发布。