



弹性负载均衡

产品介绍

发布日期 2020-05-30

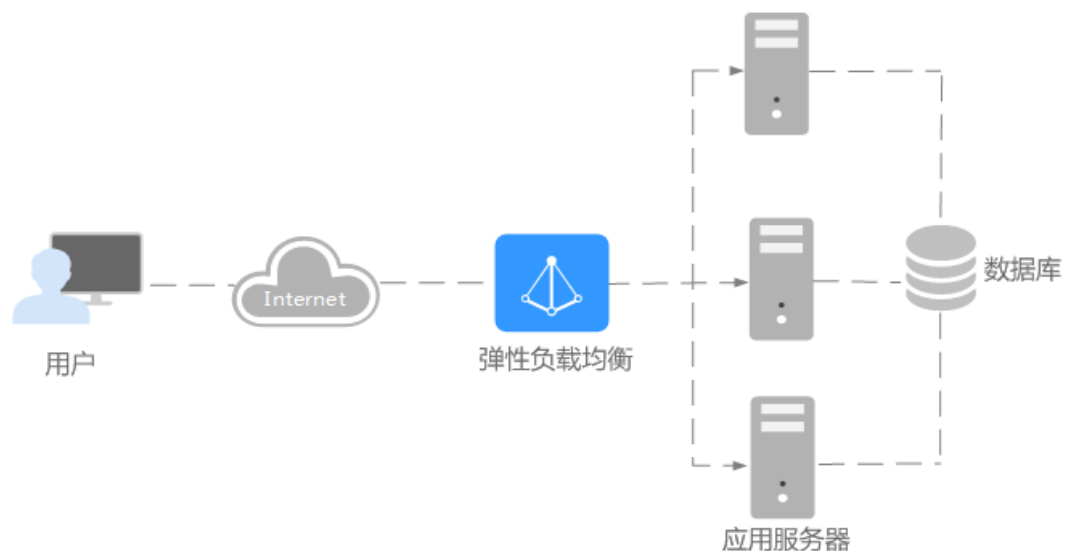
目录

1 什么是弹性负载均衡.....	1
2 产品优势.....	3
3 应用场景.....	5
4 共享型与经典型负载均衡的功能区别.....	8
5 弹性负载均衡是如何工作的.....	11
6 公网和私网负载均衡器.....	13
7 与其他服务的关系.....	15
8 计费说明.....	17
9 权限管理.....	19
10 基本概念.....	22
10.1 产品基本概念.....	22
10.2 区域和可用区.....	23
11 修订记录.....	25

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。



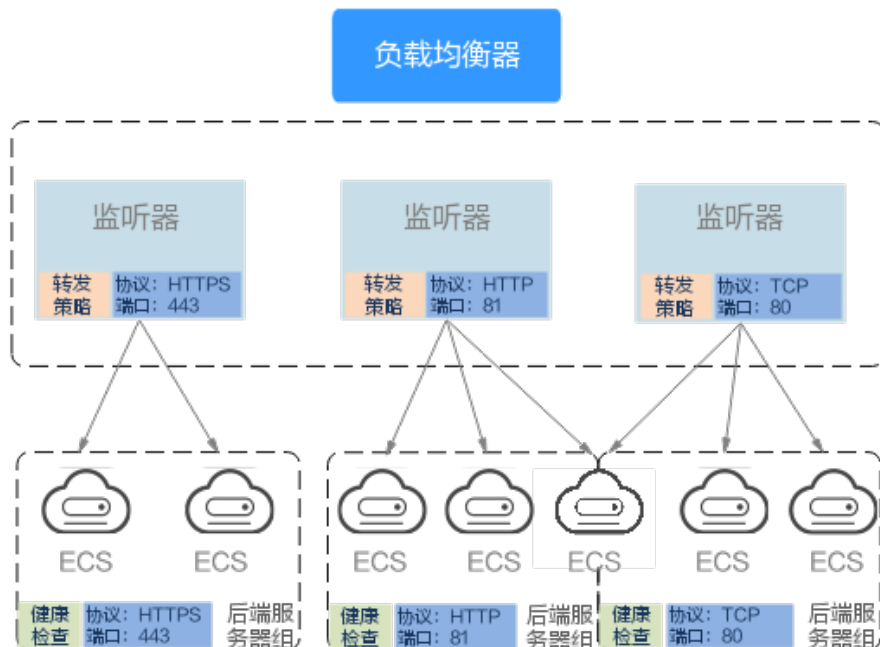
弹性负载均衡的组件

- 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 1-1 弹性负载均衡组件图



弹性负载均衡的类型

弹性负载均衡支持经典型负载均衡、共享型负载均衡。

- 经典型负载均衡：适用于访问量较小，应用模型简单的web业务。
- 共享型负载均衡：适用于访问量较大的web业务，提供基于域名和URL的路由均衡能力，实现更加灵活的业务转发。共享型负载均衡就是原增强型负载均衡。

两者的详细区别请参见[共享型与经典型负载均衡的功能区别](#)。

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：

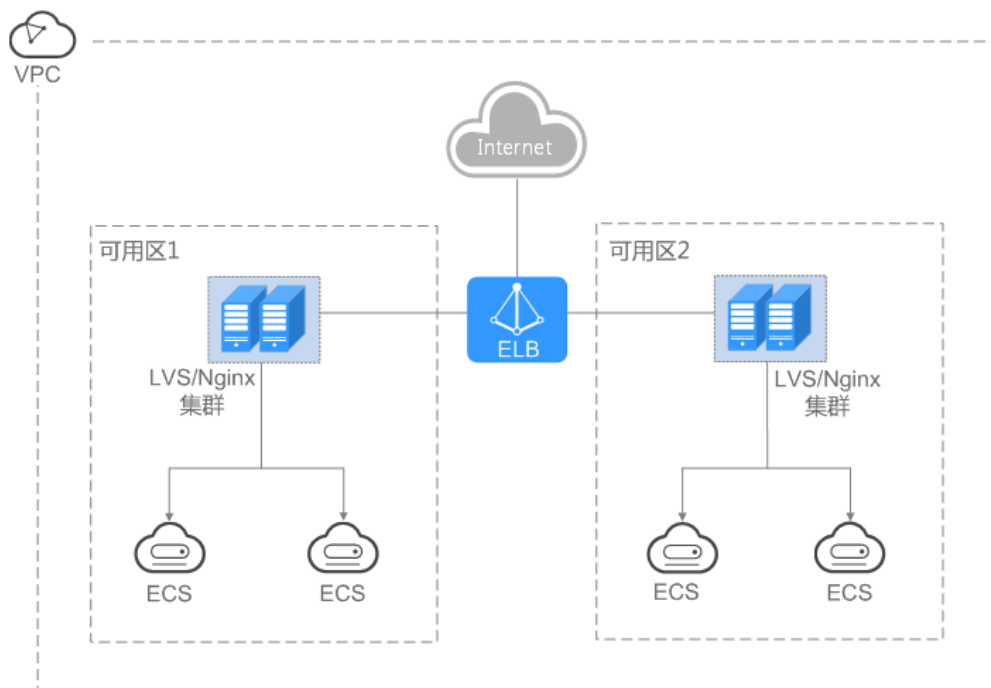
- 管理控制台
请使用管理控制台方式访问弹性负载均衡。可直接登录管理控制台，从主页选择“弹性负载均衡”。
- 查询API
通过调用API的方式访问弹性负载均衡，具体操作请参见《[弹性负载均衡API参考](#)》。

2 产品优势

弹性负载均衡具有以下优势：

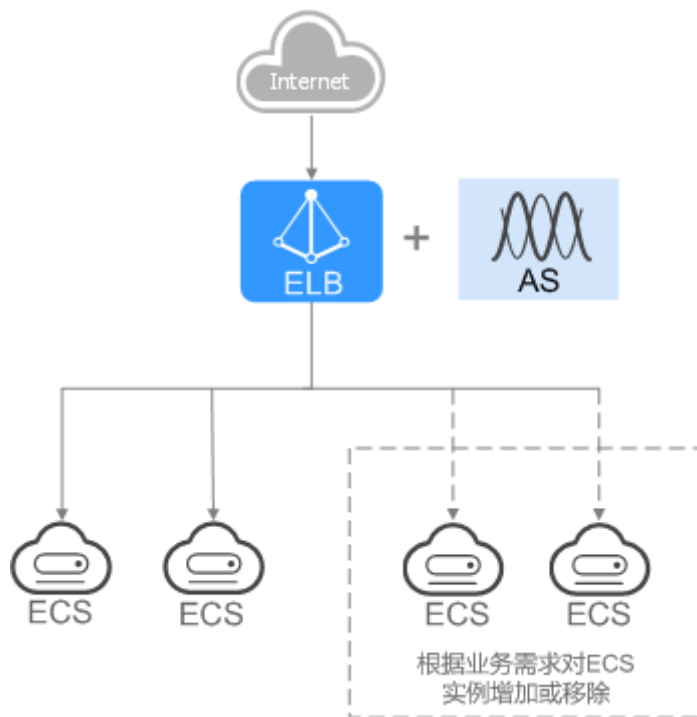
- 高性能
集群支持最高1亿并发连接，满足用户的海量业务访问需求。
- 高可用
采用集群化部署，支持多可用区的同城双活容灾，无缝实时切换。

图 2-1 高可用



- 灵活扩展
根据应用流量自动完成分发，与弹性伸缩服务无缝集成，灵活扩展用户应用的对外服务能力。

图 2-2 灵活扩展



- 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。

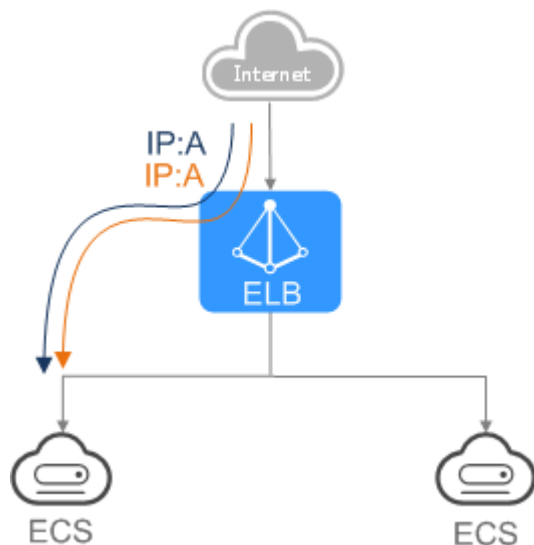
3 应用场景

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3-1所示。

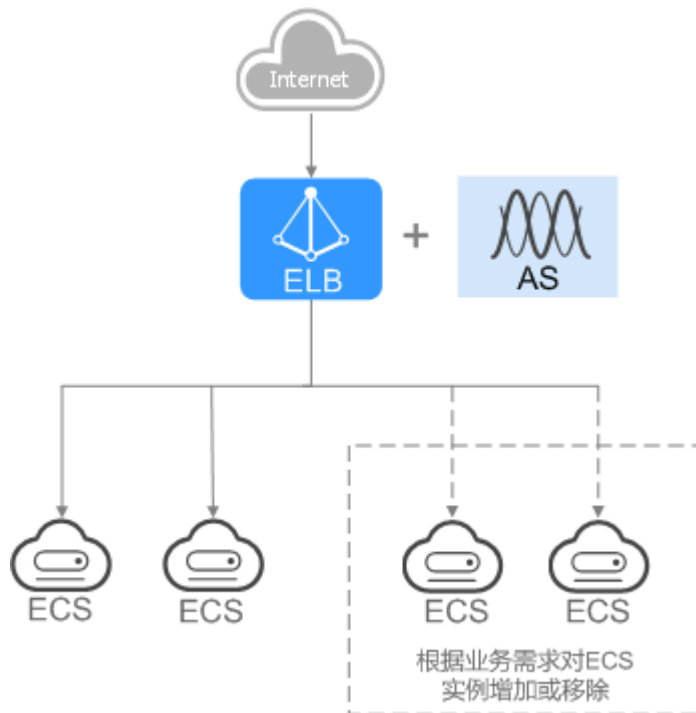
图 3-1 会话保持流量分发



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的后端云服务器组或者从ELB的后端云服务器组移除。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3-2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 3-2 灵活扩展

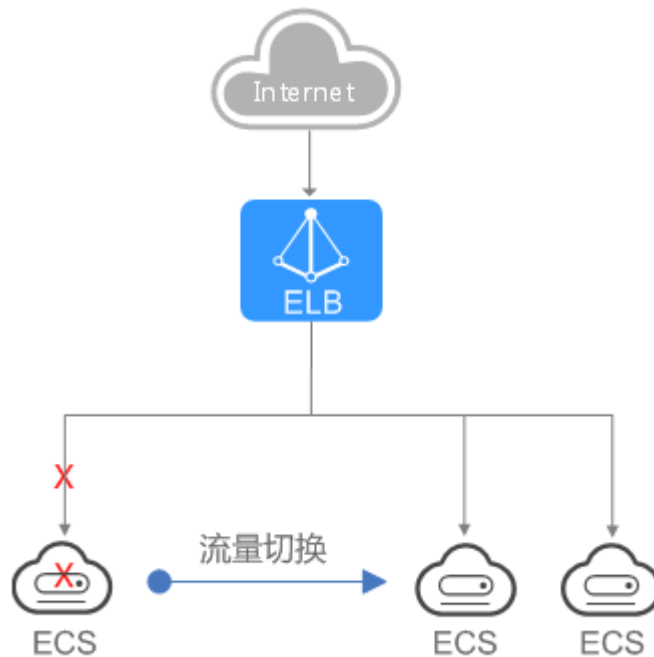


使用 ELB 消除单点故障

对于可靠性有较高要求的业务，可以在负载均衡器上添加多个后端云服务器。负载均衡器会通过健康检查及时发现并屏蔽有故障的云服务器，并将流量转发到其他正常运行的后端云服务器，确保业务不中断，如图3-3所示。

例如官网，计费业务，Web业务等。

图 3-3 消除单点故障

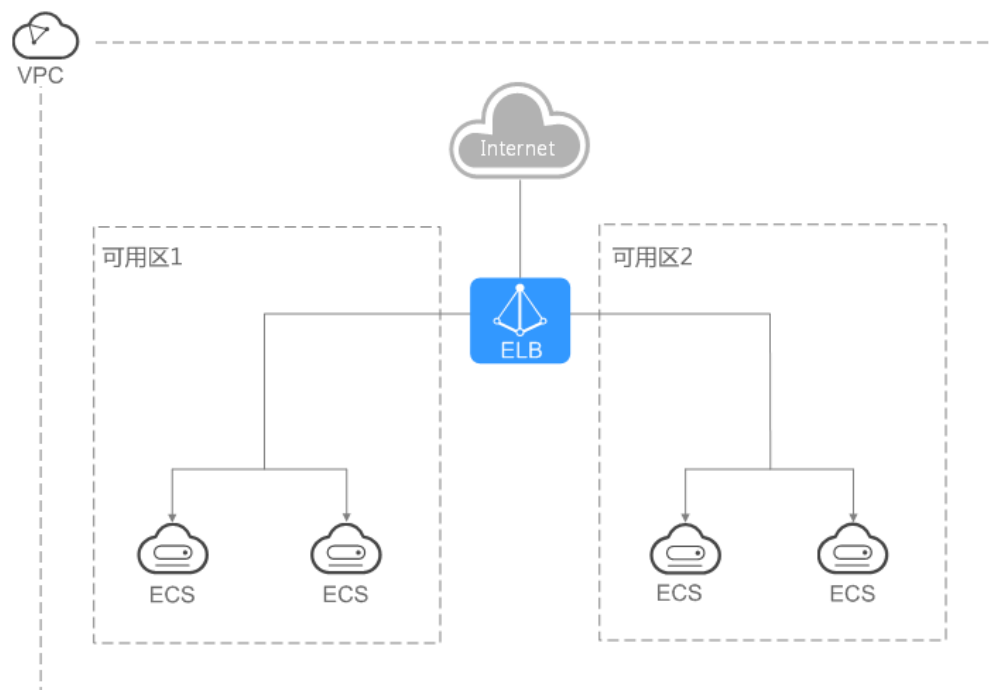


使用 ELB 跨可用区特性实现业务容灾部署

对于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云服务器进行处理，如图3-4所示。

例如银行业务，警务业务，大型应用系统等。

图 3-4 多可用区部署



4 共享型与经典型负载均衡的功能区别

弹性负载均衡有不同的负载均衡，分别是共享型负载均衡和经典型负载均衡，便于用户根据不同的应用场景和功能需求选择合适的负载均衡器类型。

- 经典型负载均衡：适用于访问量较小，应用模型简单的web业务。
- 共享型负载均衡：适用于访问量较大的web业务，提供基于域名和URL的路由均衡能力，实现更加灵活的业务转发。共享型负载均衡就是原增强型负载均衡。

共享型负载均衡对比经典型负载均衡，提供了更丰富的HTTP和HTTPS转发能力，同时在转发性能和稳定性上也有较大提升。两者具体的功能差异如表4-1所示。（“√”表示支持，“—”表示不支持。）

表 4-1 功能对比

功能	功能描述	经典型负载均衡	共享型负载均衡
支持公网和私网负载均衡	<ul style="list-style-type: none">● 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云服务器进行处理。● 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进行处理。	√	√
支持四层（TCP/UDP）和七层负载均衡（HTTP/HTTPS）	<ul style="list-style-type: none">● 四层负载均衡：支持TCP和UDP协议，监听器收到访问请求后，将请求直接转发给后端服务器。转发效率高，数据传输快。● 七层负载均衡：支持HTTP和HTTPS协议，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。转发效率不如四层负载均衡，但是支持加密传输、基于Cookie的会话保持等高级功能。	√（私网类型不支持UDP协议）	√

功能	功能描述	经典型负载均衡	共享型负载均衡
支持轮询 /最少连接/源IP	共享型负载均衡支持加权轮询、加权最小连接和源IP算法，经典型负载均衡支持轮询算法、最小连接和源IP算法。	√	√
支持会话保持	会话保持功能可以识别客户与服务器之间交互过程的关联性，在作负载均衡的同时，还保证一系列相关联的访问请求会保持分配到同一台服务器上。	√	√
支持WebSocket协议	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。	√	√
支持按域名和URL转发	通过添加转发策略支持自行设定的域名和URL，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。此功能目前仅支持协议类型为HTTP、HTTPS的监听器。	—	√
支持HTTP/2	HTTP/2，即超文本传输协议 2.0，是下一代HTTP协议。向下兼容HTTP1.X协议版本，同时性能和安全性都得到了提升。此功能目前仅支持协议类型为HTTPS的监听器。	—	√
支持后端服务器为ECS	后端服务器支持弹性云服务器。	√	√
支持访问控制（白名单）	通过添加白名单的方式控制访问负载均衡的监听器的IP，能够设置允许特定IP访问，而其他IP不许访问。	—	√
支持标准OpenStack API	可以通过标准OpenStack API接口调用负载均衡的API完成相关的操作，兼容性强。	—	√
支持后端服务器为裸机	后端服务器支持裸金属服务器。	—	√
支持SNI多证书特性	SNI (Server Name Indication) 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之后，用户需要添加域名对应的证书。	—	√
支持访问日志	通过对接云日志服务，可以查看和分析七层负载均衡HTTP和HTTPS的详细访问记录，包括请求时间、客户端IP地址、请求路径和服务器响应等。	—	√
支持权重	共享型负载均衡支持加权轮询、加权最小连接。	—	√

功能	功能描述	经典型负载均衡	共享型负载均衡
支持修改证书内容	证书创建成功后支持修改证书的内容。	—	√
支持双向认证	双向认证对通信双方的身份都要做认证，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。此功能目前仅支持协议类型为HTTPS的监听器。	—	√
支持HTTP重定向	将HTTP访问重定向至HTTPS，客户端HTTP的访问请求，后端服务器返回HTTPS的响应。目前仅共享型负载均衡支持此功能。	—	√
支持获取弹性公网IP	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。	—	√
支持按监听器粒度监控性能指标	公有云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。您可以通过管理控制台，直观地查看弹性负载均衡器的各项监控指标。	—	√

5 弹性负载均衡是如何工作的

您可以在弹性负载均衡服务中创建一个负载均衡器。该负载均衡器会接收来自客户端的请求，并将请求转发到一个或多个可用区的后端服务器中进行处理。请求的流量分发与负载均衡器配置的分配策略类型相关。

共享型负载均衡算法，支持以下三种调度算法：

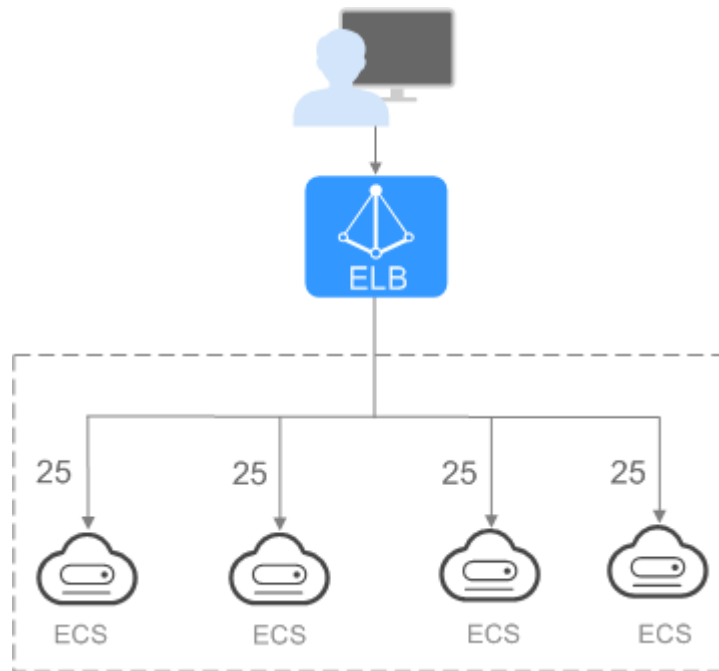
- 加权轮询算法：根据后端服务器的权重，按顺序依次将请求分发给不同的服务器。它用相应的权重表示服务器的处理性能，按照权重的高低以及轮询方式将请求分配给各服务器，相同权重的服务器处理相同数目的连接数。常用于短连接服务，例如HTTP等服务。
- 加权最少连接：最少连接是通过当前活跃的连接数来估计服务器负载情况的一种动态调度算法。加权最少连接就是在最少连接数的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权重，使其能够接受相应权值数的服务请求。常用于长连接服务，例如数据库连接等服务。
- 源IP算法：将请求的源IP地址进行一致性Hash运算，得到一个具体的数值，同时对后端服务器进行编号，按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。

经典型负载均衡算法，支持以下三种调度算法：

- 轮询算法：按顺序把每个新的连接请求分配给下一个服务器，最终把所有请求平分给所有的服务器。常用于短连接服务，例如HTTP等服务。
- 最少连接：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，系统把新的连接请求分配给当前连接数目最少的服务器。常用于长连接服务，例如数据库连接等服务。
- 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5-1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4台权重相同的后端服务器，负载均衡器节点会将25%的客户端流量分发到其可用区中的每一台后端服务器。

图 5-1 加权轮询算法流量分发



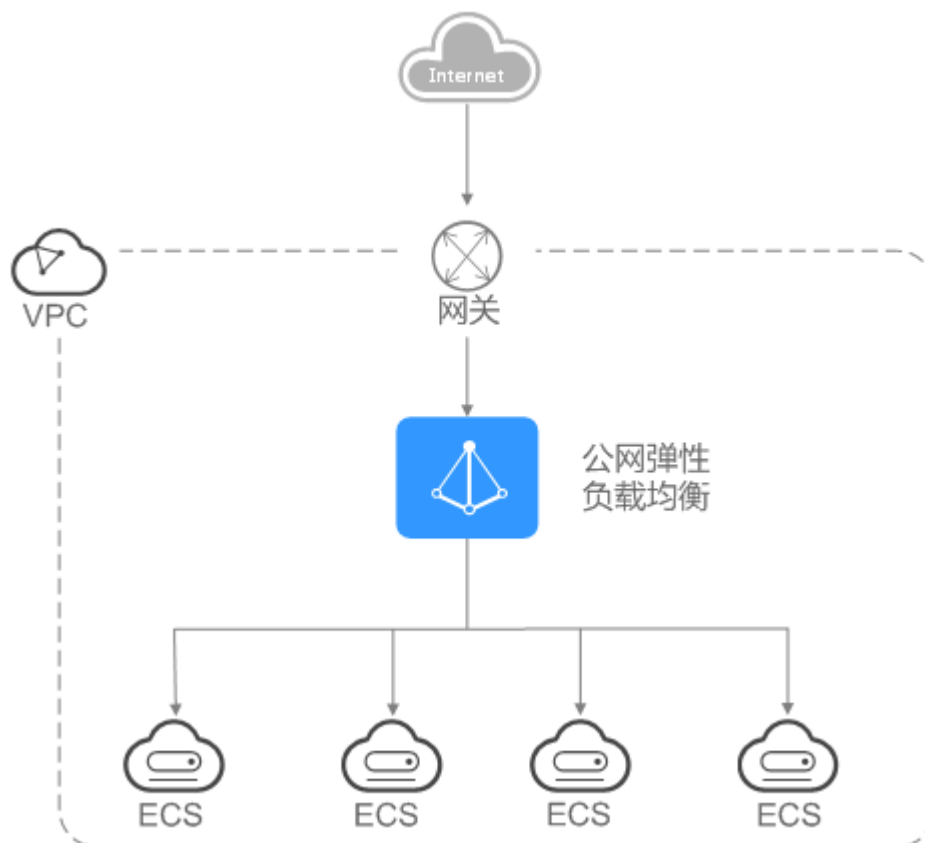
6 公网和私网负载均衡器

公网负载均衡器

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云服务器进行处理。

对于共享型公网负载均衡，不仅支持公网访问，也支持私网访问。

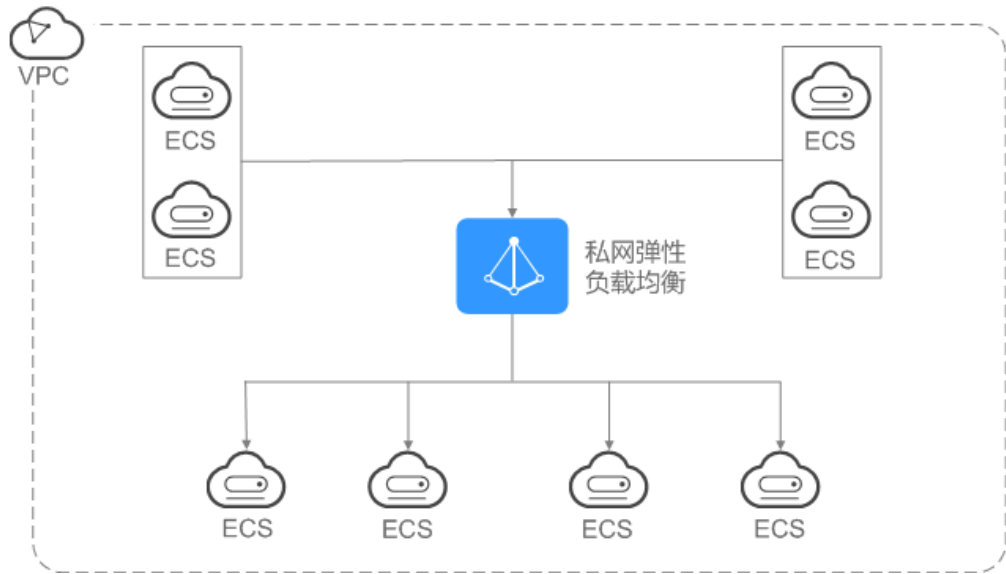
图 6-1 公网负载均衡器



私网负载均衡器

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进行处理。

图 6-2 私网负载均衡器



7 与其他服务的关系

弹性负载均衡服务与其它服务的依赖关系如图1所示：

图 7-1 弹性负载均衡服务其它服务的关系示例图

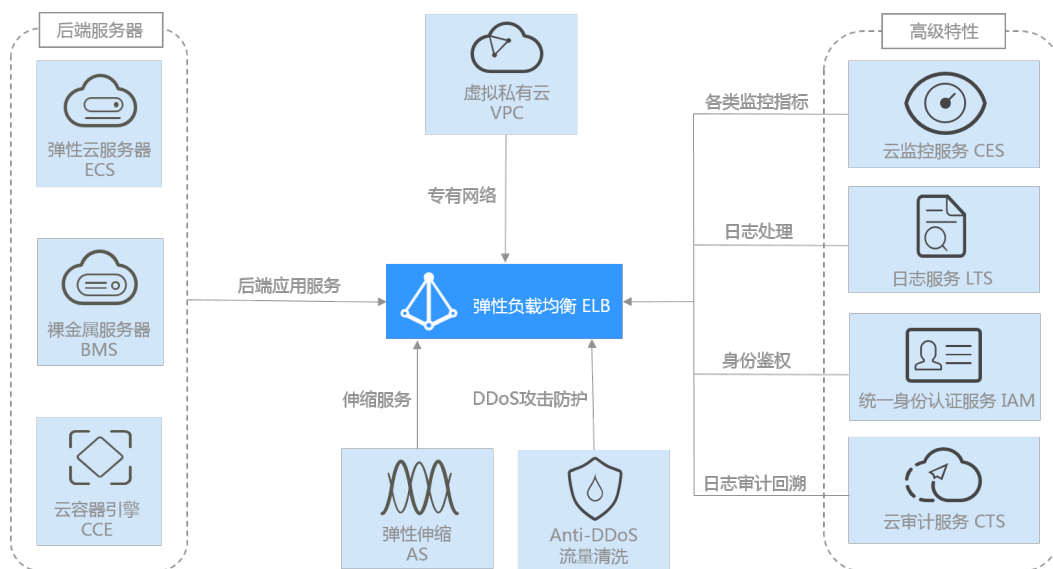


表 7-1 与其他服务之间关系

交互功能	服务名称	相关内容
通过相关服务部署用户业务，并接收ELB分发的访问流量。	弹性云服务器（Elastic Cloud Server, ECS）	购买弹性云服务器
	裸金属服务器（Bare Metal Server, BMS）	购买裸金属服务器
	云容器引擎（Cloud Container Engine, 简称 CCE）	CCE负载均衡

交互功能	服务名称	相关内容
创建ELB时需要使用虚拟私有云服务创建的弹性公网IP、带宽。	虚拟私有云（Virtual Private Cloud, VPC）	申请弹性公网IP
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。	弹性伸缩（Auto Scaling, AS）	创建弹性伸缩组
需要统一身份认证提供鉴权。	统一身份认证服务（Identity and Access Management, IAM）	创建用户组并授权
使用云审计服务记录弹性负载均衡服务的资源操作。	云审计服务（Cloud Trace Service, CTS）	查看审计日志
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。	云监控服务（Cloud Eye Service）	查看监控指标
当用户购买了Anti-DDoS服务后，配置了负载均衡器的公网IP，确保了弹性负载均衡服务免受外部攻击，提高安全性。	Anti-DDoS流量清洗服务（Anti-DDoS）	开启Anti-DDoS防护
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。	云日志服务（Log Tank Service, LTS）	访问日志

8 计费说明

计费项

弹性负载均衡提供两种类型的负载均衡：分为公网负载均衡和私网负载均衡。私网负载均衡不收取费用。公网负载均衡因为绑定了公网IP，会收取公网IP的费用以及公网IP所耗费的带宽费用。

带宽提供两种付费方式：包年包月和按需付费。

- 包年包月为预付费方式，目前只支持按固定带宽方式计费。

📖 说明

目前不支持创建负载均衡时选择预付费方式。可以先创建一个预付费的公网IP，然后创建负载均衡时选择该公网IP。或者将该公网IP绑定到一个已有的负载均衡。

- 按需付费为后付费购买方式，支持按固定带宽计费和按使用流量计费。
 - 如果选择按固定带宽计费，则根据带宽值和使用时长收费。
 - 如果选择按使用流量计费，则根据实际使用的流量收费。

计费方式

创建公网负载均衡器会绑定一个EIP，用来接收来自Internet公网的访问请求。

公网负载均衡器，还需要选择实例的计费方式：

- 按流量计费：指定带宽上限，按实际使用的出网流量计费，与使用时间无关。适用于波峰波谷效应明显的业务。
- 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。适用于带宽较为平稳的业务。

变更配置

当负载均衡器是公网类型时，通过带宽提供负载均衡器和公网之间的访问流量，您可以变更配置。

- 按需付费弹性公网IP支持修改带宽名称、大小和计费方式。
- 带宽计费方式：按带宽计费将按固定带宽大小收费，按流量计费将按带宽实际使用的流量收费。

续费

如需续费，请在管理控制台[续费管理](#)页面进行续费操作。详细操作请参考[续费管理](#)。

到期与欠费

客户欠费后，可以查看欠费详情。为防止相关资源被停止或者释放，需要客户及时进行充值，账号将进入欠费状态，需要在约定时间内支付欠款，详细操作请参考[欠费还款](#)。

公网负载均衡因为绑定了公网IP，会收取公网IP的费用以及公网IP所耗费的带宽费用。所以欠费超过宽限期后，公网负载均衡将无法继续使用、私网负载均衡仍可以继续使用。弹性负载均衡的其他操作不受影响，例如创建监听器、添加后端服务器。

9 权限管理

如果您需要对华为云上购买的弹性负载均衡（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华为云资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并使用策略来控制他们对华为云资源的访问范围。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。

如果华为云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是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见《[IAM产品介绍](#)》。

ELB 权限

默认情况下，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（如华北-北京1）对应的项目（cn-north-1）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。由于华为云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，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[策略及授权项说明](#)。

如表9-1所示，包括了ELB的所有系统权限。

表 9-1 ELB 系统权限

系统角色/ 策略名称	描述	类别
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。	细粒度策略
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。	细粒度策略
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 Guest权限。 作用范围：项目级服务。 说明 <ul style="list-style-type: none"> 此策略名称之前为ELB Service Administrator，新的策略名称于2020/3/30 22:00（北京时间）正式生效。 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RBAC策略。 	RBAC策略

表9-2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 9-2 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	可以	不可以	可以
查询负载均衡器	可以	可以	可以
查询负载均衡器状态树	可以	可以	可以
查询负载均衡器列表	可以	可以	可以
更新负载均衡器	可以	不可以	可以
删除负载均衡器	可以	不可以	可以
创建监听器	可以	不可以	可以
查询监听器	可以	可以	可以

操作	ELB FullAccess	ELB ReadOnlyAccess	ELB Administrator
创建后端服务器组	可以	不可以	可以
创建后端服务器	可以	不可以	可以
创建健康检查	可以	不可以	可以

相关链接

- [IAM产品介绍](#)。
- 创建用户组、用户并授予ELB权限请参考：[创建用户并授权使用ELB](#)。
- 策略支持的授权项[策略及授权项说明项](#)。

10 基本概念

10.1 产品基本概念

表 10-1 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云服务器。
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	会话保持就是指在负载均衡器上有这么一种机制，可以识别客户与服务器之间交互过程的关联性，在作负载均衡的同时，还保证一系列相关联的访问请求会保持分配到同一台服务器上。

名词	说明
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	如果用户的后台应用对外提供多个域名的访问，并且每个域名都使用独立的证书，则需要创建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多个域名和证书的TLS扩展，主要解决一台服务器只能使用一个证书的缺点。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 监听器支持此功能，即支持绑定多个证书。
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。

10.2 区域和可用区

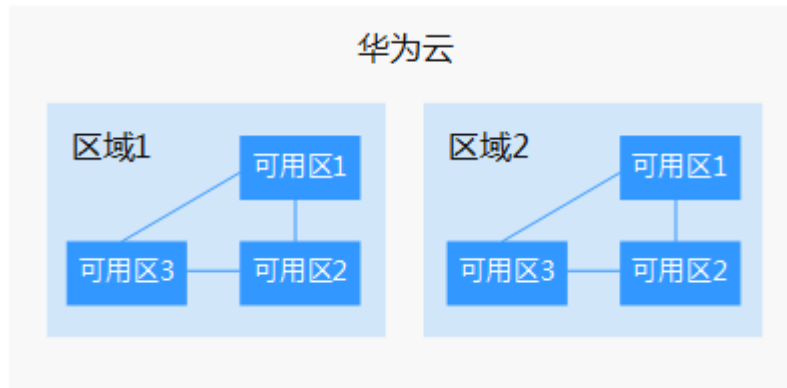
什么是区域、可用区？

我们用区域和可用区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)：从地理位置和网络时延维度划分，同一个Region内共享弹性计算、块存储、对象存储、VPC网络、弹性公网IP、镜像等公共服务。Region分为通用Region和专属Region，通用Region指面向公共租户提供通用云服务的Region；专属Region指只承载同一类业务或只面向特定租户提供业务服务的专用Region。
- 可用区 (AZ, Availability Zone)：一个AZ是一个或多个物理数据中心的集合，有独立的风火水电，AZ内逻辑上再将计算、网络、存储等资源划分成多个集群。一个Region中的多个AZ间通过高速光纤相连，以满足用户跨AZ构建高可用性系统的需求。

图10-1阐明了区域和可用区之间的关系。

图 10-1 区域和可用区



目前，华为云已在全球多个地域开放云服务，您可以根据需求选择适合自己的区域和可用区。更多信息请参见[华为云全球站点](#)。

如何选择区域？

选择区域时，您需要考虑以下几个因素：

- 地理位置

一般情况下，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。不过，在基础设施、BGP网络品质、资源的操作与配置等方面，中国大陆各个区域间区别不大，如果您或者您的目标用户在中国大陆，可以不用考虑不同区域造成的网络时延问题。

- 在除中国大陆以外的亚太地区有业务的用户，可以选择“亚太-香港”、“亚太-曼谷”或“亚太-新加坡”区域。
- 在非洲地区有业务的用户，可以选择“南非-约翰内斯堡”区域。
- 在欧洲地区有业务的用户，可以选择“欧洲-巴黎”区域。

- 资源的价格

不同区域的资源价格可能有差异，请参见[华为云服务价格详情](#)。

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。有关华为云的区域和终端节点的更多信息，请参见[地区和终端节点](#)。

11 修订记录

版本日期	变更说明
2020-05-30	第四次正式发布。 变更如下： 增强型负载均衡更名为共享型负载均衡。
2019-11-06	第三次正式发布。 变更如下： 1、调整目录； 2、增加产品基本概念章节； 3、优化各章节的内容描述。
2018-12-30	第二次正式发布。 变更如下： 1，调整目录。 2，优化各个章节内容。
2018-06-30	第一次正式发布。