阿里云
负载均衡

历史文档

文档版本：20190815

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	**禁止:** 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	**警告:** 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	**说明:** 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 确定。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
##	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

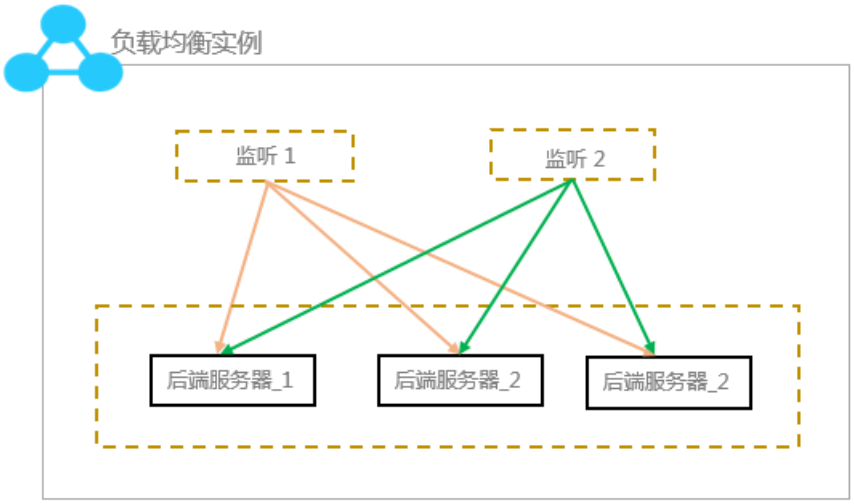
法律声明.....I
通用约定.....I
1 用户指南（旧版控制台）.....1
1.1 负载均衡实例.....1
1.1.1 什么是负载均衡实例.....1
1.1.2 性能保障型实例.....3
1.1.3 网络流量路径说明.....11
1.1.4 创建实例.....12
1.1.5 管理实例.....14
1.1.6 管理标签.....14
1.1.7 绑定EIP.....16
1.1.8 按量付费实例变配.....16
1.2 监听.....17
1.2.1 监听介绍.....17
1.2.2 四层监听.....19
1.2.3 七层监听.....19
1.2.4 健康检查.....19
1.2.5 共享实例带宽.....19
1.3 后端服务器.....21
1.3.1 后端服务器概述.....21
1.3.2 添加默认服务器.....22
1.3.3 创建虚拟服务器组.....23
1.3.4 创建主备服务器组.....24
1.4 DDoS基础防护.....25
1.5 各地域带宽峰值限制.....26
1.6 多可用区.....27
1.7 结合云解析实现跨地域负载均衡.....30

1 用户指南（旧版控制台）

1.1 负载均衡实例

1.1.1 什么是负载均衡实例

负载均衡实例是一个运行的负载均衡服务实体。使用负载均衡服务，您必须创建一个负载均衡实例，在实例中添加监听和后端服务器。



阿里云提供公网和私网两种类型的负载均衡服务。您可以根据业务场景选择配置对外公开或对内私有的负载均衡，系统会根据您的选择分配公网或私网服务地址。

公网负载均衡实例

公网类型的负载均衡实例可以通过Internet将客户端请求按照您制定的监听规则分发到添加的后端服务器ECS上。

在您创建公网负载均衡实例后，系统会为其分配一个公网服务地址，您可以将您的域名和该公网服务地址进行绑定，对外提供服务。

私网负载均衡实例

私网类型的负载均衡实例只能在阿里云内部使用，可以转发的请求只能来自具有负载均衡的私网访问权限的客户端。

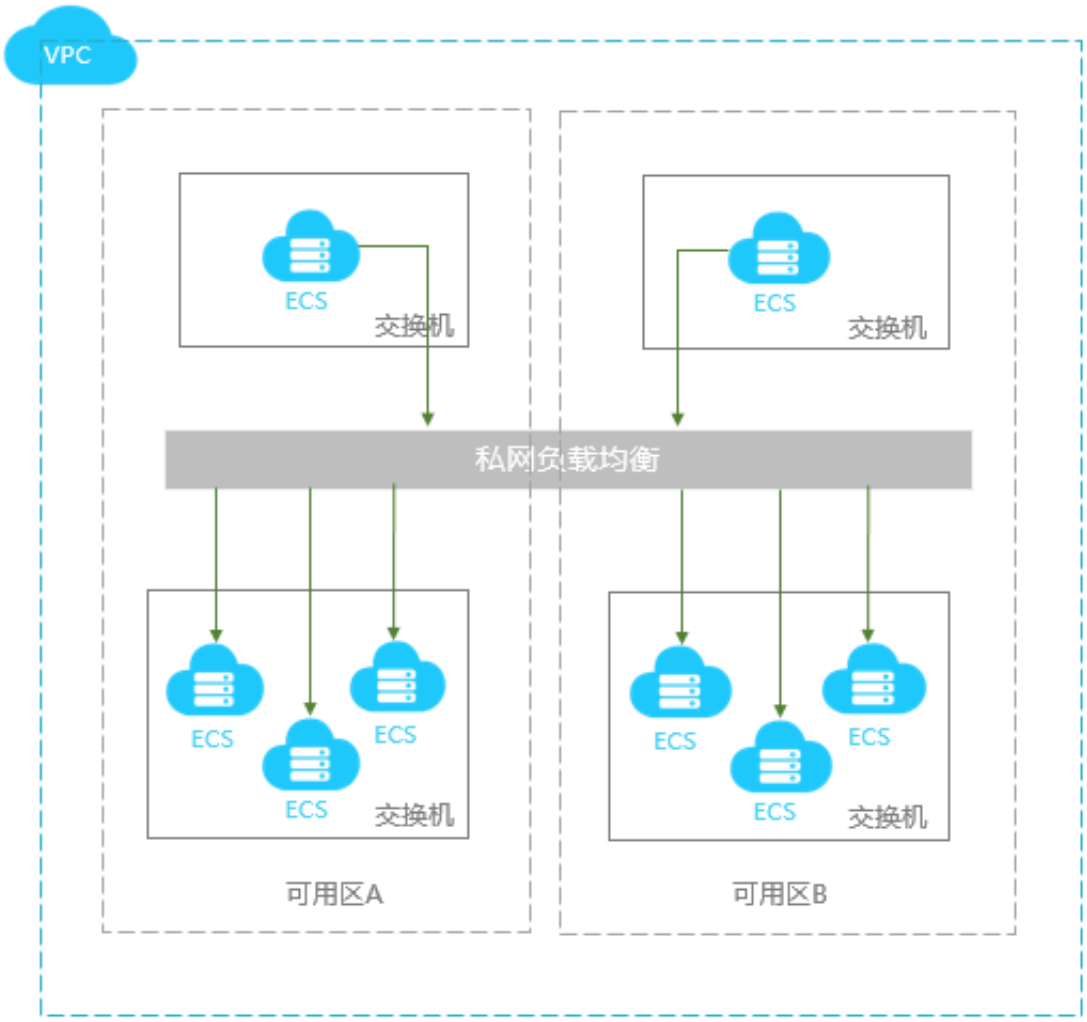
私网负载均衡实例可以进一步对网络类型进行选择：

- 经典网络

  如果您选择的私网负载均衡实例的网络类型是经典网络，那么您的私网负载均衡实例的服务地址由阿里云统一分配和管理。该私网负载均衡服务只能被经典网络ECS实例访问。

- 专有网络

  如果您选择的私网负载均衡实例的网络类型是专有网络，那么您的私网负载均衡实例的服务地址会从您指定的专有网络的交换机网段内分配。该私网负载均衡服务只能被相同VPC内的ECS实例访问。



1.1.2 性能保障型实例

阿里云负载均衡计划将于2018年4月1日开始针对性能保障型实例收取规格费，同时继续保留性能共享型实例的售卖。

1. 什么是负载均衡性能保障型实例?
2. 性能保障型实例如何收费?
3. 性能保障型实例规格费的定价
4. 如何选择性能保障型实例?
5. 是否可以调整性能保障型实例的规格?

6. 性能保障型实例何时收费?

7. 收费后性能共享型实例会收费吗?

8. 为何有时性能保障型实例看起来达不到规格中的性能指标上限?

9. 还可以购买性能共享型实例吗?

10. 私网负载均衡实例也会收取规格费吗?

1. 什么是负载均衡性能保障型实例?

负载均衡性能保障型实例提供了可保障的性能指标。与之相对的是负载均衡性能共享型实例，资源是所有实例共享的，所以不保障实例的性能指标。

在推出负载均衡性能保障型实例之前，您所有购买的实例均为性能共享型实例。在控制台上，您可以查看已购实例的类型。

把鼠标移至性能保障型实例的绿色图标，可查看具体的性能指标，如下图所示。

性能保障型实例的三个关键指标如下:

- 最大连接数-Max Connection

  最大连接数定义了一个负载均衡实例能够承载的最大连接数量。当实例上的连接超过规格定义的最大连接数时，新建连接请求将被丢弃。

- 每秒新建连接数-Connection Per Second (CPS)

  每秒新建连接数定义了新建连接的速率。当新建连接的速率超过规格定义的每秒新建连接数时，新建连接请求将被丢弃。

- 每秒查询数-Query Per Second (QPS)

  每秒请求数是七层监听特有的概念，指的是每秒可以完成的HTTP/HTTPS的查询（请求）的数量。当请求速率超过规格所定义的每秒查询数时，新建连接请求将被丢弃。

阿里云负载均衡性能保障型实例开放了如下六种实例规格（各地域因资源情况不同，开放的规格可能略有差异，请以控制台购买页为准）。

规格	规格	最大连接数	每秒新建连接数 (CPS)	每秒查询数(QPS)
规格 1	简约型I (slb.s1.small)	5,000	3,000	1,000
规格 2	标准型I (slb.s2.small)	50,000	5,000	5,000
规格 3	标准型II (slb.s2.medium)	100,000	10,000	10,000
规格 4	高阶型I (slb.s3.small)	200,000	20,000	20,000
规格 5	高阶型II (slb.s3.medium)	500,000	50,000	30,000
规格 6	超强型I (slb.s3.large)	1,000,000	100,000	50,000

如果需要更大规格，请联系您的客户经理申请。

2. 性能保障型实例如何收费?

负载均衡性能保障型实例需要收取规格费用，收费模型如下:

性能保障型费用 = 实例费 + 流量/带宽费 + 规格费

说明:

负载均衡私网实例也可以选择性能共享型实例或性能保障型实例，性能保障型私网实例，也需要收取规格费用，收费方式与公网性能保障型实例一致，但不收取流量费/带宽费和实例费。

性能保障型实例规格费按使用量收取，即不论您选择的何种规格，实例规格规格费均会按照您实际使用的规格收取。

例如，您选择了超强型I (slb.s3.large)规格（最大连接数1,000,000；CPS 500,000；QPS 50,000）。您的实例在某个小时内各项指标产生的实际峰值如下:

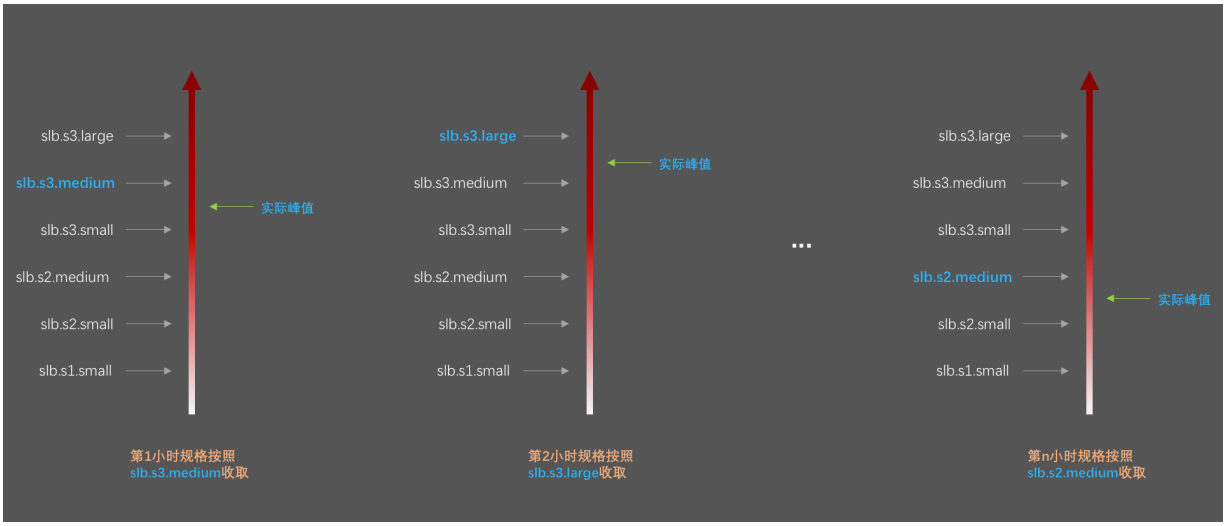
最大连接数	每秒新建连接数（CPS）	每秒查询数 （QPS）
90000	4000	11000

- 从最大连接数维度看，90,000超过slb.s2.small规格中最大连接数50,000的上限，但未达到slb.s2.medium规格中最大连接数的100,000上限，因此从最大连接数维度计算，该小时规格为slb.s2.medium。

- 从每秒新建连接数（CPS）维度看，4,000超过slb.s1.small规格中CPS 3,000的上限，但未到达slb.s2.small规格中CPS 5,000的上限，因此从CPS维度计算，该小时规格为slb.s2.small。
- 从每秒查询数（QPS）维度看，11,000超过slb.s2.medium规格中QPS 10,000的上限，但未达到slb.s3.small中QPS 20,000的上限，因此从QPS维度计算，该小时规格为slb.s3.small。

  综合以上三个维度，QPS指标的规格（slb.s3.small）最大，因此将QPS维度的规格作为该小时实例的综合规格，该小时内该实例将按照slb.s3.small规格进行计费。

以后每小时规格费均按照上述方式计算，如下图所示：



因此，按量付费的性能保障型实例具有自动弹性伸缩（或计费）的能力。您在购买时所选的规格，是性能的上限，比如您选择高阶型II (slb.s3.medium)，那么意味着，您的实例最大可以达到的规格上限就是高阶型II (slb.s3.medium)。

3. 性能保障型实例规格费的定价

下表中所列的只是规格费用。除规格费以外，负载均衡实例的实例费和流量正常收取。更多详细信息，参考按量计费。

地域	规格	最大连接数	每秒新建连接数 (CPS)	每秒查询数(QPS)	规格费（美元/小时）
华东1（杭州） 华北3（张家口） 华北5（呼和浩特） 华北1（青岛） 华北2（北京） 华东2（上海） 华南1（深圳）	规格1：简约型I (slb.s1.small)	5000	3000	1000	免费
	规格2：标准型I (slb.s2.small)	50,000	5,000	5,000	0.05
	规格3：标准型II (slb.s2.medium)	100,000	10,000	10,000	0.10
	规格4：高阶型I (slb.s3.small)	200,000	20,000	20,000	0.20
	规格5：高阶型II (slb.s3.medium)	500,000	50,000	30,000	0.31
	规格6：超强型I (slb.s3.large)	1,000,000	100,000	50,000	0.51
亚太东南1（新加坡） 亚太东南3（吉隆坡） 亚太东南5（雅加达） 亚太南部1（孟买） 美国西部1（硅谷） 美国东部1（弗吉尼亚） 香港	规格1：简约型I (slb.s1.small)	5,000	3,000	1,000	免费
	规格2：标准型I (slb.s2.small)	50,000	5,000	5,000	0.06
	规格3：标准型II (slb.s2.medium)	100,000	10,000	10,000	0.12
	规格4：高阶型I (slb.s3.small)	200,000	20,000	20,000	0.24
	规格5：高阶型II (slb.s3.medium)	500,000	50,000	30,000	0.37
	规格6：超强型I (slb.s3.large)	1,000,000	100,000	50,000	0.61

目前海外地域规格费享83折优惠。

4. 如何选择性能保障型实例?

规格费是按量（弹性）计费的，因此建议您直接选择您可以买到的最大规格，对于大多数用户而言，即高阶型I(slb.s3.large)，这样可以保证较好的业务灵活性（弹性），且不会让您额外多付出成本。但如果您认为您的业务量不太可能到达超强型I(slb.s3.large），也可以设置一个合理的弹性上限，比如高阶型II(slb.s3.medium)。

5. 是否可以调整性能保障型实例的规格?

您可在控制台对性能保障型实例进行变配，如下图所示。

说明:

- **将性能共享型实例变更为性能保障型实例后，无法再将其变更回性能共享型。**
- **由于历史存量原因，部分实例可能存在于较老的集群。此部分实例在变配到性能保障型实例时，因为需要将实例迁移，因此可能出现10-30秒的业务中断，其他变配操作均不会影响业务。因此建议在业务低谷期进行此类变配。**
- **所有的变配操作都不影响负载均衡实例的IP地址。**

6. 性能保障型实例何时收费？

阿里云负载均衡计划将于2018年4月1日开始针对性能保障型实例收取规格费，同时继续保留性能共享型实例的售卖。

性能保障型实例的规格费收取将按地域分批次生效：

- 第一批：

  生效时间：4月1日至4月10日陆续生效

  生效地域：亚太东南1（新加坡）、亚太东南3（吉隆坡）、亚太东南5（雅加达）、亚太南部1（孟买）、美国西部1（硅谷）、美国东部1（弗吉尼亚）
- 第二批：

  生效时间：4月11日至4月20日陆续生效

  生效地域：华东1（杭州）、华北3（张家口）、华北5（呼和浩特）、香港
- 第三批：

  生效时间：4月21日至4月30日陆续生效

  生效地域：华北1（青岛）、华北2（北京）、华东2（上海）、华南1（深圳）

7. 收取规格费以后，性能共享型实例需也会额外收取费用吗？

不会。

原有的性能共享型实例（如果您不将其变配性能保障型）将继续保持为性能共享型实例，不收取规格费。您也可以通过变配，将性能共享型实例升级成性能保障型实例。变更成性能保障型后，当性能保障型实例开始正式收费时，该实例将收取规格费。

8. 为何有时性能保障型实例看起来达不到规格中的性能指标上限?

短木板原理。

性能保障型实例并不保障三个指标（包含带宽指标）同时达到指定规格的指标上限。即规格中哪个指标先达到峰值，就以哪个指标开始限速。

比如某用户选择高阶型I（slb.s3.small）实例，当实例的QPS已经达到20000，但并发连接数确远未达到20万，那么该实例最大连接数可能永远都不会达到规格上限，因为新建的连接请求会因为QPS达到上限而被丢弃。

9. 还可以购买性能共享型实例吗?

可以。

当前继续开放性能共享型实例的售卖，后续性能共享型实例有可能会下线，届时会通过官网公告、邮件等方式通知。

10. 私网负载均衡实例也会收取规格费吗?

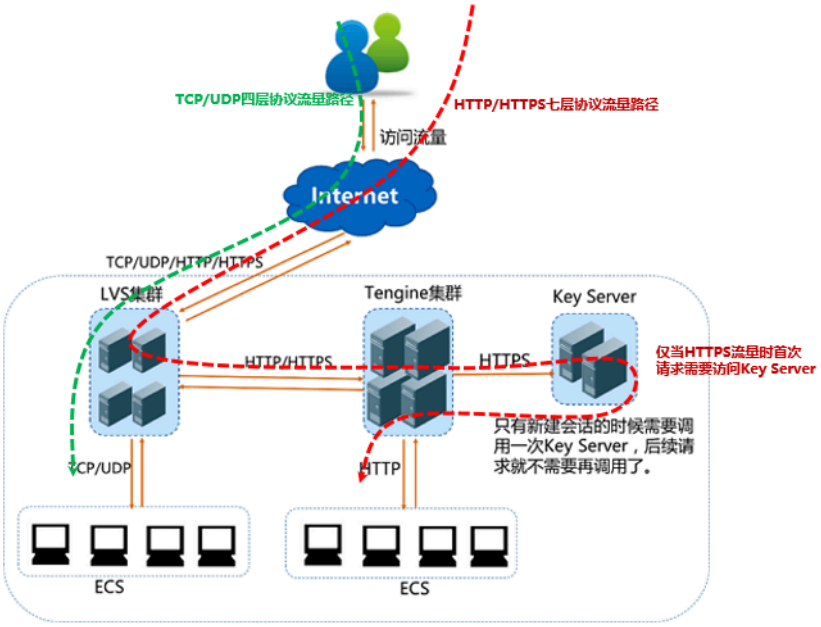
如果您选择的是性能共享型私网实例，则不会收取规格费；如果您选择的是性能保障型私网实例，则需要收取规格费。规格费收取方式与公网实例规格费计费规则一致。私网实例免收实例费和流量费。

1.1.3 网络流量路径说明

入网流量路径

对于入网流量，负载均衡会根据用户在控制台或API上配置的转发策略，对来自前端的访问请求进行转发和处理，数据流转如图 1-1: 入网流量路径所示。

图 1-1: 入网流量路径



1. TCP/UDP协议和HTTP/HTTPS协议的流量都需要经过LVS集群进行转发。
2. LVS集群内的每一台节点服务器均匀地分配海量访问请求，并且每一台节点服务器之间都有会话同步策略，以保证高可用。
   - 如果相应的负载均衡实例服务端口使用的是四层协议（TCP或UDP），那么LVS集群内每个节点都会根据负载均衡实例负载均衡策略，将其承载的服务请求按策略直接分发到后端ECS服务器。
   - 如果相应的负载均衡实例服务端口使用的是七层HTTP协议，那么LVS集群内每个节点会先将其承载的服务请求均分到Tengine集群，Tengine集群内的每个节点再根据负载均衡策略，将服务请求按策略最终分发到后端ECS服务器。
   - 如果相应的负载均衡实例服务端口使用的是七层HTTPS协议，与上述HTTP处理过程类似，差别是在按策略将服务请求最终分发到后端ECS服务器前，先调用Key Server进行证书验证及数据包加解密等前置操作。

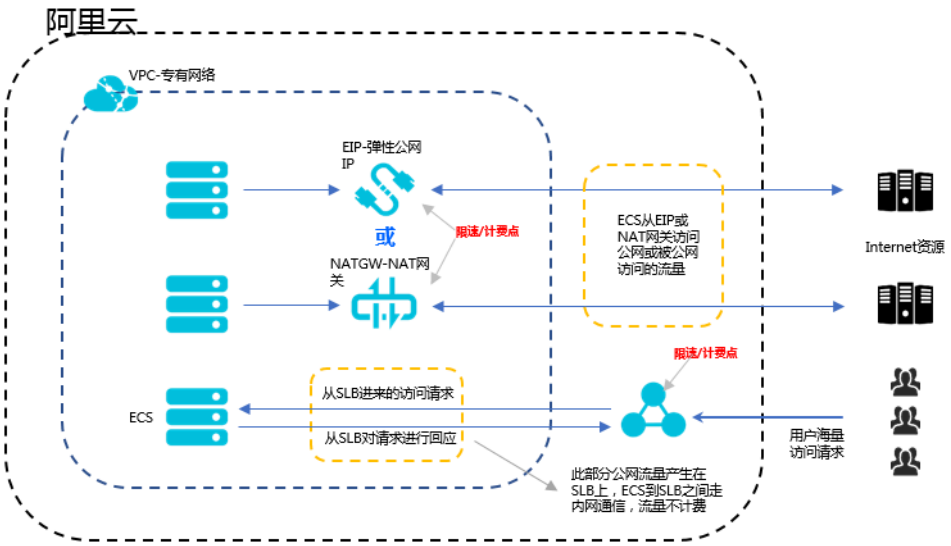
出网流量路径

负载均衡SLB和后端ECS之间是通过内网进行通信的。

- 如果ECS仅仅处理来自负载均衡的请求，可以不购买公网带宽（ECS公网IP/弹性公网IP/NAT网关等）。
- 如果需要直接通过后端ECS对外提供服务，或后端ECS有访问外网的需求， 那么需要相应的配置或购买ECS公网IP/弹性公网IP/NAT网关等服务，

ECS的公网流量访问路径如图 1-2: 出网流量路径所示。

图 1-2: 出网流量路径



总体原则：流量从哪里进来，就从哪里出去。

1. 通过负载均衡进入的流量在负载均衡SLB上限速/计费，仅收取出方向流量费用，入方向流量不收取（在未来可能会改变），SLB到ECS之间是阿里云内网通信，不收取流量费用。
2. 来自弹性公网IP/NAT网关的流量，分别在弹性公网IP/NAT网关上进行限速/计费，如果在购买ECS时选择了公网带宽，限速/计费点在ECS上。
3. 负载均衡SLB仅提供被动访问公网的能力，即后端ECS只能在收到通过负载均衡SLB转发来的公网的请求时，才能访问公网回应该请求，如后端ECS希望主动发起公网访问，则需要配置/购买ECS公网带宽、弹性公网IP或NAT网关来实现。
4. ECS公网带宽（购买ECS时配置）、弹性公网IP、NAT网关均可以实现ECS的双向公网访问（访问或被访问），但没有流量分发和负载均衡的能力。

1.1.4 创建实例

在配置负载均衡服务前，您必须先创建一个负载均衡实例。

前提条件

在您创建负载均衡实例前，确保您已经做好了相关规划，详情参考准备工作。

操作步骤

1. 登录负载均衡管理控制台。
2. 在左侧导航栏，单击实例管理，然后单击左上角的创建负载均衡。
3. 根据如下信息，配置负载均衡实例。

配置	说明
地域	选择负载均衡实例的所属地域。 说明: 确保负载均衡实例的地域和后端添加的云服务器ECS的地域相同。
可用区类型	显示所选地域的可用区类型。云产品的可用区指的是一套独立的基础设施，常用数据中心IDC表示。不同的可用区之间具有基础设施（网络、电力、空调等）的独立性，就是说一个可用区的基础设施故障不影响另外一个可用区。可用区是属于某个地域的，一个地域下可能有一个或者多个可用区。负载均衡已经在大部分地域部署了多可用区。 · 单可用区：负载均衡实例只部署在一个可用区上。 · 多可用区：负载均衡实例会部署在两个可用区上。默认启用主可用区的实例。当主可用区出现故障时，将会自动切换到备可用区继续提供负载均衡服务，可以大大提升本地可用性。
主可用区	选择负载均衡实例的主可用区，主可用区是当前承载流量的可用区。
备可用区	选择负载均衡实例的备可用区。备可用区默认不承载流量，主可用区不可用时才承载流量。
实例规格	选择一个性能规格。 不同的性能规格所提供的性能指标也不同，详情查看4. 如何选择性能保障型实例？。
实例类型	根据业务场景选择配置对外公开或对内私有的负载均衡服务，系统会根据您的选择分配公网或私网服务地址。更多详细信息，参考什么是负载均衡实例。 · 公网：公网负载均衡实例仅提供公网IP，可以通过Internet访问负载均衡。 · 私网：私网负载均衡实例仅提供阿里云私网IP，只能通过阿里云内部网络访问该负载均衡服务，无法从Internet访问。
IP版本	选择IPv4。
网络类型	如果您选择的实例类型是私网，您还需要选择该负载均衡实例的网络类型。 · 经典网络：经典网络的负载均衡实例的服务地址由阿里云统一分配和管理。 · 专有网络：专有网络的负载均衡实例的服务地址会从您指定的专有网络的交换机网段内分配。

配置	说明
计费方式	选择一种计费方式。
购买数量	选择购买数量。

4. 单击立即购买，完成支付。

1.1.5 管理实例

在实例管理页面，选择负载均衡实例的所属地域，您可以查看该地域的所有负载均衡实例。此外，您还可以：

- 修改负载均衡实例名称

  将光标移至负载均衡ID区域，单击出现的铅笔图标，输入实例名称。

- 暂停负载均衡实例

  勾选负载均衡实例，单击页面下方的停止，或选择 管理 > 停止 。

- 启动负载均衡实例

  勾选已经停止运行的负载均衡实例，单击页面下方的启动，或选择 管理 > 启动 。

- 释放负载均衡实例

  勾选负载均衡实例， 单击页面下方的释放设置，或选择 管理 > 释放设置 。在释放设置对话框，选择立即释放或在某个特定时刻释放实例。

- 设置标签

  您可以通过标签实例进行分类和统一管理。详情参考标签。

- 变更计费

  单击 更多 > 变更配置 ，修改实例的类型和带宽。

- 查看负载均衡实例详情

  单击负载均衡实例的ID链接或管理，查看负载均衡实例详情。

  - 在详情页面，您可以单击消费明细，查看负载均衡服务的费用明细。
  - 单击监听，查看或添加负载均衡监听。详情参考监听。
  - 单击服务器，查看或添加后端服务器。详情参考服务器。
  - 单击监控， 查看监控信息，设置报警机制。详情参考监控。

1.1.6 管理标签

负载均衡提供标签管理功能，方便您通过标签对负载均衡实例进行分类。

每个标签都由一对键值对组成，负载均衡标签的使用限制如下：

- 目前不支持未绑定实例的空标签存在，标签必须绑定在某个负载均衡实例上。
- 一个实例最多可以绑定10个标签。
- 一个实例上的每个标签的标签键必须唯一，相同标签键的标签会被覆盖。
- 每个地域中的的标签信息不互通，例如在华东1地域创建的标签在华东2地域不可见。

添加标签

完成以下操作，添加标签：

1. 登录负载均衡管理控制台。
2. 在实例管理页面，选择地域，然后勾选需要添加同一标签的实例。
3. 单击编辑标签。

4. 在编辑标签窗口，单击新建，然后输入新建标签的标签键和值，单击确定。

搜索实例

完成以下操作，搜索指定标签关联的实例：

1. 登录负载均衡管理控制台。
2. 在实例管理页面，选择地域，查看该地域的所有实例。
3. 单击请选择标签，然后选择要搜索的实例绑定的标签键和标签值。

   符合您选择条件的实例会显示在实例列表。

4. 单击标签旁边已选的标签键的删除图标，清除标签过滤条件。

删除标签

负载均衡不支持批量删除多个实例的标签，您只能单独对某一个实例进行标签移除。

完成以下操作，删除标签：

1. 登录负载均衡管理控制台。
2. 在实例管理页面，选择地域，查看该地域的所有实例。
3. 选择要删除标签的实例，单击更多 > 编辑标签 。
4. 在编辑标签窗口，单击要移除的标签的删除图标，然后单击确定。

> 说明：
> 当一个标签从一个实例上移除后，如果该标签没有和其他实例绑定，系统会将该标签删除。

1.1.7 绑定EIP

您可以为专有网络类型的SLB实例绑定一个EIP。绑定后，SLB实例便可以转发来自公网的请求。

操作步骤

1. 登录负载均衡管理控制台。
2. 在左侧导航栏，单击实例管理。
3. 选择负载均衡实例的地域，找到目标实例。

> 说明：
> 确保负载均衡实例的网络类型为专有网络。

4. 单击更多 > 绑定EIP。
5. 选择一个EIP，然后单击确认。

1.1.8 按量付费实例变配

您可以更改后付费实例的带宽、实例规格和计费方式（按流量计费或按带宽计费）。

背景信息

在变更前，请注意：

- 性能共享型实例变更为性能保障型实例会有10-30秒业务中断，其他变配操作均不会影响业务，也不会变更负载均衡实例IP。建议在业务低谷期进行此类变配。
- 将性能共享型实例变更为性能保障型实例后，无法再将其变更回性能共享型。

  您可以选择使用简约型I (slb.s1.small)规格，该规格免收规格费。

操作步骤

1. 登录负载均衡管理控制台。
2. 选择目标实例的所属地域。

3. 找到目标实例，然后单击更多 > 变更配置。

4. 在配置变更区域，选择新的带宽值、实例规格或计费方式后，完成支付。

1.2 监听

1.2.1 监听介绍

创建负载均衡实例后，您需要为实例配置监听。负载均衡实例监听负责检查连接请求，然后根据调度算法定义的转发策略将请求流量分发至后端服务器。

负载均衡监听包括监听配置和健康检查配置。

监听配置

负载均衡提供四层（TCP/UDP协议）和七层（HTTP/HTTPS协议）监听，您可根据应用场景选择监听协议：

协议	说明	使用场景
TCP	· 面向连接的协议，在正式收发数据前，必须和对方建立可靠的连接 · 基于源地址的会话保持 · 在网络层可直接看到来源地址 · 数据传输快	· 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 · 无特殊要求的Web应用

协议	说明	使用场景
UDP	· 面向非连接的协议，在数据发送前不与对方进行三次握手，直接进行数据包发送，不提供差错恢复和数据重传 · 可靠性相对低；数据传输快	关注实时性而相对不注重可靠性的场景，如视频聊天、金融实时行情推送
HTTP	· 应用层协议，主要解决如何包装数据 · 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、小的手机游戏等
HTTPS	· 加密传输数据，可以阻止未经授权的访问 · 统一的证书管理服务，用户可以将证书上传到负载均衡，解密操作直接在负载均衡上完成	需要加密传输的应用

说明:

负载均衡已在全部地域支持HTTP/2和WSS/WS协议，详情参见HTTP/2协议支持常见问题和WS/WSS协议支持常见问题。

健康检查配置

负载均衡对后端服务器提供健康检查，提高服务的可用性。详情参见健康检查原理和健康检查配置。

1.2.2 四层监听

1.2.3 七层监听

1.2.4 健康检查

1.2.5 共享实例带宽

负载均衡支持按带宽计费的负载均衡实例下的所有监听共享实例的总带宽。在创建监听时，您可以设置带宽峰值也可以选择不设置。

- **配置：您可以对监听的带宽进行限制，但所有监听带宽峰值的总和不能超过实例的带宽峰值。**

- 不限制：不限制带宽的情况下，实例下的监听共享实例带宽。

如何共享带宽?

假如您购买了一个带宽峰值为 10MB 的负载均衡实例，并在该实例下创建了三个监听（监听A、监听B和监听C）。监听A的带宽峰值设置为 4MB，另外两个监听没有设置带宽峰值。三个监听的带宽使用可能出现如下几种情况：

- 如果监听A和监听C一直没有出流量，那么监听B最多也只能跑满剩余的 6MB 带宽（10MB - 4MB)。
- 如果监听C一直没有出流量，而监听B的出流量很大，超过了剩余的 6MB 带宽。此时，监听B已经产生丢包，而监听A只有 4MB 的出流量，没有超过设置的带宽峰值，所以不会产生丢包。
- 如果监听A一直是满速在跑（监听峰值 4MB），而后监听B和监听C也有出流量并且两个监听的流量很大，那么监听B和监听C就会共享（竞争）剩余的 6MB 带宽。此时，监听A的流量不会受监听B和监听C的影响，始终能达到预留的 4MB 峰值；如果监听B和监听C出流量同等大小，两个监听占用的带宽去会趋近于均分。

因此，对监听带宽的限制值是资源预留，这是为了保证核心的业务始终有足够的带宽。非核心的业务可以不设置监听带宽值，它们竞争实例剩余的带宽资源。

1.3 后端服务器

1.3.1 后端服务器概述

在使用负载均衡服务前，您需要添加ECS实例作为负载均衡实例的后端服务器，用来接收负载均衡监听转发的请求。

您可以在任意时刻增加或减少负载均衡实例的后端ECS数量，还可以在不同ECS实例之间进行切换。但是为了保证您对外服务的稳定性，确保在执行上述操作时，开启了负载均衡的健康检查功能并同时保证负载均衡实例中至少有一台正常运行的ECS。

负载均衡服务通过设置虚拟服务地址，将添加的同一地域的多台ECS实例虚拟成一个高性能、高可用的应用服务池。默认后端服务器是在实例维度上维护的，即负载均衡实例下的所有监听都只能够将流量转发到相同服务器的相同端口上。

您也可以通过服务器组的方式添加ECS。不同的监听可以关联不同的服务器组，这样一个负载均衡实例的不同监听就可以将请求转发给不同的服务器组内不同端口的ECS。

> 说明:
> 如果您在配置监听时，选择使用服务器组，那么该监听会将请求转发到关联的服务器组中的ECS，而不会再将请求转发给后端服务器池中的ECS。

默认服务器组

用来接收前端请求的ECS实例。如果监听没有设置虚拟服务器组或主备服务器组，默认将请求转发至默认服务器中的ECS。

参见添加默认服务器创建一个默认服务器组。

主备服务器组

当您有传统的主备需求时，即后端服务器中有一台主机和一台备机。当主机工作正常时，流量将直接走主机；当主机宕机时，流量将走到备机。此时，可以使用主备服务器组，避免服务中断。

由于备机不会做健康检查，所以只要主机健康检查失败，系统会直接将流量切到备机。当主机健康检查成功恢复服务后，流量会自动切到主机。

主备服务器组是在监听维度上维护的，并且只支持四层监听，详情参考创建主备服务器组。

虚拟服务器组

当您需要将不同的请求转发到不同的后端服务器上时，或需要通过域名和URL进行请求转发时，可以选择使用虚拟服务器组。详情参考创建虚拟服务器组。

注意事项

添加后端服务器时，注意：

- 负载均衡不支持跨地域部署，确保ECS实例的所属地域和负载均衡实例的所属地域相同。
- 负载均衡本身不会限制后端ECS实例使用哪种操作系统，只要您的两台ECS实例中的应用服务部署是相同的且保证数据的一致性即可。建议您选择相同操作系统的ECS实例作为后端服务器，以便日后管理和维护。
- 一个负载均衡实例最多支持添加50个监听，每个监听对应后端ECS实例上的一个应用。负载均衡监听的前端端口对应后端ECS实例上的应用服务端口。
- 您可以指定后端服务器池内各ECS实例的转发权重。权重越高的ECS实例将被分配到更多的访问请求，您可以根据后端ECS实例的对外服务能力和情况来区别设定。

  说明：
  如果您同时开启了会话保持功能，那么有可能会造成对后端应用服务器的访问并不是完全相同的。如果出现了访问不均衡的情况，建议您可以暂时关闭会话保持功能，观察一下是否依然存在这种情况。

  当负载均衡服务分发请求不均匀时，可以参考以下方法检查处理：

  1. 统计一个时间段内，后端ECS实例的Web服务访问日志记录数据量。
  2. 按照负载均衡的配置，对比多台ECS实例日志的数量是否有相差。（如设置会话保持，需要剥离相同IP的访问日志。如果负载均衡配置了权重，要根据权重比例计算日志中访问比例是否正常。）
- ECS进行热迁移时，可能导致SLB长连接断开，重连即可恢复，请做好应用的重连工作。

1.3.2 添加默认服务器

在使用负载均衡服务前，必须至少添加一台默认服务器组。

前提条件

- 您已创建负载均衡实例。
- 您已创建了ECS实例并部署了相关应用，用来接收转发的请求。

操作步骤

1. 登录负载均衡管理控制台。

2. 在实例管理页面，选择目标实例的所属地域。
3. 单击目标实例的ID。
4. 选择默认服务器 。
5. 单击添加。
6. 在未添加的服务器页面，单击添加。

   系统显示我的服务器页面。
7. 单击目标ECS实例对应的添加，或者勾选多个目标ECS实例，然后单击页面下方的批量添加。
8. 在添加后端服务器对话框，指定添加的ECS实例的权重，然后单击确定。

   权重越高的ECS实例将被分配到更多的访问请求。您可以根据后端ECS实例的对外服务能力和情况来区别设定。

   > 注意:
   > 权重设置为0，该服务器不会再接受新请求。

   添加后的实例会显示在已添加的服务器页签下，您可以移除或者修改添加的ECS实例的权重。

1.3.3 创建虚拟服务器组

虚拟服务器组 （VServer group）是一组 ECS 实例。虚拟服务器组允许您在监听维度上个性化定义服务器组，即实例下的不同监听可使用不同的后端服务器组。

前提条件

- 您已经创建负载均衡实例。
- 您创建了 ECS 实例，用来接收转发的请求。

背景信息

将虚拟服务器组和一个监听关联后，监听只会将流量转发给关联的虚拟服务器组的后端服务器，不会再将流量转发给其他后端服务器。

如果您在您同时为一个七层监听添加了后端服务器、虚拟服务器组和转发规则，请求转发的顺序如下:

- 判断请求其是否能够匹配上某条转发规则，如果匹配，则将流量转发到该规则的虚拟服务器组。
- 若不匹配并且在该监听上设置了虚拟服务器组，那么将流量转发到监听关联的虚拟服务器组。
- 若您没有在该监听上设置虚拟服务器组，即将流量转发到实例级别添加的各后端服务器。

在创建虚拟服务器组时，请注意:

- 虚拟服务器组只能添加监听所在地域的后端服务器。

- 一个后端服务器可以属于多个虚拟服务器组。
- 一个虚拟服务器组可绑定在一个实例的多个监听上。
- 虚拟服务器组由服务器加端口组成。

操作步骤

1. 登录 负载均衡管理控制台。
2. 在实例管理页面，选择地域，查看该地域下的所有负载均衡实例。
3. 单击负载均衡实例的 ID 链接，进入负载均衡的详情页面。
4. 在负载均衡实例菜单栏，选择虚拟服务器组 。
5. 在 虚拟服务器组页面，单击 创建虚拟服务器组。
6. 在 创建虚拟服务器组 对话框，完成如下操作：
   a) 在 虚拟服务器组名称 文本框中，输入虚拟服务器组名称。
   b) 单击添加，在我的服务器列表选择要添加的服务器。
7. 在已添加服务器 ，输入每个 ECS 实例的端口和权重，单击 确定。
   创建的虚拟服务器组会显示在 虚拟服务器组 页面，您可以删除或者增加虚拟服务器的 ECS 实例数量 （单击 编辑）。您也可以将该虚拟服务器组和实例下的监听或者转发策略关联。

1.3.4 创建主备服务器组

当您有传统的主备需求时，即后端服务器中有一台主机和一台备机，可选择使用主备服务器组。当主机正常工作时，流量将直接走主机；当主机不可用时，流量将走到备机，避免服务中断。

前提条件

- 您已创建负载均衡实例。
- 您已创建了ECS实例并部署了相关应用，用来接收转发的请求。

背景信息

将主备服务器组和一个监听关联后，监听只会将流量转发给关联的主备服务器组中的后端服务器，不会再将流量转发给其他后端服务器。

> 注意:
> 主备服务器组只支持四层监听（TCP和UDP协议）。

操作步骤

1. 登录负载均衡管理控制台。
2. 在实例管理页面，选择目标实例的所属地域。
3. 单击目标实例的ID。

4. 选择主备服务器页签。
5. 在主备服务器组页面，单击创建主备服务器组。
6. 在创建主备服务器组对话框，完成如下操作：
   a) 在主备服务器组名称文本框中，输入主备服务器组名称。
   b) 单击添加，在我的服务器列表中，选择目标ECS实例。
   c) 在已添加服务器中，输入ECS实例的端口和权重，并选择作为主机使用的ECS实例，单击确定。

1.4 DDoS基础防护

负载均衡控制台可以查看公网负载均衡实例的云盾阈值。

DDoS基础防护介绍

阿里云免费为负载均衡服务提供最高5G的DDoS基础防护。如下图所示，所有来自Internet的流量都要先经过云盾再到达负载均衡，云盾会针对常见的攻击进行清洗过滤。云盾DDoS基础防护可以防御SYN Flood、UDP Flood、ACK Flood、ICMP Flood 和DNS Flood等DDoS攻击。

云盾DDoS基础防护根据公网负载均衡实例的带宽设定清洗阈值和黑洞阈值。当入方向流量达到阈值上限时，触发清洗和黑洞：

- 清洗：当来自Internet的攻击流量较大或符合某些特定攻击流量模型特征时，云盾将会针攻击流量启动清洗操作，清洗包括攻击报文过滤、流量限速、包限速等。
- 黑洞：当来自Internet的攻击流量非常大时，为保护整个集群的安全，流量将会被黑洞处理，即所有入流量全部被丢弃。

查看防护阈值

使用子账号登录阿里云控制台后，若无法在控制台上查看防护阈值，需要先对子账号授权。详情参见授予云盾基础防护只读权限。

完成以下操作查看防护阈值：

1. 登录负载均衡管理控制台。
2. 选择地域，查看该地域的所有实例。

3. 将鼠标移至目标实例的云盾图标，查看BPS清洗阈值、PPS清洗阈值和黑洞阈值。您可以单击DDoS控制台链接查看更多信息。
   - BPS清洗阈值：入方向流量超过了BPS清洗阈值时，触发清洗。
   - PPS清洗阈值：入方向数据包数超过了PPS清洗阈值时，触发清洗。
   - 黑洞阈值：入方向流量超过黑洞阈值时将触发黑洞。

授权云盾基础防护只读权限

完成以下操作授予子账号只读访问云盾DDoS基础防护(Anti-DDoS Basic)的权限：

> 说明:
> 使用主账号进行授权。

1. 使用主账号登录访问控制RAM管理控制台。
2. 在左侧导航栏，单击用户管理，找到目标子账号，然后单击管理。
3. 单击用户授权策略，然后单击编辑授权策略。
4. 在弹出的对话框，在可授权策略列表中搜索AliyunYundunDDosReadOnlyAccess，将其加入到已授权策略列表。单击确定。

1.5 各地域带宽峰值限制

负载均衡各地域的包年包月可售卖最大带宽和按流量实例带宽峰值如下表所示。

> 说明:
> 所有地域的私网带宽峰值都为5 Gbps。

地域	带宽峰值
华北 1 （青岛）	5 Gbps
华东 1（杭州）	5 Gbps
华北 2（北京）	5 Gbps
华东 2（上海）	5 Gbps
华南 1（深圳）	5 Gbps
华北 3（张家口）	5 Gbps

地域	带宽峰值
华北 5（呼和浩特）	5 Gbps
香港	2 Gbps
美国东部 1（弗吉尼亚）	1 Gbps
美国西部 1 （硅谷）	2 Gbps
亚太东北 1（东京）	1 Gbps
亚太东南 1（新加坡）	5 Gbps
亚太东南 2（悉尼）	1 Gbps
亚太东南 3（吉隆坡）	5 Gbps
中东东部 1（迪拜）	500 Mbps
欧洲中部 1（法兰克福）	1 Gbps
亚太南部 1 (孟买)	5 Gbps

1.6 多可用区

在创建负载均衡实例时，您可以选择将负载均衡创建在支持多可用区的地域，提高服务的可用性。

什么是多可用区

云产品的可用区指的是一套独立的基础设施，不同的可用区之间基础设施（网络，电力和空调等）相互独立，即一个可用区出现基础设施故障不影响另外一个可用区。

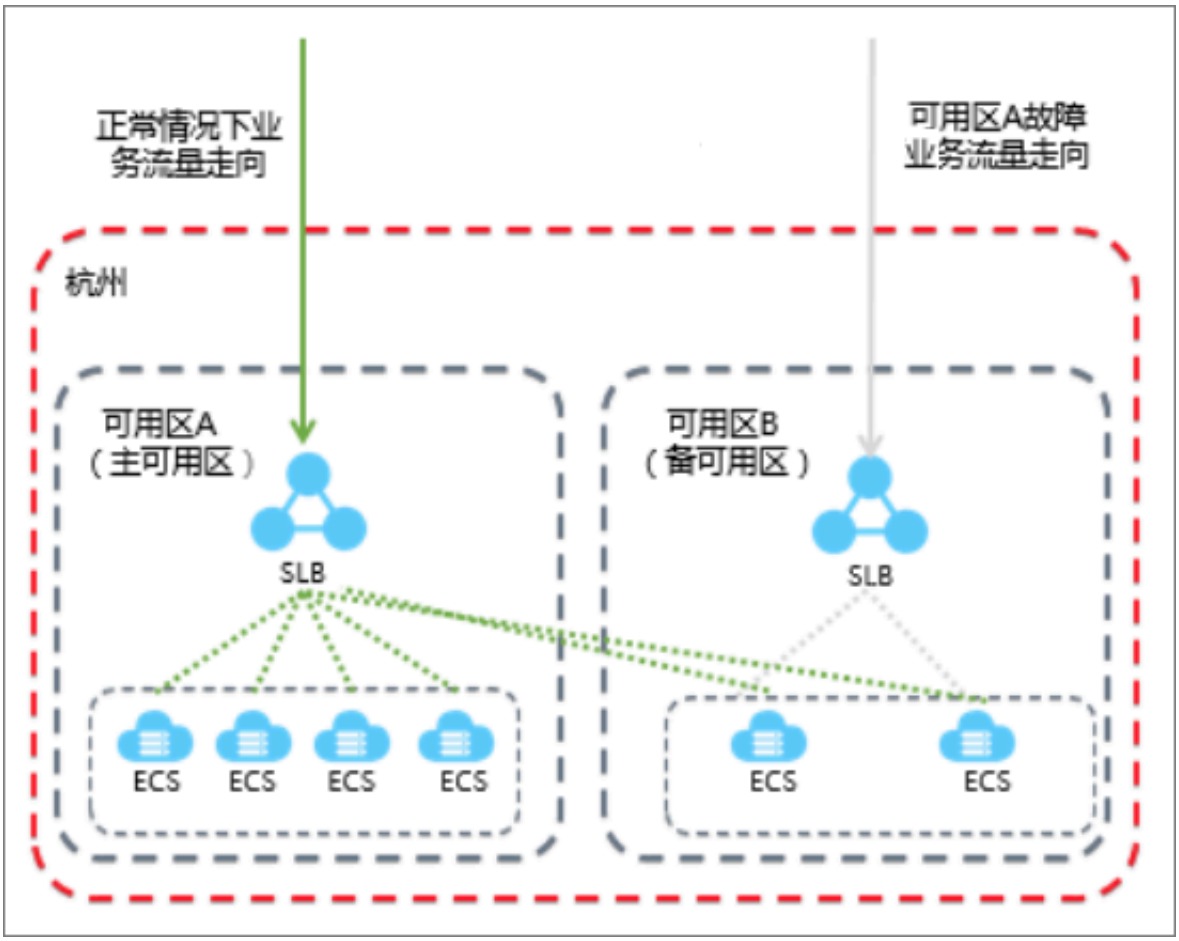
为了向广大用户提供更加稳定可靠的负载均衡服务，阿里云负载均衡已在各地域（Region）部署了多可用区以实现同地域下的跨机房容灾。当主可用区的机房故障或不可用时，负载均衡仍然有能力在非常短的时间内（约30秒）切换到另外一个备可用区的机房并恢复服务的能力；当主可用区恢复时，负载均衡同样会自动切换到主可用区的机房提供服务。

关于负载均衡主备可用区，请注意：

- SLB支持跨可用区挂载后端ECS，即只要ECS和SLB实例在同一个地域即可。SLB可以同时将流量分发至不同可用区的ECS上。
- 正常情况下，备可用区的SLB实例处于待机状态。您不可以手动切换SLB实例的主备工作状态，只有当阿里云检测到整个可用区不可用时如如机房整体断电、机房出口光缆中断等，负载均衡才会切换到备可用区。而并非某个实例出现故障，就切换到备可用区。
- SLB和ECS是不同的集群。可用区A的SLB不可用时，ECS并不一定不可用，因此如果仅因为SLB集群故障导致的SLB主备倒换，备可用区的SLB依然可以将流量分发至不同可用区的ECS。

但当整个可用区的所有集群断电或光缆中断时，那么可用区的所有服务（包括但不限于SLB、ECS等）就都无法正常工作了。

更多信息，参见负载均衡高可用最佳实践。



主备可用区列表

下表列举了各地域的主备可用区，您也可以通过DescribeZones接口查询可用的主备可用区。

地域	可用区类型	可用区	
华东1	多可用区	主可用区	备可用区
		可用区B	可用区D
		可用区D	可用区E
		可用区E	可用区F
		可用区F	可用区E
华东2	多可用区	主可用区	备可用区
		可用区A	可用区B

地域	可用区类型	可用区	
		可用区B	可用区A或可用区D
		可用区C	可用区B
		可用区D	可用区B
华南1	多可用区	主可用区	备可用区
		可用区A	可用区B
		可用区B	可用区A
		可用区C	可用区B
华北1	多可用区	主可用区	备可用区
		可用区B	可用区C
		可用区C	可用区B
华北2	多可用区	主可用区	备可用区
		可用区A	可用区B或可用区D
		可用区B	可用区A或可用区C
		可用区C	可用区B
		可用区D	可用区A
华北3	单可用区	可用区A	可用区A
华北5	单可用区	可用区A	可用区A
欧洲中部1	单可用区	可用区A	可用区A
中东东部1	单可用区	可用区A	可用区A
新加坡	多可用区	主可用区	备可用区
		可用区A	可用区B
		可用区B	可用区A
亚太东南2	单可用区	可用区A	可用区A
亚太东南3	单可用区	可用区A	可用区A
亚太东北1	单可用区	可用区A	可用区A
香港	多可用区	主可用区	备可用区
		可用区B	可用区C
		可用区C	可用区B
美东1	单可用区	可用区A	可用区A

地域	可用区类型	可用区	
美西1	多可用区	主可用区	备可用区
		可用区A	可用区B
		可用区B	可用区A

1.7 结合云解析实现跨地域负载均衡

全球负载均衡

负载均衡从其应用的地理结构上分为本地负载均衡和全局负载均衡。本地负载均衡是指对同地域的服务器群做负载均衡，全局负载均衡是指对分别部署在不同地域有不同网络结构的服务器群做负载均衡。

结合云解析（全球负载均衡版），您可在本地负载均衡上层部署云解析（全球负载均衡版），实现跨地域容灾和智能解析。

- 多线路智能化解析服务

  云解析（全球负载均衡版）支持多线智能解析，即根据网站访问者所在的网络类型，智能判断提供最佳的访问解析地址，使访问用户获得最快捷、最流畅的体验。

  如下图所示，当判断访问者来源为海外用户，则将域名解析到位于美国的负载均衡实例上；当判断访问者来源为国内用户，则将域名解析到到国内的负载均衡实例上。
- 跨地域容灾

  云解析提供监控服务，监控主机记录下的IP地址。根据网站监控探测结果，实时隔离故障集群，流量动态切换，实现跨地域容灾。

部署全球负载均衡

本操作以一个域名为aliyuntest.club的网站为例（该网站的大多数用户分布在新加坡和国内），指导您如何通过云解析和负载均衡实现全局负载均衡均衡。

步骤一 购买与配置云服务器

根据您的应用服务的用户的地域分布，在相应地域下购买并配置至少两台ECS。

本操作中，在北京、深圳、新加坡这三个地域分别购买了两台ECS，并在ECS上搭建了一个简单的静态网页。

- 北京地域ECS示例
- 深圳地域ECS示例

- 新加坡地域ECS示例

步骤二 购买与配置负载均衡实例

1. 参考创建负载均衡实例分别在北京、深圳、新加坡创建三个公网负载均衡实例。
2. 参考配置负载均衡实例添加监听并将各个地域下配置好的ECS添加到后端服务器池。

- 北京地域负载均衡实例示例
- 深圳地域负载均衡实例示例
- 新加坡地域负载均衡实例示例

步骤三 配置全球负载均衡

1. 购买云解析全球负载均衡版。
   a. 在全球负载均衡页面，单击立即购买。
   b. 配置云解析全球负载均衡。

      本操作中，除以下配置外，其它配置使用默认选项：

      - 配置类型，选择全球负载均衡版。
      - 全球负载均衡类型，选择线路间负载均衡。
      - 绑定您的网站域名。您也可以在创建后绑定域名。

   c. 单击立即购买。
   d. 在云解析DNS管理控制台的VIP产品页面，查看创建的全球负载均衡版云解析实例。
2. 设置智能解析。
   a. 登录域名服务管理控制台。
   b. 找到云解析实例绑定的域名，然后单击其操作列下的解析设置。
   c. 设置域名解析。添加三条A记录分别指向创建的负载均衡实例的公网IP，并将新加坡地域的A记录解析线路设置为海外。
3. 添加网站监控。
   a. 在解析设置页面，单击网站监控。
   b. 单击添加监控，将三条A记录的IP地址添加到监控列表中。
   c. 单击每条监控记录对应的设置选项，修改监控配置。

      将域名记录无法访问时的切换规则设置为强制暂停该记录解析。

4. 开启全球负载均衡。

   a. 在解析设置页面，单击全球负载均衡。

   b. 单击线路间负载均衡页签，选择目标域名，然后单击开启。

   c. 在弹出的对话框中，单击确定。

步骤四 测试

移除北京地域负载均衡实例的后端服务器，使该负载均衡实例的服务不可用。

访问该网站，查看访问是否正常。

说明:

云解析DNS监控到您的IP宕机后需要1-2分钟聚合判断，假设您的监控频率设置是1分钟，那么线路异常切换生效时间在2-3分钟内；如果您的监控频率设置是10分钟，那么线路异常切换生效时间在12-13分钟内。