**阿里云云计算专业认证考试（ACP）（ECS）**

什么是云服务器 ECS

云服务器 Elastic Compute Service（ECS）是阿里云提供的一种基础云计算服务。使用云服务器 ECS 就像使用水、电、煤气等资源一样便捷、高效。您无需提前采购硬件设备，而是根据业务需要，随时创建所需数量的云服务器实例，并在使用过程中，随着业务的扩展，对云服务器进行扩容磁盘、增加带宽。如果不再需要云服务器，也可以方便的释放资源，节省费用。

**优势**

云服务器 ECS 拥有传统服务器和虚拟主机无法企及的优势：

* + 稳定性： 服务可用性高达 99.95%，数据可靠性高达 99.9999999%。支持宕机迁移、数据快照备份和回滚、系统性能报警。
  + 容灾备份： 每份数据多份副本，单份损坏可在短时间内快速恢复。
  + 安全性： 支持配置安全组规则、云盾防 DDOS 系统、多用户隔离、防密码破解。
  + 多线接入： 基于边界网关协议（Border Gateway Protocol，BGP）的最优路由算法。BGP 多线机房，全国访问流畅均衡。骨干机房，出口带宽大，独享带宽。
  + 弹性扩容： 10 分钟内可启动或释放 100 台 ECS 实例；支持在线不停机升级带宽；5 分钟内停机升级 CPU 和内存。
  + 成本低：无需一次性大投入，按需购买，弹性付费，灵活应对业务变化。
  + 可控性： 作为 ECS 的用户，您拥有超级管理员的权限，能够完全控制 ECS 实例的操作系统，可以通过管理终端自助解决系统问题，并可以进行部署环境、安装软件等操作。
  + 易用性： 丰富的操作系统和应用软件，使用镜像可一键简单部署同一镜像；可在多台 ECS 实例中快速复制环境，轻松扩展；支持自定义镜像、磁盘盘快照批量创建 ECS 实例。
  + API 接口： 使用 ECS API 调用管理，通过安全组功能对一台或多台 ECS 实例进行访问设置，使开发使用更加方便。

**功能**

云服务器 ECS 提供以下功能：

* + 14 大地域中创建实例，有的地域提供多个可用区。
  + 面向多种业务场景、数十种实例规格，从 “1 核 1 GB” 到 “56 核 480 GB”，满足各种不同需求。
  + 3 种数据存储盘（普通云盘、SSD 云盘、高效云盘），并提供 I/O 优化实例。
  + 2 种 IP 地址：公网 IP 和私网 IP，实现内网互联，并能访问 Internet。
  + 2 种网络类型：经典网络和专有网络，在不同维度管理您的网络。
  + 支持多种版本的 Windows 和 Linux 操作系统。
  + 免费开通云盾并提供云监控服务。
  + 丰富的镜像资源，支持公共镜像、自定义镜像、共享镜像和镜像市场，让您免安装，并快速部署操作系统和应用软件。
  + 提供控制台、远程终端和 API 等多种管理方式，给您完全管理权限。
  + 灵活的付费方式：包月包年和按量收费。

云计算的高可用性

阿里云为您提供了如下三项支持：

* + 提升可用性的产品和服务，包括云服务器、负载均衡、多备份数据库服务以及数据迁移服务DTS等。
  + 行业合作伙伴以及生态合作伙伴，帮助您完成更高、更稳定的架构，并且保证服务的永续性。
  + 多种多样的的培训服务，让您从业务端到底层的基础服务端，在整条链路上实现高可用。

云计算的安全性

* + **在阿里云专有网络之上，可以产生更多的业务可能性。**您只需进行简单配置，就可在自己的业务环境下，与全球所有机房进行串接，从而提高了业务的灵活性、稳定性以及业务的可发展性。
  + **对于原来拥有自建的IDC机房，也不会产生问题。**阿里云专有网络可以拉专线到原有的IDC机房，形成混合云的架构。阿里云可以提供各种混合云的解决方案和非常多的网络产品，形成强大的网络功能，让您的业务更加灵活。结合阿里云的生态，您可以在云上发展出意想不到的业务生态。
  + **阿里云专有网络更加稳定和安全。  
    稳定性：**业务搭建在专有网络上，而网络的基础设施将会不停进化，使您每天都拥有更新的网络架构以及更新的网络功能，使得您的业务永远保持在一个稳定的状态。专有网络允许您自由地分割、配置和管理自己的网络。  
    **安全性：**面对互联网上不断的攻击流量，专有网络天然就具备流量隔离以及攻击隔离的功能。业务搭建在专有网络上后，专有网络会为业务筑起第一道防线。
  + 云计算的弹性

**计算弹性**

**纵向的弹性，即单个服务器的配置变更。**传统IDC模式下，很难做到对单个服务器进行变更配置。而对于阿里云，当您购买了云服务器或者存储的容量后，可以根据业务量的增长或者减少自由变更自己的配置。关于纵向弹性的具体应用，详情请参考[升降配](https://help.aliyun.com/document_detail/25437.html?spm=5176.doc25367.6.649.FMIUkj)。

**横向的弹性。**对于游戏应用或直播平台出现的高峰期，若在传统的IDC模式下，您根本无法立即准备资源；而云计算却可以使用弹性的方式帮助客户度过这样的高峰。当业务高峰消失时，您可以将多余的资源释放掉，以减少业务成本的开支。利用横向的扩展和缩减，配合阿里云的弹性伸缩，完全可以做到定时定量的伸缩，或者按照业务的负载进行伸缩。关于横向弹性的具体应用，详情请参考[弹性伸缩](https://help.aliyun.com/document_detail/25857.html?spm=5176.product25855.6.539.D0ePwz)。

**存储弹性**

阿里云拥有很强的存储弹性。当存储量增多时，对于传统的IDC方案，您只能不断去增加服务器，而这样扩展的服务器数量是有限的。在云计算模式下，将为您提供海量的存储，当您需要时可以直接购买，为存储提供最大保障。关于存储弹性的具体应用，详情请参考[磁盘扩容](https://help.aliyun.com/document_detail/35095.html?spm=5176.product25365.6.668.ygZKDu)。

**网络弹性**

云上的网络也具有非常大的灵活性。只要您购买了阿里云的专有网络，那么所有的网络配置与线下IDC机房配置可以是完全相同的，并且可以拥有更多的可能性。可以实现各个机房之间的互联互通，各个机房之间的安全域隔离，对于专有网络内所有的网络配置和规划都会非常灵活。关于网络弹性的具体应用，详情请参考[专有网络](https://help.aliyun.com/product/27706.html)。

地域和可用区

**地域**

地域是指物理的数据中心。当前可选的地域、地域所在城市和 Region ID 的对照关系如以下表格所示。您也可以通过 API 接口 [DescribeRegions](https://help.aliyun.com/document_detail/25609.html) 查看完整的地域列表。

资源创建好后，不能更换地域。

* + **中国大陆**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **地域名称** | **华北 1** | **华北 2** | **华北 3** | **华东 1** | **华东 2** | **华南 1** |
| **所在城市** | 青岛 | 北京 | 张家口 | 杭州 | 上海 | 深圳 |
| **RegionId** | cn-qingdao | cn-beijing | cn-zhangjiakou | cn-hangzhou | cn-shanghai | cn-shenzhen |

* + **其他国家和地区**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **地域名称** | **香港** | **亚太东南 1** | **亚太东南 2** | **亚太东北 1** | **美国西部 1** | **美国东部 1** | **欧洲中部 1** | **中东东部 1** |
| **所在城市** | 香港 | 新加坡 | 悉尼 | 东京 | 硅谷 | 弗吉尼亚 | 法兰克福 | 迪拜 |
| **RegionId** | cn-hongkong | ap-southeast-1 | ap-southeast-2 | ap-northeast-1 | us-west-1 | us-east-1 | eu-central-1 | me-east-1 |

* + **华东 1、华东 2、华北 1 、华北 2、华北 3、华南 1 地域**提供多线 BGP 骨干网线路，网络能力覆盖中国大陆各省市，实现稳定高速中国大陆地域访问。
  + **香港地域**提供国际带宽访问，覆盖香港、东南亚等地域。
  + **亚太东南 1 地域**合作伙伴为 SingTel（新加坡电信），是东南亚优势运营商，公司业务能力、成熟度各方面非常可靠，可以满足周边用户需求。
  + **美国西部 1 地域**位于美国西部硅谷，通过 BGP 线路直接接连多家美国运营商骨干网，可覆盖全美，同时可以很好地辐射南美洲和欧洲大陆。

**可用区（Zone）**

可用区是指在同一地域内，电力和网络互相独立的物理区域。同一可用区内的 ECS 实例网络延时更小。

在同一地域内可用区与可用区之间内网互通，可用区之间能做到故障隔离。是否将云服务器 ECS 实例放在同一可用区内，主要取决于对容灾能力和网络延时的要求。

* + 如果您的应用需要较高的容灾能力，建议您将 ECS 实例部署在同一地域的不同可用区内。
  + 如果您的应用在实例之间需要较低的网络时延，则建议您将 ECS 实例创建在相同的可用区内。

**服务器 ECS 与其他阿里云产品之间的关系**

如果您的 ECS 实例需要与其他阿里云产品一起搭配使用，需要注意：

* + 不同地域的云服务器 ECS、关系型数据库 RDS、对象存储服务 OSS 内网不互通。
  + 不同地域之间的云服务器 ECS 不能跨地域部署负载均衡，即在不同的地域购买的 ECS 实例不支持跨地域部署在同一负载均衡实例下。
  + 实例
  + 一个云服务器 ECS 实例等同于一台虚拟机，包含 CPU、内存、操作系统、网络、磁盘等最基础的计算组件。您可以方便的定制、更改实例的配置。您对该虚拟机拥有完全的控制权，和您本地服务器的区别在于，您只需要登录到阿里云，即可使用云服务器，进行独立的管理、顶级配置等操作。
  + 实例生命周期

**实例固有状态**

在这个生命周期中，实例有其固有的几个状态，如下表：

|  |  |  |  |
| --- | --- | --- | --- |
| **状态** | **状态属性** | **解释** | **API 的对应状态** |
| 准备中 | 中间状态 | 实例创建后，在进入运行中之前的状态，如果长时间处于该状态，则说明出现异常。 | Pending |
| 已创建 | 稳定状态 | 实例已经创建完成，等待启动。 | Stopped |
| 启动中 | 中间状态 | 实例在控制台或通过 API，重启、启动等操作后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。 | Starting |
| 运行中 | 稳定状态 | 实例正常运行状态，在这个状态的实例可以上运行您的业务。 | Running |
| 停止中 | 中间状态 | 实例在控制台或通过 API，停止操作后，在进入已停止之前的状态，如果长时间处于该状态，则说明出现异常。 | Stopping |
| 已停止 | 稳定状态 | 实例被正常停止，在这个状态下的实例，不能对外提供业务。 | Stopped |
| 重新初始化中 | 中间状态 | 实例在控制台或通过 API，重新初始化系统盘或数据盘后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。 | Stopped |
| 更换系统盘中 | 中间状态 | 实例在控制台或通过 API，更换操作系统等操作后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。 | Stopped |
| 已过期 | 稳定状态 | 包年包月的实例到期后因您没有及时为其续费；按量付费的实例因您的账户欠费以上情况会让实例变成已过期状态。这个状态下的实例处于停止状态，不能对外提供业务。从停机时刻起数据保留 7 天，然后实例会被释放。 | Stopped |

部署集

* + **提升业务可靠性**   
    您希望相同的应用实例尽量不要分布在同一台物理宿主机、机架或交换机上，以避免物理宿主机、机架或交换机发生故障时，对业务造成较大的影响。
  + **提升业务网络性能**   
    在一些业务场景中，实例间会有较多网络交互，您希望这些实例间网络访问延时尽可能低，或者网络访问带宽尽可能高。在这种场景下，您希望相应实例尽可能集合到同一交换机下，以减少网络延时和保障网络带宽。
  + 部署集（Deployment Set）是阿里云 ECS 提供的一种能让您感知宿主机、机架、交换机物理拓扑的能力，并且让您可以根据自己的业务类型选择符合业务要求的部署策略，提升业务整体可靠性或性能等。

Deployment Set 部署策略包括：

* + 部署策略：包括尽量聚合、严格聚合、尽量打散和严格打散等四种策略，其中，尽量聚合和严格聚合是适用更高的性能，尽量打散和严格打散是为了更高的可用性。
  + 部署域：包括宿主机、机架和交换机等三种级别，其中，宿主机和机架支持尽量打散和严格打散两种部署策略，交换机支持四种部署策略。
  + 云盘参数和性能测试方法
  + 普通云盘：面向低 I/O 负载的应用场景，为 ECS 实例提供数百 IOPS 的 I/O 性能。
  + 高效云盘：面向中度 I/O 负载的应用，为 ECS 实例提供最高 3000 随机 IOPS 的存储性能。
  + SSD 云盘：面向 I/O 密集型应用，提供稳定的高随机 IOPS 性能。

**云盘参数对比**

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **SSD 云盘** | **高效云盘** | **普通云盘** |
| 最大容量 | 32768 GB | 32768 GB | 2000 GB |
| 最大 IOPS | 20000 \* | 3000 \* | 数百 |
| 最大吞吐量 | 300 MBps | 80 MBps | 30 MBps |
| 性能计算公式 | IOPS=min{1200+30\*容量, 20000}  吞吐量=min{80+0.5\*容量, 300}MBps | IOPS=min{1000+6\*容量, 3000}  吞吐量=min{50+0.1\*容量, 80}MBps | 不适用 |
| 访问时延 | 0.5−2 ms | 1−3 ms | 5−10 ms |
| 数据可靠性 | 99.9999999% | 99.9999999% | 99.9999999% |
| API名称 | cloud\_ssd | cloud\_efficiency | cloud |
| 价格\*\* | 1.0 元/GB/月 | 0.35 元/GB/月 | 0.3 元/GB/月 |
| 典型应用场景 | * + - I/O 密集型应用     - 中大型关系型数据库     - NoSQL 数据库 | * + - 中小型数据库     - 大型开发测试     - Web 服务器日志 | 不被经常访问或者低 I/O 负载的应用场景 |

**高效云盘**

**产品特点**

高效云盘采用固态硬盘与机械硬盘的混合介质作为存储介质，具备如下特点：

* + I/O 性能：最高提供 3000 随机读写 IOPS、80 MBps 的吞吐性能。
  + 数据可靠性：采用分布式三副本机制，提供 99.9999999% 的数据可靠性。
  + 性能准则：
  + IOPS：起步 1000 IOPS、每 GB 增加 6 个 IOPS，最高 3000；
  + 吞吐量：起步 50 MBps、每 GB 增加 0.1 MBps、最高 80 MBps；  
    例如 250 GB 的高效云盘，拥有 2500 的随机读写 IOPS、75 MBps 的吞吐性能。
  + 最大容量：单块高效云盘最大提供 32768 GB 存储空间。
  + 单独挂载：高效云盘支持挂载到在相同可用区内的任意云服务器 ECS 实例上。

**应用场景**

* + MySQL、SQL Server、PostgreSQL 等中小型关系数据库应用。
  + 对数据可靠性要求高、中度性能要求的中大型开发测试应用。

**SSD 云盘**

**产品特点**

SSD 云盘利用分布式三副本机制，能够提供稳定的高随机 I/O、高数据可靠性的高性能存储，具备如下特点：

* + 高性能：最高提供 20000 随机读写 IOPS、300 MBps 的吞吐能力。
  + IOPS：起步 1200 IOPS、每 GB 容量提供 30 个随机 IOPS 能力，最大提供 20000 随机读写 IOPS 性能。比如 100 GB 的 SSD 云盘提供 4200 IOPS 性能；334 GB 的 SSD 云盘提供 11220 IOPS 性能。   
    **注意**： SSD 云盘只有挂载到 I/O 优化的实例时，才能获得期望的 IOPS 性能。挂载到非 I/O 优化的实例时无法获得期望的 IOPS 性能。
  + 吞吐量：SSD云盘的吞吐性能=min{80+0.5\*容量，300}MBps，起步 80、每 GB 增加 0.5 MBps，上限为 300 MBps 的吞吐性能。
  + 性能基准：
  + 块大小为 4KB/8KB 时，IOPS 可达最大 20000。
  + 块大小为 16 KB 时，IOPS 最大 17200 左右，吞吐量到 300 MBps 上限。
  + 块大小为 32 KB 时，IOPS 最大 9600 左右，吞吐量到 300 MBps 上限。
  + 块大小为 64 KB、IOPS 最大 4800 左右。
  + 以此类推。
  + 数据可靠性：采用分布式三副本机制，提供 99.9999999% 的数据可靠性。
  + 最大容量：单块 SSD 云盘最大提供 32768 GB 存储空间。
  + 单独挂载：SSD 云盘支持挂载到在相同可用区内的任意 ECS 实例上。

**应用场景**

SSD 云盘具备稳定的高随机 I/O 性能及高数据可靠性，适合以下场景：

* + PostgreSQL、MySQL、Oracle、SQL Server 等中大型关系数据库应用。
  + 对数据可靠性要求高的中大型开发测试环境。

**普通云盘**

**产品特点**

普通云盘采用机械磁盘作为存储介质，利用分布式三副本机制，提供高数据可靠性，具备如下特点：

* + I/O 性能：提供数百的随机读写 IOPS 能力，最大 30 ~ 40 MBps 的吞吐量。
  + 数据可靠性：采用分布式三副本机制，提供 99.9999999% 的数据可靠性。
  + 最大容量：单块普通云盘最大提供 2000 GB 存储空间。
  + 单独挂载：普通云盘支持挂载到在相同可用区内的任意 ECS 实例上。

**应用场景**

* + 适合数据不被经常访问或者低 I/O 负载的应用场景；如果应用需要更高的 I/O 性能，建议使用 SSD 云盘。
  + 需要低成本并且有随机读写 I/O 的应用环境。

云盘三副本技术介绍

**Chunk**

ECS 用户对虚拟磁盘的读写最终都会被映射为对阿里云数据存储平台上的文件的读写。阿里云提供一个扁平的线性存储空间，在内部会对线性地址进行切片，一个分片称为一个 Chunk；对于每一个 Chunk，阿里云会复制出三个副本，并将这些副本按照一定的策略存放在集群中的不同节点上，保证用户数据的可靠。

**三份副本的原理**

在阿里云数据存储系统中，有三类角色，分别称为 Master、Chunk Server，以及 Client。ECS 用户的一个写操作，经过层层转换，最终会交由 Client 来执行，执行过程简要说明如下：

* + Client 计算出这个写操作对应的 Chunk。
  + Client 向 Master 查询该 Chunk 的三份副本的存放位置。
  + Client 根据 Master 返回的结果，向这 3 个 Chunk Server 发出写请求。
  + 如果三份都写成功，Client 向用户返回成功；反之，Client 向用户返回失败。

内网

目前阿里云的服务器内网间，非 I/O 优化的实例为千兆共享的带宽，I/O 优化的实例为万兆共享的带宽，没有特殊限制。由于是共享网络，因此无法保证带宽速度是不变的。

对于内网中的 ECS 实例：

* + 如果是 **经典网络** 实例：
  + 同一账号、同一地域的实例，如果在同一个安全组，即默认内网互通。
  + 同一账号、同一地域、不同可用区的实例，如果在同一安全组，即使内网 IP 地址不是同一网段，也可以正常内网连接。
  + 不同账号、相同地域的实例，可以通过安全组实现内网互通，详情请参见 [安全组应用案例](https://help.aliyun.com/document_detail/25475.html)。
  + 如果是 **专有网络** 实例：
  + 同一账号、同一地域、同一个 VPC 网络的实例：
  + 在同一个安全组内，内网互通。
  + 在不同安全组内，需要使用安全组授权后才能内网互通。
  + 同一账号、同一地域的实例，如果在不同的 VPC 网络，需要通过高速通道实现网络互通，详情请参见 [高速通道使用场景](https://help.aliyun.com/document_detail/27787.html)。
  + 实例的内网 IP 地址不能进行修改、更换。
  + 实例的内网、外网不支持 VIP（虚拟 IP）配置。
  + 实例的网络类型不同，默认不能内网互通。您可以启用 [ClassicLink 功能](https://help.aliyun.com/document_detail/57517.html) 实现经典网络 ECS 通过私网访问 VPC 网络内的云资源。

经典网络的IP

**私网 IP**

每个实例会分配一块私网网卡，并绑定一个私网 IP。私网 IP 是必选的且无法修改。

注意不要在操作系统内部自行变更私网 IP，否则会导致私网通讯中断。

同一地域内实例之间通过私网 IP 进行的通讯流量是免费的。

私网 IP 可以用于以下情况：

* + 负载均衡
  + ECS 实例之间内网互访
  + ECS 实例与其他云服务（如 OSS、RDS）之间内网互访

**公网 IP**

如果您购买了公网带宽（即公网带宽不为 0 Mbps），阿里云会为您的实例分配一块公网网卡，并为网卡配置一个公网 IP 地址。

无论您选择何种计费方式，在购买实例的时候，您都需要选择公网的带宽限制。根据您选择的带宽限制，对公网网卡的出方向带宽进行限制。

公网流量是收费的。

公网 IP 可以用于以下情况：

* + ECS 实例与 Internet 之间互访
  + ECS 实例与云服务之间互访

**组播和广播**

云服务器 ECS 不支持组播和广播。

专有网络的IP

专有网络(VPC)类型的ECS实例有两种公网 IP 类型：

* + 公网 PublicIP
  + 弹性公网IP（EIP）

**专有网络和经典网络的 PublicIP 异同**

* + **相同点**：都是可以通过 PublicIP 进行访问 Internet 的。在产品上的所有操作都相同，随实例一起购买，可以升级带宽，不能解绑，可以随实例一起释放。
  + **区别**：专有网络的 PublicIP 是 NAT IP，在机器内部无法通过命令行查询；经典网络是 Binding IP，可以在机器中通过命令行查询。

安全组

**安全组限制**

* + 单个安全组内的实例个数不能超过 1000。如果您有超过 1000 个实例需要内网互访，可以将他们分配到多个安全组内，并通过互相授权的方式允许互访。
  + 每个实例最多可以加入 5 个安全组。
  + 每个用户的安全组最多 100 个。
  + 对安全组的调整操作，对用户的服务连续性没有影响。
  + 安全组是有状态的。如果数据包在 Outbound 方向是被允许的，那么对应的此连接在 Inbound 方向也是允许的。
  + 安全组的网络类型分为经典网络和专有网络。
  + 经典网络类型的实例可以加入同一地域（Region）下经典网络类型的安全组。
  + 专有网络类型的实例可以加入同一专有网络（VPC）下的安全组。

**安全组规则限制**

每个安全组最多有 100 条安全组规则。

镜像

云服务器 ECS 提供了以下灵活多样的方式让您方便的获取镜像：

* + 选择阿里云官方提供的公共镜像（支持 Linux 和 Windows 的多个发行版本）
  + 去镜像市场选择第三方服务商（ISV）提供的镜像
  + 根据现有的云服务器 ECS 实例创建自定义镜像
  + 选择其他阿里云用户共享给您的镜像

**快照链**

快照链是一个磁盘中所有快照组成的关系链，一个磁盘对应一条快照链，所以快照链 ID 即磁盘 ID。一条快照链会包括以下信息：

* + 快照节点：快照链中的一个节点表示磁盘的一次快照。
  + 快照容量：快照链中所有快照占用的存储空间。
  + 快照额度：每条快照链最多只能有 64 个快照额度，包括手动创建及自动创建的快照；达到额度上限后，如果要继续创建自动快照，系统会自动将最早的自动快照删掉。

应用场景

快照作为一种便捷高效的数据保护服务手段，推荐应用于以下业务场景中：

* + 系统盘、数据盘的日常备份，您可以利用快照定期的对重要业务数据进行备份，来应对误操作、攻击、病毒等导致的数据丢失风险。
  + 更换操作系统，应用软件升级或业务数据迁移等重大操作前，您可以创建一份或多份数据快照，一旦升级、迁移过程中出现任何问题，可以通过数据快照及时恢复到正常的系统数据状态。
  + 生产数据的多副本应用，用户可以通过对生产数据创建快照，从而为数据挖掘、报表查询、开发测试等应用提供近实时的真实生产数据。

使用自定义镜像创建实例

阿里云自定义镜像主要用于创建 ECS 实例。操作系统、已经预安装的应用程序和数据，都会自动复制到新实例中。您可以根据通过准备好的自定义镜像方便的创建有相同配置环境的实例，从而提高工作或交付效率。

在使用镜像时，需要考虑镜像的条件如下：

* + 所在地域：只能使用本地域的镜像。镜像不能跨地域使用。您可以将镜像复制到目标地域。详情请参见 [复制镜像](https://help.aliyun.com/document_detail/25462.html)。
  + 操作系统：Linux 还是 Windows 的。
  + 操作系统的位数：32 位还是 64 位。

公网带宽降低到 0 Mbps 对公网 IP 地址的影响：

* + 如果是经典网络实例，公网 IP 地址不变。
  + 如果是 VPC 网络实例，在新计费周期，公网 IP 地址会被释放。

升级内存

**说明：**升级 CPU 和内存后，必须通过 [控制台重启实例](https://help.aliyun.com/document_detail/25440.html) 才能生效。在实例内重启无效。

**注意**：

* + 实例停止后，仍会继续收费。如要停止收费，您需要[释放](https://help.aliyun.com/document_detail/25442.html)实例。
  + 停止操作只能在实例处于运行中时进行。
  + 停止操作会造成您的实例停止工作，从而中断您的业务，请谨慎执行。

当欠费停机但数据未释放时，ECS 实例上的数据会保留 7 天。您需要充值，自动结清欠费账单并有余额，然后将需要保留的实例进行重开机操作。**如果没有重开机，ECS 实例将在欠费之日起 7 天后自动释放，且数据不可恢复。**

按量付费转包年包月

您目前可以通过 ECS 管理控制台将按量付费实例转换成包年包月实例。

**限制**

您每次最多只能将 20 个按量付费实例转换为包年包月实例。

创建云盘

您可以从管理控制台购买云盘（即数据盘），对系统的存储空间进行扩容。

* + 最多可开通 **250** 块云盘。
  + 一个实例可以支持最多挂载 **16** 块数据盘
  + 每块高效云盘和 SSD 云盘最大支持 **32768 GB** 容量；每块普通云盘最大支持 **2000 GB**。
  + 用快照创建磁盘
  + 您可以使用**系统盘**或**数据盘**的快照创建磁盘。创建的云盘可以挂载到**同一地域、且同一可用区**的任一台实例上。

挂载磁盘时，实例需要满足以下条件：

* + 实例状态必须为 **运行中** (Running) 或者 **已停止**（Stopped）。
  + 实例状态不能为 **已锁定** (Locked)。
  + 实例不欠费。
  + 本地磁盘不能挂载、或卸载。
  + 挂载磁盘时，云盘的状态必须为 **待挂载** (Available)。
  + 云盘只能挂载在同一可用区内的实例上，不能跨可用区挂载。
  + 云盘作为实例系统盘时，不支持单独的挂载操作。
  + 您只能卸载**数据盘**，不能卸载**系统盘**。
  + 本地磁盘不能卸载。

**注意：**

* + 只有 **按量付费** 的数据盘才能释放。**包年包月** 的数据盘随主机购买，不支持卸载释放。
  + 数据盘释放后，您的手动快照会保留，**但自动快照随数据盘一起删除，并且不可恢复，因此请慎重操作**。
  + 为保持足够的快照额度来顺利完成周期性的自动快照策略，您可以删除不需要的手动快照和自动快照。

更换系统盘（非公共镜像）

* + 更换系统盘不会更改您实例的 IP 地址和 MAC 地址。
  + 系统盘类型不能更换。
  + 为保持足够的快照数量额度完成新磁盘的自动快照策略，您可以[删除不需要的快照](https://help.aliyun.com/document_detail/25458.html)。
  + 实例更换系统盘后，原系统盘会被释放。
  + *自动快照策略是与磁盘 ID 绑定的。更换了新的系统盘后，旧磁盘上应用的自动快照策略自动失效。您需要对新系统盘设置自动快照策略。*

**共享块存储扩容注意事项**

* + 共享块存储扩容时，需要所挂载到的实例均为**已停止（Stopped）**状态，或者磁盘本身为**待挂载**状态。
  + 如果共享块存储本身已挂载到实例上，需要在控制台重启实例，扩容才能生效。
  + 如果尚未挂载到实例上，扩容后直接挂载到实例上即可生效，无需重启实例。

* + 扩容系统盘时，您只能增加或保持不变磁盘容量，而不能减小。
  + 扩容系统盘不会更改您实例的 IP 地址和 MAC 地址。
  + 系统盘类型不能更换。
  + Windows 2003 不支持系统盘扩容。
  + 扩容后的系统盘，容量最小应为 40 GB，最大应为 500 GB。
  + *挂载在实例上的数据盘，只有当实例处于****运行中****(****Running****) 或****已停止****(****Stopped****) 状态时才可以扩容。****扩容这种数据盘需要重启实例后才能使扩容后的容量生效，而重启实例会停止实例，中断您的业务，所以请您谨慎操作****。*
  + *建议在扩容数据盘之前手动创建快照，以备份数据。*
  + *无论数据盘的状态是****待挂载****还是****使用中****，都可以执行磁盘扩容操作。*
  + *如果数据盘正在创建快照，则不允许执行扩容数据盘的操作。*
  + *磁盘扩容功能只能扩容数据盘，不能扩容系统盘或本地盘（本地 SSD 盘等）。*

包年包月数据盘转为按量付费

***注意****：*

* + *续费降配后，新配置会在新的计费周期内生效，当前剩余服务期限内配置不会发生改变。续费降配后，当前剩余服务期限内将不再支持升级和降配功能。****请谨慎操作。***
  + *如果包年包月实例上挂载的是按量付费数据盘，您可以使用*[*升级配置*](https://help.aliyun.com/document_detail/25437.html#upgrade)*功能将计费方式转为包年包月，转换后立即生效。*
  + *如果按量付费实例上挂载的是按量付费数据盘，您可以使用*[*按量付费实例转换为包年包月实例*](https://help.aliyun.com/document_detail/49884.html)*功能将计费方式转为包年包月，转换后立即生效。*
  + 创建快照
  + 避开业务高峰。创建快照可能会轻微降低磁盘的性能，出现短暂瞬间变慢。
  + 实例状态必须为 **运行中** 或 **已停止**。
  + 手动创建的快照会一直保留。如不再需要，请 [手动删除](https://help.aliyun.com/document_detail/25458.html)。
  + 配置自动快照随磁盘释放

默认设置下，在[更换系统盘](https://help.aliyun.com/document_detail/25448.html)、[释放磁盘](https://help.aliyun.com/document_detail/25447.html)的时候，磁盘的自动快照会随磁盘一起释放。

您也可以禁止自动快照随磁盘释放，保留自动快照

删除快照和自动快照策略

**注意：**

* + 快照删除后，不能用任何方法恢复。请谨慎操作。
  + 如果快照用于制作自定义镜像，需要先删除关联的镜像，然后才能删除。
  + 使用快照创建自定义镜像

**说明**

* + 一个账号在一个地域最多能创建 100 个自定义镜像。
  + 创建的自定义镜像不能跨地域使用。
  + 通过自定义镜像开通的 ECS 实例可以更换操作系统。更换系统后原来的自定义镜像还可以继续使用。请参见 [更换系统盘](https://help.aliyun.com/document_detail/25448.html)。
  + 使用自定义镜像开通的 ECS 实例可以升级 CPU、内存、带宽、磁盘等。
  + 自定义镜像功能不受付费模式限制，即不区分包年包月和按量付费。包年包月 ECS 实例的自定义镜像，可以用于开通按量付费的 ECS 实例；反之亦然。
  + 用于创建自定义镜像的 ECS 实例到期或数据释放后（即用于快照的系统盘到期或释放），创建的自定义镜像不会受影响，使用自定义镜像开通的 ECS 实例也不会受影响。但自动快照则会随着 ECS 实例释放而被清除。
  + 使用实例创建自定义镜像

**注意**：

* + 请将实例中的敏感数据删除之后再创建自定义镜像，避免数据安全隐患。
  + 创建镜像的过程中，不要改变实例的状态，不要进行停止、启动、重启等操作，以免创建失败。
  + 复制镜像
  + 复制镜像是将一个自定义镜像从一个地域复到到另一个地域的过程，您可以使用复制的镜像 [创建 ECS 实例](https://help.aliyun.com/document_detail/25424.html) 或 [更换系统盘](https://help.aliyun.com/document_detail/25448.html)。
  + 共享镜像
  + 您可以把自己的自定义镜像共享给其他用户，该用户可以通过管理控制台或 ECS API 查询到其他账号共享到本账号的共享镜像列表。被共享用户可以使用其他账号共享的镜像创建 ECS 实例和更换系统盘。

**限制**

* + 每个镜像最多可以共享给 **50** 个用户。
  + 共享镜像不占用自身镜像名额。
  + 共享过来的镜像在创建实例的时候有地域限制，同共享源地域相同。
  + 其他账号共享给您的账号不能再共享给其他人。只有镜像的拥有者才能共享给其他账号。

**删除共享镜像造成的影响**

* + 如果您的自定义镜像已经共享给其他账号，这个镜像依然可以删除，但首先需要把该自定义镜像的所有关系删除。
  + 如果您把某个自定义镜像的共享账号给删除了，那么该用户通过管理控制台或者 ECS API 就无法查询到该镜像，也无法使用该镜像创建 ECS 实例和更换系统盘。
  + 共享的镜像被删除后，会导致使用共享镜像创建的 ECS 实例进行重新初始化系统盘时失败。

安全组

应用案例

* + [**案例 1：内网互通**](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc58746.6.675.MZsJ4d#intranetCommunication)   
    **场景举例：**如果您需要将一台 ECS 实例上的资源拷贝到另一台 ECS 实例上，可以通过安全组设置实现两台 ECS 实例内网互通后再拷贝。
  + [**案例 2：拦截特定 IP 或端口**](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc58746.6.675.MZsJ4d#blockAccess)   
    **场景举例：**如果您的 ECS 实例因为异常的 IP 地址登录造成内存溢出、带宽跑满、CPU 跑满等情况，您可以通过安全组设置拦截这些异常 IP 地址。
  + [**案例 3：只允许特定 IP 远程登录到实例**](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc58746.6.675.MZsJ4d#specifyIpAccess)   
    **场景举例：**如果您的 ECS 实例被肉鸡，您可以修改远程登录端口号，并设置只允许特定的 IP 地址远程登录到您的 ECS 实例。
  + [**案例 4：只允许实例访问外部特定 IP**](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc58746.6.675.MZsJ4d#specifyInstanceAccess)   
    **场景举例：**如果您的 ECS 实例被肉鸡，对外恶意扫描或发包，您可以通过安全组设置您的 ECS 实例只能访问外部特定 IP。
  + [**案例 5：允许远程登录到实例**](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc58746.6.675.MZsJ4d#allowRemoteAccess)
  + [**案例 6：允许公网通过 HTTP、HTTPS 等服务访问实例**](https://help.aliyun.com/document_detail/25475.html?spm=5176.doc58746.6.675.MZsJ4d#allowHttp)
  + 添加安全组规则
  + **VPC 网络**：只需要设置出方向和入方向的规则，不区分私网和公网。
  + **经典网络**：需要分别设置公网和内网的出方向和入方向规则。
  + 删除安全组
  + 如果安全组下没有加入的ECS实例，且该安全组没有被其他安全组的安全组规则引用，才能成功删除该安全组。删除安全组，会删除该安全组下所有的安全组规则。
  + 克隆安全组
  + 跨地域克隆，假设您已经在地域 A 里创建了一个安全组 SG1，此时您需要对地域 B 里的实例使用与 SG1 完全相同的规则，您可以直接将 SG1 克隆到地域 B，而不需要在地域 B 从零开始创建安全组。
  + 从经典网络克隆到VPC，假设您已经创建了一个适用于经典网络的安全组 SG2，此时您需要对一些处于 VPC 网络里的实例使用与 SG2 完全相同的规则，您可以在克隆 SG2 时将网络类型改为 VPC，生成一个适用于 VPC 网络的安全组。
  + 如果需要对一个线上业务执行新的安全组规则，您可以先克隆原来的安全组作为备份，再修改安全组里的规则。如果新的安全组规则对线上业务产生了不利影响，您可以全部或部分还原安全组规则。

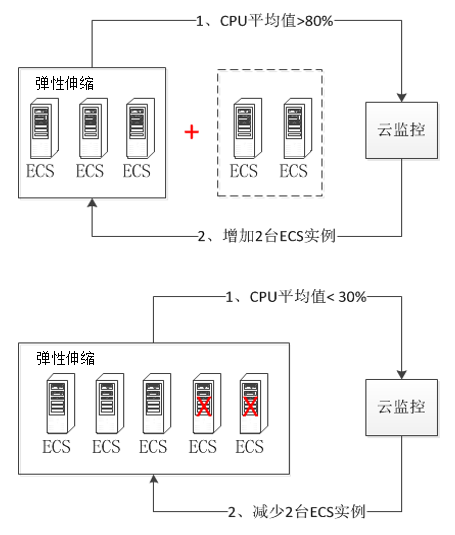
**阿里云云计算专业认证考试（ACP）（弹性伸缩）**

**弹性伸缩**

**产品概述**

弹性伸缩是根据用户的业务需求和策略，自动调整其弹性计算资源的管理服务。根据业务需求，您可以自动地调整其弹性计算资源。

* + 在业务需求增长时无缝地增加ECS实例。
  + 在业务需求下降时自动减少ECS实例以节约成本。



**名词解释**

**弹性伸缩**

弹性伸缩是根据用户的业务需求和策略，自动调整其弹性计算资源的管理服务。其能够在业务增长时自动增加ECS实例，并在业务下降时自动减少ECS实例。

**伸缩组**

伸缩组是具有相同应用场景的ECS实例的集合。伸缩组定义了组内ECS实例数的**最大值、最小值**及其相**关联的负载均衡实例**和**RDS实例等**属性。

**伸缩配置**

伸缩配置定义了用于弹性伸缩的**ECS实例的配置信息**。

**伸缩规则**

伸缩规则定义了具体的**扩展或收缩操作**，例如**加入或移出N个ECS实例**。

**伸缩活动**

伸缩规则成功触发后，就会产生一条伸缩活动。伸缩活动主要用来描述伸缩组内ECS实例的变化情况。

**伸缩触发任务**

用于触发伸缩规则的任务，如**定时任务**、**云监控的报警任务**。

**冷却时间**

冷却时间是指，在**同一伸缩组内**，一个伸缩活动执行完成后的一段锁定时间。在这段锁定时间内，该伸缩组不执行其他的伸缩活动。

**备注**

* + 伸缩组包含**伸缩配置、伸缩规则、伸缩活动。**
  + 伸缩配置、伸缩规则、伸缩活动依赖伸缩组的生命周期管理，删除伸缩组的同时会删除与伸缩组相关联的伸缩配置、伸缩规则和伸缩活动。
  + 伸缩触发任务有定时任务、云监控报警任务等类型。
  + **定时任务独立于伸缩组存在**，不依赖伸缩组的生命周期管理，删除伸缩组不会删除定时任务。
  + **云监控报警任务独立于伸缩组存在**，不依赖伸缩组的生命周期管理，删除伸缩组不会删除报警任务。

**产品优势**

**产品功能**

* + 根据客户业务需求横向扩展ECS实例的容量，即**自动增加和减少ECS实例**。
  + **支持负载均衡配置**：在增加或减少ECS实例时，**自动向负载均衡实例中添加或移除相应的ECS实例**。
  + **支持RDS访问白名单**：在增加或减少ECS实例时，**自动向RDS访问白名单中添加或移出该ECS实例的IP**。

**产品特点**

* + **随需应变：**根据需求“恰到好处”地分配资源，无需担心需求预测的准确性，无需担心突增的业务变化。
  + **自动化：**无需人工干预，自动创建和释放ECS实例，自动配置负载均衡和RDS访问白名单。
  + **伸缩模式丰富：**多模式兼容，可同时配置**定时、动态、自定义、固定、健康模式**，可通过**API方便对接外**在监控系统。
  + **智能：**智能调度应对各种复杂场景。

**应用场景**

* + 某**视频公司**：春晚或每周五热门节目来临时，如临大敌，需要按负载自动弹性伸缩。
  + 某视频直播公司：无法预估业务负载情况，需要根据CPU利用率、Load、带宽利用率，自动弹性伸缩。
  + 某游戏公司：每天中午12点，每天晚上6点~9点，需要定时扩容。

**伸缩模式**

弹性伸缩模式主要分六类：

* + 定时模式：配置周期性任务（如每天 13：00），定时地增加或减少 ECS 实例。
  + 动态模式：基于云监控性能指标（如 CPU 利用率），自动增加或减少 ECS 实例。
  + 固定数量模式：通过 **最小实例数** （MinSize）属性，可以让您始终保持健康运行的 ECS 实例数量，以保证日常场景实时可用。
  + 自定义模式：根据用户自有的监控系统，通过 API 手工伸缩 ECS 实例。
  + 手工执行伸缩规则。
  + 手工添加或移出既有的 ECS 实例。
  + 手工调整 MinSize、MaxSize 后，弹性伸缩会自动创建或释放 ECS 实例，尽可能将当前 ECS 实例维持在MinSize～MaxSize之间。
  + 健康模式：如 ECS 实例为非 running 状态，弹性伸缩将自动移出或释放该不健康的 ECS 实例。
  + 多模式并行：以上所有模式都可以组合配置,客户预期每天 13：00 ~ 14：00 会出现业务高峰，所以设置定时创建 20 台 ECS 实例的伸缩模式，当客户不确定业务高峰期的实际需求是否会高于客户预期时,如某天实际需要 40 台 ECS 实例，可同时配置动态伸缩模式以应付不可预期的变化。

限制条件

弹性伸缩对用户有以下限制：

* + 弹性伸缩的 ECS 实例中部署的应用需要是无状态、可横向扩展的。
  + 由于弹性伸缩会自动释放 ECS 实例，所以用于弹性伸缩的 ECS 实例不可以保存应用的状态信息（如 session）和相关数据（如数据库、日志等）。如果应用中需要保存状态信息，可以考虑把状态信息保存到独立的状态服务器、数据库（如 RDS）及集中日志存储（如 Log）。
  + 弹性伸缩自动扩展出来的实例暂不支持直接自动添加到 OCS 访问白名单中，需要您自行添加。
  + 弹性伸缩目前不支持纵向扩展，即弹性伸缩暂时无法自动升降ECS实例的 CPU、内存和带宽。
  + 每个用户所能创建的伸缩组、伸缩配置、伸缩规则、伸缩 ECS 实例、定时任务的数量都有一定的限制。

**常见问题**

**进行弹性伸缩的ECS实例如何保证配置环境的一致，或者如何把之前的系统环境直接添加到新创建的弹性伸缩实例中？**

创建弹性伸缩配置 ECS 实例时，您可以通过 ECS 自定义镜像模板来创建实例。在 ECS 实例运行过程中，如果需要做系统内部数据同步，建议您自定义安装 rsync 进行同步。

**购买了 100 个 product code 同样值的镜像，是否支持在所有的地域可用？**

不支持，目前镜像市场的镜像已经具备 region 属性，请您购买需要使用的地域镜像。

**1. 弹性伸缩自动创建的实例如何查看密码并进行登录？**

* + 由于弹性伸缩自动创建的实例没有统一的登录密码，在 Linux 环境下，推荐您通过设置 **公私钥** 认证的方式来进行 SSH 免密码登录。
  + 如果您不需要通过设置 **公私钥** 认证的方式来进行 SSH 免密码登录，则需要在控制台重置密码（重启生效）才能登录。
  + **弹性伸缩创建伸缩配置的时候为什么不能选择已经购买过的云服务器？**

弹性伸缩支持添加已有 ECS 实例，需要满足如下条件：

* + 加入的ECS实例必须与伸缩组在同一个 region；
  + 加入的ECS实例规格必须与生效伸缩配置的实例规格一致；
  + 加入的ECS实例的状态必须是 **运行中** 状态；
  + 加入的ECS实例不能已加入到其它伸缩组中；
  + 加入的ECS实例支持包年包月和按量付费两种类型；
  + 加入的ECS实例在移出伸缩组时不会被释放。

**弹性伸缩是否支持已有的包年包月实例添加？**

目前弹性伸缩默认自动创建按量付费实例，但是同时也支持用户已有的包年包月和按量付费的实例添加。

**同一台 ECS 实例可以添加进入不同的伸缩组吗？**

目前无法支持。

**弹性伸缩组是否可以添加不同规格的 ECS 实例？**

每个伸缩组中只能设置一种伸缩配置的规格（CPU 和 Memory）。不过您可以设置多个伸缩组，在每个伸缩组设置不同的配置。

**弹性伸缩服务内，一个伸缩组的最大实例数只能为 100 个，是否可以调高最大实例数？**

您可以提交工单申请更高配额，我们的工作人员会在工作日时间内协助您添加，周末节假日顺延。

**将 ECS 实例移出伸缩组并释放，ECS 实例上的数据能否保存？**

* + ECS 实例移除之后会释放，数据无法保留。
  + 由于弹性伸缩会自动释放 ECS 实例，所以用于弹性伸缩的 ECS 实例不可以保存应用的状态信息（如 session）和相关数据（如数据库、日志等）。如果应用中需要保存状态信息，可以考虑把状态信息保存到独立的状态服务器、数据库（如 RDS）及集中日志存储（如 Log Service）。

**通过 API 接口调用 DisableScalingGroup 方法，会不会导致已经自动加入的 ECS 实例被释放？**

执行 DisableScalingGroup 时，伸缩组内自动创建的按量付费实例不会被自动释放。

**弹性伸缩是否支持设置周期性的重复任务，在指定的时间内重复性地执行？**

支持，您可以在弹性伸缩控制台的 **自动触发任务管理** > **定时任务** 中设置定时任务，如果未设置重复周期，则按指定的日期和时间执行一次。如果设置了重复周期，任务按照指定的时间点重复执行。

**如何保证手工添加的 ECS 实例不被移出伸缩组？**

假设您需要添加 N 台包年包月的 ECS 实例加入伸缩组，并且希望这些 ECS 实例不会自动被移出，那么您需要做以下两个配置（同时满足）：

* + 将最小实例数（MinSize）设置为等于或者大于 N。
  + 将移出策略（RemovalPolicy）的第一条挑选规则设置为 **最早伸缩配置对应的实例** （OldestScalingConfiguration）。
  + 根据弹性伸缩的规则，手工添加的 ECS 实例不会对应任何伸缩配置（因为它们不是通过伸缩配置创建的），所以弹性伸缩会先挑选自动创建的 ECS 实例移出伸缩组，只有当自动创建的 ECS 实例移出伸缩组完毕，才会挑选手工添加的 ECS 实例移出伸缩组。
  + 请注意自动创建的 ECS 实例，被移出伸缩组会自动释放；手工添加的 ECS 实例，移出伸缩组不会被释放。

**弹性伸缩在添加伸缩组时，能绑定多个负载均衡实例吗？**

一个伸缩组默认只能绑定一个负载均衡实例，如果需要使用多个负载均衡实例，您可以提交工单申请更高配额，我们的工作人员会在工作日时间内协助您添加，周末节假日顺延。

**弹性伸缩新增的ECS实例是否可以添加进入多个负载均衡中？**

可以，目前仅支持一个负载均衡，多个负载均衡需要提交工单申请。

**我现在有公网的负载均衡，在创建伸缩配置时ECS实例是否需要公网带宽？**

创建伸缩配置时，ECS实例可以不选择公网带宽。但是为了方便管理ECS实例，建议您购买至少1Mbps的ECS带宽。

**弹性伸缩一定要搭配负载均衡、云监控、RDS才能使用吗？**

弹性伸缩是一个开放的灵活扩展收缩平台。弹性伸缩可以单独扩展和收缩ECS实例，既可以搭配负载均衡和RDS一起部署，也可以不搭配负载均衡和RDS。弹性伸缩支持通过云监控触发任务扩展和收缩ECS实例，也可以通过弹性伸缩的Open API对接客户自己的监控系统，客户可以通过自己的监控系统，触发弹性伸缩的伸缩活动。

**弹性伸缩告警触发条件有几种？**

可以针对CPU负载、内存利用率、系统平均负载、外网和内网出入流量等进行监测告警，自动增加或减少ECS实例。

**弹性伸缩是否可以根据云监控中自定义报警项进行动态伸缩？**

目前不支持根据自定义监控进行动态伸缩。

**自动释放ECS实例之前，如何保证应用运行完成？**

如您需要确保在ECS实例上任务处理完成才让弹性伸缩释放该实例，您需要在自定义镜像中存放执行脚本，并设置命令在操作系统关机时自动执行该脚本。

创建弹性伸缩失败

创建弹性伸缩失败，显示“The resource is out of usage.”，如下图所示。

或者显示错误“The specified region is in resource control , please try later.”，如下图所示。

这一般都是由于后端没有ECS资源导致，建议更换可用区进行测试。

创建伸缩组时选择的负载均衡实例和RDS实例的具体用途

 悻 组 
沣 悻 霏 组 名 称 ： 
地 悻 缩 大 实 例 敦 （ 台 冫 ． 
： 《 默 认 冷 却 的 闾 （ 秒 } 0 
负 载 均 衛 ． 
名 称 为 2 ． 4 伊 i 、 ， 以 大 小 写 享 母 ， 享 或 中 文 开 ， 可 包 自 0 
小 力 O 。 大 为 100 
小 力 O 。 大 为 100 
300 
小 为 O 。 必 须 力 翌 
罕 悻 蒹 蚍 过 应 的 实 例 ， 在 中 再 
何 证 李 工 加 的 ECS 列 不 满 移 出 悻 蒹 
远 菽 
罕 创 的 实 

**负载均衡**

如果在伸缩组中创建了负载均衡实例，伸缩组会自动将加入伸缩组的ECS实例添加到指定的负载均衡实例中。指定的负载均衡实例必须开启健康检查，否则会导致创建失败，同时加入负载均衡的ECS实例权重默认是50。

**RDS数据库**

如果在伸缩组中指定了RDS实例，伸缩组会自动将加入伸缩组的ECS实例的内网IP添加到指定的RDS实例的访问白名单中，允许ECS通过内网进行连接。

工作流程

 ． 监 控 健 康 情 况 
《 伸 缩 触 发 任 务 
定 时 
健 康 检 
自 定 义 

**伸缩活动**

* + 同一伸缩组内、同一时刻只能有一个伸缩活动在执行。
  + 伸缩活动不可以中断。例如，某个创建 20 台 ECS 实例的伸缩活动正在执行中，当创建到第 5 台 ECS 实例时，您无法强行终止该伸缩活动。
  + 伸缩活动有 ECS 实例加入伸缩组失败时，需要保持 ECS 实例级事务的完整性，而非伸缩活动级事务的完整性，即只进行 ECS 实例级回滚，而不是伸缩活动级回滚。例如，当伸缩组创建了 20 台 ECS 实例，但只有 19 台 ECS 实例成功加入负载均衡时，则只对不成功的 1 台 ECS 实例进行自动释放操作。
  + 由于弹性伸缩是借助阿里云的 RAM （Resource Access Management）服务，通过 ECS Open API代替用户弹性伸缩 ECS 实例资源，所以回滚的 ECS 实例仍然会被扣费。

**冷却时间**

* + 在冷却时间内，伸缩组只会拒绝云监控报警任务类型的伸缩活动请求，其他类型的触发任务（如用户手工执行伸缩规则、定时任务等）可以绕过冷却时间立即执行伸缩活动。
  + 每个伸缩活动的最后一个 ECS 实例加入或移出伸缩组成功后，整个伸缩组冷却时间才开始计时。

**释放策略类型**

* + OldestInstance：取最早创建的 ECS 实例，如果作为一级筛选条件，不区分是用户手工加入的 ECS 实例还是自动创建的 ECS 实例，直接按最早创建的实例进行移出。
  + NewestInstance：取最新创建的实例，如果作为一级筛选条件，不区分是用户手工加入的 ECS 实例还是自动创建的 ECS 实例，直接按最新创建的 ECS 实例进行移出。
  + OldestScalingConfiguration：取最早伸缩配置创建的 ECS 实例；对于用户手工移入的 ECS 实例，自动跳过此条件；如果与伸缩配置相关联的 ECS 实例已经全部移出，还要继续移出 ECS 实例时，则在用户手工加入的 ECS 实例中（没有与伸缩配置关联的 ECS 实例），随机选一台。

**实例状态**

ECS 实例在伸缩组中的生命周期，通过以下几种状态描述：

* + Pending – 表示 ECS 实例正在加入伸缩组，包括创建实例、加入负载均衡、添加 RDS 访问名单等过程。
  + InService – 表示 ECS 实例已成功加入伸缩组，并正常提供服务。
  + Removing – 表示 ECS 实例正在移出伸缩组。

**实例的健康状态**

ECS 实例在伸缩组中的健康状态为：

* + Healthy(健康)
  + Unhealthy(不健康)
  + 冷却时间

冷却时间锁定的是同一个伸缩组的伸缩活动，对于多伸缩组内的不同伸缩活动，不做以上要求。

伸缩规则的冷却时间和伸缩组默认冷却时间同时存在时，以伸缩规则的冷却时间为伸缩活动的执行标准。

停用伸缩组，然后再启用伸缩组后，冷却时间的效果将会失效。例如一个伸缩活动在 0 点 0 分执行完成，冷却时间是 15 分钟。假如这时用户停用再启用伸缩组，冷却时间效果失效，即如果在 0 点 3 分云监控触发伸缩活动，请求将被接受执行。

伸缩组状态

**伸缩组状态**

伸缩组具有以下三种状态：Active、Inacitve 和 Deleting。

|  |  |
| --- | --- |
| **状 态** | **Open API 对应的状态** |
| 创建中 | Inactive |
| 已创建 | Inactive |
| 启用中 | Inactive |
| 运行中 | Active |
| 停用中 | Inactive |
| 已停止 | Inactive |
| 删除中 | Deleting |

伸缩活动的流程

**伸缩组自动伸缩**

**自动扩展**

* + 判断伸缩组的健康状态和边界条件。
  + 分配 ActivityId 和执行伸缩活动。
  + 创建 ECS 实例。
  + 修改 Total Capacity。
  + 分配 IP。
  + 添加 RDS 白名单。
  + 启动 ECS 实例，等待启动完成。
  + 挂载负载均衡，将权重设为创建伸缩配置时指定的 **负载均衡权重**。
  + 伸缩活动完成，启动 cooldown。

**自动收缩**

* + 判断伸缩组的健康状态和边界条件。
  + 分配 ActivityId 和执行伸缩活动。
  + 从负载均衡移除 ECS 实例。
  + 停止 ECS 实例，等待停止完成。
  + 从 RDS 白名单移除。
  + 释放 ECS 实例。
  + 修改 Total Capacity。
  + 伸缩活动完成，启动 cooldown。

伸缩活动的状态

伸缩活动执行完成后，有 3 种状态：

* + 成功（**Successful**）：根据 Minsize、MaxSize 调整后的伸缩规则，伸缩活动成功将所有 ECS 实例加入或移出伸缩组。  
    **说明**：加入伸缩组成功表明创建 ECS 实例、加入负载均衡实例、配置 RDS 访问白名单每一步骤都成功。以上任何步骤失败都认为是该 ECS 实例加入伸缩组失败。
  + 部分成功（**Warning**）：根据 Minsize、MaxSize 调整后的伸缩规则，伸缩活动至少有一台ECS实例加入或移出伸缩组不成功。
  + 全部失败（**Failed**）：根据 Minsize、MaxSize 调整后的伸缩规则，伸缩活动没有一台ECS实例成功加入或移出伸缩组。
  + 移出不健康ECS实例
  + 移出不健康 ECS 实例，不受 MinSize 的限制，即移出不健康 ECS 实例后，有可能导致 Total Capacity 低于 MinSize，此时弹性伸缩服务会自动创建差额的 ECS 实例使得 Total Capacity 等于 MinSize。

通知

需要短信息和邮件通知的伸缩活动条件如下：

* + 只有定时任务、云监控报警任务、健康检查触发的伸缩活动才通知。
  + 只有创建了 ECS 实例或释放了 ECS 实例才通知。
  + 数量限制

每个用户所能创建的伸缩组、伸缩配置、伸缩规则、伸缩 ECS 实例、定时任务的数量都有一定的限制。

* + 一个用户最多能创建 20 个伸缩组。
  + 一个伸缩组内最多只能创建 10 个伸缩配置。
  + 一个伸缩组内最多只能创建 10 个伸缩规则。
  + 对于所有地域和所有伸缩组，一个用户最多能弹性伸缩 100 台 ECS 实例。此数量只包含自动创建的 ECS 实例，不包含手工添加的 ECS 实例。
  + 一个用户最多能创建 20 个定时任务。

创建伸缩组

**负载均衡注意事项**

如果在伸缩组中指定了负载均衡实例，伸缩组会自动将加入伸缩组的 ECS 实例添加到指定的负载均衡实例当中。

* + 指定的负载均衡实例必须是已启用状态。
  + 指定的负载均衡实例所有配置的监听端口必须开启健康检查，否则创建失败。
  + 如果负载均衡实例已挂载了 VPC 类型的 ECS 实例，则不支持该负载均衡实例加入伸缩组。
  + 加入负载均衡的 ECS 实例的权重默认为 50。

**RDS注意事项**

如果在伸缩组中指定了 RDS 实例，伸缩组会自动将加入伸缩组的 ECS 实例的内网 IP 添加到指定的 RDS 实例的访问白名单当中。

* + 指定的 RDS 实例必须是运行中状态。
  + 指定的 RDS 实例访问白名单的 IP 个数不能达到上限。
  + 伸缩组创建成功后，伸缩组不会立即生效，只有启用伸缩组，才能接受伸缩规则的触发和执行相关的伸缩活动。

**限制**

* + 伸缩组、负载均衡实例和 RDS 实例必须在同一个地域。
  + 修改伸缩组

在创建伸缩组的属性中，有以下属性不可以修改：

* + 地域
  + 负载均衡
  + RDS 数据库实例
  + 当伸缩组为生效（Active）和失效（Inactive）状态，才可以执行此功能。
  + 查询伸缩组

伸缩组具有以下几种状态（LifecycleState）：

* + 生效（Active）：生效状态，在该状态下才能接收执行伸缩规则的请求并触发伸缩活动。
  + 失效（Inacitve）：失效状态，在该状态下不接收任何执行伸缩规则的请求。
  + 删除中（Deleting）：伸缩组正在删除，在该状态下不接收任何执行伸缩规则的请求。
  + 启用伸缩组

启用伸缩组成功以后（active 状态），如果当前 ECS 实例数量小于 MinSize，则弹性伸缩服务会自动创建差额的按量付费的 ECS 实例。

例如，创建伸缩组时，指定 MinSize = 5，启用前伸缩组已有 2 台 ECS 实例，则启动后会自动创建 3 台 ECS 实例。

当伸缩组为失效（Inactive）状态，才可以执行此功能。

当伸缩组有生效的伸缩配置时，才可以执行此功能。

一个伸缩组在同一时刻只能有一个 **生效** 的伸缩配置。

删除伸缩组

控制台删除一个伸缩组时，会默认采用强制删除的模式

您手工添加的 ECS 实例会被移出伸缩组，弹性伸缩自动创建的 ECS 实例会被自动删除。

删除伸缩组，包含删除相关联的伸缩配置、伸缩规则、伸缩活动、伸缩请求的信息。

删除伸缩组，不会删除以下任务或实例：定时任务、云监控报警任务、负载均衡实例、RDS 实例。

创建伸缩配置

* + 创建伸缩配置中指定的 ECS 实例规格（InstanceType）必须与伸缩组内生效的伸缩配置的实例规格一致。
  + 伸缩配置不可以编辑更改。如需使用新的伸缩配置，只能重新创建新的伸缩配置，并在伸缩组中激活。  
    在伸缩组中加入新的伸缩配置，不会影响通过早前的伸缩配置创建并正在运行的 ECS 实例。
  + 一个伸缩组内最多只能创建 10 个伸缩配置。
  + 删除伸缩配置
  + 伸缩配置在伸缩组中属于生效状态，则该伸缩配置不允许删除。
  + 某个伸缩配置创建的任意一个 ECS 实例仍存在于伸缩组中，则该伸缩配置不允许删除。
  + 创建伸缩规则

伸缩规则（Scaling Rule）定义了具体的扩展或收缩操作，例如加入或移出 N 个实例。如果伸缩规则的执行会造成伸缩组的 ECS 实例数低于 MinSize 或高于 MaxSize 时，则弹性伸缩会自动调整需要加入或移出的 ECS 实例数，使之按照 **将伸缩组的实例数调整到 MinSize** 或 **将伸缩组的实例数调整到 MaxSize** 的规则执行。

例如：某个伸缩组，MaxSize = 3，当前实例数 Total Capacity =2，伸缩规则指定 **加 3 台 ECS 实例**，则在实际执行过程中只会 **加 1** 台 ECS 实例。（伸缩规则当中的值不会调整）

例如：某个伸缩组，MinSize = 2，当前实例数 Total Capacity = 3，伸缩规则指定 **减去 5 台 ECS 实例**，则在实际执行过程中只会 **减 1** 台 ECS 实例。（伸缩规则当中的值不会调整）

查看伸缩活动

最多只能返回30天前的伸缩活动。

触发任务，添加ECS实例

可以为经典网络类型或 VPC 类型，但有以下限制：

* + 当伸缩组为经典网络类型时，只能添加经典网络类型的 ECS 。
  + 当伸缩组为 VPC 类型时，只能添加同一 VPC 下的 ECS 。

执行此功能，必须同时满足以下条件：

* + 当伸缩组为生效（Active）状态。
  + 当伸缩组没有伸缩活动正在执行。
  + 如果该功能指定的实例数加上当前伸缩组的实例数（Total Capacity）大于 MaxSize 时，则执行失败。
  + 触发任务，移出ECS实例

执行该功能，必须同时满足以下条件：

* + 伸缩组为生效（Active）状态。
  + 当伸缩组没有伸缩活动正在执行。
  + 执行功能返回成功，只是表示弹性伸缩服务接受了该功能的调用请求，伸缩活动可以执行，但不代表伸缩活动能够执行成功。用户需要通过返回的 ScalingActivityId 查看该伸缩活动的执行状态。
  + 如果当前伸缩组的实例数（Total Capacity）减去该功能指定的实例数小于 MinSize 时，则执行失败。
  + 创建定时任务
  + 每个用户最多能创建 20 个定时任务。
  + 如果指定了同一分钟的定时任务，会选取最近创建的定时任务执行。
  + 创建报警任务
  + 根据传入参数创建报警任务。报警任务名称在用户账号下唯一。如果该报警任务名称已经存在，则用新的值进行替换。
  +  报 任 务 
    在 0 甲 彗 鰹 之 前 栾 要 在 ECS 的 像 里 新 本 的 亓 监 崆 enta 
    地 任 名 称 ： 
    地 监 崆 室 源 ． 
    地 监 崆 项 ： 
    绗 计 周 期 （ 分 〕 ． 
    地 绗 计 办 法 ． 
    望 总 几 次 后 彗 0 
    地 发 规 则 ． 
    http://jiankong.aliyun.com/readme.htm 
    菽 享 或 中 文 开 。 可 包 
    名 称 为 2 O 个 享 。 以 大 小 写 享 母 
    层 少 要 2 个 享 
    干 均 匾 

* **阿里云云计算专业认证考试（ACP）（SLB）**

基础架构

* + 四层采用开源软件LVS（Linux Virtual Server）+ keepalived的方式实现负载均衡，并根据云计算需求对其进行了个性化定制。
  + 七层采用Tengine实现负载均衡。Tengine是由淘宝网发起的Web服务器项目，它在Nginx的基础上，针对有大访问量的网站需求，添加了很多高级功能和特性。
  +  引 回 
    L く S 冊 
    回 訓 図 A 
    Te コ 旦 コ e 冊 
    一 コ met 
    凹
  + LVS集群内的每台LVS都会进行会话，通过组播报文同步到该集群内的其它LVS机器上，从而实现LVS集群内各台机器间的会话同步。
  + **注意**：对于连接未建立（三次握手未完成），或者已建立连接但未触发会话同步机制，热升级不保证连接不中断，需要依靠客户端重新发起连接。
  + 功能概述
  + **协议支持**当前提供四层（TCP协议和UDP协议）和七层（HTTP和HTTPS协议）的负载均衡服务。
  + **健康检查**支持对后端ECS进行健康检查。负载均衡服务会自动屏蔽异常状态的ECS，待该ECS恢复正常后自动解除屏蔽。
  + **会话保持**提供会话保持功能。在会话的生命周期内，可以将同一客户端的会话请求转发到同一台后端ECS上。
  + **调度算法**支持轮询、加权轮询（WRR）、加权最小连接数（WLC）三种调度算法。
  + 轮询：按照访问次数依次将外部请求依序分发到后端ECS上。
  + 加权轮询：用户可以对每台后端服务器设置权重值。权重值越高的服务器，被轮询到的次数（概率）也越高。
  + 加权最小连接数：除了根据对每台后端服务器设定的权重值来进行轮询，同时还考虑后端服务器的实际负载（即连接数）。当权重值相同时，当前连接数越小的后端服务器被轮询到的次数（概率）也越高。
  + **域名URL转发**针对七层协议（HTTP协议和HTTPS协议），支持按设定的访问域名和URL将请求转发到不同的虚拟服务器组。
  + **多可用区**支持在指定可用区创建负载均衡实例。在多可用区部署的地域还支持主备可用区，当主可用区出现故障时，负载均衡可自动切换到备可用区上提供服务。
  + **访问控制**通过添加负载均衡监听的访问白名单，仅允许特定IP访问负载均衡服务。
  + **安全防护**结合云盾，可提供5G以下的防DDoS攻击能力。
  + **证书管理**针对HTTPS协议，提供统一的证书管理服务。证书无需上传到后端ECS，解密处理在负载均衡上进行，降低后端ECS的CPU开销。
  + **带宽控制**支持根据监听设置其对应服务所能达到的带宽峰值。
  + **实例类型**提供公网和私网类型的负载均衡服务。您可以根据业务场景来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。  
    公网类型的负载均衡默认使用经典网络；私网类型的负载均衡服务可以选择使用经典网络或专有网络。
  + **监控**提供丰富的监控数据，实时了解负载均衡运行状态。
  + **管理方式**提供控制台、API、SDK多种管理方式。
  + 使用场景
  + **场景一：应用于高访问量的业务**
  + **场景二：横向扩张系统**
  + **场景三：消除单点故障**
  + **场景四：同城容灾 （多可用区容灾）**
  + 为了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部署了多可用区以实现同地域容灾。当主可用区出现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（大约30s中断）切换到另外一个备可用区恢复服务能力；当主可用区恢复时，负载均衡同样会自动切换到主可用区提供服务。
  +  载 均 衡 实 例 
    ECS 实 例 
    認 主 可 用 区 ： 可 用 区 A 
    负 鞔 均 衡 实 例 
    ECS 实 例 
    备 可 用 区 ： 可 用 区 B 
  + **场景五：跨地域容灾**
  + 您可以在不同地域下部署负载均衡实例，并分别挂载相应地域内不同可用区的ECS。上层利用云解析做智能DNS，将域名解析到不同地域的负载均衡实例服务地址下，可实现全局负载均衡。当某个地域出现不可用时，暂停对应解析即可实现所有用户访问不受影响。
  +  载 均 衡 妄 例 
    ECS 实 例 
    認 主 可 厍 区 ： 可 区 A 
    路 由 器 
    华 东 1 
    Internet 
    N 
    云 析 DNS 
    负 载 均 衡 实 例 
    ECS 实 例 
    备 可 厍 区 ： 可 区 B 
    路 由 器 
    负 载 均 衡 妄 例 
    ECS 实 例 
    可 用 区 A 
    华 北 1 
  + 使用限制

|  |  |  |
| --- | --- | --- |
| **限制项** | **普通用户限制描述** | **例外申请方式（例外上限）** |
| 创建负载均衡实例的财务限制 | 账户余额大于等于100元现金 | 工单 |
| 创建负载均衡实例的用户限制 | 在控制台上创建，需要实名认证  使用Open API创建，无限制 | 没有例外 |
| 创建负载均衡实例的ECS限制 | 至少拥有一个ECS实例，且只能选择有ECS实例的地域创建负载均衡实例 | 没有例外 |
| 创建负载均衡实例可用的类型 | 公网/私网 | 没有例外 |
| 负载均衡实例的计费方式 | 按使用流量或按固定带宽 | 没有例外 |
| 负载均衡实例公网带宽可选范围（按固定带宽） | 1-5000 MB，缺省 1 MB | 没有例外 |
| 用户默认按量付费的实例配额 | 60 | 工单 |
| 负载均衡实例名称输入范围 | 长度限制为1-80个字符，允许包含字母、数字、‘-’、‘/’、‘.’、‘\_’这些字符 | 没有例外 |
| 负载均衡实例监听数量 | 50 | 没有例外 |
| 一个七层监听最多可添加的域名、URL转发策略数量 | 20 | 工单申请 |
| 负载均衡实例监听可选择的协议类型 | HTTP/HTTPS/TCP/UDP | 没有例外 |
| 负载均衡实例监听可选择的前端/后端端口范围（公共云） | 1-65535 | 没有例外 |
| 负载均衡实例监听可选择的前端/后端端口范围（金融云） | 80，443，2800-3300，5000-10000，13000-14000 | 没有例外 |
| 支持的调度算法 | 轮询、加权轮询、加权最小连接数 | 没有例外 |
| HTTP协议会话保持Cookie处理方式 | insert/server，缺省insert | 没有例外 |
| HTTP协议会话保持Cookie超时时间 | 1-86400，缺省3600 | 没有例外 |
| HTTP协议会话保持Cookie名称 | 长度限制为1-200，cookie必须遵守[RFC2965](https://www.ietf.org/rfc/rfc2965.txt)。这意味着它只能包含ASCII英文字母和数字，不能包含逗号、分号或空格，也不能以$字符开头 | 没有例外 |
| HTTP健康检查端口 | 输入范围为1-65535，缺省使用后端服务端口 | 没有例外 |
| HTTP健康检查域名 | 长度限制为1-80，只能使用字母、数字、‘-’、‘.’这些字符 | 没有例外 |
| HTTP健康检查URI | 长度限制为1-80，只能使用字母、数字、‘-’、‘/’、‘.’、‘%’、‘?’、‘#’、‘&’这些字符 | 没有例外 |
| HTTP健康检查超时时间 | 输入范围1-300秒，缺省5  注意：如果健康检查Timeout < Interval，则健康检查Timeout无效，超时时间为Interval | 没有例外 |
| HTTP健康检查间隔 | 输入范围1-50秒，缺省2 | 没有例外 |
| HTTP健康检查健康阈值 | 输入范围2-10，缺省3 | 没有例外 |
| HTTP健康检查不健康阈值 | 输入范围2-10，缺省3 | 没有例外 |
| TCP会话保持超时时间 | 1-3600秒 | 没有例外 |
| TCP健康检查端口 | 输入范围为1-65535，缺省使用后端服务端口 | 没有例外 |
| TCP健康检查超时时间 | 输入范围1-300秒，缺省5 | 没有例外 |
| TCP健康检查间隔 | 输入范围1-50秒，缺省2 | 没有例外 |
| TCP健康检查健康阈值 | 输入范围2-10，缺省3 | 没有例外 |
| TCP健康检查不健康阈值 | 输入范围2-10，缺省3 | 没有例外 |
| UDP健康检查端口 | 输入范围为1-65535，缺省使用后端服务端口 | 没有例外 |
| UDP健康检查超时时间 | 输入范围1-300秒，缺省10 | 没有例外 |
| UDP健康检查间隔 | 输入范围1-50秒，缺省5 | 没有例外 |
| UDP健康检查健康阈值 | 输入范围2-10，缺省3 | 没有例外 |
| UDP健康检查不健康阈值 | 输入范围2-10，缺省3 | 没有例外 |
| 后端ECS批量添加/删除数量 | 20 | 没有例外 |
| 单个key的API访问频率限制 | 5000次/天 | 联系客服或BD提高限制 |
| 每个用户上传证书数量限制 | 100 | 联系客服或BD提高限制 |

* + 负载均衡常见问题

**1. 负载均衡支持哪些调度算法？**

可以使用轮询、加权轮询（WRR）、加权最小连接数（WLC）这三种调度方法：

* + **轮询**：按照访问次数依次将外部请求依序分发到后端ECS上。
  + **加权轮询**：您可以对每台后端服务器设置权重值，权重值越高的服务器，被轮询到的次数（概率）也越高。
  + **加权最小连接数**：除了根据每台后端服务器设定的权重值来进行轮询，同时还考虑后端服务器的实际负载（即连接数）。当权重值相同时，当前连接数越小的后端服务器被轮询到的次数（概率）也越高。

**公网负载均衡和私网负载均衡的区别是什么？**

* + 公网类型的负载均衡可以对公网提供服务，实例创建后的服务地址是公网IP地址。
  + 私网类型的负载均衡仅可以在阿里云内部使用，实例创建后的服务地址是内网IP地址。

**3. 是否可以修改负载均衡实例类型？**

不可以。

因为负载均衡系统会根据负载均衡实例类型来分配不同的服务地址（公网IP或私网IP），所以当您要进行负载均衡实例类型切换操作时，必须通过先删除后创建的方式，在新建实例时重新选择负载均衡实例类型。

**7. 负载均衡是否依赖外网带宽？**

负载均衡和后端ECS之间是通过内网进行通信的，所以ECS无需配置外网带宽。

**9. 如何避免负载均衡服务本身的故障问题？**

* + 添加不同可用区的ECS实例作为负载均衡实例的后端服务器，从而提高本地可用性。
  + 在同一地域创建多个负载均衡实例，通过DNS轮询的方式对外提供服务，从而提高本地可用性。
  + 在不同地域创建多个负载均衡实例，通过DNS轮询的方式对外提供服务，从而提高跨地域的可用性。

**1. 负载均衡实例最多支持添加多少ECS实例？**

目前没有限制在负载均衡实例上配置的ECS数量。

**2. 负载均衡运行中是否可调整ECS数量？**

可以。

**3. 后端ECS的操作系统是否可以不同？**

可以。

**4. 可以使用不同地域的ECS实例作为后端服务器么？**

不可以，负载均衡不支持跨地域部署，一个负载均衡实例后端的多台ECS必须属于同一账号且同一地域才可以配置。

**10. ECS设置权重的作用是什么？**

您可以指定后端服务器池内各ECS的转发权重，权重比越高的ECS将被分配到更多的访问请求，根据后端ECS的对外服务能力和需求来区别设定。

比如将负载均衡池中的A、B两台ECS权重均配置成100和均配置成50都是一样的，在 A、B两台ECS健康检查状态均正常的情况下，A、B均分前端请求的流量。 但是如果 A、B权重设置为10和100，则在A、B两台ECS健康检查状态均正常的情况下，A承担10/(10+100)的流量，B承担 100/(10+100)的流量。

**2. 健康检查的参数配置是否有相对合理的推荐值？**

* + 关于TCP/HTTP/HTTPS健康检查的建议使用如下配置：

|  |  |
| --- | --- |
| **配置** | **推荐值** |
| 响应超时时间 | 5秒 |
| 健康检查间隔 | 2秒 |
| 不健康阈值 | 3 |
| 健康阈值 | 3 |

在此配置下有利于用户服务及应用状态的尽快收敛。如果您有更高要求，可以适当地降低响应超时时间值，但必须先保证自己服务在正常状态下的处理时间小于这个值。

* + 健康检查失败时间窗=(2秒检查间隔+5秒响应超时时间)×3次检查=21秒
  + 健康检查成功时间窗=5秒检查间隔×3次检查=15

**4. TCP监听服务如何选择健康检查方式？**

TCP监听支持HTTP和TCP两种健康检查方式。

* + TCP模式的健康检查是基于网络层探测，利用传统的三次握手机制来判断后端服务是否异常。
  + HTTP模式的健康检查是检测head请求，Tengine节点服务器通过发送HTTP的header请求，然后对比返回码参数来校验后端服务是否异常。

TCP的健康检查方式对服务器的性能资源消耗相对要少一些，如果您对后端服务器的负载高度敏感，则选择TCP方式进行健康检查，如果负载不是很高，则选择HTTP方式进行健康检查。

**5. ECS权重设置为0对健康检查有什么影响？**

将负载均衡后端ECS的权重置零，相当于将该ECS手工下线，用于对相应ECS进行重启、配置调整等主动运维。而由于该状态下，业务数据是无法经负载均衡转发到该服务器的，所以健康检查相应的自然会显示异常。

**2. 为什么访问域名发现在HTTP的header头部多了一个字段Transfer-Encoding: chunked？**

将域名解析到七层负载均衡的服务地址后，访问域名发现在HTTP的header头部多了一个字段Transfer-Encoding: chunked，但是本地主机绑定负载均衡的后端ECS上访问时是没有这个字段的。

由于七层负载均衡的原理是Tengine反向代理，会加入这个字段。Transfer-Encoding字段表示Web服务器自己对本响应消息体作了怎样的编码，Transfer-Encoding: chunked表示做了分块传输。

在四层负载均衡服务中，负载均衡就是单纯的转发流量，不会存在这个字节。

**4. HTTPS监听使用什么端口？**

没有特殊要求，建议使用443端口。

**5. 负载均衡支持哪些类型的证书？**

目前支持服务器证书和CA证书的上传，服务器证书需要上传证书内容和私钥，CA证书只需要上传证书内容；这两种类型的证书都只支持PEM编码格式的上传。

**20. HTTP/HTTPS监听访问后端ECS服务器HTTP协议的版本是什么？**

使用HTTP1.0协议。

**28. 每个监听可以添加多少条转发规则？**

每个监听限制10条转发规则。

**2. 如何开启会话保持？**

在您进行负载均衡服务监听配置时就可以选择是否开启会话保持。您可以针对不同的监听配置不同的会话保持策略。会话保持的最长时间是86400秒（24小时）。

七层（HTTP/HTTPS协议）服务，负载均衡系统是基于cookie的会话保持。植入cookie的会话保持的最长时间是86400秒（24小时）。

**4. 会话保持可设置哪种类型的cookie？**

HTTP/HTTPS监听可使用植入cookie和重写cookie来进行会话保持。

* + 植入cookie: 此种方法下，您只需要指定cookie的过期时间。客户端第一次访问时，负载均衡服务在返回请求中植入cookie（即在HTTP/HTTPS响应报文中插入SERVERID字串），下次客户端携带此cookie访问，负载均衡服务会将请求定向转发给之前记录到的ECS实例上。
  + 重写cookie：此种方式下，您可以根据需要指定HTTPS/HTTP响应中插入的cookie。您需要在后端ECS上维护该cookie的过期时间和生存时间。负载均衡服务发现用户自定义了cookie，将会对原来的cookie进行重写，下次客户端携带新的cookie访问，负载均衡服务会将请求定向转发给之前记录到的ECS实例上。

**5. 是否支持针对不同的域名配置不同的会话保持规则？**

支持。

**2. 负载均衡是否对入流量计费？**

负载均衡目前只对出流量计费，入流量不计费。

**3. 健康检查产生的流量是否会被计费？**

不会。

**6. 负载均衡实例的所有后端ECS都停止，或者没有挂载ECS，是否会被计费？**

负载均衡实例还是会计费，具体说明如下：

* + 按流量计费  
    按使用流量计费只有在实例停止、被释放、或无任何访问的情况下才不会产生流量费用。  
    负载均衡是放置于ECS之前的负载均衡服务设备，通过服务地址的方式提供服务。负载均衡实例的所有后端ECS停止，但负载均衡实例本身服务并未停止，当有请求发生时，入流量还是会到负载均衡的服务地址，负载均衡健康检查发现后端没有可用的ECS，会进行响应。  
    对于四层负载均衡服务，响应的仅是三次握手的包。对于七层负载均衡服务，由于负载均衡七层服务是通过Tengine提供的，因此响应的是一个Tengine的503错误页。如果不停地有访问进来，负载均衡不停地响应，这些响应流量会被计费。  
    对于没有挂载ECS的负载均衡实例也是这样的情况。因此，为避免这种情况下被计费，您不使用负载均衡实例的时候，可以停止这个负载均衡实例。
  + 按固定宽带计费  
    按固定带宽计费时，费用和实例状态及使用流量无关，只要开通就按固定带宽进行计费，只有在实例被释放后才不会被计费。

实例与网络类型

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来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 （IP address）。



结合网络类型，您可以创建以下三种类型的负载均衡实例：

* + 经典网络公网负载均衡实例
  + 经典网络私网负载均衡实例
  + 专有网络私网负载均衡实例

**变更计费**

* + 预付费模式下，您可以变更购买实例的带宽规格，但只支持升级带宽，不支持降低带宽：单击**更多**>**变更带宽规格**。
  + 后付费模式下，您可以在按使用流量和按公网带宽两种计费方式间切换，单击**更多**>**变更计费方式**。
  + 变更配置
  + 负载均衡服务提供预付费（包年包月）和后付费（按量付费）两种购买方式。
  + 后付费实例支持按流量计费和按带宽计费。您可以变更后付费实例的计费方式，计费方式的变更会在次日零点生效。
  + 后端服务器概述

在添加ECS作为后端服务器时，请注意：

* + 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。  
    ***注意****：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。*

**管理后端服务器**

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。

***注意****：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。*

添加后端服务器

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**。

***注意****： ECS实例的网络类型要和该负载均衡实例的类型匹配。详情查看*[*负载均衡实例类型*](https://help.aliyun.com/document_detail/52390.html)*。*

* + 经典网络的公网负载均衡实例，可添加经典网络类型的ECS或者同属于同一VPC的ECS；
  + 专有网络的私网负载均衡实例，仅能添加和负载均衡实例相同VPC内的ECS；
  + 经典网络的私网负载均衡实例，仅能添加经典网络类型的ECS。
  + 创建虚拟服务器组

**使用限制**

* + 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+端口组成。
  + 监听介绍

**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](https://help.aliyun.com/document_detail/53856.html) | * + - 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 |
| [UDP](https://help.aliyun.com/document_detail/53856.html) | * + - 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](https://help.aliyun.com/document_detail/53858.html) | * + - 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](https://help.aliyun.com/document_detail/53858.html) | * + - 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 |

如果选择使用服务器组，选择您要使用的服务器组类型：

* + 虚拟服务器组：一个虚拟服务器组（VServer group）由多个后端服务器组成，且后端服务器的端口可以不同。您可以为不同的监听配置不同的虚拟服务器组，这样就可以将请求转发至不同的后端服务器。详情参考[创建虚拟服务器组](https://help.aliyun.com/document_detail/52795.html)。
  + 主备服务器组：一个主备服务器组由两台后端服务器组成，即一台主服务器，一台备服务器。当您有传统的主备需求时，可以使用主备服务器组。当主机工作正常时，将流量转发至主服务器；当主机宕机时，会将流量转发至备服务器，避免服务中断。

|  |  |
| --- | --- |
| 获取真实IP | 针对四层监听，后端服务器可直接获得来访者的真实IP，无需采用其它手段获取。  **注意**：经典网络的负载均衡实例的UDP监听暂不支持查看源地址。 |
| 会话保持 | 是否开启会话保持。开启会话保持后，负载均衡监听会把来自同一客户端的访问请求分发到同一台后端服务器上。  针对四层（TCP协议和UDP协议）监听，负载均衡是基于IP地址的会话保持，即来自同一IP地址的访问请求转发到同一台后端服务器上。 |
| 连接超时时间 | 指定TCP连接的超时时间。可选值为10-900秒。  **注意**：此配置只适用于TCP监听。 |
| Gzip数据压缩 | 开启该配置对特定文件类型进行压缩。  目前Gzip支持压缩的类型包括：text/xml、text/plain、text/css、application/javascript、application/x-javascript application/rss+xml、application/atom+xml、application/xml。 |
| 附加HTTP头字段 | 选择您要添加的自定义HTTP header字段：   * + - * X-Forwarded-For: 添加该字段获取客户端的IP地址。       * X-Forwarded-Proto: 添加该字段获取客户端与监听连接时所用的协议 (HTTP或HTTPS)。       * SLB-IP: 添加该字段获取负载均衡实例的公网IP。       * SLB-ID: 添加该字段获取负载均衡实例的ID。 |

|  |  |
| --- | --- |
| Gzip数据压缩 | 开启该配置对特定文件类型进行压缩。  目前Gzip支持压缩的类型包括：text/xml、text/plain、text/css、application/javascript、application/x-javascript application/rss+xml、application/atom+xml、application/xml。 |
| 附加HTTP头字段 | 选择您要添加的自定义HTTP header字段：   * + - X-Forwarded-For: 添加该字段获取客户端的IP地址。     - X-Forwarded-Proto: 添加该字段获取客户端与监听连接时所用的协议 (HTTP或HTTPS)。     - SLB-IP: 添加该字段获取负载均衡实例的公网IP。     - SLB-ID: 添加该字段获取负载均衡实例的ID。 |

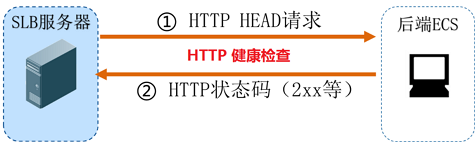
健康检查原理

**健康检查机制**

**HTTP/HTTPS监听**

针对七层（HTTP或HTTPS协议）监听，健康检查通过HTTP HEAD探测来获取状态信息，如下图所示。

*对于HTTPS监听，证书在负载均衡系统中进行管理。负载均衡与后端ECS之间的数据交互（包括健康检查数据和业务交互数据），不再通过HTTPS进行传输，以提高系统性能。*

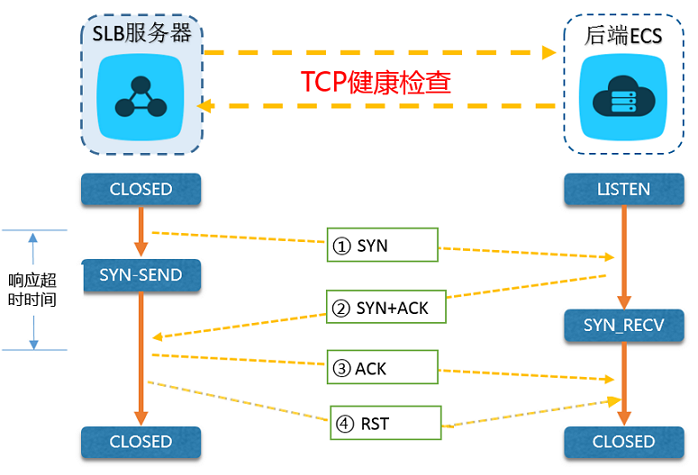


七层监听的检查机制如下：

* + Tengine节点服务器根据监听的健康检查配置，向后端ECS的内网IP+【健康检查端口】+【检查路径】发送HTTP HEAD请求（包含设置的【域名】）。
  + 后端ECS收到请求后，根据相应服务的运行情况，返回HTTP状态码。
  + 如果在【响应超时时间】之内，Tengine节点服务器没有收到后端ECS返回的信息，则认为服务无响应，判定健康检查失败。
  + 如果在【响应超时时间】之内，Tengine节点服务器成功接收到后端ECS返回的信息，则将该返回信息与配置的状态码进行比对。如果匹配则判定健康检查成功，反之则判定健康检查失败。

**TCP监听**

针对四层TCP监听，为了提高健康检查效率，健康检查通过定制的TCP探测来获取状态信息，如下图所示。



TCP监听的检查机制如下：

* + LVS节点服务器根据监听的健康检查配置，向后端ECS的内网IP+【健康检查端口】发送TCP SYN数据包。
  + 后端ECS收到请求后，如果相应端口正在正常监听，则会返回SYN+ACK数据包。
  + 如果在【响应超时时间】之内，LVS节点服务器没有收到后端ECS返回的数据包，则认为服务无响应，判定健康检查失败；并向后端ECS发送RST数据包中断TCP连接。
  + 如果在【响应超时时间】之内，LVS节点服务器成功收到后端ECS返回的数据包，则认为服务正常运行，判定健康检查成功，而后向后端ECS发送RST数据包中断TCP连接。

***注意****：正常的TCP三次握手，LVS节点服务器在收到后端ECS返回的SYN+ACK数据包后，会进一步发送ACK数据包，随后立即发送RST数据包中断TCP连接。*

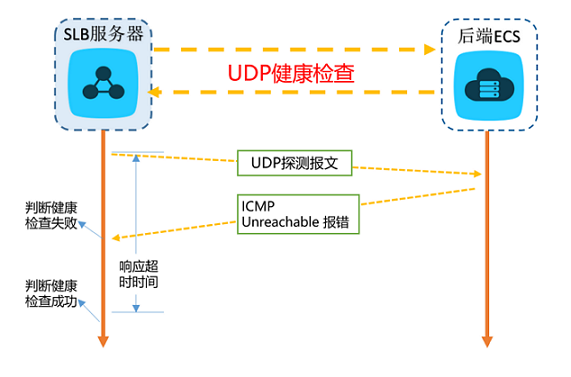
*该实现机制可能会导致后端ECS认为相关TCP连接出现异常（非正常退出），并在业务软件如Java连接池等日志中抛出相应的错误信息（如“Connection reset by peer”）。*

*解决方案：*

* + *TCP监听采用HTTP方式进行健康检查。*
  + *在后端ECS配置了获取客户端真实IP后，忽略来自前述负载均衡服务地址段相关访问导致的连接错误。*

**UDP监听**

针对四层UDP监听，健康检查通过UDP报文探测来获取状态信息，如下图示。



UDP监听的检查机制如下：

* + LVS节点服务器根据监听的健康检查配置，向后端ECS的内网IP+【健康检查端口】发送UDP报文。
  + 如果后端ECS相应端口未正常监听，则系统会返回类似返回 “port XX unreachable”的ICMP报错信息；反之不做任何处理。
  + 如果在【响应超时时间】之内，LVS节点服务器收到了后端ECS返回的上述错误信息，则认为服务异常，判定健康检查失败。
  + 如果在【响应超时时间】之内，LVS节点服务器没有收到后端ECS返回的任何信息，则认为服务正常，判定健康检查成功。

*注意：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ECS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。*

*解决方案：*

*负载均衡通过发送您指定的字符串到后端服务器，必须得到指定应答后才认为检查成功。但该实现机制需要客户端程序配合应答。*

设置白名单访问控制

白名单是一种访问控制方式，可以为负载均衡监听设置仅允许哪些IP访问，适用于应用只允许特定IP访问的场景。

注意：

* + 设置白名单存在一定业务风险。一旦设置白名单，就只有白名单中的IP可以访问负载均衡监听。
  + 如开启访问控制而不设置白名单列表，则这个负载均衡监听就无人可以访问。
  + 设置白名单的过程中可能会引起访问负载均衡监听短时中断。
  + 设置报警规则
  + 云监控支持“短信”、“邮件”、“旺旺” 三种报警方式，暂不支持电话报警。

**阿里云云计算专业认证考试（ACP）（VPC）**

创建VPC时，系统会自动为VPC创建一个路由器和一个路由表。您不可以直接删除专有网络的路由器或路由表，但可以在路由表中添加新的路由条目，来转发流量。当删除VPC后，关联的路由器和路由表也会随之删除。

**安全隔离**

* + 不同用户的云服务器部署在不同的专有网络里。
  + 不同专有网络之间通过隧道ID进行隔离。专有网络内部由于交换机和路由器的存在，所以可以像传统网络环境一样划分子网，每一个子网内部的不同云服务器使用同一个交换机互联，不同子网间使用路由器互联。
  + 不同专有网络之间内部网络完全隔离，只能通过对外映射的IP（弹性公网IP和NAT IP）互联。
  + 由于使用隧道封装技术对云服务器的IP报文进行封装，所以云服务器的数据链路层（二层MAC地址）信息不会进入物理网络，实现了不同云服务器间二层网络隔离，因此也实现了不同专有网络间二层网络隔离。
  + 专有网络内的ECS使用安全组防火墙进行三层网络访问控制。

名词解释

|  |  |  |
| --- | --- | --- |
| **术语** | **英文** | **说明** |
| 专有网络 | VPC | 专有网络是您基于阿里云创建的自定义私有网络，不同的专有网络之间彻底逻辑隔离。您可以在自己创建的专有网络内创建和管理云产品实例，例如ECS，SLB，RDS等。 |
| 交换机 | VSwitch | 交换机是组成专有网络的基础网络设备。它可以连接不同的云产品实例。在专有网络内创建云产品实例时，必须指定云产品实例所连接的交换机。 |
| 路由器 | VRouter | 路由器是专有网络的枢纽。它可以连接专有网络的各个交换机，同时也是连接专有网络与其它网络的网关设备。路由器根据具体的路由条目的设置来转发网络流量。 |
| 路由条目 | Route Entry | 路由表中的每一项是一条路由条目。路由条目定义了通向指定目标网段的网络流量的下一跳地址。路由条目包括系统路由和自定义路由两种类型。 |
| 路由表 | Route Table | 路由表是指路由器上管理路由条目的列表。 |

**路由器和交换机**

|  |  |
| --- | --- |
| **分类** | **限制说明** |
| 路由器 | * + - 每个VPC有且只有1个路由器。     - 路由器不支持BGP和OSPF等动态路由协议。     - 每个路由器有且只有1个路由表。     - 路由表的路由条目会影响VPC中的所有云产品实例。目前不支持指定交换机和云产品实例的源地址策略路由。 |
| 交换机 | * + - VPC的交换机不支持二层广播和组播。     - 交换机本身对云产品实例数量没有限制。它能够容纳多少实例，取决于所在专有网络当前的云产品实例数量，即10,000减去当前已保有的云产品数量。     - 交换机的网段不可以进行修改。 |

**专有网络ECS实例的迁移限制**

您可以在VPC网络中将ECS实例从某一路由器下的一台交换机转移到另一台交换机，但不支持以下操作：

* + ECS实例不支持跨路由器切换。
  + ECS实例不支持网络类型之间的切换，不支持从专有网络迁移到经典网络和不支持从经典网络迁移到专有网络。

在阿里云各项资源中，目前只有ECS实例支持绑定弹性公网IP。

**弹性公网IP和ECS公网IP的区别**

下表描述了弹性公网IP与ECS公网IP的区别。

|  |  |  |
| --- | --- | --- |
| **比较点** | **弹性公网IP** | **ECS公网IP** |
| 支持的网络环境 | 专有网络 | 专有网络和经典网络 |
| 是否能够单独持有 | 支持 | 不支持 |
| 是否支持在ECS上的弹性插拔 | 支持 | 不支持 |
| ECS实例网卡上是否能看到该IP | 不能 | 能 |

使用限制

* + 目前有且只有ECS实例支持绑定弹性公网IP。
  + 弹性公网IP只能绑定在专有网络类型的ECS实例上，不支持绑定在经典网络类型的ECS实例上。
  + 如果要绑定的ECS实例已经配分配了一个公网IP，则该实例不能再绑定弹性公网IP。
  + 一个ECS实例只能绑定一个弹性公网IP。
  + 一个弹性公网IP只能绑定一个ECS实例。
  + 弹性公网IP只能绑定在同地域的ECS实例上。
  + 单个账户下的弹性公网IP配额为20个。
* **阿里云云计算专业认证考试（ACP）（OSS）**

**OSS具备的其他各项优势**

* + **方便、快捷的使用方式**
  + 提供标准的RESTful API接口、丰富的SDK包、客户端工具、控制台。您可以像使用文件一样方便地上传、下载、检索、管理用于Web网站或者移动应用的海量数据。
  + 不限文件数量和大小。您可以根据所需存储量无限扩展存储空间，解决了传统硬件存储扩容问题。
  + 支持流式写入和读出。特别适合视频等大文件的边写边读业务场景。
  + 支持数据生命周期管理。您可以自定义将到期数据批量删除或者转入到低成本的归档服务。
  + **强大、灵活的安全机制**
  + 灵活的鉴权，授权机制。提供STS和URL鉴权和授权机制，以及白名单、防盗链、主子账号功能。
  + 提供用户级别资源隔离机制和多集群同步机制（可选）。
  + **丰富、强大的增值服务**
  + 图片处理：支持jpg、png、bmp、gif、webp、tiff等多种图片格式的转换，以及缩略图、剪裁、水印、缩放等多种操作。
  + 音视频转码：提供高质量、高速并行的音视频转码能力，让您的音视频文件轻松应对各种终端设备。
  + 内容加速分发：OSS作为源站，搭配CDN进行加速分发，具有稳定、无回源带宽限制、性价比高、一键配置的特点。
  + 使用场景

**图片和音视频等应用的海量存储**

OSS可用于图片、音视频、日志等海量文件的存储。各种终端设备、Web网站程序、移动应用可以直接向OSS写入或读取数据。OSS支持流式写入和文件写入两种方式，如下图所示：

elay. 
oss 
MTS 
ECS 
CDN 

**网页或者移动应用的静态和动态资源分离**

利用BGP带宽，OSS可以实现超低延时的数据直接下载。也可以配合阿里云CDN加速服务，为图片、音视频、移动应用的更新分发提供最佳体验，如下图所示：

O 
SSO 

**云端数据处理**

上传文件到OSS后，可以配合媒体转码服务(MTS)和图片处理服务（IMG）进行云端的数据处理，如下图所示：

00 
(IMG) 
oss 
(MT) 

OSS适合谁用？

OSS适合于社区、多媒体分享站、电子商务、等各类形式规模的网站站长，APP应用、软件应用开发者、游戏开发企业、硬件厂商配套应用开发及大规模数据存储需求的用户使用。

OSS适合存什么？

OSS适合存储论坛网站与软件应用中的附件、高清图片、音视频、备份文件等，及各种APP应用、多终端同步软件、网盘下载站的文件存储

OSS存储容量是多少？

OSS总存储容量不限制，单个Bucket容量也不限制。

OSS是不是解决了不同运营商之间的互联互通问题？

阿里云服务器和OSS，都是BGP多线接入的，不同运营商的网站都可以快速访问的。

OSS中的目录/文件夹概念

OSS中文件夹的概念仅是一个逻辑概念，在通过API/SDK的方式设置文件夹的时候可以指定object对应的key值包括前面的目录即可实现该功能。

OSS中的文件夹其实是一个大小为0KB的空文件。因此，用户创建一个key值为1/的object就会定义文件夹1；并且如果用户创建文件abc/1.jpg，系统是不会创建abc/这个文件的，因此在删除abc/1.jpg后将不会再存在abc这个文件夹。

对象存储OSS的特点

**1. 安全稳定，数据可靠**

**安全稳定**

* + 加密用户身份验证，提供签名权限控制与防盗链功能。
  + 强大的安全防护系统，抵御洪水攻击，保障合法请求通过。提供访问日志记录，及时掌握流量动向；

**数据可靠**

* + 分布式存储，保障存储数据安全，存储数据三重备份可靠性99.99999999%；
  + 与阿里巴巴、淘宝彼邻，服务可用性99.9%

**2. 海量存储、弹性计费**

**海量存储**

* + 文件存储数量无上限；根据上传数据方式不同而不同,Put Object方式最大不能超过5GB, 使用multipart上传方式object大小不能超过48.8TB
  + 基于云计算平台动态扩展，存储容量与流量按实际用量自由伸缩，无需手动扩容；

**弹性计费**

* + 用多少付多少，不用提前为峰值买单；
  + 阶梯式计费模式，越用越省钱；

**3.简单易用、性能卓越**

**简单易用**

* + 不断更新的API、与丰富的SDK开发包，开发者快速上手；
  + 提供易用的用户Web控制台文件管理界面，与更多第三方工具、插件，满足各种应用需求；

**性能卓越**

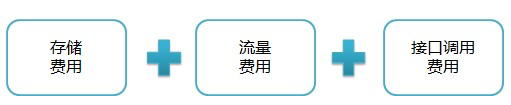
* + 大规模数据处理，文件读写、I/O性能更强；
  + BGP多线（中国电信、联通、移动、教育网等）骨干网络接入，南北互联互通，全国各地流畅访问；

如何对OSS的数据进行迁移

|  |  |  |  |
| --- | --- | --- | --- |
| **应用场景** | **推荐迁移方式** | **优点** | **缺点** |
| 用户本地、OSS、七牛、百度对象存储、金山对象存储的数据迁移到OSS上来。 | OSS-import工具，详细的配置与使用请参考[OssImport](https://help.aliyun.com/document_detail/56990.html)。 | - 您仅需将该工具下载下来，再通过简单的指令即可完成上传操作。  - 该工具使用方便、简单。 |  |
| 两个分属不同区域的存储空间，通过**跨区域复制**将源存储空间的数据同步到目标存储空间。 | OSS提供的跨区域复制功能，详细操作请参考[设置跨区域复制](https://help.aliyun.com/document_detail/31905.html)。 | - 操作方便，通过控制台即可完成。  - 源和目标Bucket自动同步，控制台配置对源Bucket中的对象的改动（新建、覆盖、删除等）同步到目标Bucket。 | 暂不支持对Bucket的备份功能。 |
| 同一数据中心的bucket间的数据迁移。 | 使用 OSS API/SDK的copyObject的接口。 | - 采用内网的方式来迁移。  - 速度较稳定。 | - 需要用户自己去编写代码实现迁移逻辑，详细请参考[Copy Object](https://help.aliyun.com/document_detail/31979.html)。  - 大文件的迁移请使用Upload Part Copy接口，详细请参考[Upload Part Copy](https://help.aliyun.com/document_detail/31994.html)。 |
| 对于从任何地方迁移到OSS或者从OSS迁移到任何地方都是可行的。 | 使用OSS API/SDK的putObject、postObject接口。 | 适用于所有的应用场景。 | - 需要用户自己编写代码实现逻辑。  - 迁移量较大的话会较为麻烦，详细请参考[PutObject](https://help.aliyun.com/document_detail/31978.html) 以及[PostObject](https://help.aliyun.com/document_detail/31988.html)。  - 大文件的迁移请使用Multipart Upload，详细请参考[Multipart Upload](https://help.aliyun.com/document_detail/31991.html)。 |
| 海量数据迁移至OSS。 | 驻云提供的第三方工具，请参考[海量数据迁移至OSS](https://market.aliyun.com/products/52738004/cmfw000394.html?spm=5176.7739628.2.7.bqGh3O#sku=1tb)。 |  |  |

OSS价格总览

OSS 产品的账单费用由三部分组成：



OSS下载有峰值吗，如何收费？

一个OSS用户在一个Region内的上传或下载的带宽缺省阈值是5Gbps，如达到阈值会收到DownloadTrafficRateLimitExceeded或者UploadTrafficRateLimitExceeded错误响应。

OSS最小计费单位是什么？

存储和流量按实际使用量，最小计费单位是GB。请求次数按PUT型每千次，GET型每万次整数计量计费收费。

云服务器与OSS 上传文件，流量与请求次数是否收费？

云服务器与OSS之间通过内网地址上传或下载数据，属内网流量，是免费的。

云服务器与OSS每次请求所产生的请求次数，不分内外网都会计费。

OSS服务如何停止计费？

只要您在OSS上的所有数据都删除完就没有费用产生了。

您可以为您的存储空间设置生命周期（Lifecycle）来清除掉所有数据。

OSS可以上传同名文件吗？

OSS允许用户上传同名的文件，但是会对源文件直接进行覆盖操作。

oss上传提示Name or service not known报错

此报错目前有两种原因。1、上传的host地址输入错误。2、服务器dns无法解析。

OSS下载rar格式文件，打开后乱码

举例说明，http://\*\*\*\*\*\*-cn-hangzhou.aliyuncs.com/l\*\*\*\*\*\*\_1.4.1.rar 是OSS中的文件，但访问下载后为乱码。遇到该问题，可以修改一下Object属性中的http表头，将设置中的 **Content-Type** 值改成 **application/x-rar-compressed** 即可。

OSS通过range参数获取部分文件

大文件（超过100M）的上传和下载会受到网络环境的影响导致文件上传失败。上传的过程中可以Multipart Upload接口进行分快上传，而对于下载的过程则可以通过range参数来决定获取文件的大小以及位置。

OSS文件通过链接地址访问提示下载不能直接在浏览器显示

       OSS出于安全考虑，直接在浏览器地址栏输入OSS通信域名(oss.aliyuncs.com)，如访问：http://bucketname.oss.aliyuncs.com/a.jpg (文件类型包括：txt、html、htm、图片格式、视频格式、音频格式等安全敏感文件)，均限制在浏览器客户端以“另存为”下载的方式打开文件。若需要直接打开文件，请您绑定三级域名。

       OSS Bucket域名绑定可以参考：[点击查看](https://help.aliyun.com/document_detail/31902.html)。图片处理服务绑定后，使用“域名/objectname@规则”的形式即可访问。

       如果使用绑定域名也是直接下载，请查看下Object的HTTP头信息中Content-Type设置是否正确。比如：访问http://bucketname.oss.aliyuncs.com/index.html，如果Object的Content-Type=application/octet-stream，访问会提示下载。设置为正确的格式例如Content-Type=audio/mp3就可以正常访问。

OSS有子用户吗?

不支持

OSS变通实现IP黑白名单

oss服务本身是不提供IP黑名单的功能的，OSS这边提供的安全机制主要包含两个方面：

1）权限管理：bucket可以设为私有、公有读、公共读写三种权限根据用户应用场景而设置。

2）设置refer防盗链设置：可以设置refer访问白名单，详细设置可参考：[点击查看](https://help.aliyun.com/document_detail/31901.html)

注意：当bucket私有的情况下，得使用签名URL才可以访问object

Access key对OSS的bucket控制权限太高，需求对bucket更细粒度的权限控制

RAM的子账户功能可以实现该需求

OSS数据是如何备份？

OSS提供三重备份，故障自动恢复能力，保障数据可靠性99.99999999%。

OSS 校验上传文件的完整性

OSS上的object会有etag标签，etag主要是用来判断服务端数据是否存在变化。但是，这个标签并不一定等同于文件的MD5值，所以不建议作为校验数据一致性的依据。

如果需要校验上传到OSS的文件和本地文件是否一致，可以在上传文件时携带文件的Content-MD5值，这时OSS服务端会帮用户进行MD5校验，只有在OSS服务器接收到的文件MD5值和Content-MD5一致时才可以上传成功，从而保证上传数据的一致性。

ECS用户如何正确使用OSS内网地址？

SS 的每一个 Bucket 都属于一个特定区域，在创建 Bucket 时需要设置区域，创建之后这个区域不可更改。

ECS 用户使用 OSS 内网地址只能访问与 ECS 实例所在区域相同的 Bucket。如果访问的 Bucket 与 ECS 实例不在同一个区域，必须使用 Bucket 所在区域的公网地址。

例如，用户购买了北京的 ECS 实例，该用户有两个 Bucket，访问方式如下所示：

|  |  |  |
| --- | --- | --- |
| **Bucket 名称** | **创建区域** | **访问方式** |
| beijingres | 北京 | 使用内网地址 beijingres.oss-cn-beijing-internal.aliyuncs.com 去访问 beijingres 的资源。 |
| qingdaores | 青岛 | 使用公网地址 qingdaores.oss-cn-qingdao.aliyuncs.com 去访问 beijingres 的资源。因为 ECS 实例在北京，用户不能使用内网地址去访问。 |

正确使用 OSS 内网地址，一方面可以使用同区域 OSS 到 ECS 流量之间的免费流量，另一方面 ECS 内网访问 OSS 的网络质量较好，能够有效提升应用的上传和下载质量。

OSS是否支持bucket作为三级域名的访问方式？

OSS目前已不支持二级域名访问，仅支持三级域名的访问。三级域名外链访问：

<http://bucketname.oss.aliyuncs.com/object>

例如：http://cloudstorage.oss.aliyuncs.com/pujing.jpg

OSS域名绑定（CNAME）功能目前仅支持OSS以三级域名访问方式进行绑定。即您的OSS访问方式为（Bucket name）.${region}.aliyuncs.com。

OSS是否支持HTTPS访问

**OSS支持HTTPS访问**，其采用的是单向HTTPS证书，无需在客户端安装证书。

OSS中可以重命名bucket吗？是否支持object迁移？

OSS的bucket暂不支持重命名，若需要其他名称建议您重新创建bucket。

OSS 提供了COPY objcet的功能，您可以将原bucekt下的文件COPY到新bucket即可。

OSS如何一键删除bucket

1. 下载驻云开发的客户端工具[点此查看](http://market.aliyun.com/product/12-121590002-cmgj000281.html?spm=5176.383663.9.19.2pBcq1) 输入accessid和key后就可以进入管理bucket，在bucket上右键“属性管理”页面可以查看到“删除bucket”选项，就可以一键删除。

2.  使用bucket属性中的lifecycle进行删除：lifecycle 设置1天，1天内不要进行任何操作，1天后之前的所有object都会清空，就可以来删除bucket了

OSS对object和目录的移动与修改操作

OSS实际存储文件实际是没有目录结构的概念，正常看到的OSS目录结构目录是通过“/”来模拟的。

如果OSS上的一个文件oss:/bucket/123/456/789.jpg/ 以“/”结尾就会被识别为目录。

所以移动和改名object原理相同：

1、手动上传object到新的路径位置，删除老文件（删除目录需要先删除目录下的文件）。

2、通过第三方工具复制文件后至目标地址黏贴 ，工具可使用

   Windows版本：[http://market.aliyun.com/product/12-121590002-cmgj000281.html](http://market.aliyun.com/product/12-121590002-cmgj000281.html?spm=5176.383663.9.17.dDfrnG)

   Mac版本：[http://market.aliyun.com/product/12-121590002-cmgj000282.html](http://market.aliyun.com/product/12-121590002-cmgj000282.html?spm=5176.383663.9.18.dDfrnG)

3、可以用ossimport2工具来操作

OSS数据操作常见问题FAQ

问题：OSS支持压缩图片吗？

回答：目前OSS产品可使用“图片处理”功能自定义设置样式，在样式中可以设定缩放类型，实现图片尺寸修改；调整图片质量参数实现压缩功能。

问题：OSS API返回结果没有JSON格式的吗?

回答：只有XML格式

问题：咨询一下java sdk中的 ossclient对象是否是线程安全的？

回答：java sdk中的 ossclient对象是线程安全的

问题：比如多个文件object, 直接指定多个object，生成一个url，打包一起下载？

回答：目前oss不支持多个object打包下载。此需求可以由上层应用逻辑进行处理。OSS服务提供平台级基础资源的存储支持。

问题：OSS不想让其它人访问，那有什么办法能实现么?

回答：把bucket设成私有读写就可以了

问题：OSS存储文件，能为多个用户设置访问权限吗？

回答：OSS目前不支持针对文件进行多用户访问权限的设置。

问题：OSS可以限制 IP 或者域名访问？

回答：OSS支持防盗链功能，即支持域名访问的白名单。暂时不支持限制IP的访问。

问题：OSS服务是否限制用户的并发请求数？

回答：每个区域（例如：北京），每个用户上限为1万QPS

问题：OSS能否避免迅雷感知下载?

回答：暂时不防迅雷。

问题：OSS上传有流量限制吗？

回答：OSS不会限制上传流量

OSS静态文件的打开方式

目前使用OSS自带域名，如http://bucketname.oss.aliyuncs.com/objcet 访问静态文件时（文件类型包括：txt、html、htm、图片格式、视频格式、音频格式等），均限制在浏览器中以“另存为”下载的方式打开文件。

【注意】：网站或应用程序中进行嵌入式引用，则不受强制下载影响，可在网页和应用中直接正常显示。

对象存储OSS控制台操作—图片鉴黄

1）查询条件：每一张图片下方会有绿网图片鉴黄服务检测出的黄图分值，分值从0-100，分值越高，代表该图片是黄图的概率越大。用户可以根据“黄图分值”、“Bucket”、“时间”进行查询。

2）图片操作：用户将鼠标放置在每一张图片上，会显示对该图片相应的操作项，如图所示。用户点击“删除”视为将图片从绿网控制台和OSS Bucket中一并删除；点击“忽略”视为忽略此图片在鉴黄服务中的检测结果，忽略后不再在控制台展示，并不影响保存在OSS Bucket中的图片；当将鼠标放置在上面可查看图片的基本信息，包括图片创建时间、文件名和Bucket名，如下图：

OSS图片处理对原图和缩略图的限制说明

OSS的图片处理服务，对原图和缩略图的使用限制因使用场景不同而有所差异：

* + **使用一般缩略图：**1).对于原图：  
    a. 图片格式只支持jpg, png, bmp, gif, webp, tiff；  
    b. 文件大小不能超过20MB；

2)对于缩略图：对缩略后的图片大小有限制，目标缩略图宽与高的乘积不能超过4096 \* 4096， 单边长度不能超过4096 \* 4。

* + **使用图片旋转：**1).对于原图：  
    a. 图片的宽或者高不能超过4096;  
    b. 图片格式只支持jpg, png, bmp, gif, webp, tiff ；  
    c. 文件大小不能超过20MB；  
    2).对于缩略图：对缩略后的图片大小有限制，目标缩略图宽与高的乘积不能超过4096 \* 4096， 单边长度不能超过4096 \* 4
  + 在OSS图片处理服务绑定域名验证时提示"域名绑定在自己的其他Bucket上，一个域名只能绑定在一个bucket上"。
  + OSS图片处理服务不支持ICO格式图片
  + OSS图片处理URL访问报错NoSuchStyle, NoSuchKey, NoSuchFile
  + 1. NoSuchStyle：使用规则名错误
  + 2. NoSuchKey：没有该object文件
  + 3. NoSuchFile：使用参数名错误
  + OSS图片服务域名与bucket绑定域名的区别

1.定义：

oss图片服务域名：<https://help.aliyun.com/document_detail/oss/getting-started/img.html?spm=5176.docoss/getting-started/bucket-attributes/cname.6.151.Mgn9AS>

bucket绑定域名：<https://help.aliyun.com/document_detail/oss/getting-started/bucket-attributes/cname.html?spm=5176.docoss/api-reference/bucket/GetBucket.6.142.MMpLIa>

2. 区别

1）oss图片服务域名，要在域名服务商那边cname 解析到昆仑域名上 或者bucketname.img-cn-region.aliyuncs.com上

bucket域名：要在域名服务商那边cname 解析到bucketname.oss-cn-region.aliyuncs.com上

2）图片服务域名：可以使用图片服务，并且只能访问图片格式的文件

参考：<https://help.aliyun.com/document_detail/oss/oss-img-guide/concept.html?spm=5176.docoss/oss-img-guide/region_endpoint.6.440.8s2vgN>

bucket域名：只要浏览器支持，可以访问任意格式的object，不仅限于图片格式的文件，不能进行图片处理

图片处理服务是否可以支持跨域操作

图片处理服务目前是支持跨域请求的，具体要在OSS属性中进行CORS设置。但是目前OSS只支持请求头中带有origin头信息的请求，其他类型的跨域请求暂时无法支持。

为什么访问OSS图片在浏览器端为下载

**1. 由于直接使用OSS域名访问的限制**

如果直接使用OSS Bucket域名访问如：bucketname.oss-cn-hangzhou.com/sample.jpg访问图片，产品这里会直接强制返回“content-disposition”头，导致图片会强制下载(这个是产品的默认限制，要求客户必须绑定已经接入阿里云备案的域名才可以在浏览器直接显示)。

这里建议用户使用自定义域名方式进行访问。[点击查看](http://help.aliyun.com/document_detail/oss/getting-started/bucket-attributes/cname.html)

**2. 由于没有正确设置content-type头**

 Content-Type，在web服务中定义文件的类型，决定浏览器将以什么形式、什么编码读取这个文件。如果一张图片文件的content-type没有正确设置，可能导致图片为下载而不是直接在浏览器端进行展示。

OSS提供的安全防护功能介绍

OSS适用于存储各种类型的静态资源，为防止OSS的资源被恶意盗用，OSS提供了几种安全防护功能，用户还可以集成安全类产品进行安全防护。本文会从OSS本身提供的安全防护功能（Referer、跨域、权限控制、安全排查）和结合安全产品两方面进行安全防护（高防、WAF结合OSS使用）的介绍。

**异常流量安全排查**

您可以通过[OSS 管理控制台](https://oss.console.aliyun.com/)->**Bucket名称**->**热点统计**，查看哪些IP发起的请求，是否存在异常IP发起了请求。

* + 对于异常IP发起了请求
  + 如果Bucket私有，建议迁移数据到新的Bucket中。若新的Bucket私有，通过[**开启waf/高防防护**](https://help.aliyun.com/document_detail/57685.html?spm=5176.doc57868.6.1080.j6A6pG#details)的自定义域名对外服务。
  + 如果Bucket公共读： 可以使用Bucket私有对外提供的签名URL来访问（需要业务端集成签名算法，有一定开发成本）。或者迁移数据到另外的Bucket中，方式一通过[**开启waf/高防防护**](https://help.aliyun.com/document_detail/57685.html?spm=5176.doc57868.6.1080.j6A6pG#details)的自定义域名对外服务，方式二使用自定义域名对外服务，开启CDN加速，利用CDN的IP黑名单进行限制访问。数据迁移请参考：[OSSimport](https://help.aliyun.com/document_detail/56990.html)，OSS开启CDN加速：[CDN加速OSS](https://help.aliyun.com/document_detail/31936.html)。
  + 排查Refer监控，查看哪些Refer发起了请求。看看是否存在恶意的盗链行为，若存在，OSS层面要设置Refer白名单
  + 要详细排查哪些IP访问了什么资源，需要分析OSS日志。
  + OSS日志开启看： [OSS日志开启](https://help.aliyun.com/document_detail/31900.html)
  + 日志字段说明：[OSS日志字段说明](https://help.aliyun.com/document_detail/31868.html)
  + 图片样式

所有对图片的变换都会加在URL后面，会导致URL变得冗长，不方便管理与阅读。图片处理服务提供用户将常见的操作保存成一个别名，即样式（Style）。一个复杂操作，利用样式功能后，只要用一个很短的URL就能实现相同的效果。

一个Bucket下面有多个样式，样式的作用范围只在一个Bucket下，目前一个Bucket允许最多有50个样式。

创建存储空间

* + 在**所属地域**框中，下拉选择该存储空间的数据中心。  
    订购后不支持更换地域。如需要通过ECS内网访问OSS，需要选择与您ECS相同的地域。更多信息请参阅[访问域名](https://help.aliyun.com/document_detail/31834.html)。
  + 在**存储类型**框中，下拉选择所需要的存储类型。
  + **标准存储**：高可靠、高可用、高性能，数据会经常被访问到。
  + **低频访问**：数据长期存储、较少访问，存储单价低于标准类型。低频访问存储类型的Object有最短存储时间，存储时间短于30天的文件提前删除会产生一定费用。低频访问存储Object有最小计量空间，Object大小低于128KB，会按照128KB计算存储空间，数据获取会产生费用。
  + **归档存储**：适合需要长期保存（建议半年以上）的归档数据，在存储周期内极少被访问，数据进入到可读取状态需要等待1分钟的解冻时间。适合需要长期保存的档案数据、医疗影像、科学资料、影视素材。
  + 在 **读写权限** 框中，下拉选择对应的权限。
  + **私有**：只有该存储空间的拥有者可以对该存储空间内的文件进行读写操作，其他人无法访问该存储空间内的文件。
  + **公共读**：只有该存储空间的拥有者可以对该存储空间内的文件进行写操作，任何人（包括匿名访问者）可以对该存储空间中的文件进行读操作。
  + **公共读写**：任何人（包括匿名访问者）都可以对该存储空间中的文件进行读写操作。**所有这些操作产生的费用由该存储空间的拥有者承担，请慎用该权限**。
  + 删除存储空间
  + 如果您不再需要存储空间，请将其删除以免进一步产生费用。
  + 设置防盗链
  + OSS是按使用量收费的服务，为了减少您存储于OSS的数据被其他人盗链而产生额外费用，OSS支持设置Referer基于HTTP和HTTPS header 中表头字段Referer的防盗链方法。您可以通过控制台为您的存储空间设置Referer字段的白名单和是否允许 Referer 字段为空的请求访问。

**举例**

对于一个名为 test-1-001 的存储空间，设置其Referer白名单为 <http://www.aliyun.com。则只有Referer为> <http://www.aliyun.com> 的请求才能访问 test-1-001 这个存储空间中的对象。

设置生命周期

您可以通过控制台定义和管理存储空间内所有对象或对象的某个子集（通过指定关键词名称前缀）的生命周期。设置生命周期一般用于文件的批量管理和自动碎片删除等操作。

* + 对于规则匹配的对象，系统会保证在生效日期两日内将数据清除。
  + 应用生命周期规则而批量删除的数据将无法恢复，请谨慎配置该规则。
  + 设置跨区域复制
  + 当前跨区域同步可以支持异名存储空间的同步，如果您拥有两个分属不同区域的存储空间，您可以使用控制台配置同步将源存储空间的数据同步到目标存储空间。
  + ***注意****：目前只有中国大陆地区支持跨区域复制功能。海外地域暂不支持。*
  + 设置回源规则

您可以设置回源规则对获取数据的请求以多种方式进行回源读取。设置回源规则可以满足您对于数据热迁移、特定请求的重定向等需求。回源类型主要分为镜像方式和重定向的方式。您可以最多配置 5 条回源规则，系统将按规则顺序依次执行。

***说明：****镜像回源不支持内网 endpoint，外网流量按正常收费。*

选择 **镜像** 或者 **重定向方式** 配置对应的规则。

* + **镜像**方式下，当请求在OSS中没找到文件时，会自动到源站抓取对应文件保存到OSS，并将内容直接返回到用户。
  + **重定向方式**下，满足响应条件的请求会通过http重定向的方式返回重定向的地址，浏览器或客户端再到源站获取内容。
  + 文件概览
  + 在 OSS 服务中，用户操作的基本数据单元是 object（对象），单个对象的大小限制为 48.8TB，一个存储空间中可以有无限量的对象。

上传文件

创建了存储空间之后，您可以通过以下两种途径上传任意类型文件到存储空间中：

* + 您可以通过OSS控制台上传小于5GB的文件。
  + 您可以通过SDK或API使用Multipart Upload的分片上传的方法，上传大于 5GB 的文件，您可以参阅[开发人员指南](https://help.aliyun.com/document_detail/31991.html)中的相关章节进行上传操作。
  + 新建文件夹
  + 阿里云 OSS 服务是没有文件夹这个概念的，所有元素都是以对象来存储。OSS 控制台中的文件夹本质上来说是创建了一个大小为 0 并以”/“结尾的对象用于同类文件的归类操作和批处理，同时控制台默认将以”/“结尾的对象作为文件夹形式展现。该对象同样可以上传及下载。用户可以在 OSS 控制台中，采用同 Windows 文件夹的基本操作使用 OSS 文件夹。

搜索文件

本部分介绍如何使用阿里云 OSS 管理控制台在存储空间或文件夹中搜索具有相同的名称前缀的对象。

按名称前缀进行搜索时，搜索字符串将区分大小写，并且不能包含正斜杠“/”字符。搜索的范围限定为当前存储空间的根级别或当前文件夹中的对象（不包括子文件夹及其所包含的对象）。有关阿里云 OSS 如何使用正斜杠“/”字符的信息，

删除文件

如果您不再需要存储所上传的文件，请将其删除以免进一步产生费用。您可以通过 OSS 控制台删除单个文件或批量删除文件。

***注意：***

* + *文件删除后无法恢复，请谨慎操作。*
  + *控制台批量删除文件，限制数量为****1000****个文件。*
  + *如果想更灵活的选择删除的文件，实现更大批量的删除，请参见*[*阿里云 OSS 开发人员指南*](https://help.aliyun.com/document_detail/31824.html)*中的对应章节。*
  + 删除文件夹
  + 使用控制台删除文件夹时，会自动删除此文件夹中包含的所有文件或文件夹。如果您想保留其中的文件，必须先将它们移到别处再删除该文件夹。
  + 文件统计

您可以通过[OSS管理控制台](https://oss.console.aliyun.com/overview)查看您的资源监控报表。OSS资源监控报表有以下功能：

* + 基于Bucket进行PV，UV，流量，ISP分布，地区分布统计。
  + 基于Object，IP，referer访问量TOP排名榜。
  + 返回码统计，爬虫收录统计等。

**阿里云云计算专业认证考试（ACP）（CDN）**

产品概述

**多场景的业务支持，多组件配合服务**

* + 视频流媒体直播服务，媒资存储、切片转码、访问鉴权、内容分发加速一体化解决方案(即将上线)
  + 视音频渐进式点播服务，低缓冲时间，高流畅度播放体验，支持 mp4、flv 视频格式
  + 支持资源链接鉴权，可自定义鉴权KEY，保障您的媒体资源安全，免去盗链担忧
  + 自由集成阿里云多种云服务，系统无缝配合提升云端资源访问下载速度
  + 天然无缝配合对象存储OSS使用，提高网站访问速度，有效降低OSS的外网流量费用
  + 结合云服务器ECS使用，提高网站可用性，保护服务器源站信息，降低带宽使用成本
  + 也可使用负载均衡做为源站地址回源，降低回源带宽压力
  + 同时也支持非阿里云源站，对接无门槛，经过资源审核即可快速部署加速服务
  + 名词解释

**域名**

域名是Internet网络上的一个服务器或一个网络系统的名字，全世界，没有重复的域名

**CNAME记录**

它是一个别名记录( Canonical Name )；当 DNS 系统在查询 CNAME 左面的名称的时候，都会转向 CNAME 右面的名称再进行查询，一直追踪到最后的 PTR 或 A 名称，成功查询后才会做出回应，否则失败。

**CNAME域名**

CDN的域名加速需要用到CNAME记录，在阿里云控制台配置完成CDN加速后，您会得到一个加速后的域名，称之为CNAME域名（该域名一定是\*.\*kunlun.com）， 用户需要将自己的域名作CNAME指向这个\*.\*kunlun.com的域名后，域名解析的工作就正式转向阿里云，该域名所有的请求都将转向阿里云CDN的节点。

**DNS**

DNS即Domain Name System，是域名解析服务的意思。它在互联网的作用是：把域名转换成为网络可以识别的ip地址。人们习惯记忆域名，但机器间互相只认IP地址，域名与IP地址之间是一一对应的，它们之间的转换工作称为域名解析，域名解析需要由专门的域名解析服务器来完成，整个过程是自动进行的。

比如：上网时输入的www.baidu.com会自动转换成为220.181.112.143

**边缘节点**

也称CDN节点、Cache节点等；是相对于网络的复杂结构而提出的一个概念，指距离最终用户接入具有较少的中间环节的网络节点，对最终接入用户有较好的响应能力和连接速度。其作用是将访问量较大的网页内容和对象保存在服务器前端的专用cache设备上，以此来提高网站访问的速度和质量。

产品优势

**稳定快速**

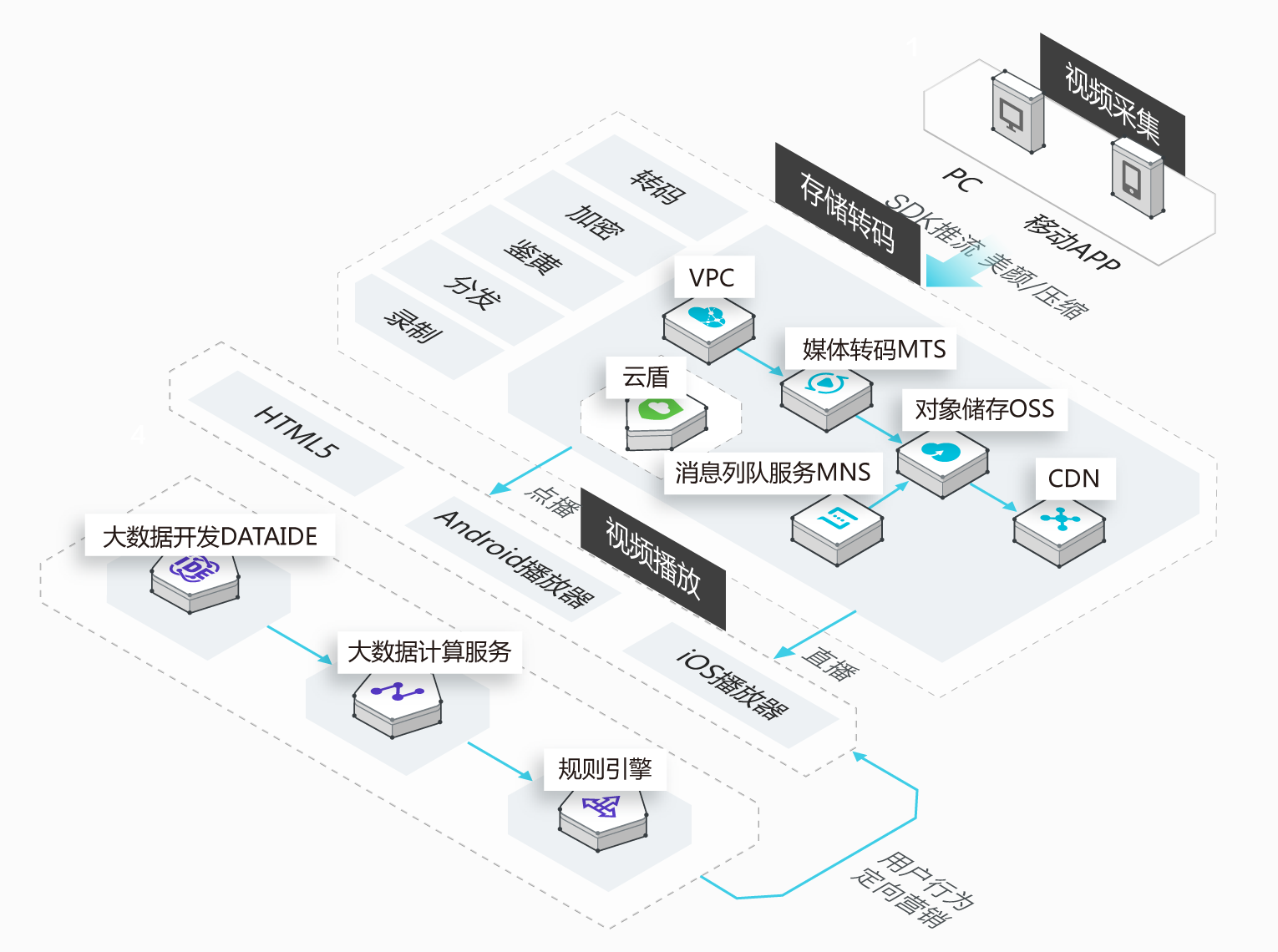
* + 先进的分布式系统架构：国内 700+ 节点，海外 300+ 节点
  + 充足的带宽、存储资源：单节点带宽 40Gbps+，全网带宽输出能力 40Tbps
  + 稳定高效的性能指标：95%+命中率，ms级响应时间，视频95%+流畅率
  + 完善的监控体系&服务体系：7 ＊ 24 小时全网监控，基于服务质量智能监控和调度
  + 使用场景
  + 站点或者应用中大量静态资源的加速分发，建议将站点内容进行动静分离，动态文件可以结合云服务器ECS，静态资源如各类型图片、html、css、js文件等，建议结合 对象存储OSS 存储海量静态资源，可以有效加速内容加载速度，轻松搞定网站图片、短视频等内容分发
  +  用 户 
    静 态 数 据 返 回 
    用 户 动 态 
    动 态 数 据 
    静 请 
    态 求 
    用

**视音频点播/大文件下载分发加速**

支持各类文件的下载、分发，支持在线点播加速业务，如mp4、flv视频文件或者平均单个文件大小在20M以上，主要的业务场景是视音频点播、大文件下载（如安装包下载）等，建议搭配对象存储OSS使用，可提升回源速度，节约近2/3回源带宽成本。

**视频直播加速**

阿里云一站式 [视频直播服务](http://https/cn.aliyun.com/product/live) 服务已正式上线发布！基于领先的内容接入与分发网络和大规模分布式实时转码技术打造的音视频直播平台，提供便捷接入、高清流畅、低延迟、高并发的音视频直播服务。支持多直播场景，全景数据统计，丰富角度分析；直播功能丰富，录制回放、实时封面、实时转码、连麦混流。



**移动应用加速**

移动APP更新文件（apk文件）分发，移动APP内图片、页面、短视频、UGC等内容的优化加速分发。提供httpDNS服务，避免DNS劫持并获得实时精确的DNS解析结果，有效缩短用户访问时间，提升用户体验。

算机生成了可选文字:


产品优势

**阿里云的带宽处理能力**

阿里云CDN目前单节点的服务能力是 40Gbps+；总处理能力为 40Tbps。

**阿里云CDN目前在国内多少个节点，是否有海外节点**

阿里云CDN国内节点 700+，会覆盖到电信，联通，移动，铁通，教育，长宽、各运营商；单节点的服务能力是 40Gbps+；海外节点 300+，部署在全球60多个国家和地区，覆盖六大洲。

**CDN的使用场景都有哪些**

访问量大的网站、如：在线视频、游戏、图片、音频、社交、电商、下载站等。CDN就是将内容缓存在离用户更近的节点上，以此来提高用户体验；所以CDN适用于具有一定量级的静态资源访问，包括html，js，css，apk，mp3，flv，jpg，gif、mp4、flv等所有的静态资源。

**如何使用阿里云CDN可以使加速效果达到最优**

结合阿里云其他产品使用，能够优化回源速度，进一步优化加速效果。

* + 天然无缝配合OSS使用，提高网站访问速度，有效降低OSS的外网流量费用。
  + 结合ECS使用，提高网站可用性，保护服务器源站信息，降低带宽使用成本。PS：可使用SLB做为源站地址回源。

计费相关

**CDN的计费方式**

阿里云CDN主要有两种计费方式供用户选择，分别是按流量小时结算和按峰值带宽计费按日结算

**如何选择合适的计费方式？**

您可以根据自己的业务情况选择更适合您更节省成本的计费方式。

* + 估算下自己业务的带宽利用率，主要参考每日的流量及峰值带宽。计算出自己的带宽利用率。带宽利用率高的用户适合使用按峰值带宽计费，反之则适合使用按流量计费。
  + 大致选择标准为：带宽利用率大于30%的用户适合使用按带宽计费（因计费价格为阶梯计价，实际计算会略有不同，以实际计算结果为准）。

*注：*

* + *带宽利用率=实际使用流量/峰值带宽*
  + *100%利用率产生的流量，1Mbps带宽每日100%利用率产生的流量约为10.54GB*

**是否可以切换计费模式？**

支持用户在按流量计费和按带宽峰值计费两种计费方式间切换；您可在CDN产品控制台直接操作变更。

*注：在一个自然日内，限制用户变更计费方式次数为1次，变更和计费将在第二日零点后生效；例：用户在今天6月16日的10:00提交了变更计费方式那么该变配申请将在明天6月17日00:00后生效。*

**发生账户欠费后，CDN会停止服务么？**

* 1. 当您未付清CDN服务产生的账单，则服务处于欠费状态
  2. 服务欠费后延时24小时停服，在欠费后24小时内会以短信/邮件的方式提醒用户尽快支付账单，在欠费后24小时内进行充值，您的服务将不会受到停服影响；
  3. 如您未在欠费后24小时内未能及时充值。CDN服务将停止服务；停止服务后，CDN也将停止计费。您所占用的Cache资源将被释放，配置信息保留12个月。

**仅停用CDN加速域名，仍会有费用产生么？**

* 1. 停用CDN加速域名后，系统会保留该域名的配置信息，但是边缘节点停止服务，访问CDN节点会直接报错，因此不会产生其他费用。
  2. 删除加速域名后，系统将不会保留该域名任何配置信息。

加速内容

**CDN支持的文件类型有哪些？**

目前支持所有格式的文件，针对不同的文件可以选择不同的加速方式，目前主要提供四种分发方式：

* 1. 图片及小文件分发
  2. 大文件下载加速
  3. 视频流媒体加速
  4. 直播流媒体加速

**源站域名可以和加速域名一致么？**

不可以 源站类型为域名时，源站域名不允许和加速域名相同；若用户请求某资源，该CDN节点上没有缓存相应的内容，CDN节点则会回到源站获取，然后再返回给用户，若加速域名与源站域名一致，会导致请求反复解析到CDN节点上，CDN节点无法回源拉取内容。因此建议，如果加速域名为:example.aliyun.com，源站可以考虑 src.example.aliyun.com，以做区分。

**针对动态文件可以进行加速分发么？**

如果加速域名下面的文件类型多为动态文件，强烈建议做动静分离；动态内容采用独立域名，不使用CDN加速

**采用CDN服务对源站点是否需要改造？**

基本不需要，但建议用户先做动静分离，以达到更优的加速效果，动态文件和静态文件放在不同的域名下面，静态资源加速。

**CDN是对网站所在的服务器加速，还是对域名加速？**

是针对某个域名下面加速的，如果该域名完成了加速配置，那么这个域名下面的文件的访问都会通过CDN

缓存相关

**CDN文件的缓存时间是多久？**

* 1. 文件的缓存时间，是指文件在浏览器中的缓存时间周期。
  2. CDN 缓存服务器严格遵守http的标准协议，具体的缓存时间由http 响应头里面的 Cache-Control 和 Expires 响应头控制
  3. Html的文件缓存时间查看方式：查看http消息头里面的 Cache-Control ，如“Cache-Control max-age=2592000（秒）”，表示文件将会缓存30天。这时除非使用手动刷新，否则新开的浏览器页面在文件缓存周期内都不会去回源重新获取文件。

**是否支持源站的Cache-Control设置？**

支持，CDN的cache软件本身是遵循http协议的，所以具体的缓存时间由用户的文件header头来控制的，由 Cache-Control 或 Expires 来控制的。

*注意：变更缓存配置规则后，新规则不会立即生效，在“原有缓存配置”时间失效后，新的配置会自动生效；若要立即生效，需要手动刷新，注意：目录刷新请仔细评估资源回源量，以免造成源站压力*

**缓存刷新操作有限制么？**

目前支持单条URL刷新和目录刷新两种方式，其中单条URL刷新生效时间在5分钟，每天最多刷新2000条URL；目录刷新的生效时间大概在5分钟，每天支持100次目录刷新操作

**缓存刷新、缓存预热的区别和使用场景是什么？**

* 1. 缓存刷新：强制将分发节点上缓存的资源标记为过期，当用户再次对该资源发起请求时，节点会回源拉取资源，并缓存一份更新后的资源在分发节点
  2. 缓存预热：主动触发将源站资源推送到边缘节点，用户访问资源时，可以直接命中缓存，缓解突增回源流量给源站造成的压力

性能相关

**如何判断请求是否命中？**

回包中的哪些数据证明是命中了CDN缓存,可以查看Response Headers中关于X-Cache的信息

* 1. 若：X-Cache:HIT TCP\_MEM\_HIT 表示命中缓存
  2. 若：X-Cache:MISS TCP\_MISS 则表示未命中缓存
  3. **命中率如果较低的原因是什么？**
  4. 网站访问量较低，文件热度不够，CDN收到请求较少无法有效命中缓存
  5. 缓存配置不合理，针对某些文件设置较短，导致CDN节点频繁回源
  6. HTTP Header设置导致无法缓存，需要用户检查源站的 Cache-Control 设置或者 Expires 的设置
  7. 源站动态资源较多，多为不可缓存的内容，也会导致频繁回源拉取

**CDN什么情况下会回源拉取资源？**

* 1. 用户访问时，如节点上无缓存，则会回源拉取资源
  2. CDN节点上的文件超时过期，会回源拉取资源
  3. 若为不缓存文件，用户访问时，会直接回源

**CDN回源失败的场景有哪些？**

* 1. 源站带宽或计算过载，导致无法正常回源
  2. 源站存储的资源有问题，不能正常回源访问资源
  3. 可以考虑通过设置缓存时间的方式，降低回源率，并对源站进行升级扩容

功能相关

**是否支持HTTPS加速？**

支持，

**是否支持泛域名加速？**

目前图片小文件加速、大文件下载加速、视音频点播加速均支持泛域名添加；直播业务和HTTPS安全加速暂不支持泛域名。

故障检测

**加速域名无法访问或者访问速度慢，如何定位故障？**

* 1. 测试您网站是否可以访问
  2. 请您在访问异常的地区提供以下测试结果：
  + *访问*[*http://tool.cdnpe.com/doc.html*](http://tool.cdnpe.com/doc.html)
  + *ping 您的加速域名*
  + *tracert -d 您的加速域名*

**使用CDN，通常的故障状况有那些？**

* + 本地网络异常
  + CDN节点异常
  + 缓存配置错误
  + 源站异常
  + 源站采集异常

**使用CDN，第一次回源拉数据访问是否会出现慢的情况？**

会比直接回源相对还慢些，因为第一次缓存Cache上没有数据，要回源取数据，需要经过两层网络，分别是CDN的边缘层和中间层。

常见FAQ

**CDN的源站地址和回源Host的区别。**

源站地址：源站决定了回源时，请求到哪个IP。

回源host：源站上可能存在若干Web站点，回源host决定回源请求访问到哪个站点。

*示例1：源站是域名源站为www.a.com*

*回源host为www.b.com*

*那么实际回源是请求到www.a.com解析到的IP，对应的主机上的站点www.b.com*

*示例2：源站是IP*

*源站为1.1.1.1 回源host为www.b.com*

*那么实际回源的是1.1.1.1对应的主机上的 站点www.b.com*

**CDN的回源端口如何设置，是否支持自定义端口。**

目前只支持80和443端口，不支持自定义端口。

**协议跟随回源在什么样的场景需要开启？对源站有什么要求？**

当需要CDN回源时使用和客户端访问资源时一样的协议时（即客户端使用HTTPS请求资源，CDN节点也使用HTTPS回源；客户端使用HTTP请求资源，CDN节点也用HTTP回源），则需要开启协议跟随回源。（注：开启协议回源后，在CDN控制台上设置的固定回源端口将失效。）

*对源站的要求是：源站需要同时支持 80 端口和 443 端口。*

**CDN是否支持内网回源？**

不支持，CDN回源全部走公网。

**CDN资源加载速度较慢如何优化？**

* + 首次访问资源，会比直接访问源站相对还慢些，因为第一次CDN节点没有缓存，要回源取数据。此情况推荐使用【预热】功能，将源站的内容主动预热到CDN节点上，用户首次访问可直接命中缓存，提高加载速度。
  + 非首次访问仍然加载速度慢，可针对以下几点进行排查:
  + 资源访问量较低，文件热度不够，CDN收到请求较少无法有效命中缓存。
  + 缓存配置不合理，缓存时间过短，CDN节点频繁回源。
  + HTTP Header设置导致无法缓存，建议检查源站的Cache-Control设置或者Expires 的设置。
  + 访问资源的URL带参，并且参数不断变化，当用不同的URL去访问CDN的时候，CDN会认为这是一个新请求（即便这两个不同的URL其实是访问到了同一个文件，并且该文件已经缓存在节点上），会回源去拉取所请求的内容，建议开启【过滤参数】功能。
  + 强烈建议做动静分离，动态文件和静态文件放在不同的域名下面，只加速静态资源。
  + 大文件建议开启range回源。

**CDN有哪些安全防护功能可以避免资源被恶意访问？**

目前有Refer防盗链，鉴权配置，IP黑名单等访问控制功能，防止盗链或者恶意请求造成不必要流量损失。

**CDN的智能压缩对哪些特定的文件生效？**

当前支持的压缩类型有： text/xml text/plain text/css application/javascript application/x-javascript application/rss+xml text/javascript image/tiff image/svg+xml application/json application/xml

*注：不支持IE6，且原始资源大小要超过1KB，且控制台要把gzip压缩功能打开。*

产品限制

所有接入CDN的域名都要经过审核，发现以下任意一种情况，CDN目前不支持接入的加速域名类型包括，

* + 无法正常访问或内容不含有任何实质信息
  + 游戏私服类
  + 传奇类游戏、纸牌类游戏
  + 盗版软件等无版权下载网站
  + P2P类网站
  + 彩票类网站
  + 违规医院和药品类网站
  + 涉黄、涉毒、涉赌等
  + 类型1：图片小文件加速

**应用场景介绍**

网站或者应用的静态内容分发，例如各种类型的图像文件，html文件、flash动画、css、javascript文件等。适用于各种门户网站、电子商务类网站、新闻资讯类站点或应用、政府/企业官网站点、娱乐游戏类站点或应用等。

类型2：大文件下载加速

**应用场景介绍**

网站或者应用的静态大文件分发，例如游戏安装包.apk文件、应用更新文件.rar、补丁程序文件、音视频文件等相对较大的文件。适用于下载类站点和音视频的应用

类型3：视音频点播加速

**应用场景介绍**

各类视音频站点，如影视类视频网站、在线教育类视频网站、新闻类视频站点、短视频社交类网站以及音频类相关站点和应用。

类型4：直播设置说明

**直播流媒体加速**

**应用场景介绍**

为各类视频直播平台提供高性能稳定直播技术支持，如交互性在线教育网站、游戏竞技类直播站点、个人秀场直播、事件类和垂直行业的直播平台等。当前支持RTMP，HLS，FLV三种格式直播内容加速

移动加速，概述

**功能特性**

移动加速服务将主要通过以下几种技术手段来实现移动应用网络加速：

* + 协议优化：采用深度优化定制的私有协议替换传统的HTTP协议，收获多路复用、请求头压缩、请求优先级支持等收益，防止内容劫持现象发生。同时我们也为云加速服务终端与加速节点间长连复用，最小化TCP的建连开销，提高连接利用率和请求响应速度；
  + 链路优化：以阿里云遍布全国的优质边缘节点，海量的带宽资源为基础设施，结合HTTPDNS智能路由精准的调度，实现就近、就快接入；云加速节点会缓存热点内容，大大提高访问效率；云加速节点和ECS间搭建专线进行链路加速，如果您已经在使用阿里云ECS作为服务后端，加速结果更是锦上添花；
  + 当前功能包括：HTTPDNS和HTTP的动态和静态加速