

ALIBABA CLOUD

# 阿里云

负载均衡  
快速入门

文档版本：20210127

 阿里云

## 法律声明

阿里云提醒您在使用或阅读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
 危险	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。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会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，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	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。	单击设置> 网络> 设置网络类型。
<b>粗体</b>	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[ ] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

# 目录

1.入门概述	05
2.准备工作	06
3.创建负载均衡实例	09
4.配置负载均衡实例	11
5.域名解析	14
6.释放负载均衡实例	15

# 1.入门概述

本教程指引您快速创建一个公网负载均衡实例，将来自客户端的请求转发到两台后端ECS上。

 **说明** 在开始搭建负载均衡服务前，您需要确定负载均衡实例的地域、类型、付费模式等配置，更多信息，请参见[准备工作](#)。

本教程包含以下操作：

## 1. 创建负载均衡实例

创建负载均衡实例。负载均衡实例是一个运行的负载均衡服务实体。

## 2. 添加监听和后端服务器

配置负载均衡实例，添加监听规则和后端服务器。

## 3. 域名解析。（可选）

使用云解析服务将一个域名解析到负载均衡的服务地址上，提供服务。

## 4. 释放负载均衡实例

如果您不需要负载均衡服务了，为避免不必要的计费，可以将其删除。

## 视频教程

[负载均衡快速入门](#)

## 2.准备工作

在使用负载均衡前，您需要根据您的业务确定负载均衡的监听类型和网络类型等。

### 规划实例地域

在选择地域时，请注意：

- 为了减少延迟并提高下载速度，建议选择离您客户最近的地域。
- 为了提供更加稳定可靠的服务，阿里云负载均衡已在大部分地域提供主备可用区，实现同地域下的跨机房容灾。建议您选择提供主备可用区的地域。
- 由于负载均衡不支持跨地域部署，因此应选择与后端ECS实例相同的地域。

### 选择实例的网络类型（公网或私网）

负载均衡提供面向公网和内网的负载均衡服务：

- 如果您需要使用负载均衡分发来自公网的请求，选择创建公网负载均衡实例。

公网负载均衡实例提供一个公网IP，用来接收来自Internet的请求。

对于公网负载均衡实例，您还需要选择实例的计费方式：

- 按流量计费：适用于波峰波谷效应明显的业务。
  - 按带宽计费：适用于带宽较为平稳的业务。
- 如果您需要使用负载均衡分发来自内网的请求，选择创建私网负载均衡实例。

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。

### 选择实例规格

负载均衡于2018年4月1日推出了性能保障型实例，您可以独享已购实例的资源，更好地保障服务的可用性。负载均衡提供六种实例规格供您选择：

- 对于按量付费实例，建议您直接选择可以买到的最大规格，这样可以保证业务的灵活性（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）。
- 对于预付费实例，您需要评估您的实际业务量，然后选择一个较合适的规格，对于业务量评估来说，主要参考下面几个原则：
  - 如果是四层监听，关注的重点是长连接的并发连接数，那么最大（并发）连接数应当作为一个关键指标来参考。根据不同的业务场景，您需要预估一个负载均衡实例需要承载的最大并发连接数，并选择相应的规格。
  - 如果是七层监听，关注的重点是QPS的性能，QPS决定了一个七层应用系统的吞吐量。同样，您也需要根据经验对QPS进行预估。在初步选定一个规格后，在业务压测和实测过程中对规格进行微调。
  - 结合与性能保障型实例一起推出的其它关键监控指标，查看实际业务流量的走势、峰值情况，对性能规格进行更加精确的选取。

### 选择协议类型

阿里云提供基于四层（TCP协议和UDP协议）和七层（HTTP协议和HTTPS协议）的应用的负载均衡：

- 四层监听将请求直接转发给后端服务器，不会修改报头。客户端请求到达负载均衡监听后，负载均衡服务器会使用监听中配置的后端端口与后端服务器建立TCP连接。

- 七层监听原理上是反向代理的一种实现。客户端请求到达负载均衡监听后，负载均衡服务器会通过与后端服务器建立TCP连接，即再次通过新TCP连接HTTP协议访问后端，而不是直接转发报文到后端ECS。

由于七层监听比四层监听在底层实现上多了一个T engine处理环节，因此，七层监听性能没有四层好。此外，客户端端口不足、后端服务器连接过多等场景也可能导致七层服务性能不高，如果您对性能有很高的要求，建议您使用四层监听。

更多详细信息，参见[协议说明](#)。

## 准备后端服务器

在使用负载均衡服务前，您需要创建ECS实例并部署相关应用，然后将ECS实例添加到负载均衡实例中来处理转发的客户端请求。

创建ECS时，请注意：

- ECS实例的地域和可用区

确保ECS实例的地域和负载均衡实例的地域相同。此外，建议您将ECS部署在不同的可用区内，提高可用性，ECS实例的创建详情请参见[使用向导创建实例](#)。

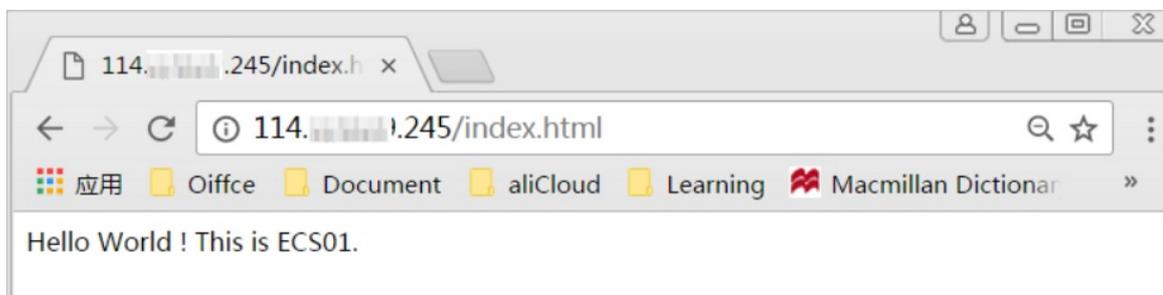
本教程中，在华东1地域创建了两个ECS实例，为了便于辨识，将实例分别命名为ECS01和ECS02，如下图所示。

实例ID/名称	标签	监控	IP地址	状态	网络类型	配置	付费方式	操作
i-bp1hv7 ECS01			114.114.114.114	运行中	专有网络	2 vCPU 8 GB (I/O优化) ecs.g5.large 5Mbps (峰值)	按量 2018年11月13日 17:00 创建	<a href="#">管理</a>   <a href="#">远程连接</a> <a href="#">更改实例规格</a>   <a href="#">更多</a>
i-bp1car ECS02			114.114.114.114	运行中	专有网络	2 vCPU 8 GB (I/O优化) ecs.g5.large 5Mbps (峰值)	按量 2018年11月13日 16:58 创建	<a href="#">管理</a>   <a href="#">远程连接</a> <a href="#">更改实例规格</a>   <a href="#">更多</a>

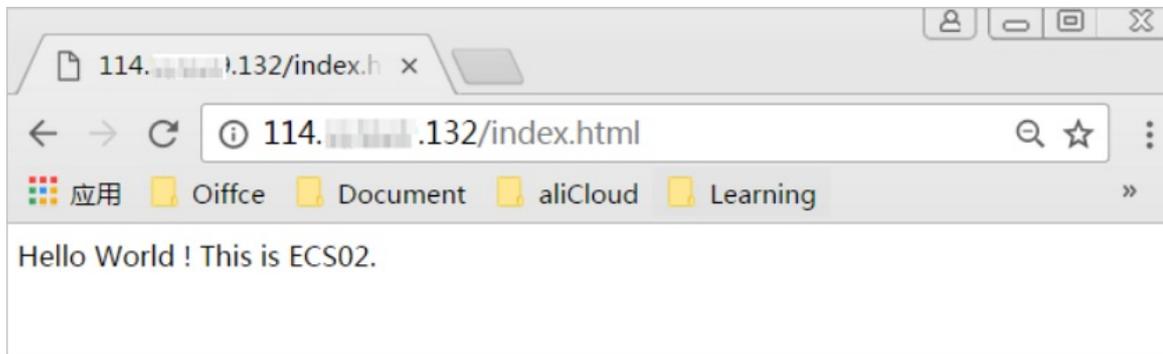
- 应用配置

本教程中，分别在ECS01和ECS02两个实例上使用Apache搭建了两个静态网页，如下图所示。

- 在浏览器中输入ECS01实例绑定的弹性公网IP地址：



- 在浏览器中输入ECS02实例绑定的弹性公网IP地址：



在ECS上部署好应用后，不需要再进行特别的配置。但如果您要配置一个四层监听（TCP协议或UDP协议），并且ECS使用的是Linux系统，确保ECS实例上/etc/sysctl.conf目录下net.ipv4.conf文件中的以下三个参数的值为零：

```
net.ipv4.conf.default.rp_filter = 0  
net.ipv4.conf.all.rp_filter = 0  
net.ipv4.conf.eth0.rp_filter = 0
```

# 3.创建负载均衡实例

本文介绍如何创建一个公网负载均衡实例。负载均衡实例创建后，系统会自动分配一个服务地址，您可以将您的域名解析到该负载均衡实例的服务地址。

## 背景信息

一个负载均衡实例可以添加多个监听和后端服务器。

## 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 > 实例管理。
3. 在实例管理页面，单击创建负载均衡。
4. 在负载均衡SLB购买页面，根据需要配置负载均衡实例。本教程中的负载均衡实例配置如下：

负载均衡SLB



- **付费模式**：选择一种付费模式。本示例中选择**按量付费**。
- **地域和可用区**：由于负载均衡不支持跨地域部署，因此创建负载均衡实例时应选择与ECS实例相同的地域。本教程选择**华北1（青岛）**。
- **可用区类型**：负载均衡在大部分地域都开通了多可用区功能，实现同城容灾。当主可用区的机房故障、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的机房恢复服务能力；当主可用区恢复时，负载均衡同样会自动切换到主可用区的机房提供服务。

本教程中的主可用区为华北1 可用区C，备可用区为华北1 可用区B。

- **实例名称**：输入实例名称，或者使用系统自动创建的实例名称。
  - **实例规格**：选择**简约型(slb.s1.small)**。该规格最大可以支持连接数为5000，新建连接数（CPS）为3000，每秒查询数（QPS）为1000。
  - **实例类型**：选择**公网**。
  - **IP版本**：选择**IPv4**。
  - **后端服务器类型**：选择**本地域**，允许挂载本地域后端服务器。
  - **计费方式**：选择**按使用流量计费**。
5. 单击**立即购买**。
6. 在**确认订单**页面，选择**我已阅读并同意《负载均衡SLB（按量付费）服务协议》**，然后单击**去支付**。

您可以将域名解析到负载均衡实例的公网服务地址上，提供服务，详情参考[域名解析](#)。

## 4. 配置负载均衡实例

创建负载均衡实例后，您需要对负载均衡实例进行配置才能进行流量转发，您需要添加至少一个监听和一组后端服务器。本指南指引您配置一个TCP监听并添加部署了静态网页的两个ECS实例（ECS01和ECS02）作为后端服务器。

### 操作步骤

1. 登录[负载均衡管理控制台](#)。
2. 在实例管理页面，单击要配置监听的实例操作列的[监听配置向导](#)。



3. 在[协议&监听](#)页签下，根据如下信息，配置监听规则，其它配置保持默认选项。
  - **选择负载均衡协议：**本教程选择TCP协议。
  - **监听端口：**用来接收请求并向后端服务器进行请求转发的负载均衡系统的前端协议和端口。  
本教程端口设置为80。

#### 高级配置：

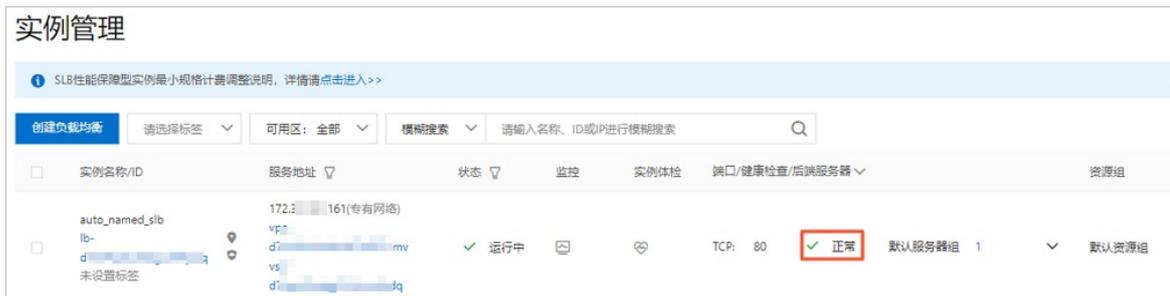
- **开启监听带宽限速：**设定不同的带宽峰值来限定后端ECS实例的不同应用所能对外提供的服务能力。  
本教程创建的公网负载均衡实例是按流量计费的，不受带宽峰值限制，所以不进行配置。
- **调度算法：**负载均衡支持如下三种调度算法，本教程选择**轮询**。
  - **加权轮询模式：**将访问请求依序分发后端服务器，后端服务器的权重越高，被分发的几率也越大。
  - **加权最小连接数模式：**除了根据每台后端服务器设定的权重值来进行轮询，同时还考虑后端服务器的实例连接数。当权重值相同时，当前连接数越小的后端服务器被轮询到的次数也越高。
  - **轮询模式：**按照访问顺序将访问请求分发给后端服务器。



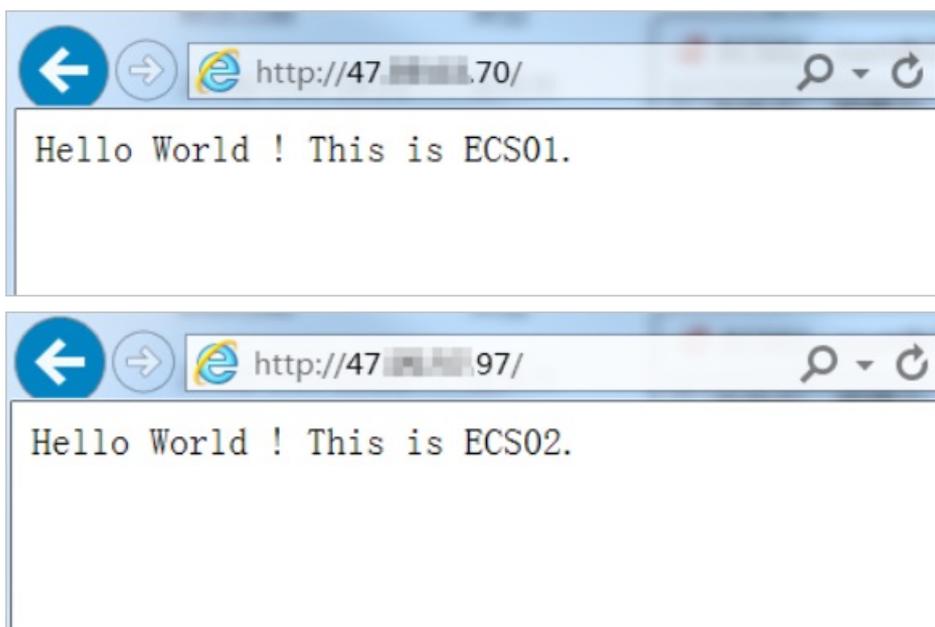


7. 单击知道了，返回实例管理页面，单击 。

当后端ECS的健康检查状态为正常时，表示后端ECS可以正常处理负载均衡转发的请求了。



8. 在浏览器中输入负载均衡实例的服务地址，测试负载均衡服务。



## 5. 域名解析

域名解析是域名和IP地址相互映射的分布式数据库。阿里云负载均衡支持将域名解析到负载均衡实例的公网服务地址上，使用户更加方便地访问互联网。

### 背景信息

例如，您网站的域名为www.aliyun.com，运行在公网IP地址为1.1.1.1的ECS实例上。创建负载均衡实例后，系统分配的负载均衡实例的公网IP地址为2.2.2.2。您需要将ECS实例添加到负载均衡实例的后端服务器池，并将域名www.aliyun.com解析到2.2.2.2。建议您使用“A记录解析”（即将域名解析到一个IP地址）即可。

### 操作步骤

1. 登录[域名解析控制台](#)。
2. 单击[添加域名](#)，添加您的域名。
3. 在[域名解析](#)页面，单击目标域名操作列下的[解析设置](#)，完成域名解析设置。

## 6. 释放负载均衡实例

您可以根据需要删除负载均衡实例，避免不必要的计费。删除负载均衡实例不会删除后端ECS，也不会影响后端ECS的运行。

### 背景信息

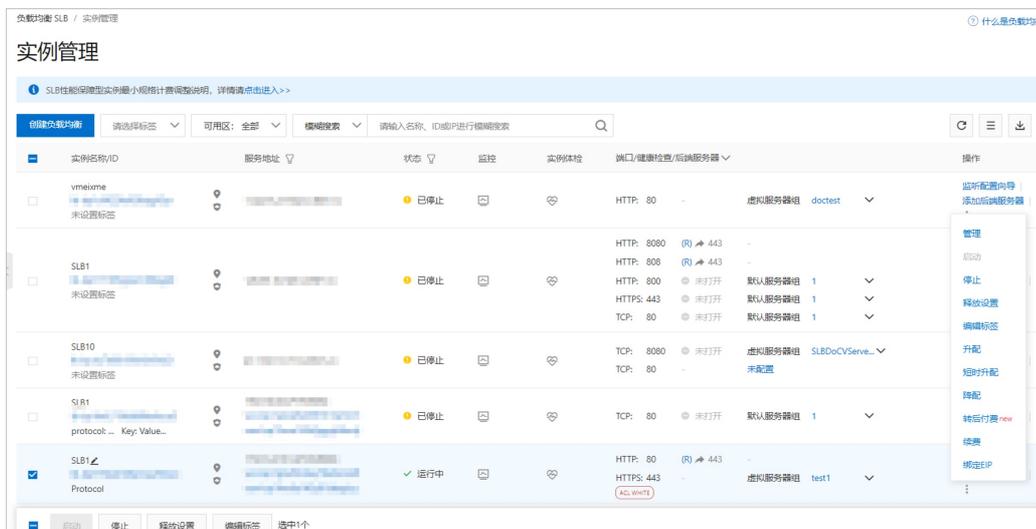
#### 说明

- 如果您已经将一个域名解析到负载均衡实例的公网服务地址，在释放负载均衡实例之前，需要先将该域名解析到另外一个负载均衡实例的公网服务地址，避免业务中断。
- 只能释放按量计费的负载均衡实例。包年包月实例在到期后若未按时续费会自动释放。
- 释放负载均衡实例不会影响后端ECS的运行。您可以根据需要释放ECS。

### 操作步骤

1. 登录[负载均衡管理控制台](#)。
2. 在左侧导航栏，选择实例 > 实例管理。
3. 在实例管理页面，选择目标实例操作列的  > 释放设置，或者单击列表底部的释放设置。

#### 释放设置



4. 在释放设置页面，根据需要选择立即释放或者定时释放。

如果您选择定时释放，设置自动释放的时间。

**说明** 系统执行释放时间是每整点和每半点，但是系统会按照您设置的释放时间停止计费。

5. 单击下一步。
6. 单击确定，完成负载均衡实例的释放。

**说明** 按量计费（后付费）实例释放不会进入回收站，请谨慎操作。