ALIBABA CLOUD

# Alibaba Cloud

## 负载均衡
## 教程专区

文档版本：20201013

阿里云

# 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 |
| --- | --- | --- |
| 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<br>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<br>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<br>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<br>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**。 |
| **粗体** | 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 `cd /d C:/window` 命令，进入Windows系统文件夹。 |
| *斜体* | 表示参数、变量。 | `bae log list --instanceid`<br>*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。 | `switch {active\|stand}` |

# 目录

# 1.负载均衡快速入门

本教程介绍什么是负载均衡以及配置和使用负载均衡的操作步骤，通过视频的方式直观的指导您如何通过阿里云负载均衡将流量分发给后端服务器。

## 相关文档

- 什么是负载均衡
- 入门概述

# 2.使用SLB部署HTTPS业务（单向认证）

要配置HTTPS单向认证的监听，您仅需要在配置监听时上传服务器证书。

## 步骤一：上传服务器证书

在配置HTTPS监听（单向认证）前，您需要购买服务器证书，并将服务器证书上传到负载均衡的证书管理系统。上传后，无需在后端ECS上进行其它证书配置。

1. 登录负载均衡管理控制台。
2. 在左侧导航栏，选择**证书管理**，单击**创建证书**。
3. 单击**上传第三方签发证书**。
4. 按照以下信息，配置证书：
   - 证书名称：长度限制为1~80个字符，只允许包含字母、数字、短横线（-）、正斜杠（/）、点号（.）、下划线（_）和注释符（*）。
   - 证书部署地域：选择**华东1（杭州）**。

     > 说明 证书的地域和负载均衡实例的地域要相同。

   - 证书类型：选择**服务器证书**。
   - 证书内容和私钥：复制服务器证书的内容和私钥。单击**导入样例**查看合法的证书格式。上传的证书必须是PEM格式，详情请参见证书要求。
5. 单击**确定**，完成上传。

## 步骤二：配置负载均衡实例

1. 登录负载均衡管理控制台。
2. 在**实例管理**页面，单击**创建负载均衡**。
3. 配置负载均衡实例，单击**立即购买**完成支付。 实例类型选择**公网**，地域选择**华东1（杭州）**。详细配置信息请参见创建负载均衡实例。
4. 创建成功后，返回**实例管理**页面，选择**华东1（杭州）**地域。
5. 单击已创建的负载均衡实例ID链接，或者直接单击**监听设置向导**。
6. 在**监听**页签下，单击**添加监听**。
7. 在**协议&监听**页签下，完成如下配置。
   - **选择负载均衡协议**：HTTPS。
   - **监听端口**：443。
   - **调度算法**：轮询（RR）。
8. 单击**下一步**，在**SSL证书**页签下，选择已经上传的服务器证书和TLS安全策略。
9. 单击**下一步**，选择**默认服务器组**，单击**继续添加**，添加ECS服务器，后端协议监听端口设置为80。
10. 其他参数保持默认值，单击**下一步**至**确定**，完成负载均衡实例配置。

## 步骤三：测试负载均衡服务

1. 负载均衡实例配置完成后，在**实例管理**页面，查看健康检查状态。

   当状态为**正常**时，表示后端服务器可以正常接收处理负载均衡监听转发的请求。

2. 在浏览器中输入负载均衡的公网服务地址。

# 3.使用SLB部署HTTPS业务（双向认证）

要配置HTTPS双向认证的监听，您需要在配置监听时上传服务器证书和CA证书。

本指南中使用自签名的CA证书为客户端证书签名，完成以下操作配置HTTPS监听（双向认证）：

1. 准备服务器证书
2. 使用OpenSSL生成CA证书
3. 生成客户端证书
4. 上传服务器证书和CA证书
5. 安装客户端证书
6. 配置负载均衡双向认证监听
7. 测试负载均衡服务

## 步骤一：准备服务器证书

服务器证书用于用户浏览器检查服务器发送的证书是否是由自己信赖的中心签发的，服务器证书可以到阿里云云盾证书服务购买，也可以到其他服务商处购买。

## 步骤二：使用OpenSSL生成CA证书

1. 运行以下命令在*/root*目录下新建一个*ca*文件夹，并在*ca*文件夹下创建四个子文件夹。

   ```
   sudo mkdir ca
   cd ca
   sudo mkdir newcerts private conf server
   ```

   其中：

   - *newcerts*目录：用于存放CA签署过的数字证书（证书备份目录）。
   - *private*目录：用于存放CA的私钥。
   - *conf*目录：用于存放一些简化参数用的配置文件。
   - *server*目录：存放服务器证书文件。
2. 在*conf*目录下新建一个包含如下信息的*openssl.conf*文件。

```
[ ca ]
 default_ca = foo
 [ foo ]
 dir = /root/ca
 database = /root/ca/index.txt
 new_certs_dir = /root/ca/newcerts
 certificate = /root/ca/private/ca.crt
 serial = /root/ca/serial
 private_key = /root/ca/private/ca.key
 RANDFILE = /root/ca/private/.rand
 default_days = 365
 default_crl_days= 30
 default_md = md5
 unique_subject = no
 policy = policy_any
 [ policy_any ]
 countryName = match
 stateOrProvinceName = match
 organizationName = match
 organizationalUnitName = match
 localityName = optional
 commonName      = supplied
 emailAddress    = optional
```

3. 运行以下命令生成私钥key文件。

```
cd /root/ca
 sudo openssl genrsa -out private/ca.key
```

运行结果如下图所示。

4. 运行以下命令并按命令后的示例提供需要输入的信息，然后回车，生成证书请求*csr*文件。

```
sudo openssl req -new -key private/ca.key -out private/ca.csr
```

> **说明** Common Name请输入您的负载均衡服务的域名。

5. 运行以下命令生成凭证*crt*文件。

```
sudo openssl x509 -req -days 365 -in private/ca.csr -signkey private/ca.key -out private/ca.crt
```

6. 运行以下命令为CA的key设置起始序列号，可以是任意四个字符。

```
sudo echo FACE > serial
```

7. 运行以下命令创建CA键库。

```
sudo touch index.txt
```

8. 运行以下命令为移除客户端证书创建一个证书撤销列表。

```
sudo openssl ca -gencrl -out /root/ca/private/ca.crl -crldays 7 -config "/root/ca/conf/openssl.conf"
```

输出为：

```
Using configuration from /root/ca/conf/openssl.conf
```

## 步骤三：生成客户端证书

1. 运行以下命令在*ca*目录内创建一个存放客户端key的目录*users*。

```
sudo mkdir users
```

2. 运行以下命令为客户端创建一个key：

```
sudo openssl genrsa -des3 -out /root/ca/users/client.key 1024
```

> 说明 创建key时要求输入pass phrase，这是当前key的口令，以防止本密钥泄漏后被人盗用。两次输入同一个密码。

3. 运行以下命令为客户端key创建一个证书签名请求*csr*文件。

```
sudo openssl req -new -key /root/ca/users/client.key -out /root/ca/users/client.csr
```

输入该命令后，根据提示输入上一步输入的pass phrase，然后根据提示，提供对应的信息。

> 说明 `A challenge password` 是客户端证书口令（请注意将它和 `client.key` 的口令区分开，本教程设置密码为test），可以与服务器端证书或者根证书口令一致。

4. 运行以下命令使用步骤二中的CA Key为刚才的客户端key签名。

```
sudo openssl ca -in /root/ca/users/client.csr -cert /root/ca/private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/users/client.crt -config "/root/ca/conf/openssl.conf"
```

当出现确认是否签名的提示时，两次都输入*y*。

5. 运行以下命令将证书转换为大多数浏览器都能识别的*PKCS12*文件。

```
sudo openssl pkcs12 -export -clcerts -in /root/ca/users/client.crt -inkey /root/ca/users/client.key -out /root/ca/users/client.p12
```

按照提示输入客户端client.key的pass phrase。

再输入用于导出证书的密码。这个是客户端证书的保护密码，在安装客户端证书时需要输入这个密码。

6. 运行以下命令查看生成的客户端证书。

```
cd users
 ls
```

## 步骤四：上传服务器证书和CA证书

1. 登录负载均衡管理控制台。
2. 在**实例管理**页面，单击**创建负载均衡**。
3. 配置负载均衡实例，单击**立即购买**完成支付。

   本操作中网络类型选择**公网**，地域选择**华东1（杭州）**，详细配置信息请参见创建负载均衡实例。
4. 创建成功后，在**实例管理**页面，将鼠标移至实例名称区域，单击出现的铅笔图标，修改负载均衡实例名称。
5. 在选左侧导航栏，单击**证书管理**页签。
6. 单击**创建证书**。
7. 在**创建证书**页面，完成如下配置后，单击**确定**。
   - **证书部署地域**：本教程中选择**华东1**。

     > **说明** 证书的地域和负载均衡实例的地域要相同。

   - **证书类型**：选择**服务器证书**。
   - **证书内容和私钥**：复制您的服务器证书内容和私钥。

     > **说明** 在复制内容前，您可以单击**导入样式**，查看正确的证书和私钥格式。更多详细信息请参见证书要求。
8. 在负载均衡左侧导航栏，单击**证书管理**，然后单击**创建证书**，上传CA证书。
9. 在**创建证书**页面，完成如下配置后，单击**确定**。
   - **证书部署地域**：本教程中选择**华东1（杭州）**。

     > **说明** 证书的地域和负载均衡实例的地域要相同。

   - **证书类型**：选择**CA证书**。
   - **证书内容**：复制您的CA证书内容。

     > **说明** 在复制内容前，您可以单击**导入样式**，查看正确的证书和私钥格式。更多详细信息请参见证书要求。

## 步骤五：安装客户端证书

将生成的客户端证书安装到客户端。本教程以Windows客户端，IE浏览器为例。

1. 打开Git Bash命令行窗口，运行以下命令导出步骤三中生成的客户端证书。

```
scp root@IPaddress:/root/ca/users/client.p12 ./
```

> 说明 IPaddress是生成客户端证书的服务器的IP地址。

2. 在IE浏览器中导入下载的客户端证书。
   i. 打开IE浏览器，单击**设置** > **Internet选项**。
   ii. 单击**内容**页签，然后单击**证书**，导入下载的客户端证书。在导入证书时需要输入在步骤三时生成*PKCS12*文件的密码。

## 步骤六：配置HTTPS双向认证监听

1. 登录负载均衡管理控制台。
2. 选择**华东1（杭州）**地域，单击已创建的负载均衡实例ID链接，或者单击**监听配置向导**。
3. 选择**监听**页签，单击**添加监听**。
4. 在**协议&监听**页签下，配置监听。
   - **选择负载均衡协议**：HTTPS
   - **监听端口**：443
   - **调度算法**：轮询（RR）
5. 单击**下一步**，在**SSL证书**页签下，配置SSL证书信息，启用双向认证。
   - **服务器证书**：选择已上传的服务器证书。
   - **CA证书**： 选择已上传的CA证书。
6. 单击**下一步**，选择**默认服务器组**页签，单击**添加**，添加ECS服务器，并将后端协议端口设置为80。
7. 单击**下一步**，开启健康检查。
8. 单击**下一步**，查看监听配置信息。
9. 单击**提交**，提交审核。
10. 单击**确定**。

## 步骤七：测试HTTPS双向认证

1. 在**实例管理**页面，查看健康检查状态。当状态为**正常**时，表示后端服务器可以正常接收处理负载均衡监听转发的请求。
2. 在浏览器中，输入负载均衡的公网服务地址，当提示是否信任客户端证书时，选择信任。
3. 刷新浏览器，您可以观察到请求在两台ECS服务器之间转换。

# 4.将HTTP访问重定向至HTTPS

HTTPS是加密数据传输协议，安全性高。负载均衡支持将HTTP访问重定向至HTTPS，方便您进行全站HTTPS部署。负载均衡已经在全部地域开放了HTTP重定向功能。

80转443 重定向 HTTP到HTTPS

## 前提条件

已创建了HTTPS监听，详情参见添加HTTPS监听。

## 背景信息

本教程以将HTTP 80访问重定向转发至HTTPS 443为例。

> 说明 仅负载均衡新版控制台在创建HTTP监听时支持配置监听转发。

## 操作步骤

1. 登录负载均衡管理控制台。
2. 在顶部菜单栏选择负载均衡实例的所属地域。
3. 在**实例管理**页面，单击目标实例的ID链接。
4. 在**监听**页签下，单击**添加监听**。
5. 在**协议&监听**页签下，负载均衡协议选择**HTTP**， 监听端口输入**80**。
6. 单击**高级配置**后的**修改**。
7. 开启**监听转发**，选择目的监听为**HTTPS:443**。
8. 单击**下一步**。
9. 确认后，单击**提交**，然后单击**知道了**。

   转发开启后，所有来自HTTP的访问都会转发至HTTPS，并根据HTTPS的监听配置进行转发。

# 5.单SLB实例配置多域名HTTPS网站（HTTPS多域名）

本教程介绍配置扩展域名的详细操作步骤。

## 场景描述

本教程以华东1（杭州）地域的性能保障型负载均衡实例SLB1为例。在本教程中您会创建一个七层HTTPS监听，认证方式为单向认证，您需要将来自域名为*.example1.com的前端请求转发至虚拟服务器组test1上，将来自域名为www.example2.com的前端请求转发至虚拟服务器组test2上。

您需要完成以下操作：

1. 添加HTTPS监听。
2. 配置转发规则。
3. 添加扩展域名。

## 前提条件

- 在华东1（杭州）地域创建性能保障型实例SLB1，具体操作请参见创建负载均衡实例。
- 上传本教程中需要使用的证书，具体操作请参见概述。
  - 监听使用的默认证书为default。
  - 域名*.example1.com使用的证书为example1。
  - 域名www.example2.com使用的证书为example2。

## 步骤一：添加HTTPS监听

完成以下操作，添加七层HTTPS监听：

1. 在左侧导航栏，选择**实例 > 实例管理**。
2. 在**实例管理**页面，单击性能保障型实例SLB1**操作**列的**监听配置向导**。

   首次配置监听，也可以单击**端口/健康检查/后端服务器**列的**点我开始配置**。
3. 配置监听。

   本操作的主要配置如下，其他配置请参见添加HTTPS监听。
   - 双向认证：关闭。
   - SSL证书：选择服务器证书default。
   - 后端服务器：需要创建test1和test2两个虚拟服务器组。

## 步骤二：配置转发规则

完成以下操作，配置转发规则：

1. 单击SLB1实例ID，进入**实例详情**页面。
2. 在**监听**页签下，找到已创建的HTTPS监听，单击**添加转发策略**。
3. 在**转发策略**页面，配置转发策略，详情请参见基于域名或URL路径进行转发。

   本教程中配置域名转发规则，URL不进行设置。

- 设置规则名称，在域名操作列输入*.example1.com，选择test1虚拟服务器组，单击添加转发策略+。
- 设置规则名称，在域名操作列输入www.example2.com，选择test2虚拟服务器组，单击确认。

> 说明 转发规则中设置的域名，必须与证书中和步骤三：添加扩展域名中添加的扩展域名保持一致。

## 步骤三：添加扩展域名

完成以下操作，添加扩展域名：

1. 单击SLB1实例ID，进入实例详情页面。
2. 在监听页签下，找到已创建的HTTPS监听，选择 > 扩展域名管理。
3. 在扩展域名管理页面，单击添加扩展域名，配置扩展域名。
   - 输入域名。域名只能使用字母、数字、连字符（-）和点号（.）。

     域名转发策略支持精确匹配和通配符匹配两种模式：
     - 精确域名：www.aliyun.com。
     - 通配符域名（泛域名）：*.aliyun.com和*.market.aliyun.com 。

       当前端请求同时匹配多条域名策略时，策略的匹配优先级为：精确匹配高于小范围通配符匹配，小范围通配符匹配高于大范围通配符匹配，如下表所示。

| 模式 | 请求测试URL | 配置的转发域名策略 | | |
|---|---|---|---|---|
| | | www.aliyun.com | *.aliyun.com | *.market.aliyun.com |
| 精确匹配 | www.aliyun.com | ✓ | × | × |
| 泛域名匹配 | market.aliyun.com | × | ✓ | × |
| 泛域名匹配 | info.market.aliyun.com | × | × | ✓ |

   - 选择该域名关联的证书。

     > 说明 证书中的域名和您添加的扩展域名必须一致。

> 注意 配置完成后，如果出现问题，请尝试重启浏览器后再测试，避免缓存对结果的影响。

# 6.基于域名或URL路径进行转发

负载均衡支持配置基于域名或URL路径的转发策略。您可以将来自不同域名或URL路径的请求转发给不同的后端服务器组，合理分配服务器资源。

> 说明 只有七层监听（HTTPS/HTTP协议）支持配置转发策略。

## 域名或URL路径转发

七层负载均衡服务支持配置域名或者URL路径转发策略，将来自不同域名或者URL路径的请求转发给不同的ECS处理。

URL路径转发支持字符串匹配，按照前缀最长匹配原则。例如您配置了/abc和/abcd两个规则，当您访问/abcde时，系统优先匹配/abcd规则。

域名转发策略支持精确匹配和通配符匹配两种模式：

- 精确域名：www.aliyun.com
- 通配符域名（泛域名）：*.aliyun.com、*.market.aliyun.com

  当前端请求同时匹配多条域名策略时，策略的匹配优先级为：精确匹配高于小范围通配符匹配， 小范围通配符匹配高于大范围通配符匹配，如下表所示。

| 模式 | 请求测试URL | 配置的转发域名策略 | | |
|---|---|---|---|---|
| | | www.aliyun.com | *.aliyun.com | *.market.aliyun.com |
| 精确匹配 | www.aliyun.com | ✓ | × | × |
| 泛域名匹配 | market.aliyun.com | × | ✓ | × |
| 泛域名匹配 | info.market.aliyun.com | × | × | ✓ |

您可以在一个监听下添加多条转发策略，每条转发策略关联不同的虚拟服务器组（一个虚拟服务器组由一组ECS实例组成）。例如您可以将所有读请求转发到一组后端服务器上而将写请求转发到另一组后端服务器上，这样可以更灵活地适配业务需求，合理分配资源。

如下图所示，在配置了转发策略后，负载均衡系统将按照以下两种方式匹配策略，转发前端请求：

- 方式一：前端请求中存在域名，则根据域名匹配转发策略。
  - 存在匹配该域名的转发策略，则继续匹配URL路径部分。

    若URL路径部分也能匹配，则将请求转发到对应的虚拟服务器组；若URL路径部分未能命中该域名下的任何规则，则将请求转发给域名根路径转发策略（转发策略中只配置了域名，没有配置URL路径）。

    当用户没有为该域名配置根路径转发策略时，将向客户端返回404错误。
  - 不存在匹配该域名的转发策略，则按照方式二匹配转发策略。
- 方式二：前端请求中不存在域名或者转发策略中不存在与之相匹配的域名，则直接匹配无域名转发策略（转发策略中只配置了URL，没有配置域名）。

  成功匹配到转发策略时，将请求转发到对应的虚拟服务器组；未能匹配到任何转发策略时，将请求转发到负载均衡实例默认服务器组。

## 添加域名或URL路径转发策略

在配置域名或URL路径转发策略前，确保您已经：

- 添加HTTP监听或添加HTTPS监听。
- 添加ECS实例作为虚拟服务器

完成以下步骤，配置基于域名或URL路径的转发策略：

1. 登录负载均衡管理控制台。
2. 选择地域，查看该地域的所有负载均衡实例。
3. 单击负载均衡实例的ID。
4. 单击监听页签。
5. 单击目标七层监听操作列下的配置转发策略。
6. 在转发策略页面，根据以下信息配置转发策略：
   i. 域名：输入要转发的请求域名。域名只能使用字母、数字、短横线（-）和英文句点（.）。
   ii. URL：输入请求路径。路径必须以正斜线（/）开头，只能包含字母、数字、短横线（-）、英文句点（.）、正斜线（/）、百分号（%）、问号（?）、井号（#）和and（&）。

   > 说明
   > - 如果请求的URL路径中包含特殊字符，您需要使用URL特殊字符转义编码。例如，如果配置的转发策略使用包含特殊字符“/#/”的URL路径，那么在访问对应的服务时，需要使用特殊字符井号（#）的转义编码“%23”，即请求的URL路径中必须是“/%23/”，这样才能按设定的转发规则转发请求。

   iii. 虚拟服务器组：选择关联的虚拟服务器组。
   iv. 备注：输入描述。
   v. 单击添加转发策略。
7. 单击添加域名或添加规则再添加一个域名或URL策略。

   一个HTTP或HTTPS监听最多可添加转发策略个数请参见使用限制。

## 编辑转发策略

您可以修改转发策略关联的后端服务器。

完成以下操作，编辑转发策略：

1. 登录负载均衡管理控制台。
2. 选择地域，查看该地域的所有负载均衡实例。
3. 单击负载均衡实例的ID。
4. 单击监听页签。
5. 单击目标七层监听操作列下的配置转发策略。
6. 在转发策略页面的转发策略列表区域，单击目标转发策略的编辑选项。
7. 根据以下信息自定义转发策略的调度算法、会话保持和健康检查等配置。

**说明** 当前仅支持在以下地域自定义已有转发策略的高级配置：

- 华北2（北京）
- 华东1（杭州）
- 华东2（上海）
- 华北3（张家口）
- 华北5（呼和浩特）
- 中国香港
- 新加坡
- 日本

| 高级配置 | 说明 |
| --- | --- |
| 调度算法 | 负载均衡支持轮询、加权轮询（WRR）、加权最小连接数（WLC）三种调度算法。<br>• **加权轮询**：权重值越高的后端服务器，被轮询到的次数（概率）也越高。<br>• **轮询**：按照访问顺序依次将外部请求依序分发到后端服务器。<br>• **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。 |
| 开启会话保持 | 选择是否开启会话保持。<br>开启会话保持功能后，负载均衡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进行处理。<br>HTTP协议会话保持基于Cookie。负载均衡提供了两种Cookie处理方式：<br>• **植入Cookie**：您只需要指定Cookie的过期时间。<br>客户端第一次访问时，负载均衡会在返回请求中植入Cookie（即在HTTP/HTTPS响应报文中插入SERVERID），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后端服务器上。<br>• **重写Cookie**：可以根据需要指定HTTPS/HTTP响应中插入的Cookie。您需要在后端服务器上维护该Cookie的过期时间和生存时间。<br>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后端服务器。详情参考会话保持规则配置。 |

| 高级配置 | 说明 |
| --- | --- |
| 开启健康检查 | ○ **健康检查端口**：健康检查服务访问后端时的探测端口。<br>默认值为配置监听时指定的后端端口。<br>○ **健康检查路径**：用于健康检查页面文件的URI，建议对静态页面进行检查。<br>○ **健康检查域名（可选）**：默认使用各后端服务器的内网IP为域名。<br>○ **正常状态码**：选择健康检查正常的HTTP状态码。<br>默认值为http_2xx和http_3xx。<br>○ **健康检查响应超时时间**：接收来自运行状况检查的响应需要等待的时间。如果后端ECS在指定的时间内没有正确响应，则判定为健康检查失败。<br>○ **健康检查间隔时间**：进行健康检查的时间间隔。<br>默认为2秒。<br>○ **健康不检查健康阈值**：同一LVS节点服务器针对同一ECS服务器，从成功到失败的连续健康检查失败次数。<br>可选值2-10，默认为3次。<br>○ **健康检查健康阈值**：同一LVS节点服务器针对同一ECS服务器，从失败到成功的连续健康检查成功次数。<br>可选值2-10，默认为3次。 |

8. 单击**确定**。

## 删除转发策略

完成以下操作，删除转发策略：

1. 登录负载均衡管理控制台。
2. 选择地域，查看该地域的所有负载均衡实例。
3. 单击负载均衡实例的ID。
4. 单击**监听**页签。
5. 单击目标七层监听操作列下的**配置转发策略**。
6. 在**转发策略**页面的**转发策略列表**区域，单击目标转发策略的**删除**选项。

# 7.相同域名不同路径的流量转发

负载均衡支持配置基于域名和路径的转发策略。您可以将来自相同域名不同路径的请求转发给不同的后端服务器组，合理分配服务器资源。

## 背景信息

> 说明 只有7层监听（HTTPS/HTTP协议）支持配置转发策略。

本操作以四个部署了Nginx服务器的ECS为例，演示如何通过配置域名加URL转发规则，完成如下表所示的流量转发。

| 前端请求 | 流量转发至 |
| --- | --- |
| www.aaa.com/tom | 后端服务器SLB_tom1和SBL_tom2，属于虚拟服务器组TOM。 |
| www.aaa.com/jerry | 后端服务器SLB_jerry1和SBL_jerry2，属于虚拟服务器组JERRY。 |

## 域名和路径转发介绍

七层负载均衡服务支持配置域名或者URL转发策略，将来自不同域名或者URL的请求转发给不同的ECS处理。

URL转发支持字符串匹配，按照前缀最长匹配原则，例如有/abc和/abcd两个规则，访问/abcde，优先匹配/abcd规则。

域名转发策略支持精确匹配和通配符匹配两种模式：

- 精确域名：www.aliyun.com。
- 通配符域名（泛域名）：*.aliyun.com和 *.market.aliyun.com

  当前端请求同时匹配多条域名策略时，策略的匹配优先级为：精确匹配高于小范围通配符匹配， 小范围通配符匹配高于大范围通配符匹配，如下表所示。

| 模式 | 请求测试URL | 配置的转发域名策略 | | |
| --- | --- | --- | --- | --- |
| | | www.aliyun.com | *.aliyun.com | *.market.aliyun.com |
| 精确匹配 | www.aliyun.com | ✓ | × | × |
| 泛域名匹配 | market.aliyun.com | × | ✓ | × |
| 泛域名匹配 | info.market.aliyun.com | × | × | ✓ |

您可以在一个监听下添加多条转发策略，每条转发策略关联不同的虚拟服务器组（一个虚拟服务器组由一组ECS实例组成）。例如您可以将所有的请求转发到一组后端服务器上而将写请求转发到另一组后端服务器上，这样可以更灵活地适配业务需求，合理分配资源。

如下图所示，在配置了转发策略后，负载均衡系统将按照以下两种方式匹配策略，转发前端请求：

- 前端请求中存在域名（domain），则根据域名匹配转发策略。

- 方式一：存在匹配该域名的转发策略，则继续匹配URL部分。

  若URL部分也能匹配，则将请求转发到对应的虚拟服务器组；若URL部分未能命中该域名下的任何规则，则将请求转发给域名根路径转发策略（转发策略中只配置了域名，没有配置URL）。

  当用户没有为该域名配置根路径转发策略时，将向客户端返回404错误。

- 不存在匹配该域名的转发策略，则按照方式二匹配转发策略。

- 方式二：前端请求中不存在域名或者转发策略中不存在与之相匹配的域名，则直接匹配无域名转发策略（转发策略中只配置了URL，没有配置域名）。

  成功匹配到转发策略时，将请求转发到对应的虚拟服务器组；未能匹配到任何转发策略时，将请求转发到负载均衡实例默认服务器组。

## 配置路径转发策略

在配置域名和路径转发策略前，确保您已经：

1. 已创建一个公网负载均衡实例，详情请参见创建负载均衡实例。
2. 已创建一个七层监听，调度算法为轮询，详情请参见添加HTTP监听或添加HTTPS监听。
3. 已创建两个虚拟服务器组TOM和JERRY，详情请参见添加ECS实例作为虚拟服务器。
   - TOM虚拟服务器组中已添加服务器SLB_tom1和SBL_tom2，将端口设置为80，权重使用默认值100。
   - JERRY虚拟服务器组中已添加服务器SLB_jerry1和SBL_jerry2，将端口设置为80，权重使用默认值100。

执行下面的操作步骤配置路径转发策略：

1. 登录负载均衡管理控制台。
2. 在顶部菜单栏选择负载均衡实例的所属地域。
3. 在**实例管理**页面，单击目标实例的ID链接。
4. 在**监听**页签下，单击**操作**列的**转发策略**。
5. 配置两条转发规则：将来自www.aaa.com/tom的请求转发至TOM虚拟服务器组，以及将来自www.aaa.com/jerry的请求转发至JERRY虚拟服务器组。

   参数说明如下：

   - **域名**：输入要转发的请求域名。域名只能使用字母、数字、连字符（-）和点号（.）。
   - **URL**：输入请求路径。路径必须以正斜杠（/）开头，只能包含字母、数字和以下特殊字符。

     -./%?#&

     > 说明 如果您只想配置域名转发策略，则不需要配置URL。

   - **虚拟服务器组**：选择关联的虚拟服务器组。
   - **备注**：输入描述。

   > 说明 一个HTTP或HTTPS监听最多可添加转发策略个数请参见使用限制。

6. 单击**添加转发策略**。
7. 单击**确定**。
8. 测试路径转发策略是否配置成功。
   - 在浏览器中输入wwww.aaa.com/jerrry，将返回如下结果。
   - 在浏览器中输入wwww.aaa.com/tom，将返回如下结果。
   - 在浏览器中输入wwww.aaa.com，将返回如下结果。

# 8.获取客户端真实IP

您可以通过设置负载均衡的7层监听服务获取客户端真实IP地址。

## 背景信息

七层负载均衡（HTTP/HTTPS协议）服务需要对应用服务器进行配置，然后使用 `X-Forwarded-For` 的方式获取客户端的真实IP地址。真实的客户端IP存放在HTTP头部的X-Forwarded-For字段，格式如下：

```
X-Forwarded-For: 用户真实IP, 代理服务器1-IP， 代理服务器2-IP，...
```

当使用此方式获取客户端真实IP时，获取的第一个地址就是客户端真实IP。

> **说明** 负载均衡的HTTPS监听是在负载均衡服务上的加密控制，后端仍旧使用HTTP协议，因此，在Web应用服务器上配置HTTPS和HTTP监听没有区别。

## 配置IIS7/IIS8服务器

1. 下载并解压 *F5XForwardedFor*文件。
2. 根据自己的服务器操作系统版本将*x86\Release*或*x64\Release*目录下的*F5XFFHttpModule.dll*和*F5XFFHttpModule.in*拷贝到某个目录，例如*C:\F5XForwardedFor*\。确保IIS进程对该目录有读取权限。
3. 打开**IIS管理器**，双击**模块**功能。
4. 单击**配置本机模块**，然后在弹出的对话框中，单击**注册**。
5. 添加下载的*.dll*文件。
6. 为添加的两个文件授权允许运行ISAPI和CGI扩展。

   > **说明** 确保您已经安装了ISAPI和CGI应用程序。

7. 重启IIS服务器，等待配置生效。

## 配置Apache服务器

1. 执行如下命令，安装Apache的一个第三方模块mod_rpaf。

   ```
   wget https://github.com/gnif/mod_rpaf/archive/v0.6.0.tar.gz
   tar zxvf mod_rpaf-0.6.tar.gz
   cd mod_rpaf-0.6
   /alidata/server/httpd/bin/apxs -i -c -n mod_rpaf-2.0.so mod_rpaf-2.0.c
   ```

2. 修改Apache的配置文件*/alidata/server/httpd/conf/httpd.conf*，在最末尾添加以下配置信息。

```
LoadModule rpaf_module modules/mod_rpaf-2.0.so
RPAFenable On
RPAFsethostname On
RPAFproxy_ips  <IP_address>
RPAFheader X-Forwarded-For
```

> **说明** 如果您要获取代理服务器的地址，可以将代理服务器的网段添加到 `RPAFproxy_ips <IP_address>` ，如负载均衡的IP地址段100.64.0.0/10（100.64.0.0/10 是阿里云保留地址，其他用户无法分配到该网段内，不会存在安全风险）和高防IP地址段。多个IP地址段用逗号分隔。

3. 添加完成后重启Apache。

```
/alidata/server/httpd/bin/apachectl restart
```

## 配置Nginx服务器

1. 执行如下命令，安装http_realip_module。

```
wget http://nginx.org/download/nginx-1.0.12.tar.gz
tar zxvf nginx-1.0.12.tar.gz
cd nginx-1.0.12
./configure --user=www --group=www --prefix=/alidata/server/nginx --with-http_stub_status_m
odule --without-http-cache --with-http_ssl_module --with-http_realip_module
make
make install
kill -USR2 `cat /alidata/server/nginx/logs/nginx.pid`
kill -QUIT `cat /alidata/server/nginx/logs/ nginx.pid.oldbin`
```

2. 执行如下命令，打开*nginx.conf*文件。

```
vi /alidata/server/nginx/conf/nginx.conf
```

3. 在以下配置信息后添加新的配置字段和信息。

```
fastcgi connect_timeout 300;
fastcgi send_timeout 300;
fastcgi read_timeout 300;
fastcgi buffer_size 64k;
fastcgi buffers 4 64k;
fastcgi busy_buffers_size 128k;
fastcgi temp_file_write_size 128k;
```

需要添加的配置字段和信息为：

```
set_real_ip_from IP_address;
real_ip_header X-Forwarded-For;
```

> **说明** 如果您要获取代理服务器的地址，可以将代理服务器的网段添加到 `set_real_ip_from <IP_address>` ，如负载均衡的IP地址段100.64.0.0/10（100.64.0.0/10 是阿里云保留地址，其他用户无法分配到该网段内，不会存在安全风险）和高防IP地址段。多个IP地址段用逗号分隔。

4. 执行如下命令，重启Nginx。

```
/alidata/server/nginx/sbin/nginx -s reload
```

# 9.压力测试的方法

四层负载均衡采用开源软件LVS（Linux Virtual Server）+ Keepalived的方式实现负载均衡，七层负载均衡由Tengine实现负载均衡。

## 压力测试性能概述

四层监听经过LVS后直接到达后端服务器，而七层监听经过LVS后，还需要再经过Tengine，最后到达后端服务器。七层监听比四层监听多了一个处理环节，因此，七层性能没有四层性能好。

如果您使用七层监听进行压力测试，发现压测性能比较低。挂了两台ECS的七层负载均衡监听性能不如挂了一台ECS的四层负载均衡监听性能，除了七层本身的性能比四层低外，以下情况也可能会造成七层压测性能低：

- 客户端端口不足 。

  在进行压力测试时，客户端端口不足会导致建立连接失败。负载均衡会默认抹除TCP连接的timestamp属性，Linux协议栈的tw_reuse（time_wait 状态连接复用）无法生效，time_wait状态连接堆积导致客户端端口不足。

  解决方法：客户端使用长连接代替短连接。使用RST报文断开连接，即socket设置SO_LINGER属性。

- 后端服务器accept队列满 。

  后端服务器accept队列满，导致后端服务器不回复syn_ack报文，客户端超时。

  解决方法：默认net.core.somaxconn的值为128，执行 `sysctl -w net.core.somaxconn=1024` 命令更改net.core.somaxconn的值，并重启后端服务器上的应用。

- 后端服务器连接过多。

  由于架构设计的原因，使用七层负载均衡时，用户长连接经过Tengine后变成短连接，可能导致后端服务器连接过多，从而表现为压测性能低。

- 后端服务器依赖的应用成为瓶颈。

  请求经过负载均衡到达后端服务器后，后端服务器本身负载正常，但由于所有的后端服务器上的应用又依赖其它应用，例如数据库，当数据库成为瓶颈时，也会引起性能降低。

- 后端服务器的健康检查状态异常。

  在压测时，容易忽略后端服务器的健康检查状态，如果有后端服务器健康检查失败或者健康检查状态经常跳跃（好到坏，又从坏到好，反复变化），也会导致压测性能低。

## 压力测试建议

在进行压力测试时，请注意如下配置：

- 压测负载均衡转发能力建议使用短连接 。

  一般来说压测除了验证会话保持和均衡性等功能外，主要想验证负载均衡的转发能力，因此使用短连接比较合适，用于测试负载均衡和后端服务器的处理能力。使用短连接测试时，需要注意客户端端口不足的问题。

- 压测负载均衡吞吐量建议使用长连接，用于测试带宽上限或特殊业务。

  压测工具的超时时间建议设置为一个较小值，如5秒。超时时间太大的话，测试结果会体现在平均响应时间加长，不利于判断压测水位是否已到达。超时时间调小，测试结果会体现在成功率上，便于快速判断压测水位。

- 后端服务器提供一个静态网页用于压测，以避免应用逻辑带来的损耗。

- 压测时，监听配置建议如下：
  - 不开启会话保持功能，否则压力会集中在个别后端服务器。
  - 关闭健康检查功能，减少健康检查对后端服务器的访问请求。
  - 性能测试服务的5000并发规格能够提供5个及5个以上的公网IP。

## 压力测试工具建议

不建议您使用Apache ab作为压力测试工具。

Apache ab在大量并发场景下存在3s、6s、9s阶梯式停顿的现象。Apache ab会通过判断content length来确定请求是否成功，而负载均衡挂载多台后端服务器时，返回的content length会不一致，导致测试结果有误。

建议使用阿里云PTS。

可以设置足够高的并发，PTS会分配来自全国各地的公网IP，压力来源足够分散，并且可以在PTS中集成云监控，实时查看端到端的全部性能数据。

## 使用PTS简单压测示例

创建一个负载均实例，添加两台ECS实例作为后端服务器，分别创建一个TCP监听和HTTP监听，后端端口设置为80。ECS服务器的配置为CPU 1核，内存512M使用CentOS 6.3 64位的操作系统。

1. 安装Apache Web Server提供Web服务。

   ```
   yum install -y httpd
   ```

2. 初始化默认首页index.html。

   ```
   echo "testvm" > /var/www/html/index.html
   ```

3. 启动HTTP服务。

   ```
   service httpd start
   ```

4. 访问本地的80端口，确认Web服务可用。

   ```
   curl localhost
   ```

5. 在PTS中创建测试场景，开始压力测试。

# 10.使用访问日志快速定位异常后端服务器

某段时间客户端访问延迟时，您可以结合阿里云日志服务，通过仪表盘巡检，分析负载均衡的响应时间，快速定位异常后端服务器。

本教程介绍如何使用访问日志快速定位异常后端服务器，更多访问日志详情请参见配置访问日志。

## 步骤一：配置负载均衡访问日志

在配置访问日志前，请确保：

1. 您已经创建了七层负载均衡。
2. 您已经开通了日志服务。

完成以下操作，配置访问日志：

1. 登录负载均衡管理控制台。
2. 在左侧导航栏，选择 **日志管理 > 访问日志** 。
3. 选择实例的所属地域。
4. 单击**立即授权**，然后在弹出的对话框，单击**同意授权**授权SLB访问日志服务。

   如果您使用的是子账号，需要主账号进行授权。详情参见授权子账号使用访问日志。

   > **说明** 该操作只有首次配置时需要。

5. 在**访问日志**页面，找到目标SLB实例，然后单击**设置**。
6. 选择日志服务（LogProject）和日志库（LogStore），然后单击**确定**

   > **说明** 确保Project的名称全局唯一，且Project的地域和负载均衡实例的地域相同。

## 步骤二：查看访问日志

完成以下操作，查询访问日志：

1. 进入日志查询页面。您可以通过负载均衡控制台和日志服务控制台进入日志查询页面。
   - 负载均衡控制台

     在**访问日志**页面，单击**查看日志**。
   - 日志服务控制台

     在**日志库**页面，单击日志库的**查询/分析**选项。
2. 单击目标日志字段，查看对应的日志信息。
3. 输入SQL语句查询特定的访问日志。

   例如输入如下SQL语句查询Top20的客户端，用于分析请求访问来源，辅助商业决策。

```
*| select ip_to_province(client_ip) as client_ip_province, count(*) as pv group by
    client_ip_province order by pv desc limit 50
```

## 步骤三：定位异常后端服务器

您可以通过日志服务的仪表盘定位异常后端服务器。

1. 登录日志服务控制台，单击负载均衡的Project链接。
2. 在左侧导航栏，单击 。
3. 单击负载均衡访问日志的名称链接。

   
4. 在仪表盘中，查看**top upstream响应时间**页签下负载均衡SLB的响应时间，可以将参数**平均upstream响应时间(s)**设置降序排列，查看是否有后端服务器的响应时间超过1s。

   如果有响应时间超过1s的后端服务器，执行ssh命令，登录该后端服务器，查看CPU是否持续高位运行，进行高负载处理。

# 11.配置访问控制

通过视频模式介绍配置访问控制的详细操作步骤，负载均衡提供监听级别的访问控制，您可以为不同的监听配置不同的访问控制策略。

# 12.配置会话保持

通过视频模式介绍如何配置会话保持，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。

# 13.将流量转发到虚拟服务器组

通过视频模式介绍将流量转发给关联的虚拟服务器组中的后端服务器的详细操作。

虚拟服务器组是一组 ECS 实例。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。

# 14.通过OpenAPI Explorer创建VPC类型实例时指定IP

使用OpenAPI explorer创建VPC类型负载均衡实例时，支持在负载均衡实例所属交换机支持的网段中，指定其中一个地址作为负载均衡实例的私网IP地址。

## 操作步骤

1. 登录OpenAPI Explorer控制台。
2. 搜索负载均衡产品的CreateLoadBalancer接口。
3. 设置创建负载均衡实例的参数。此处设置部分参数作为示例，详细参数说明请参见创建负载均衡实例：
   - RegionId：表示负载均衡实例的地域，此处设置为*cn-hangzhou*。
   - VpcId：表示负载均衡实例所属VPC的ID。

     此处可登录专有网络VPC控制台，选择华东1（杭州）区域，查看VPC的ID。
   - VSwitchId：表示负载均衡所属交换机的ID，如果需要指定负载均衡IP地址，该参数必须要设置。

     此处可在专有网络VPC控制台，单击负载均衡实例所属VPC的ID，在**网络资源**页面下，单击交换机的个数，查看交换机的ID。

     单击交换机ID，查看交换机的目标网段，如192.168.0.0/24。
   - Address：指定负载均衡实例的私网IP地址，该地址必须包含在交换机的目标网段下，例如192.168.0.3。
4. 单击**发起调用**。返回结果如下：
   - XML格式

```
<?xml version="1.0" encoding="UTF-8" ?>
  <NetworkType>vpc</NetworkType>
  <LoadBalancerName>auto_named_slb</LoadBalancerName>
  <Address>192.168.0.3</Address>
  <ResourceGroupId>rg-acfmxazb4ph*****</ResourceGroupId>
  <RequestId>09197EEB-7013-4F56-A5CE-A756FFE5B75D</RequestId>
  <AddressIPVersion>ipv4</AddressIPVersion>
  <LoadBalancerId>lb-bp1h66tp5uat84kh******</LoadBalancerId>
  <VSwitchId>vsw-bp14cagpfysr29fe*****</VSwitchId>
  <VpcId>vpc-bp18sth14qii3pn*****</VpcId>
```

   - JSON格式

```
{
  "NetworkType": "vpc",
  "LoadBalancerName": "auto_named_slb",
  "Address": "192.168.0.3",
  "ResourceGroupId": "rg-acfmxazb4******",
  "RequestId": "09197EEB-7013-4F56-A5CE-A756FFE5B75D",
  "AddressIPVersion": "ipv4",
  "LoadBalancerId": "lb-bp1h66tp5uat84******",
  "VSwitchId": "vsw-bp14cagpfysr29*******",
  "VpcId": "vpc-bp18sth14qii3*********"
}
```


5. 登录负载均衡管理控制台，选择华东1（杭州）区域，查看IP为192.168.0.3的负载均衡实例是否创建成功。

# 15.查看流量使用情况

用户需要查看某一时间段内云账号下负载均衡实例流量使用情况。

## 操作步骤

1. 登录负载均衡管理控制台。
2. 在菜单栏右上角选择 **费用 > 进入费用中心**。
3. 在**费用中心**页面，选择 **消费记录 > 使用记录**。
4. 在**使用记录**页面，选择**负载均衡**产品，配置需要查看的负载均衡流量使用情况的使用期间和计量粒度。
5. 单击**导出CSV**，在本地生成 *.CSV*格式的流量使用表格。 该表格包含以下信息，可根据实例、地域或者服务地址等查看具体流量使用情况。