

阿里云 负载均衡

扩展阅读

文档版本：20181108

法律声明

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

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

通用约定

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

目录

法律声明.....	I
通用约定.....	I
1 最佳实践.....	1
1.1 使用Open API配置负载均衡.....	1
1.2 负载均衡+ECS多站点部署.....	3
1.3 健康检查导致大量日志的处理.....	4
1.4 配置服务器Cookie.....	8
1.5 获取客户端真实IP.....	9
1.6 将按量付费实例切换为包年包月实例.....	12
1.7 从负载均衡服务中移除ECS实例.....	13
1.8 高可用最佳实践.....	14
1.9 SLB实例克隆.....	16
2 常见问题.....	17
2.1 负载均衡实例增加重试机制公告.....	17
2.2 如何进行压力测试.....	17
2.3 负载均衡常见问题.....	20
2.4 后端服务器常见问题.....	25
2.5 健康检查常见问题.....	28
2.6 性能保障型实例收费常见问题.....	34
2.7 如何使用负载均衡性能保障型实例？.....	37
2.8 为什么请求不均衡.....	46
2.9 如何实现相同域名不同路径的流量转发.....	47
2.10 会话保持常见问题.....	49
2.11 HTTP/2协议支持常见问题.....	51
2.12 HTTPS/HTTP监听常见问题.....	52
2.13 WS/WSS协议支持常见问题.....	56

1 最佳实践

1.1 使用Open API配置负载均衡

前提条件

您已经创建了两个ECS实例，并授予SSH和Web端口的访问权限。

背景信息

以下示例中使用URL表示请求参数，并且没有将公共请求参数包含在内。构造完整的请求URL，请参见[负载均衡Open API](#)。



说明：

为方便阅读，示例中请求URL的参数值没有进行URL编码。

```
server.modules = ( "mod_setenv" )
$http["host"] == "test.example.com" {
    server.document-root = "/var/www/html/"
    setenv.add-response-header = ( "Set-Cookie" => "name=XXXXXX"
    }
}
```

操作步骤

1. 调用**CreateLoadBalancer**接口创建一个负载均衡实例。

请求：

```
https://slb.aliyuncs.com/?Action=CreateLoadBalancer&RegionId=cn-hangzhou
```

响应：

```
{
  "RequestId": "3DE96B24-E2AB-4DFA-9910-1AADD60E13A5",
  "LoadBalancerId": "LoadBalancerId",
  "Address": "SLBIPAddress"
}
```

2. 调用**CreateLoadBalancerHttpListener**接口为创建好的负载均衡实例添加一个端口为80的HTTP监听。

请求：

```
https://slb.aliyuncs.com/?Action=CreateLoadBalancerHttpListener&
LoadBalancerId=LoadBalancerId&ListenerPort=80&BackendServerPort=80&
ListenerStatus=active
```

3. 调用**SetLoadBalancerStatus**接口激活负载均衡实例。

请求：

```
https://slb.aliyuncs.com/?Action=SetLoadBalancerStatus&LoadBalancerId=LoadBalancerId&LoadBalancerStatus=active
```

4. 调用**AddBackendServers**接口将一个ECS实例添加到负载均衡实例中。

请求：

```
https://slb.aliyuncs.com/?Action=AddBackendServers&LoadBalancerId=LoadBalancerId&BackendServers=[{"ServerId":"ECS1InstanceID"}]
```

响应

```
{
  "RequestId" : "FA2F2172-63F2-409D-927C-86BD1D536F13",
  "LoadBalancerId" : "LoadBalancerId",
  "BackendServers" : {
    "BackendServer" : [
      {
        "ServerId" : "ECS1InstanceId",
        "Weight" : 100
      }
    ]
  }
}
```

5. 再次调用**AddBackendServers**接口将另外一个ECS实例添加到负载均衡实例中。

请求：

```
https://slb.aliyuncs.com/?Action=AddBackendServers&LoadBalancerId=LoadBalancerId&BackendServers=[{"ServerId":"ECS2InstanceID"}]
```

响应

```
{
  "RequestId" : "C61FAD0A-2E87-4D0C-80B0-95AB758FCA70",
  "LoadBalancerId" : "LoadBalancerId",
  "BackendServers" : {
    "BackendServer" : [
      {
        "ServerId" : "ECS1InstanceId",
        "Weight" : 100
      },
      {
        "ServerId" : "ECS2InstanceId",

```

```
        "Weight" : 100
      }
    ]
  }
}
```

6. 调用**DescribeLoadBalancerAttribute**接口查询负载均衡实例的配置信息。

请求：

```
https://slb.aliyuncs.com/?Action=DescribeLoadBalancerAttribute&
LoadBalancerId=LoadBalancerId
```

响应：

```
{
  "RequestId" : "4747E9AE-ADFD-412D-B523-C1CBD45A2154",
  "LoadBalancerId" : "LoadBalancerId",
  "Address" : "SLBIPAddress",
  "IsPublicAddress" : "true",
  "ListenerPorts" : {
    "ListenerPort" : [
      80
    ]
  },
  "BackendServers" : {
    "BackendServer" : [
      {
        "ServerId" : "ECS1InstanceId",
        "Weight" : 100
      },
      {
        "ServerId" : "ECS2InstanceId",
        "Weight" : 100
      }
    ]
  }
}
```

配置完成后，在浏览器中输入负载均衡实例的IP地址，查看服务是否正常。

1.2 负载均衡+ECS多站点部署

ECS上可以通过设置虚拟子目录搭建多个站点。负载均衡服务支持添加这种类型的ECS实例作为后端服务器，进行流量转发。HTTP和TCP负载均衡服务都可以挂载多站点ECS，并且不受会话保持功能影响。

背景信息

本教程概述了如何创建一个TCP监听，并挂载两个不同站点的ECS实例。

操作步骤

1. 在ECS01上通过设置虚拟子目录，搭建两个域名为www.aaa.com和www.bbb.com的站点。

www.aaa.com的站点内容，如下图所示。

www.bbb.com的站点内容，如下图所示。

您可以在本地host文件中，将这两个域名和ECS实例的公网IP地址绑定，然后在浏览器中访问www.aaa.com和www.bbb.com，如果能正常显示设置的网站，则说明配置成功。

2. 为ECS01创建一个自定义镜像，使用该镜像创建一个相同配置的ECS02。
3. 创建一个公网负载均衡实例，然后添加一个TCP监听。
4. 添加ECS01和ECS02作为后端服务器。
5. 在本地host文件中，将www.aaa.com和www.bbb.com这两个域名解析到负载均衡的公网IP地址上。
6. 在浏览器中输入www.aaa.com和www.bbb.com，如果可以正常访问到对应的后端服务，证明负载均衡服务正常。

1.3 健康检查导致大量日志的处理

负载均衡的日志管理功能会自动保存三天内的健康检查日志，如果健康检查日志过多，对您的运维工作造成不便，您可以选择以下方案来减少或禁止某些场景下健康日志的产生。



说明：

减少健康检查日志的数量可能会导致您无法及时发现负载均衡实例运行时所出现的问题，请您谨慎权衡每种方案所带来的风险，根据您的实际情况进行选择。

- [获取访问日志](#)
- [调整健康检查频率](#)
- [关闭7层负载均衡下的健康检查](#)

- 将7层负载均衡切换为4层负载均衡
- 关闭健康检查页面的应用日志

获取访问日志

HTTP健康默认使用HEAD请求方法（后续会开通GET方法），因此过滤掉HEAD的请求，即可获实际的访问日志。

调整健康检查频率

通过延长健康检查的间隔时间来减少健康检查的次数，降低健康检查产生的日志数量。

方案风险说明

延长健康检查的间隔时间后，后端ECS实例出现故障时，负载均衡发现故障ECS实例的时间也会变长。

操作步骤

1. 登录[负载均衡管理控制台](#)。
2. 在实例管理页面中找到相应的负载均衡实例，单击管理。
3. 单击监听进入监听管理页面，找到对应监听，单击配置。
4. 在配置监听对话框中单击下一步，进入健康检查配置。
5. 调整健康检查间隔，范围为1~50秒，间隔越大，健康检查的频率就越低，后端服务器产生的日志也会相应减少。请根据您的实际情况进行修改。
6. 单击确认完成修改。

关闭7层负载均衡下的健康检查

使用7层HTTP/HTTPS负载均衡模式时，健康检查由HTTP Head请求实现，后端服务器的应用日志会记录相应的健康检查请求信息，可能导致大量的日志信息。

风险说明

HTTP/HTTPS模式下关闭健康检查后，负载均衡不再检查后端服务器，一旦某台后端服务器发生故障，则无法实现访问流量自动切换至其它正常的后端服务器。

操作步骤

1. 登录[负载均衡管理控制台](#)。
2. 在实例管理页面中找到对应的负载均衡实例，单击管理。
3. 单击监听进入监听管理页面，找到对应监听，单击配置。
4. 在配置监听对话框中单击下一步，进入健康检查配置。
5. 关闭是否开启健康检查。
6. 单击确认完成修改。

将7层负载均衡切换4层负载均衡

4层TCP模式下的的健康检查仅仅使用TCP的三次握手实现，不会生成应用日志。如果您的业务可以切换为4层TCP模式，采用该方法可以减少应用日志的产生。

风险说明

将HTTP/HTTPS模式的负载均衡修改为TCP模式后，负载均衡将只检查监听端口状态，不检查HTTP状态，会导致负载均衡无法实时获知HTTP应用是否出现问题。

操作步骤

1. 登录[负载均衡管理控制台](#)。
2. 在实例管理页面中找到对应的负载均衡实例，单击管理。
3. 单击监听进入监听管理页面，找到对应监听，单击配置。
4. 在配置监听对话框中单击下一步，进入健康检查配置。
5. 将健康检查模式修改为**TCP**。
6. 单击确认完成修改。

关闭健康检查页面的应用日志

在业务站点之外，独立配置健康检查站点，并关闭健康检查页面的应用日志，可以减少健康检查日志的数量。例如业务站点为`abc.123.com`，则使用`test.123.com`作为健康检查站点，并关闭`test.123.com`站点的日志记录。

风险说明

如果健康检查的站点正常，但是业务站点出现异常时，健康检查则无法检测到业务站点的异常。

操作步骤

1. 在后端服务器上新建一个健康检查站点和健康检查页面，并关闭日志记录。本操作以nginx为例进行说明。
2. 登录 [负载均衡管理控制台](#)。
3. 在实例管理页面中找到对应的负载均衡实例，单击管理。
4. 单击监听进入监听管理页面，找到对应监听，单击配置。
5. 在配置监听对话框中单击下一步，进入健康检查配置。
6. 在域名中输入健康检查站点的域名，在检查路径中输入健康检查页面的相对路径。

高级配置 收起

● 健康检查协议

TCP

HTTP

● 健康检查方法

GET

● 健康检查端口

默认使用后端服务器端口进行检查，除非您希望指定特定的端口，否则建议留空

端口输入范围为1-65535。

● 健康检查路径

/test.html

长度限制为1-80个字符，只能使用字母、数字、'-'、'_'、'/'、':'、'%'、'?','#','&','='这些字符。

● 健康检查域名（可选）

test.123.com

只能使用字母、数字、'-'、'_'。默认使用各后端服务器的内网IP为域名。

● 正常状态码

☒

http_2xx

☒

http_3xx

☐

http_4xx

☐

http_5xx

● 健康检查响应超时时间

5

秒

7. 单击确认完成修改。

1.4 配置服务器Cookie

负载均衡服务提供会话保持功能。开启会话保持功能后，负载均衡会将会话期间内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。

四层监听的会话保持是基于IP地址的会话保持，负载均衡监听器会将来自同一IP地址的请求转发到同一个后端ECS上；而七层监听是基于Cookie的会话保持。

如果您选择使用重写Cookie的方式，需要在后端服务器上配置Cookie。假如您的负载均衡服务地址下有两个域名：vip.a.com和img.a.com，当您想为vip.a.com配置会话保持时，您可以设置Cookie名称为name，然后在后端服务器上为域名vip.a.com设置key为name的Cookie。

本文档介绍了如何在Apache、Nginx和Lighttpd应用服务器上配置Cookie。

Apache

1. 打开`httpd.conf`配置文件，确保如下配置没有被注释。

```
LoadModule usertrack_module modules/mod_usertrack.so
```

2. 在virtual host中添加以下配置。

```
CookieName name
CookieExpires "1 days"
CookieStyle Cookie
CookieTracking on
```

Nginx

参考以下配置，设置Cookie。

```
server {
    listen 8080;
    server_name wqwq.example.com;
    location / {
        add_header Set-Cookie name=xxxx;
        root    html;
        index   index.html index.htm;
    }
}
```

Lighttpd

参考以下配置，设置Cookie。

```
server.modules = ( "mod_setenv" )
$HTTP["host"] == "test.example.com" {
    server.document-root = "/var/www/html/"
    setenv.add-response-header = ( "Set-Cookie" => "name=XXXXXX" )
}
```

1.5 获取客户端真实IP

负载均衡服务获取真实IP说明

负载均衡提供获取客户端真实IP地址的功能，该功能默认是开启的。

- 四层负载均衡（TCP协议）服务可以直接在后端ECS上获取客户端的真实IP地址，无需进行额外的配置。

- 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会被负载均衡放在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应用服务器配置上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* 或者 *x64\Release* 目录下的 *F5XFFHttpModule.dll* 和 *F5XFFHttpModule.ini* 拷贝到某个目录，比如 *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. 运行以下命令安装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. 运行以下命令安装`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_module --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. 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. 重启Nginx。

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

1.6 将按量付费实例切换为包年包月实例

为了节省费用，您可以将按量付费的负载均衡实例切换为成预付费（包年包月）。本操作介绍如何在OpenAPI Explorer中调用**ModifyLoadBalancerPayType**接口修改负载均衡实例的计费方式。

前提条件

- 负载均衡实例没有处于欠费状态，且帐户余额足以支付预付费的费用。
- 目前只支持将按带宽计费的公网负载均衡实例切换为预付费。

操作步骤

1. 登录[OpenAPI Explorer](#)。
2. 在左侧导航栏，单击负载均衡 **SLB**。
3. 在搜索框中输入**ModifyLoadBalancerPayType**，并单击该接口。根据以下信息，输入接口的请求参数。
 - ReigonId：输入负载均衡实例所在的地域的ID。
 - LoadBalancerId：输入负载均衡实例的ID。
 - PayType：输入**PrePay**。
 - PricingCycle：选择付费周期，**year**为按年付费，**month**为按月付费。

- **Duration**：选择购买的时长，按月付费时，取值范围为[1，9]（超过9个月时，选择按年付费更优惠），按年付费时，取值范围为[1，3]。
- 4. 单击发送请求，当请求状态为“SUCCESS”时，表示成功生成了订单。
- 5. 登录费用中心。
- 6. 在左侧导航栏，单击订单管理。
- 7. 找到该负载均衡实例的未支付订单，并完成支付。

1.7 从负载均衡服务中移除ECS实例

直接从后端服务器池中移除ECS实例可能会造成业务中断，建议您先将ECS的权重设置为零后，再移除。

移除ECS实例

直接将ECS实例从负载均衡实例中移除，可能会造成业务闪断。建议您按照以下步骤移除ECS实例：

1. 登录[负载均衡管理控制台](#)。
2. 单击负载均衡实例的ID链接，进入实例详情页。
3. 在详情左侧导航栏，单击服务器 > 后端服务器。

如果您的ECS实例在虚拟服务器组或主备服务器组内，则选择对应的服务器组。

4. 将鼠标移至要移除的ECS实例的权重区域，单击出现的铅笔图标，然后将权重修改为0。
5. 当负载均衡服务不再转发流量给该ECS实例后，单击移除。

流量排查

把ECS从负载均衡实例中移除后，发现还有持续的业务请求，可以从以下两个方面排查。

- 查看该ECS是否加入了其他的负载均衡实例中。

登录负载均衡管理控制台，可以根据ECS实例ID和内网IP查询该ECS实例所属的负载均衡实例。

- 登录ECS实例，用netstat命令查看服务器端口监听情况。根据端口监听情况查看ECS本身是否对公网提供了其他服务。

- Windows：运行`netstat -ano`，可以查看本机开放的全部端口。
- Linux：也可以使用该命令查看本机开放端口，或使用`netstat`命令的其他参数。

1.8 高可用最佳实践

负载均衡（SLB）从系统设计、产品配置等多个方面提供了可用性保障。此外，您可以根据业务需求，配合使用云解析DNS等产品实现跨地域容灾。

SLB系统的高可用

负载均衡实例采用集群部署，可实现会话同步，以消除服务器单点故障，提升冗余，保证服务的稳定性。其中四层负载均衡通过LVS（Linux Virtual Server）+ keepalived的方式实现，七层负载均衡通过Tengine（淘宝网发起的Web服务器项目，在Nginx的基础上，针对有大访问量的网站需求进行了优化）实现。

来自公网的请求通过等价多路径路由（ECMP）到达LVS集群，LVS集群内的每台LVS通过组播报文将会话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。同时，LVS集群会对Tengine集群进行健康检查，将异常机器从Tengine集群移除，保证七层负载均衡的可用性。

最佳实践：

会话同步可以保证长连接不受集群内服务器故障的影响，但是对于短连接或连接未触发会话同步规则时（未完成三次握手），集群内的服务器故障仍可能会影响用户请求。为了防止集群中某台机器故障导致的会话中断，您可以在业务逻辑中加入重试机制，降低对用户访问造成的影响。

单SLB实例的高可用

为了向广大用户提供更稳定可靠的负载均衡服务，阿里云负载均衡已在大部分地域部署了多可用区以实现同地域下的跨机房容灾。当主可用区出现故障或不可用时，负载均衡有能力在非常短的时间内（约30秒）切换到备可用区并恢复服务；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。



说明：

负载均衡的主备可用区是可用区级别的容灾。只有当主可用区整体不可用时，如机房整体断电、机房出口光缆中断等，负载均衡才会切换到备可用区。而并非某个实例出现故障，就切换到备可用区。

最佳实践：

1. 为了更好的利用负载均衡的主备可用区机制，建议您在支持主备可用区的地域创建负载均衡实例，即在购买负载均衡实例时选择可用区类型为多可用区的地域。
2. 当您选择SLB的主备可用区时，可以根据ECS实例的可用区分布进行选择。大部分ECS实例位于哪个可用区，就将哪个可用区选择为SLB的主可用区，以获取最小的访问延迟。

但是并不建议您将全部ECS实例都部署在一个可用区内，您也需要在SLB的备可用区部署少量ECS实例，以便在极端情况下（主可用区整体不可用时），切换到备可用区后仍旧可以正常处理负载均衡转发的请求。

多SLB实例的高可用

如果您对可用性的要求特别高，负载均衡实例自身的可用性保障机制可能无法满足您的需求。例如当网络攻击或配置错误等情况导致负载均衡实例不可用时，由于未出现可用区级故障，不会触发负载均衡实例的可用区切换。此时，您可以创建多个SLB实例，通过云解析DNS对访问进行调度，或通过全球负载均衡解决方案实现跨地域容灾备份。

最佳实践：

您可以在一个地域内的多个可用区或多个地域内部署负载均衡实例和后端ECS实例，然后使用云解析DNS对访问进行调度。

后端ECS实例的高可用

负载均衡通过健康检查来判断后端ECS实例的可用性。健康检查机制提高了前端业务整体可用性，避免了后端ECS异常对总体服务的影响。

开启健康检查功能后，当后端某个ECS实例健康检查出现异常时，负载均衡会自动将新的请求分发到其他健康检查正常的ECS实例上；而当该ECS实例恢复正常运行时，负载均衡会将其自动恢复到负载均衡服务中。关于健康检查的详细机制，请参见[健康检查原理](#)。

最佳实践：

为了使健康检查功能正常运作，您需要开启并正确配置健康检查。具体的操作流程请参见[配置健康检查](#)。

1.9 SLB实例克隆

当您需要构建实例级别的高可用架构或做ABTest等场景，要克隆实例配置时，您可以使用aliyun-slb-helper工具将SLB实例的配置导出到本地，然后再使用导出的配置文件克隆一个新的SLB实例。

在使用该工具前，请注意：

- 本工具非官方提供的克隆工具，请谨慎在生产环境使用。
- 无论原实例是包年包月实例还是按量付费实例，克隆后的新实例的计费方式统一为按量付费。
- 克隆的新实例和原实例的地域相同。
- 该工具当前仅支持克隆实例的监听配置、后端服务器、虚拟服务器组和主备服务器组的配置。暂无法克隆转发规则和访问配置白名单。

具体的使用方法，请参考[SLB实例克隆](#)。

2 常见问题

2.1 负载均衡实例增加重试机制公告

尊敬的阿里云用户，

您好！

为了减少机器故障或升级对用户业务的影响，负载均衡采用了集群部署架构，并支持会话同步，可最大限度保证客户端访问负载均衡实例不中断。但是，会话同步无法保证所有连接不中断。因此，在如下两种情况下：

1. 连接未建立（三次握手未完成）
2. 已建立连接但未触发会话同步机制

没有重试机制的业务所使用的负载均衡实例会存在闪断，用户在负载均衡实例监控上可能会看到流量下跌。为了优化用户体验，建议没有重试机制的业务尽快调整业务逻辑，增加重试机制。

由于没有重试机制的业务所使用的负载均衡实例较少，因此，2017年04月15日之前，我们将对此类升级继续进行公告和通知；2017年04月15日之后，我们将不再对此类升级进行公告和通知，请在此之前完成业务逻辑调整。

注：单击[这里](#)查看负载均衡集群部署和会话同步原理。

非常感谢您的理解！

阿里云计算有限公司

2017年2月8日

2.2 如何进行压力测试

压力测试性能概述

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。其中四层监听经过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，最后达到后端服务器。七层比四层多了一个处理环节，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，发现根本跑不上去，挂了两台ECS的七层负载均衡监听性能还不如一台ECS的性能，除了七层本身的性能比四层低外，以下情况也可能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

请求经过负载均衡达到后端服务器后，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，比如数据库，数据库成为瓶颈，也会引起性能低。

- 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，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，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处理能力。但使用短连接测试时，需要注意客户端端口不足问题。

- 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RT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的后端服务器。
 - 监听关闭健康检查功能，减少健康检查请求对后端服务器的访问请求。
 - 用多个客户端进行压测最好多于5个，源IP分散，能够更好的模拟线上实际情况。

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](#)。

可以选择多个客户端作为压力测试源，测试结果清晰，并且可以通过配置监控，获取压力测试时后端服务器的性能数据。

使用PTS简单压测示例

创建一个负载均衡实例，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中创建测试脚本，开始压力测试。

注意关闭长连接和设置超时时间：

- 设置超时时间为5秒：PTS.HttpUtilities.setTimeout(5000)
- 关闭长连接：PTS.HttpUtilities.setKeepAlive(False)

2.3 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。

2. 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。

4. 负载均衡分配的IP是否为独占？

在您购买使用负载均衡服务的整个生命周期内，该服务地址都是由您所购买的服务独占的。更改负载均衡配置和监听策略不会导致负载均衡服务地址的变更。

如果您的负载均衡服务地址已经正常解析到域名且对外提供服务，除非必要请不要删除对应的负载均衡实例。负载均衡实例删除后，相应的服务配置和服务地址将会被释放掉，数据一旦删除，不可恢复。如果您重新创建负载均衡服务，系统会重新给您分配一个新的服务地址。

5. 负载均衡是否支持端口跳转？

支持。

负载均衡支持HTTP重定向至HTTPS即80端口跳转到443，详情参见[HTTP重定向至HTTPS](#)。

6. ping负载均衡的IP和ping负载均衡后端ECS的IP有什么区别？

ping负载均衡的IP是由负载均衡集群响应，不会转发给后端的ECS；ping负载均衡后端ECS是由后端ECS直接去响应，与前端的负载均衡没有关联关系。

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。

如果您的业务存在同时通过负载均衡和ECS提供对外服务的需要，但那么相应的ECS需要具备足够支撑业务的外网带宽。负载均衡的服务能力与后端ECS的公网带宽规格无关。

8. 禁用公网网卡是否影响负载均衡服务？

如果ECS有公网IP，禁用公网网卡就会影响负载均衡服务。

因为在有公网网卡的情况下，默认路由会走公网，如禁用就无法回包从而影响负载均衡服务。建议不要禁止，如一定要这么做，需要修改默认路由为私网才会不影响服务。但需要考虑业务是否对公网有依赖，如通过公网访问RDS等。

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。

10. 负载均衡均衡的是什么？

负载均衡是按特定调度算法把流量分发到后端ECS上，其中：

- 四层（TCP和UDP）是基于连接做流量调度。TCP和UDP创建一个socket访问负载均衡实例，这个源和目的IP和端口就是一个连接。
- 七层（HTTP/HTTPS）是基于请求做调度。比如http get请求访问一个页面。

11. 为什么请求不均衡？

负载均衡请求不均衡的原因可能有以下几种可能：

- 开启了会话保持功能

配置了会话保持，当访问负载均衡实例的客户端又很少时，容易导致不均衡，尤其在使用少量客户端对负载均衡进行测试的时候。比如TCP监听，开启了会话保持（四层是基于来源地址做会话保持），使用一台客户端对负载均衡实例进行压测，就会导致不均衡。

- ECS健康检查异常

后端服务器ECS的健康状态异常会导致不均衡，尤其在压测的时候容易忽略后端服务器ECS的健康检查状态，如果有后端服务器ECS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）必然会导致不均衡。

- TCP Keepalive保持长连接

后端服务器ECS有些开启了TCP Keepalive保持长连接，而有些又没有开启，则连接会在保持长连接的后端服务器上堆积，造成不均衡。

- 排查和解决方法

- 查看后端各台ECS的权重是否相同；
- 在相关时间段内是否有健康检查失败或波动现象，查找波动的原因；或者健康检查没有配置正确的响应码2xx，3xx导致了健康检查显示正常，但后端服务有异常；
- 是否同时使用了加权最小连接数（WLC）调度方式和会话保持，如果是，尝试改为加权加权轮询（WRR）算法和会话保持。

12. 为什么每个连接达不到带宽峰值？

因为负载均衡系统通过集群部署的方式为负载均衡实例提供服务，所有外部的访问请求都将平均分散到这些负载均衡系统服务器上进行转发。所以，设定的带宽峰值将被平均设定在多台系统服务器上。

单个连接下载的流量上限计算方法为：单个连接下载峰值=设置的负载均衡总带宽/(N-1)。N为流量转发分组个数，当前值为4。比如您在控制台上设置的是10Mb带宽上限，那么单个客户端可下载的最大流量为 $10/(4-1)=3.33\text{Mb}$ 。

基于负载均衡的实现原理，建议在配置单个监听的带宽峰值时根据您的业务情况并结合其实现方式来设定一个较为合理的值，从而确保您业务的正常对外服务不会受到影响和限制。

13. 为什么负载均衡压缩失败？

查看文件的content-type属性，如果文件类型不是text/xml、text/plain、text/css、application/javascript、application/x-javascript、application/rss+xml、application/atom+xml或application/xml，则压缩失败。

解决方法：

- 在源站修改该文件的content-type属性，改成负载均衡支持的文件类型。
- 将七层负载均衡实例监听修改为四层实例监听。

14. 如何查询负载均衡带宽和流量使用情况？

消费明细

打开费用中心管理控制台，在费用中心导航栏，单击消费记录 > 消费明细。产品选择负载均衡，查看负载均衡的消费明细。

流量使用情况

在负载均衡实例详情页面，单击监控，查看流量使用情况。

负载均衡的监控数据和您的消费明细中的流量数据会有些差距，详情查看[监控数据与计量数据](#)。

15. 为什么调用API修改带宽失败？

如果调用API修改带宽失败，报错Code:"InvalidParameter", "Message": "The specified parameter bandwidth is not valid."，则说明控制台上配置监听设置的带宽大于通过API修改的负载均衡实例带宽，这时是变更失败的。您需要在控制台上更改带宽限制。

16. 负载均衡监控数据与实际账单数据为什么不同？

- 负载均衡控制台监控指标中展示的实例流量数据是按照一分钟一个采集粒度，由负载均衡系统采集后上报云监控系统，再由云监控系统计算出每15分钟所有采集点的平均值；而负载均衡实例的账单计量数据是按照同样粒度的采集频度采集数据，在一个消费帐期后，以1个小时的累加值向账单计量系统上报并结算费用的。

账单数据是每一分钟的累加值，而监控数据是每15分钟平均值的累加值。所以，这两组数据在实际计算生成方式上是有区别的，两种数据不具备对比性。

- 负载均衡系统从采集数据到向云监控上报数据，然后云监控对数据进行平均值计算，最后展示在监控控制台的整个过程中，不可避免的存在一定的延迟。虽然这个延迟很小，我们也会尽量保证数据的实时性，但这种延迟也会导致其数据本身与账单计量数据存在一定程度的差异。用作计费的账单计量数据是可以容忍最多三小时延迟的，比如在01:00-02:00产生的账单计量数据，正常情况下会在03:00之前由负载均衡上报账单计量系统并进行计费，但是系统允许该上报时间最晚于05:00之前完成并计费。所以，从数据对实时性的要求不同来看，这两组数据也是不具备可比性的。
- 从监控和账单计量的产品定义上也是有区别的。监控的目的是为了通过一种手段来观察被监控实例的运行状态是否会出现异常，从而针对性地采取一定措施来解决由于异常导致的问题。账单计量的目的是根据实例的实际资源消耗情况进行计费。所以，从账单核算的出发点来看应该是以账单计量系统生产的数据来作为计费的判断依据而不应该以监控数据作为计费的判断依据。

17. 负载均衡各监听连接超时时间是多少？

- TCP监听：900秒
- UDP监听：90秒
- HTTP监听：60秒
- HTTPS监听：60秒

18. 为什么负载均衡服务地址会连接访问超时？

从服务端分析，以下情况会导致服务地址链接访问超时：

- 服务地址被安全防护

如流量黑洞和清洗，WAF防护（WAF的特点是建连后向客户端和服务器集群双向发送RST报文）。

- 客户端端口不足

尤其容易发生在压测的时候，客户端端口不足会导致建立连接失败，负载均衡默认会抹除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（socket设置SO_LINGER属性），而不是发FIN包这种方式断开。

- 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：默认的net.core.somaxconn的值为128，执行sysctl -w net.core.somaxconn=1024更改它的值，并重启后端服务器上的应用。

- 从四层负载均衡后端服务器访问该四层负载均衡的服务地址

四层负载均衡，在该负载均衡的后端服务器上去访问该负载均衡的服务地址会导致连接失败，常见的场景是后端应用使用URL拼接的方式跳转访问。

- 对连接超时的RST处理不当

负载均衡上建立TCP连接后，如果900s未活动，则会向客户端和服务器双向发送RST断开连接，有的应用对RST异常处理不当，可能会对已关闭的连接再次发送数据导致应用超时。

2.4 后端服务器常见问题

1. 负载均衡实例最多支持添加多少ECS实例？

默认每个SLB实例可以添加200个ECS实例，一个ECS实例可以被50个SLB实例关联。



说明：

如果您需要添加更多ECS实例，请提交工单申请。通过审核后，为您提升配额。

为了保证您对外服务的稳定与高效，建议您根据业务分类或应用服务的模块划分将提供不同服务或执行不同任务的应用服务器配置在不同的负载均衡实例上。

2. 负载均衡运行中是否可调整ECS数量？

可以。

您可以在任意时刻增加或减少负载均衡的后端ECS实例数量并且支持不同ECS实例之间的切换。但是为了保证您对外服务的稳定，确保在执行上述操作时，开启了负载均衡的健康检查功能并保证负载均衡后端至少有一台正常运行的ECS实例。

3. 后端ECS实例的操作系统是否可以不同？

可以。

负载均衡本身不会限制后端ECS实例使用哪种操作系统，只要确保后端ECS实例中的应用服务部署相同且数据一致即可。但建议使用相同的操作系统，以便您日后的管理维护。

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。负载均衡实例的地域和后端ECS实例的地域必须相同。

5. 后端ECS实例为什么访问不了负载均衡服务？

这和负载均衡TCP的实现机制有关。在四层TCP协议服务中，不支持后端ECS实例既作为Real Server又作为客户端向所在的负载均衡实例发送请求。因为返回的数据包只在云服务器内部转发，不经过负载均衡，所以在后端ECS实例上去访问负载均衡的服务地址是不通的。

6. ECS实例在接入负载均衡服务前添加了白名单，对负载均衡服务有影响么？

若ECS实例在接入负载均衡前设置了访问白名单，那么接入负载均衡后需要将负载均衡的白名单添加到ECS实例中，否则可能会引起网站403报错。

需要添加的负载均衡地址为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)。

7. 为什么有100开头的IP在频繁访问ECS实例？

负载均衡系统除了会通过系统服务器的内网IP将来自外部的访问请求转到后端ECS实例之外，还会对ECS实例进行健康检查和可用性监控，这些访问的来源都是由负载均衡系统发起的。

负载均衡系统的地址段为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)，所以会有很多100开头的IP地址访问ECS实例。

为了确保您对外服务的可用性，确保对上述地址的访问配置了放行规则。

8. 要使用负载均衡服务，ECS实例是否需要进行特殊配置？

添加到负载均衡实例的后端ECS实例，原则上不需要进行特别的配置。

但是，如果后端ECS实例的操作系统为Linux并且使用了TCP监听，确保系统配置文件`/etc/sysctl.conf`中以下三个配置的值为零：

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

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

如果部署在同一内网网段下的ECS实例之间有通信需求，需要检查如下参数的配置是否正确。更改配置后，执行**sysctl -p**更新配置。

```
net.ipv4.conf.default.arp_announce =2  
net.ipv4.conf.all.arp_announce =2
```

9. 部署了多站点的ECS实例可以添加到同一个负载均衡实例中么？

可以，详情参考[负载均衡+ECS多站点部署](#)。

一个负载均衡实例最多支持添加50个监听，每个监听配置对应后端ECS实例上的一个应用。您可以通过对后端ECS实例上的多个应用配置不同主机头的方式来满足这一需求。

10. ECS实例设置权重的作用是什么？

权重比越高的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需求来区别设定。

将负载均衡池中的A、B两台ECS实例的权重均配置成100和均配置成50都是一样的，在A、B两台ECS实例健康检查状态均正常的情况下，ECS实例A和B均分前端请求的流量。但是如果A和B的权重分别设置为10和100，在A、B两台ECS实例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。

如果您同时开启了会话保持功能，有可能会造成后端ECS实例访问不均衡，建议您暂时关闭会话保持功能再观察一下是否依然存在这种情况。

当负载均衡分发到后端ECS实例的请求不均匀时，参考以下方法检查处理：

1. 统计一个时间段内后端ECS实例的Web服务访问日志的数据量。

访问日志查看方法：

- Nginx和Apache：打开日志目录/*access.log*。
- IIS：打开IIS管理界面，鼠标放在需要设置开启日志的站点上，单击右键，选择属性，选择网站选项卡，单击启用日志记录。

2. 按照负载均衡的配置，对比多个ECS实例日志的数量是否有差距。

如果设置了会话保持，需要剥离相同IP的访问日志。如果负载均衡配置了权重，要根据权重比例计算日志中访问比例是否正常。

11. ECS实例上没有配置压缩，为什么从负载均衡返回的响应却被压缩了？

可能是客户端浏览器端支持压缩。您可以在控制台上创建监听时关闭Gzip压缩功能，或改用TCP监听。

12. ECS实例使用了HTTP1.0是否支持chunked transfer传输编码？

支持。

2.5 健康检查常见问题

1. 健康检查的原理是什么？

负载均衡通过健康检查来探测后端ECS实例的可用性。开启健康检查功能后，当后端某个ECS实例健康检查出现异常时，来自客户端的新请求将不会再被转发到该ECS实例，直到该ECS实例恢复正常。

如下图所示，负载均衡健康检查使用的地址段是100.64.0.0/10，后端服务器务必不能屏蔽该地址段。您无需在ECS安全组中额外针对该地址段配置放行策略，但如有配置iptables等安全策略，请务必放行（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）。

更多详细信息，参考[负载均衡健康检查原理](#)。

2. 推荐的健康检查配置是什么？

为了避免频繁的健康检查失败引起的切换对系统可用性的冲击，健康检查只有在健康检查时间窗内连续多次检查成功或失败后，才会进行状态切换。更多详细信息参见[健康检查配置](#)。

以下是TCP/HTTP/HTTPS监听建议使用的健康检查配置。

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

以下是UDP监听建议使用的健康检查配置。

配置	推荐值
响应超时时间	10秒

配置	推荐值
健康检查间隔	5秒
不健康阈值	3
健康阈值	3



说明：

此配置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，但必须先保证服务在正常状态下的处理时间小于这个值。

3. 是否可以关闭健康检查？

您只可能关闭HTTP/HTTPS监听的健康检查，不能关闭TCP/UDP监听的健康检查。具体操作，参考[关闭健康检查](#)。



说明：

如果关闭健康检查，当后端某个ECS实例健康检查出现异常时，负载均衡还是会把请求转发到该异常的ECS实例上，造成部分业务不可访问。建议您不要关闭健康检查。

4. TCP监听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Tengine节点服务器通过发送HTTP Head请求，对比返回码来校验后端服务是否异常。

TCP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健康检查；如果负载不是很高，则选择HTTP健康检查。

5. ECS实例权重设置为零对健康检查有什么影响？

该状态下，负载均衡不会再将流量转发给该ECS实例，且四层监听的后端服务器健康检查会显示异常（七层监听不会显示异常）。

将负载均衡后端ECS实例的权重置为零，相当于将该ECS实例移出负载均衡。一般是在ECS实例进行重启、配置调整等主动运维时将其权重设置为零。

6. HTTP监听向后端ECS实例执行健康检查使用的方法是什么？

HEAD方法。

如果后端ECS实例的服务关闭HEAD方法，会导致健康检查失败。建议在ECS实例上用HEAD方法访问自己IP地址进行测试：

```
echo -e "HEAD /test.html HTTP/1.0\r\n\r\n" | nc -t LAN_IP port
```

7. HTTP监听向后端ECS实例执行健康检查的IP地址是什么？

负载均衡健康检查使用的地址段是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)。如果后端ECS实例启用了iptables等访问控制，需要在内网网卡对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) 做访问放行。

8. 为什么健康检查监控频率与web日志记录不一致？

负载均衡健康检查服务也是集群方式的，这样可以避免单点故障。负载均衡的代理分布到很多节点上，因此看到的健康检查日志访问频率和控制台设置的频率不一致，这是正常现象。

9. 如何排查502 Bad Gateway错误？

问题现象：

访问负载均衡服务提示502 Bad Gateway并且健康检查提示异常。尽管Web日志提示404，但是ECS实例内网测试应用可以正常访问，ECS实例内部80端口监听也正常。

问题原因：

明配置的健康检查页面不存在，导致健康检查异常。

解决方案：

修改健康检查配置。

10. 健康检查是否会消耗系统资源？

HTTP模式的健康检查对后端ECS实例的资源消耗不大。

11. 为什么负载均衡后端ECS实例频繁收到UA为KeepAliveClient的请求？

问题现象：

负载均衡后端的ECS实例即使在没有用户访问时也会频繁收到GET请求，来源的IP是阿里云的内网IP，User-Agent显示为KeepAliveClient。

问题原因：

监听协议选择的是TCP，而健康检查选择了HTTP协议。此时负载均衡对后端ECS实例的健康检查就是以GET方式来进行，而不是HEAD方式了。

解决方案：

建议您将监听协议和健康检查协议统一设置为同一个协议。

12. 负载均衡因后端数据库故障导致健康检查失败，如何处理？

问题现象：

ECS实例内配置了两个网站，www.test.com是静态网站，app.test.com是动态网站，都配置了负载均衡。后端数据库服务异常，导致访问www.test.com静态网站出现502错误。

问题原因：

负载均衡健康检查配置的检查域名是app.test.com，RDS或者自建数据库故障导致app.test.com访问异常，所以健康检查失败。

解决方案：

将负载均衡健康检查域名配置为www.test.com即可。

13. 负载均衡服务TCP端口健康检查成功，为什么在后端业务日志中出现网络连接异常信息？

问题现象：

负载均衡后端配置TCP服务端口后，后端业务日志中频繁出现类似如下网络连接异常错误信息。经进抓包分析，发现相关请求来自负载均衡服务器，同时负载均衡主动向服务器发送了RST数据包。

问题原因：

该问题和负载均衡的健康检查机制有关。

由于TCP对上层业务状态无感知，同时，为了降低负载均衡健康检查成本和对后端业务的冲击，当前负载均衡针对TCP协议服务端口的健康检查只会做简单的TCP三次握手，而后直接发送RST包断开TCP连接。数据交互流程如下：

1. 负载均衡服务器向后端负载均衡服务端口发送SYN请求包；

2. 后端服务器收到请求后，如果端口状态正常，则按照正常的TCP机制返回相应的SYN+ACK应答包；
3. 负载均衡服务器成功收到后端服务端口应答后，则认为端口监听是正常的，判定健康检查成功；
4. 负载均衡服务器向相应TCP服务端口直接发送RST包主动关闭连接，结束本次健康检查操作，且没有继续发送业务数据。

如上所述，由于健康检查成功后，负载均衡服务器直接发送TCP RST包中断了连接，并没有做进一步的业务数据交互，导致上层业务（比如Java连接池等）认为相应的连接是异常的，所以会出现 `Connection reset by peer` 等错误信息。

解决方案：

- 更换TCP协议为HTTP协议。
- 在业务层面，对来自SLB服务器IP地址段的相关请求做日志过滤，忽略相关错误信息。

14. 为什么业务本身没有异常但是健康检查显示异常？

问题现象：

负载均衡HTTP方式的健康检查始终失败，但测试 `curl -I` 得到的状态码是正常的。

```
echo -e 'HEAD /test.html HTTP/1.0\r\n\r\n' | nc -t 192.168.0.1 80
```

问题原因：

如果返回的状态与控制台配置的正常状态码不一致，则判定健康检查异常。如果您配置的正常状态码为 `http_2xx`，则所有返回的非HTTP 2xx状态码均被认为是健康检查失败。

Tengine/Nginx配置会发现curl没有问题，但是echo测试会匹配到默认站点，导致测试文件 `test.html` 返回404错误，如下图所示。

解决方案：

- 修改主配置文件，将默认站点注释掉。
- 在健康检查配置中添加检查域名。

15. 如何排查负载均衡健康检查异常？

您可以通过以下方法进行排查：

1. 查看后端ECS实例是否能够正常提供服务。如果配置了HTTP模式的健康检查，确保服务返回的状态码与负载均衡控制台配置的正常状态码一致。
2. 负载均衡健康检查与后端ECS实例之间通过内网通信。登录服务器检查应用服务器端口是否正常监听在内网地址上，如果没有监听在内网地址，将应用服务器端口监听到内网上。
3. 确保ECS实例上10.0.0.0/8、100.64.0.0/10和11.0.0.0/8网段的路由配置的下一跳为内网网关。如果ECS实例只有内网地址，确保默认路由（0.0.0.0/0）指向内网网关。

各操作系统的路由配置如下：

- Windows: 登录ECS实例，下载[Windows添加路由工具](#)，双击执行下载文件。
- Linux: 登录ECS实例，下载[添加路由工具](#)，执行`bash linux_add_routes.sh`命令。
- FreeBSD: 登录ECS实例，下载[FreeBSD添加路由工具](#)，执行`bash freebsd_add_routes.sh`命令。

运行添加路由工具后内网路由显示正常，如下图所示。

4. 确保后端服务器开启了相应的端口，该端口必须与您在负载均衡监听配置中配置的后端端口保持一致。
5. 查看后端ECS实例权重值是否为零，如果权重值设置为零，会导致健康检查显示异常。
6. 检查后端ECS实例内部是否有防火墙或其他的安全类防护软件，这类软件可能屏蔽负载均衡系统的本地IP地址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），导致负载均衡系统无法和后端ECS实例通信。建议关闭防火墙或者卸载安全类防护软件进行测试。
7. 查看后端业务本身的响应时间是否超出健康检查设置的响应超时时间。

执行以下命令查看四层健康检查的业务响应时间：

```
time telnet <SLB IP address> <SLB Port>
```

执行以下命令查看七层健康检查的业务响应时间：

```
time echo -e 'HEAD <健康检查的检查路径> HTTP/1.0\r\n\r\n'|nc -t <端口>
```



说明：

在同账号同地域下的另外一台机器上执行该命令。比如健康检查的检查路径为/，ECS实例内网IP为192.168.0.1，端口为80，则您使用的命令为：`time echo -e 'HEAD / HTTP/1.0\r\n\r\n' | nc -t 192.168.0.1 80`

当结果中显示的**real**的参数值超过健康检查中配置的响应超时时间，健康检查就会提示异常。

解决方案：

- 优化应用或者程序，减少响应时间。
- 增加监控检查中设置的响应超时时间。

8. 七层的健康检查，执行`echo -e "HEAD /test.html HTTP/1.0\r\n\r\n" | nc -t LAN_IP 80`查看返回值是不是2XX或者3XX。



说明：

其中/test.html是健康检查配置的检查路径，需要根据实际情况进行替换。

9. 查后端ECS实例资源是否有较高负载导致ECS实例对外提供服务响应时间过长。
10. 建议提供健康检查的文件使用HTML静态文件，只用于检查返回结果，不建议使用PHP等动态脚本语言。

2.6 性能保障型实例收费常见问题

1. 何时开始收费？

负载均衡（SLB）于2017年5月开始推出性能保障型实例并开启免费公测，已经公测了10个月，计划在2018年3月31日结束公测，4月1日进入正式收费阶段。更多关于性能保障型实例的信息，参见[如何使用性能保障型实例#](#)。

性能保障型实例的规格费收取将按地域分批次生效：

- 第一批：

生效时间：4月1日至4月10日陆续生效

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

- 第二批：

生效时间：4月11日至4月20日陆续生效

生效地域：华东1（杭州）、华北3（张家口）、华北5（呼和浩特）、香港

- 第三批：

生效时间：4月21日至4月30日陆续生效

生效地域：华北1（青岛）、华北2（北京）、华东2（上海）、华南1（深圳）

2. 已有包年包月保障型实例4月1日开始会收费吗？

在到期前无需补交性能规格费，但是从4月1日起如果您进行了升级或续费操作，会收取规格费。

3. 私网负载均衡实例也会收取规格费吗？

- 如果您选择的是性能共享型私网实例，则不收取规格费。
- 如果您选择的是性能保障型私网实例，则需要收取规格费。

规格费收取方式与公网实例规格费计费规则一致。私网实例免收实例费和流量费。

4. 收取规格费以后，共享型实例也会额外收取费用吗？

不会。

原有的共享型实例（如果您不将其变更成性能保障型）将继续保持为性能共享型实例，不收取规格费。

但如果您将其变更成性能保障型，将从4月1日起收取规格费。

5. 共享型实例可以变更成保障型实例么？

可以。

变更以后将从4月1起收取规格费，且不能再变更回性能共享型，即性能保障型实例无法变更为共享型。

6. 共享型实例性能规格如何？

性能共享型实例不提供性能保障，没有可提供的规格。

7. 性能保障型实例有很多规格，怎么选？

- 后付费实例（按量付费），可以选择最大规格，因为后付费实例的规格费按照实际使用量收取，闲时免费。
- 预付费实例（包年包月），需要根据您的实际业务量选择，您可以从以下几个方面对业务量进行评估：

- 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。

建议您先使用按量付费的实例进行业务测试，确认好规格后再购买所需规格的包年包月实例。更多信息参见[如何选择性能保障型实例](#)。

8. 性能保障型实例各规格的费用是多少？

负载均衡性能保障型实例开放了六种实例规格，其中规格1简约型l (slb.s1.small)免收规格费。详细信息参见[性能保障型实例规格费的定价](#)。

9. 性能保障型实例的流量费和实例费与共享型一样吗？

一样。

10. 性能保障型实例的购买上限是多少？

和性能共享型一样，默认可购买60个按量付费的实例。如有更多需求，请提交工单申请。

11. 是否可以调整性能保障型实例的规格？

可以。

- 按量付费的性能保障型实例的规格可以升配也可以降配，详情参见[后付费实例变配](#)。
- 包年包月的性能保障型实例需要开通白名单才可以降配。详情参考[包年包月实例变配](#)。



说明：

- 将性能共享型实例变更为性能保障型实例后，无法再将其变更回性能共享型。
- 变更性能保障型实例规格时，如果同时变更计费方式(按流量计费或按带宽计费)，则规格变更需要到次日零点才能生效。如果仅仅是对实例规格进行变更，变更立即生效。建议您在变更规格时，尽量不要变更计费方式。
- 由于历史存量原因，部分实例可能存在于较老的集群。此部分实例在变配到性能保障型实例时，因为需要将实例迁移，因此可能出现10-30秒的业务中断，因此建议在业务低谷期进行此类变配，再进行变配。

12. 还可以购买性能共享型实例吗？

可以。

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

2.7 如何使用负载均衡性能保障型实例？

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

1. 什么是负载均衡性能保障型实例？

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

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

把鼠标移至性能保障型实例的绿色图标，可查看具体的性能指标，如下图所示。

负载均衡ID/名称	可用区	服务地址(全部)	状态	网络类型(全部)	端口/健康检查	后端服务器	实例规格	带宽计费方式(全部)	付费方式(全部)	操作
☐ lb-bp1rlw0l...	华东 1 可用区 F(主)	120.55	运行中	经典网络	未配置(配置)	未配置(配置)	性能共享型	按固定带宽	包年包月	管理 更多
	华东 1 可用区 E(备)	公网							2017-09-08 00:00:00 到期	
☐ lb-wz9nznspn...	华南 1 可用区 A(主)	120.78	运行中	经典网络	未配置(配置)	未配置(配置)	性能保障型 slb.s1.small	连接数: 5000 CPS: 3000 QPS: 1000	按量付费	管理 更多
	华南 1 可用区 B(备)	公网							2017-08-07 21:48:36 创建	

性能保障型实例的三个关键指标如下：

- 最大连接数-Max Connection

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

- 每秒新建连接数-Connection Per Second (CPS)

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

- 每秒查询数-Query Per Second (QPS)

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

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

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

如果需要更大规格，请联系您的客户经理申请。

2. 性能保障型实例如何收费？

负载均衡性能保障型实例需要收取规格费用，收费模型如下：

性能保障型费用 = 实例费 + 流量/带宽费 + 规格费



说明：

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

负载均衡分为两种计费模式，预付费和按量付费。在不同的计费模式下，性能保障型实例的规格费收取规则不同：

- 预付费模式

性能保障型实例规格费按照预付费模式收取，即在实例的付费周期内，实例规格费按照固定的价格收取。假设您选择的是高阶型I (slb.s3.small)规格，并且选择购买时长为3个月，则规格费用 = slb.s3.small规格费月价 x 3月。

- 按量付费模式

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

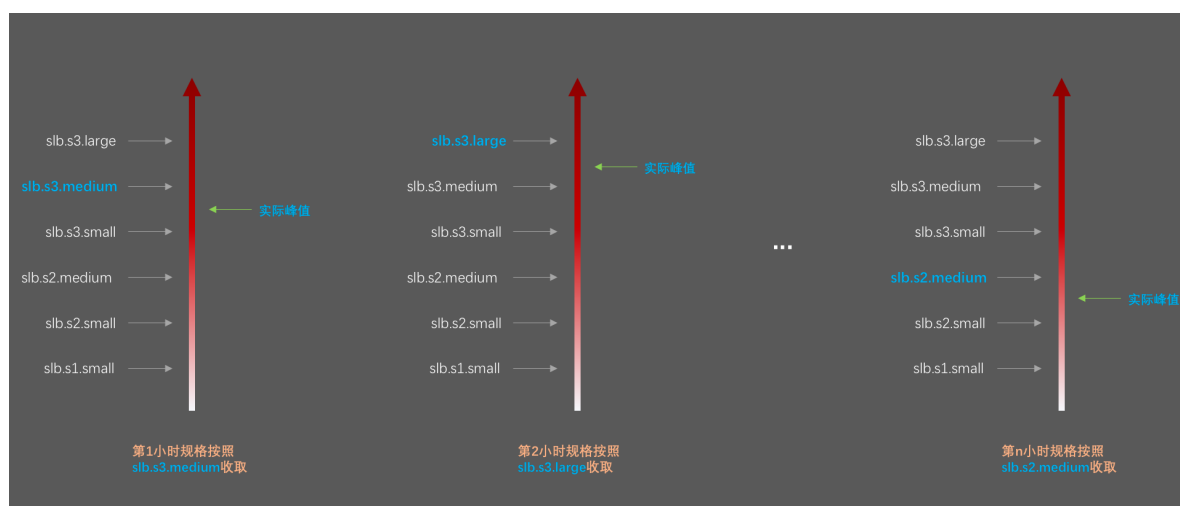
例如，您选择了超强型I (slb.s3.large)规格 (最大连接数1,000,000；CPS 500,000；QPS 500,00)。您的实例在某个小时内各项指标产生的实际峰值如下：

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

- 从最大连接数维度看，90,000超过slb.s2.small规格中最大连接数50,000的上限，但未达到slb.s2.medium规格中最大连接数的100,000上限，因此从最大连接数维度计算，该小时规格为slb.s2.medium。
- 从每秒新建连接数 (CPS) 维度看，4,000超过slb.s1.small规格中CPS 3,000的上限，但未到达slb.s2.small规格中CPS 5,000的上限，因此从CPS维度计算，该小时规格为slb.s2.small。
- 从每秒查询数 (QPS) 维度看，11,000超过slb.s2.medium规格中QPS 10,000的上限，但未达到slb.s3.small中QPS 20,000的上限，因此从QPS维度计算，该小时规格为slb.s3.small。

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

以后每小时规格费均按照上述方式计算，如下图所示：



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

3. 性能保障型实例规格费的定价

下表中所列的只是规格费用。除规格费以外，负载均衡实例的实例配费和流量/带宽费正常收取。更多详细信息，参考[计费说明](#)。

地域	规格	规格	最大连接数	每秒新建连接数 (CPS)	每秒查询数(QPS)	包年包月(元/月)	按量付费(元/时)
华东1 (杭州) 华北3 (张家口) 华北5 (呼和浩特) 华北1 (青岛) 华北2 (北京) 华东2 (上海) 华南1 (深圳)	规格 1	简约型I (slb.s1.small)	5,000	3,000	1,000	免费	免费
	规格 2	标准型I (slb.s2.small)	50,000	5,000	5,000	190.00	0.32
	规格 3	标准型II (slb.s2.medium)	100,000	10,000	10,000	380.00	0.63
	规格 4	高阶型I (slb.s3.small)	200,000	20,000	20,000	760.00	1.27
	规格 5	高阶型II (slb.s3.medium)	500,000	50,000	30,000	1,143.00	1.91
	规格 6	超强型I (slb.s3.large)	1,000,000	100,000	50,000	1,908.00	3.18
亚太东南1 (新加坡) 亚太东南3 (吉隆坡) 亚太东南5 (雅加达) 亚太南部1 (孟买)	规格 1	简约型I (slb.s1.small)	5,000	3,000	1,000	免费	免费
	规格 2	标准型I (slb.s2.small)	50,000	5,000	5,000	228.00	0.38
	规格 3	标准型II (slb.s2.medium)	100,000	10,000	10,000	456.00	0.76

地域	规格	规格	最大连接数	每秒新建连接数 (CPS)	每秒查询数(QPS)	包年包月(元/月)	按量付费(元/时)
美国西部1 (硅谷) 美国东部1 (弗吉尼亚) 香港	规格 4	高阶型I (slb.s3.small)	200,000	20,000	20,000	912.00	1.52
	规格 5	高阶型II (slb.s3.medium)	500,000	50,000	30,000	1,372.00	2.29
	规格 6	超强型I (slb.s3.large)	1,000,000	100,000	50,000	2,290.00	3.82

目前海外地域规格费享83折优惠。

4. 如何选择性能保障型实例？

- 如果您购买的是按量付费实例，如上文所描述，规格费是按量（弹性）计费的，因此建议您直接选择您可以买到的最大规格，对于大多数用户而言，即高阶型I(slb.s3.large)，这样可以保证较好的业务灵活性（弹性），且不会让您额外多付出成本。但如果您认为您的业务量不太可能到达超强型I(slb.s3.large)，也可以设置一个合理的弹性上限，比如高阶型II(slb.s3.medium)。
- 如果您购买的是预付费实例，情况会略微复杂一点。因为规格费按照固定费率恒定收取，而您不希望购买一个超出您实际业务量很多的规格，并因此付出不必要的成本，因此您需要评估您的实际业务量，并合理的考虑一些冗余，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。更多详细信息，参考[监控数据](#)。

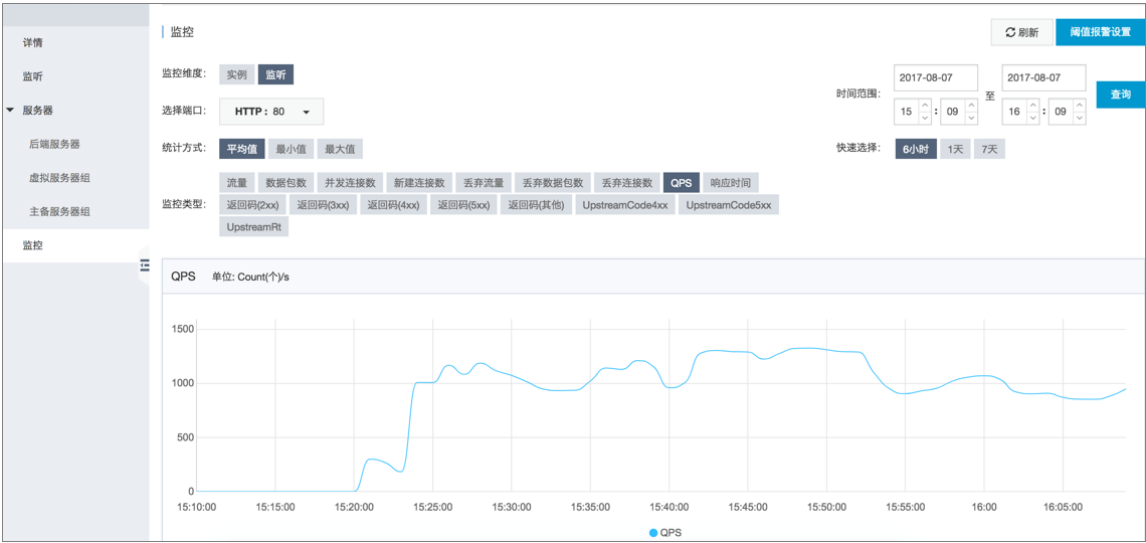
并发连接数监控示例



新建连接数监控示例



QPS监控示例



5. 是否可以调整性能保障型实例的规格？

您可在控制台对性能保障型实例进行变配，如下图所示。

负载均衡ID/名称	可用区	服务地址(全部)	状态	网络类型(全部)	端口/健康检查	后端服务器	实例规格	带宽计费方式(全部)	付费方式(全部)	操作
slb-2z6w9ub0...	华北 2 可用区 B(主)	47.136.107.140	运行中	经典网络	未配置(配置)	未配置(配置)	性能保障型 slb.s1.small	按固定带宽	包年包月 2017-10-05	管理 更多
slb-2z6w9ub0...	华北 2 可用区 B(主)	47.136.107.140	运行中	经典网络	未配置(配置)	未配置(配置)	性能共享型	按使用流量	按量付费 2017-09-01	启动 停止 编辑标签 变更配置 续费
slb-2z6w9ub0...	华北 2 可用区 B(主)	47.136.107.140	运行中	经典网络	未配置(配置)	未配置(配置)	性能共享型	按使用流量	按量付费 2017-09-01	

当前配置

实例名称: slb-2z6w9ub0...

计费项: 配置费(公网IP+性能规格) + 带宽费	实例规格: 简约型I (slb.s1.small)	主可用区: cn-beijing-b	计费周期: 1个月
备可用区: cn-beijing-a	带宽值: 483Mbps	流量: 按固定带宽计费	实例: 公网
地域: 华北 2(北京)	slb服务: 是	云盾: 是	可用区类型: 多可用区

到期时间: 2017-10-05 00:00:00

配置变更

实例规格：

标准型 (slb.s2.small)

连接数: 50000, 新建连接数 (CPS): 5000, 每秒查询数 (QPS): 5000 负载均衡推出性能保障型实例, 提供多种规格的实例, 并提供相应规格下的性能保障。各地域将陆续升级为性能保障型实例, 暂对用户免收性能规格费, 具体费用和收费计划 [详情点击>>](#)

实例类型：

公网

[实例类型详解>>](#)

公网带宽：

按固定带宽计费

开通后即开始按固定带宽计费, 和实例状态及使用流量无关

带宽值：

1250Mbps

2500Mbps

5000Mbps

483 Mbps

开通后即开始按固定带宽计费, 和实例状态及使用流量无关

按量付费的性能保障型实例的规格可以升配也可以降配，包年包月的性能保障型实例需要开通白名单才可以降配。详情参考[包年包月实例变配](#)。

因此，建议您先使用按量付费的实例进行业务测试，确认好规格后再购买所需规格的包年包月实例。



说明：

- 将性能共享型实例变更为性能保障型实例后，无法再将其变更回性能共享型。
- 变更性能保障型实例规格时，如果同时变更计费方式(按流量计费或按带宽计费)，则规格变更需要到次日零点才能生效。如果仅仅是对实例规格进行变更，变更立即生效。建议您在变更规格时，尽量不要变更计费方式。
- 由于历史存量原因，部分实例可能存在于较老的集群。此部分实例在变配到性能保障型实例时，因为需要将实例迁移，因此可能出现10-30秒的业务中断，其他变配操作均不会影响业务。因此建议在业务低谷期进行此类变配，或通过[GSLB](#)来做实例间的负载均衡后，再进行变配。
- 所有的变配操作都不影响负载均衡实例的IP地址。

警告!!!

请注意：

如将性能共享型实例变更为性能保障型实例，SLB将有小概率出现短暂的业务中断（10秒-30秒），建议您在业务低谷期进行变配，或者使用[GSLB](#)将业务调度至其他的SLB实例后，再进行变配，仅对计费方式和带宽进行变更，业务不受上述影响。

性能共享型实例变配为性能保障型实例后，无法再变回性能共享型实例！

我已知晓上述风险，继续

取消

6. 性能保障型实例何时收费？

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

性能保障型实例的规格费收取将按地域分批次生效：

- 第一批：

生效时间：4月1日至4月10日陆续生效

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

- 第二批：

生效时间：4月11日至4月20日陆续生效

生效地域：华东1（杭州）、华北3（张家口）、华北5（呼和浩特）、香港

- 第三批：

生效时间：4月21日至4月30日陆续生效

生效地域：华北1（青岛）、华北2（北京）、华东2（上海）、华南1（深圳）

7. 收取规格费以后，性能共享型实例需也会额外收取费用吗？

不会。

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

8. 为何有时性能保障型实例看起来达不到规格中的性能指标上限？

短板板原理。

性能保障型实例并不保障三个指标（包含带宽指标）同时达到指定规格的指标上限。即规格中哪个指标先达到峰值，就以哪个指标开始限速。同样，如果购买了按带宽付费的实例，当实例带宽达到峰值上限时，也可能出现因为带宽限速而导致某些指标达不到规格上限的情况。

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

9. 还可以购买性能共享型实例吗？

可以。

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

10. 私网负载均衡实例也会收取规格费吗？

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

2.8 为什么请求不均衡

原因

负载均衡请求不均衡的原因可能有以下几种可能：

- 开启了会话保持功能

配置了会话保持，当访问负载均衡实例的客户端又很少时，容易导致不均衡，尤其在使用少量客户端对负载均衡进行测试的时候。比如TCP监听，开启了会话保持（四层是基于来源地址做会话保持），使用一台客户端对负载均衡实例进行压测，就会导致不均衡。

- ECS健康检查异常

后端服务器ECS的健康状态异常会导致不均衡，尤其在压测的时候容易忽略后端服务器ECS的健康检查状态，如果有后端服务器ECS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）必然会导致不均衡。

- TCP Keepalive保持长连接

后端服务器ECS有些开启了TCP Keepalive保持长连接，而有些又没有开启，则连接会在保持长连接的后端服务器上堆积，造成不均衡。

排查和解决方法

- 查看后端各台ECS的权重是否相同；
- 在相关时间段内是否有健康检查失败或波动现象，查找波动的原因；或者健康检查没有配置正确的响应码2xx，3xx导致了健康检查显示正常，但后端服务有异常；
- 是否同时使用了加权最小连接数（WLC）调度方式和会话保持，如果是，尝试改为加权加权轮询（WRR）算法和会话保持。

2.9 如何实现相同域名不同路径的流量转发

背景信息

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	服务器SLB_tom1和SBL_tom2
www.aaa.com/jerry	服务器SLB_jerry1和SBL_jerry2

操作步骤

1. 创建一个公网负载均衡实例。

详情参考[创建负载均衡实例](#)。

2. 使用DNS将域名解析为负载均衡实例的公网IP。

为了方便演示，本操作中，将负载均衡实例的公网IP和域名www.aaa.com在host文件中进行了绑定。

3. 创建两个虚拟服务器组。

a) 在负载均衡管理控制台，找到新建的目标实例，然后单击实例ID进入详情页面。

b) 在左侧导航栏，单击服务器 > 虚拟服务器组。

c) 单击创建虚拟服务器组。

- d) 在弹出的对话框中，选择要添加的后端服务器，并设置端口和权重。虚拟服务器组内的ECS的端口可以不同。

本示例中，输入**TOM**作为服务器组的名称，选择添加服务器SLB_tom1和SBL_tom2，将端口设置为80，权重保持默认（100）。

- e) 重复上述步骤，添加另外一个名为JERRY的虚拟服务器组，包含服务器SLB_jerry1和SBL_jerry2。

4. 添加监听。

- a) 在左侧导航栏，单击监听，然后单击添加监听。

- b) 配置监听。本示例中的监听配置如下：

- 前端协议 [端口]：HTTP：80
- 后端协议 [端口]：HTTP：80
- 调度算法：轮询。
- 其它配置使用默认选项。

- c) 在监听页面，单击更多 > 添加转发策略。

- d) 在转发策略页面，单击添加转发策略。

- e) 配置三条转发规则。

5. 测试：

- 在浏览器中输入www.aaa.com/jerry，将返回如下结果。
- 在浏览器中输入www.aaa.com/tom，将返回如下结果。

在浏览器中输入www.aaa.com，将返回如下结果。

2.10 会话保持常见问题

1. 会话保持的作用是什么？

将同一客户端的会话请求转发给指定的一个后端服务器处理。

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）。

3. 负载均衡支持什么类型的会话保持？

- 四层（TCP协议）服务，负载均衡系统是基于源IP的会话保持。四层会话保持的最长时间是3600秒。
- 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字符串），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。服务器配置参考[配置会话保持规则](#)。

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。

您可以通过负载均衡系统提供的重写cookie的会话保持方式来实现这一需求。

6. cookie的超时时间应设置为多少？

- 植入cookie可在控制台上配置1-86400秒。

- 重写cookie需要您在后端ECS上维护超时时间。

7. 如何查看会话保持字串？

可以在浏览器中用F12查看回应报文中是否含有SERVERID字串或用户指定的关键字，或者运行 `curl www.xxx.com -c /tmp/cookie123` 保存一下cookie，再用 `curl www.xxx.com -b /tmp/cookie123` 访问。

8. 为什么有时候会话保持失败？

- 查看是否在监听配置中已经开启了会话保持功能。
- HTTP/HTTPS监听在后端服务器返回 4xx 响应码的报文中无法插入会话保持所需cookie。

解决方案：改用TCP监听，因为TCP监听是以源客户端的IP来做会话保持的，另外后端ECS上也可以插入cookie，并增加cookie的判断来多重保障。

- 302重定向会改变会话保持中的SERVERID字串。

负载均衡植入cookie时，如果后端ECS中有回复302重定向的报文，将改变会话保持中的SERVERID字串，导致会话保持失效。

排查方法：在浏览器端捕获请求与响应的回复，或用抓包软件抓包后分析是否存在302的响应报文，对比前后报文的cookie中的SERVERID字串是否不同了。

解决方案：改用TCP监听，因为TCP监听是以源客户端的IP来做会话保持的，另外后端ECS上也可以插入cookie，并增加cookie的判断来多重保障。

- 会话保持时间设置过小，会话保持时间过小也会导致会话保持失败。

9. 如何使用Linux curl测试负载均衡会话保持？

1. 创建测试页面。

在负载均衡所有后端ECS中创建测试页面，如下图所示页面中能显示本机内网IP。内网IP用于判断相应请求被指派到的物理服务器。通过观察该IP的一致性，来判断负载均衡会话保持的有效性。

2. Linux下curl测试。

假设负载均衡服务IP地址是 1.1.1.1，创建的测试页面URL为：<http://1.1.1.1/check.jsp>

- a. 登录用来测试的Linux服务器。

- b. 执行以下命令负载均衡服务器cookie值。

```
curl -c test.cookie http://1.1.1.1/check.jsp
```



说明：

阿里云负载均衡会话保持默认模式是植入cookie，而curl测试默认是不会保存和发送cookie的。所以必须先保存相应的cookie，用于cookie测试。否则，curl测试结果是随机的，会误认为负载均衡会话保持无效。

- c. 执行以下命令持续测试。

```
for ((a=1;a<=30;a++)); do curl -b="" -c test.cookie http://1.1.1.1/check.jsp >/dev/null | grep '10.170.*'; sleep 1; done
```



说明：

a<=30是重复测试次数，可以按需修改；grep '10.170.*' 是筛选显示的IP信息，根据后端ECS内网IP情况进行相应修改；

- d. 观察上述测试返回的IP，如果是同一台ECS内网IP，则证明负载均衡会话保持有效；反之则证明负载均衡会话保持有问题。

2.11 HTTP/2协议支持常见问题

什么是HTTP/2？

HTTP/2 (Hypertext Transfer Protocol Version 2) 是超文本传输协议的第二版，向下兼容HTTP1.X协议版本，同时带来性能的大幅提升。

与HTTP/1.X相比HTTP/2的优势如下：

- 多路复用 (Multiplexing): 多路复用允许同时通过单一的HTTP/2连接发起多重的请求-响应消息。
- 二进制分帧与首部压缩: 提升了数据在网络中传输的效率。
- 服务端推送(Server Push): 服务端可以主动往客户端推送数据，以减少请求次数提升效率。
- 除此以外还有流量控制、主动重置请求、请求优先级等众多特性使得Web服务性能得以大幅提升，如下图所示。
-

如何在阿里云负载均衡上启用HTTP/2协议？

无需配置，HTTPS监听上默认支持HTTP/2协议。



说明：

需要将实例升级为性能保障型实例。详细参见[如何使用负载均衡性能保障型实例](#)。

负载均衡HTTPS监听会侦测客户端发起的握手报文ClientHello中的ALPN字段，进行协议版本的协商。若客户端ClientHello报文中没有携带ALPN字段，则按照HTTP/1.x的请求逻辑来处理，反之，按照HTTP/2的请求逻辑处理。



说明：

HTTP/2协议仅在HTTPS监听上支持，不支持明文HTTP/2(HTTP/2 Cleartext)。

支持的地域

全部地域已支持HTTP/2协议。

使用限制

HTTP/2协议支持的约束如下：

- HTTP/2协议仅在HTTPS监听上支持，不支持明文HTTP2(HTTP/2 Cleartext)
- 当前HTTP/2只启用在客户端到SLB之间的链路，SLB到后端服务器之间的连接依然是HTTP/1.X
- HTTP/2的请求会统计到监听/实例的QPS中，与HTTP/1.X的QPS请求统一计数
- 在HTTP/2链路中所有后端发给SLB的回应，根据协议要求，会将head中所有字段的名称改为小写字母，如Content-Type会被改写为content-type
- 单个连接支持的最大并发stream数为128
- HTTP/2的连接超时时间为180秒，没有例外

收费政策

HTTP/2协议支持不额外收取费用。

2.12 HTTPS/HTTP监听常见问题

1. 为什么请求经过七层负载均衡转发后，后端服务器的响应头中的某些参数会被删除？

为了实现会话保持，负载均衡会修改后端服务器响应头中的Date、Server、X-Pad和X-Accel-Redirect等参数值。

解决方案：

- 在自定义的报文头部中加入一个前缀，如xl-server或xl-date，以避免负载均衡的处理。

- 将七层HTTP监听改为四层TCP监听。

2. 为什么在HTTP请求的头部增加了Transfer-Encoding: chunked字段？

将域名解析到七层负载均衡的服务地址后，从本地主机访问域名时发现在HTTP请求的头部增加了一个Transfer-Encoding: chunked字段，但是从本地主机直接访问后端服务器时是没有这个字段的。

由于七层负载均衡基于Tengine反向代理实现。Transfer-Encoding字段表示Web服务器如何对响应消息体编码，比如Transfer-Encoding: chunked表示Web服务器对响应消息体做了分块传输。



说明：

在四层负载均衡服务中，负载均衡仅转发流量，不存在该字段。

3. 为什么HTTP监听访问正常但HTTPS监听打开网址不加载样式？

现象：

分别创建HTTP和HTTPS监听，两个监听使用同样的后端服务器。以HTTP方式访问监听端口对应的网站时，网站正常显示，但使用HTTPS监听访问时，网站排版显示错乱。

原因：

负载均衡默认是不会屏蔽JS文件加载传输的，可能原因：

- 证书和浏览器安全级别不兼容导致。
- 证书是非正规第三方证书需要联系证书发布者检查证书问题。

解决方案：

1. 打开网站时，按照浏览器提示加载脚本。
2. 在客户端中添加对应证书。

4. HTTPS监听使用什么端口？

HTTPS监听对端口无特殊要求，建议您使用443端口。

5. 负载均衡支持哪些类型的证书？

支持上传PEM格式的服务器证书和CA证书。

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。

6. 负载均衡是否支持keytool创建的证书？

支持。

但在上传证书前，您需要将证书转换为PEM格式，详情参见[转换证书格式](#)。

7. 可以使用PKCS#12 (PFX) 格式的证书么？

可以。

但在上传证书前，您需要将证书转换为PEM格式，详情参见[转换证书格式](#)。

8. 一个账号可以上传多少个证书？

每个账号最多可上传100个证书，包含CA证书和服务器证书。

9. 添加证书时，为什么会出现KeyEncryption的错误？

该错误由于私钥内容有误导致。关于私钥格式说明，参见[证书要求](#)。

10. 一个HTTPS监听可绑定多少证书？

性能共享型负载均衡实例若使用HTTPS单向认证，则一个监听只能绑定一个服务器证书；若使用HTTPS双向认证，则一个监听需要绑定一个服务器证书和一个CA证书。

性能保障型负载均衡HTTPS监听支持挂载多个证书，将来自不同访问域名的请求转发至不同的后端服务器组，详细操作请参见[配置教程](#)。

11. 负载均衡HTTPS支持哪些SSL协议版本？

TLSv1、TLSv1.1以及TLSv1.2。

12. 为什么HTTPS协议实际产生的流量会比账单流量多一些？

HTTPS协议会使用一些流量用于协议握手，因此其实际产生的流量会多于账单流量。

13. HTTPS session ticket的保持时间是多久？

HTTPS session ticket保持时间为300秒。

14. 可以上传包含DH PARAMETERS字段的证书么？

HTTPS监听使用的ECDHE算法簇支持前向保密技术，不支持将DHE算法簇所需要的安全增强参数文件上传，即不支持将PEM证书文件中含BEGIN DH PARAMETERS字段的证书上传。

15. HTTPS监听是否支持SNI？

SNI (Server Name Indication) 是为了解决一个服务器使用多个域名和证书的SSL/TLS扩展，负载均衡HTTPS监听支持SNI功能，具体请参见[配置教程](#)。

16. HTTP/HTTPS监听访问后端服务器的HTTP协议版本是什么？

HTTP/1.0。

17. 后端服务器能否获取客户端访问HTTP/HTTPS监听的协议版本？

可以。

18. 一个请求通过负载均衡到达后端服务器，如果客户端在未收到后端服务器的回复前主动断开和负载均衡的连接，负载均衡会同时断开和后端服务器的连接么？

负载均衡在读写过程中不会断开与后端服务器的连接。

19. HTTP/HTTPS监听是否支持WebSocket/SSL WebSocket？

全部地域都已支持WSS/WS协议，详情参见[WS/WSS协议支持常见问题](#)。

20. HTTP/HTTPS连接的超时时间是如何规定的？

- HTTP长连接的请求数量限定是最多连续发送100个请求，超过限定将关闭这条连接。
- HTTP长连接两个HTTP/HTTPS请求之间的超时时间为15秒（存在误差1-2秒），超过后会关闭TCP连接，如果用户有长连接使用需求请尽量保持在13秒之内发送一个心跳请求。
- 负载均衡与后端一台ECS实例TCP三次握手完成过程的超时时间为5秒，超时后选择下一台ECS实例；查询访问日志的upstream响应时间可以定位。
- 负载均衡等待一台ECS实例回复请求的响应时间是60秒，超过后一般会返回504响应码或408响应码给客户端；查询访问日志的upstream响应时间可以定位。
- HTTPS session重用超时间为300秒，超过后同一客户端需要重新进行完整的SSL握手过程。

21. 负载均衡是否支持配置域名和URL转发策略？

支持，详情参见[配置域名URL转发策略](#)。

22. 每个监听可以添加多少条域名和URL转发规则？

每个监听最多可添加20条转发规则。

2.13 WS/WSS协议支持常见问题

什么是WS/WSS？

WebSocket (WS)是HTML5一种新的协议。它实现了浏览器与服务器全双工通信，能更好地节省服务器资源和带宽并达到实时通讯。WebSocket建立在TCP之上，同HTTP一样通过TCP来传输数据，但是它和HTTP最大不同是：

WebSocket是一种双向通信协议，在建立连接后，WebSocket服务器和Browser/Client Agent都能主动的向对方发送或接收数据，就像Socket一样；WebSocket需要类似TCP的客户端和服务端通过握手连接，连接成功后才能相互通信。

WSS (Web Socket Secure) 是WebSocket的加密版本。

为何使用WS/WSS？

随着互联网的蓬勃发展，各种类型的Web应用层出不穷，很多应用要求服务端有能力进行实时推送能力（比如直播间聊天室），以往很多网站为了实现推送技术，所用的技术都是轮询。轮询是在特定的时间间隔（如每1秒），由浏览器对服务器发出HTTP请求，然后由服务器返回最新的数据给客户端的浏览器。这种传统的模式带来很明显的缺点，即浏览器需要不断地向服务器发出请求，然而HTTP请求可能包含较长的头部，其中真正有效的数据可能只是很小的一部分，显然这样会浪费很多的带宽资源。

在这种情况下，HTML5定义了WebSocket协议，能更好地节省服务器资源和带宽，并且能够更实时地进行通讯。WebSocket实现了浏览器与服务器全双工(full-duplex)通信，允许服务器主动发送信息给客户端。

WebSocket协议的交互过程如下图所示。

如何在阿里云负载均衡上启用WS/WSS支持？

无需配置，当选用HTTP监听时，默认支持无加密版本WebSocket协议（WS协议）；当选择HTTPS监听时，默认支持加密版本的WebSocket协议（WSS协议）。



说明：

需要将实例升级为性能保障型实例。详细参见[如何使用负载均衡性能保障型实例](#)。

支持的地域

全部地域都已开放WS/WSS支持。

限制

WSS/WS协议支持的约束如下：

- 负载均衡与ECS后端服务的连接采用HTTP/1.1，建议后端服务器采用支持HTTP/1.1的Web Server。
- 若负载均衡与后端服务超过60秒无消息交互，会主动断开连接，如需要维持连接一直不中断，需要主动实现保活机制，每60秒内进行一次报文交互。

收费政策

WSS/WS协议支持不额外收取费用。