

负载均衡

最佳实践

最佳实践

什么是负载均衡性能保障型实例？

性能保障型实例提供了可保障的性能指标（性能SLA）。与之相对的是性能共享型实例，即不保障实例的性能指标，资源是所有实例共享的。

在阿里云负载均衡推出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-bp1inw0...	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
lb-wz9nzsprn...	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

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提供了如下六种实例规格。

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)	5000	3000	1000
规格 2	标准型I	50000	5000	5000

	(slb.s2.small)			
规格 3	标准型II (slb.s2.medium)	100000	10000	10000
规格 4	高阶型I (slb.s3.small)	200000	20000	20000
规格 5	高阶型II (slb.s3.medium)	500000	50000	30000
规格 6	超强型I (slb.s3.large)	1000000	100000	50000

如何选择性能保障型实例？

您需要结合具体业务来选择性能规格，整体原则如下：

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



性能保障型实例的变配操作限制

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
lb-2zcn09y...	华北2 可用区 B(主)	47.96.101.100	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
lb-2zcn09y...	华北2 可用区 B(主)	47.96.101.100	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
lb-2zcn09y...	华北2 可用区 B(主)	47.96.101.100	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	变更配置 续费

当前配置

实例名称: lb-2zcn09y...

计费项: 配置费(公网IP+性能规格)+带宽费 实例规格: 简约型 (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



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目前暂时只允许升配，不允许降配。

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。

另外，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。

注意：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，或通过GSLB来做实例间的负载均衡后，再进行变配。

警告!!!

请注意，在变更配置时，如果您变更实例的规格，或者将共享型实例变更为保障型实例，SLB将有小概率可能性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，再进行变配操作。(仅对计费方式和带宽进行变配，业务不受任何影响)

我已知晓上述风险，继续
取消

性能保障型实例售卖开放计划

阿里云负载均衡从5月份开始在美东1地域小范围公测性能保障型实例，目前已经开放性能保障型实例地域和后续计划如下：

- 美东1（弗吉尼亚）：已上线
- 华北2（北京）：已上线
- 华东1（杭州）：已上线
- 华东2（上海）：已上线
- 华北1（青岛）：已上线
- 华北3（张家口）及海外各地域：后续陆续上限开放，敬请关注

性能保障型实例的定价

性能保障型实例根据不同规格收取规格费，如下表所示。阿里云负载均衡提供了一种免费的规格，可满足以前绝大部分共享型实例用户的需求。

注意：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置费用和流量费保持不变。更多详细信息，参考计费说明。

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		按量付费
					月价(元/月)	年价(元/年)	小时价(元/时)
规格 1	简约型I (slb.s1.small)	5000	3000	1000	免费	免费	免费
规格 2	标准型I (slb.s2.small)	50000	5000	5000	190.00	1,938.00	0.32
规格 3	标准型II (slb.s2.medium)	100000	10000	10000	380.00	3,876.00	0.63
规格 4	高阶型I (slb.s3.small)	200000	20000	20000	760.00	7,752.00	1.27
规格 5	高阶型II (slb.s3.medium)	500000	50000	30000	1,143.00	11,658.60	1.91
规格 6	超强型I (slb.s3.large)	1000000	100000	50000	1,908.00	19,461.60	3.18

如何处理原有的性能共享型实例？

原有的性能共享型实例将继续保持性能共享型实例，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，后续在保障型实例正式收费时，该实例将收取规格费。



为何有时性能保障型实例看起来达不到规格中的性能指标上限？

短木板原理。

性能保障型实例并不保障三个指标（包含带宽指标）同时达到指定规格的指标上限。即规格中那个指标先达到峰值，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，也可能会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。

注意：如果您使用了高防产品导致您的流量源IP变的非常少，有可能产生HTTPS业务流量负载不均衡的情况，此类情况下建议工单联系售后处理。

为什么有时候低配的性能保障型实例比共享型实例的性能还差一些？

性能共享型实例在资源上是共享的，在业务高峰期性能指标没有保障；而性能保障型实例可以保障任何时刻的性能。

何时开放API来创建、变配性能保障型实例？

性能保障型实例暂时还未开放API来进行创建和变配。开放时间请关注后续官网公告、邮件及短信通知。

还可以购买性能共享型实例吗？

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下线计划会通过官网公告、邮件等方式通知。

性能保障型实例何时开始收费？

当前性能保障型实例处于公测期，性能规格免费。后续收费计划将通过官网公告、短信、邮件方式进行通知。

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ECS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。当您想为其中的一个域名配置会话保持时，比如vip.a.com，则可以使用**重写Cookie**的方式进行配置，在**Cookie名称**中输入name，然后在应用服务器上为域名vip.a.com设置key为name的Cookie。



收起高级配置

获取真实IP： 已开启(默认开启)

会话保持： 已开启
HTTP 协议会话保持基于cookie

Cookie处理方式： 重写Cookie

Cookie名称：

\$不能作为起始字符，不可以有';','`'(空格)这三种字符

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

Apache

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
```

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" )
}
```

操作说明

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，参考负载均衡Open API。

为方便阅读，示例中请求URL的参数值没有进行URL编码。

前提条件

您已经创建了两个ECS实例，并授权SSH和Web端口访问权限。

操作步骤

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-dg-a01>

响应：

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

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>

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>

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=\[{"ServerId":"ECS1InstanceId"}\]](https://slb.aliyuncs.com/?Action=AddBackendServers&LoadBalancerId=LoadBalancerId&BackendServers=[{)

响应：

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

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
      }
    ]
  }
}
```

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。

移除ECS实例

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

登录负载均衡管理控制台。

单击负载均衡实例的ID链接，进入实例详情页。

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。

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。



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。

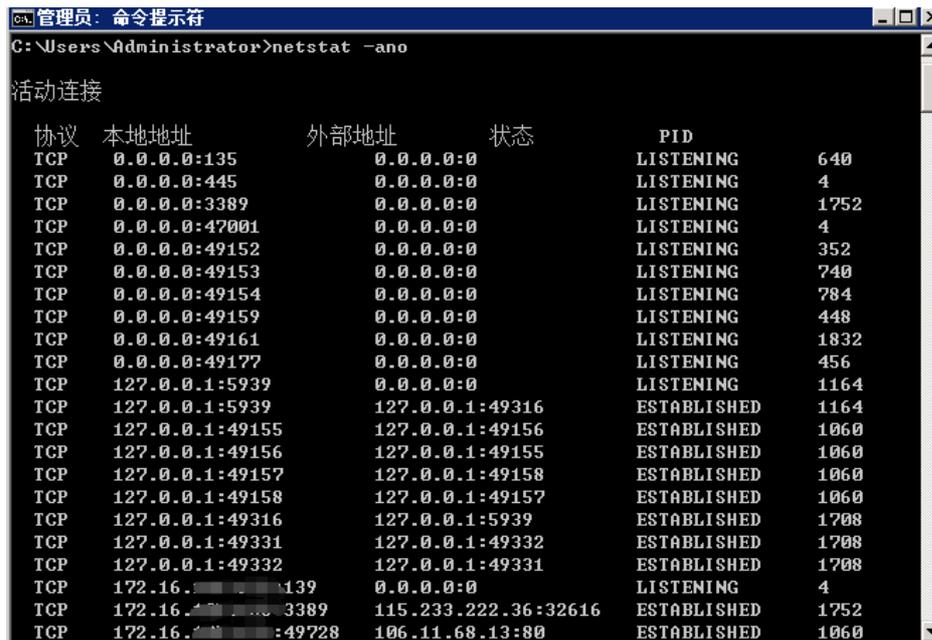
查看该ECS是否加入了其他的负载均衡实例中。

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。

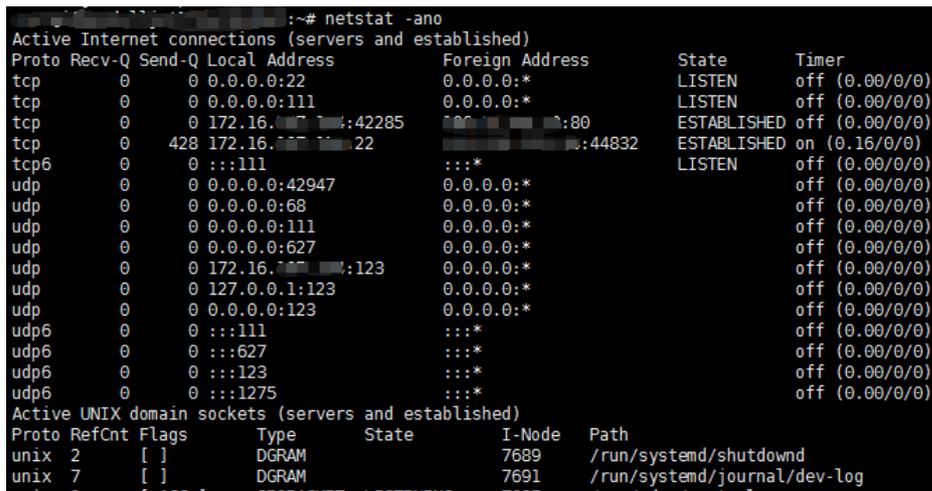


登录ECS实例，用netstat命令查看服务器端口监听情况。根据端口监听情况查看ECS本身是否对公网提供了其他服务。

Windows：运行netstat -ano，可以查看本机开放的全部端口。



Linux系统也可以使用该命令查看本机开放端口，也可以使用netstat命令的其他参数。



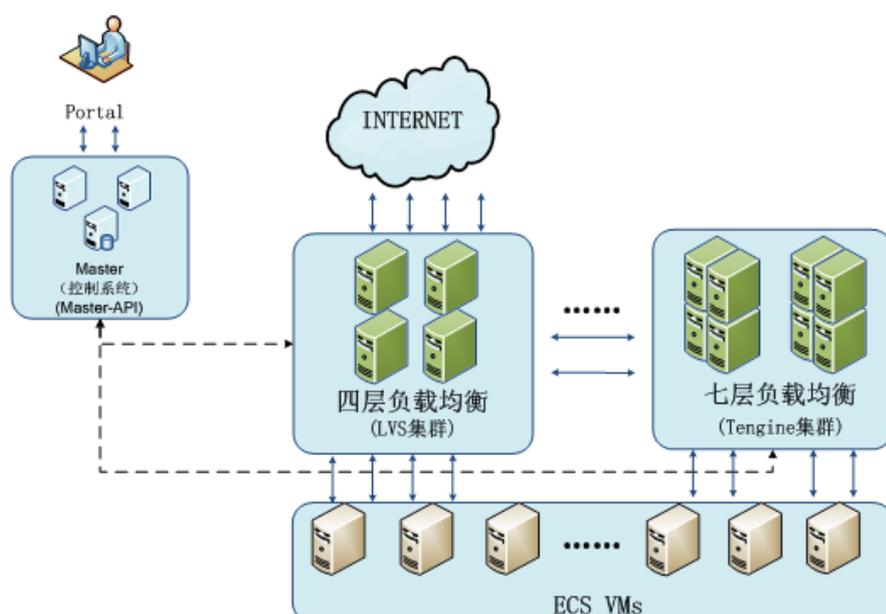
负载均衡高可用

负载均衡高可用特性介绍

集群化部署

如下图所示，负载均衡系统中四层负载均衡（LVS）服务、七层负载均衡（Tengine）服务和控制系统等关键组件都采用集群化部署，以提高扩展性和可用性。

SLB集群模块、组件图



多可用区部署

负载均衡支持多可用区部署，以实现同地域下的跨机房容灾。通过该特性可实现当整个机房故障时，负载均衡能在较短时间内，将前端访问流量切换到同一地域下的其它可用区，以恢复服务能力。

相应负载均衡实例在所对应主备可用区后端都必须部署了实例，才能保障负载均衡实例主备切换后业务不中断。

会话保持

负载均衡会话保持可以使来自同一客户端的相关访问请求，分发到后端同一台ECS上进行数据交互和处理，以保障同一会话交互的连续性。

针对7层（HTTP协议和HTTPS协议）服务，负载均衡系统是基于Cookie的会话保持。负载均衡系统提供了两种Cookie处理方式：

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。

健康检查

负载均衡通过健康检查来避免后端ECS发生故障对负载均衡整体服务造成影响。开启健康检查功能后，当后端某台ECS健康检查出现问题后，负载均衡会将新请求转发到其它健康检查正常的ECS上，而当该ECS恢复正常运行时，负载均衡会将新请求自动恢复到该ECS上，以提供负载均衡整体服务的可用性。

负载均衡后端服务器必须至少配置2台及以上ECS，才能保障单台ECS异常时，前端请求能被分发到其他正常服务器上。

HTTPS协议支持

负载均衡支持创建HTTPS协议监听，以满足用户安全加密传输数据的需求，提高业务安全性和可用性。

负载均衡提供了对提供证书的集中管理和存储，并在前端请求时自动进行加解密处理。用户无需再将证书上传到后端ECS进行手工配置。用户上传到证书管理系统的私钥都会加密存储。

负载均衡高可用配置最佳实践

建议您在负载均衡使用和配置过程中采用如下措施以提高负载均衡业务可用性：

将负载均衡后端服务器部署在同一地域的不同可用区内，然后结合负载均衡多可用区特性实现本地容灾。

可用区访问会增加访问延迟，需要根据业务敏感性进行评估。

开启负载均衡后端ECS的健康检查，以便实现负载均衡对异常ECS自动进行隔离或恢复。

负载均衡后端最少配置两台及以上ECS。这样才能在其中某一台出现异常的时候，将前端访问调度到其它正常ECS，否则单台ECS故障，会导致负载均衡业务整体不可用。

在业务层面配置自动重联机制，从而在访问请求落到异常的负载均衡后端ECS时，客户端能够自动重联，以便负载均衡将新的请求调度到正常ECS，进而提高业务可用性。

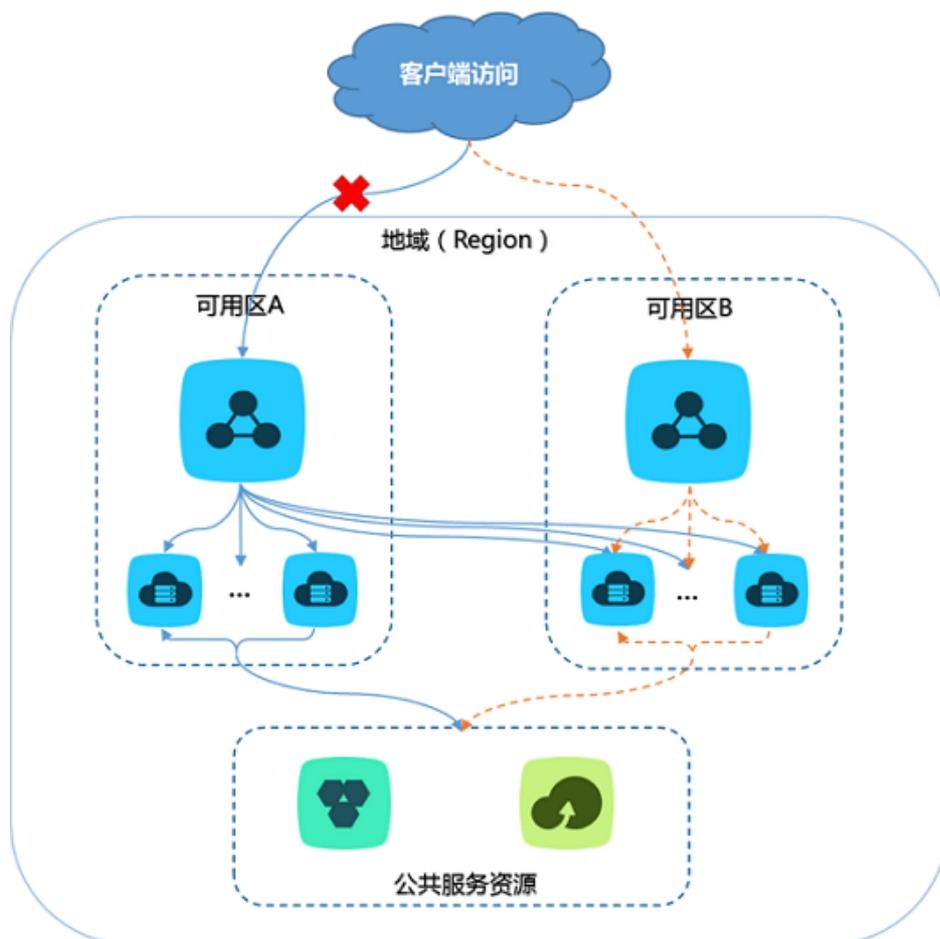
业务层面超时时间必须在负载均衡当前各协议服务的超时时间之内。

- TCP监听：900秒
- UDP监听：300秒
- HTTP监听：60秒
- HTTPS监听：60秒

负载均衡高可用常见架构

1. 本地容灾

如下图所示，负载均衡可以结合后端ECS及RDS、OSS等底层公共服务资源以实现业务的本地容灾。



方案说明：

负载均衡实例后端ECS同时分布在A、B可用区。

创建负载均衡实例的时候选择多可用区，设置可用区A为主可用区，可用区B为备可用区。

负载均衡后端ECS均为无状态ECS，业务数据存储在后端RDS、OSS等公共服务资源内。

正常情况下，客户端请求被调度到主负载均衡，进而分发到后端A、B可用区内的所有ECS上。

系统侦测到可用区A因故障无法访问时，自动将客户端请求调度到可用区B上的负载均衡备份实例上。

可用区B内的负载均衡和ECS继续为前端提供服务。

方案优点：

- 部署简单。
- 低成本。
- 同城容灾自动侦测、自动切换。
- 容灾切换业务无感知。

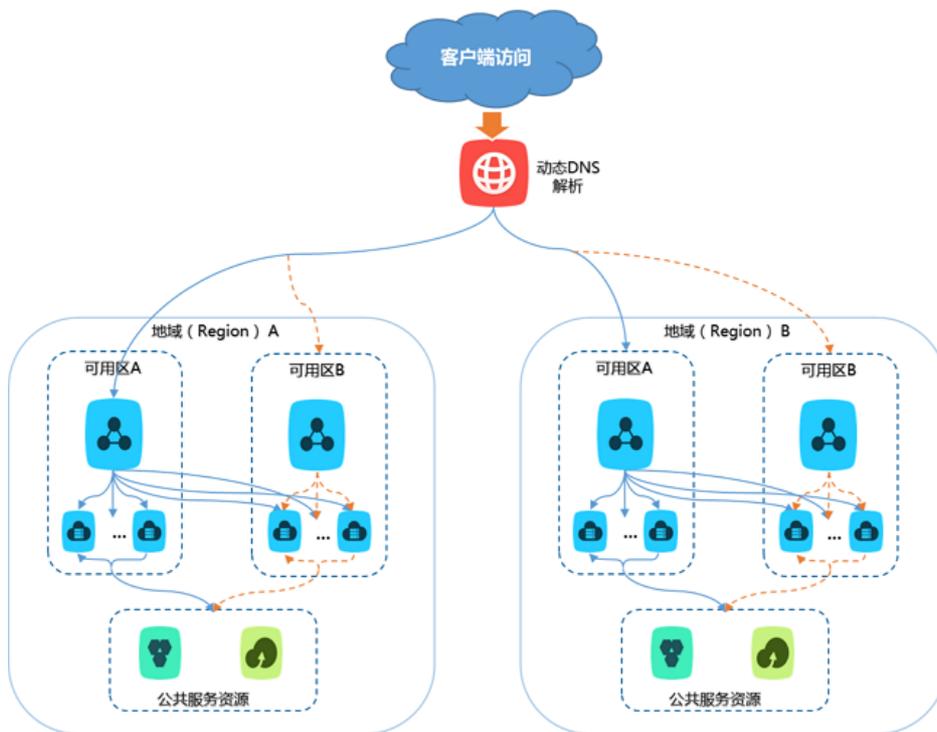
方案缺点：

无法解决异地容灾问题。

2. 异地容灾

负载均衡自身目前不支持异地容灾，需要结合外部智能DNS解析等手段来实现异地容灾。

如下图所示，您需要在不同地域购买负载均衡实例（本身已配置本地容灾）后，将最终业务访问域名配置智能DNS解析。最终根据客户端地理位置、链路质量等因素将访问请求动态解析到不同区域的负载均衡实例上。具体配置，参考结合云解析实现跨地域负载均衡。



方案优点：

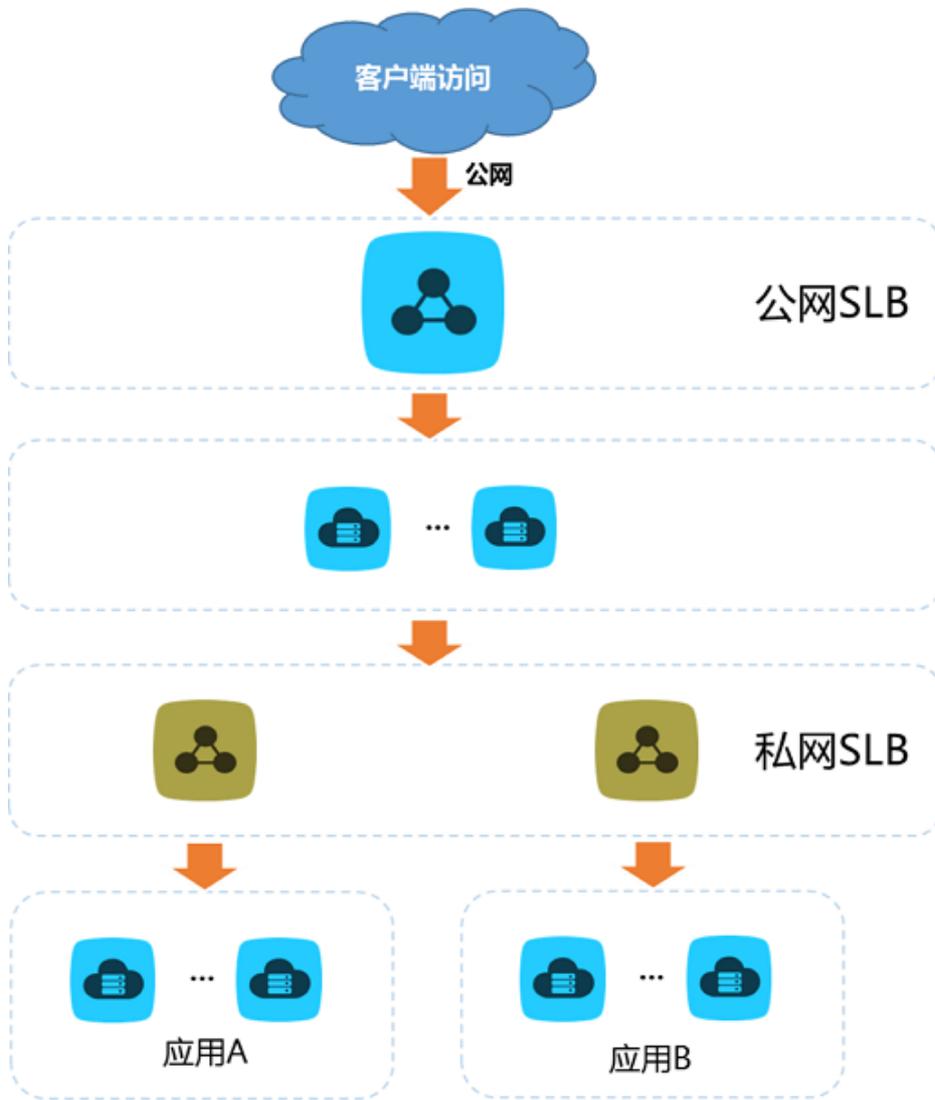
- 异地容灾，实现更高业务可用性。
- 容灾切换业务无感知。

方案缺点：

- 更高的成本。
- 需要结合外部应用联动。
- 需要了解业务分布、访问源分布等业务属性后配置定制策略。

3. 多级负载

如果有部分业务只在内部访问而同时又需要避免单点（比如数据库访问代理等），则可以部署多级负载均衡。



方案说明：

同时创建多个公网负载均衡实例和私网负载均衡实例。

公网负载均衡实例面向最终用户提供服务，私网负载均衡实例为后端服务提供高可用。

公网负载均衡实例后端ECS上，将后端服务的调用指向私网负载均衡实例。

方案优点：

- 低成本。
- 实现简单，业务逻辑清晰。
- 更高的安全性：私网负载均衡对公网无法访问，保护内部敏感应用。同时，还可在私网实例上设置访问控制白名单，进一步提高安全性。

方案缺点：

多级转发会对性能有影响。

全球负载均衡

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

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

多线路智能解析服务

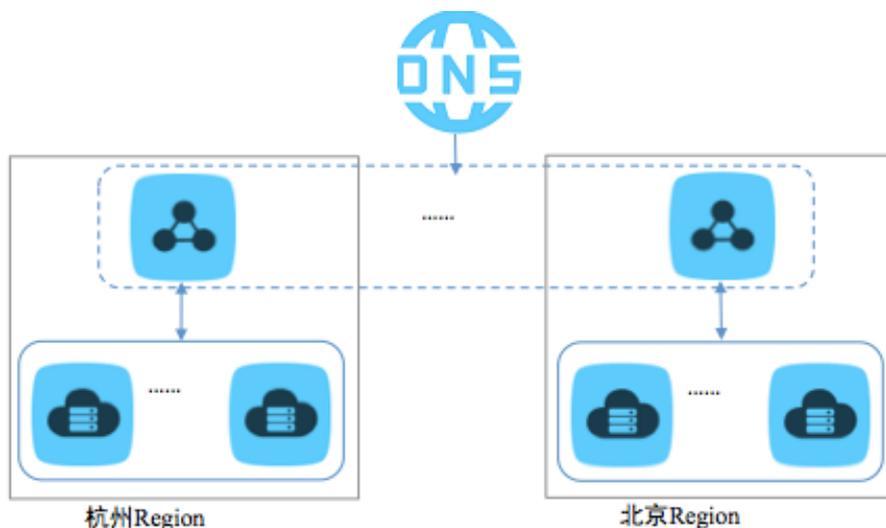
云解析（全球负载均衡版）支持多线智能解析，即根据网站访问者所在的网络类型，智能判断提供最佳的访问解析地址，使访问用户获得最快捷、最流畅的体验。

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



跨地域容灾

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



部署全球负载均衡

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

步骤一 购买与配置云服务器

根据您的应用服务的用户的地域分布，在相应地域下购买并配置至少两台ECS。

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

北京地域ECS示例

实例ID/名称	监控	所在可用区	IP地址	状态(全部)	网络类型(全部)	配置	专有网络属性	付费方式(全部)	操作
i-Zee1h8 TEST_bj2		华北 2 可用区 A	192.168.1.1 (公有)	运行中	专有网络	CPU: 1核 内存: 1 GB (1.00核/4) 5Mbps (峰值)	vpc-Zzeu vsw-Zze6	按量	17-09-12 17:20 创建 管理 远程连接 更多
i-Zee86ju TEST_bj1		华北 2 可用区 A	192.168.1.2 (私有)	运行中	专有网络	CPU: 1核 内存: 1 GB (1.00核/4) 5Mbps (峰值)	vpc-Zzeu vsw-Zze6	按量	17-09-12 17:16 创建 管理 远程连接 更多

深圳地域ECS示例

实例ID/名称	监控	所在可用区	IP地址	状态(全部)	网络类型(全部)	配置	专有网络属性	计费方式(全部)	操作
i-w99f7i-TEST_shenzhen1		华南 1 可用区 A	120.172. (公有)	运行中	专有网络	CPU: 1核 内存: 1 GB (1/0优化) 5Mpps (峰值)	vpc-wz vsw-wz	按量	管理 远程连接 更多
i-w99f7i-TEST_shenzhen2		华南 1 可用区 A	120.172. (私有)	运行中	专有网络	CPU: 1核 内存: 1 GB (1/0优化) 5Mpps (峰值)	vpc-wz vsw-wz	按量	管理 远程连接 更多

新加坡地域ECS示例

实例ID/名称	监控	所在可用区	IP地址	状态(全部)	网络类型(全部)	配置	专有网络属性	计费方式(全部)	操作
I4-Neg Singapore1		亚太东南 1 可用区B	47.17. (公有)	运行中	专有网络	CPU: 1核 内存: 1 GB (1/0优化) 5Mpps (峰值)	vpc-t4mww vsw-t4tbbk	按量	管理 远程连接 更多
I4-Neg Singapore2		亚太东南 1 可用区A	47.17. (私有)	运行中	专有网络	CPU: 1核 内存: 1 GB (1/0优化) 5Mpps (峰值)	vpc-t4mww vsw-t4t0g	按量	管理 远程连接 更多

步骤二 购买与配置负载均衡实例

参考创建负载均衡实例分别在北京、深圳、新加坡创建三个公网负载均衡实例。

参考配置负载均衡实例添加监听并将各个地域下配置好的ECS添加到后端服务器池。

北京地域负载均衡实例示例

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	标签	实例规格	带宽计费方式(全部)	计费方式(全部)	操作
lb-2ze3t49f... slb_beijing	华北 2 可用区 B(主)	47.17. (公有)	运行中	经典网络	TCP: 80 正常	TEST_lb1 TEST_lb2	未绑定	性能保障型 slb.s1.small	按使用流量计费	按量付费	管理 更多

深圳地域负载均衡实例示例

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	标签	实例规格	带宽计费方式(全部)	计费方式(全部)	操作
lb-w99f7i... slb_shenzhen	华南 1 可用区 A(主)	120.172. (公有)	运行中	经典网络	TCP: 80 正常	TEST_shenzhen2 TEST_shenzhen1	未绑定	性能共享型	按使用流量计费	按量付费	管理 更多

新加坡地域负载均衡实例示例

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	标签	实例规格	带宽计费方式(全部)	计费方式(全部)	操作
lb-t4mww... slb_singapor	亚太东南 1 可用区A(主)	47.17. (公有)	运行中	经典网络	TCP: 80 正常	Singapore2 Singapore1	未绑定	性能共享型	按使用流量计费	按量付费	管理 更多

步骤三 配置全球负载均衡

购买云解析全球负载均衡版。

在全球负载均衡页面，单击**立即购买**。

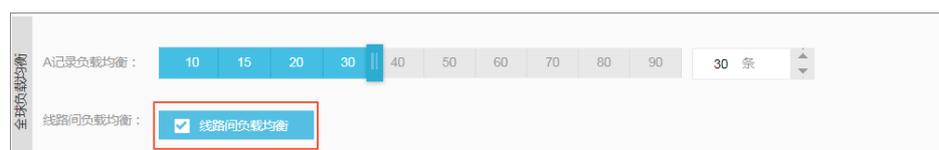
配置云解析全球负载均衡。配置详情参考云解析全球负载均衡。

本操作中，除以下配置外，其它配置使用默认选项：

配置类型，选择**全球负载均衡版**。



全球负载均衡类型，选择**线路间负载均衡**。



绑定您的网站域名。您也可以在创建后绑定域名。



单击**立即购买**。

在云解析DNS管理控制台的**VIP产品**页面，查看创建的全球负载均衡版云解析实例。



设置智能解析。

登录域名服务管理控制台。

找到云解析实例绑定的域名，然后单击**解析**。



在**解析设置**页面，单击**进入高级设置**。

添加三条A记录分别指向创建的负载均衡实例的公网IP。

将新加坡地域的A记录解析线路设置为**海外**。



添加网站监控。

在**解析设置**页面，单击**网站监控**。

单击**添加监控**，将三条A记录的IP地址添加到监控列表中。



单击每条监控记录对应的**设置**选项，修改监控配置。

将域名记录无法访问时的切换规则设置为**强制暂停该记录解析**。



开启全球负载均衡。

在**解析设置**页面，单击**全球负载均衡**。

单击**线路间负载均衡**页签，选择目标域名，然后单击**开启**。



在弹出的对话框中，单击**确定**。

步骤四 测试

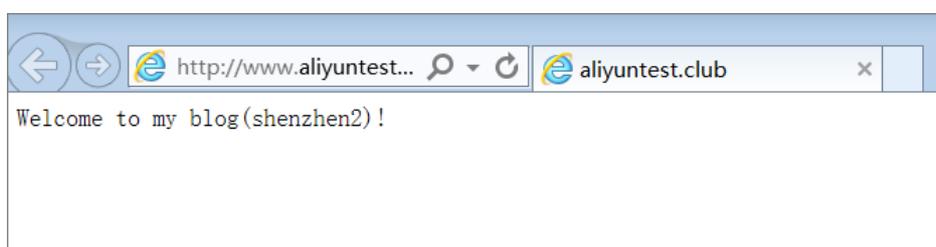
移除北京地域负载均衡实例的后端服务器，使该负载均衡实例的服务不可用。



网站域名	监控对象	监控状态	操作
www.aliyuntest.club	120.███.39	异常	停止 删除 详情 设置
www.aliyuntest.club	47.███.63	监控中	停止 删除 详情 设置
www.aliyuntest.club	47.███.19	监控中	停止 删除 详情 设置

访问该网站，查看访问是否正常。

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



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。

注意：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。

添加监听

1. 基本配置 2. 健康检查配置 3. 配置成功

前端协议 [端口] : * TCP :
端口输入范围为1-65535。
四层监听请选择TCP、UDP；七层监听请选择HTTP、HTTPS；[查看详情](#)

后端协议 [端口] : * TCP :
端口输入范围为1-65535。

带宽峰值 : * M 可用: 2M (已用0M,共2M)
固定带宽计费方式的实例，不同监听分配的带宽峰值总和不能超出在创建负载均衡实例时设定的带宽总值

调度算法 : 加权轮询

使用服务器组 : [什么是服务器组？](#)

创建完毕自动启动监听: 已开启

收起高级配置

获取真实IP : 已开启(默认开启)

会话保持: 关闭
TCP协议会话保持基于IP地址,将同一IP地址的请求转发到同一台后端云服务器处理

下一步 取消

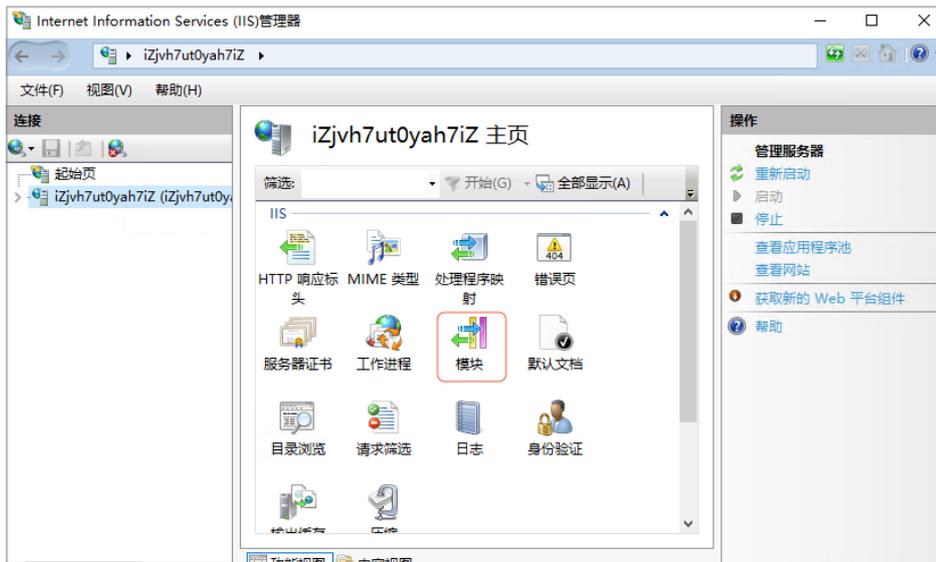
常见Web应用服务器配置方法

配置IIS7/IIS8

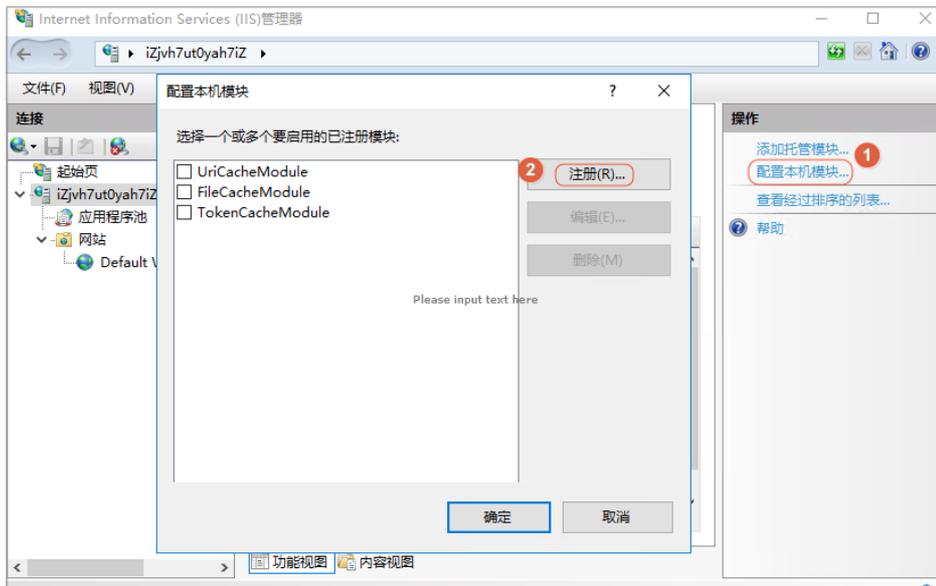
下载并解压 F5XForwardedFor文件。

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\。确保对IIS进程对该目录有读取权限。

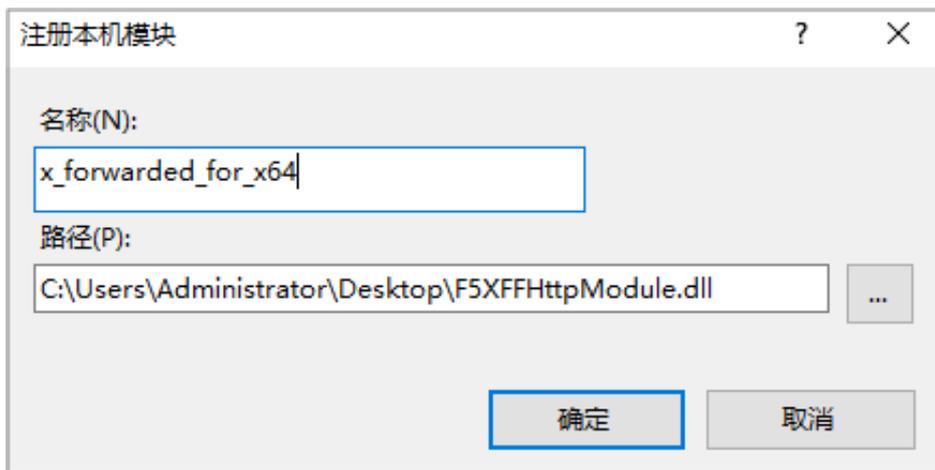
打开**IIS管理器**，双击**模块功能**。



单击配置本机模块，然后在弹出的对话框中，单击注册。

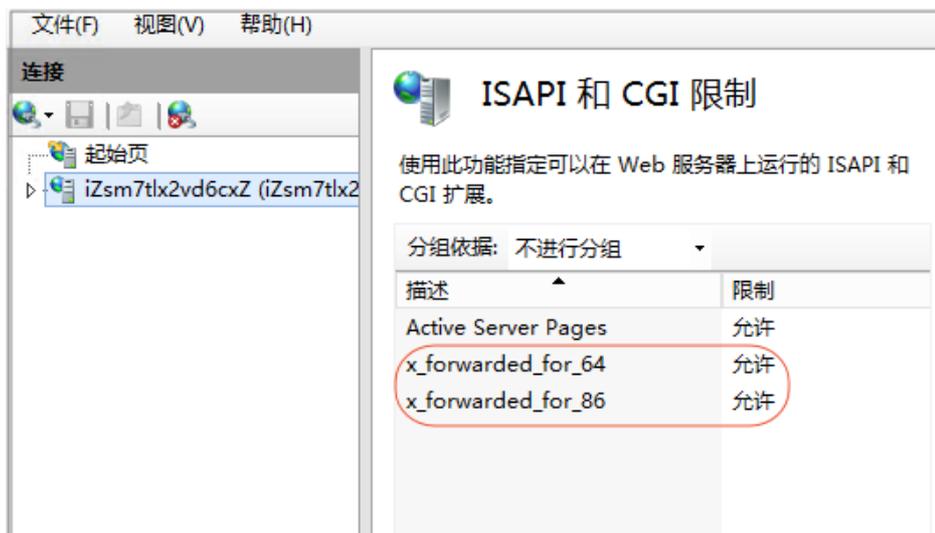


添加下载的.dll文件。



为添加的两个文件授权允许运行ISAPI和CGI扩展。

确保您已经安装了ISAPI和CGI应用程序。



重启IIS服务器，等待配置生效。

配置Apache

运行以下命令安装Apache的一个第三方模块mod_rpaf。

```
wget http://nginx.org/download/nginx-1.0.12.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
```

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 rpafl_module modules/mod_rpafl-2.0.so
RPAFenable On
RPAFsethostname On
RPAFproxy_ips IP_address
RPAFheader X-Forwarded-For
```

RPAFproxy_ips：此IP地址不是负载均衡提供的公网IP，具体IP请检查Apache日志，通常两个IP地址都要写上。

添加完成后重启Apache。

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

配置Nginx

运行以下命令安装http_realip_module。

```
wget http://soft.phpwind.me/top/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`
```

打开nginx.conf文件。

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

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;
```

set_real_ip_from IP：此IP地址不是负载均衡提供的公网IP，具体IP请检查Nginx日志，通常两个IP地址都要写上。

重启Nginx。

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

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

实例ID/名称	监控	所在可用区	IP地址	状态(全部)	网络类型(全部)
i-bp1vp9axky8i SLB_jerry1		华东 1 可用区 E	120 (弹性) 172 (私有)	运行中	专有网络
i-bp1vp9axky8i SLB_jerry2		华东 1 可用区 E	172 (私有)	运行中	专有网络
i-bp1vp9axky8i SLB_tom1		华东 1 可用区 E	116 (弹性) 172 (私有)	运行中	专有网络
i-bp1vp9axky8i SLB_tom2		华东 1 可用区 E	121 (弹性) 172 (私有)	运行中	专有网络

操作步骤

创建一个公网负载均衡实例。

详情参考创建负载均衡实例。

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。

创建两个虚拟服务器组。

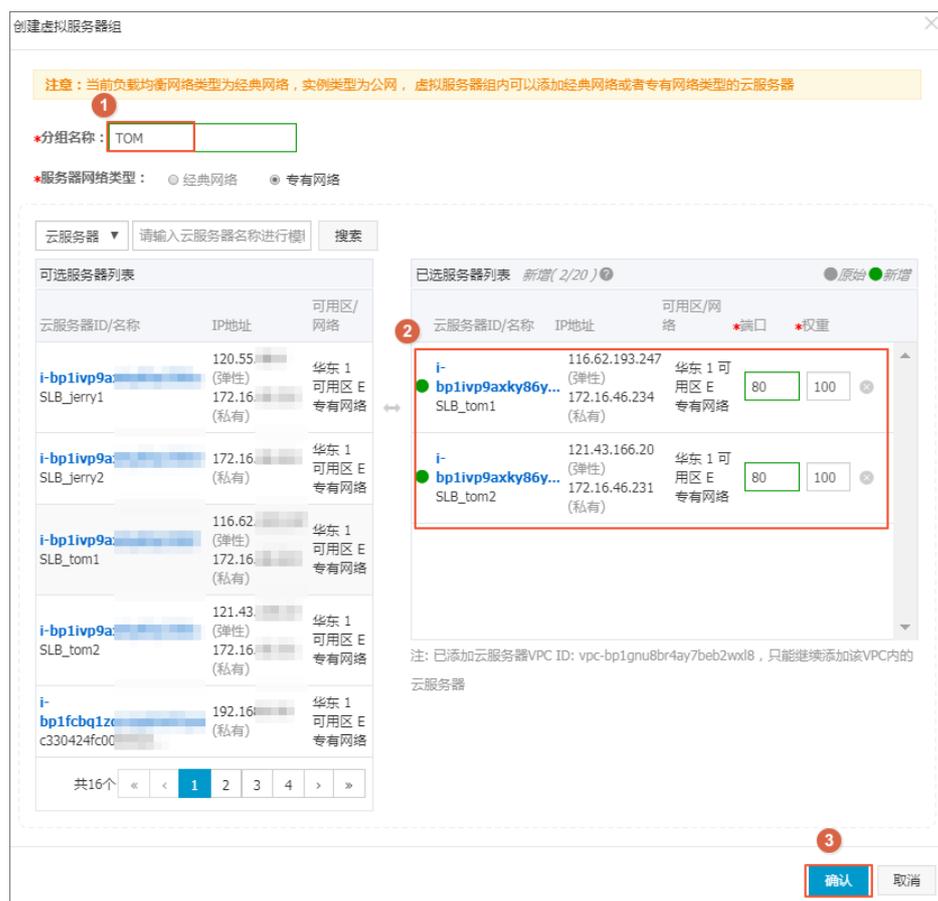
在负载均衡管理控制台，找到新建的目标实例，然后单击实例ID进入详情页面。

在左侧导航栏，单击**服务器** > **虚拟服务器组**。

单击**创建虚拟服务器**。

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）。



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。

添加监听。

在左侧导航栏，单击**监听**，然后单击**添加监听**。

配置监听。本示例中的监听配置如下：

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

在监听页面，单击**更多** > **添加转发策略**。



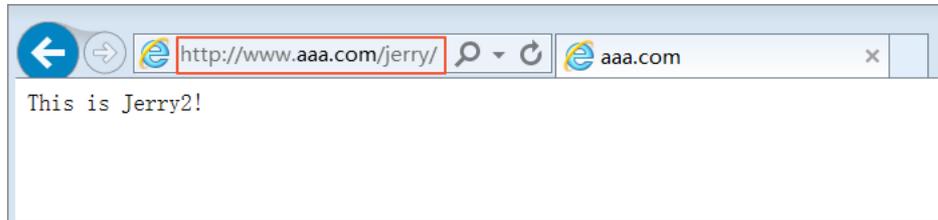
在转发策略页面，单击**添加转发策略**。

配置三条转发规则。



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。

