

负载均衡

最佳实践

最佳实践

如何使用负载均衡性能保障型实例？

1. 什么是负载均衡性能保障型实例？
2. 性能保障型实例如何收费？
3. 性能保障型实例规格费的定价
4. 如何选择性能保障型实例？
5. 是否可以调整性能保障型实例的规格？
6. 性能保障型实例何时收费？
7. 收费后性能共享型实例会收费吗？
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？
9. 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？
10. 还可以购买性能共享型实例吗？
11. 私网负载均衡实例也会收取规格费吗？

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-bp1rhw0...	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-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

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		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 (slb.s2.small)	50000	5000	5000
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

注意：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 500,000）。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

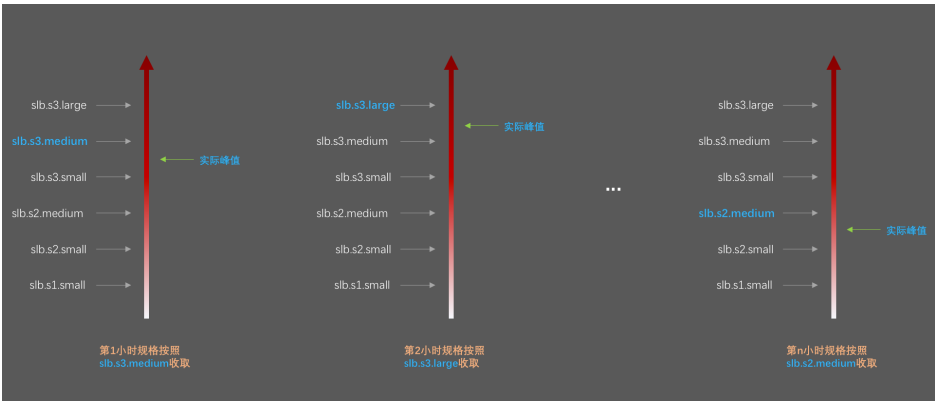
从最大连接数维度看，90000超过slb.s2.small规格中最大连接数50000的上限，但未达到slb.s2.medium规格中最大连接数的100000上限，因此从最大连接数维度计算，该小时规格为slb.s2.medium。

从每秒新建连接数（CPS）维度看，4000超过slb.s1.small规格中CPS 3000的上限，但未达到slb.s2.small规格中CPS 5000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，11000超过slb.s2.medium规格中QPS 10000的上限，但未达到slb.s3.small中QPS 20000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。

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)	小时价(元/时)
规格 1	简约型I (slb.s1.small)	5000	3000	1000	免费	免费
规格 2	标准型I (slb.s2.small)	50000	5000	5000	190.00	0.32
规格 3	标准型II (slb.s2.medium)	100000	10000	10000	380.00	0.63
规格 4	高阶型I (slb.s3.small)	200000	20000	20000	760.00	1.27
规格 5	高阶型II (slb.s3.medium)	500000	50000	30000	1,143.00	1.91
规格 6	超强型I (slb.s3.large)	1000000	100000	50000	1,908.00	3.18

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

(未设置)

lb-2z9w90bq

可用区

华北 2 可用区 B(主)

服务地址(全部)

47.96.161.100

网络类型(全部)

公网

端口/健康检查

● 运行中

后端服务器

经典网络

实例规格

未配置(配置)

带宽计费方式(全部)

性能保障型

付费方式(全部)

slb.s1.small

操作

按固定带宽

包年包月

2017-10-05

管理

更多

负载均衡ID/名称

(未设置)

lb-2z9w90bq

可用区

华北 2 可用区 B(主)

服务地址(全部)

47.96.161.100

网络类型(全部)

公网

端口/健康检查

● 运行中

后端服务器

经典网络

实例规格

未配置(配置)

带宽计费方式(全部)

性能共享型

付费方式(全部)

按使用流量

操作

按量付费

2017-09-01

启动

停止

编辑标签

负载均衡ID/名称

(未设置)

lb-2z9w90bq

可用区

华北 2 可用区 B(主)

服务地址(全部)

47.96.161.100

网络类型(全部)

公网

端口/健康检查

● 运行中

后端服务器

经典网络

实例规格

未配置(配置)

带宽计费方式(全部)

性能共享型

付费方式(全部)

按使用流量

操作

按量付费

2017-09-01

变更配置

续费

当前配置

实例名称: lb-2z9w90bq

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

配置变更

实例规格: 标准型I (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。

注意：

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，具体收费时间和计划、后续请关注官网公告、短信、邮件。

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。

配置变更

实例规格：

性能共享型实例

继续使用性能共享，该类型实例不提供性能保障
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。

实例类型：

公网

实例类型详解>> ?

计费方式：

按使用流量计费

按固定带宽计费

开通即按使用流量计费，停止或释放实例不会产生流量费用
进行变配操作时，若选择的计费方式与当前计费方式不同，则代表变更计费方式，变更计费方式将在第二天0点生效
阿里云最高提供5Gbps的恶意流量攻击防护，[了解更多>>提升防护能力>>](#)
[阿里云现已开通共享流量包，可同时抵扣 ECS、EIP、SLB、NAT 产生的流量。点击购买特惠共享流量包套餐>>](#)

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。

9. 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，如果您是后付费实例，因为规格费是按量收取的，所以建议您选择您可以选择的最大规格。

10. 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。

11. 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计费规则一致。私网实例免收实例费和流量费。

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。

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" }
}
```

使用Open API配置负载均衡

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，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。

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。

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"}]
```

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。

从负载均衡服务中移除ECS实例

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：

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。

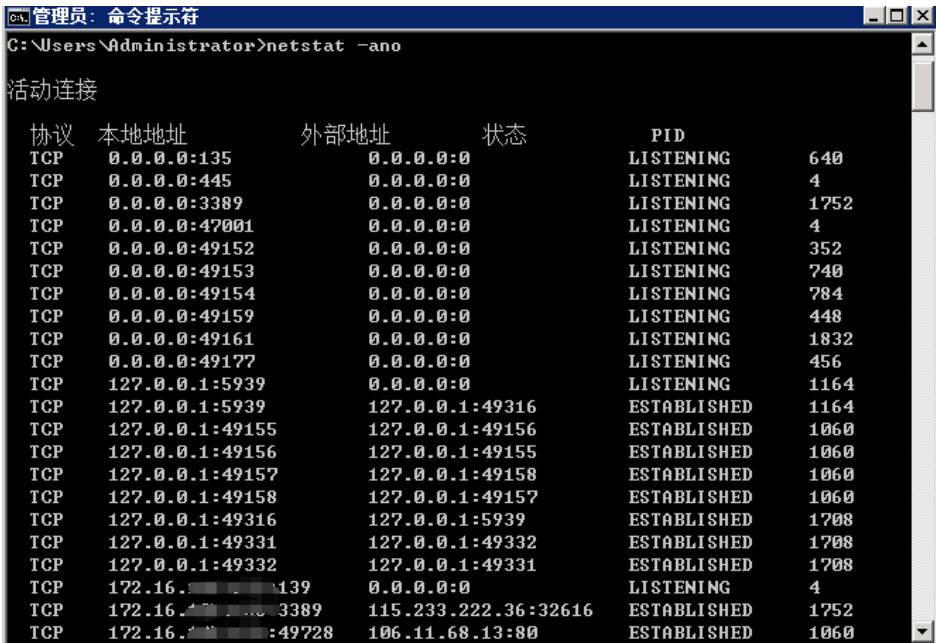
查看该ECS是否加入了其他的负载均衡实例中。

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。

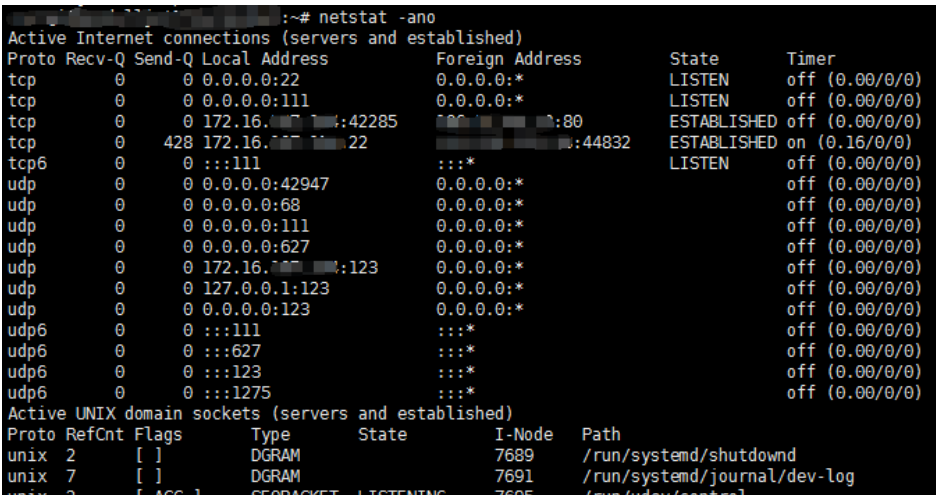


登录ECS实例，用netstat命令查看服务器端口监听情况。根据端口监听情况查看ECS本身是否对公网提供了其他服务。

Windows：运行netstat -ano，可以查看本机开放的全部端口。



Linux：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。



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

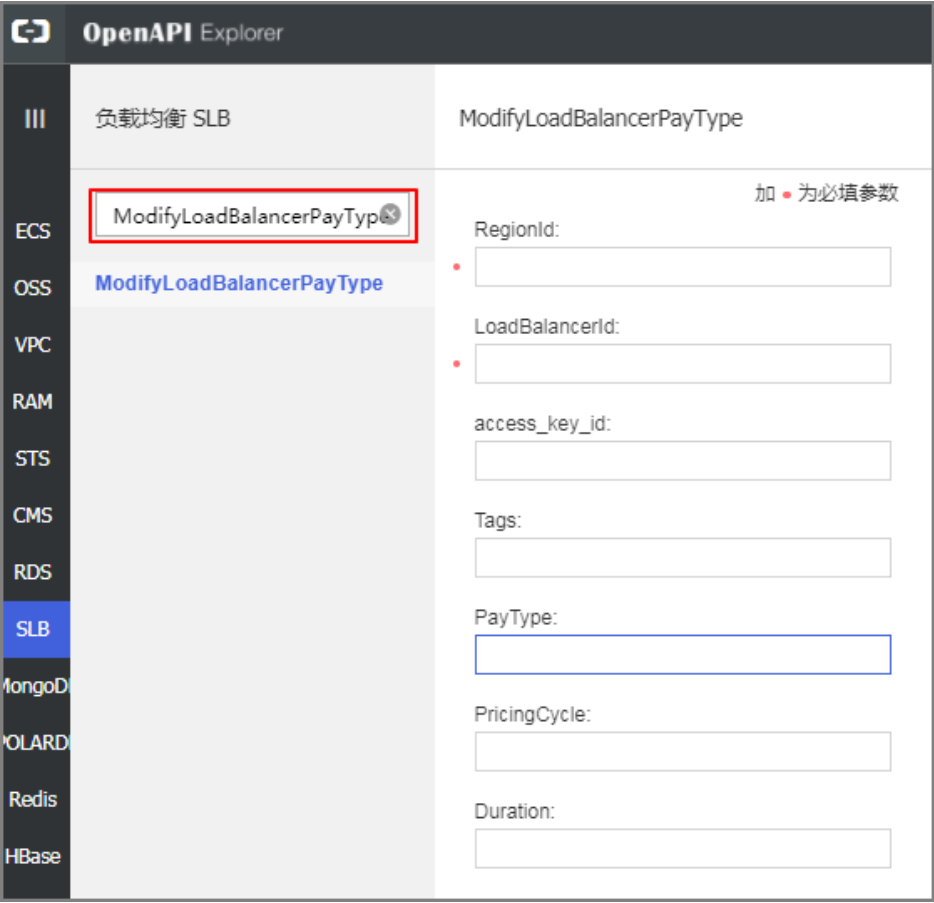
登录OpenAPI Explorer。

在左侧导航栏，单击**负载均衡 SLB**。



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。

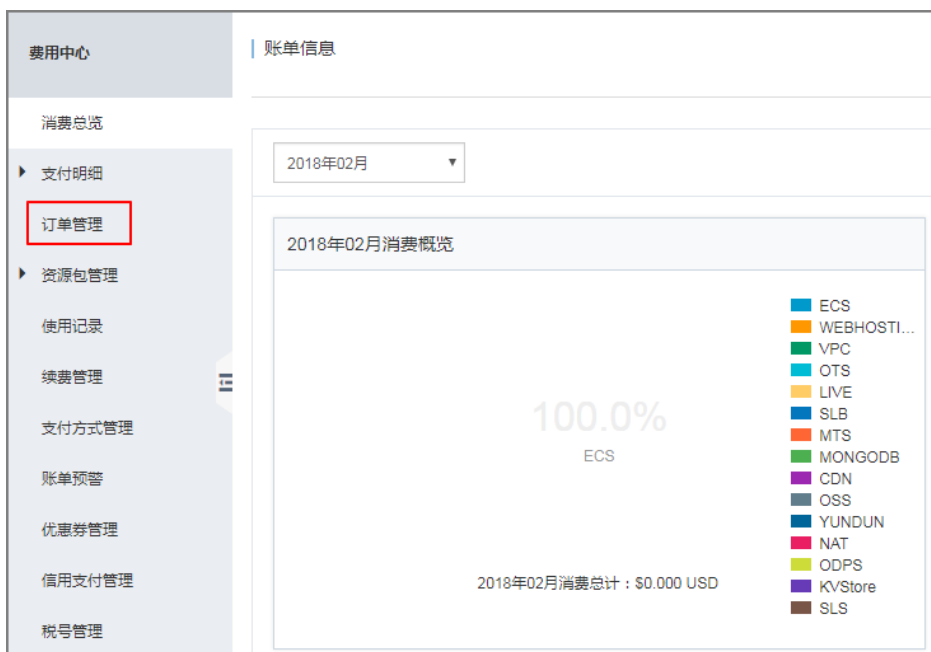
PricingCycle：选择付费周期，year为按年付费，month为按月付费。

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]。

单击**发送请求**，当请求状态为“SUCCESS”时，表示成功生成了订单。

登录费用中心。

在左侧导航栏，单击**订单管理**。



找到该负载均衡实例的未支付订单，并完成支付。

负载均衡高可用最佳实践

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，但是对于短链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，降低对方用户访问造成的影响。

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。

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。

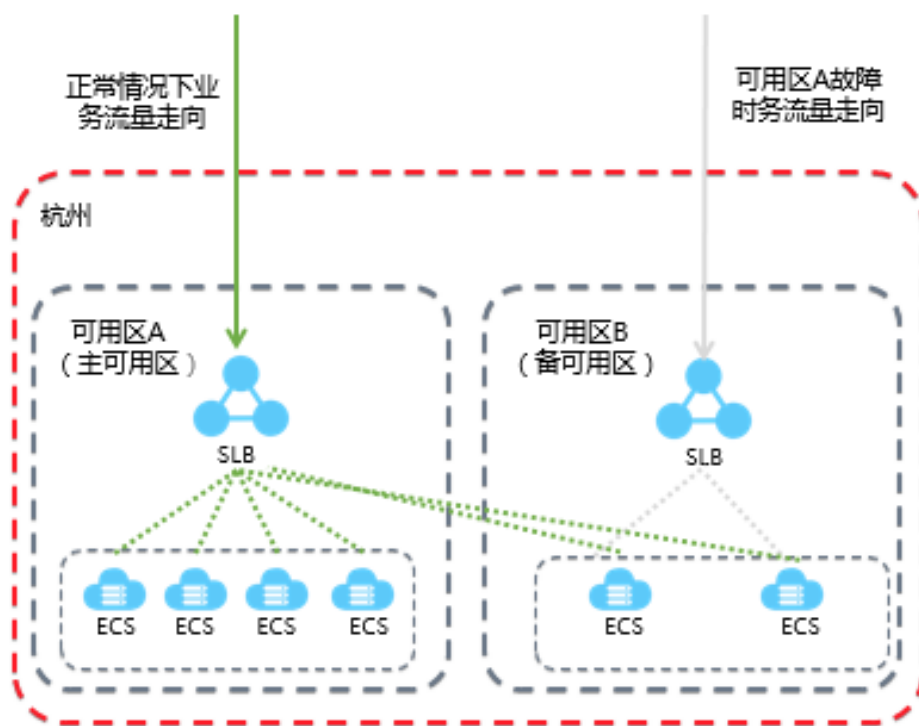
说明：负载均衡的主备可用区是可用区级别的容灾。只有当主可用区整体不可用时，如机房整体断电、机房出口光缆中断等，负载均衡才会切换到备可用区。而非某个实例出现故障，就切换到备可用区。

最佳实践：

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。

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。

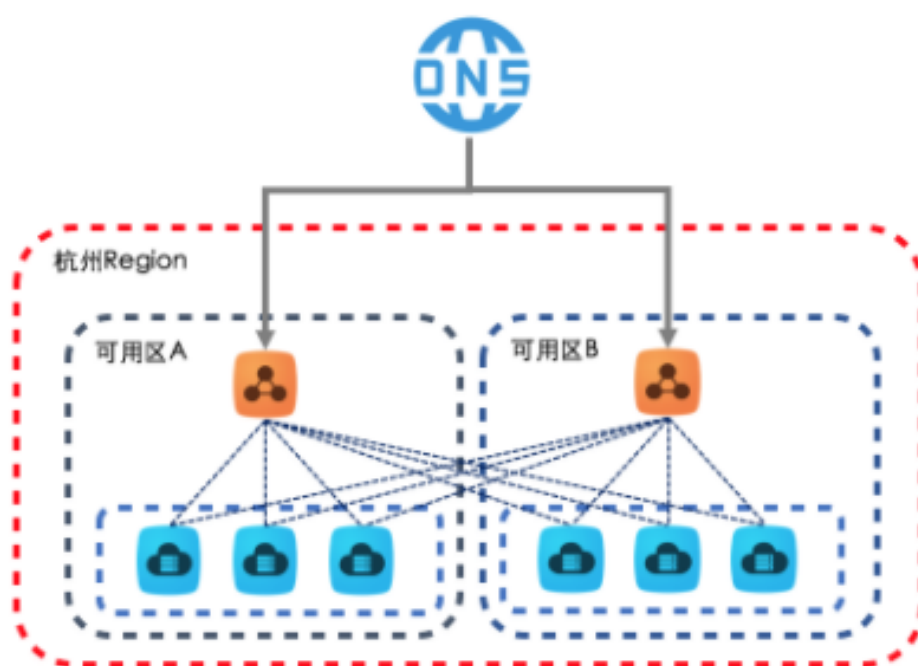


多SLB实例的高可用

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，详情参见[结合云解析实现跨地域负载均衡](#)。



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。

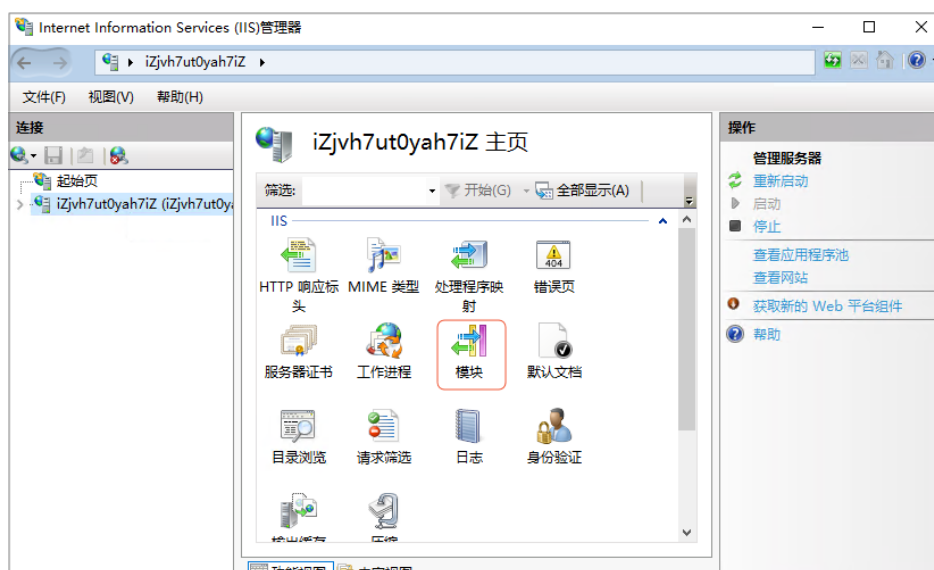
注意：负载均衡的HTTPS监听是在负载均衡服务上的加密控制，后端仍旧使用HTTP协议，因此，在Web应用服务器配置上HTTPS和HTTP监听没有区别。

配置IIS7/IIS8服务器

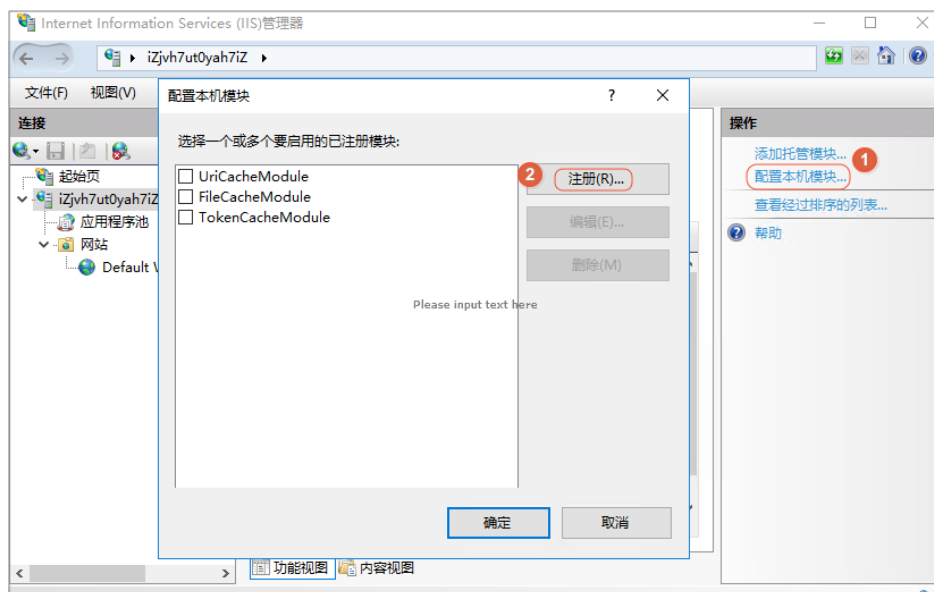
下载并解压 F5XForwardedFor文件。

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。

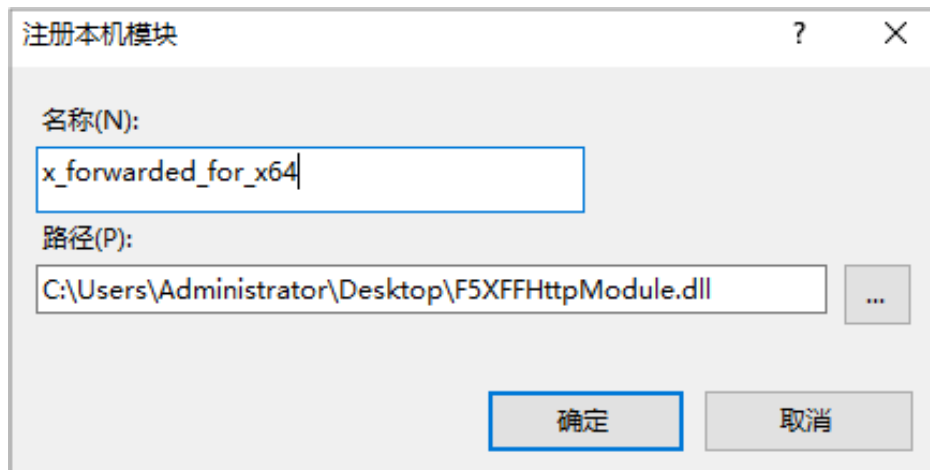
打开**IIS管理器**，双击**模块**功能。



单击**配置本机模块**，然后在弹出的对话框中，单击**注册**。

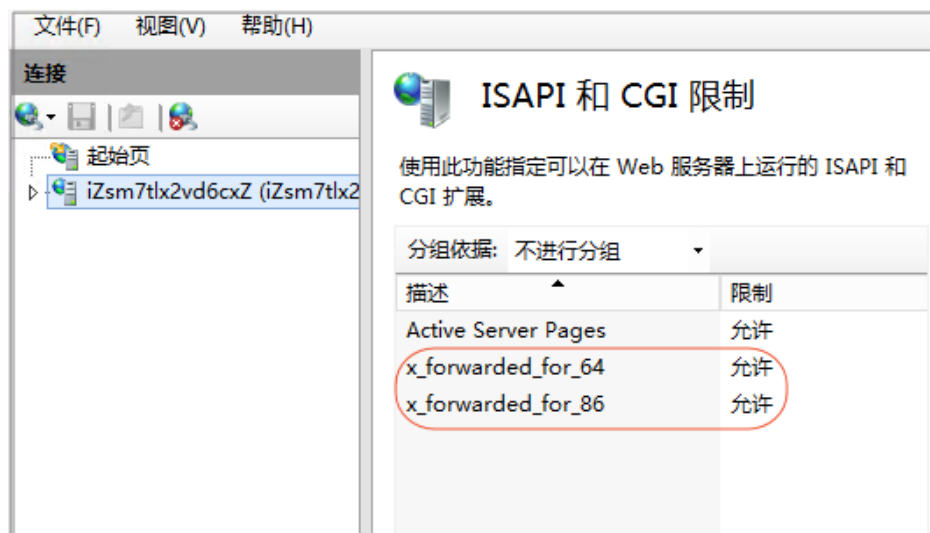


添加下载的.dll文件。



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服务器

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
```

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、10.158.0.0/16、10.159.0.0/16和10.49.0.0/16）和高防IP地址段。多个IP地址段用逗号分隔。

添加完成后重启Apache。

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

配置Nginx服务器

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`
```

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;
```

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、10.158.0.0/16、10.159.0.0/16和10.49.0.0/16）和高防IP地址段。多个IP地址段用逗号分隔。

重启Nginx。

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

如何实现相同域名不同路径的流量转发

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-bp1vp9axky8... SLB_jerry1		华东 1 可用区 E	120... (弹性) 172... (私有)	运行中	专有网络
i-bp1vp9axky8... SLB_jerry2		华东 1 可用区 E	172... (私有)	运行中	专有网络
i-bp1vp9axky8... SLB_tom1		华东 1 可用区 E	116... (弹性) 172... (私有)	运行中	专有网络
i-bp1vp9axky8... 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。

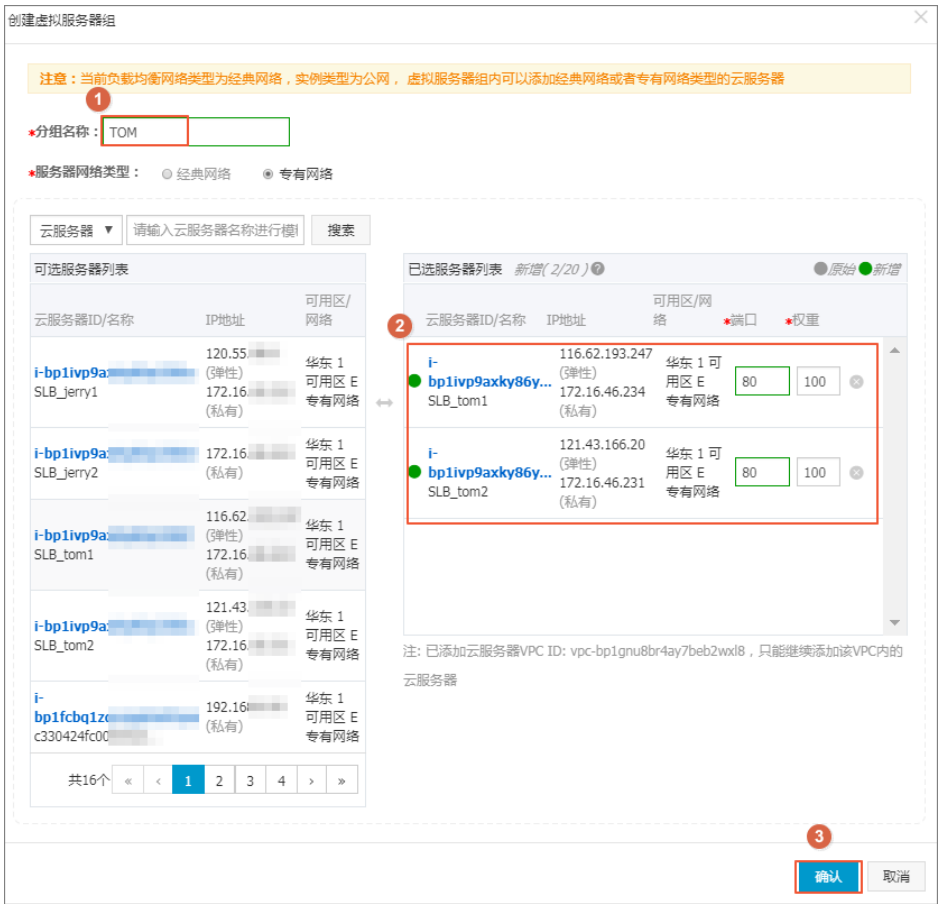
在负载均衡管理控制台，找到新建的目标实例，然后单击实例ID进入详情页面。

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。

添加转发策略

规则名称	域名	URL	虚拟服务器组	操作
rule1	www.aaa.com	/jerry	JERRY ▼	删除
rule2	www.aaa.com	/tom	TOM ▼	删除
rule3	www.aaa.com		TOM ▼	删除

添加转发策略 +

* 域名规范：

只能使用字母、数字、'-'、'.'，只支持以下两种形式的domain形式

- 标准域名：www.test.com;

- 泛解析域名：*.test.com，*一定在第一个字符，并且是*.的格式，*不能在最后。

* URL规范：

长度限制为2-80个字符，只能使用字母、数字、'-'、'/'、'.'、'%'、'?'、'#'、'&'这些字符；URL不能只为/，但必须以/开头。

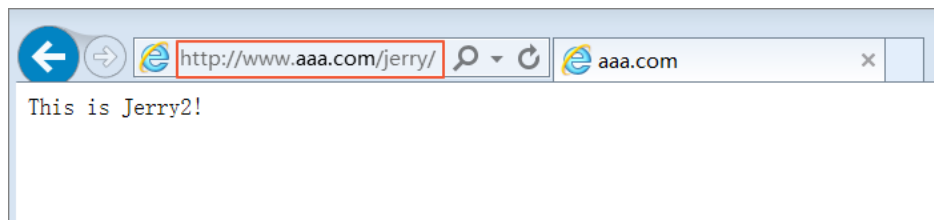
* 域名与URL请至少填写一项。

确认

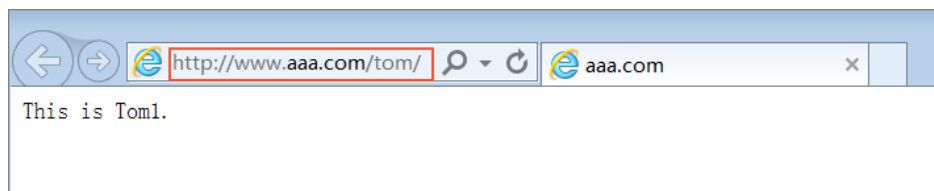
取消

测试：

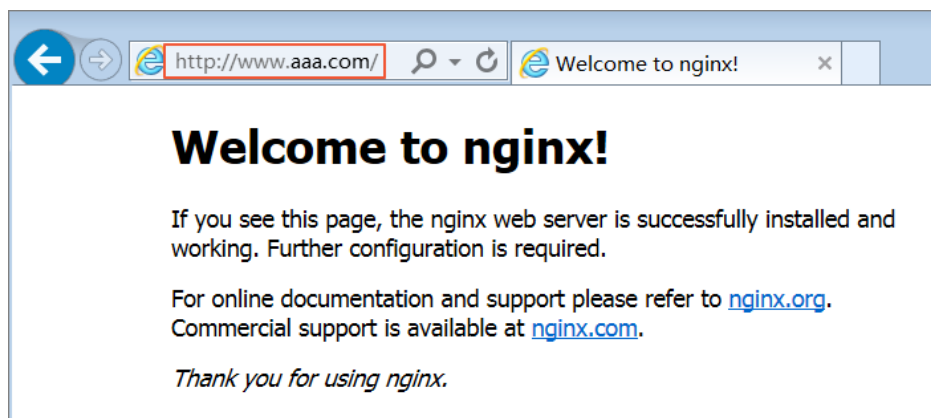
在浏览器中输入www.aaa.com/jerry，将返回如下结果。



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。



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。

结合云解析实现跨地域负载均衡

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

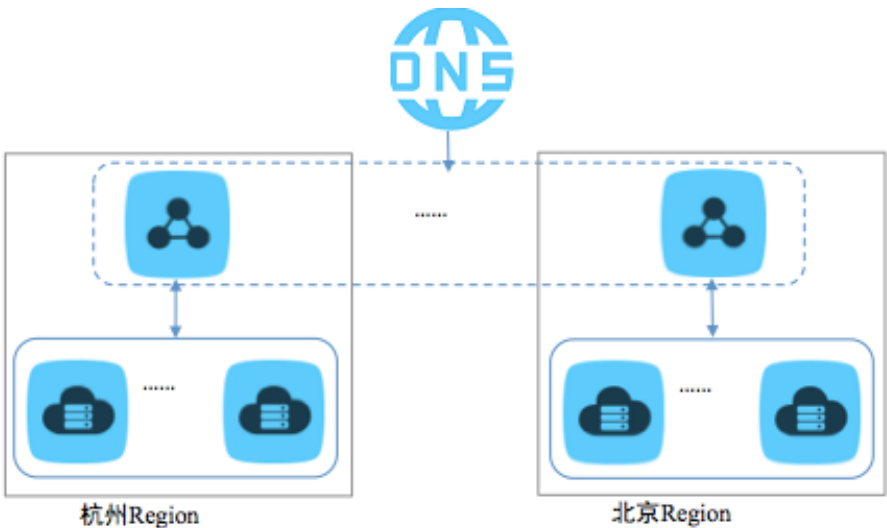
云解析（全球负载均衡版）支持多线智能解析，即根据网站访问者所在的网络类型，智能判断提供最佳的访问解析地址，使访问用户获得最快捷、最流畅的体验。

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

实例列表	华北1	华北2	华北3	华东1	华东2	华南1	香港	亚太东北1 (东京)	亚太东南1 (新加坡)	亚太东南2 (悉尼)	美国东部1 (弗吉尼亚)	创建实例
	美国西部1 (硅谷)	中东东部1 (迪拜)	欧洲中部1 (法兰克福)									
实例名称	输入实例名称模糊查询		搜索	标签	即日起，共享实例升级为独享实例，差价补贴 50%! 高级搜索							
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-2ze1h8h TEST_bj2		华北2 可用区 A	10.10.10.10 (公有) 10.10.10.11 (私有)	运行中	专有网络	CPU : 1核 内存 : 1 GB (I/O优化) 5Mbps (峰值)	vpc-2zeu vsw-2ze6	按量	17-09-12 17:20	创建	管理	远程连接 更多
i-2ze86ju TEST_bj1		华北2 可用区 A	10.10.10.12 (公有) 10.10.10.13 (私有)	运行中	专有网络	CPU : 1核 内存 : 1 GB (I/O优化) 5Mbps (峰值)	vpc-2zeu vsw-2ze6	按量	17-09-12 17:16	创建	管理	远程连接 更多

深圳地域ECS示例

实例列表	华北1	华北2	华北3	华东1	华东2	华南1	香港	亚太东北1 (东京)	亚太东南1 (新加坡)	亚太东南2 (悉尼)	美国东部1 (弗吉尼亚)	创建实例
	美国西部1 (硅谷)	中东东部1 (迪拜)	欧洲中部1 (法兰克福)									
实例名称	输入实例名称模糊查询		搜索	标签	即日起，共享实例升级为独享实例，差价补贴 50%! 高级搜索							
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-w99f7 TEST_shenzhen1		华南1 可用区 A	120.172.10.10 (公有) 120.172.10.11 (私有)	运行中	专有网络	CPU : 1核 内存 : 1 GB (I/O优化) 5Mbps (峰值)	vpc-wz vsw-wz	按量	17-09-12 18:08	创建	管理	远程连接 更多
i-w99f7 TEST_shenzhen2		华南1 可用区 A	120.172.10.12 (公有) 120.172.10.13 (私有)	运行中	专有网络	CPU : 1核 内存 : 1 GB (I/O优化) 5Mbps (峰值)	vpc-wz vsw-wz	按量	17-09-12 18:08	创建	管理	远程连接 更多

新加坡地域ECS示例



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



深圳地域负载均衡实例示例



新加坡地域负载均衡实例示例



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**。

推荐配置：

基础版

标准版

旗舰版

尊享版

全球负载均衡版

自定义

适合于多个数据中心异地容灾的网站，SLA不低于99.999%

全球负载均衡类型，选择**线路间负载均衡**。

A记录负载均衡：

10

15

20

30

40

50

60

70

80

90

30 条

线路间负载均衡：

☒ 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。

绑定域名：

选择已有域名

输入域名

暂不绑定

绑定规则>>

aliyuntest.club

此处最多能显示20个域名，如果域名不在列表中，你可以选择输入域名

此域名可用

域名更换次数：

2

3

4

5

4 次

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。

云解析 DNS	VIP产品				
我的云解析	① 提醒：云解析企业版需要绑定域名，才可正常使用！首次绑定域名，不计算在域名变更次数以内，请及时的绑定域名！				
VIP产品	产品类型	配置	产品到期日	绑定域名	操作
HTTPDNS		运营商线路解析-普通（默认） 域名备案次数-2次 海外线路-海外 DNS劫持-防劫持 DNS劫持-防劫持	2017-12-12	aliyuntest.club	续费 升级 更换域名 更多
操作记录	alidns-cn-v09081af81				

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，然后单击其**操作**列下的**解析设置**。

根据设置域名解析中的步骤设置域名解析。添加三条A记录分别指向创建的负载均衡实例的公网IP，并将新加坡地域的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。



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。

