

# 阿里云云计算ACP认证培训

阿里云负载均衡服务SLB

---

# 课程目标

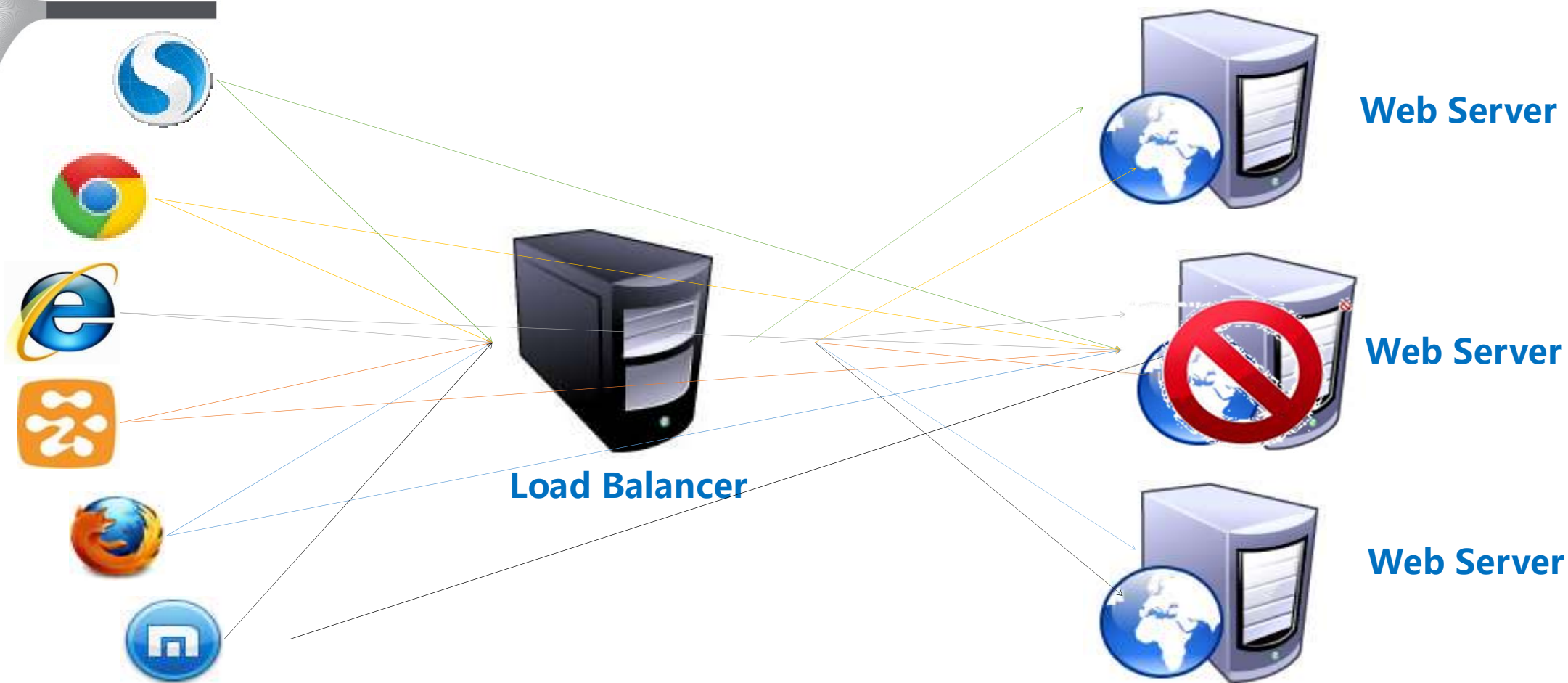
学习完本课程后，你将能够：

1. 了解负载均衡服务的产生、发展、技术原理
2. 掌握阿里云SLB的功能、组成、使用场景
3. 掌握SLB的基本操作

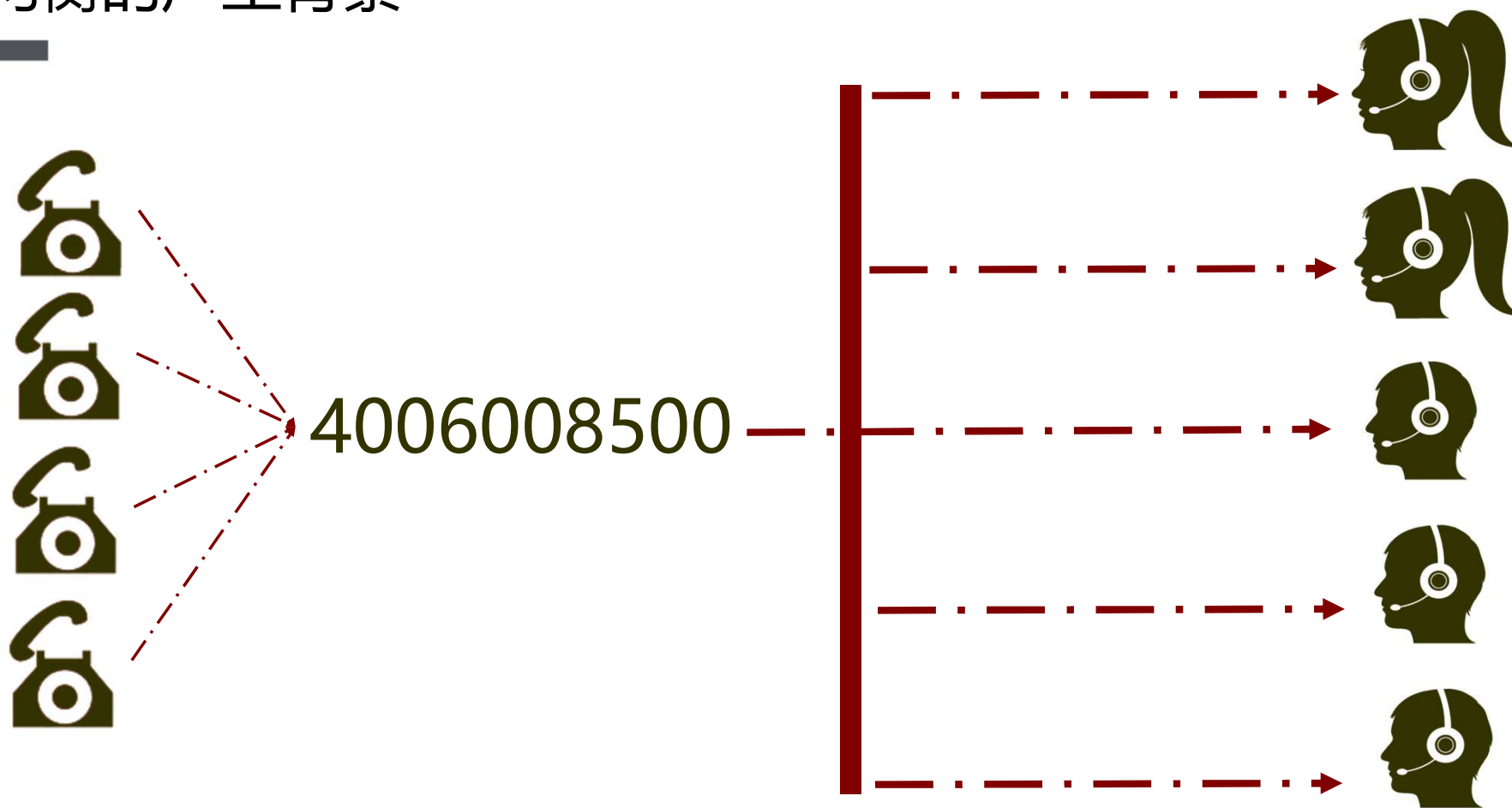
# 目录

- 1. 负载均衡的产生及发展**
2. 阿里云负载均衡服务SLB介绍
3. SLB的配置流程
4. SLB的最佳实践
5. SLB的典型应用案例

# 负载均衡的产生背景



## 负载均衡的产生背景



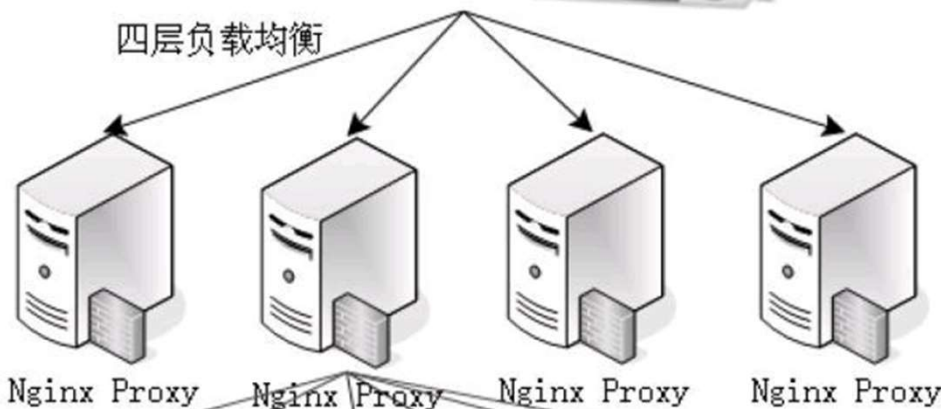
# 负载均衡的发展

硬件负载均衡 (F5)

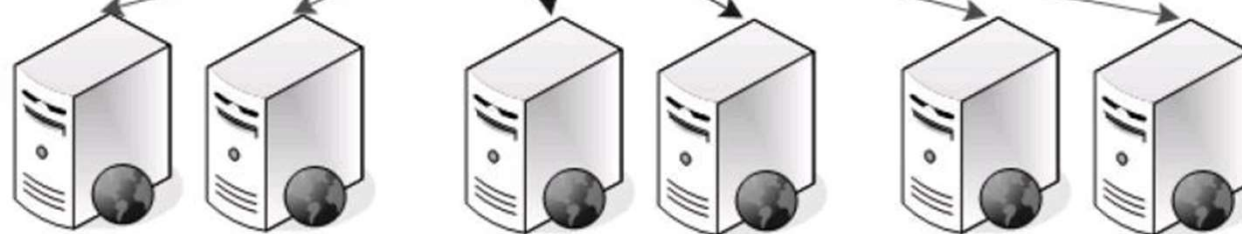


四层负载均衡

软件负载均衡 (Nginx集群)



服务器 (集群)



# 目录

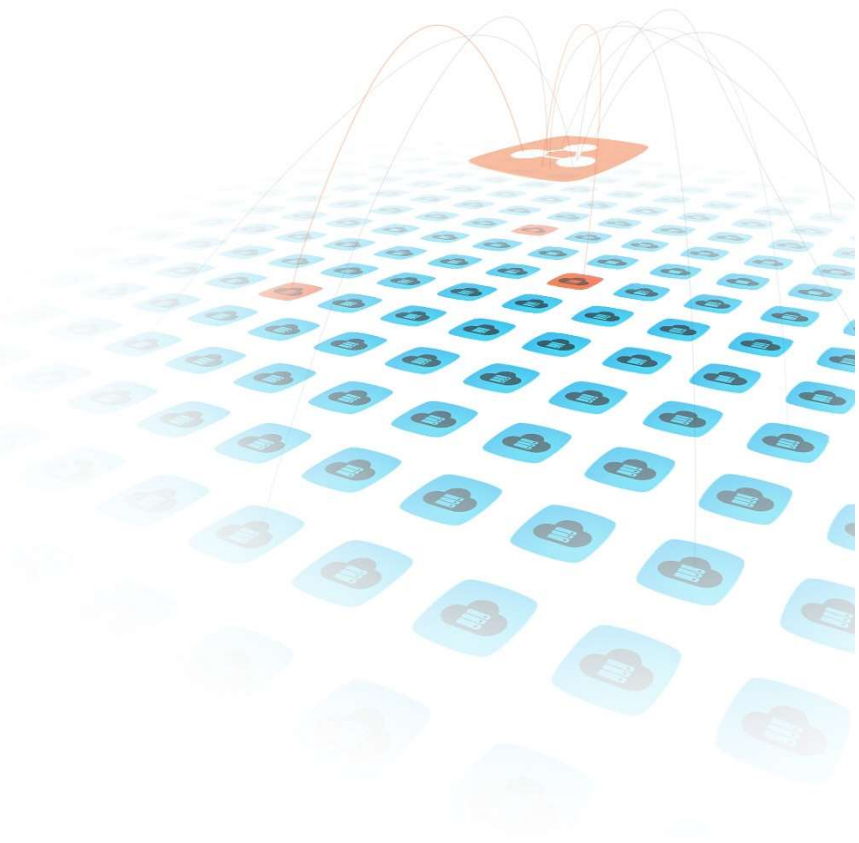
1. 负载均衡服务的产生背景
- 2. 阿里云负载均衡服务SLB介绍**
  - 2.1 SLB的定位
  - 2.2 SLB的实现原理
  - 2.3 SLB的组成
  - 2.4 SLB的功能
  - 2.5 SLB的优势
3. SLB的配置流程
4. SLB的最佳实践
5. SLB的典型应用案例

# SLB的定位

**阿里云SLB：**是阿里云结合自身弹性计算平台的特点以及强大的技术优势，提供的一套软件负载均衡解决方案，以更好的满足弹性计算平台负载均衡的需求。

## 什么情况下使用SLB？

- 简单说就是单台云服务器不能满足需求的时候
- 使用多台云服务器进行流量分发，**提升服务能力**
- 使用多台云服务器消除单点故障，**提升可用性**



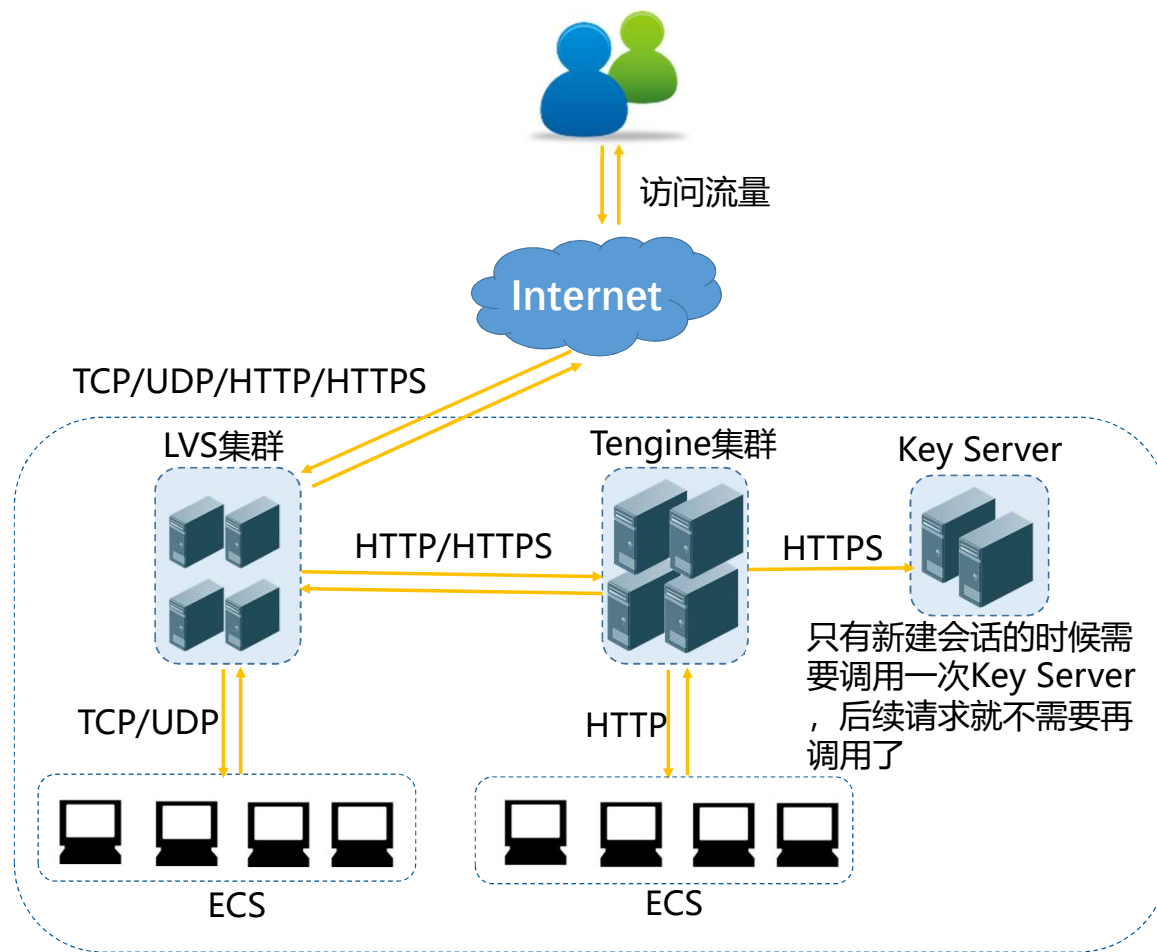


# SLB的实现原理：访问流量路径

采用集群部署，当前提供四层（TCP协议和UDP协议）和七层（HTTP和HTTPS协议）的负载均衡服务。

系统由三部分构成：四层负载均衡、七层负载均衡、控制系统。

- 四层：LVS+Keepalived
- 七层：Tengine



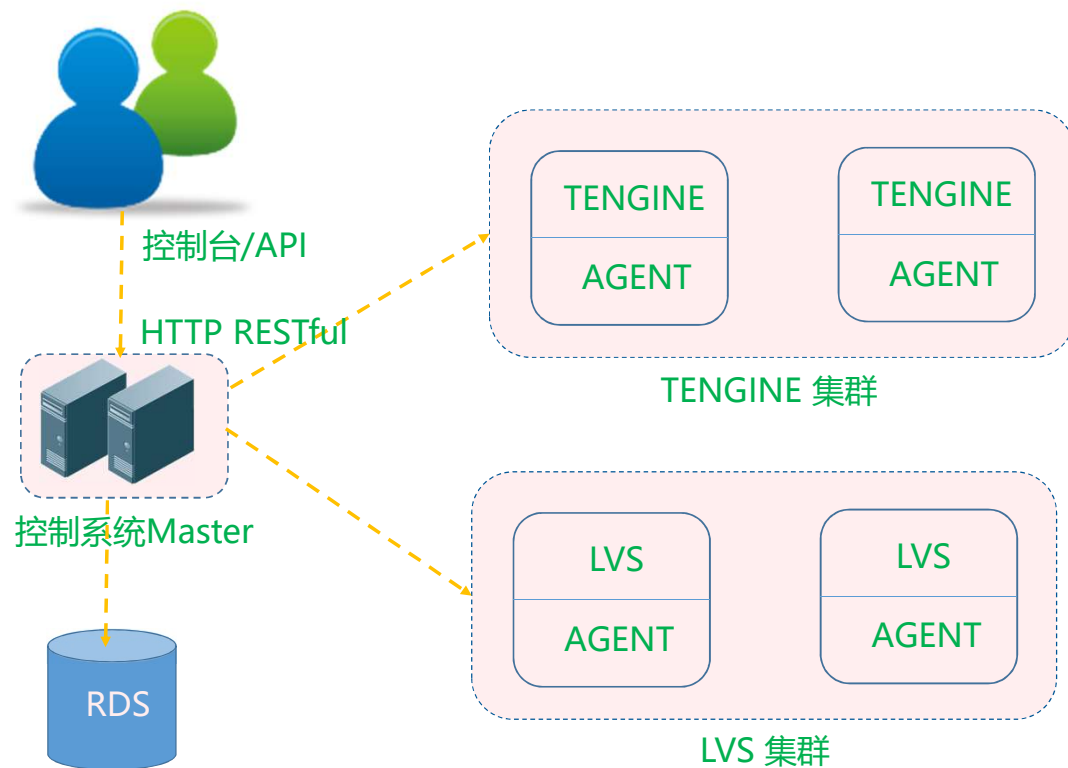
# SLB的实现原理：控制系统

控制系统：

用于监控和配置整个负载均衡系统

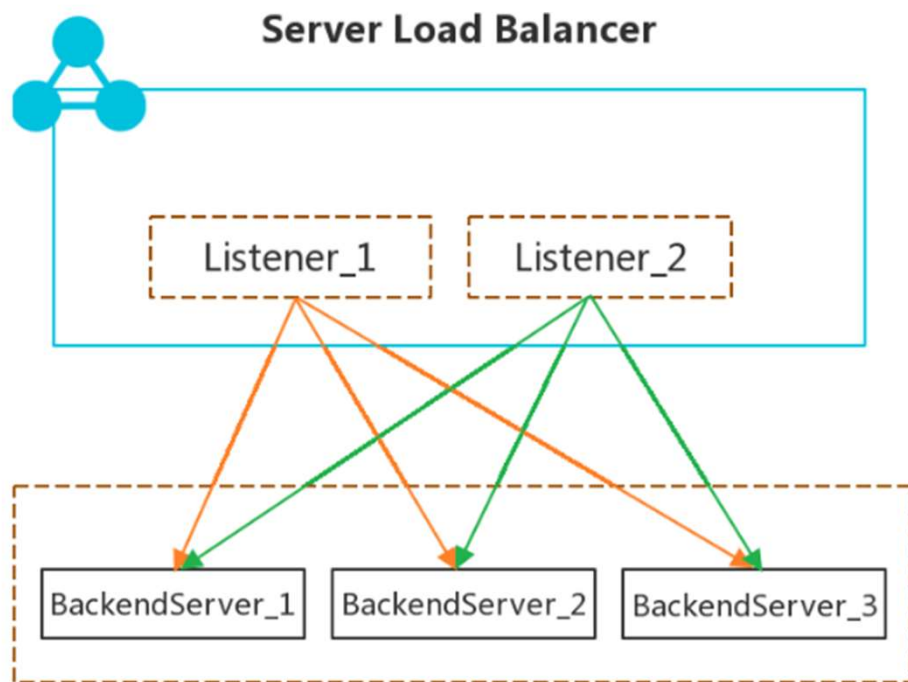
Master+Agent:

- Master接受用户的配置请求
- Agent部署在LVS和Tengine集群上



# SLB的组成

来自外部的访问请求，通过负载均衡实例并根据相关的策略和转发规则分发到后端云服务器进行处理。



**负载均衡实例**

Server Load Balancer instances

**监听**

Listeners

**后端服务器**

Backend Servers

# SLB实例规格

目前SLB支持两种规格：

- 性能共享型实例：资源是所有实例共享的，不保障实例的性能指标。
- 性能保障型实例：提供了可保障的性能指标。当指标超过限定值时，新建连接请求将被丢弃。

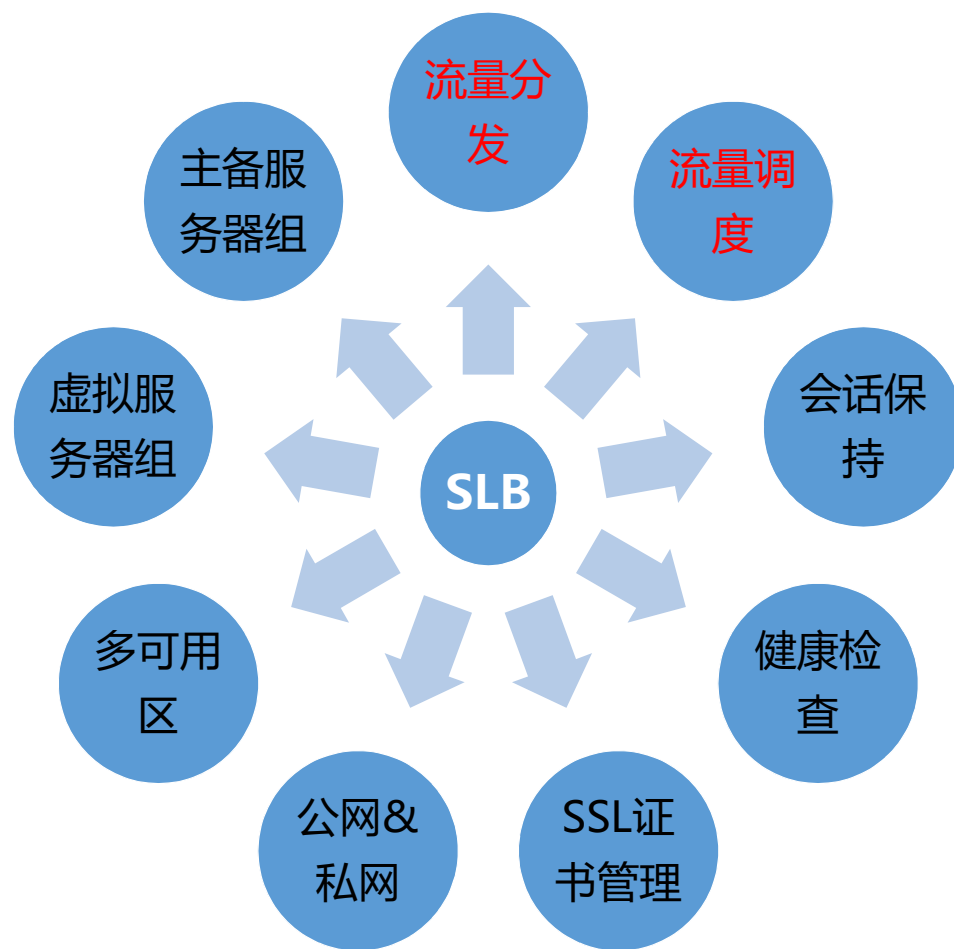


最大连接数

每秒新建  
连接数  
(CPS)

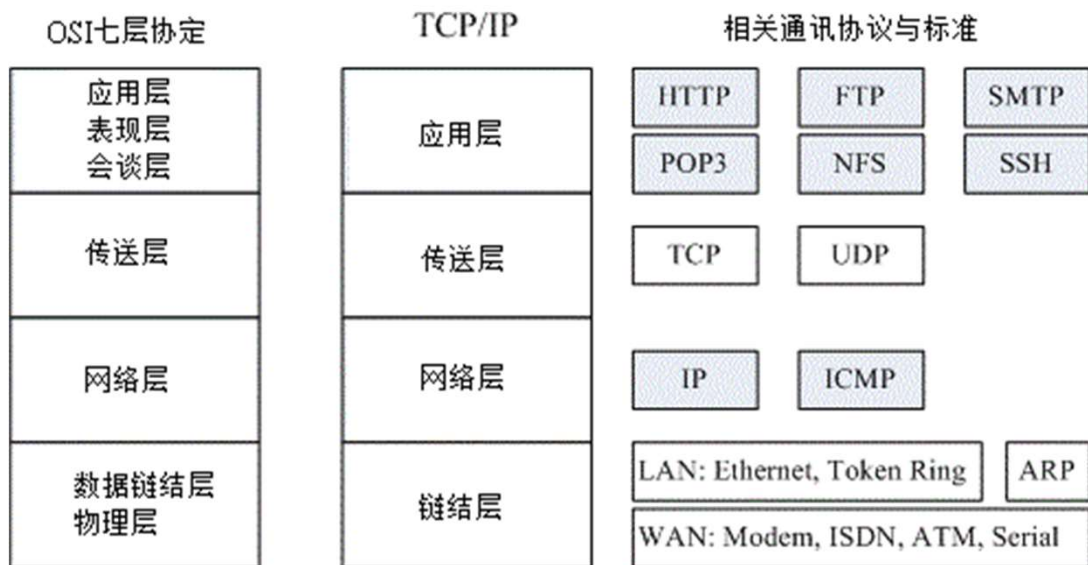
每秒查询数  
(QPS)

# SLB的功能



# 协议支持

SLB当前提供四层和七层的负载均衡服务。



- **四层：**支持TCP/UDP

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：**支持HTTP/HTTPS

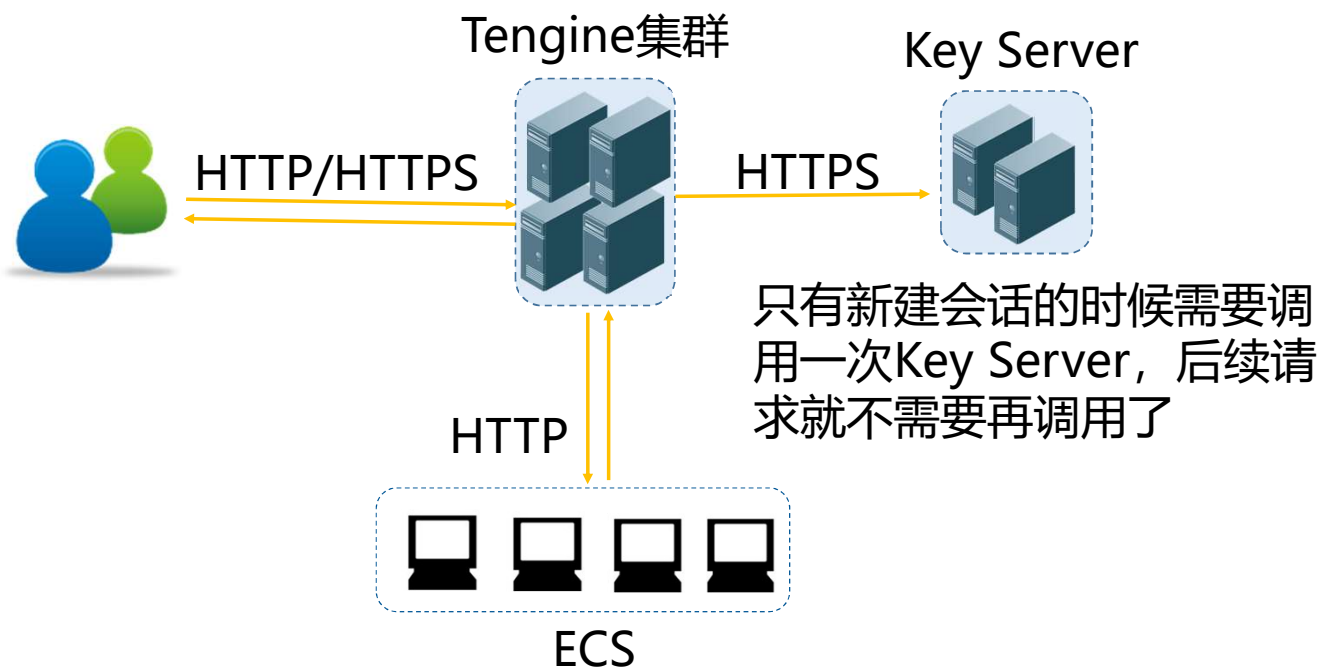
基于请求做调度。比如：http get请求访问一个页面。

- **WS/WSS协议支持：**

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）。

# SSL证书管理

针对HTTPS协议，提供统一的证书管理服务。证书无需上传到后端ECS实例，解密处理在负载均衡上进行，降低后端ECS实例的CPU开销。

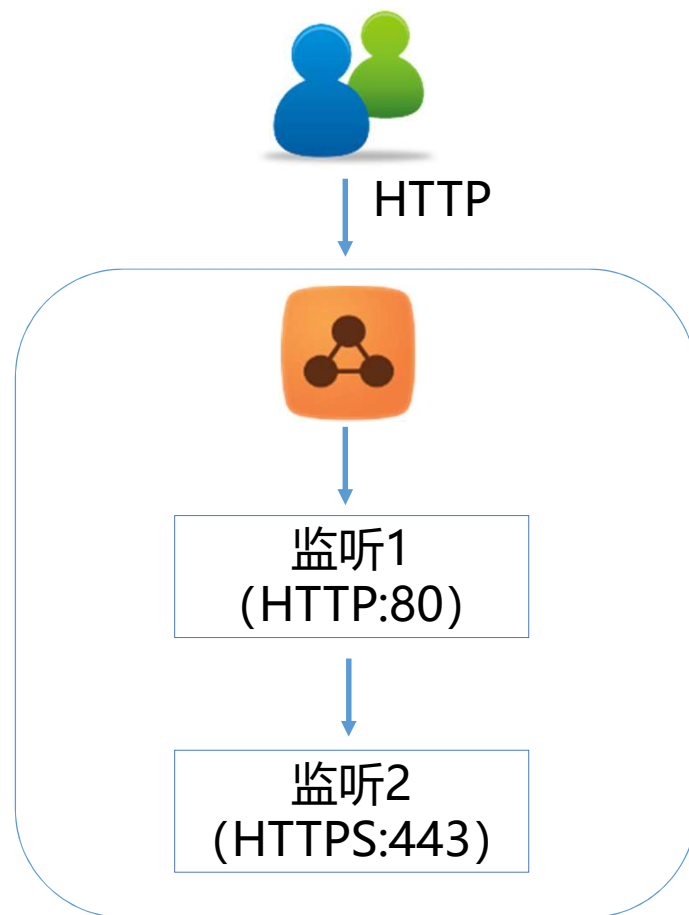


# 监听转发/HTTPS重定向

SLB支持通过将HTTP监听转发至HTTPS监听，实现HTTPS重定向。

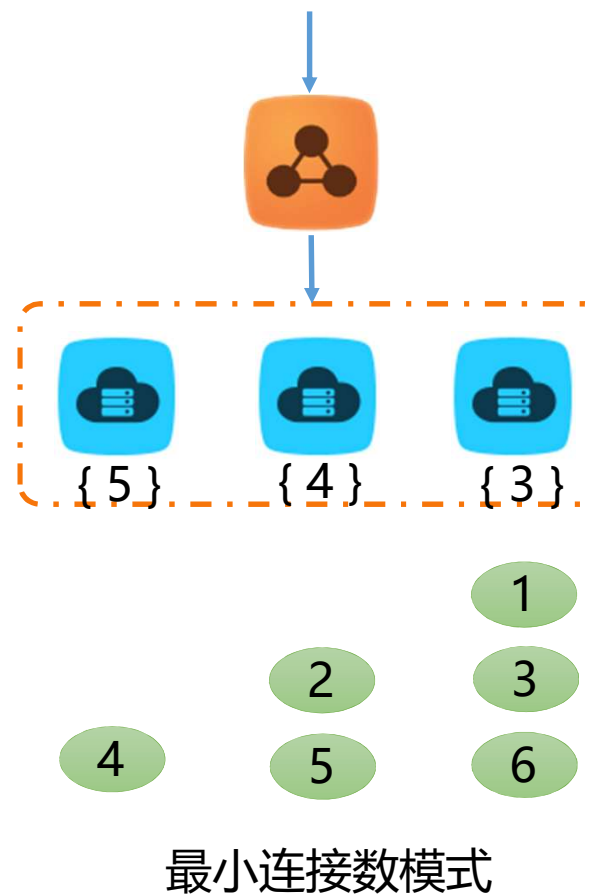
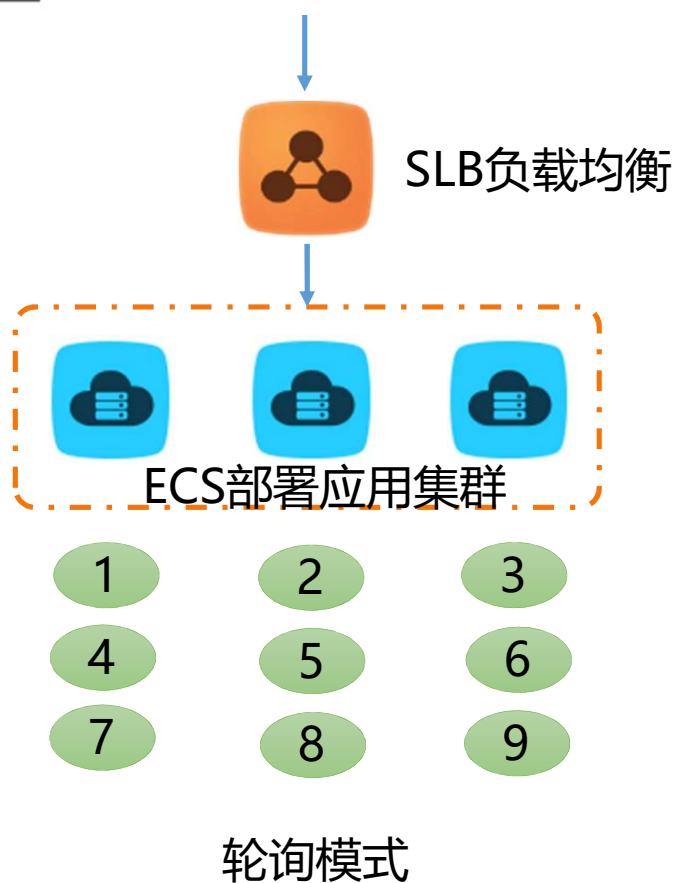
配置方法：

- 1、创建HTTPS监听
- 2、创建HTTP监听，在配置监听时，点击开启“监听转发”功能，选择目的监听为上一步中创建的HTTPS监听。

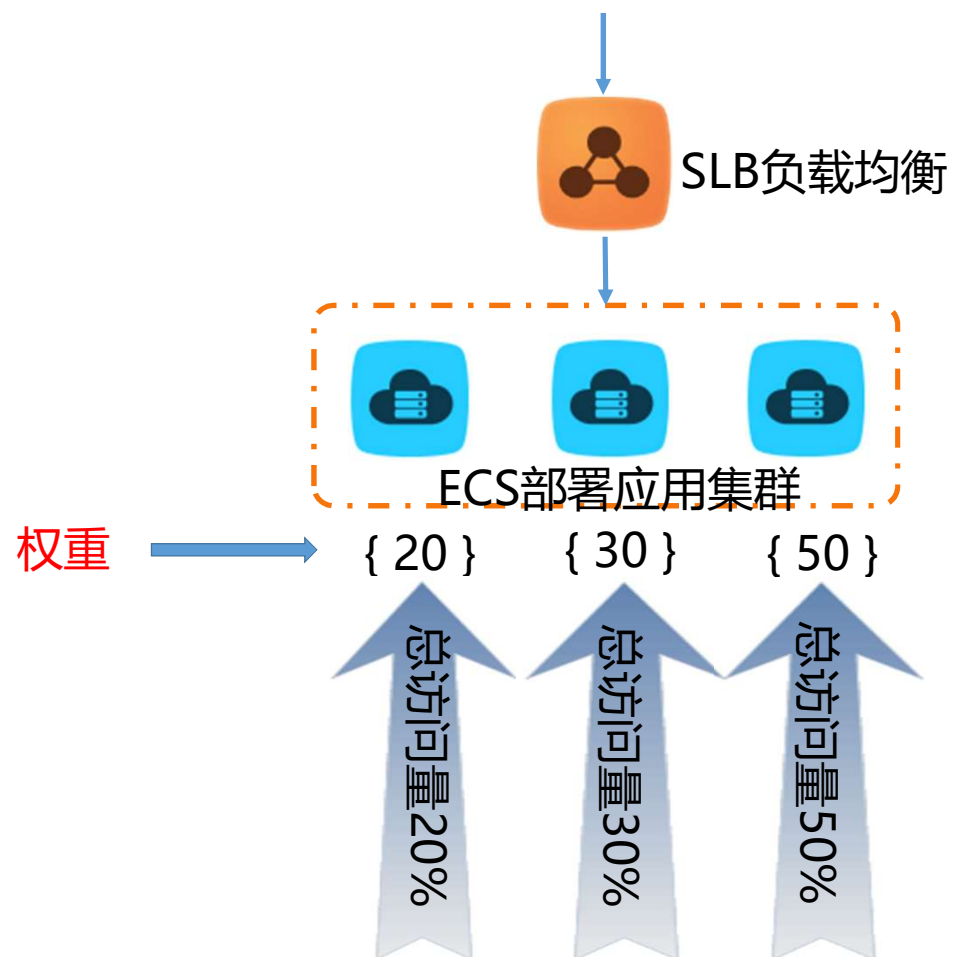




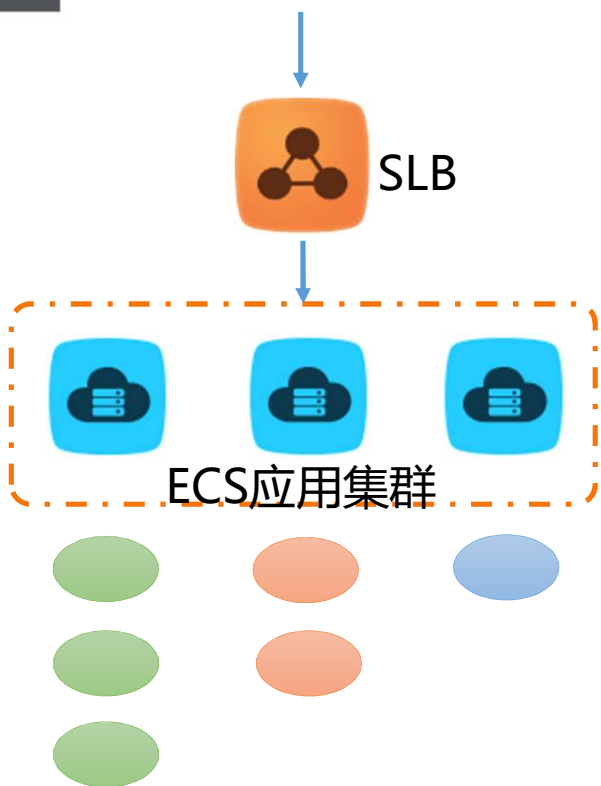
# 调度算法



# 调度算法



# 会话保持



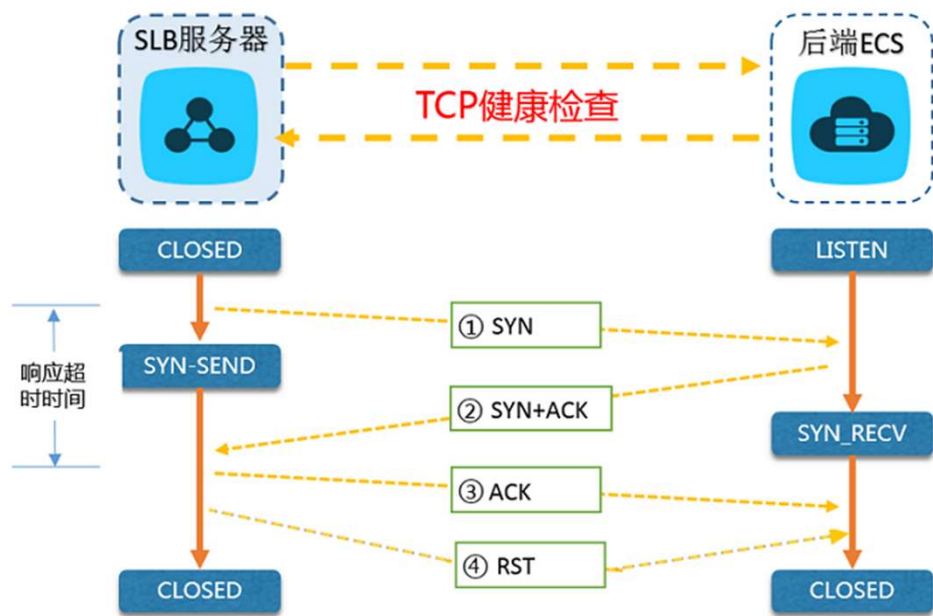
会话保持：在Session的生命周期内，可以将同一客户端请求转发到同一台后端ECS上

- 四层：同一IP地址的请求持续发往一台服务器
- 七层：相同cookie的请求发往一台服务器

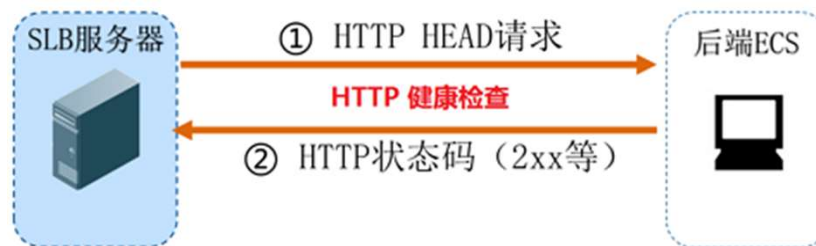
# 健康检查

SLB通过健康检查来判断后端服务器（ECS实例）的业务可用性。

健康检查机制避免了后端ECS异常对总体服务的影响，提高了前端业务整体可用性。

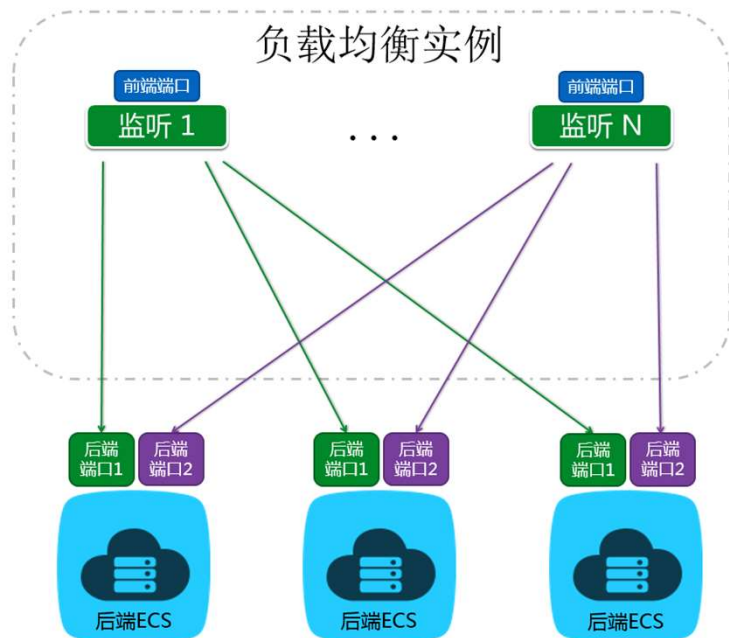


四层健康检查原理

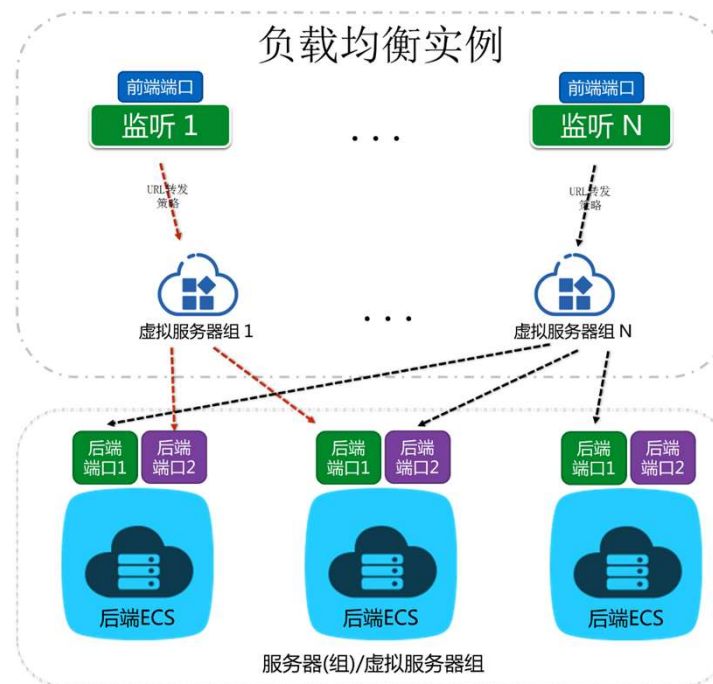


七层健康检查原理

# 虚拟服务器组



未使用虚拟服务器组

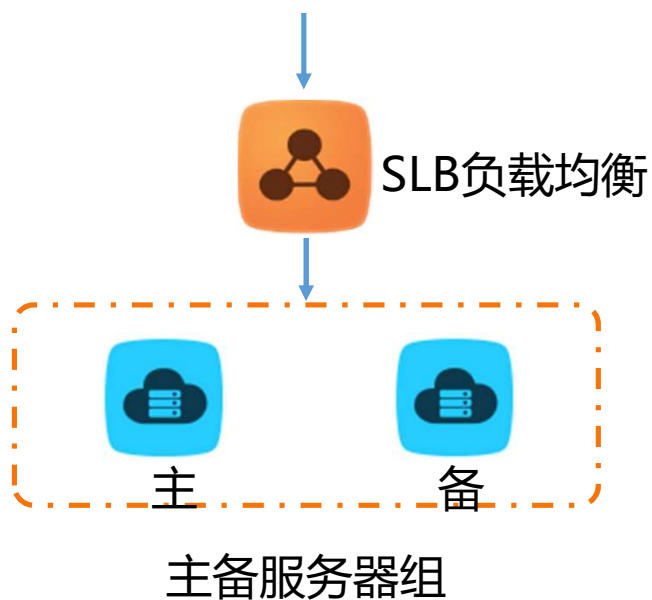


使用虚拟服务器组

# 主备服务器组

**应用场景：**传统的主备需求，即后端服务器中有一台主机和一台备机。

当主机正常工作时，流量将直接走主机；当主机不可用时，流量将走到备机，避免服务中断。



## 注意：

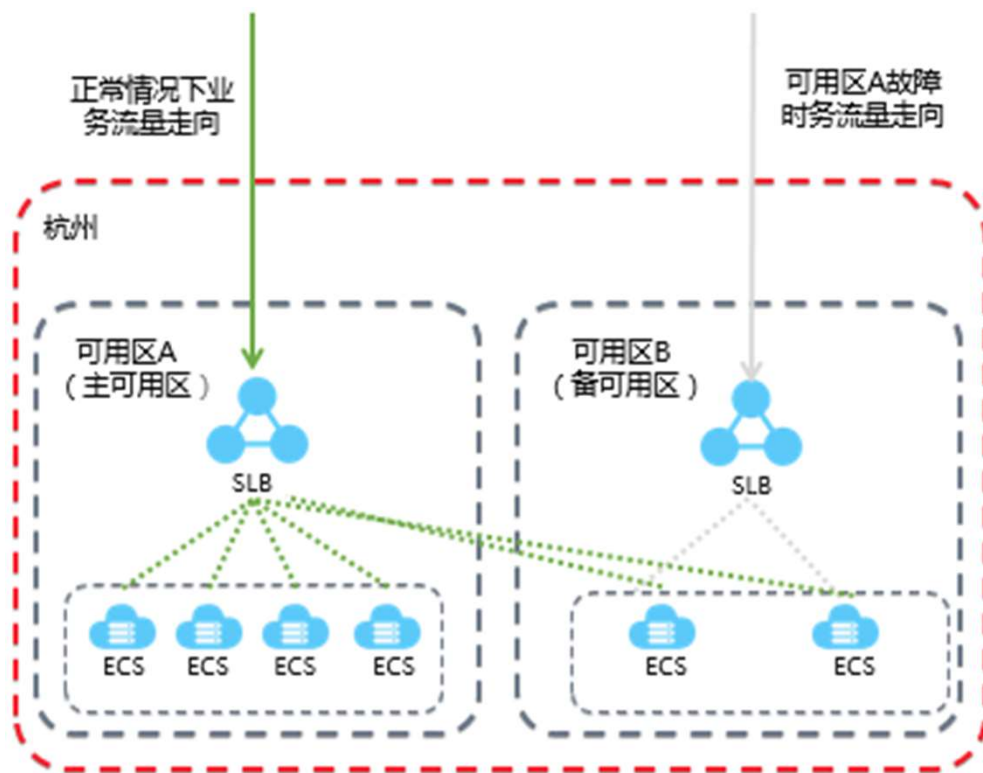
- 1、主备服务器组只支持四层监听（TCP/UDP）
- 2、一个主备服务器组只允许添加两个ECS实例，其中一个作为主机，另外一个作为备机

# 多可用区

阿里云SLB有两种类型：

- (1) “单可用区” 类型
- (2) “多可用区” 类型。

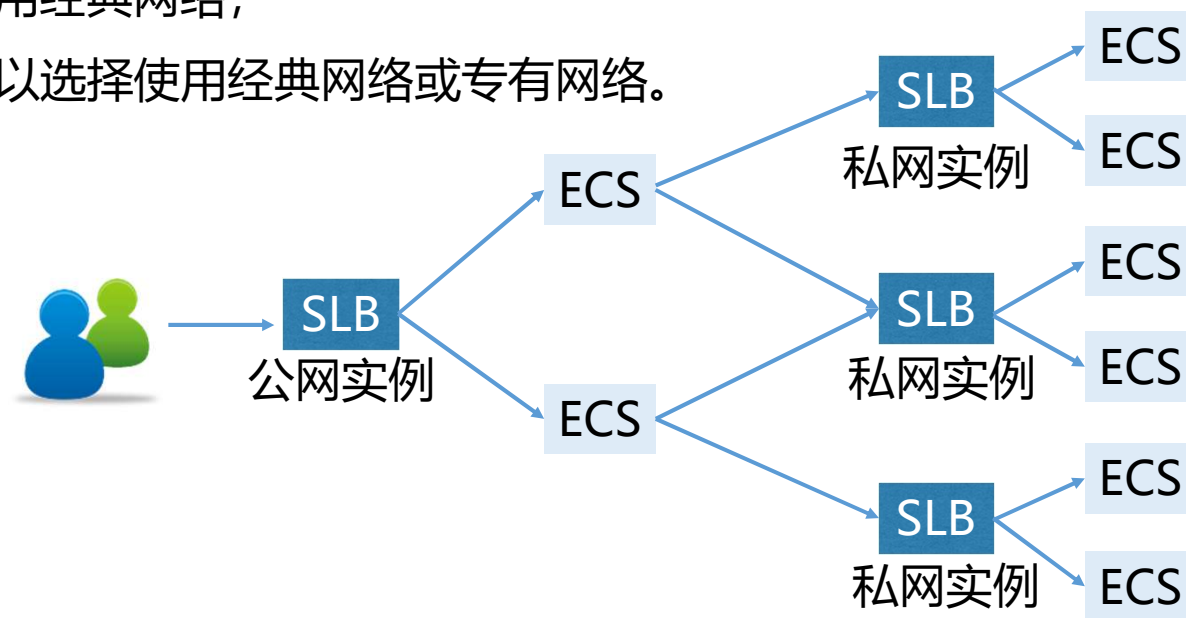
在某些地域，可以选择创建“多可用区”类型的SLB实例，以提高负载均衡服务的可用性。



# 公网&私网实例

根据业务场景来选择配置对外公开或对内私有的负载均衡服务，系统会分配相应的公网或私网服务地址。

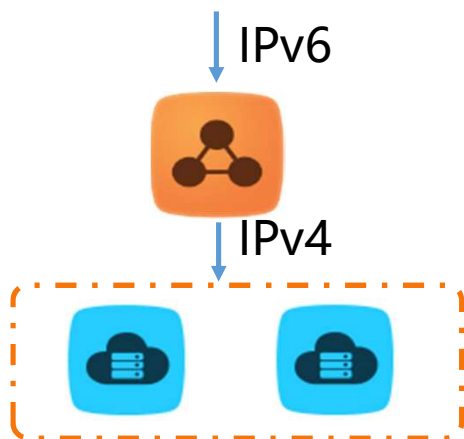
- 公网类型的负载均衡默认使用经典网络；
- 私网类型的负载均衡服务可以选择使用经典网络或专有网络。



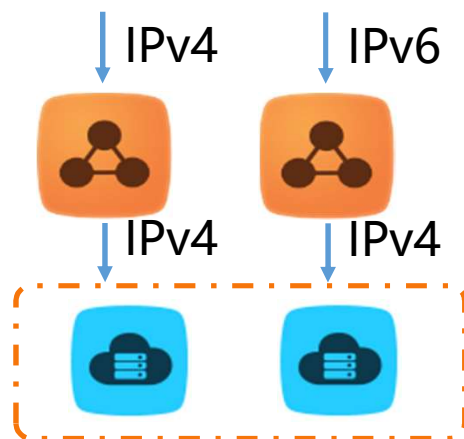


## IPv6实例

SLB支持创建IPv6实例。系统会为实例分配一个公网IPv6地址，转发来自IPv6客户端的请求。



平滑迁移IPv6，业务无感知



新增IPv6入口

# 访问控制

SLB提供监听级别的访问控制，即：为不同监听配置不同的访问控制策略。

支持黑名单、白名单两种访问控制策略，通过“访问策略组”进行黑/白名单的IP管理。



**黑名单**



**白名单**

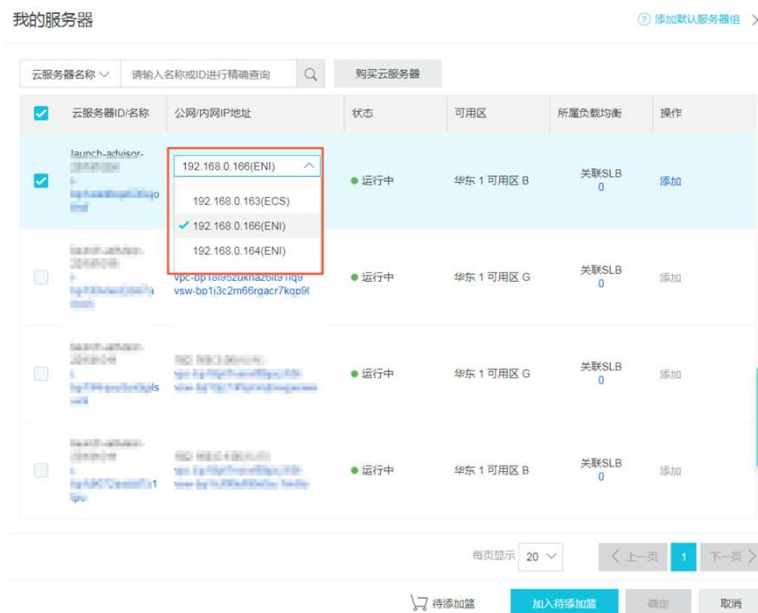
注意：若开启了访问控制，但访问策略组中没有添加任何IP，则负载均衡监听会转发全部请求。

# 后端服务器支持添加ECS弹性网卡ENI

弹性网卡（ENI）是一种可以附加到专有网络VPC类型ECS实例上的虚拟网卡，通过弹性网卡，可以实现高可用集群搭建、低成本故障转移和精细化的网络管理。性能保障型负载均衡实例后端服务器支持挂载ECS弹性网卡ENI。

## 前提条件

负载均衡实例添加后端服务器组时，如果ECS实例绑定多个弹性网卡，支持挂载弹性网卡ENI。



# SLB日志管理

查看负载均衡实例、HTTP监听和服务  
器证书资源一个月内的操作日志。



操作日志：

需要在ActionTrail控制台查看。操作审计(ActionTrail)记录云账户资源操作，提供操作记录查询，并可以将记录文件保存到指定的OSS存储空间。

访问日志：

结合阿里云日志服务，可以通过分析负载均衡的访问日志了解客户端用户行为、客户端用户的地域分布，排查问题等。

健康检查日志：

可以在日志管理页面，查看三天内的健康检查日志。如需要更久的健康检查日志，需要将健康检查日志存储到OSS中，并可以下载完整的健康检查日志。

# 闲置实例与配额管理

闲置实例向您展示超过7天未投入使用的后付费实例，保持关注闲置实例，有助于更好的管理成本。

## 闲置实例

闲置实例向您展示超过7天未投入使用的后付费实例，保持关注闲置实例，有助于更好的管理成本。

实例名称/ID	服务地址	地域	闲置原因	操作
实例名称/ID	服务地址	地域	没有监听	释放设置
实例名称/ID	服务地址	地域	没有监听	释放设置

负载均衡 SLB

实例

实例管理

回收站

证书管理

访问控制

日志管理

操作日志

访问日志

健康检查日志

SLB 实验室

闲置实例

配额管理

快捷链接

负载均衡 SLB

闲置实例

闲置实例向您展示超过7天未投入使用的后付费实例，保持关注闲置实例，有助于更好的管理成本。

实例名称/ID

批量释放

可以通过负载均衡控制台查询当前资源配额使用情况。如果某个资源的剩余配额不满足业务需求，也可以直接申请增加配额。

阿里云

负载均衡 SLB

配额管理

配额名称	描述	类型	用量/配额	申请历史	操作
slb_quota_backendservers_num	每个slb实例后端可以挂载的服务器数量	配额	0/200	-	申请
slb_quota_acl_entries_num	每个访问控制策略组中保有的访问控制条目数量	配额	0/300	-	申请
slb_quota_certs_num	每个地域可保有的服务器证书数量	配额	0/100	-	申请
slb_quota_ca_certs_num	每个地域可保有的客户端CA证书数量	配额	0/100	-	申请
slb_quota_acl_attached_num	每个访问控制策略组关联不同监听的次数	配额	0/50	-	申请
slb_quota_acls_num	每个账户下可配置访问策略组数量	配额	0/50	-	不能申请
slb_quota_domain_extensions_num	每个监听上可以保有的扩展域名数量	配额	0/3	-	申请
slb_quota_rules_num	每个监听上可以保有的域名URL转发策略数量	配额	0/40	-	不能申请
slb_quota_listeners_num	每个实例可以保有的监听数量	配额	0/50	-	申请
slb_quota_instances_num	用户可保有的slb实例个数	配额	0/60	-	申请
slb_quota_backendserver_attached_num	同一台服务器可以重复添加为slb后端服务器的次数	配额	0/50	-	不能申请
slb_privilege_listener_same_port	TCP/UDP协议监听同端口公网权限	特权	-	-	申请
slb_privilege_allow_all_https_redirect	拥有任意https端口跳转到任意https端口的权限	特权	-	-	申请

# SLB的优势



## 安全稳定

SLB目前已经在阿里云内部多个系统使用，系统安全性和稳定性很高，可胜任生产要求。



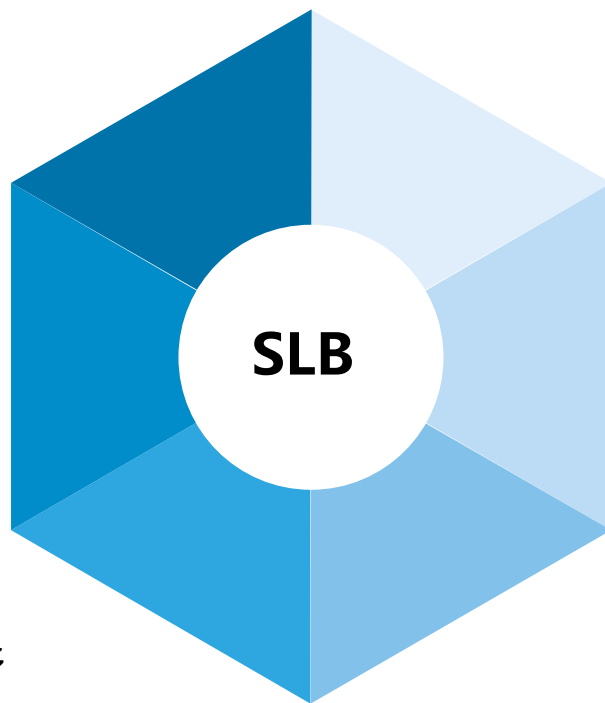
## 无限扩容

无性能、容量的限制，且自动扩容。



## 无单点故障

集群部署，无单点故障，并且支持多可用区，从而保证应用高可用。



## 按需计费

只为使用的流量付费。



## 功能强大

提供4层和7层的负载均衡，无论是Web服务器，还是中间件服务器，都可完美支持。



## 配置简单

调用API或使用portal都可以随时配置负载均衡，立即生效，灵活简单。

# 目录

1. 负载均衡服务的产生背景
2. 阿里云负载均衡服务SLB介绍
- 3. SLB的配置流程**
4. SLB的最佳实践
5. SLB的典型应用案例

# SLB的配置流程





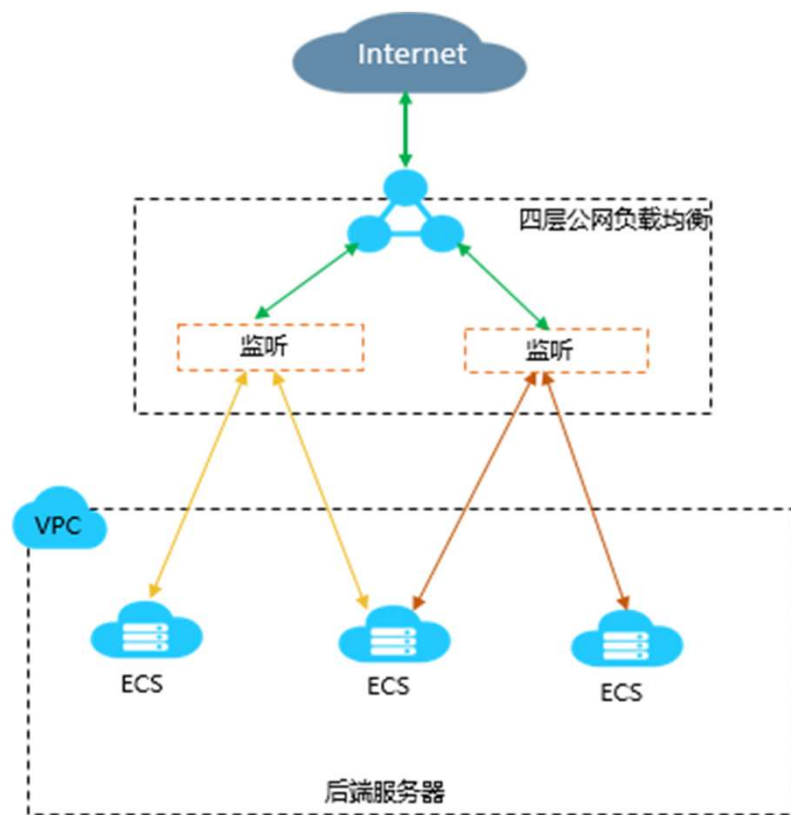
# 目录

1. 负载均衡服务的产生背景
2. 阿里云负载均衡服务SLB介绍
3. SLB的使用方法
- 4. SLB的最佳实践**
  - 4.1 作为公网流量入口
  - 4.2 多级负载均衡
  - 4.3 虚拟服务器组
  - 4.4 同城容灾架构
5. SLB的典型应用案例

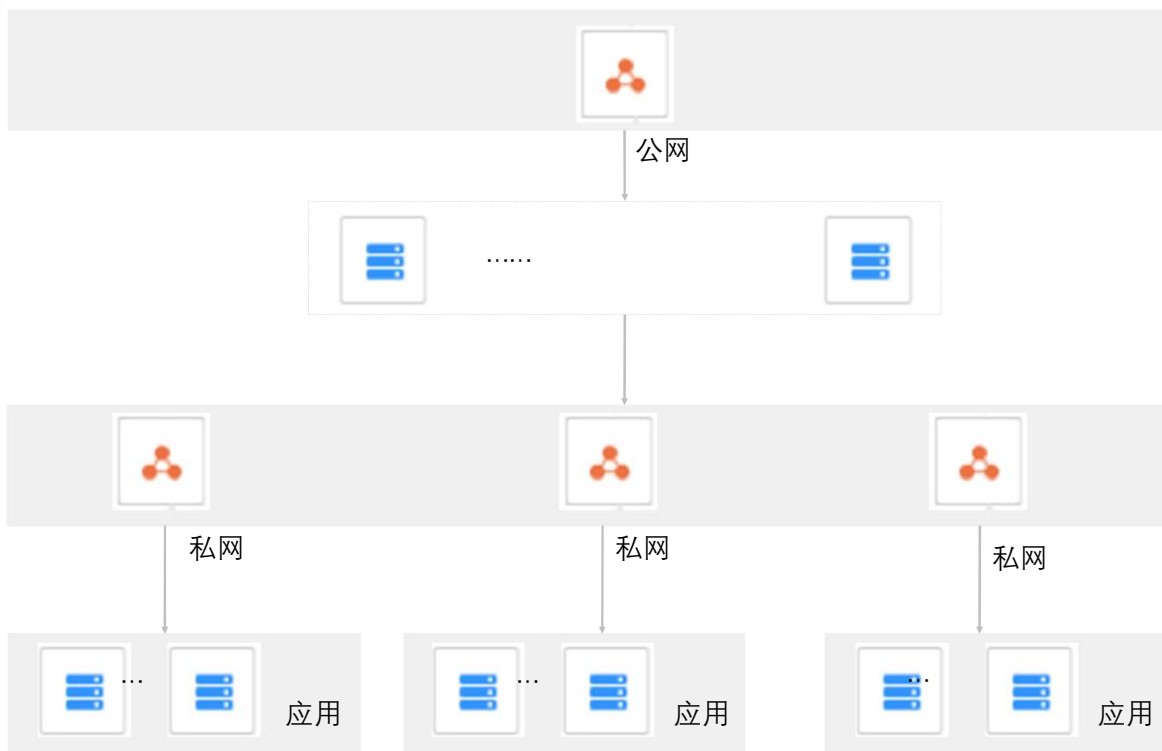
## 最佳实践1：作为公网流量入口

架构解读：

这是一种典型的使用方式。在ECS上部署无状态的应用，使用SLB服务作为公网流量的入口，可以提高系统的可用性，同时，后端ECS不再需要公网IP和公网带宽。



## 最佳实践2：多级负载均衡



**场景：**有对公应用，也有很多应用不需要对公网访问，但都需要负载均衡

**解决方案：**选用公网SLB+私网SLB 搭建两级负载均衡。公网SLB承接对公访问，公网SLB的后端ECS调用私网SLB

**优势：**

- 节约成本：私网SLB免费
- 更加安全：私网SLB对公无法访问，保护内部敏感应用
- 配置简单

# 最佳实践3：实现相同域名不同路径的流量转发

使用 **虚拟服务器组 + URL转发规则**，实现将相同域名不同路径的流量转发至不同的后端虚拟服务器组。

详情

监听

服务器

监控

training\_slb

返回负载均衡列表

使用限制和注意事项

添加监听

刷新

监听

前端协议/端口	后端协议/端口	状态	转发规则	会话保持	健康检查	带宽峰值	服务器组	操作
HTTP: 80	HTTP: 80	运行中	加权轮询	关闭	已开启	无限制	无	配置   详情   更多

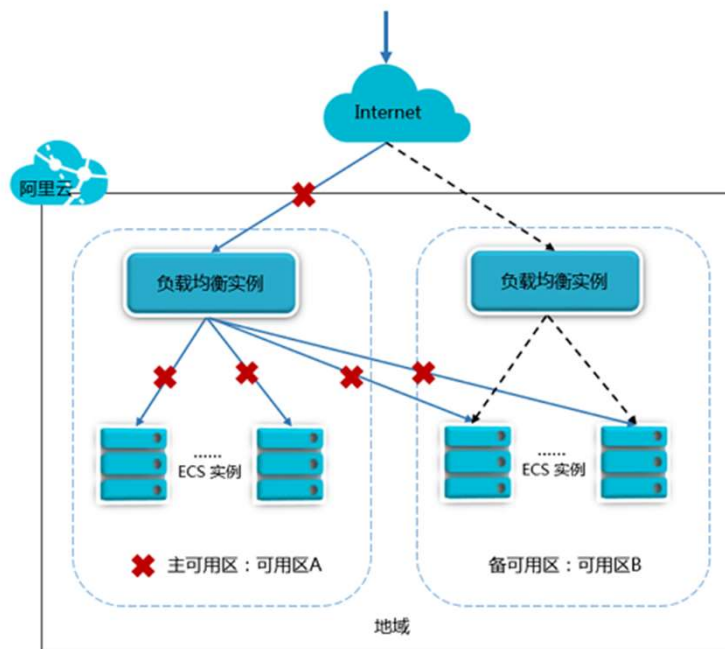
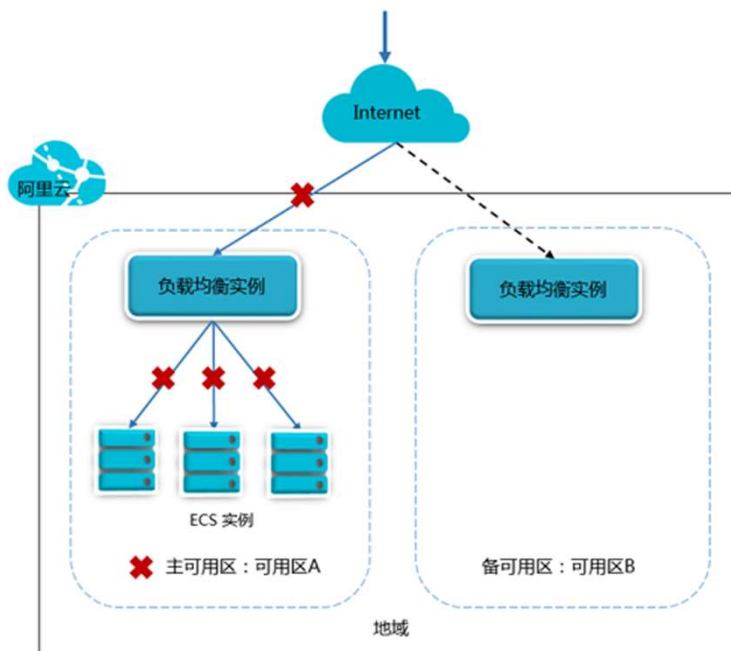
规则名称	域名	URL	虚拟服务器组	操作
rule1	www.aaa.com	/jerry	JERRY	删除
rule2	www.aaa.com	/tom	TOM	删除
rule3	www.aaa.com		TOM	删除

添加转发策略 +

- 启动
- 停止
- 删除
- 设置访问控制
- 添加转发策略



## 最佳实践4：同城容灾架构



1. 购买负载均衡实例时，选择可用区类型为多可用区的地域，在SLB的主、备可用区均部署ECS；
2. 在极端情况下（主可用区整体不可用时），负载均衡30秒内切换到备可用区并恢复服务；
3. 当主可用区恢复时，负载均衡自动切换到主可用区。

# 目录

- 1. 负载均衡服务的产生背景
- 2. 阿里云负载均衡服务SLB介绍
- 3. SLB的使用方法
- 4. SLB的最佳实践
- 5. SLB的典型应用案例**
  - 5.1 可扩展的应用架构
  - 5.2 金融级两地三中心架构

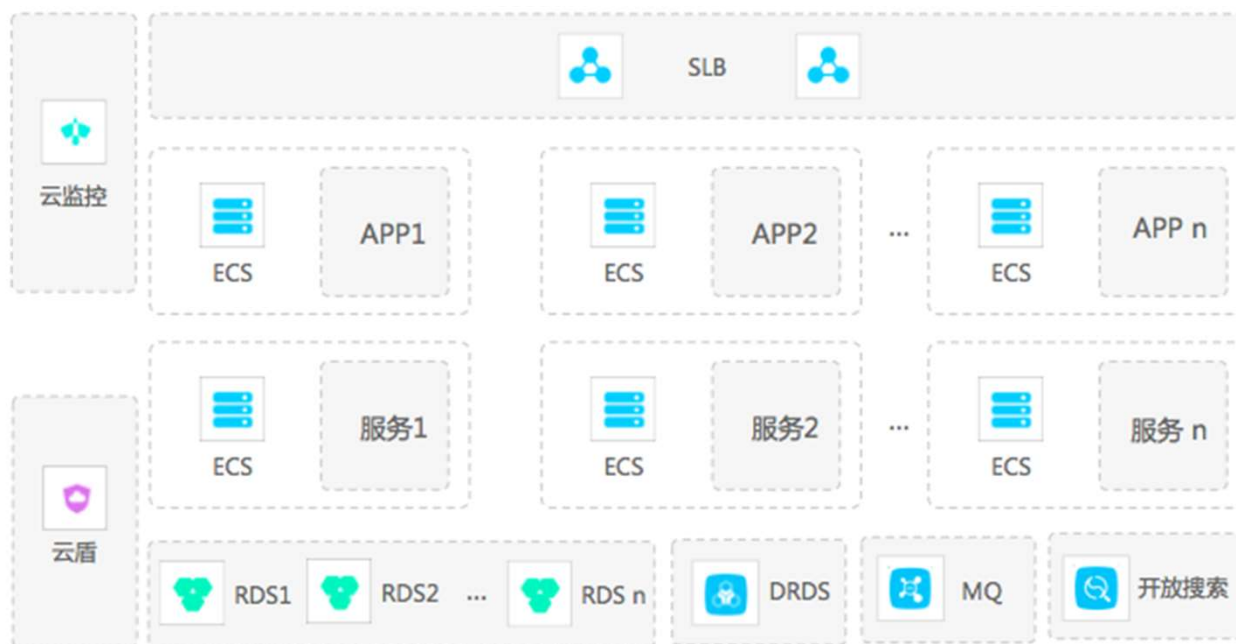
# 应用案例1：可扩展的应用架构

某家园共育平台，专注于3-6岁孩子的成长教育

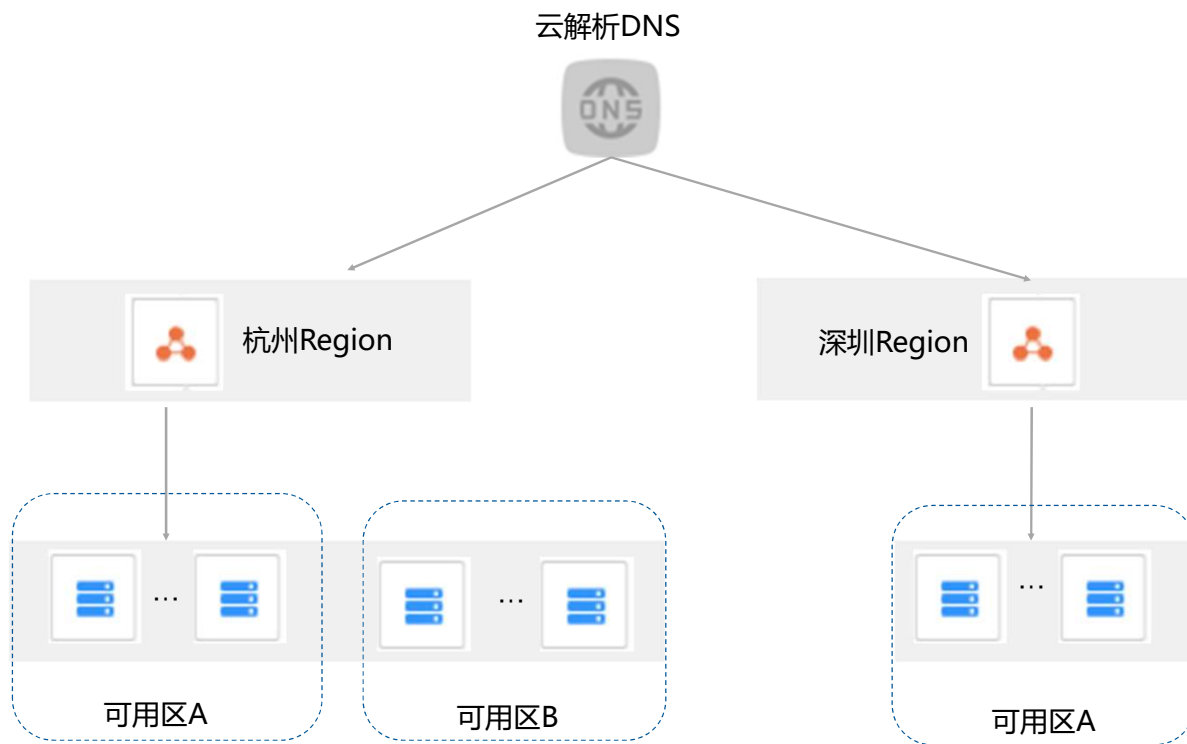
**面临挑战：**用户大规模的增长，系统需要应对高并发的访问量

## 架构解读：

- 分布式部署、服务化架构，易扩展、高可用；
- SLB实现负载均衡，统一带宽流量入口，消除单点故障，实现高可用，提高系统可用性；



## 应用案例2：金融级两地三中心架构



**场景：**金融应用对可靠性要求极高，合规也要求两地三中心。

### 解决方案：

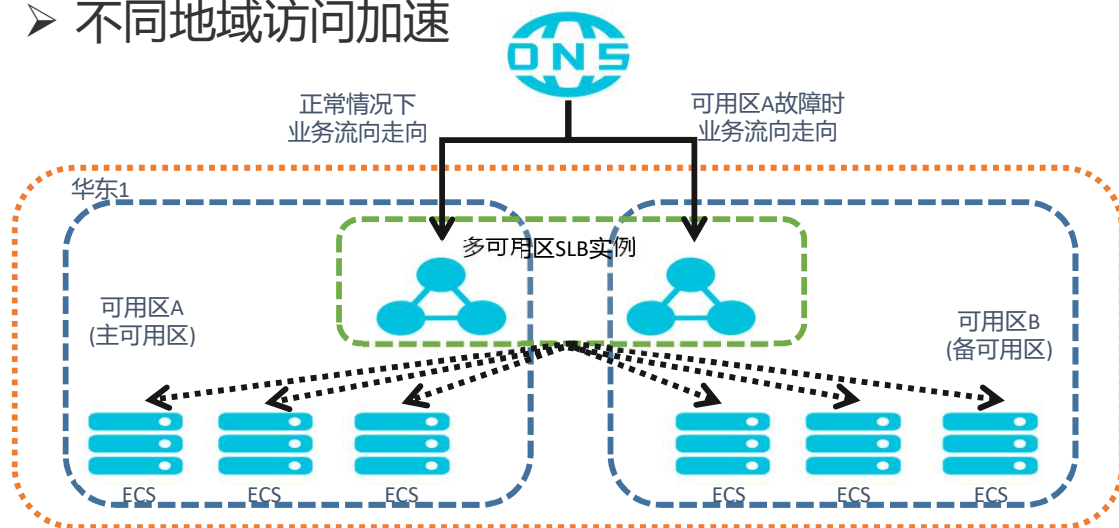
- SLB后端挂载多台同Region下不同可用区的ECS，结合SLB自身的同城容灾，实现一个可用区故障时业务自动切换到另外一个可用区继续服务。
- 配合使用云解析智能DNS，可做到多Region冗余。



## 应用案例3：结合全局流量管理实现跨地域负载均衡

结合全局流量管理，您可在地域负载均衡上层部署全局流量管理，实现跨地域容灾、不同地域访问加速和智能解析。

- 多线路智能解析服务
- 跨地域容灾
- 不同地域访问加速



新增地址池

\* 地址池名称:

华北\_北京

\* 地址池类型 ?

IP

\* 最小可用地址数量 ?

1

地址	模式
4	智能返回

+ 新增一行

取消

确认

## 小结

- 1、SLB的典型应用场景是什么？
- 2、SLB有哪些功能？
- 3、SLB有哪些优势？

为了无法计算的价值 |  阿里云

