

目 录

| | |
|-------------------|---|
| 目录 | 1 |
| 什么是应用型负载均衡ALB | 2 |
| 什么是应用型负载均衡ALB | 2 |
| 经典型负载均衡与应用型负载均衡对比 | 2 |
| 实例性能指标 | 2 |
| ALB组成 | 2 |
| ALB类型 | 2 |
| 使用流程&限制 | 2 |
| 配额限制 | 2 |
| 支持的地域与可用区 | 2 |
| 产品架构 | 2 |
| 基础架构说明 | 2 |

什么是应用型负载均衡ALB

什么是应用型负载均衡ALB

应用型负载均衡ALB（Application Load Balancer）是专门面向HTTP、HTTPS的应用层负载场景的负载均衡服务，可实现性能独享，资源隔离，具备超强弹性及大规模应用层流量处理能力。

经典型负载均衡与应用型负载均衡对比

| 对比项 | 经典型负载均衡SLB | 应用型负载均衡ALB |
|---------|---|--|
| 产品定位 | <ul style="list-style-type: none">- 提供基础的四层、七层处理能力- 面向TCP、UDP、HTTP和HTTPS | <ul style="list-style-type: none">- 强大的七层处理能力- 聚焦HTTP、HTTPS应用层协议- 面向应用层交付 |
| 产品架构与性能 | <ul style="list-style-type: none">- 共享型实例，容易产生邻居吵闹效应- 性能不做承诺 | <ul style="list-style-type: none">- 实例级隔离，无租户间影响- 单实例QPS默认为10w，最大上限100w- 单实例并发连接数200w- 单实例新建连接数默认10w，最大100w |
| 运维能力 | <ul style="list-style-type: none">- 无弹性- 需要预估业务峰值- 支持日志上传到KS3 | <ul style="list-style-type: none">- 自动弹性- 根据业务峰值自动伸缩，无需人工干预- 控制台提供日志分析能力 |

实例性能指标

| 默认指标 | 最大每秒请求数（QPS） | 最大新建连接数（CPS） | 最大并发连接数 | 最大私网带宽 | 最大公网带宽 |
|------|--------------|--------------|---------|--------|--------|
| 10W | 10w | 100w | 10Gbps | 15Gbps | |

说明：如有更高性能需求，请联系售后人员

ALB组成

实例：负载均衡实例，通过将流量分发到不同的后端服务器来扩展应用系统的服务吞吐能力。
监听器：监听器是ALB的最小业务单元，监听上需要配置协议与端口。
后端服务器组：服务器组是一个逻辑组，包含多个后端服务器用于处理ALB分发的业务请求。ALB服务器组支持云服务器，裸金属服务器的后端服务器。

ALB类型

目前仅提供公网类型的ALB

使用流程&限制

请联系您的客户经理提配额管理流程申请权限

配额限制

配额是指一个主账号可以使用的云资源的最大值或操作次数的最大值。本文介绍应用型负载均衡ALB的配额项、默认值、是否支持配额提升等概况。

| 配额项 | 说明 | 默认值 | 是否支持配额提升 |
|-------------------------|---------------|-----------|----------|
| quota_slb_alb | 是否开通ALB权限 | 1 不支持 2支持 | 是 |
| quota_slb_albrulegroups | 配额监听器下的转发策略数量 | 40 | 是 |

支持的地域与可用区

5月中旬俄罗斯（莫斯科）公测 6月中旬华北1（北京）公测

产品架构

基础架构说明

ALB采用Tengine实现负载均衡。Tengine在Nginx的基础上，针对有大访问量的网站需求，添加了很多高级功能和特性。



- 可用区灾备：实例多可用区部署，机房级容灾
- 故障域隔离：ALB实例L7资源独享，不受他人影响
- 弹性伸缩平滑：扩容连接合理分配，缩容已有连接保持
- 配置生效时间提升显著：L7网关配置生效时间从180s提示到10s
- 快速迭代，灰度发布：镜像独立管理，新特性可按租户灰度迭代
- 资源扩容不依赖域名：负载均衡双层架构，L7资源扩容优雅