

目录

目录	1
产品概述	2
负载均衡是什么？	2
负载均衡的组成部分	2
为什么选择负载均衡？	2
负载均衡的工作原理	2
访问SLB	2
负载均衡收费说明	3
产品功能	3
基本转发服务	3
灵活绑定解绑弹性IP	3
健康检查服务	3
会话保持	3
支持HTTPS	3
产品优势	3
高性能	3
高稳定性	3
低成本	3
安全	3
使用场景	3
高访问量业务流量分发	3
消除单点故障	3
跨可用区容灾部署	3
名词解释	3
名词解释	4
产品规格	4
产品规格	4
最大并发连接数（Max Connection）	4
每秒新建连接数（Connection Per Second）	4
每秒查询速率（Query Per Second（QPS））	4
带宽（Mbps）	4

产品概述

负载均衡是什么？

负载均衡（Server Load Balancing，简称SLB）是对多台后端服务器进行流量分发的网络服务。通过设置虚拟服务地址（VIP），将位于同一地域的多台后端服务器资源虚拟成一个高性能、高可用的应用服务池，然后根据应用指定的方式，将来自客户端的网络请求分发到服务池中。

负载均衡的组成部分



一个提供服务的负载均衡通常由以下三个部分组成：

- 负载均衡实例：用来接收流量并将其分配给后端服务器。
- 监听器：用来检查客户端请求并将请求转发给后端服务器。监听也会对后端服务器进行健康检查。
- 后端服务器：用于实际处理请求。

为什么选择负载均衡？

高可用

金山云负载均衡可根据您指定的方式，将您多台后端服务器组成一个高性能、高可用的应用服务集群，将网络请求分发到集群中的服务器中。

您开启健康检查后，负载均衡服务会检查服务器的健康状态，自动将请求负载到状态健康的服务器中，从而提高服务的可靠性和稳定性。

灵活

金山云负载均衡的监听器及真实服务器的端口均可通过控制台自助调整。

安全

负载均衡与虚拟私有网络配合使用，提供稳定的联网和安全功能。金山云负载均衡免费提供防5Gbps的SYN-FLOOD攻击功能。您可以在证书管理轻松上传SSL证书，并通过创建HTTPS监听器轻松为您的站点或应用程序启用SSL。

负载均衡的工作原理

基本工作原理

负载均衡器接受来自客户端的传入流量并将请求路由到后端服务器上进行处理。

负载均衡服务主要由负载均衡监听器提供。监听器负责监听负载均衡实例上的请求、执行策略分发至后端服务器等任务，通过配置 **客户端-负载均衡** 和 **负载均衡-后端服务器** 两个维度的转发协议及协议端口，负载均衡可以将请求直接转发到后端服务上。

请求路由选择

客户端请求通过域名访问服务，在请求发送到负载均衡器之前，域名解析服务器将会解析负载均衡域名，并将对应的负载均衡IP地址返回到客户端。

当负载均衡监听器收到请求时，将会使用不同的负载均衡算法将请求分发到后端服务器中。目前金山云支持轮询、加权轮询、最小连接数、加权最小连接数等多种均衡算法。

后端服务状态

负载均衡器还可以检测后端实例的运行状况，从而确保只将流量路由到正常运行的实例上去。

当负载均衡器检测到运行不正常的实例时，它会停止向该实例路由流量，然后会在它再次检测到实例正常运行之后重新向其路由流量。

访问SLB

通过注册金山云账号，您可以通过以下方式访问和管理SLB：

- [负载均衡SLB控制台](#)：具有交互式操作的Web服务页面。您可登录控制台完成SLB实例的创建、使用或释放，具体操作，请参见[创建和管理SLB实例](#)。
- [金山云SDK](#)：提供Java、Python、PHP、Go等多种编程语言的SDK。

- [OpenApi开发者门户](#)：提供快速检索接口、在线调用API和动态生成SDK示例代码等服务。

负载均衡收费说明

当前负载均衡提供免费服务。负载均衡绑定的EIP单独收取公网带宽费用。详见[EIP计费](#)。

产品功能

基本转发服务

支持公网、私网类型服务；提供4层（TCP、UDP）和7层（HTTP、HTTPS）的网络业务服务。

灵活绑定解绑弹性IP

支持绑定解绑弹性IP，增加了灵活性，提升业务的可用性。

健康检查服务

分布式集群根据配置对后端服务器完成健康检查，自动隔离状态异常的后端服务器。

会话保持

对 TCP 和 HTTP 协议都支持会话保持功能，将同一个客户的请求分发到后端同一个服务器上。

支持HTTPS

HTTPS 协议提供集中化的证书管理系统，满足您对数据可靠传输、快速传输与安全传输等多样化需求。

产品优势

高性能

分布式集群满足大规模业务分发的性能要求

高稳定性

冗余设计，无单点故障；后端服务器可以随业务量方便的扩展

低成本

使用金山云的负载均衡实例，可以快速低成本搭建业务；对于内网负载均衡产品，目前实行免费创建实例并分配IP地址

安全

免费的 DDoS 攻击防御功能，无延时动态启动

使用场景

高访问量业务流量分发

对于业务量访问较大的业务，可以通过设置监听器规则，将访问量均匀的分到多个后端服务器处理。

消除单点故障

您可以在负载均衡实例下添加多台云服务器实例。当某台或某几台云服务器不可用时，负载均衡可自动屏蔽故障的实例，将请求分发给正常运行的实例，保障确保业务正常运行。

跨可用区容灾部署

对于可靠性和容灾有很高要求的业务，负载均衡可将流量跨可用区进行分发，建立实时的业务容灾部署。

名词解释

名词解释

术语	简称	全称	说明
负载均衡	SLB	Server Load Balancing	金山云提供的一种网络负载均衡服务，可以为后端服务器提供流量分发
监听器	Listener	Listener	负载均衡服务监听器，包括监听端口、负载均衡策略、健康检查和转发方式配置等，每个监听项对应后端的一个应用服务，可添加真实服务器
真实服务器	RS	Real Server	接受流量分发的后端服务器实例，流量将会按照配置的规则转发
虚拟服务地址	VIP	Virtual IP	系统分配的服务地址，当前为 IP 地址。用户可以选择该服务地址是否对外公开，来分别创建公网和私网类型的负载均衡服务

产品规格

产品规格

经典型负载均衡SLB支持的性能指标和使用限制如下，当实际业务超过指标性能上限时，会导致新建连接请求受限以及丢包问题。

最大并发连接数（Max Connection）

并发最大连接数是指从客户端到负载均衡建立的所有连接的数量。当SLB实例上的连接数超过上限时，新建连接请求将被丢弃。当经典型负载均衡的监听器协议类型为TCP时，最大并发连接数限制为100万；当监听器协议类型为UDP时，限制为50万。

每秒新建连接数（Connection Per Second）

每秒新建连接数是指从客户端到负载均衡每秒新建立的TCP连接数。经典型负载均衡SLB的每秒新建连接数没有限制。

每秒查询速率（Query Per Second（QPS））

每秒查询速率是针对七层（HTTP/HTTPS）监听而言，表示负载均衡SLB每秒可以处理查询请求的数量。当请求速率超过上限时，新建连接请求将被丢弃。SLB每秒处理的请求数上限为1万。

带宽（Mbps）

SLB支持的最大内网带宽为8Gbps。