目录

裸金属服务器托管

金山云提供裸金属服务器无缝接入VPC网络的托管服务，可以和您的云服务器或裸金属服务器在同一个VPC下，提供完整的生命周期管理服务，可以像云服务器一样方便的使用您的托管裸金属服务器。

生命周期管理

1. 控制台或openAPI提供以下功能。

   开机、关机、重启、分配公网IP、调整带宽、更换内网IP、配置DNS

运维管理

1. 提供标准的带外管理服务。
2. 提供标准的流量监控告警服务，可通过短信、邮件通知。

VIP服务

1. 提供7*24小时的售后运维服务。
2. 提供每天3次的现场巡检服务。
3. 超过50台同机型托管的客户，金山云可提供适配服务，适配后可在控制台重装系统及制作镜像。

托管流程

1. 联系金山云商务沟通详细需求，联系产品经理。
2. 下载裸金属服务器托管业务需求信息表填写相关内容。
3. 邮寄裸金属服务器至金山云指定机房及接收人,详情商务与IDC沟通。
4. 商务联系裸金属服务器团队安排上架交付客户。

裸金属服务器监控

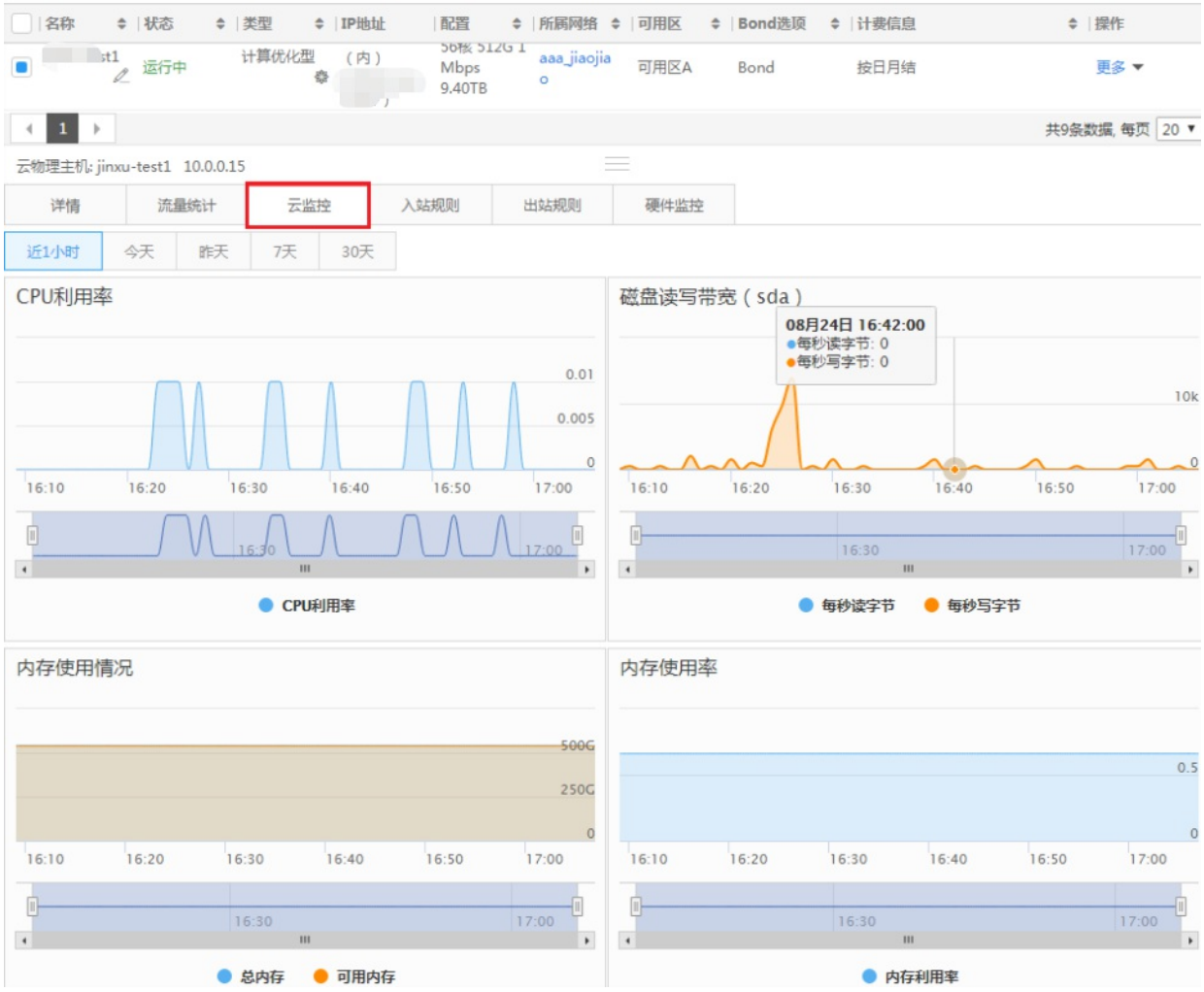
监控指标

监控指标(Metrics)	描述(Description)	单位
cpu.utilizition.total	CPU利用率	%
load.1min	CPU 1分钟平均负载(每核)	空
load.5min	CPU 5分钟平均负载(每核)	空
load.15min	CPU 15分钟平均负载(每核)	空
vm.memory.free	可用内存	B
vm.memory.size	总内存	B
vm.memory.util	内存利用率	%
disk.read.Bps[盘符]	磁盘读盘符	Bps
disk.read.ops[盘符]	磁盘每秒读次数盘符	Ops
disk.write.Bps[盘符]	磁盘写盘符	Bps
disk.write.ops[盘符]	磁盘每秒写次数盘符	Ops
vfs.fs.size[/]	磁盘使用率/	%
net.if.in_bps[bond0]	网卡进流量bond0	bps
net.if.out_bps[bond0]	网卡出流量bond0	bps
proc.num[]	运行进程个数	个

操作指南

点击您的裸金属服务器进入详情页，在详情页可以对您的机器状况进行监控，包括机器详情、流量统计、云监控、入站规则、出站规则和硬件监控。

如果您在创建裸金属服务器时选择了免费开通云监控功能，点击您的裸金属服务器详情页中的 “云监控” 按钮，就可以监控到30天内的CPU利用率、磁盘读写带宽、内存使用情况和内存使用率。



自助安装云监控

如果您在创建时未选择免费开通云监控功能，也可在需要时手动安装云监控服务。

安装步骤参考云监控：https://docs.ksyun.com/documents/6788

概述

NAT （Network Address Translation）网络地址转换是一种将虚拟私有网络中内网 IP 地址和公网 IP 地址进行转换的网关，能够让虚拟私有网络内无公网 IP 的云服务器或云物理主机访问Internet （但不支持 Internet 主动访问虚拟私有网络内的云服务器或云物理主机）。

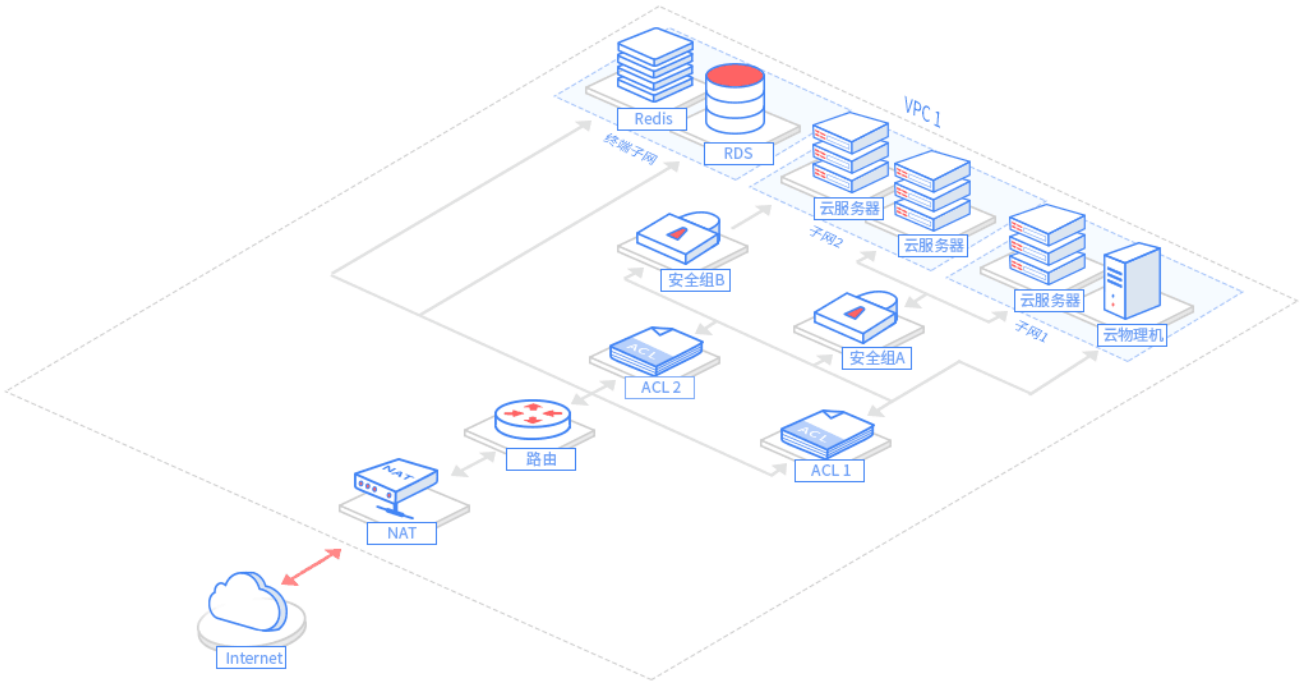
应用场景

金山云虚拟私有网络NAT网关的典型应用场景如下：

- **大出口、高可用Internet访问**：针对用户需要超大带宽、公网 IP 使用量大、部署服务较多的公网访问应用场景，金山云 NAT均可以满足需求。
- **安全的Internet访问**：金山云虚拟私有网络的 NAT提供 IP 的安全转换。如果用户希望隐藏虚拟私有网络内主机的公网 IP 以避免暴露其网络部署，同时又希望访问公网，那么使用金山云 NAT可以满足这类需求。

网络拓扑

如下图所示，NAT 网关是一个处于 Internet 和 VPC 边界的网关，并接在 VPC 的路由器上。由这样的拓扑图可知，VPC 内云物理主机等资源通过 NAT 网关向外发送数据包时，数据会先经过路由器，按照路由策略进行路由选择。然后 NAT 网关通过NAT IP地址作为源IP地址，将流量发送到 Internet：



配置类型

金山云NAT单IP最大支持15Gbps带宽，最大可支持20个IP，1亿以上并发连接数。

关键特性

- **SNAT**：源网络地址转换，用于 VPC 内的云服务器或云物理主机访问互联网。
- **高性能**：单IP可支撑最大15Gbps级别的转发能力。
- **高可用**：多机热备，单机出故障自动切换业务无感知。

NAT限速策略

原则上金山云会分配给用户与购买出流量带宽1：1的公网入带宽。但由于入流量带宽普遍比出流量带宽小，所以金山云在当前可用区整体入流量带宽低于出流量带宽时，会放开用户入流量带宽的限制，允许一定量的超出，增强用户体验。购买带宽小于50Mbps，入机房最大放开到50Mbps，购买带宽大于等于50Mbps，出机房和入机房仍 1:1 限速。

当金山云当前可用区整体入流量带宽大于出流量带宽时，会重新限制用户的入流量带宽，且优先限制入流量带宽与出流量带宽有极端差异的用户。

有关NAT网关的更多详细介绍，请参见NAT网关

概述

负载均衡（Server Load Balancing，简称SLB）是对云物理主机进行流量分发的网络服务设备。它可以通过流量分发，快速提高应用系统对外的服务能力；隐藏实际服务端口，增强内部系统的安全性；通过消除服务单点故障，提升应用系统的可靠性。

负载均衡服务通过设置虚拟服务地址（VIP），将位于同一地域的多台云物理主机资源虚拟成一个高性能、高可用的应用服务池；根据应用指定的方式，将来自客户端的网络请求分发到物理服务器池中。

负载均衡服务会检查云物理主机池中云物理主机实例的健康状态，自动隔离异常状态的实例，从而解决了云物理主机的单点问题，同时提高了应用的整体服务

能力。

关键特性

1.**高性能**。分布式集群满足大规模业务分发的性能要求
2.**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
3.**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
4.**安全**。免费的 5G DDoS攻击防御功能，无延时动态启动

使用场景

- 横向扩展应用系统服务能力，适用于各种 web server 和 app server
- 消除应用系统单点故障，当其中一部分物理服务器宕机后，应用系统仍能正常工作

协议支持

典型的 Web 应用程序之间的通信需要经由网络的各个分层，每层都会提供特定的通信功能。依据开放式系统互联（Open System Interconnect，OSI）网络模型，各个分层中都有标准的通信格式。金山云负载均衡涉及网络模型中的 四层（传输层） 和 七层（应用层）。

金山云负载均衡支持以下协议的请求转发：

- HTTP（应用层）
- HTTPS（应用层）
- TCP（传输层）

1.四层协议

如果使用四层协议转发，负载均衡实例会直接将请求转发到后端实例，而不修改任何数据包。负载均衡收到请求之后，会尝试在监听器配置中指定的端口上打开与后端实例的 TCP 连接。

2.七层协议

如果前端和后端连接均使用七层协议转发，负载均衡器会解析请求中有意义的应用层内容，并根据其内容选择后端物理服务器。因此，七层负载均衡器需要先代理后端服务器和客户端建立连接(三次握手)后，才可能接受到客户端发送的真正应用层内容的报文，然后再根据该报文中的特定字段，再加上负载均衡设备设置的服务器选择方式，决定最终选择的内部服务器。

均衡方式

均衡方式是指金山云负载均衡向后端物理服务器分配流量的算法：

1.加权轮询算法

以轮询方式依次将请求调度到不同服务器，用权值表示服务器的处理性能，按权值的高低和轮询方式分配请求，权值高的服务器先收到连接，权值高的服务器比权值低的服务器处理更多的连接，相同权值的服务器处理相同数目的连接数。

2.加权最小连接数算法

一种动态调度算法，它通过服务器当前活跃的连接数来估计服务器的负载情况。调度器需要记录各个服务器已建立连接的数目，当一个请求被调度到某台服务器，其连接数加一；当连接中止或超时，其连接数减一。同时根据服务器的不同处理能力，给每个服务器分配不同的权值，使其能够接受相应权值数的服务请求。

健康检查

金山云负载均衡实例定期向后端物理主机发送 ping、尝试连接或发送请求来测试后端物理主机运行的状况。

当后端物理主机实例被判定为不健康时，负载均衡实例将不会把请求转发到该实例上，但健康检查会对所有后端物理主机（不管是判定为健康的还是不健康的）进行，当不健康实例恢复正常状态时，负载均衡实例将恢复把新的请求转发给它。

1.四层协议

由负载均衡向配置中指定的物理机端口发起访问请求，如果端口访问正常则视为后端物理机运行正常，否则视为后端物理机运行异常。对于 TCP 业务，使用 SYN 包进行探测。

- 健康检查：开启
- 检查间隔： 2-300 秒，默认为5s
- 健康阀值：2-10 次，默认为5次（不健康后端物理机出现此指定次数响应超时后，视为健康）
- 不健康阀值：2-10 次，默认为4次 （健康后端物理机出现此指定次数响应超时后，视为不健康）

2.七层协议

由负载均衡器向后端物理机发送 HTTP 请求来检测后端服务，负载均衡器会通过 HTTP 返回值是否为预设的值来判断服务是否正常。

- 健康检查：开启
- 检查间隔： 2-300 秒，默认为5s
- 健康阀值：2-10 次，默认为5次（不健康后端物理机出现此指定次数响应超时后，视为健康）
- 不健康阀值：2-10 次，默认为4次 （健康后端物理机出现此指定次数响应超时后，视为不健康）

会话保持

可使得来自同一 IP 请求被转发到同一台后端物理机上；

1.四层协议

四层转发支持简单会话保持能力，可以设置会话保持的时间，超过该时间阀值，会话中无新请求则断开连接。

2.七层协议

HTTP/HTTPS 监听可使用植入 cookie 和重写 cookie 来进行会话保持。

Client IP 获取

- 4层负载均衡后端服务自动获取真实客户端IP
- 7层负载均衡支持通过HTTP Header X-forwarded-for 获取真实客户端IP

有关负载均衡的更多详细介绍，请参见负载均衡（SLB）

概述

VPC 对等连接是一种用于跨VPC网络数据同步的互联服务，打通对等连接的两个VPC之间就像同一个VPC网络一样。您可以实现同地域或跨地域的相同/不同账号的VPC互联，通过在两端配置路由策略，可以实现不同VPC的流量互通。对等连接不依赖某个独立硬件，因而不存在单点故障或带宽瓶颈。

有关对等连接的详细介绍，请参见对等连接

概述

弹性IP（Elastic IP，简称EIP）是与用户账户相关联的IP地址，可以绑定到用户的任何一台云服务器、云物理主机或负载均衡上；借助弹性IP地址，您可以快速的将地址重新映射到账户中的另一个云服务器、云物理主机或负载均衡上，从而屏蔽实例故障。金山云弹性IP拥有多种灵活的计费方式，可以满足各种业务场景的需求。

使用范围

弹性IP可以随时和云服务器、负载均衡进行关联，VPC环境下还可以和物理机、云服务器进行关联。弹性IP只能与在同一地域内的资源进行绑定，支持动态的绑定和解绑，您需要注意的是：

- 1个弹性IP同一时间只能绑定到1个资源上
- 1个资源同一时间只能绑定1个弹性IP
- 私有网络中一个 KEC 实例绑定了弹性IP，又处于关联了NAT网关的子网内，访问公网的数据包优先经过的是弹性IP

将弹性IP与资源绑定时，资源当前公网IP地址会释放到基础网络公网IP地址池。

如果将弹性IP与资源解绑时选择了重新分配公网IP，资源会很快自动分配到新的公网IP地址。此外，销毁资源也会断开与弹性IP的关联。

使用说明

弹性IP可以选择加入共享带宽，弹性IP加入共享带宽的方法可以查看 共享带宽 相关文档 加入同一共享带宽的每个弹性IP带宽上限与共享带宽的范围相同；但加入同一共享带宽的所有弹性IP在同一时间点带宽总和不得超过共享带宽的带宽上限，加入同一共享带宽的多个弹性IP在共享带宽的带宽达到上限时，会出现资源争抢现象。

弹性IP不通原因排查方法

弹性IP不通一般有如下原因：

1）弹性IP没有绑定到资源上，具体绑定方法见弹性IP产品使用文档

2）查看弹性IP绑定的资源内部是否有安全策略，如果有安全组策略，例如：禁止8080端口访问，那么弹性IP的8080端口也是无法访问的。