ZStack 的 VPC 特性详解及实战

今天我们详细的了解一下 ZStack 的 VPC 特性，本文比较长，目录如下：

在正文之前，我们先了解一下什么是 VPC 网络，VPC（Virtual Private Cloud）主要是一个网络层面的功能，是一块可我们自定义的网络空间，其目的是让我们可以

在云平台上构建出一个隔离的、自己能够管理配置和策略的虚拟网络环境，从而进一步提升我们在云环境中的资源的安全性。

我们可以在 VPC 环境中管理自己的子网结构，IP 地址范围和分配方式，网络的路由策略等。由于我们可以掌控并隔离 VPC 中的资源，因此对我们而言这就像是一个自己私有的云计算环境。

在 ZStack 支持三种基本网络架构模型：扁平网络、云路由网络、VPC。ZStack 的 VPC 是基于 VPC 路由器和 VPC 网络共同组成的自定义私有云网络环境，帮助企业用户构建一个逻辑隔离的私有云。VPC 具有灵活的网络配置、安全可靠的隔离、东西向网络流向优化等特点。

可以说，ZStack 的 VPC 以 VPC 路由器为核心，同时 VPC 网络作为 VPC 路由器的子网，使用 VPC 路由器提供各种网络服务。VPC 路由器/云路由器的创建都需要规格。规格带有 DNS 特性，定义了 VPC 路由器使用的 CPU，内存，镜像，网络等信息。其中路由镜像需要从官网下载，它封装了多种网络服务的定制化内容。

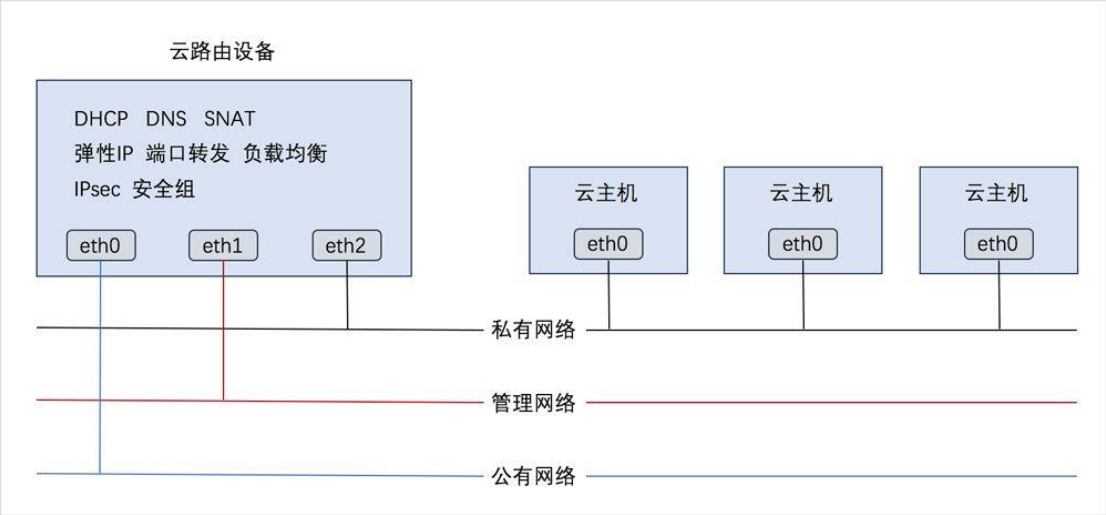
VPC 建立的大致过程如下：【云路由镜像--->云路由规格--->云路由/VPC 路由器--->创建若干 VPC 网络】，VPC 整体是由若干个子网构成，子网不可以跨可用区，一个可用区可以有多个子网，并且默认一个 VPC 的子网是互通的。

接下来正式介绍 ZStack 的 VPC 网络特性：它支持以下网络服务：DHCP、DNS、SNAT、弹性 IP、端口转发、负载均衡、IPsec 隧道、安全组等。

ZStack VPC 技术特性

说明：以下部分图摘自 ZStack 官方文档

特性一：二层隔离（vxlan/vlan）



一般来说网络隔离的话，二层算是基本能满足要求的，传统隔离是 VLAN 方式，有限 4096 个子网，但在大规模集群的今天，云计算一般采用 vxlan 的方式，子网可以多至 2^24 个，VXLAN 提供和 VLAN 相同的 2 层网络服务,但相比 VLAN 有更大的扩展性和灵活性。

ZStack 一个 VPC 路由器可以挂载多种不同的网络：公有网络（提供底层的云主机直接访问互联网的权利）、管理网络（管理节点和 vpc 路由器进行通信）、VPC 网络可

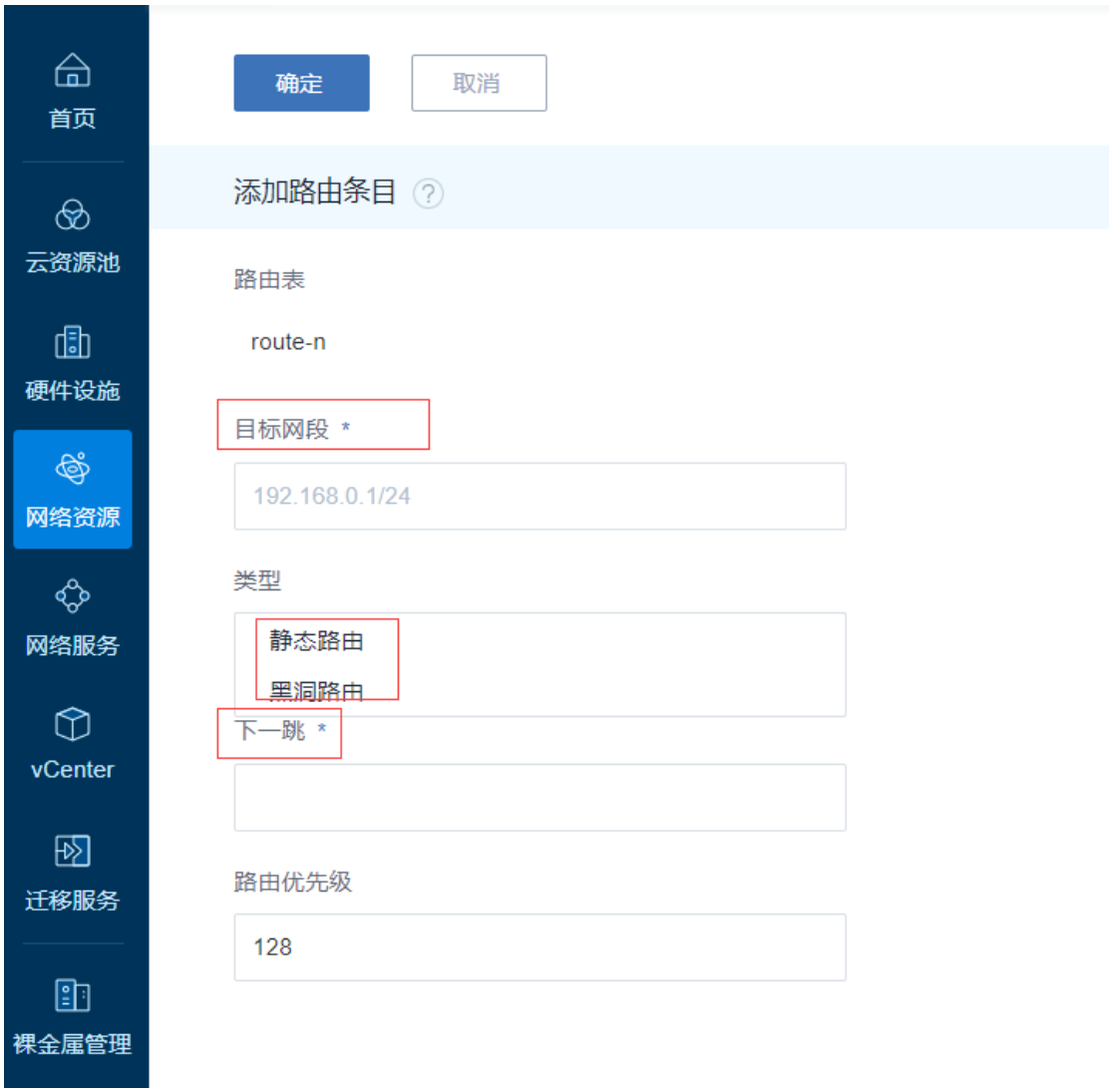
以分别对应多个 vxlan 的网络，就是说 ZStack 在 vxlan 的网络中把 underlay 网络标记成为 vxlan 的 network pool，把 overlay 网络标记成 vxlan 的 network（VxlanNetwork），这样的 underlay 网络可以加载到集群，在加载集群的时候会去配置匹配 vtep 的 ip 地址段，这样 ZStack 在加载集群的时候，通过这个 ip 地址段会去自动寻找合适的 vtep 的 ip 地址进行 vxlan 的 underlay 配置。

说明：上图中三个 vxlan 的网络中主机二层隔离，三层中通过加载到 VPC 路由器的网关实现互通。

特性二：基于源的安全组

一般 web 对外提供服务时，ZStack 可以配置基于 web 服务访问的应用服务，在这里可以提前设置安全组来保护部分主机，即使 web 服务主机被攻击，但仍有被保护的主机来撑起服务，保证整个项目的安全性，比如在数据库中，可设置安全组约束：只允许应用服务的部分主机通过设置的安全组访问数据库，所以就算攻击即使到了应用服务也到不了数据库，保证了服务的安全。

特性三：自定义路由

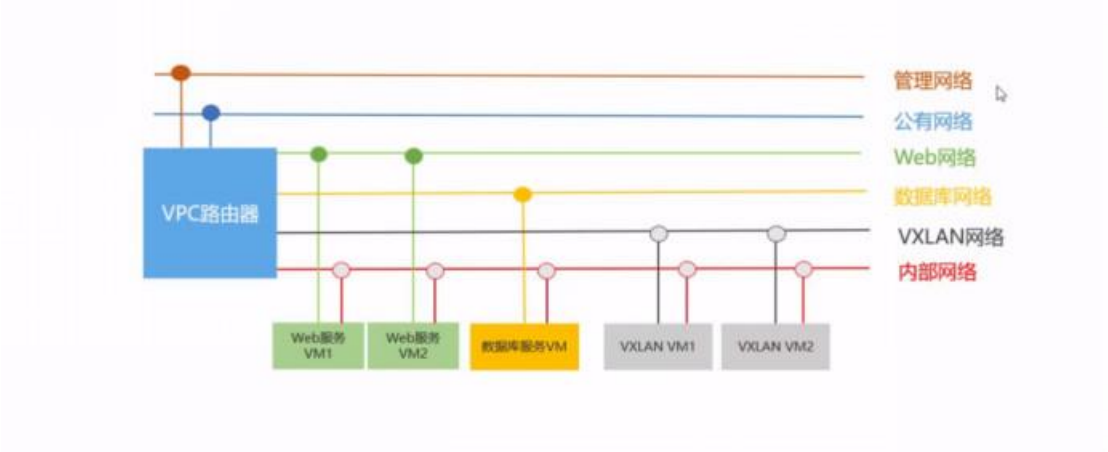


可以设置目的网段地址和下一跳路径，确保网络的正确访问。

特性四：自定义黑洞路由

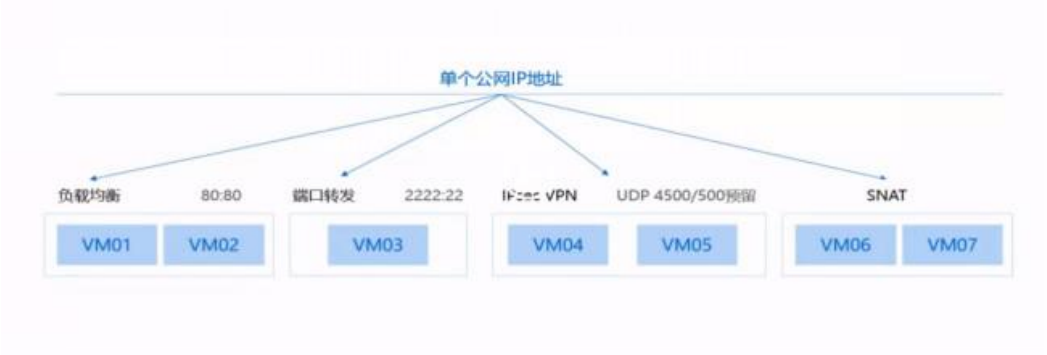
避免路由器陷入环路时，可将一些无关的路由吸入其中，让他们有来无回，丢弃匹配的数据包。

特性五：灵活子网搭配

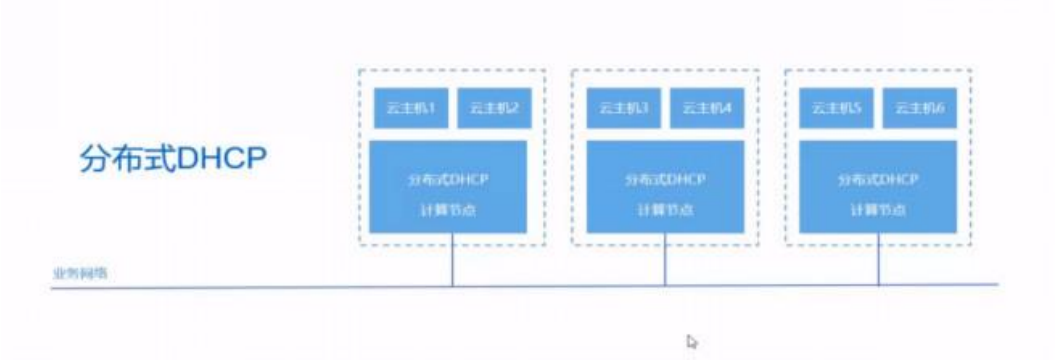


特性六：多网络服务复用公网 ip 地址

可以将单个公网 ip 地址做到多种服务特性，注意端口转发地址不能重叠，可以实现在同一 ip 的不同的端口进行不同的服务

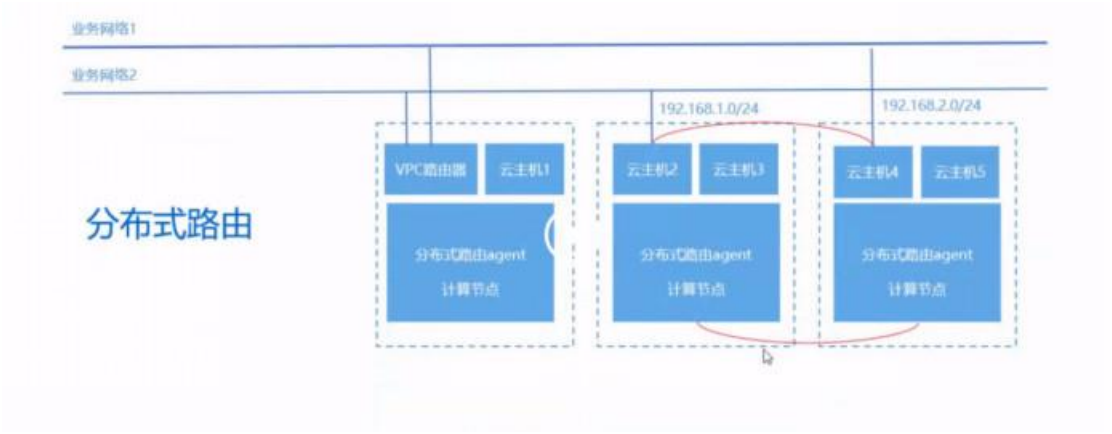


特性七：分布式 DHCP



ZStack 为了提高性能，采取了这种分布式的 DHCP，更高效的自动分配 IP 地址。

特性八：分布式路由



ZStack 可以做到东西向流量的一个优化。比如主机 2 和 4，网络段不同，正常情况下通信需要 vpc 路由器的网关进行通信，但 ZStac 的优化可以做到跨网段通信可以不走 vpc 的路由器，即在每个计算几点维护了一个分布式路由的一个代理（agent）代理去做分布式路由的一个流量优化。

特点：优先数据的通信，不采用消息队列，ZStack 研发了一套协议，直接在 ip 地址层传递消息，（一般的分布式路由会采用网关欺骗去实现，ZStack 没有采用这些，这样会存在真网关和假网关的判断问题）而且 ZStack 的分布式路由没有集中式的控制器，流量优化通过云路由的 agent 发起，每个 agent 只关心自己在云路由上面的一个情况，所以不存在系统单点的问题，类似旁挂的机制，即使旁边的计算节点 agent 挂掉，也不会网络失效，而是已有的，之前的流量走传统的，集中式的流量再退回到集中式的路径上面去，再走 vpc 路由器。

在 ZStack 中打开分布式路由功能（默认关闭），系统会尝试优化东西向网络流量，以提高吞吐量和降低网络延迟。分布式路由功能还可加强云主机之间通信的可靠性，内网跨三层流量不会因为云路由故障而失效。

特性九：公网 IP Qos

QoS(Quality of Service，服务质量)指一个网络能够利用各种基础技术，为指定的网络通信提供更好的服务能力，是网络的一种安全机制， 是用来解决网络延迟和阻塞等问题的一种技术。在 ZStack 中保证网络服务的能力，公有网络一般都是配有 QoS 的，为了内部保障业务能力，也会去配置内部的 QoS，ZStack 支持对某个虚拟 ip 的某个端口进行配置 QoS。

至此，我们对于 ZStack 的 vpc 网络特性，应该有了大致的了解，总结一下：

1. **灵活网络配置：**每个 vpc 网络可以自定义对立的网络段和网关，灵活加载和卸载 vpc，并配置路由表和路由条目
2. **安全可靠的隔离，**不同账号下 vpc 不通，互不影响
3. **在同一个 vpc 路由下，多子网互通。**
4. **网络流量的优化：**分布式路由优化了东西向的流量，有效的降低网络延时，不走集中式的路径，不会把网络瓶颈集中到 vpc 路由器上，不走 vpc 路由的直接通信

以上简单介绍了 ZStack 的 VPC 特性，但文字始终不如实践来的深刻，尤其对于我们刚接触云计算网络的程序员们来说，接下来我们通过 ZStack 平台来看看它们到底是怎么实现的，以及操作流程，相信实践过后，每个人都会有不同的体会。

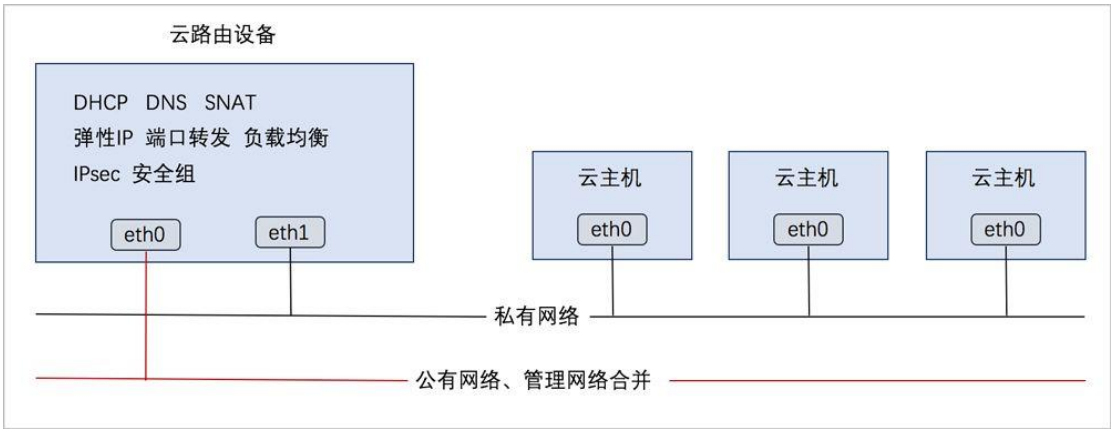
VPC 部分特性实战

在云路由网络拓扑中：云路由主要涉及以下 3 个基本网络：

- **公有网络：**用于提供弹性 IP、端口转发、负载均衡、IPsec 隧道等网络服务需要提供虚拟 IP 的网络，公有网络一般要求可直接接入互联网。
- **管理网络：**用于管理控制对应的物理资源，例如物理机、镜像服务器、主存储等需提供 IP 进行访问的资源时使用的网络。
- **私有网络：**也称之为业务网络或接入网络，是云主机使用的内部网络。包含 VPC

本次实践的部署方式：

- 公有网络和管理网络合并，私有网络独立部署

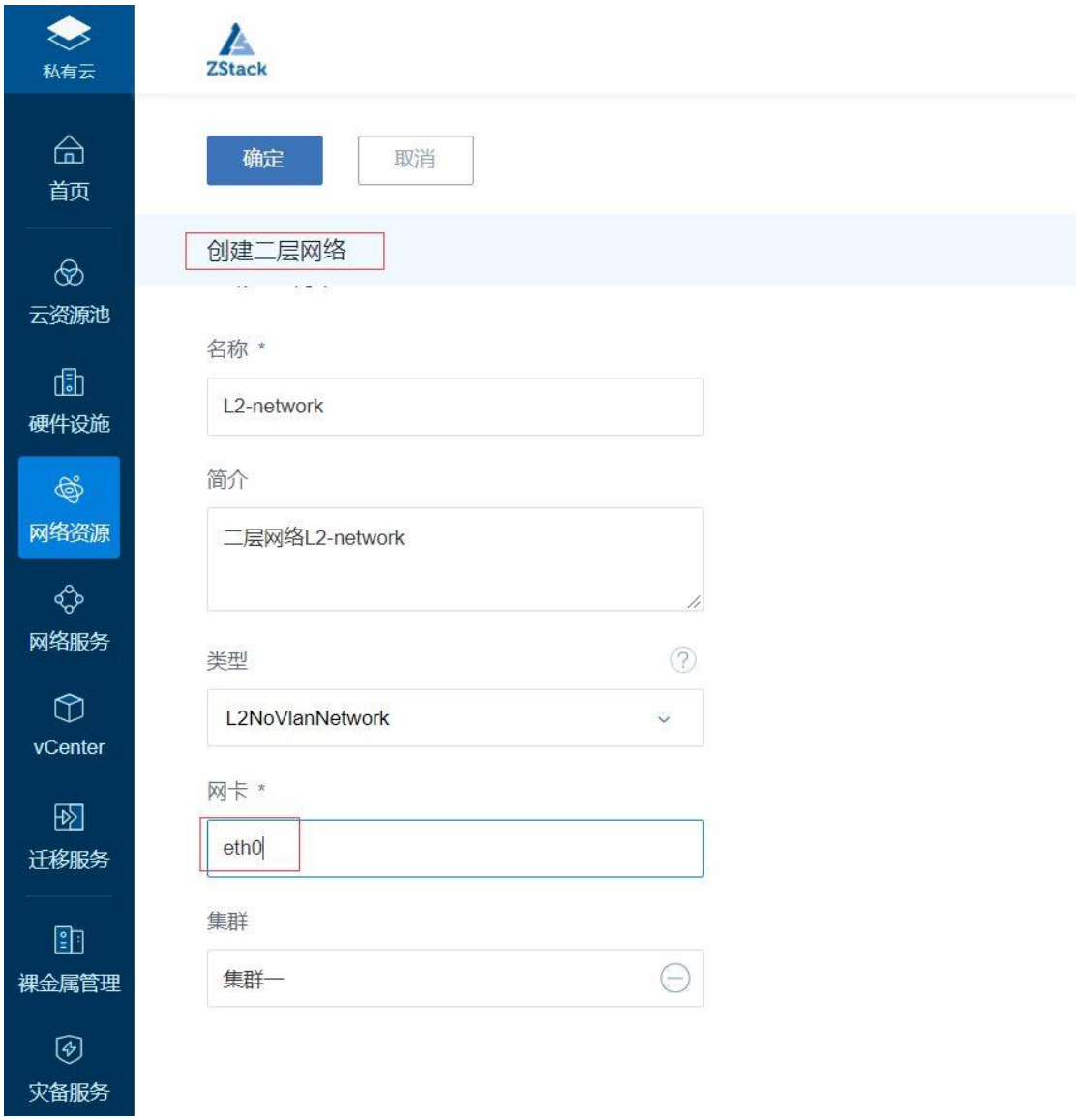


ZStack VPC 网络创建过程

一、公有网络的部署

专有网络 VPC 的基本部署流程如下：

1．创建二层公有网络，添加物理网卡名，并加载此二层网络到相应集群。要确保二层网络加载到有物理机的集群，ZStack 有区域，集群，二层网络、主存储等资源，这里的二层网络要在集群上进行加载。



我们可以在宿主机的命令行输入 brctl show 显示目前的桥接接口：

```
]# brctl show
bridge name     bridge id               STP enabled     interfaces
br_eth0         8000.fa2569294000       no              eth0
```

可以看到已经创建完成。

2.创建三层公有网络，注意网关不可以在 ip 段中，这里需要解释一下：

网关实质上是一个网络通向其他网络的 IP 地址。比如有网络 A 和网络 B，网络 A 的 IP 地址范围为“192.168.1.1 - 192. 168.1.254”，子网掩码为 255.255.255.0；网络 B 的 IP 地址范围为“192.168.2.1 - 192.168.2.254”，子网掩码为 255.255.255.0。

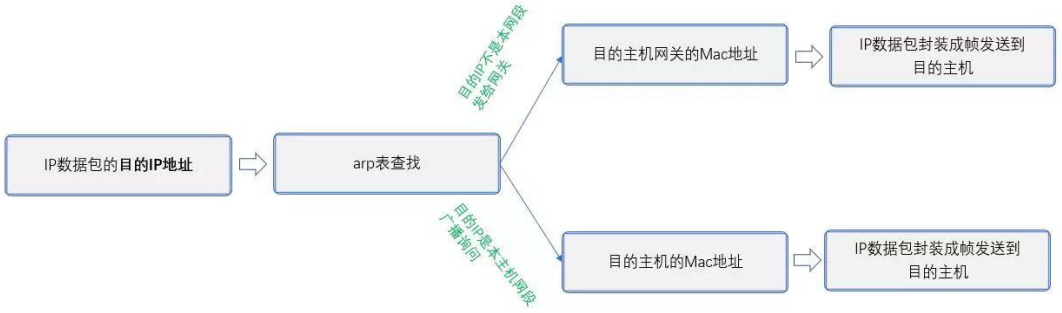
在没有路由器的情况下，两个网络之间是不能进行 TCP/IP 通信的，即使是两个网络连接在同一台交换机(或集线器)上，TCP/IP 协议也会根据子网掩码(255.255.255.0)判定两个网络中的主机处在不同的网络里。而要实现这两个网络之间的通信，则必须通过网关。如果网络 A 中的主机发现数据包的目的主机不在本地网络中，就把数据包转发给它自己的网关，再由网关转发给网络 B 的网关，网络 B 的网关再转发给网络 B 的某个主机。网络 B 向网络 A 转发数据包的过程。 所以说，只有设置好网关的 IP 地址，TCP/IP 协议才能实现不同网络之间的相互通信。下图简单示范了一下，中间省去了其他步骤。

那么网关不在 ip 段中可以通信吗?

不同网络之间的相互通信需要 ARP 寻址，这时候，请求设备需要查看自己的路由表，没有的话，只能将这个 ARP 请求先发送到路由表中指定的路由上再说。至于这个路由如何找到被请求的设备 B，可能需要下一个路由再次转发，不断的传播寻找，这其

中牵涉到路由传播算法，直到寻找到目的路由为止，路由收到数据包后，会进行拆包，得到目的设备的 IP 地址信息，然后查询自己的路由表。在表中的话返回目的主机 Mac 信息给源主机。

整个过程可以简单理解为：（不代表全部）



即源主机先发送 ARP 请求帧以获取网关 IP 地址对应的 MAC 地址，再由网关 IP 地址对应的 MAC 地址经过一系列操作得到目的 MAC 地址。

所以获取目标设备的 MAC 地址时使用的是二层广播，和 IP 地址是否为同一个网段并没有任何关系，即网关不在 ip 段中可以通信。

创建三层公有网络过程如下图：

确定 取消

创建公有网络

名称 *

L3-public

简介

二层网络 *

L2-network

关闭DHCP服务

添加网络段

方法

IP 范围 CIDR

起始IP *

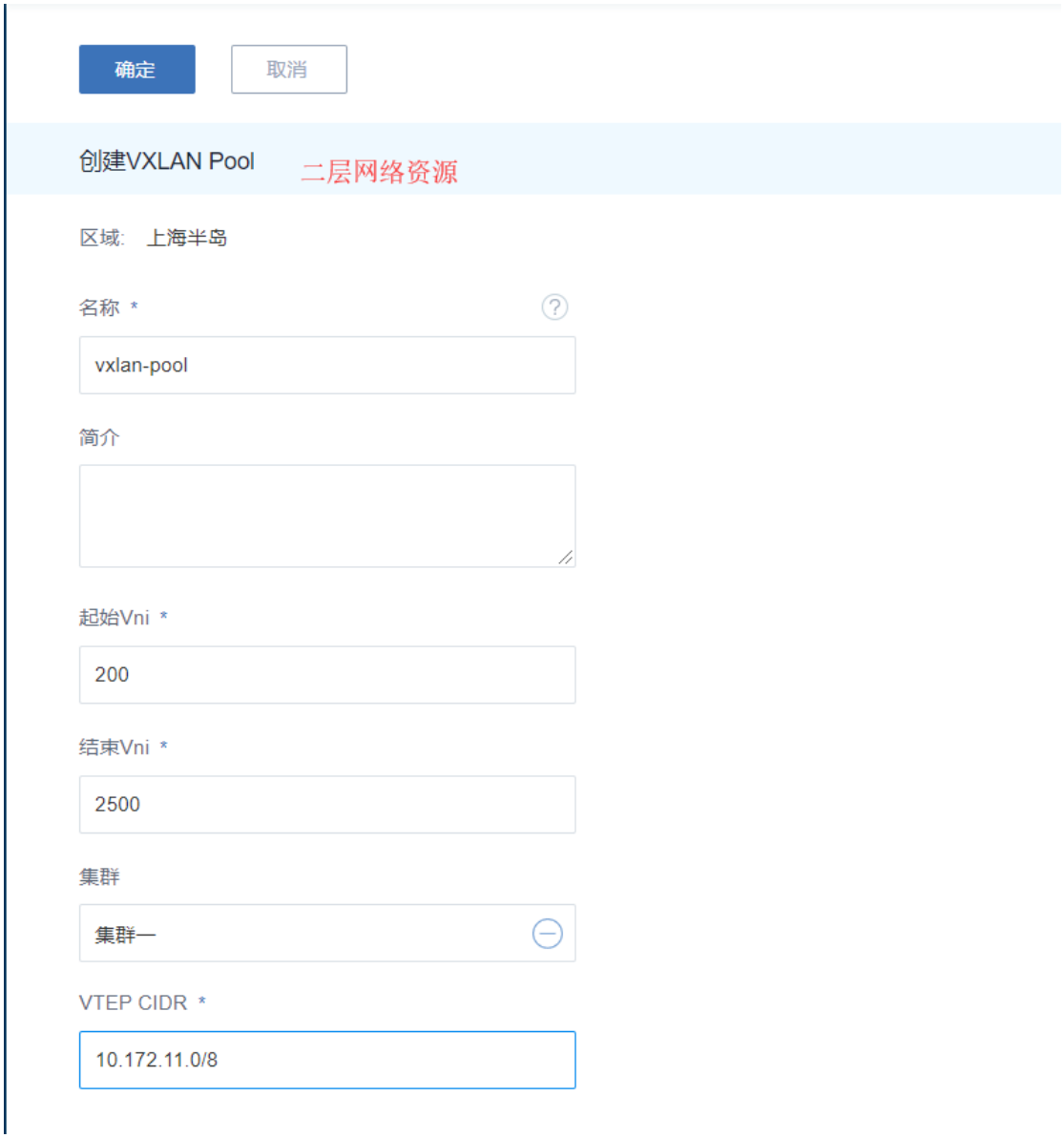
二、VPC 网络部署

VXLAN Pool 和 VxlanNetwork 共同提供了 VxlanNetwork 类型的配置，在使用 VxlanNetwork 前，需要先建立 VXLAN Pool。创建完毕 VXLAN Pool 后，可指定或随机选择 Vni 来创建 VxlanNetwork。创建二层网络资源：vxlan pool，要和集群网络段相配，这里的 VTEP （VXLAN 隧道终端（VXLAN Tunneling End Point)，用于

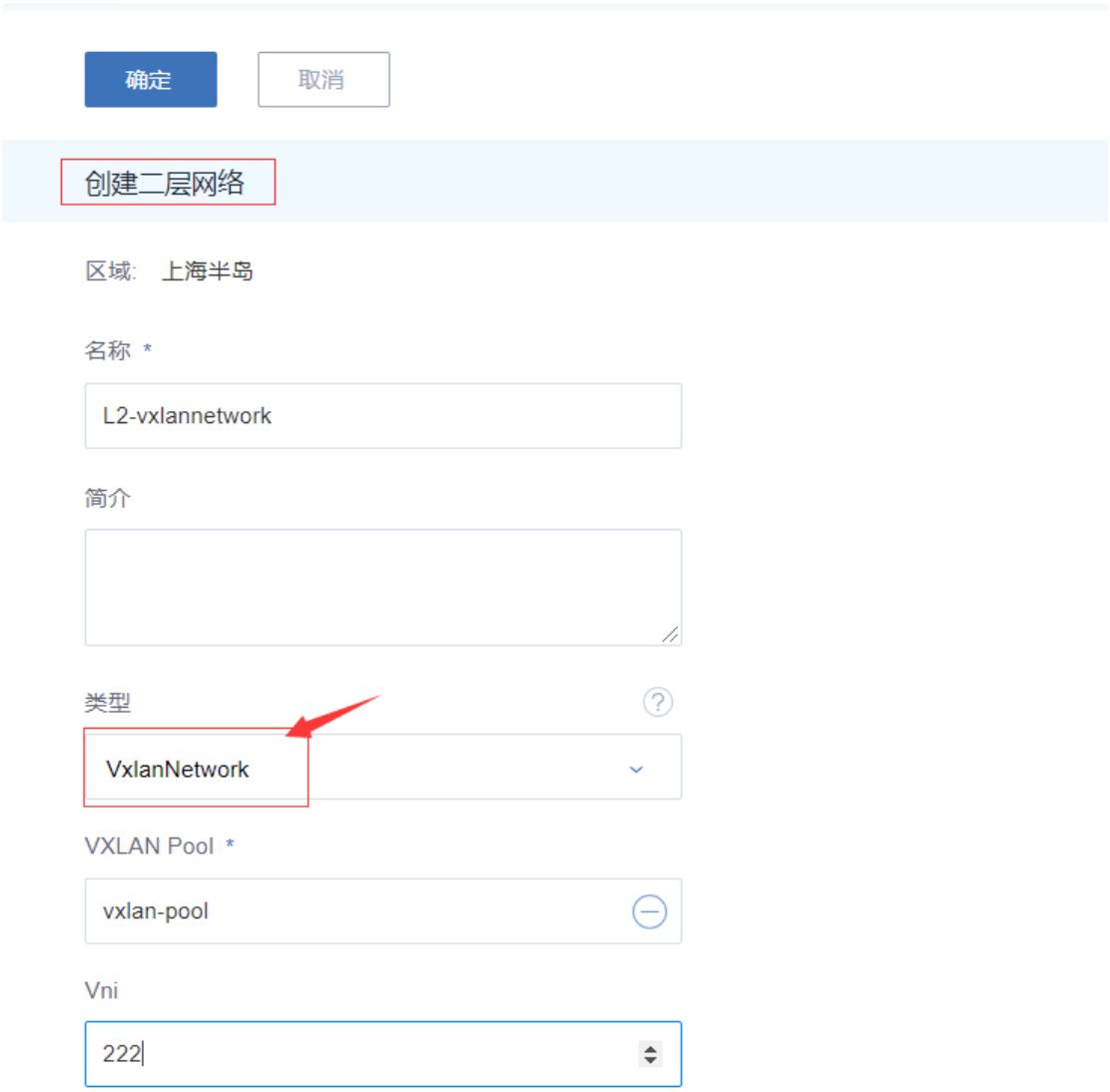
多 VXLAN 报文进行封装/解封装，包括 mac 请求报文和正常 VXLAN 数据报文）。VTEP 的 CIDR 用于让 ZStack 选择正确的 underlay 网络，通常为管理网络、或专门的 underlay 网络、或公有网络。

CIDR（Classless Inter-Domain Routing，无类域间路由选择）它消除了传统的 A 类、B 类和 C 类地址以及划分子网的概念，因而可以更加有效地分配 IPv4 的地址空间。它可以将好几个 IP 网络结合在一起，使用一种无类别的域际路由选择算法，使它们合并成一条路由从而较少路由表中的路由条目减轻 Internet 路由器的负担，简单来说就是将网络地址一致进行 CIDR 汇总。

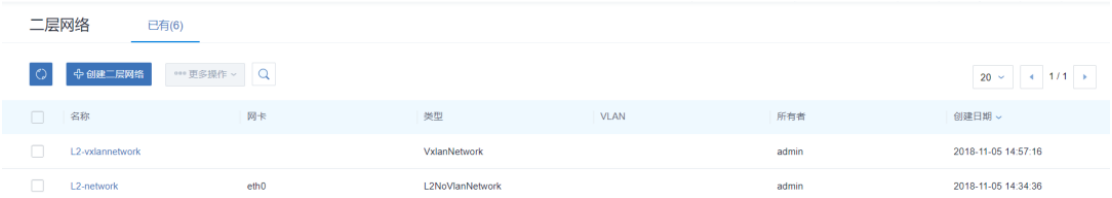
这里，我们在 ZStack 平台上，进行如下配置：



创建二层 vxlan 网络：L2-vxlannetwork



目前二层网络资源如下：二个二层网络 + 一个 vxlan-pool

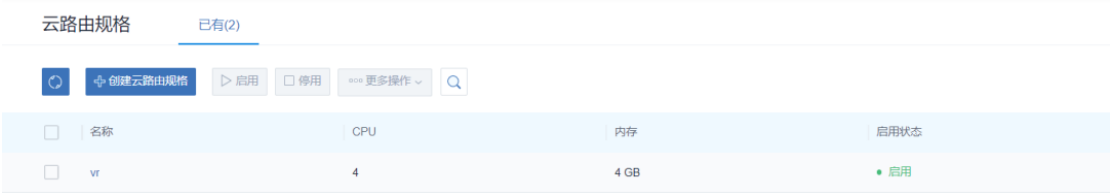


1. 添加云路由镜像。

可从 ZStack 官网下载，下载时需要填写信息，邮件接收下载地址。

2. 创建云路由规格。

其中管理网络与公用网络可以共同使用之前创建的三层公有网络 L3-public。



3. 基于刚才的云路由规格 4c4g 创建 VPC 路由器。

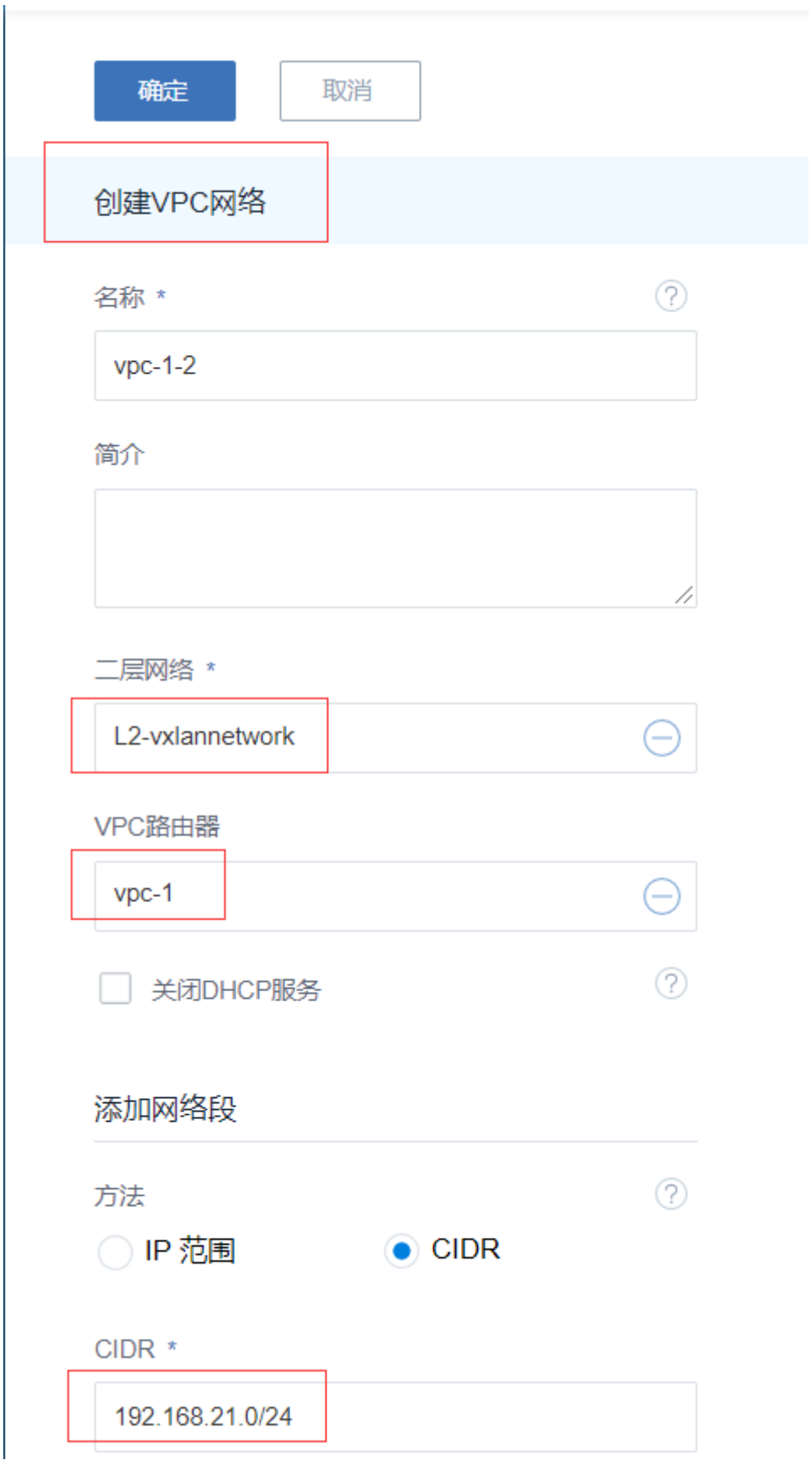
VPC 路由器可提供各种网络服务。本过程创建较慢，请耐心等待。。。

4. 创建二层私有网络

本次实践用于创建三层的 VPC 网络，并加载此二层网络到相应集群。当前集群不能为空，要有宿主物理机。分别创建二个 VPC 网络，新建的网络段不可与 VPC 路由器内任一网络的网络段重叠。即 VPC 路由器下所有 VPC 网络（子网）的网络段不可重叠。

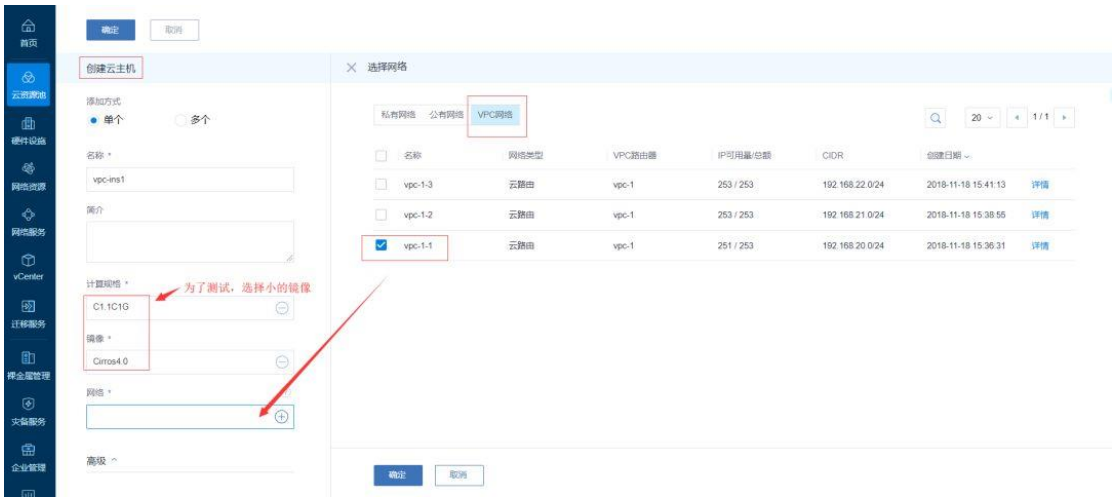
选择二层网络和 vpc 路由器


确定
取消
创建VPC网络
名称 *
vpc-1-1
简介
二层网络 *
L2-vxlannetwork
VPC路由器
vpc-1
关闭DHCP服务
添加网络段
方法
IP 范围
CIDR
CIDR *
192.168.20.0/24

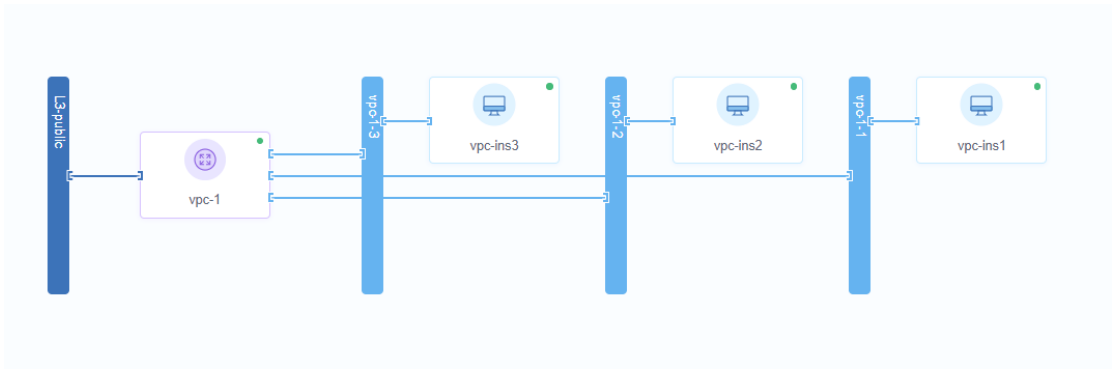

确定
取消
创建VPC网络
名称 *
vpc-1-2
简介
二层网络 *
L2-vxlannetwork
VPC路由器
vpc-1
关闭DHCP服务
添加网络段
方法
IP 范围
CIDR
CIDR *
192.168.21.0/24

5.上面步骤指定 VPC 路由器，创建三个 VPC 网络。

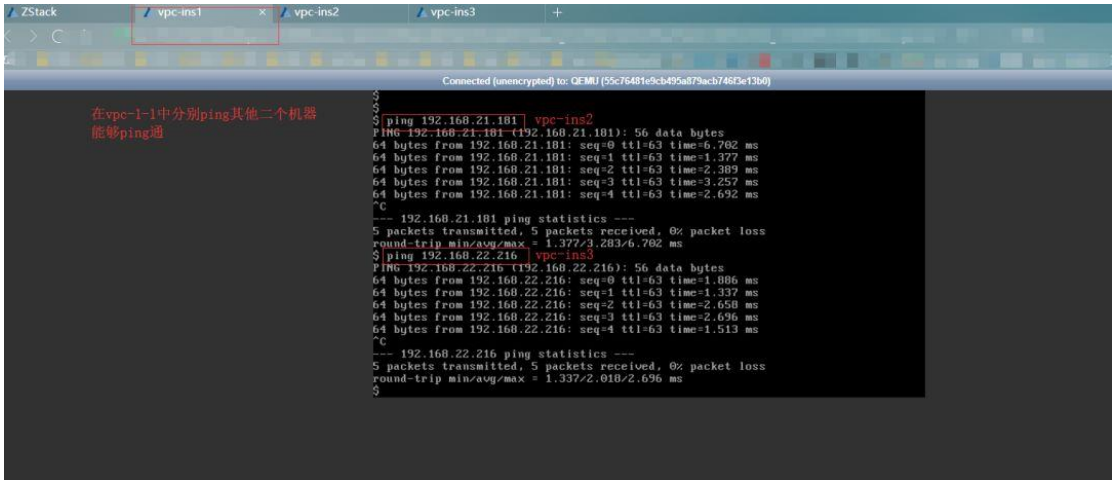
6.使用 VPC 网络创建云主机。创建三台如图 vpc-ins1，vpc-ins2，vpc-ins3，网络分别选取 vpc 网络中的 vpc-1-1、vpc-1-2、vpc-1-3



网络拓扑如下：



创建完成后，打开一台主机的命令行，去 ping 另外一台主机的 ip，证明他们是可以互通的，虽然处于不同的网段，但是他们通过云路由三层连接，达到互通。



可知，3 个 vpc 私有网络分别对应三个云主机，则三个的网络段都不同，在二层是完全隔离的网络，但在三层上来看，它们的网关是放在 vpc 路由器上面的。则它们通过网关实现互联互通。

这种机制实现了底层网络的隔离，比如创建多个 vpc 路由器，再连接不同的 vpc 网络，分别连接不同的主机，这些不同的 vpc 路由器下的主机是不同的，同一个 vpc 路由器下的不同 vpc 网络段中主机也是不通的，实现二层隔离，实现了租户的隔离，在三层上面，同一个路由器下面，不同的子网又可以实现互通，做到了一种灵活的组网与搭配。

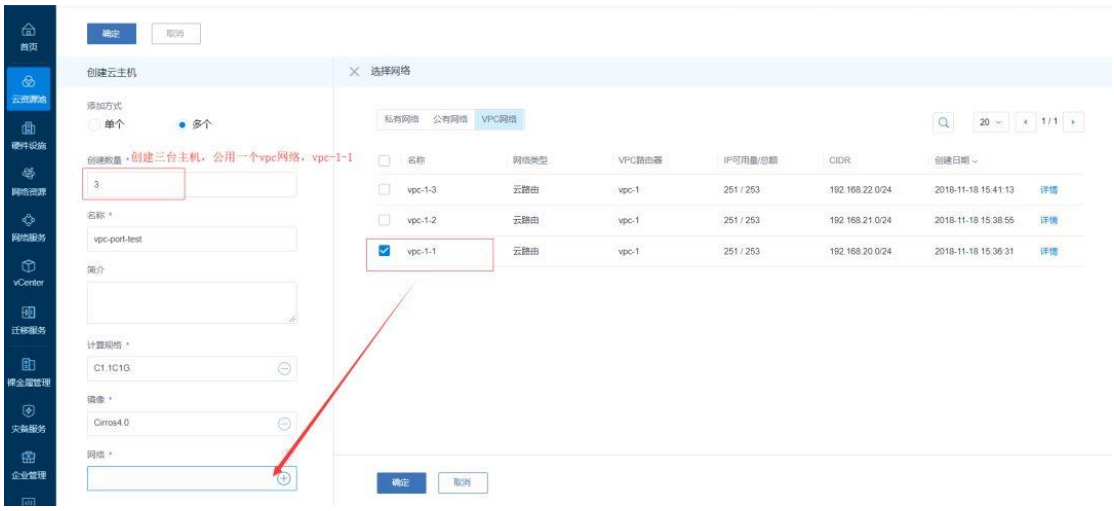
同理分别通过创建 admin 和 admin1 二个账户，各创建二个主机，在同一个账户下的主机能够互通，在不同账号下的主机不能互通。

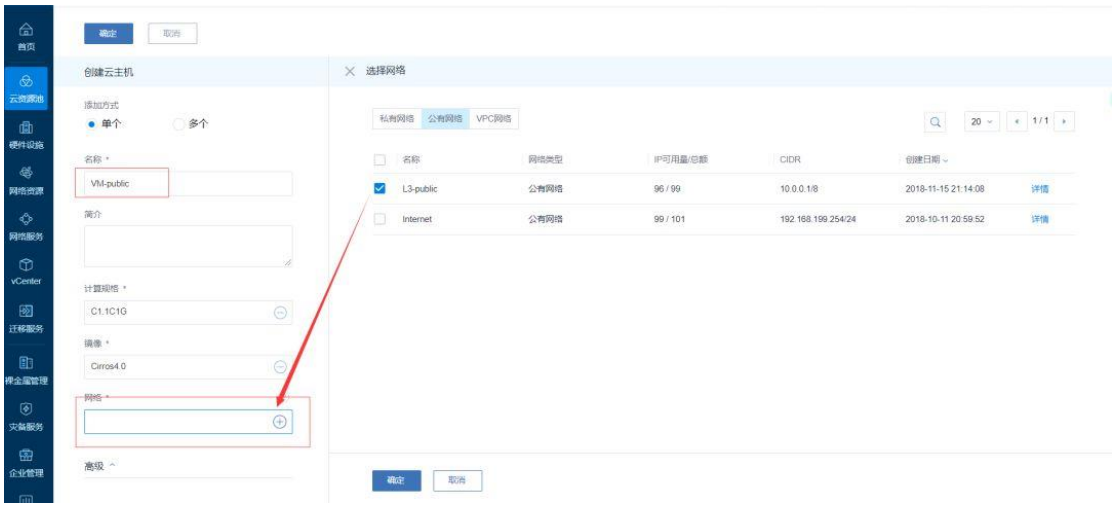
三、端口转发测试

官方解释：端口转发（PF）是基于云路由器/VPC 路由器提供的三层转发服务，可将指定公有网络的 IP 地址端口流量转发到云主机对应协议的端口。在公网 IP 地址紧缺的情况下，通过端口转发可提供多个云主机对外服务，节省公网 IP 地址资源。

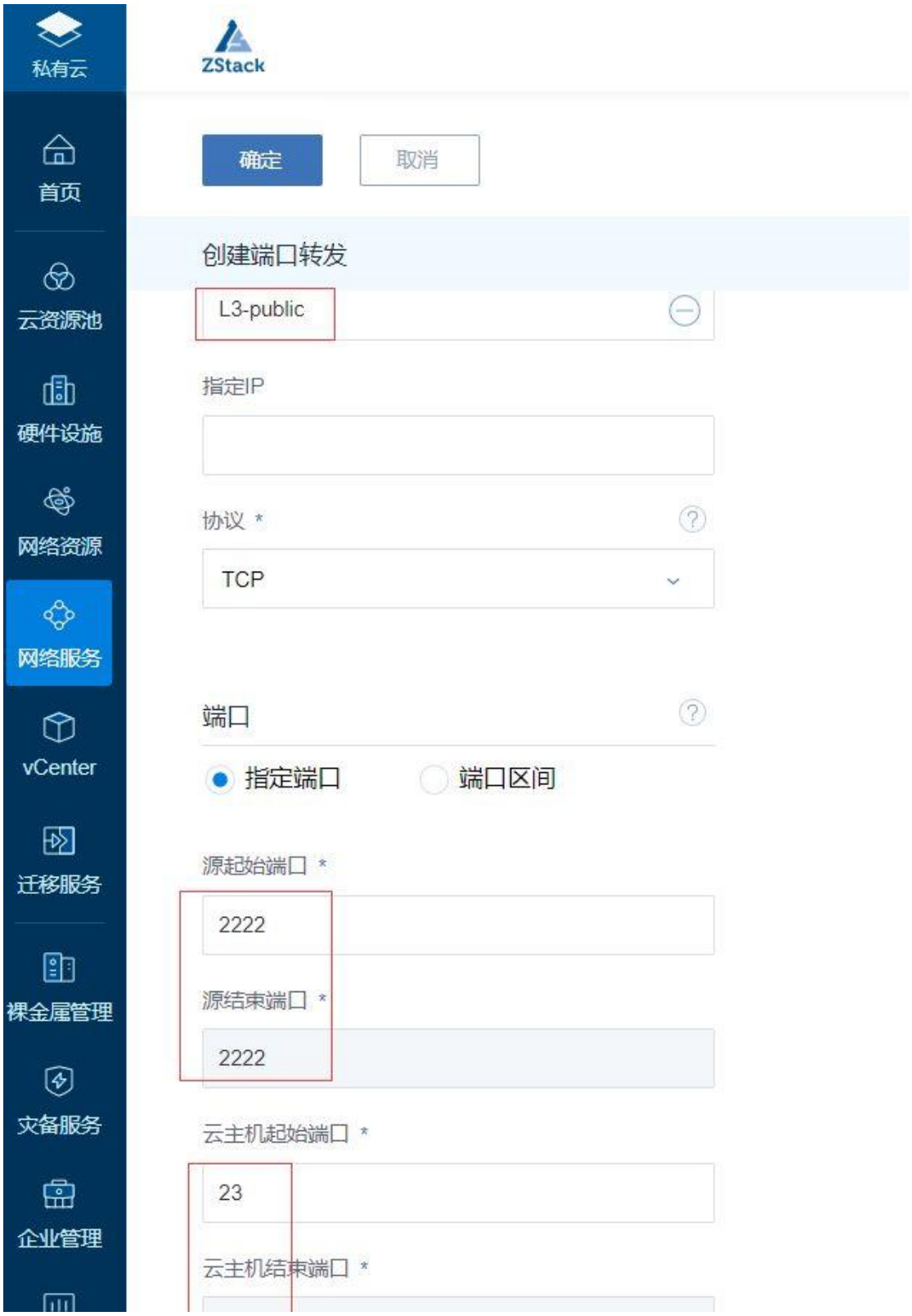
端口转发服务限于云路由器/VPC 路由器提供。端口转发规则创建于云路由器/VPC 路由器公有网络和云主机私有网络之间。

首先创建 VPC 网络主机三台，再用公有网络创建一台，注意这里选择的镜像需要有 iptables 规则。

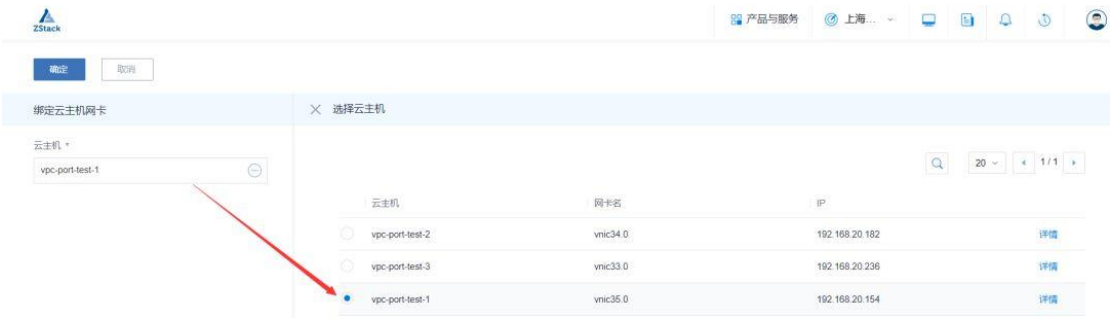




在网络服务选项中，点击端口转发，输入名称，选择新建立虚拟 ip，选择公有网络


私有云
首页
云资源池
硬件设施
网络资源
网络服务
vCenter
迁移服务
裸金属管理
灾备服务
企业管理
ZStack
确定
取消
创建端口转发
L3-public
指定IP
协议 *
TCP
端口
指定端口
端口区间
源起始端口 *
2222
源结束端口 *
2222
云主机起始端口 *
23
云主机结束端口 *

同时绑定云主机网卡



同样的，再根据第一个示例，创建二个端口转发，这是需要选择已有的虚拟 ip，即第一次创建的，源起始端口分别为 2223 和 2224，云主机起始端口为 23，再绑定云主机网卡。最后三个端口转发配置如下：

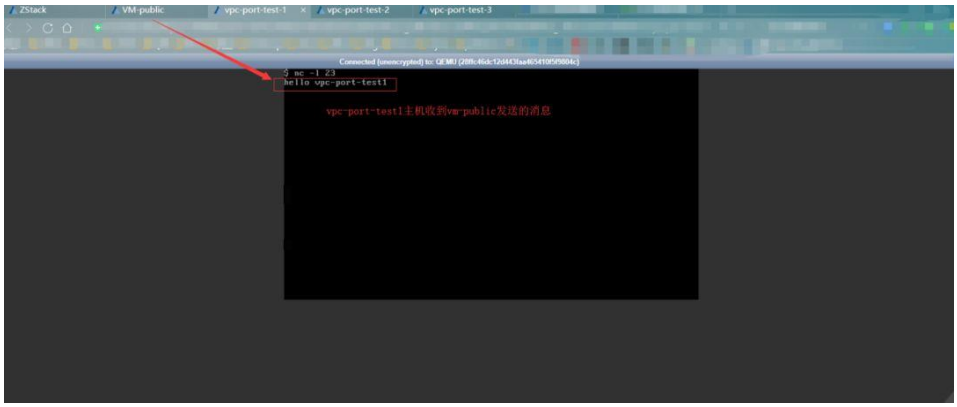
打开四个云主机，使用 iptables 命令清除相关规则，防止端口占用，注意执行 iptables -F 命令会清除所有规则，需要小心。

VPC 网络下的三个云主机均执行 nc -l 23 来监听 23 号端口，公有网络下的云主机，使用 nc 端口转发 ip+端口（2222/2223/2224），来给云主机发送消息，测试对端云主机是否能接收到消息。

如：在公有网络下的云主机输入 nc 10.172.11.93 2222,接下来发送消息 hello vpc-port-test1，在 vpc 网络中的对应 2222 端口的主机 vpc-port-test1 收到这个消息。

公有网络下的云主机输入 nc 10.172.11.93 2223，发送消息 hello vpc-port-test2，在 vpc 网络中的对应 2223 端口的主机 vpc-port-test2 收到这个消息。

公有网络下的云主机输入 nc 10.172.11.93 2224，发送消息 hello vpc-port-test3，在 vpc 网络中的对应 2224 端口的主机 vpc-port-test3 收到这个消息。

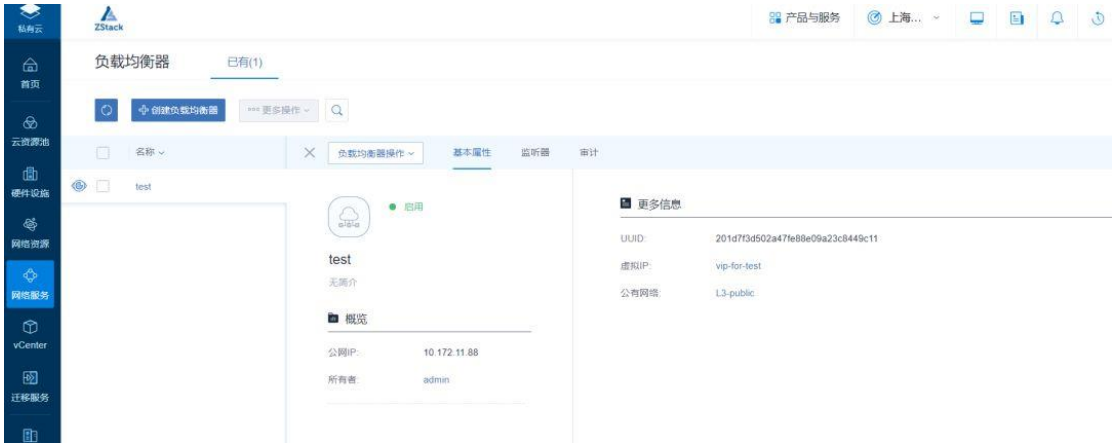


四、负载均衡

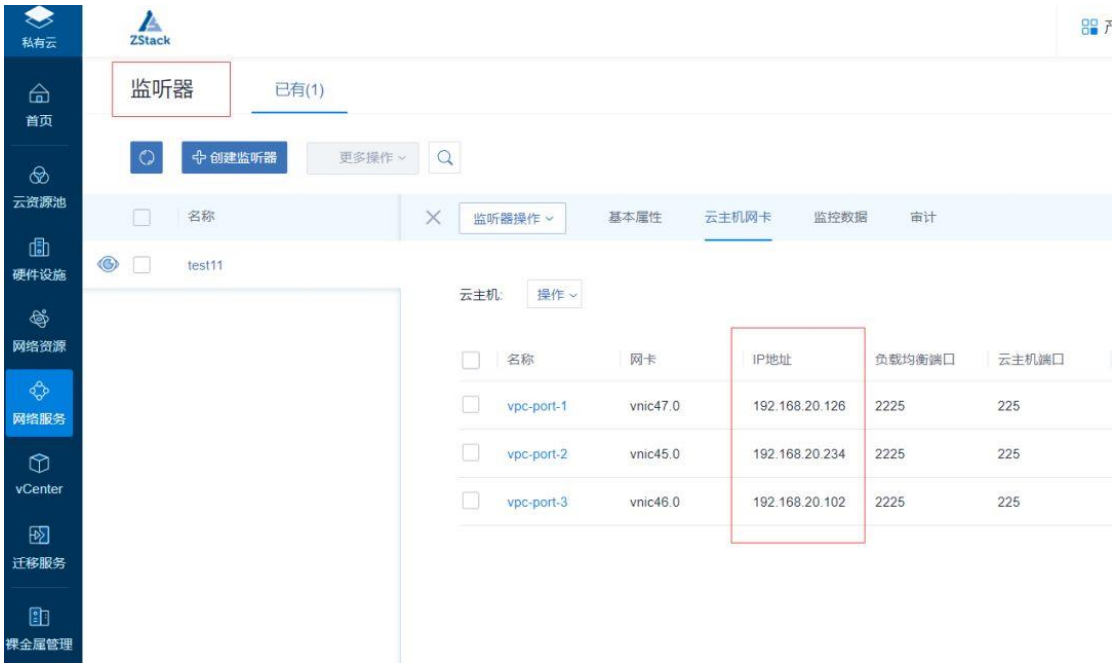
负载均衡（LB）：将公网地址的访问流量分发到一组后端的云主机， 并支持自动检测并隔离不可用的云主机，从而提高业务的服务能力和可用性。

- 负载均衡自动把访问用户应用的流量分发到预先设置的多个后端云主机，以提供高并发高可靠的访问服务。
- 根据实际情况，动态调整负载均衡监听器中的云主机来调整服务能力，且不会影响业务的正常访问。

在 ZStack 私有云主菜单，点击网络服务中的负载均衡器，创建时选择公有网络，创建如下：



在网络服务界面，点击创建监听器，创建过程中选择我们刚才创建的负载均衡器，端口选择可以如下图，之后在监听器操作中选择绑定云主机网卡，完成如下：



这时候打开四个云主机，模仿上面的端口转发过程，向负载均衡器公网 IP 的端口发送三条信息，这时候同一 vpc 网络下的三台主机会采取轮询的方式各接受一条信息。完成了将公网地址的访问流量分发到一组后端的云主机的操作，减轻了业务的压力。

至此，关于 ZStack 的 vpc 特性，相信大家应该有了初步的认识，本次关于其他特性没去一一实现，有兴趣的可以下载搭建 ZStack 平台后自己去测试下，除了 VPC 外还

有其他更多的功能，比如：互连、灾备、服务、一键迁云等，需要我们一个个地去体验。

ZStack 是谁?

大道至简·极速部署，ZStack 致力于产品化私有云和混合云。

ZStack 是新一代创新开源的云计算 IaaS 软件，由英特尔、微软、CloudStack 等世界上最早一批虚拟化工程师创建，拥有 KVM、Xen、Hyper-V 等成熟的技术背景。

ZStack 创新提出了云计算 4S 理念，即 Simple（简单）、Strong（健壮）、Smart（智能）、Scalable（弹性），通过全异步架构，无状态服务架构，无锁架构等核心技术，完美解决云计算执行效率低，系统不稳定，不能支撑高并发等问题，实现 HA 和轻量化管理。

ZStack 发起并维护着国内最大的自主开源 IaaS 社区——zstack.io，吸引了 6000 多名社区用户，对外公开的 API 超过 1000 个。基于这 1000 多个 API，用户可以自由组装出自己的私有云、混合云，甚至利用 ZStack 搭建公有云对外提供服务。

ZStack 拥有充足的知识产权储备，积极申报多项软著和专利，参与业内标准、白皮书的撰写，入选云计算行业方案目录，还通过了工信部云服务能力认证和信通院可信云认证。

ZStack 面向企业用户提供基于 IaaS 的私有云和混合云，是业内唯一一家实现产品化，并领先业内首家推出同时打通数据面和控制面无缝混合云的云服务商。选择 ZStack，用户可以官网直接下载、1 台 PC 也可上云、30 分钟完成从裸机的安装部署。

目前已有 1000 多家企业用户选择了 ZStack 云平台。