### PVE虚拟化使用手册

**前言：**

1.蓝色字体表示命令行命令，正式执行时不要复制前面的#号，#号只是提示应该使用root权限操作

2.绿色字体表示注释，有时注释太多就不用绿色表示了

3.注意：本文档的所有操作请先在测环境进行实践，请不要直接在真实的服务器中操作！

**版权声明**：

本文档以开源的形式发布，所有条款如下：

（1）无担保：作者不保证文档内容的准确无误，亦不承担由于使用此文档所导致的任何后果

（2）自由使用：任何人可以出于任何目的而自由地 阅读/链接/打印/转载/引用/分发/再创作 此文档，无需任何附加条件

若您 阅读/链接/打印/转载/引用/分发/再创作 本文档，则说明接受以上2个条款。

作者：李茂福

更新日期：2023-12-20

**★第1章、安装PVE虚拟化集群**

**★pve简介及下载安装镜像文件**

Proxmox VE是商业公司Proxmox Server Solutions的开源产品，Proxmox VE软件和社区支持都是免费的，企业用户则可以通过订阅制获得付费商业支持。

根据Proxmox VE的官网介绍，Proxmox VE是一个集计算、网络及存储于一体的解决方案，这就是我们通常所说的超融合解决方案。

Proxmox VE是一个企业级虚拟化平台，该平台集成了基于内核的虚拟机管理程序（KVM）和基于Linux容器（LXC）这两种虚拟化技术，同时还集成了软件定义存储和虚拟网络功能。借助Web的管理界面工具，可以轻松地管理和配置虚拟机、容器、高可用性集群、软件定义存储、虚拟网络以及备份等。

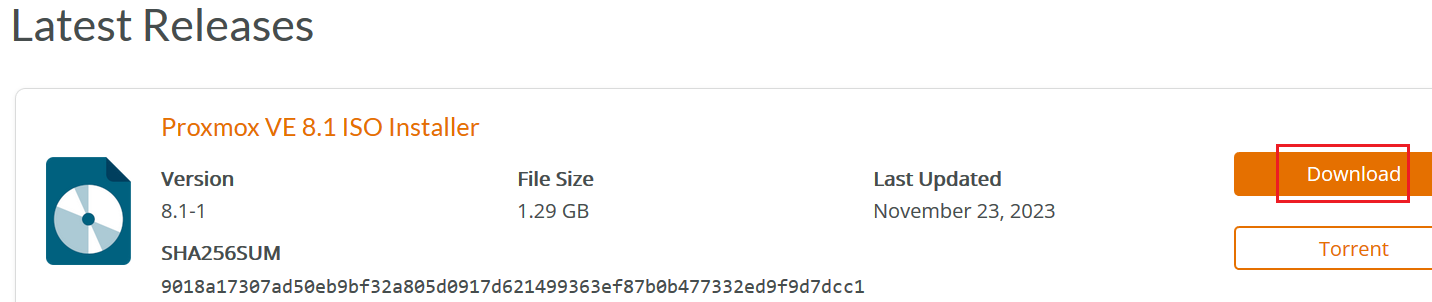
管理方面：

Proxmox VE是去中心化架构的，无论是单节点还是集群，都只需要安装一个Proxmox VE套件。

Proxmox VE基于Web界面进行去中心化管理，登录任何一个节点，就可以管理整个集群的节点，但是Proxmox VE目前只能实现单一集群管理，无法实现多集群管理，无法实现多云管理

Proxmox VE最多能管理32台物理节点，大概在1000个虚拟机左右的量

下载地址1： https://www.proxmox.com/en/downloads



下载地址2： https://enterprise.proxmox.com/iso/

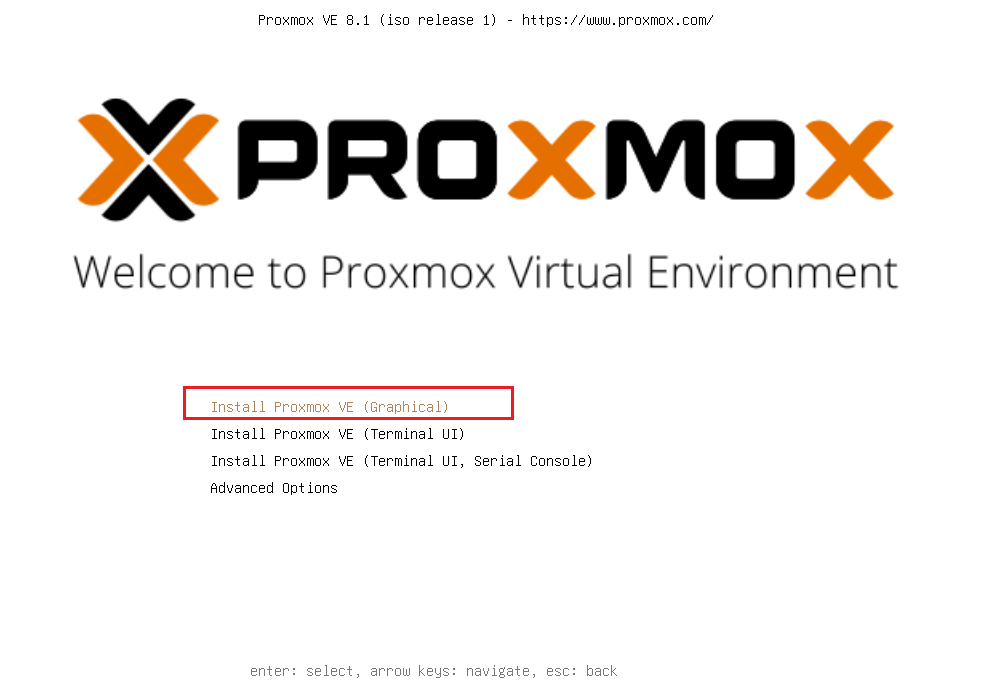
将iso光盘文件刻录到U盘，对物理机进行安装系统

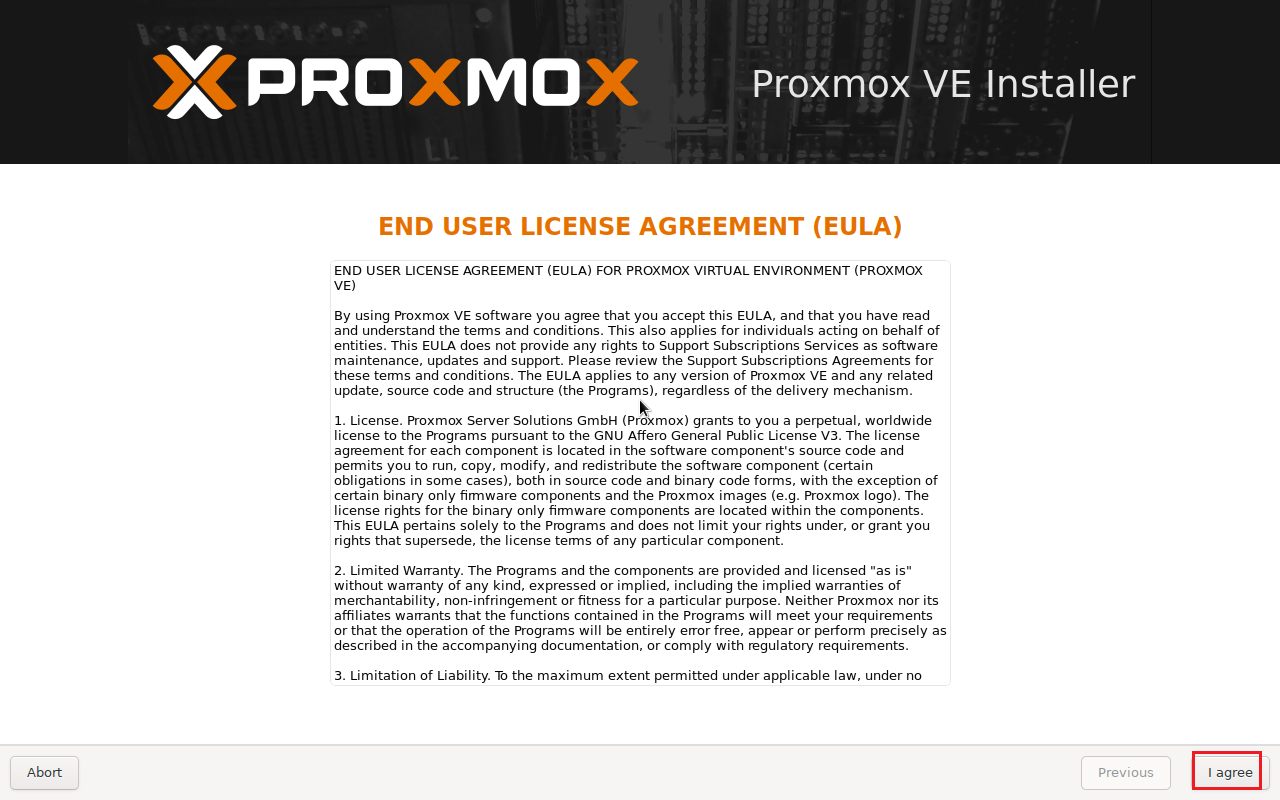
刻录工具有Rufus，官网： https://rufus.ie/en/

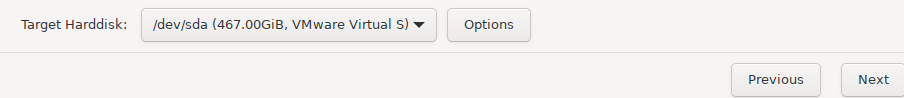
**★规划集群相关信息**

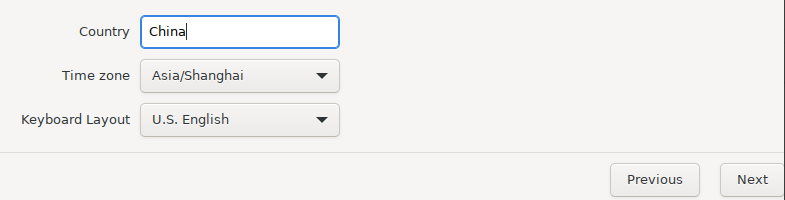
|  |  |  |  |
| --- | --- | --- | --- |
| 主机名 | 系统管理ip（同cluster网络） | 虚拟机业务口 |  |
| pve1.cof-lee.com | 10.99.1.21 （vmbr0→eth0） | vmbr1（bond0→eth1&eth2） |  |
| pve2.cof-lee.com | 10.99.1.22 （vmbr0→eth0） | vmbr1（bond0→eth1&eth2） |  |
| pve3.cof-lee.com | 10.99.1.23 （vmbr0→eth0） | vmbr1（bond0→eth1&eth2） |  |

**★安装pve系统**

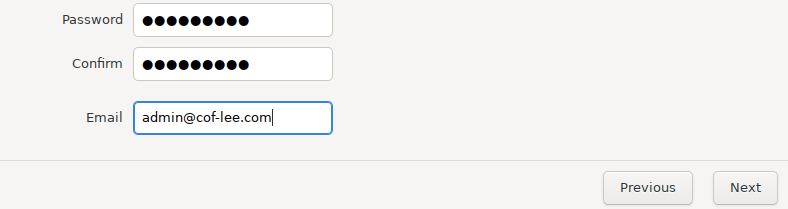




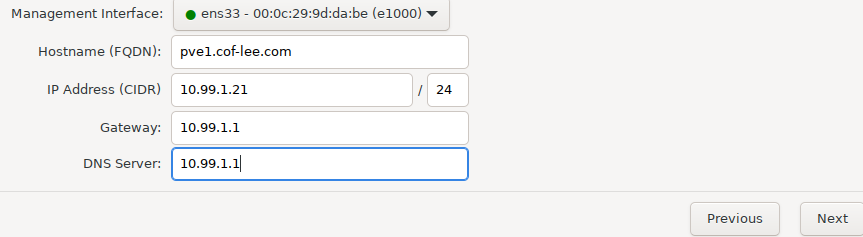


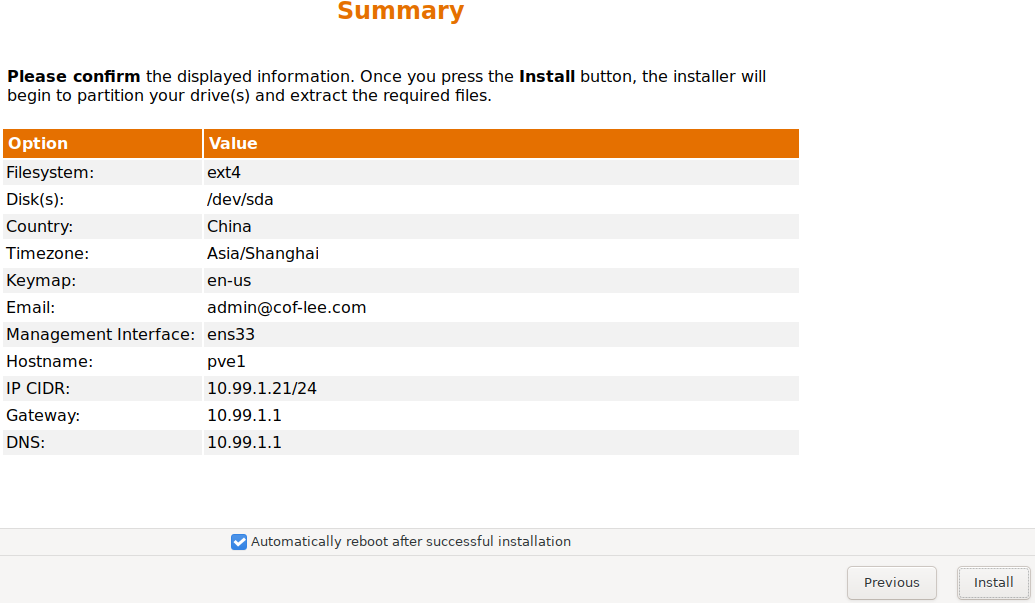


设置root管理员密码及邮箱

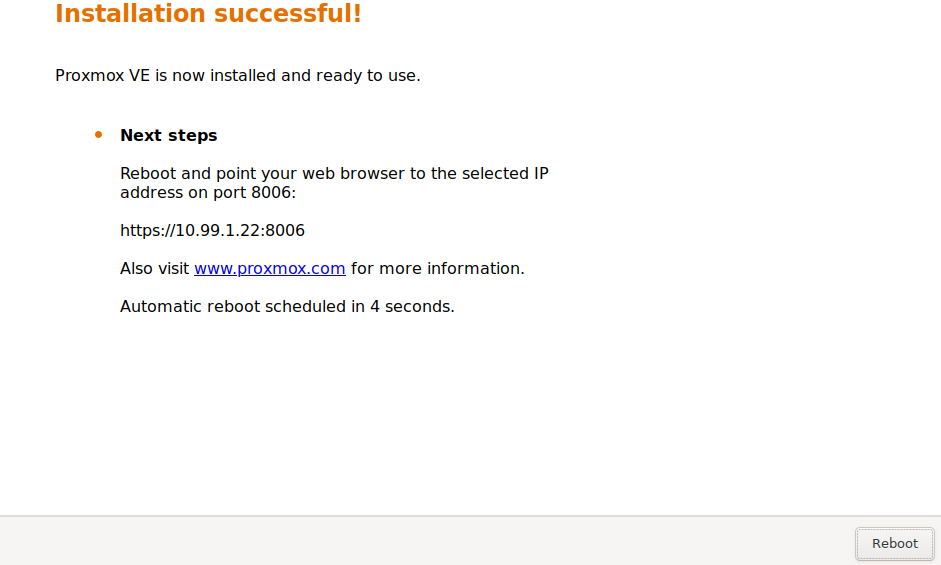


设置管理ip





确认信息无误后，点击“Install”安装



安装完成后，会自动重启，重启后可进入pve系统

可从console登录，也可从ssh远程登录，或者从浏览器登录 https://pve-ip:8006/

用户名为root，密码为前面设置的管理员密码

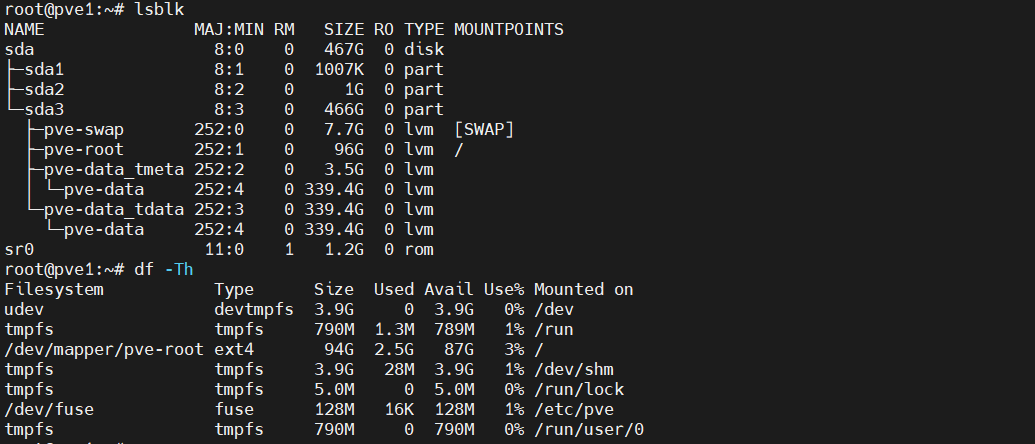
其他几台物理机也按照上面的步骤进行安装，主机名和管理ip等信息要确保能无误

**★系统基本情况查看**

467GB的系统盘，自动分区情况：

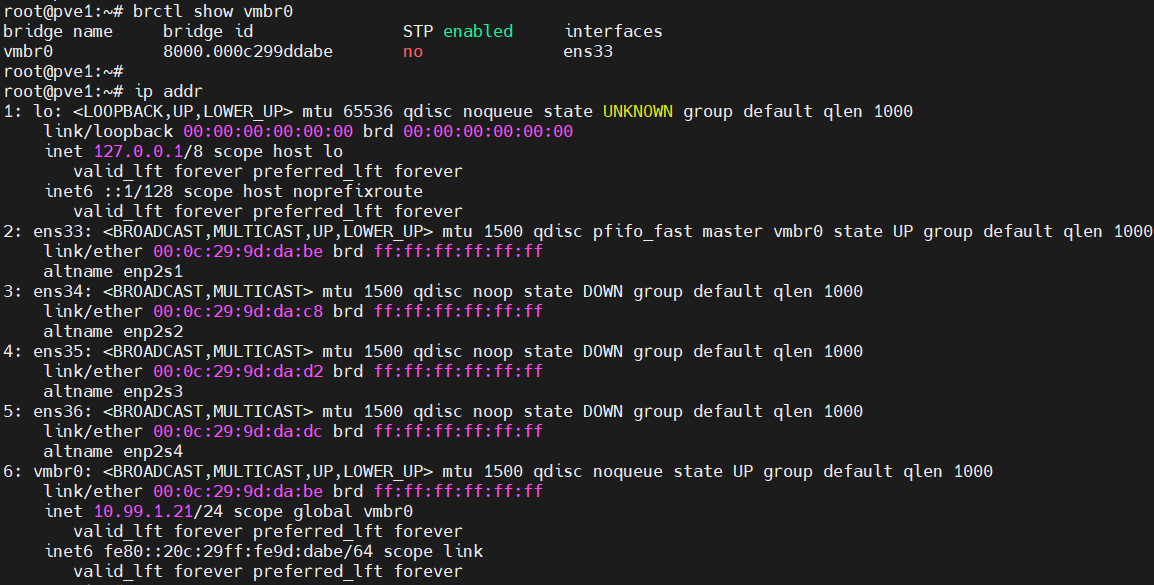
# lsblk

# df -Th #自动创建了一个名为pve-data的lvm-thin逻辑卷，用于创建虚拟机的本地磁盘，系统根分区pve-root大小只占总磁盘的20%的存储空间



# ip addr

# brctl show vmbr0 #自动将我们安装时选择的网卡加入到一个名为vmbr0的网桥里，系统管理ip是配置在vmbr0网桥上的

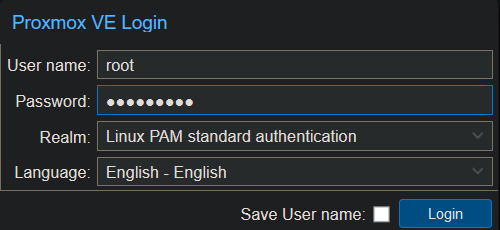


**★创建pve集群**

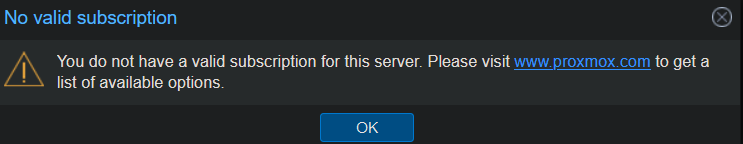
将刚刚安装好的3台宿主机添加到同一个集群里

使用浏览器登录第一台宿主机 https://10.99.1.21:8006

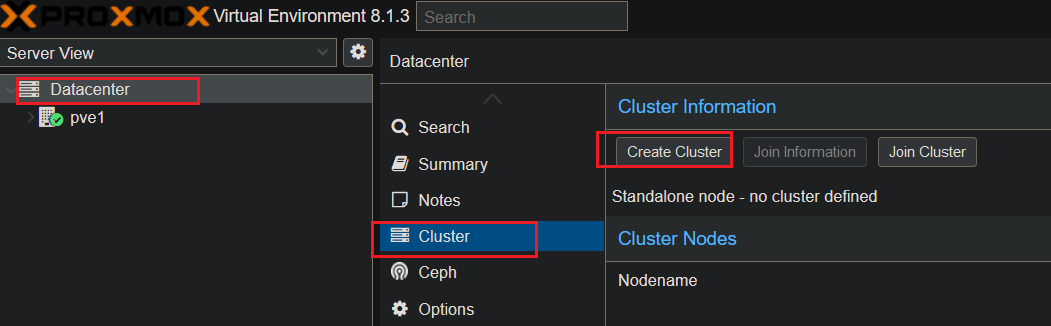
输入账号root ，密码为安装时指定的密码，其他保持默认，点击右下角“Login”登录



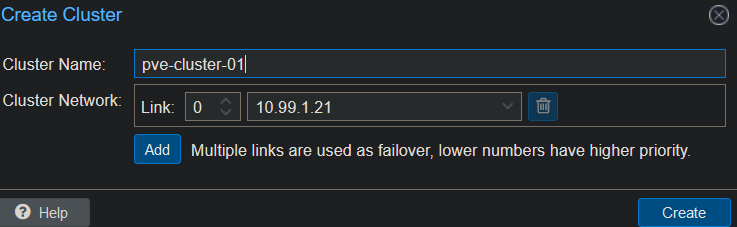
首次登录会有个弹窗提示：



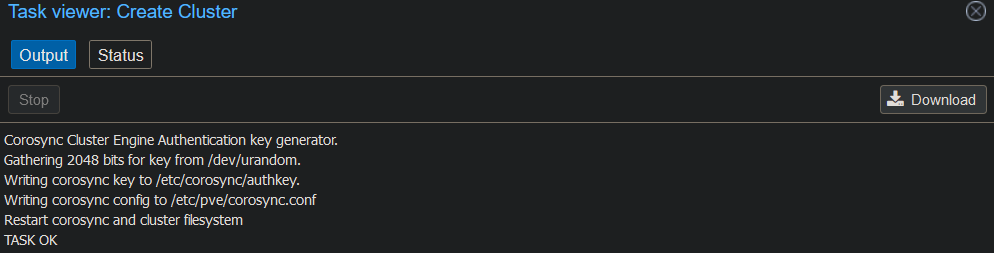
因为pve有商业支持，如果需要商业支持，可购买他的相关订阅，这里我们不使用，点击“OK”或右上角的关闭按钮即可。



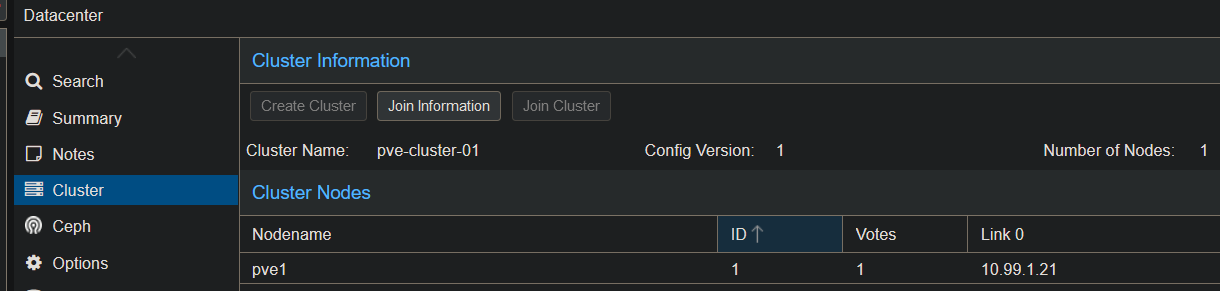
点击 Datacenter→Cluster→Create Cluster 创建集群



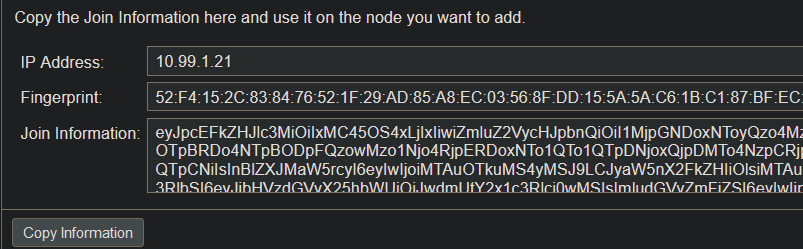
集群网络，按照规划的选择（本例使用的集群网络同系统管理网络），点击“Create”创建集群



出现TASK OK后，表示创建成功，点击右上角的“X”关闭此窗口

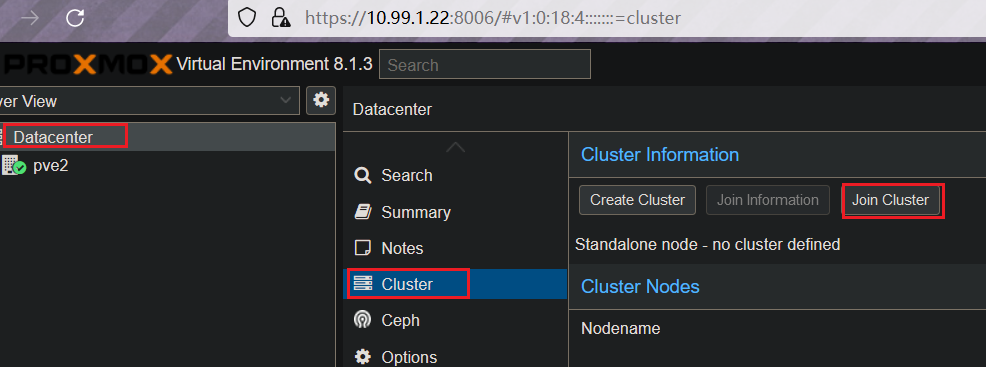


点击“Join Information”



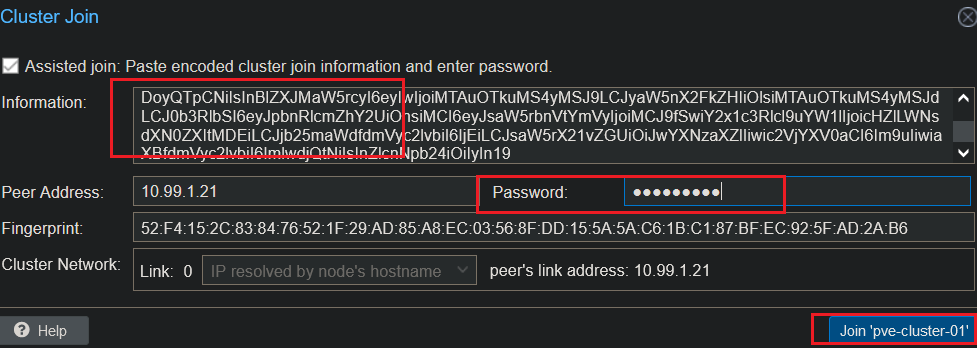
点击“Copy Information”复制加入集群的验证信息，以供其他几台宿主机加入此集群

用浏览器分别登录其他几台pve宿主系统，执行以下步骤：

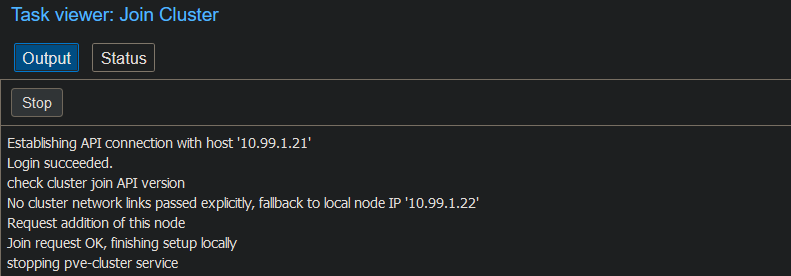


Datacenter→Cluster→Join Cluster 加入集群

将pve1第一台宿主系统上的join information复制粘贴过来，并填写第一台宿主系统的密码，点击右下角加入集群



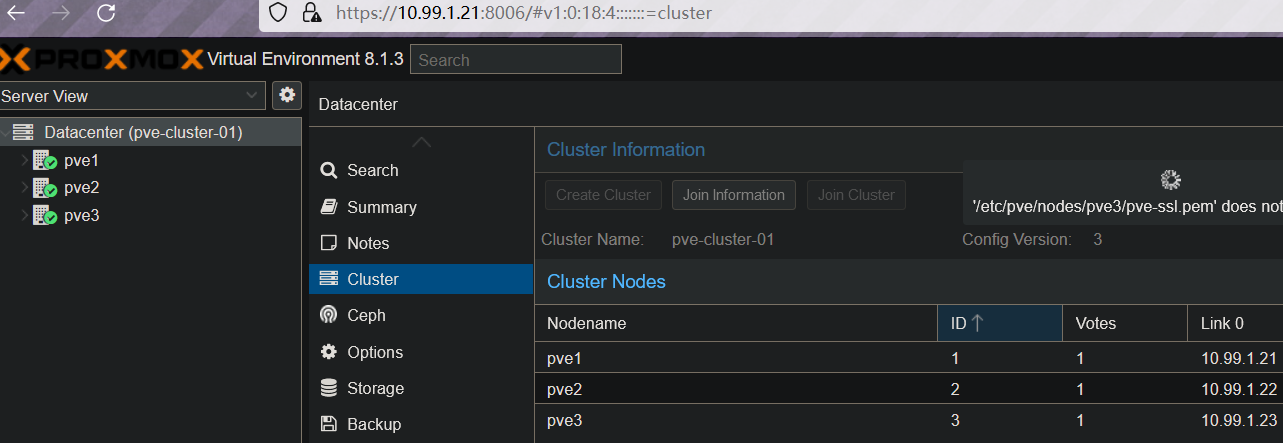
Link: 0这里要选择集群网络的网卡，若未选择则默认使用系统管理ip



上图表示加入成功，此界面会卡住，需要刷新浏览器，重新登录pve2

在pve1上可见pve2已加入了集群

其他节点也按上面的步骤加入到集群，过程省略



**★将pve节点从集群移除**

# pvecm status #查看集群状态

# pvecm nodes #查看集群节点（离线的不显示）

# pvecm delnode pve3 #移除节点（要求剩余的节点数量能够正常选举出master）

Could not kill node (error = CS\_ERR\_NOT\_EXIST) #离线节点报这个错是正常的

Killing node 3

**★删除pve集群**

#在创建集群的节点上停止pve-cluster服务

# systemctl stop pve-cluster.service

# systemctl stop corosync.service

#将节点的集群文件系统设置为本地模式的命令

# pmxcfs -l

#删除 corosync 配置文件

# rm /etc/pve/corosync.conf

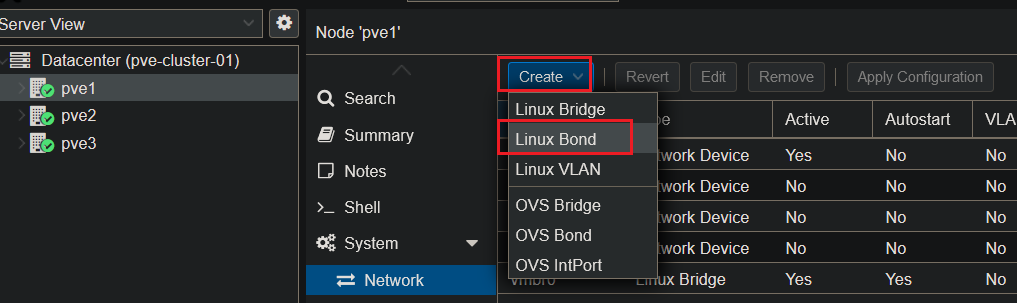
# rm -rf /etc/corosync/\*

**★第2章、pve网络配置**

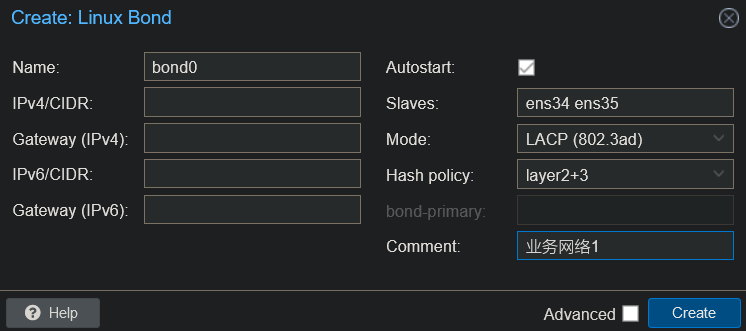
**★创建聚合口**

根据前面的规划，在每台pve宿主机上创建虚拟机业务网络vmbr1（bond0→eth1&eth2）

即先创建名为bond0的聚合口（成员为eth1及eth2），再创建vmbr1网桥，将bond0加入到此网桥中



点击单台pve节点→System→Network→Create→Linux Bond

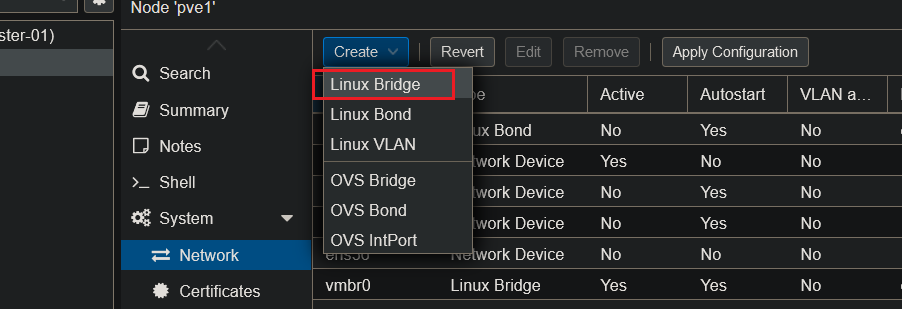


Name为聚合口名称，写bond0，Slaves填写成员口名称，多个网口之间使用空格隔开

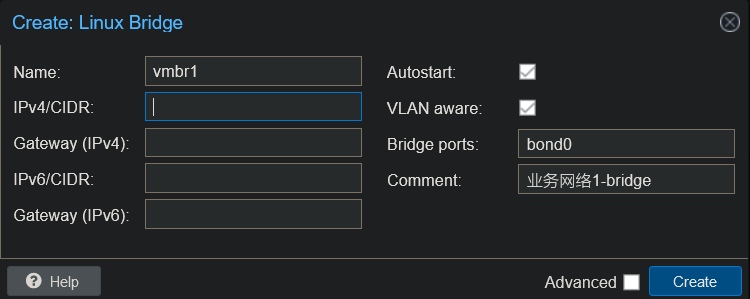
Ipv4/cidr填写聚合口的ip，这里不填写（因为要加入到网桥，若要配置ip也是在网桥上配置）

Mode为聚合模式，可为balance-xor balance-rr active-backup lacp 等聚合模式，本例使用lacp(802.3ad)

**★创建网桥**



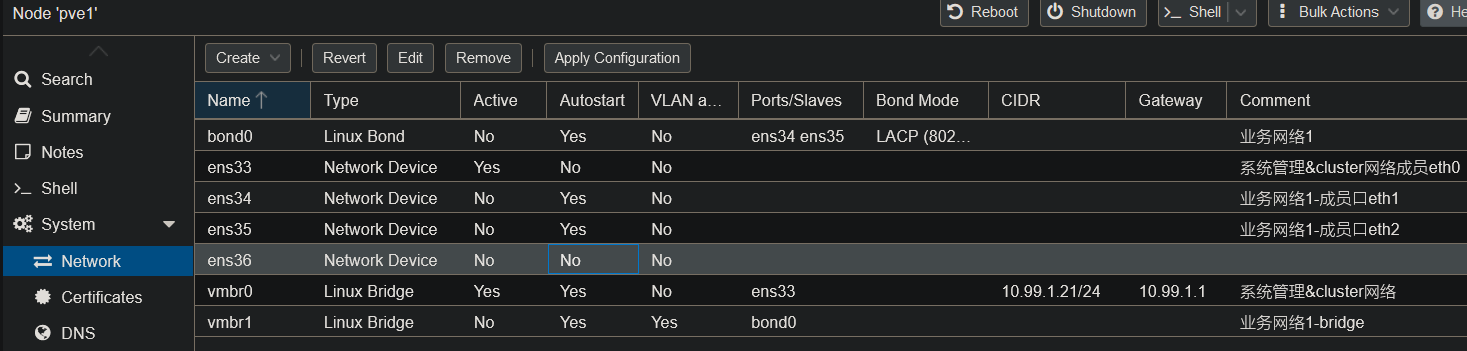
System→Network→Create→Linux Bridge



Name填写网桥名称，Autostart和VLAN aware都勾选上

IPv4/cidr可不填写，因为这是直通模式，虚拟机的网络直通到接入交换机，网关在交换机上，pve的网桥可以不用配置ip及网关

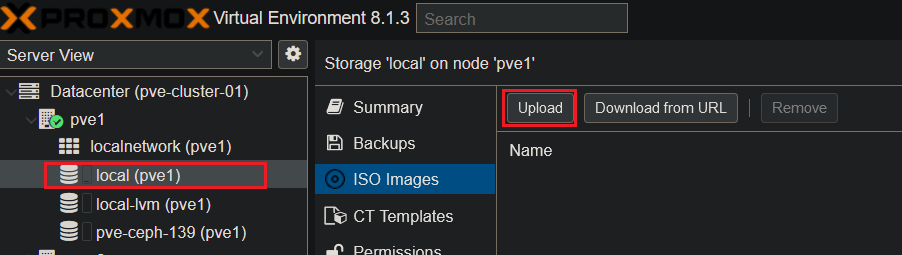
Bridge ports填写成员口，把前面创建的bond0聚合口放入此网桥中



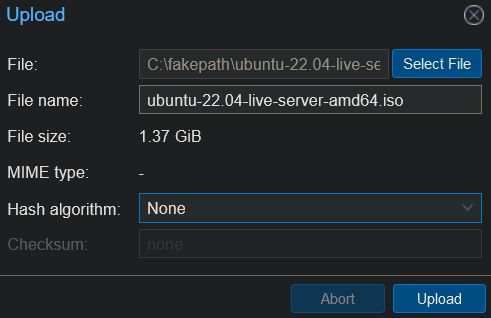
**★第3章、虚拟机操作**

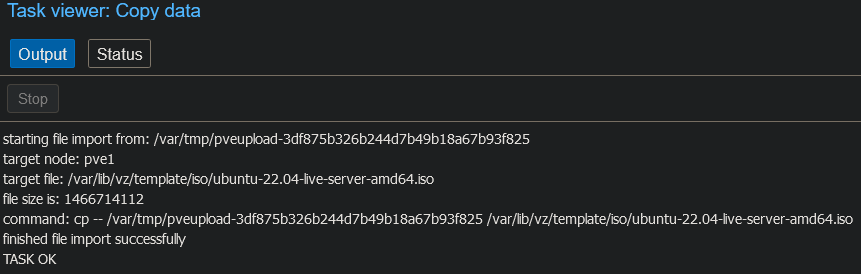
**★上传iso镜像到pve宿主机**

上传虚拟机系统安装光盘文件到pve宿主节点上



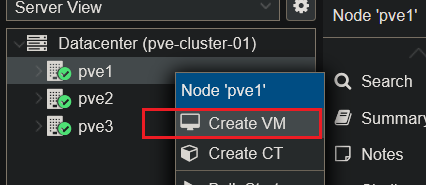
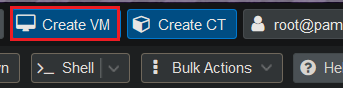
展开pve节点，点击Local(pve1)→ISO Images→Upload 选择要上传的镜像文件

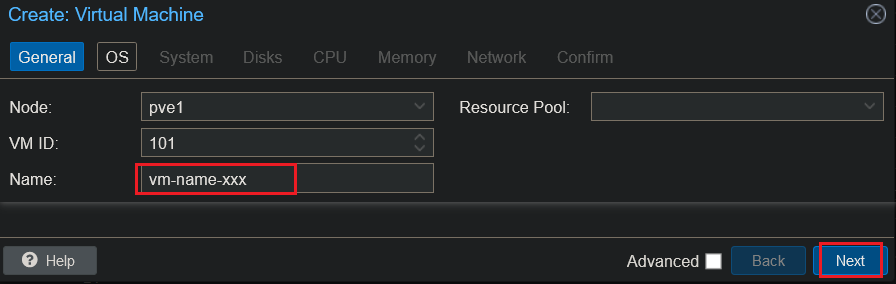


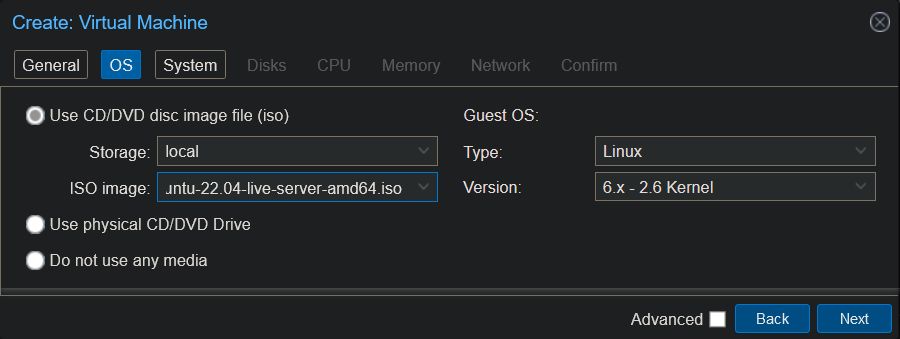


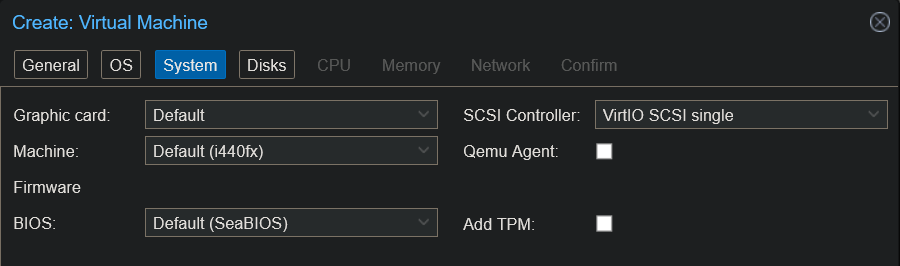
**★创建虚拟机**

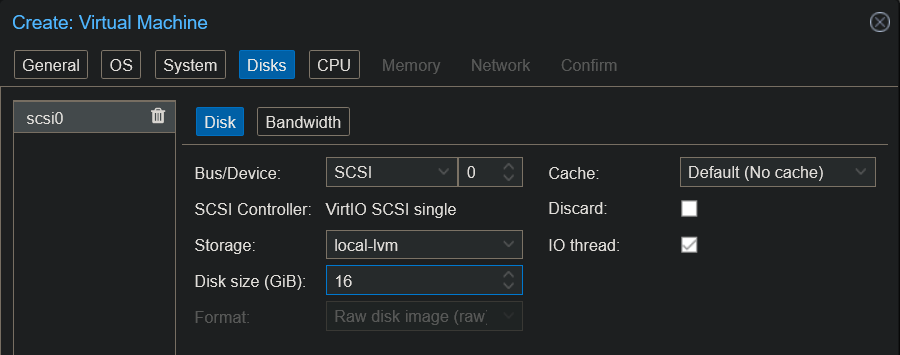
可以右键单击pve节点，弹出二级菜单里点击“Create VM”，也可直接在主界面右上角点击“Create VM”

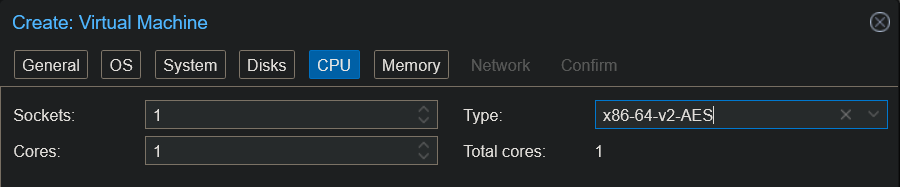
 

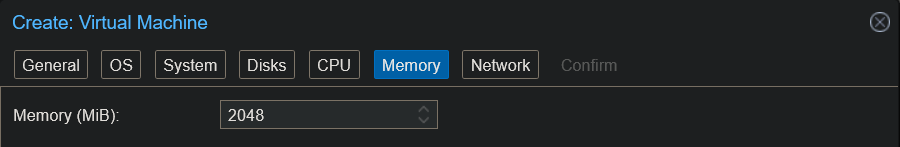


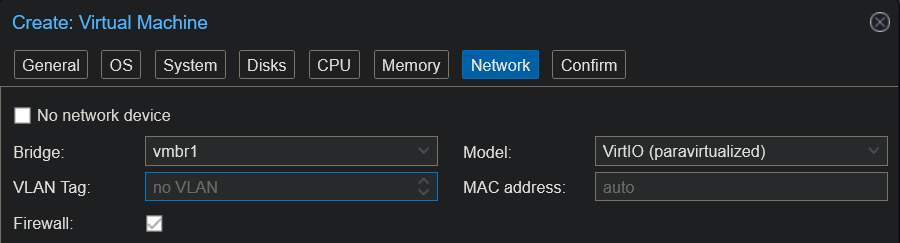




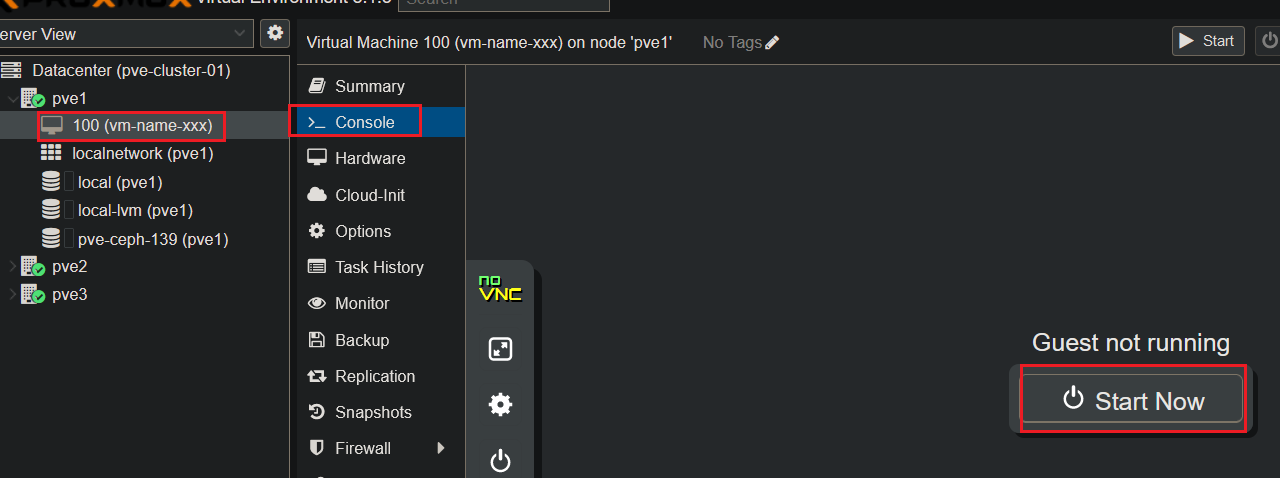


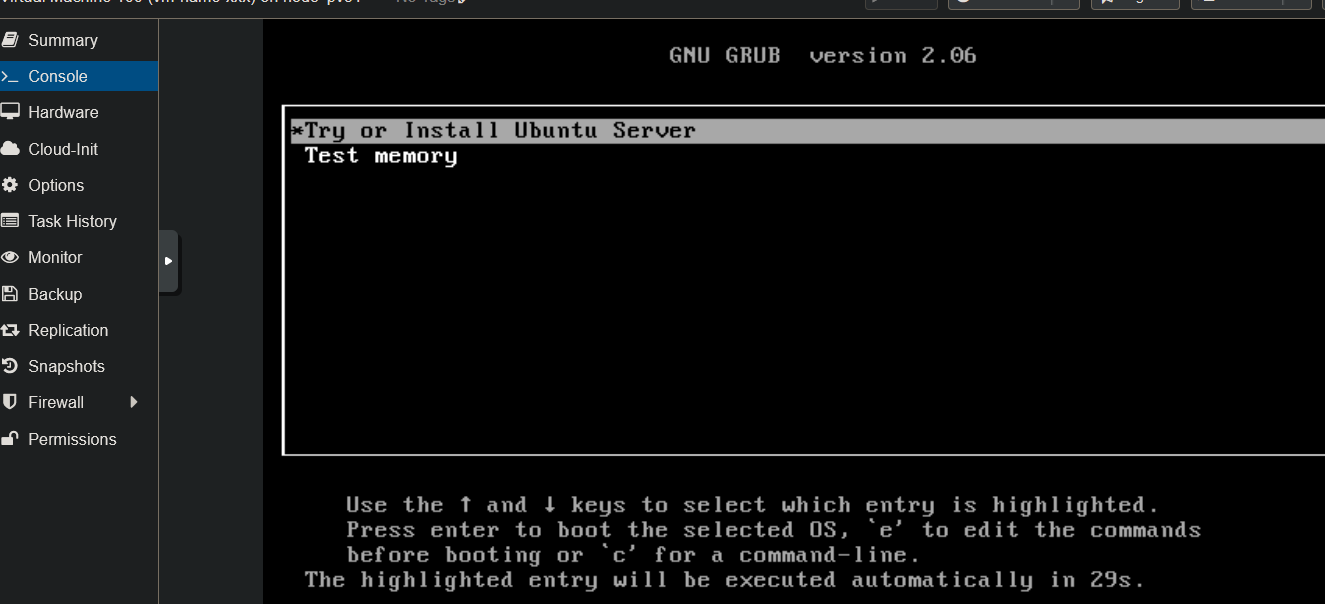






网络虚拟化技术，Virtio是半虚拟化而E1000是全虚拟化。半虚拟化技术的隔离度是没有全虚拟化好的，这就是为什么在使用Virtio时，OpenWRT网络出现问题会导致整个Proxmox的网络都不能用了的原因。除了这两种虚拟化方式外，还有些更为先进的虚拟化技术，如SR-IVO等

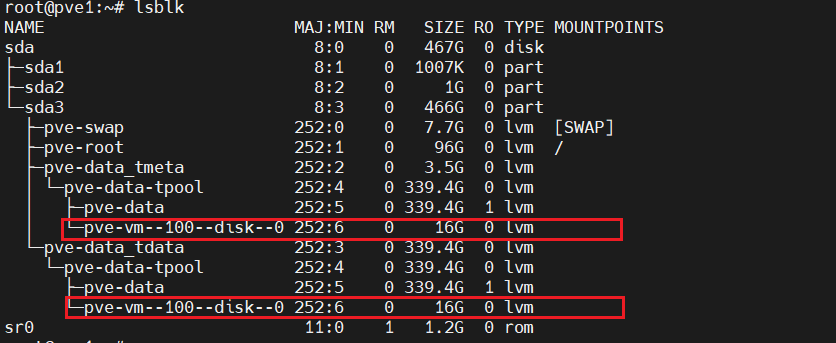




按正常流程安装系统即可

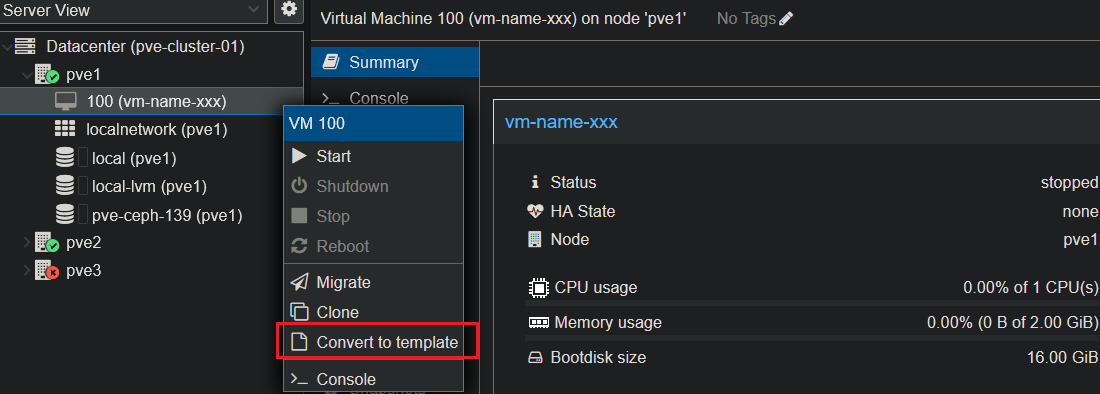
由于使用的是local存储（pve宿主机上的存储），登录pve1上查看存储分配情况：

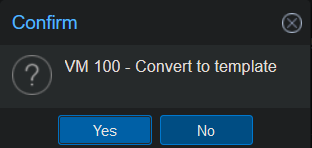
分配了一个逻辑卷 pve-vm--100--disk--0 给刚刚创建的虚拟机使用



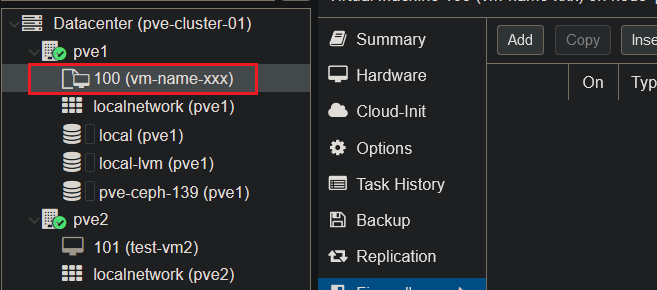
**★创建虚拟机模板**

根据已创建好的虚拟，制作模板，以供之后快速创建虚拟机

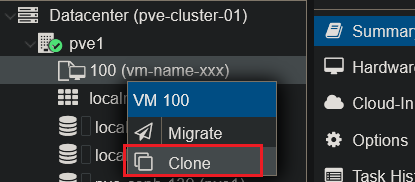


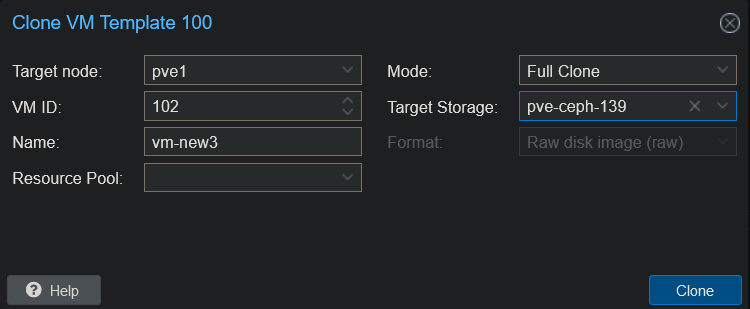


转换为template模板后，这台虚拟机不可直接使用了，没有了“Console”界面，无法开机，只能基于此模板创建新的虚拟机



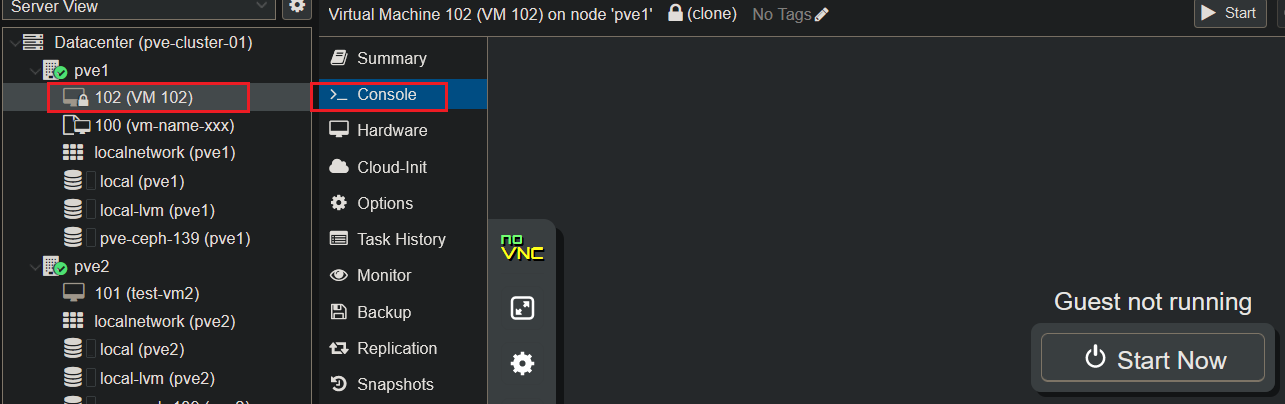
右键单击模板，弹出二级菜单点击“Clone”克隆

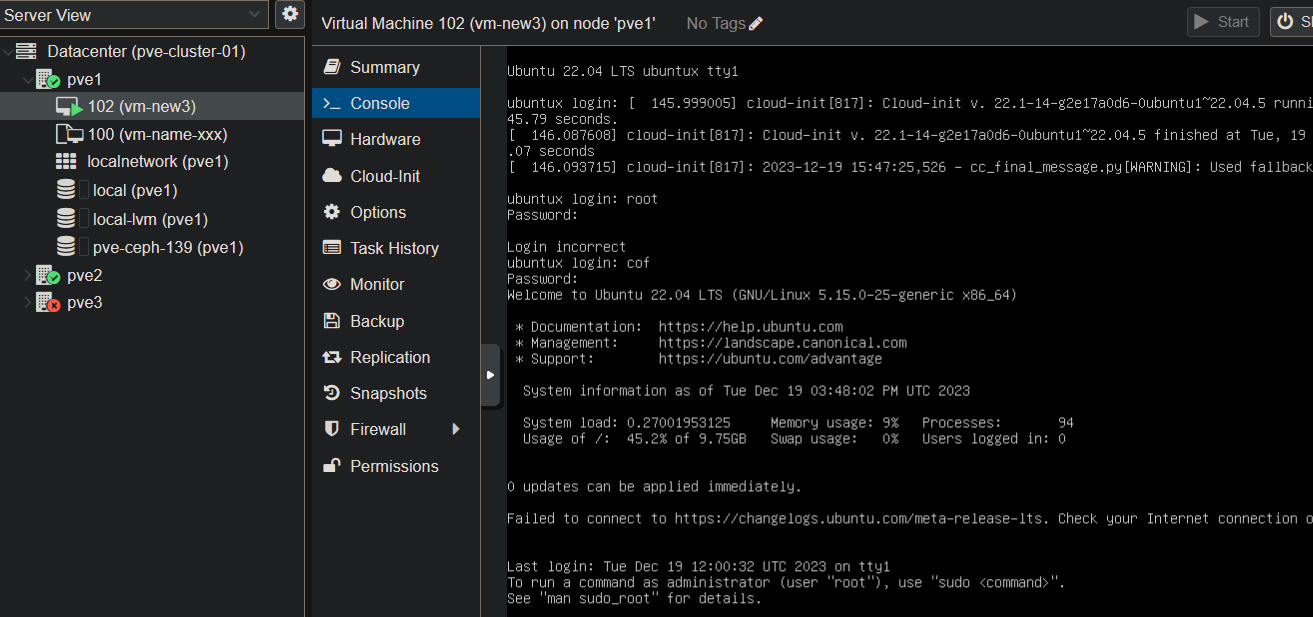




Mode使用“Full Clone”完全克隆，目标存储Target Storage可以按需选择

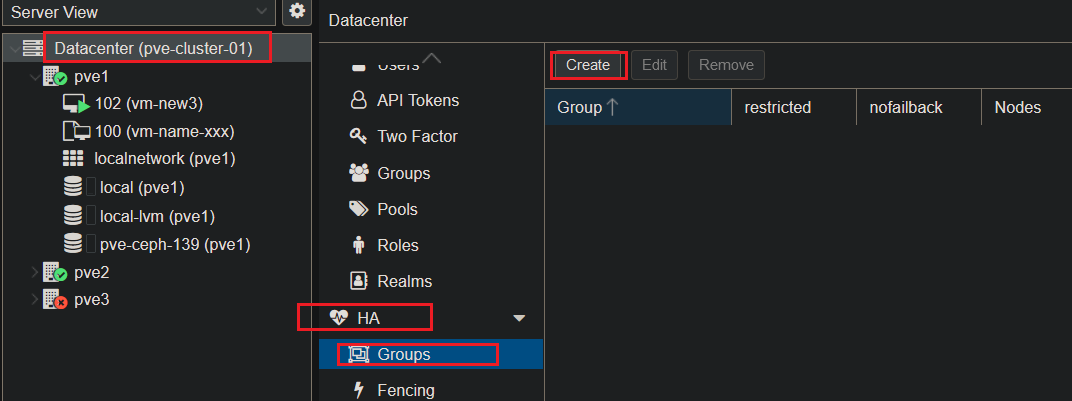
克隆后的虚拟机有console界面，可开机（如果虚拟机有个锁的图标，说明被上锁了，数据还在后端迁移复制之中，需要等待数据迁移完成，才能进一步操作）

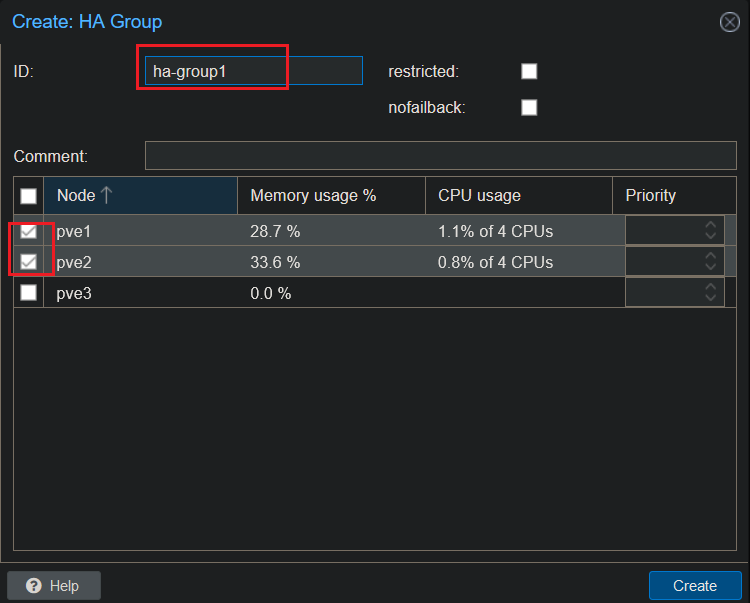




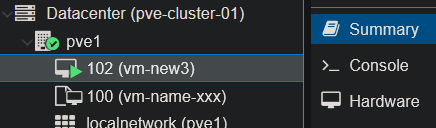
**★创建宿主机HA高可用组**

单击Datacenter→HA→Groups→Create

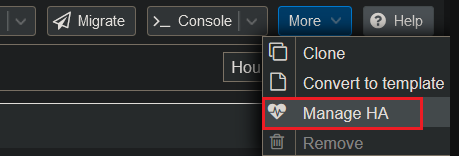


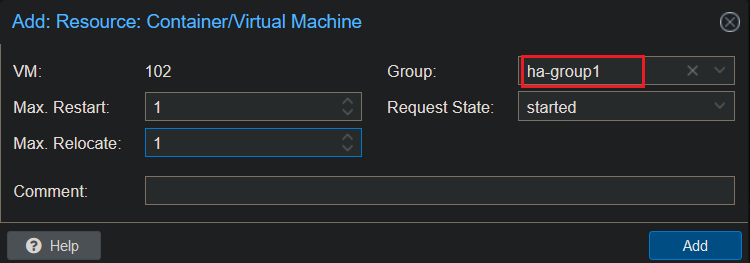


HA组创建完成后，选中目标虚拟机

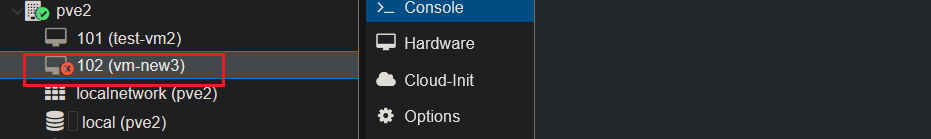


右上角点击“More”→Manage HA





将虚拟机所在宿主机pve1关机，会自动迁移到HA组中的其他pve宿主机去



迁移到pve2后，开机失败，报错如下：

ask started by HA resource agent

TASK ERROR: volume 'local:iso/ubuntu-22.04-live-server-amd64.iso' does not exist

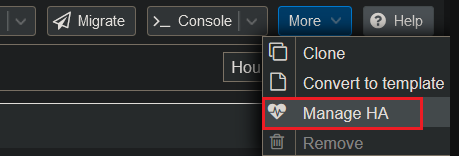
原因是虚拟机的光盘文件之前是用的pve1宿主机上的，现在pve2上没有此光盘文件，上传光盘文件到pve2后再开机，仍然报错：

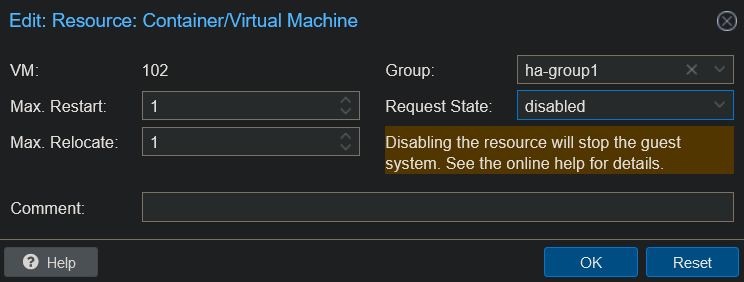
Requesting HA start for VM 102

service 'vm:102' in error state, must be disabled and fixed first

TASK ERROR: command 'ha-manager set vm:102 --state started' failed: exit code 255

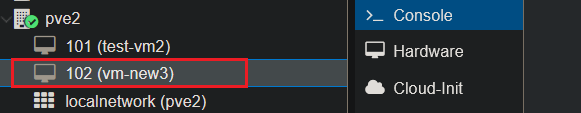
只能先把此虚拟机从HA组里删除，解除error状态后，再加入到HA组中





Request State设置为disabled，点击“OK”

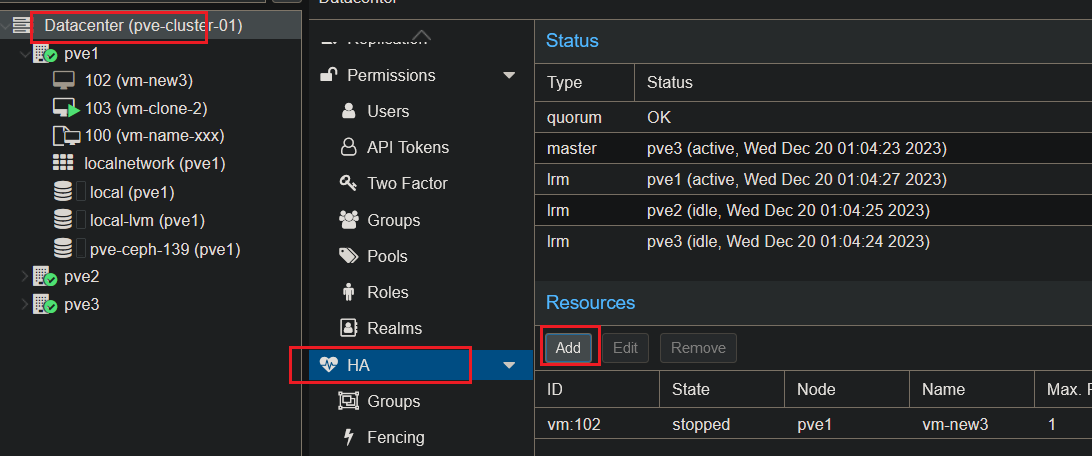
然后虚拟机的error状态解除



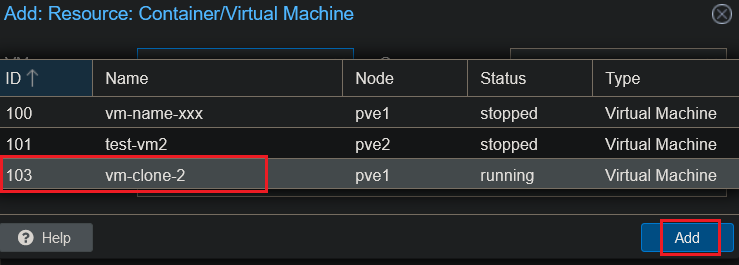
再重新加入HA，Request State设置为started即可

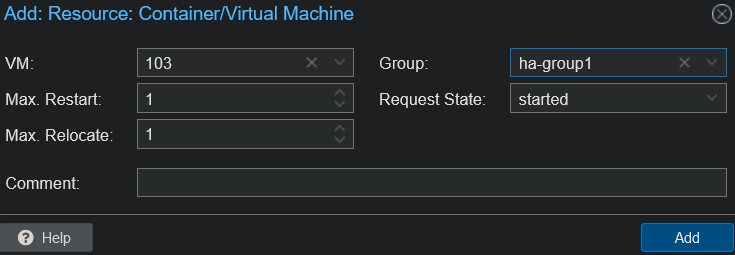
**也可在HA界面操作：**

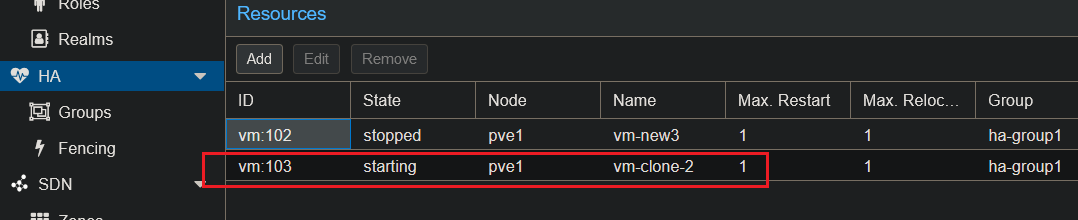
点击Datacenter→HA→Add



选择目标虚拟机







**★禁用pve宿主网卡的offload功能**

网络虚拟化技术，Virtio是半虚拟化而E1000是全虚拟化。半虚拟化技术的隔离度是没有全虚拟化好的，这就是为什么在使用Virtio时，OpenWRT网络出现问题会导致整个Proxmox的网络都不能用了的原因。除了这两种虚拟化方式外，还有些更为先进的虚拟化技术，如SR-IVO等

有时为了防止虚拟机的网络干扰到宿主网卡，需要使用E1000这种全虚拟化网卡，且禁用pve宿主网卡的offload功能

# apt install ethtool

# ethtool -K bond0 tso off gso off #禁用tcp分段卸载和通用分段卸载

# vi /etc/network/interfaces

#在对应的网卡配置下, 与address同级添加配置:

iface bond0 inet manual

offload-gso off

offload-gro off

offload-tso off

offload-rx off

offload-tx off

offload-rxvlan off

offload-txvlan off

offload-sg off

offload-ufo off

offload-lro off

**★后台关闭虚拟机**

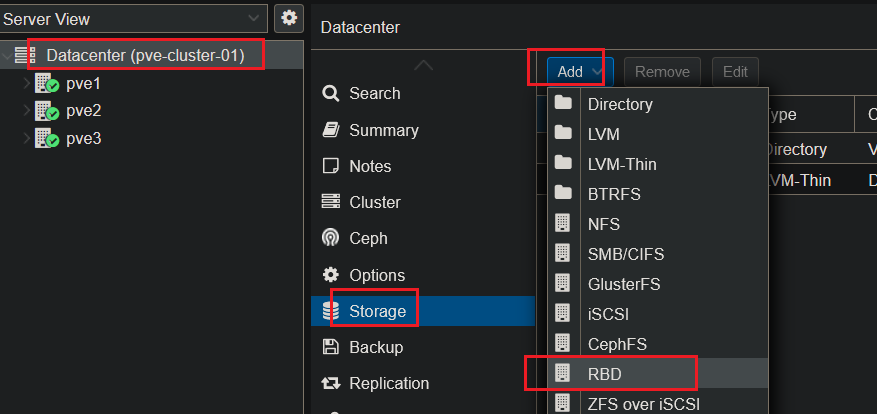
有时windows虚拟机无法从console下发shutdown指令，可以从后台执行关机操作

# rm /var/lock/qemu-server/lock-102.conf #102为目标虚拟机id

# qm stop 102 #102为目标虚拟机id

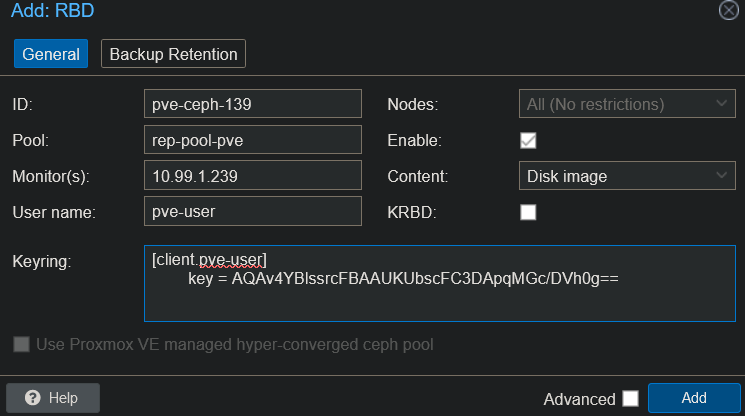
**★第4章、pve对接ceph存储**

点击“Datacenter”→Storage→Add→RBD



ID随便写，Nodes不指定则表示所有pve宿主机节点都可以使用此ceph集群

Pool为ceph存储池名称，Monitor为ceph的mon-ip，username及keyring为此存储池的验证账号信息



确认信息无误，点击右下“Add”添加

ceph配置：

# ceph osd pool create rep-pool-pve 32 32 replicated

# ceph osd pool application enable rep-pool-pve rbd

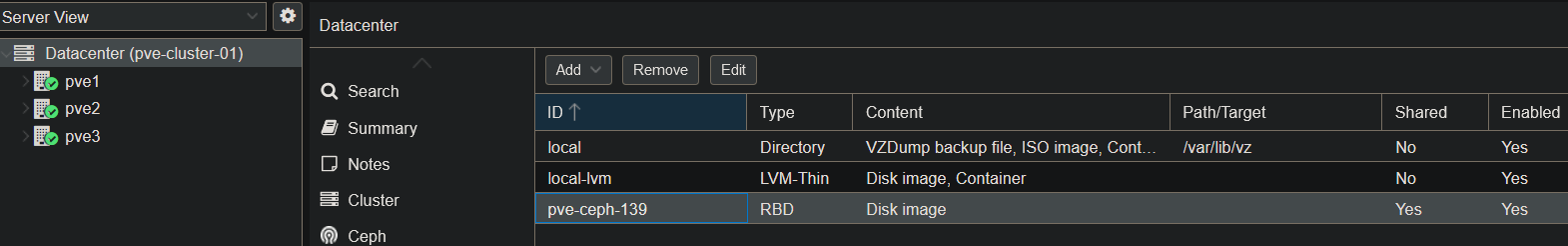
# ceph auth get-or-create client.pve-user \

# mon 'allow r' \

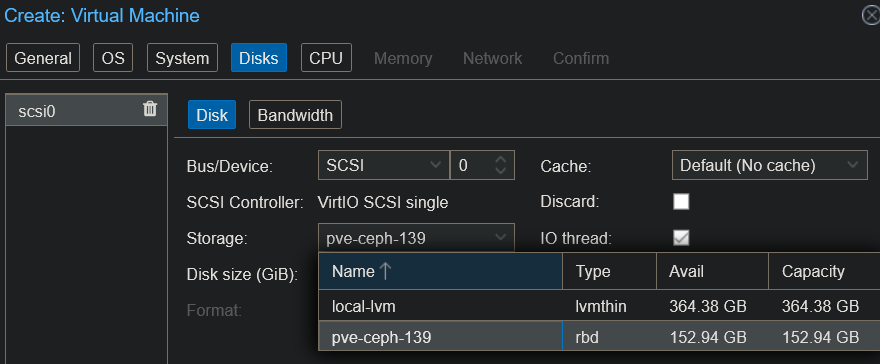
osd 'allow rwx pool=rep-pool-pve' \

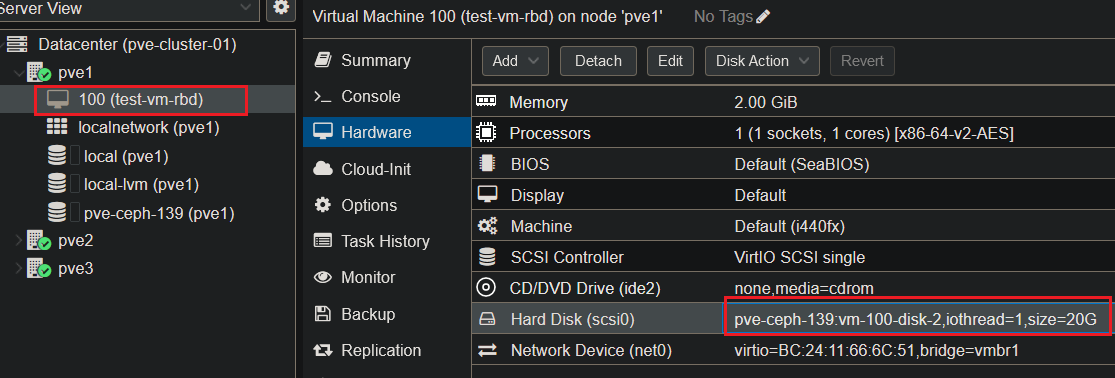
-o /etc/ceph/ceph.client.pve-user.keyring

添加后可见：

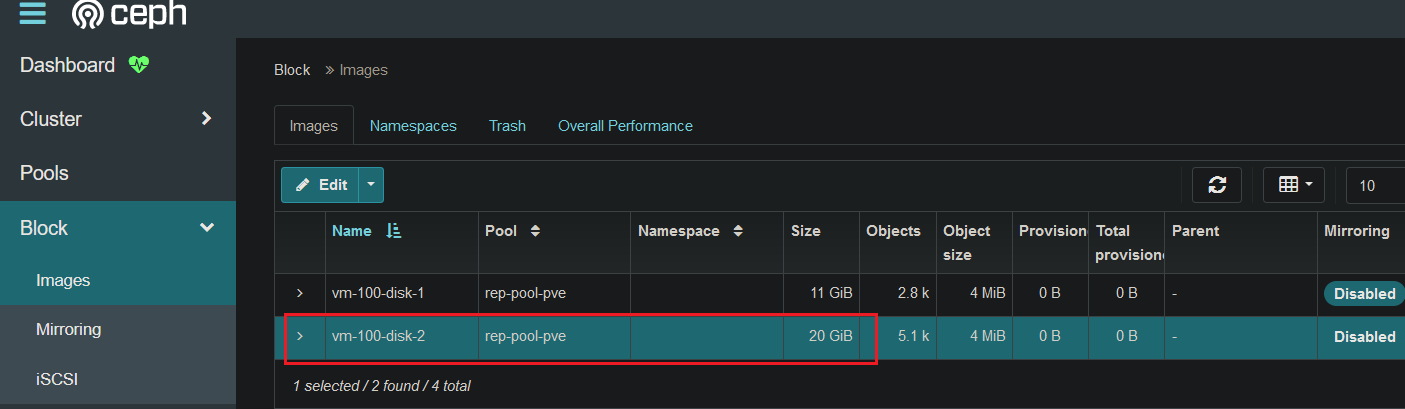


创建虚拟机时可选到此存储：





在ceph集群上可见已由pve集群创建了此磁盘：



**★附录、常见报错处理**

虚拟机启动失败，提示如下

kvm: -device ide-hd,bus=ide.0,unit=1,drive=drive-ide1,id=ide1,bootindex=101: Can't create IDE unit 1, bus supports only 1 units

TASK ERROR: start failed: QEMU exited with code 1

原因是使用了“q35”作为机器类型，则只有 ide0 和 ide2 可用（这是 qemu的限制）

解决方法：转到虚拟机的“硬件”选项下，把硬盘设备中的ide1去掉，重新添加为ide2，或者将ide1修改为ide2即可