云联壹云

物理机管理使用场景和实践

李泽玺

2020-12-30

- 背景介绍
- 物理机管理架构
- 注册场景介绍
- 装机实践

背景介绍

什么是物理机(服务器)?

更好的硬件配置

- 内存: 512G
- CPU: Intel Xeon 系列
- 磁盘: RAID 控制器 & N 块HDD + SSD
- 网卡: 多块万兆网卡

具备带外控制

- BMC 控制器: IPMI & Redfish API
  - 控制引导顺序
  - 远程开关机
  - 挂载 ISO 镜像

物理机管理的痛点

- 交付效率低
  - 从机器上架 -> 记录硬件信息 -> 安装操作系统 -> 部署业务
- 配置麻烦
  - 网络分配，磁盘做 RAID 分区
  - 后期重装操作系统等操作

云平台提供的管理功能

硬件信息记录

- 内存大小
- CPU 型号
- 硬盘信息
- 网卡信息

安装操作系统

- 配置网络：分配 ip，配置bonding
- 配置磁盘: RAID 配置，磁盘分区
- 安装操作系统

生命周期管理

- 开关机
- 调整配置，比如添加内存

云平台物理机管理的优势

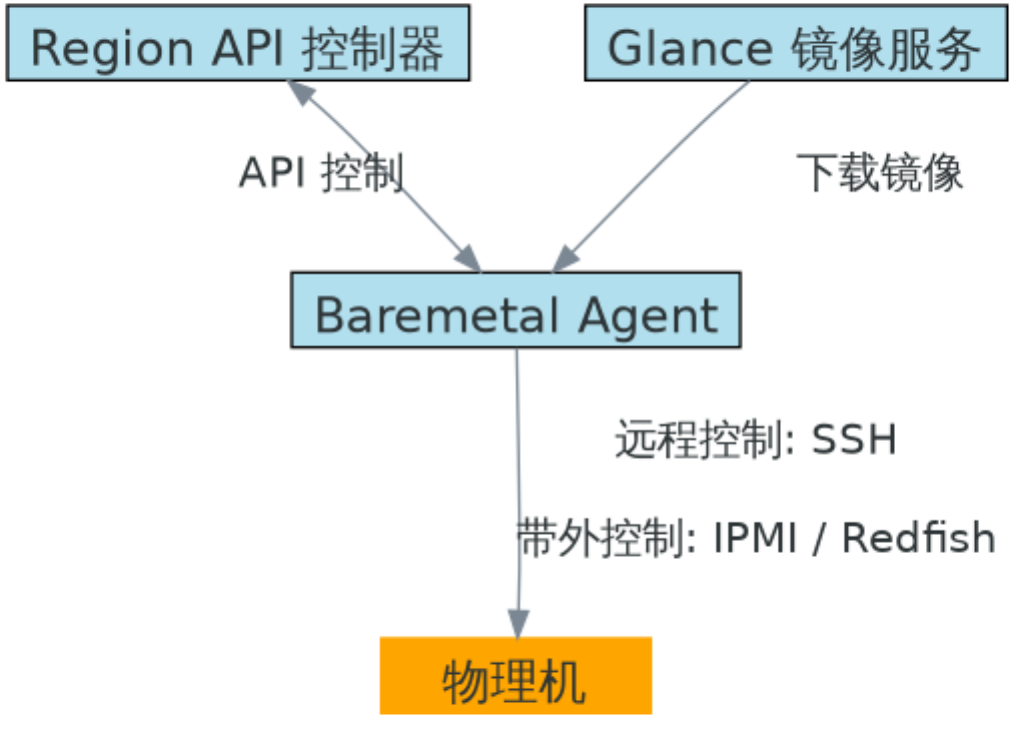
- 自动化装机提升效率、缩短交付时间
- 易于管理

	传统方式	物理机管理系统
交付 100 台物理机	1天+	30分钟
配置硬件	手动配置磁盘 raid、分区、网卡 bonding	提供 API 自动化配置
机器信息维护	Excel 或者第三方系统记录	云平台记录，API 查询

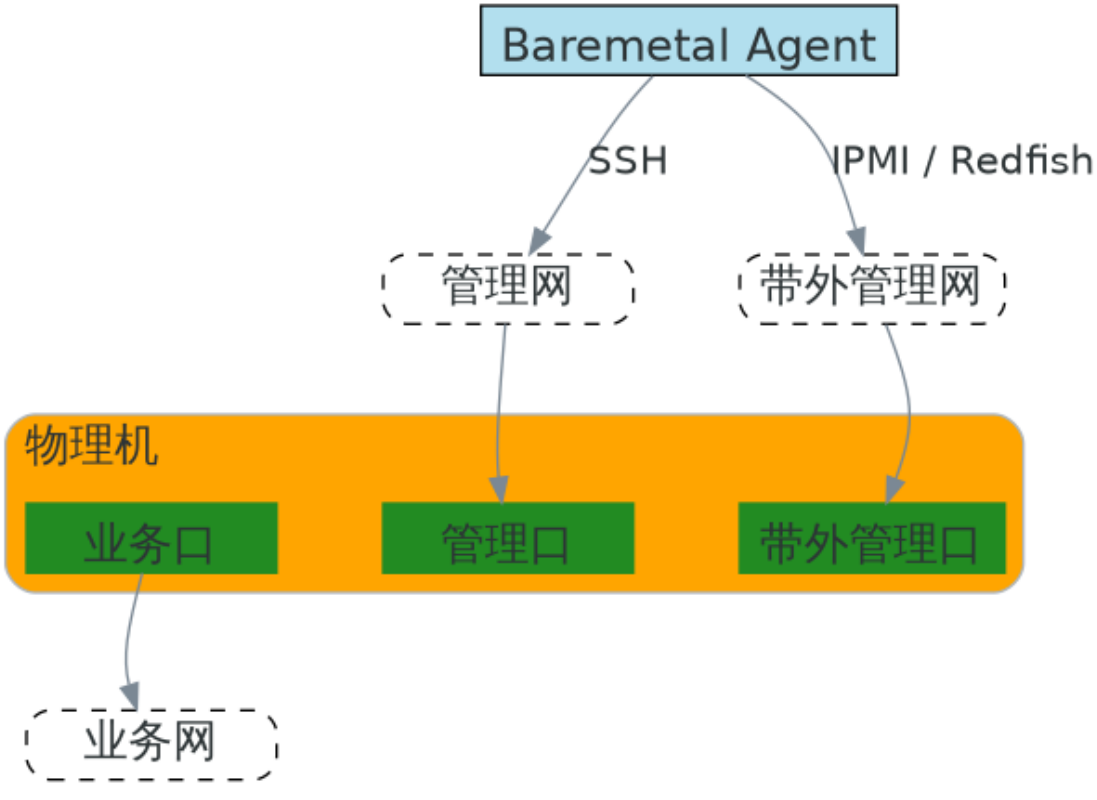
物理机管理架构

介绍物理机管理相关的架构

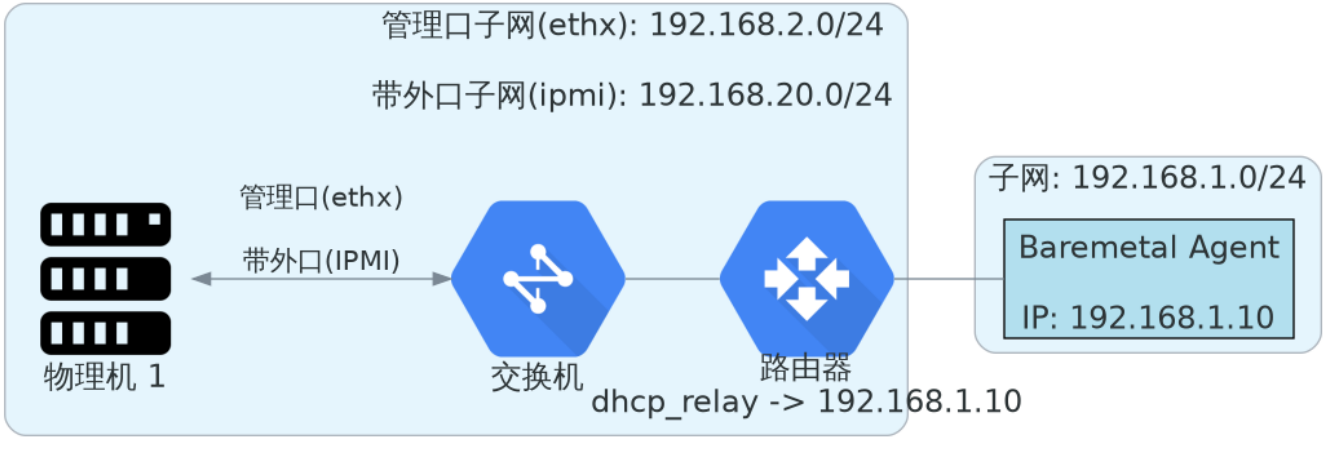
服务架构


Region API 控制器
Glance 镜像服务
API 控制
下载镜像
Baremetal Agent
远程控制: SSH
带外控制: IPMI / Redfish
物理机

物理机网络结构


Baremetal Agent
SSH
IPMI / Redfish
管理网
带外管理网
物理机
业务口
管理口
带外管理口
业务网

网络环境管理拓扑


管理口子网(ethx): 192.168.2.0/24
带外口子网(ipmi): 192.168.20.0/24
管理口(ethx)
带外口(IPMI)
物理机 1
交换机
路由器
dhcp_relay -> 192.168.1.10
子网: 192.168.1.0/24
Baremetal Agent
IP: 192.168.1.10

注册场景介绍

将物理机注册到云平台

注册物理机

注册方式	已有IPMI信息	启动引导方式	DHCP Relay	网络分配方式	依赖 Redfish API
ISO 引导注册	是	ISO	否	静态分配	是
PXE 引导注册	是	PXE	是	DHCP	否
预注册	否	PXE	是	DHCP	否

ISO 引导注册

* 物理机名称

字母开头，数字和字母大小写组合，长度为2-128个字符，不含"."，"_"，"@"

该物理机在系统中显示的名字

* IPMI地址

请输入IP地址

请输入已配置好的BMC的信息

* IPMI用户名

请输入用户名

请输入已配置好的BMC的信息

* IPMI密码

请输入密码

请输入已配置好的BMC的信息

* 管理口IP

请输入IP地址

会根据输入的IP子网或者IP地址设置物理机的管理口IP

PXE 引导注册

* 物理机名称

字母开头，数字和字母大小写组合，长度为2-128个字符，不含"."，"_"，"@"

该物理机在系统中显示的名字

* IPMI地址

请输入IP地址

请输入已配置好的BMC的信息

* IPMI用户名

请输入用户名

请输入已配置好的BMC的信息

* IPMI密码

请输入密码

请输入已配置好的BMC的信息

管理口IP

请输入IP地址

会根据输入的IP子网或者IP地址设置物理机的管理口IP，留空则使用DHCP自动分配的IP作为管理口IP

预注册

云联壹云

* MAC地址

请输入MAC地址

请输入物理机管理口的MAC地址，一般为eth0

* 物理机名称

字母开头，数字和字母大小写组合，长度为2-128个字符，不含"."，"_"，"@"

该物理机在系统中显示的名字

IPMI地址

为空时，系统会自动分配IP

为空时，系统默认自动分配IP；不为空时，则使用用户输入信息。
新机器建议留空，老机器建议输入旧IP。

IPMI用户名

为空时，系统默认使用 root

为空时，系统默认使用 root ；不为空时，则使用用户输入信息。
新机器建议留空，老机器建议输入旧用户名。

IPMI密码

为空时，系统默认使用 YunionDev@123

为空时，系统默认使用 YunionDev@123 ；不为空时，则使用用户输入信息。
新机器建议留空，老机器建议输入旧密码。

自动注册

- 适用自动化大规模，无需人为干预的场景
- 网络要求高，需要提前配置好各个子网的 DHCP Relay

命令行开启方法

```
$ climc service-edit-config baremetal
...
  # 自动注册选项，默认为 false
  auto_register_baremetal: true
...
```

注册完成的物理机(列表)

名称	启用	状态	IP	规格	品牌	分配	初始账号	MAC	IPMI
yfhost02	启用	运行中	10.127.30.254（管理） 10.127.30.27（带外）	40C256GRAID	DELL	未分配		24:6e:96:c6:25:18	
bmtest	启用	运行中	10.127.30.99（管理） 10.127.30.3（带外）	24C64GRAID	DELL	hcy-test		b8:2a:72:e0:ff:26	

注册完成的物理机(详情1)

云联壹云

CPU

核数： 24核

插槽数： 2个

超售比上限： 8

当前超售比例： 0

描述： Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.60GHz

内存

容量： 64G

系统预留： 0B

超售比上限： 1

当前超售比例： 0

磁盘

容量： 7.6T

类型： 机械硬盘

当前超售比例： 0

无效存储： 0B

适配器	驱动	型号	类型	容量	插槽序号
	MegaRaid	HITACHI HUS154530VLS300 B590JLWSK32C		285568	
	MegaRaid	HITACHI HUS154530VLS300 B590JLWTN2HC		285568	1
	MegaRaid	HITACHI HUS154530VLS300 B590JLWRZXWC		285568	2
	MegaRaid	HITACHI HUS154530VLS300 B590JLWSJBHC		285568	3
	MegaRaid	HITACHI HUS154530VLS300 B590JLWS08UC		285568	4

注册完成的物理机(详情2)

网络接口

IP	mac地址	子网掩码	类型	速率
10.127.30.3	b0:83:fe:ea:d1:eb	24	ipmi	100
10.127.30.99	b8:2a:72:e0:ff:26	24	admin	1000
	b8:2a:72:e0:ff:27			1000
	b8:2a:72:e0:ff:28			
	b8:2a:72:e0:ff:29			

品牌信息

名称： Dell Inc.

型号： PowerEdge R720

序列号： HZTJ742

装机实践

介绍采用预注册的方式进行装机的过程

云平台添加物理机信息

云联壹云

主机快照
自动快照策略
网络
安全组
IP子网
弹性公网IP
密钥
密钥
套餐
套餐
基础资源
宿主机
物理机
透传设备
回收站
虚拟机
硬盘
镜像

添加物理机

在添加物理机之前，请确保已经在平台创建物理机所需要的IPMI类型的IP子网和物理机类型的IP子网等

添加方式：ISO引导注册 | PXE引导注册 | 预注册 | 托管

预注册：用于预上架未配置 BMC 信息的服务器，通过预注册功能配置服务器基本信息，待服务器上电后， MAC 信息匹配即可进行注册并配置 BMC 信
服务器处于 DHCP relay 网络环境

指定域：Default

录入方式：单条录入 | 批量录入 | 模板导入

* MAC地址：ec:f4:bb:da:0f:58

请输入物理机管理口的MAC地址，一般为eth0

* 物理机名称：a48

该物理机在系统中显示的名字

IPMI地址：

确 定

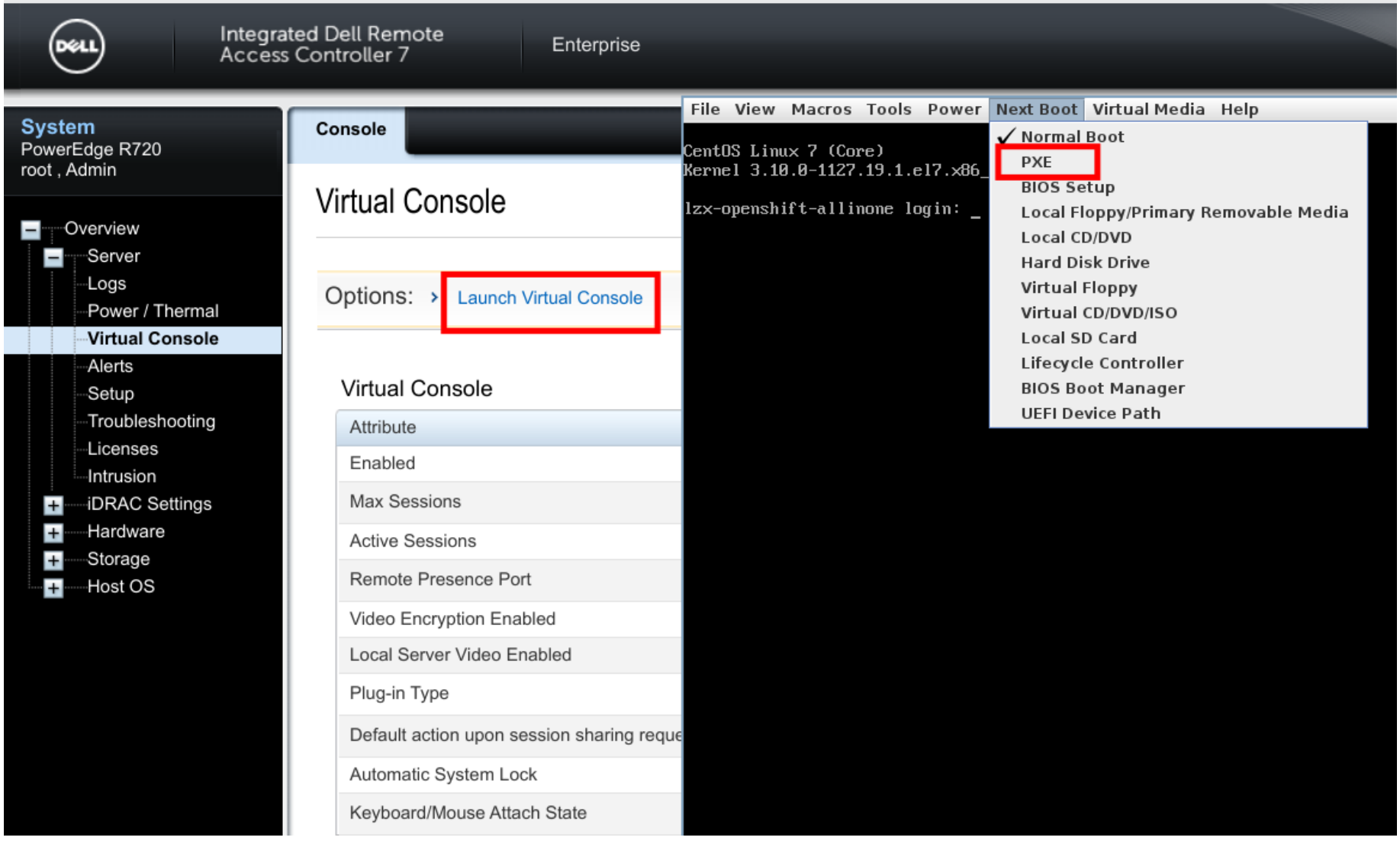
添加筛选项

名称	状态	启用状态	IP	IPMI	初始账号	规格	维护模式	品牌	分配	SN	共享范围	所属域
a48 -	初始化	禁用	10.127.5.48（带外）				正常		-	-	全局共享	Default
a45-1 -	运行中	启用	10.127.100.18（管理） 10.127.5.45（带外）			48C128GRAID	维护模式	DELL	-	62CNF52	全局共享	Default
a30 -	运行中	启用	10.127.100.23（管理） 10.127.5.30（带外）			32C128GRAID	正常	hp	-	6CU3505M2G	全局共享	Default

设置物理机 PXE 启动

```
$ ipmitool -I lanplus -H 10.127.5.48 -U root -P xxx chassis bootdev pxe
Set Boot Device to pxe

$ ipmitool -I lanplus -H 10.127.5.48 -U root -P xxx power reset
```


Integrated Dell Remote Access Controller 7
Enterprise
System
PowerEdge R720
root , Admin
Overview
Server
Logs
Power / Thermal
Virtual Console
Alerts
Setup
Troubleshooting
Licenses
Intrusion
iDRAC Settings
Hardware
Storage
Host OS
Console
Virtual Console
Options:
Launch Virtual Console
Virtual Console
Attribute
Enabled
Max Sessions
Active Sessions
Remote Presence Port
Video Encryption Enabled
Local Server Video Enabled
Plug-in Type
Default action upon session sharing reque
Automatic System Lock
Keyboard/Mouse Attach State
File View Macros Tools Power Next Boot Virtual Media Help
CentOS Linux 7 (Core)
Kernel 3.10.0-1127.19.1.el7.x86_
lzx-openshift-allinone login: _
Normal Boot
PXE
BIOS Setup
Local Floppy/Primary Removable Media
Local CD/DVD
Hard Disk Drive
Virtual Floppy
Virtual CD/DVD/ISO
Local SD Card
Lifecycle Controller
BIOS Boot Manager
UEFI Device Path

物理机 PXE 下载 引导 系统

云联壹云

```
iPXE (http://ipxe.org) 00:04.0 C980 PCI2.10 PnP PMM+BFF94590+BFEF4590 C980



Booting from ROM...
iPXE (PCI 00:04.0) starting execution...ok
iPXE initialising devices...ok



iPXE 1.0.0+ (0418) -- Open Source Network Boot Firmware -- http://ipxe.org
Features: DNS HTTP iSCSI TFTP AoE ELF MBOOT PXE bzImage Menu PXEXT

net0: 00:16:3e:26:75:a2 using 82540em on 0000:00:04.0 (open)
  [Link:up, TX:0 TXE:0 RX:0 RXE:0]
Configuring (net0 00:16:3e:26:75:a2)...... ok
net0: 10.0.4.254/255.255.255.0 gw 10.0.4.1
Next server: 10.168.26.182
Filename: lpxelinux.0
tftp://10.168.26.182/lpxelinux.0... ok
lpxelinux.0 : 75731 bytes [PXE-NBP]

PXELINUX 6.03 lwIP  Copyright (C) 1994-2014 H. Peter Anvin et al
Loading http://10.168.26.182:9879/tftp/kernel... ok
Loading http://10.168.26.182:9879/tftp/initramfs...
```

物理机 PXE 进入 引导 系统

云联壹云

```
[   35.010624] sr 2:0:0:0: Attached scsi generic sg1 type 5
Load fuse ...
[   35.104311] fuse init (API version 7.22)
Initializing random number generator... done.
Starting network: OK
Starting dropbear sshd: OK
Initialize eth0 ...
udhcpc: started, v1.27.2
[   37.495343] e1000: eth0 NIC Link is Up 1000 Mbps Full Duplex, Flow Control: R
X
udhcpc: sending discover
udhcpc: sending select for 10.0.4.254
udhcpc: lease of 10.0.4.254 obtained, lease time 2147483
deleting routers
adding dns 10.168.26.182
Init root password...
passwd: no record of root in /etc/shadow, using /etc/passwd
Changing password for root
New password:
Retype password:
passwd: password for root changed by root
Notify baremetal-agent successfully

Welcome to Buildroot
Yunion login:
```

物理机进入 引导 系统采集信息

添加筛选项

名称	状态	启用状态	IP	IPMI	初始账号	规格	维护模式	品牌	分配
a48 -	准备中	禁用	10.127.100.254（管理） 10.127.5.48（带外）				正常		-
a45-1 -	运行中	启用	10.127.100.18（管理） 10.127.5.45（带外）			48C128GRAID	维护模式		-
a30 -	运行中	启用	10.127.100.23（管理） 10.127.5.30（带外）			32C128GRAID	正常		-

物理机注册完成

添加筛选项

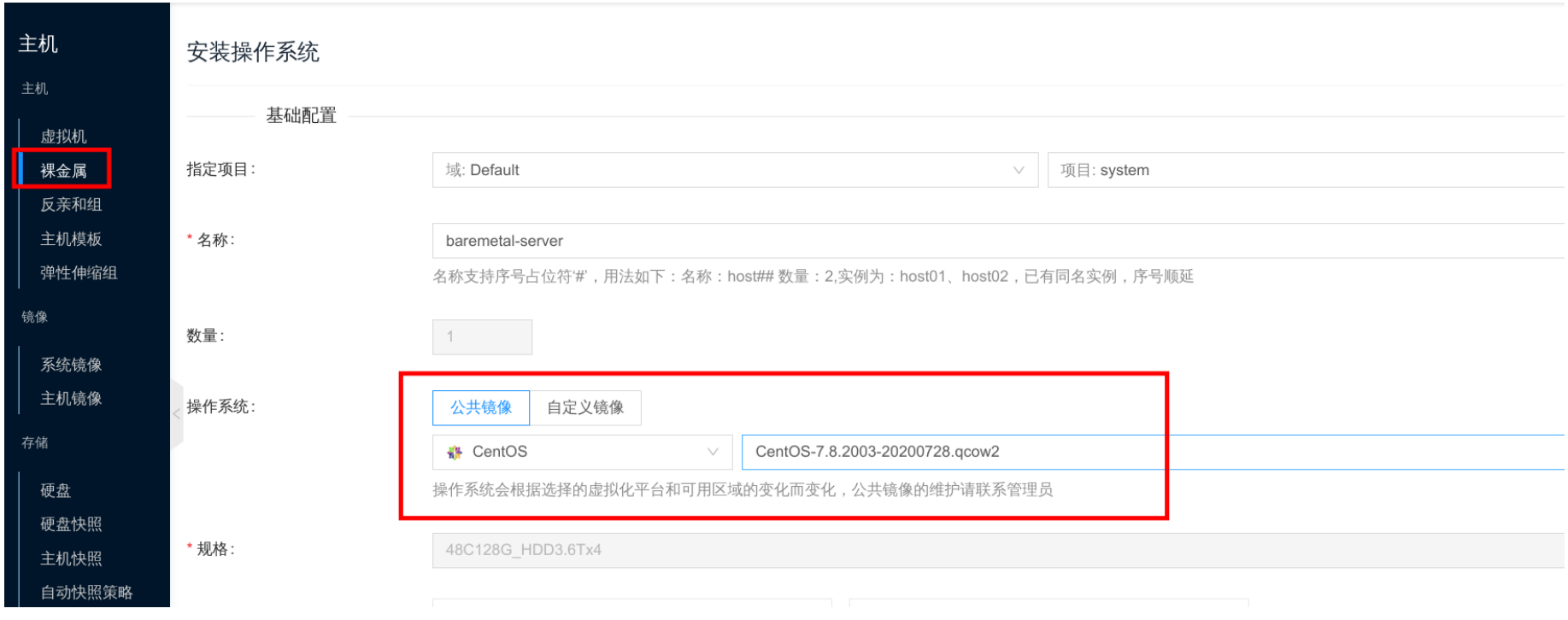
名称	状态	启用状态	IP	IPMI	初始账号	规格	维护模式	品牌	分配	SN
a48 -	运行中	禁用	10.127.100.254（管理） 10.127.5.48（带外）			48C128GRAID	正常	DELL	-	31CNF52
a45-1 -	运行中	启用	10.127.100.18（管理） 10.127.5.45（带外）			48C128GRAID	维护模式	DELL	-	62CNF52
a30 -	运行中	启用	10.127.100.23（管理） 10.127.5.30（带外）			32C128GRAID	正常	hp	-	6CU3505M2G
a9 -	运行中	启用	10.127.100.22（管理） 10.127.5.9（带外）			48C256GRAID	正常	DELL	-	GR3KBD2

安装操作系统

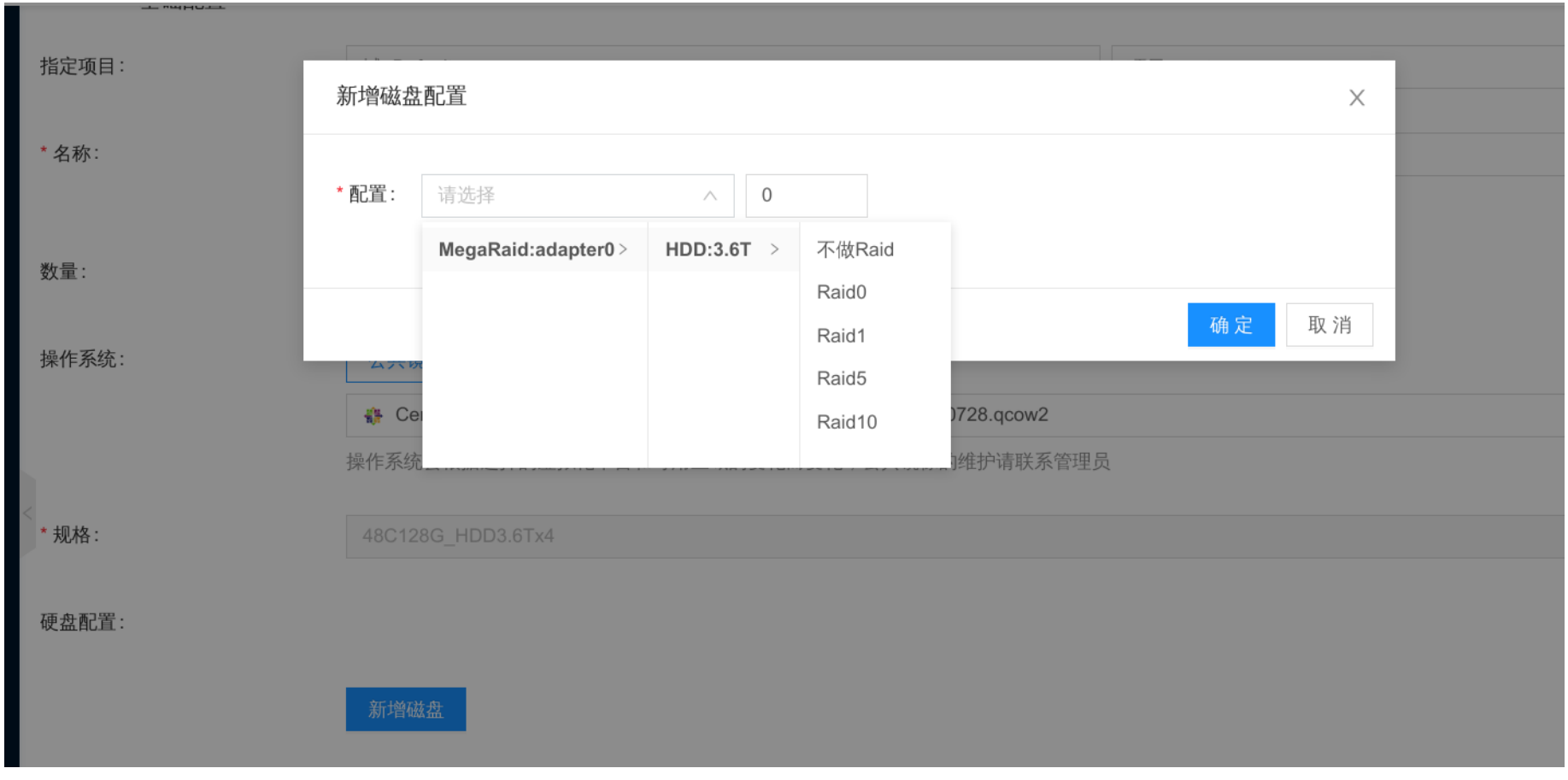
	启用状态	IP	IPMI	初始账号	规格	维护模式	品牌	分配	SN	共享范围	所属域	
中	启用	10.127.100.254（管理） 10.127.5.48（带外）			48C128GRAID	正常	DELL	baremetal-s…	31CNF52	全局共享	Default	
中	启用	10.127.100.18（管理） 10.127.5.45（带外）			48C128GRAID	维护模式	DELL	-	62CNF52	全局共享	Default	
中	启用	10.127.100.23（管理） 10.127.5.30（带外）			32C128GRAID	正常	hp	-	6CU3505M2G	全局共享	Default	
中	启用	10.127.100.22（管理） 10.127.5.9（带外）			48C256GRAID	正常	DELL	-	GR3KBD2	全局共享	Default	更多
		10.127.100.19（管理）										远程终端

更改域
设置共享
安装操作系统
转换为宿主机
同步硬件配置
开机
关机
进入维护模式
退出维护模式
删除

选择镜像


主机
主机
虚拟机
裸金属
反亲和组
主机模板
弹性伸缩组
镜像
系统镜像
主机镜像
存储
硬盘
硬盘快照
主机快照
自动快照策略
安装操作系统
基础配置
指定项目：
域: Default
项目: system
* 名称：
baremetal-server
名称支持序号占位符'#'，用法如下：名称：host## 数量：2,实例为：host01、host02，已有同名实例，序号顺延
数量：
1
操作系统：
公共镜像
自定义镜像
CentOS
CentOS-7.8.2003-20200728.qcow2
操作系统会根据选择的虚拟化平台和可用区域的变化而变化，公共镜像的维护请联系管理员
* 规格：
48C128G_HDD3.6Tx4

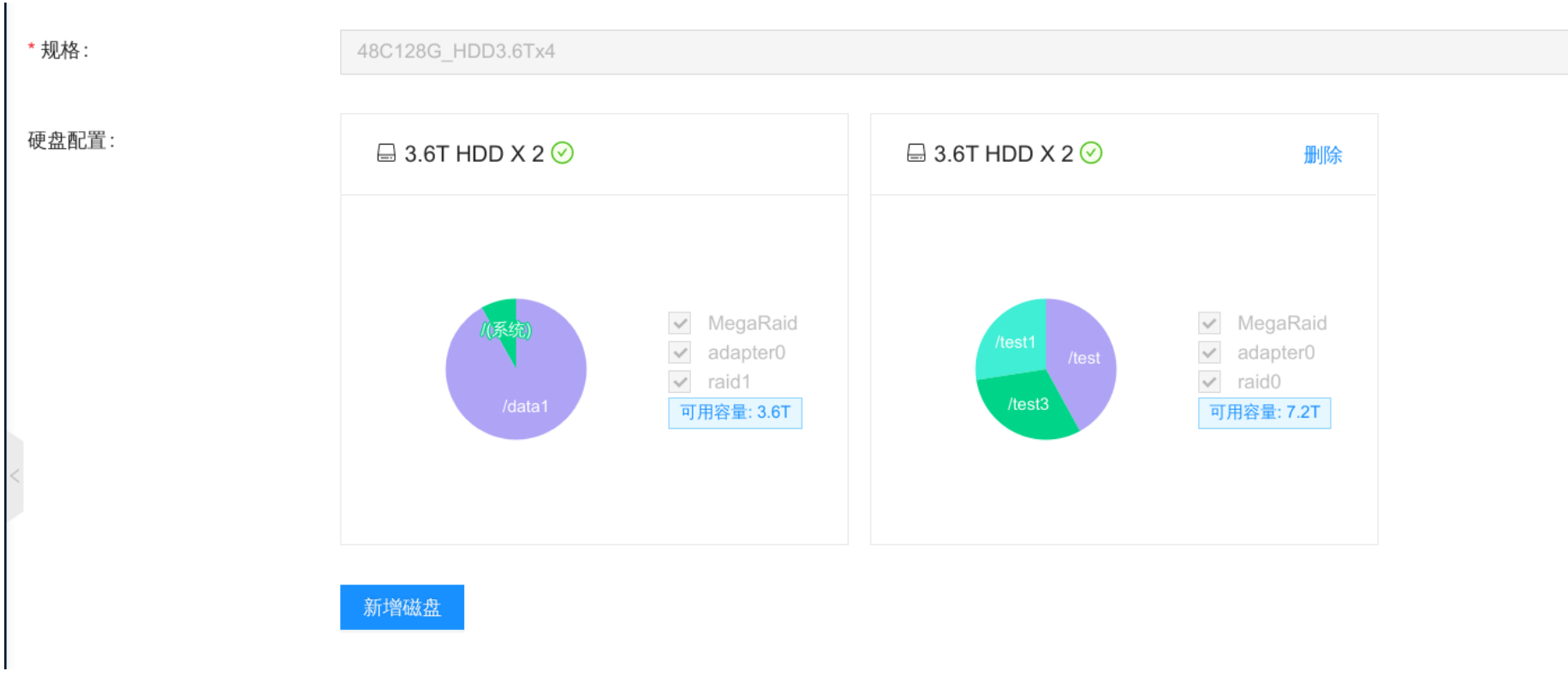
配置 RAID


指定项目：
* 名称：
数量：
操作系统：
* 规格：
48C128G_HDD3.6Tx4
硬盘配置：
新增磁盘
新增磁盘配置
* 配置：
请选择
0
MegaRaid:adapter0
HDD:3.6T
不做Raid
Raid0
Raid1
Raid5
Raid10
确 定
取 消

新建分区

数量:
操作系统:
* 规格:
硬盘配置:
创建新分区
* 挂载点:
/data1
* 分区格式:
ext4
* 分区大小:
最大容量
手动输入
100G
确 定
取 消
adapter0
raid5
可用容量: 558G
剩余

分区结果


* 规格：
48C128G_HDD3.6Tx4
硬盘配置：
3.6T HDD X 2
/(系统)
/data1
MegaRaid
adapter0
raid1
可用容量: 3.6T
3.6T HDD X 2
删除
/test1
/test
/test3
MegaRaid
adapter0
raid0
可用容量: 7.2T
新增磁盘

配置网络

管理员密码：
随机生成
关联密钥
保留镜像设置
手工输入
高级配置
网络：
指定IP子网
指定调度标签
网卡0
DEFAULT-VPC
HOST-NET(10.127.100.2 - 10.127.100.254, vlan=1) 可用: 169
手动配置IP
没有您想要的？可以新建
添加网卡
您还可以添加7个
启用bonding
备注：
名称：baremetal-server
数量：1
区域：------
类型：通用虚拟机
配置：48核CPU、128 GB内存
操作系统：CentOS-7.8.2003-20200728.qcow2
新 建

裸金属记录

裸金属
全部
新建
开机
关机
重启
批量操作
标签
添加筛选项
名称
状态
标签
IP
配置
系统
密码
物理机
项目
区域
操作
baremetal-server
部署中
10.127.100.254(私有)
48C128G4.5T
a48
system
Default
Beijing
Wangjing
远程控制
更多
a51
Baremetal conver...
运行中
10.127.100.248(私有)
0C0B
a51
system
Default
Beijing
Wangjing
远程控制
更多
shangyikun-do...
公司安全测评使用...
运行中
10.127.100.247(私有)
24C128G3.6T
a54
system
Default
Beijing
Wangjing
远程控制
更多

装机完成

添加筛选项

名称	状态	标签	IP	配置	系统	密码	物理机	项目	区域	操作
baremetal-server -	运行中		10.127.100.254(私有)	48C128G11T			a48	system Default	Beijing Wangjing	远程控制 更多
a51 Baremetal conver...	运行中		10.127.100.248(私有)	0C0B			a51	system Default	Beijing Wangjing	远程控制 更多
shangyikun-do... 公司安全测评使用...	运行中		10.127.100.247(私有)	24C128G3.6T			a54	system Default	Beijing Wangjing	远程控制 更多

获取登陆信息

裸金属
全部
新 建
开 机
关 机
重 启
批量操作
标签
添加筛选项
名称
状态
标签
IP
配置
用户名：root
密 码：crP7JFjPA43y
项目
区域
操作
baremetal-server
运行中
10.127.100.254(私有)
48C128G11T
a48
system
Default
Beijing
Wangjing
远程控制
更多
a51
Baremetal conver...
运行中
10.127.100.248(私有)
0C0B
a51
system
Default
Beijing
Wangjing
远程控制
更多

SSH 登录裸金属

```
$ ssh root@10.127.100.254
Warning: Permanently added '10.127.100.254' (ECDSA) to the list of known hosts.
root@10.127.100.254's password:
X11 forwarding request failed
Last login: Tue Jul 28 18:14:14 2020
[root@baremetal-server ~]# cat /etc/fstab
UUID=01326bbd-064f-497c-9362-a505490597d3       /       xfs     defaults        0       0
UUID=903162f3-aee6-4d7e-8c0d-f2235710087c       /data1  ext4    defaults        2       2
UUID=c6db6a70-6839-4a84-8195-cc443b72784d       /test3  ext4    defaults        2       2
UUID=86312e92-c411-4cbf-8220-a1e2935a6222       /test1  ext4    defaults        2       2
UUID=a8bd75c0-9474-48e3-8668-e302d78a9708       /test   ext4    defaults        2       2
[root@baremetal-server ~]# lsblk
NAME   MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda      8:0    0  3.7T  0 disk
├sda1    8:1    0    1M  0 part
├sda2    8:2    0  300G  0 part /
└sda3    8:3    0  3.4T  0 part /data1
sdb      8:16   0  7.3T  0 disk
├sdb1    8:17   0  2.2T  0 part /test3
├sdb2    8:18   0    2T  0 part /test1
└sdb3    8:19   0  3.1T  0 part /test
[root@baremetal-server ~]#
```

常见问题

云联壹云

1. 开源版本怎么启用物理机管理服务？

    - 参考文档: https://docs.yunion.io/docs/setup/baremetal/

2. 装机镜像从哪里下载？

    - 公网下载地址: https://iso.yunion.cn/vm-images/
    - 镜像制作脚本: https://github.com/yunionio/service-images

3. 有哪些限制？

    - 装机镜像
        - 不支持 lvm 的分区
        - 不支持 windows 系统
        - 文件系统仅限 ext4, xfs
    - 暂时不支持 UEFI 引导

4. PXE 引导系统 怎么制作的？

    - 使用 buildroot (https://buildroot.org/) 做的精简 linux
    - 制作代码在: https://github.com/yunionio/yunionos

5. 支持哪些 raid 控制器?

    - 支持多种主流 RAID 控制器 (MegaRaid、 HP SmartArray、 MPT2/3Sas、 MarvelRaid)

未来计划

- 完善物理机开源文档: https://docs.yunion.io/
- 物理机硬件监控
- Baremetal Agent 服务无状态改造

云联壹云

Q & A

// reveal.js plugins