

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
[-取消关注](#)

(案例一) Proxmox VE + Ceph + 物理网络 超融合生产环境部署案例

参考: <https://blog.csdn.net/hanziqing0630/article/details/114262035>

一、Proxmox Virtual Environment介绍

Proxmox VE 是用于企业虚拟化的开源服务器管理平台。它在单个平台上紧密集成了KVM虚拟机管理程序和LXC，软件定义的存储以及网络功能。借助基于Web的集成用户界面，您可以轻松管理VM和容器，高可用性群集或集成的灾难恢复工具。

同时Proxmox VE对接Proxmox备份服务器，可实现对VM、容器的全量和增量备份，可显著减少网络负载并节省宝贵的存储空间。

二、环境介绍

本文章所述环境采用3台物理服务器（含4口万兆网卡），1台万兆业务交换机，1台万兆管理交换机。

注意：由于使用ceph搭建分布式存储，为保证性能，管理网络及业务网络强烈建议采用全万兆部署。

2.1 软件版本

本次部署采用当前最新版本：Proxmox VE 6.3，可根据当时情况选择最新版本。

2.2 服务器配置

型号	配置	数量
DELL R930	CPU : Intel® Xeon® CPU E7-8870 v4 @ 2.10GHz 内存 : 512GB RAID卡 : PERC H730P Adapter SSD : 1.5TB 4块 HDD : 1.2TB 10块	3

三、软件获取

访问官网下载ISO镜像文件，根据需要选择版本：
<https://www.proxmox.com/en/downloads/category/iso-images-pve>

四、RAID规划

4.1 系统盘

- 建议采用两块SSD组成raid 1，作为系统盘。

不同品牌服务器设置方法不同，请参照各品牌官方文档。

4.2 ceph盘

- ceph不推荐基于raid盘创建osd，所以除系统盘外，其余磁盘全部设置为非raid模式。

不同品牌服务器设置方法不同，请参照各品牌官方文档。

五、物理网络

生产环境对可靠性要求较高，所以建议采用链路聚合方式，将多个物理网卡绑定为一个逻辑接口，从而提高网络的性能及可靠性。

链路聚合可在以下三种模式中选择：

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

最新随笔

- 1.Ceph 故障域配置（修改CRUSH MAP）
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类（Octopus版本）

随笔分类

[Alpine\(2\)](#)
[Ansible\(23\)](#)
[CentOS\(23\)](#)
[Ceph\(118\)](#)
[Debian\(19\)](#)
[Docker\(28\)](#)
[ELK日志分析\(15\)](#)
[Helm\(3\)](#)
[Java\(1\)](#)
[Jenkins\(10\)](#)
[K8s\(209\)](#)
[KubectI\(4\)](#)
[Linux集群技术\(3\)](#)
[Linux企业常用服务\(9\)](#)
[Linux企业级安全\(5\)](#)
[Linux系统调优和监控\(8\)](#)
[Linux系统基础及管理\(38\)](#)
[MongoDB\(12\)](#)
[MySQL/MariaDB\(6\)](#)
[NFS\(3\)](#)
[PostgreSQL\(1\)](#)

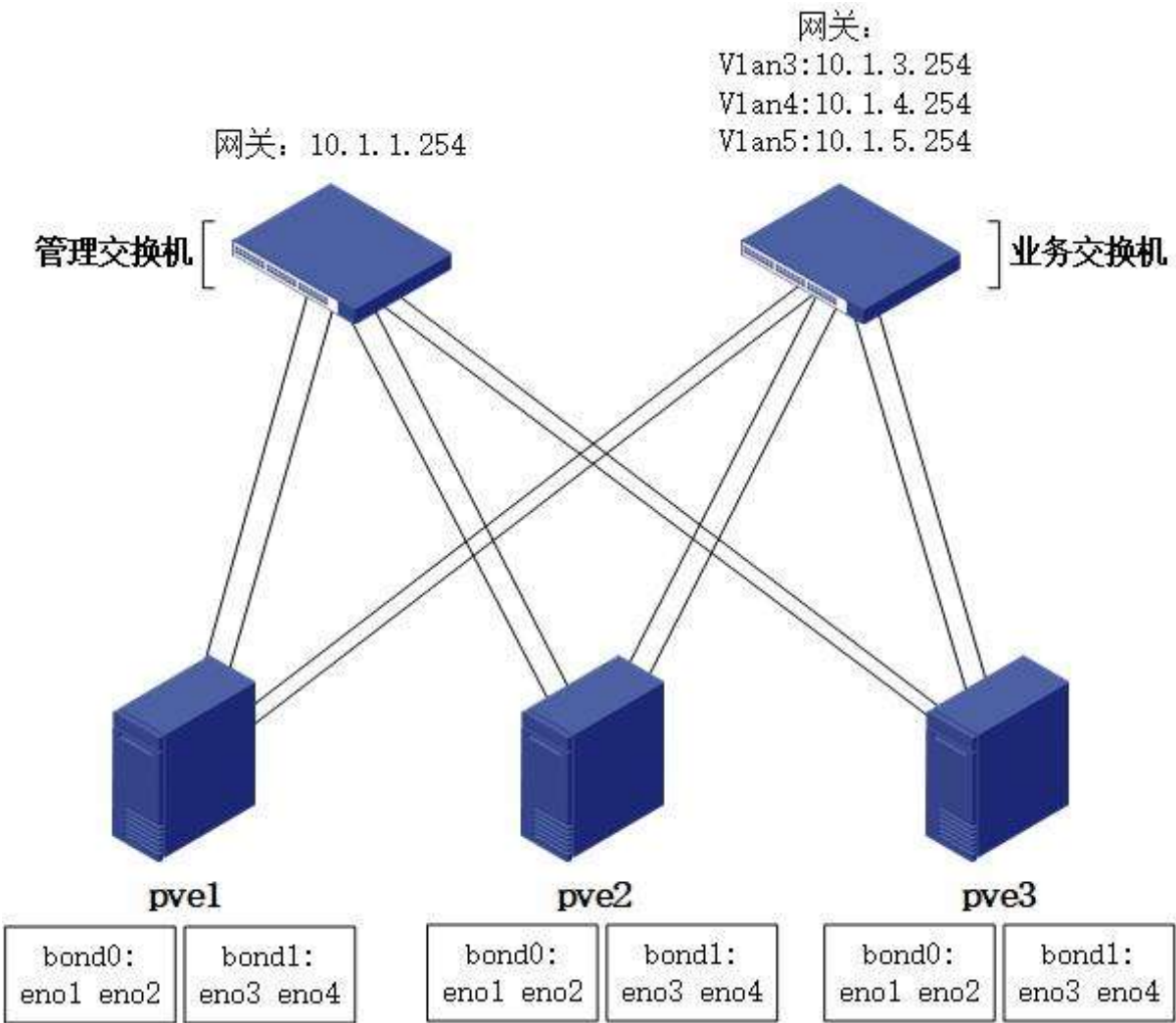
- LACP（802.3ad）：基于LACP协议的链路聚合，此模式交换机需配置链路聚合，模式为lACP。
- 静态聚合：静态链路聚合，此模式交换机需配置链路聚合，模式为静态（华为、H3C默认为此模式；思科为on模式）。
- active-backup：主备模式，网卡一主一备，主网卡故障时备份网卡接替主网卡工作，此模式交换机无需做链路聚合配置。

若交换机支持LACP，推荐采用LACP模式。否则你通常应该使用 active-backup 模式。此处采用LACP模式进行演示。

如果您打算在绑定接口上运行PVE集群网络，那么您必须在绑定接口上使用 active-passive 模式，其他模式是不支持的。

5.1 网络规划

5.1.1 物理网络拓扑



5.1.2 vlan及网段规划

vlan号	用途	IP网段	网关
1	管理	10.1.1.0/24	10.1.1.254
3	业务	10.1.3.0/24	10.1.3.254
4	业务	10.1.4.0/24	10.1.4.254
5	业务	10.1.5.0/24	10.1.5.254

5.2 网络配置

以华为交换机为例。

5.2 网络配置

以华为交换机为例。

5.2.1 创建vlan：

```
1 | vlan batch 3 to 5
```

5.2.2 创建vlan-if（网关）：

```
1 interface Vlanif1 //vlan1, 管理网络网关
2 ip address 10.1.1.254 255.255.255.0
3 interface Vlanif3 //vlan3, 业务网络网关
4 ip address 10.1.3.254 255.255.255.0
5 interface Vlanif4 //vlan4, 业务网络网关
6 ip address 10.1.4.254 255.255.255.0
7 interface Vlanif5 //vlan5, 业务网络网关
8 ip address 10.1.5.254 255.255.255.0
```

[Prometheus\(52\)](#)
[Rancher\(44\)](#)
[Redis\(1\)](#)
[ServiceMesh\(1\)](#)
[Ubuntu\(18\)](#)
[Windows\(2\)](#)
[Zabbix\(1\)](#)
[常用工具\(24\)](#)
[代码管理\(5\)](#)
[更多](#)

随笔档案

[2022年10月\(4\)](#)
[2022年6月\(3\)](#)
[2022年5月\(17\)](#)
[2022年4月\(25\)](#)
[2022年3月\(68\)](#)
[2022年2月\(18\)](#)
[2021年12月\(1\)](#)
[2021年11月\(1\)](#)
[2021年10月\(2\)](#)
[2021年9月\(50\)](#)
[2021年8月\(273\)](#)
[2021年7月\(79\)](#)
[2020年11月\(58\)](#)
[2020年10月\(58\)](#)

阅读排行榜

1. CentOS7安装源配置详解(9876)
2. leaf-spine（叶脊）网络架构介绍(7105)
3. PVE 备份和恢复介绍(5386)
4. Proxmox Virtual Environment 介绍(4947)
5. Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
6. Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
7. PVE 虚拟机模板和克隆介绍(3936)
8. 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
9. CentOS7安装配置远程桌面详解(3631)
10. Debian安装源配置详解(3457)
11. Apollo配置中心使用详解(3356)
12. Proxmox VE 防火墙介绍(3323)
13. Gitea 简单使用说明(3313)
14. Linux系统路由配置说明(3270)
15. Chrony时间同步服务器安装配置说明(3256)
16. Centos7防火墙配置详解(3231)
17. 【Rancher】简单介绍(3230)
18. Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
19. PVE 超融合架构介绍(3096)
20. 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

1. Proxmox VE 存储介绍(1)
2. Proxmox Virtual Environment 介绍(1)
3. Debian Jessie Preseed自动化安装示例(1)
4. Debian live-build官方教程学习笔记(1)
5. Ceph存储池pg_num配置详解(1)

推荐排行榜

1. leaf-spine（叶脊）网络架构介绍(1)
2. Proxmox VE 防火墙介绍(1)
3. Proxmox Virtual Environment 介绍(1)
4. VIM 常用快捷键(1)
5. Debian7 bonding配置详解(1)

最新评论

1. Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从"Default (No cache)"修改为"Write back (unsafe)", 对磁盘数据有影响么?
--大熊猫子
2. Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的EMC FC-SAN存储（不支持ISCSI），如果改用X做集群，FC-SAN应该以什么方式挂载给PROXMOX好呢?

5.2.3 创建聚合接口 (eth-trunk) :

```
1 interface Eth-Trunk101    //用于节点1管理口
2   description to-pve1-mgmt
3   port link-type access
4   mode lacp
5   #
6 interface Eth-Trunk102    //用于节点2管理口
7   description to-pve2-mgmt
8   port link-type access
9   mode lacp
10  #
11 interface Eth-Trunk103    //用于节点3管理口
12   description to-pve3-mgmt
13   port link-type access
14   mode lacp
15  #
16 interface Eth-Trunk104    //用于节点1业务口
17   description to-pve1-yewu
18   port link-type trunk
19   port trunk allow-pass vlan 2 to 4094
20   mode lacp
21  #
22 interface Eth-Trunk105    //用于节点2业务口
23   description to-pve2-yewu
24   port link-type trunk
25   port trunk allow-pass vlan 2 to 4094
26   mode lacp
27  #
28 interface Eth-Trunk106    //用于节点3业务口
29   description to-pve3-yewu
30   port link-type trunk
31   port trunk allow-pass vlan 2 to 4094
32   mode lacp
```

5.2.4 物理接口加入聚合接口

```
1 interface GigabitEthernet0/0/10    //连接节点1管理口eno1
2   description pve1-mgmt-eno1
3   eth-trunk 101
4   #
5 interface GigabitEthernet0/0/11    //连接节点1管理口eno2
6   description pve1-mgmt-eno2
7   eth-trunk 101
8   #
9 interface GigabitEthernet0/0/12    //连接节点2管理口eno1
10  description pve2-mgmt-eno1
11  eth-trunk 102
12  #
13 interface GigabitEthernet0/0/13    //连接节点2管理口eno2
14  description pve2-mgmt-eno2
15  eth-trunk 102
16  #
17 interface GigabitEthernet0/0/14    //连接节点3管理口eno1
18  description pve3-mgmt-eno1
19  eth-trunk 103
20  #
21 interface GigabitEthernet0/0/15    //连接节点3管理口eno2
22  description pve3-mgmt-eno2
23  eth-trunk 103
24  #
25 interface GigabitEthernet0/0/16    //连接节点1业务口eno3
26  description pve1-data-eno3
27  eth-trunk 104
28  #
29 interface GigabitEthernet0/0/17    //连接节点1业务口eno4
30  description pve1-data-eno4
31  eth-trunk 104
32  #
33 interface GigabitEthernet0/0/18    //连接节点2业务口eno3
34  description pve2-data-eno3
35  eth-trunk 105
36  #
37 interface GigabitEthernet0/0/19    //连接节点2业务口eno4
38  description pve2-data-eno4
39  eth-trunk 105
40  #
41 interface GigabitEthernet0/0/20    //连接节点3业务口eno3
42  description pve3-data-eno3
43  eth-trunk 106
44  #
45 interface GigabitEthernet0/0/21    //连接节点3业务口eno4
46  description pve3-data-eno4
47  eth-trunk 106
```

六、 PVE安装

6.1 选择第一个选项， 安装PVE

- [3. Re:Proxmox Virtual Environment 介绍](#)
很全面!
--willblog
- [4. Re:Ceph存储池pg_num配置详解](#)
mark
--穷死了
- [5. Re:Debian live-build官方教程学习笔记](#)
执行成功之后为什么没找看iso
--abraham_liuhai



Welcome to Proxmox Virtual Environment

Install Proxmox VE

Install Proxmox VE (Debug mode)

Rescue Boot

Test memory (Legacy BIOS)

6.2 同意协议



END USER LICENSE AGREEMENT (EULA)

END USER LICENSE AGREEMENT (EULA) FOR PROXMOX VIRTUAL ENVIRONMENT (PROXMOX VE)

By using Proxmox VE software you agree that you accept this EULA, and that you have read and understand the terms and conditions. This also applies for individuals acting on behalf of entities. This EULA does not provide any rights to Support Subscriptions Services as software maintenance, updates and support. Please review the Support Subscriptions Agreements for these terms and conditions. The EULA applies to any version of Proxmox VE and any related update, source code and structure (the Programs), regardless of the the delivery mechanism.

1. License. Proxmox Server Solutions GmbH (Proxmox) grants to you a perpetual, worldwide license to the Programs pursuant to the GNU Affero General Public License V3. The license agreement for each component is located in the software component's source code and permits you to run, copy, modify, and redistribute the software component (certain obligations in some cases), both in source code and binary code forms, with the exception of certain binary only firmware components and the Proxmox images (e.g. Proxmox logo). The license rights for the binary only firmware components are located within the components. This EULA pertains solely to the Programs and does not limit your rights under, or grant you rights that supersede, the license terms of any particular component.

2. Limited Warranty. The Programs and the components are provided and licensed "as is" without warranty of any kind, expressed or implied, including the implied warranties of merchantability, non-infringement or fitness for a particular purpose. Neither Proxmox nor its affiliates warrants that the functions contained in the Programs will meet your requirements or that the operation of the Programs will be entirely error free, appear or perform precisely as described in the accompanying documentation, or comply with regulatory requirements.

3. Limitation of Liability. To the maximum extent permitted under applicable law, under no

Abort

Previous

I agree

6.3 选择系统盘

注意：此处一定要选择RAID规划中创建的RAID 1系统盘。



Proxmox Virtualization Environment (PVE)

- The Proxmox Installer** automatically partitions your hard disk. It installs all required packages and finally makes the system bootable from hard disk. All existing partitions and data will be lost.

Press the Next button to continue installation.
- Please verify the installation target**
The displayed hard disk is used for installation.
Warning: All existing partitions and data will be lost.
 - Automatic hardware detection**
The installer automatically configures your hardware.
 - Graphical user interface**
Final configuration will be done on the graphical user interface via a web browser.

①

Target Harddisk: /dev/sda (100GB, Virtual disk) Options

②

Abort Previous Next

6.4 配置国家、地区



Location and Time Zone selection

- The Proxmox Installer** automatically makes location-based optimizations, like choosing the nearest mirror to download files from. Also make sure to select the correct time zone and keyboard layout.

Press the Next button to continue the installation.
- Country:** The selected country is used to choose nearby mirror servers. This will speed up downloads and make updates more reliable.
 - Time Zone:** Automatically adjust daylight saving time.
 - Keyboard Layout:** Choose your keyboard layout.

Country China

Time zone Asia/Shanghai

Keyboard Layout U.S. English

Abort Previous Next

6.5 配置密码、邮箱

如果为生产环境务必配置为强口令。邮箱建议配置为真实邮箱。



Administration Password and Email Address

- Proxmox Virtual Environment** is a full featured highly secure GNU/Linux system based on Debian.

Please provide the *root* password in this step.
- Password:** Please use a strong password. It should have 8 or more characters. Also combine letters, numbers, and symbols.
 - Email:** Enter a valid email address. Your Proxmox VE server will send important alert notifications to this email account (such as backup failures, high availability events, etc.).
- Press the Next button to continue installation.

Password

Confirm

Email

••••••

••••••

••••••@qq.com

Abort

Previous

Next

6.6 配置网卡、主机名、IP地址、DNS等信息

- 此处选择服务器第一块网卡，用于管理网络，按我们的规划应为eno1
- 每个节点主机名不能相同，主机名可以参考以下规划，域名请按需配置。

序号	主机名	域名
1	pve1	test.local
2	pve2	test.local
3	pve3	test.local

- 主机名、IP地址等信息按实际情况配置，配置公网DNS，用于pve连接互联网。

•

•

•

Please verify the displayed network configuration. You will need a valid network configuration to access the management interface after installing.

After you have finished, press the Next button. You will be shown a list of the options that you chose during the previous steps.

- IP address (CIDR):** Set the main IP address and netmask for your server in CIDR notation.
- Gateway:** IP address of your gateway or firewall.
- DNS Server:** IP address of your DNS server.

Management Interface:

Hostname (FQDN):

IP Address (CIDR)

Gateway:

DNS Server:

ens192 - 00:50:56:b0:ec:59 (vmxnet3)

pve1.test.local

10.1.1.101 / 24

10.1.1.254

114.114.114.114

Abort

Previous

Next

6.7 检查无误后点击安装



Summary

Please confirm the displayed information. Once you press the **Install** button, the installer will begin to partition your drive(s) and extract the required files.

Option	Value
Filesystem:	ext4
Disk(s):	/dev/sda
Country:	China
Timezone:	Asia/Shanghai
Keymap:	en-us
Email:	735273287@qq.com
Management Interface:	ens192
Hostname:	pve1
IP CIDR:	10.1.1.101/24
Gateway:	10.1.1.254
DNS:	114.114.114.114

☒ Automatically reboot after successful installation

Abort

Previous

Install

安装结束后重启，通过浏览器访问，出现以下页面内容证明安装成功。

```
地址: https://你的IP地址:8006
用户名: root
密码: 为安装时配置的密码。
域: 选择Linux PAM
```

6.8 按照此方法安装全部服务器

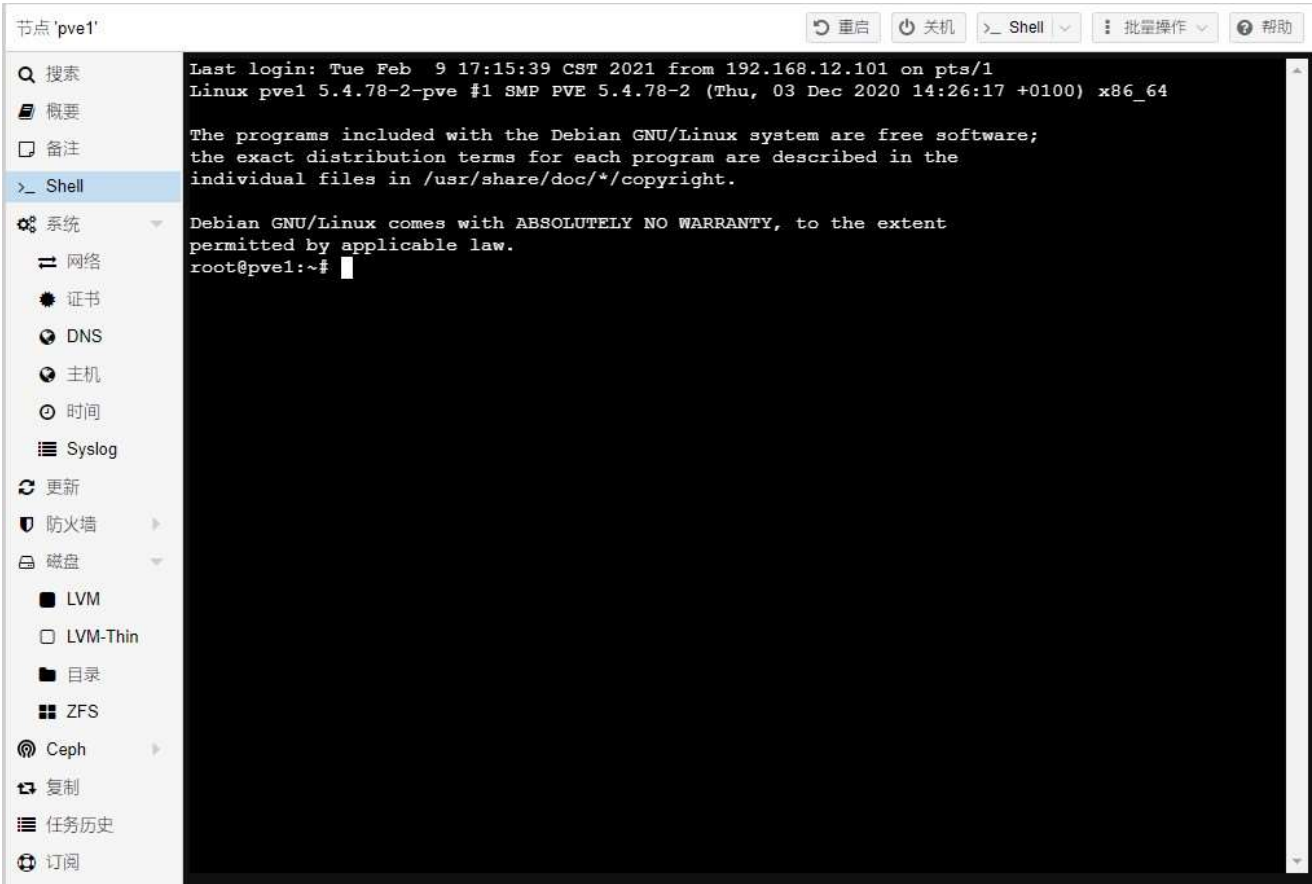
七、更换软件源（三个节点都需要配置）

由于自带的软件源速度较慢，此处选择科大源：<http://mirrors.ustc.edu.cn>

涉及的软件源有三个，分别为debain、pve、ceph，需要分别修改下列文件。

名称	source文件路径
debain	/etc/apt/sources.list
pve	/etc/apt/sources.list.d/pve-no-subscription.list
ceph	/etc/apt/sources.list.d/ceph.list

7.1 通过web管理页面中的shell或ssh登录后台



7.2 更换debian源

```
vi /etc/apt/sources.list
```


删除默认的源，替换为：

```
deb http://mirrors.ustc.edu.cn/debian buster main contrib non-free
# deb-src http://mirrors.ustc.edu.cn/debian buster main contrib non-free
deb http://mirrors.ustc.edu.cn/debian buster-updates main contrib non-free
# deb-src http://mirrors.ustc.edu.cn/debian buster-updates main contrib non-free

deb http://mirrors.ustc.edu.cn/debian-security/ buster-security main non-free contrib
# deb-src http://mirrors.ustc.edu.cn/debian-security/ buster-security main non-free
contrib
```

7.3 更换pve源为非订阅源

```
mv /etc/apt/sources.list.d/pve-enterprise.list /etc/apt/sources.list.d/pve-
enterprise.list.bak
CODENAME=`cat /etc/os-release |grep CODENAME |cut -f 2 -d "="`
echo "deb http://mirrors.ustc.edu.cn/proxmox/debian $CODENAME pve-no-subscription" >
/etc/apt/sources.list.d/pve-no-subscription.list
```

7.4 添加 Ceph 镜像源

注意：因后面 PVE 安装 Ceph 可能会使用官方源覆盖 /etc/apt/sources.list.d/ceph.list 这个文件的内容，此处使用 ceph_mirror.list 这个文件替代。

创建 ceph_mirror.list 文件：

```
vi /etc/apt/sources.list.d/ceph_mirror.list
```

增加以下内容：

```
deb http://mirrors.ustc.edu.cn/proxmox/debian/ceph-octopus/ buster main
```

八、更新PVE至最新版本（按需）

根据实际需求决定是否更新为最新版本。

如需更新三台节点都需要执行：

```
apt upgrade
```

九、基础配置（三个节点都需要配置）

9.1 安装ifupdown2

此组件用于配置网卡后重启网卡操作：

```
apt install ifupdown2
```

9.2 安装vim

```
apt install vim
```

9.3 配置ntp服务

配置统一的时钟源

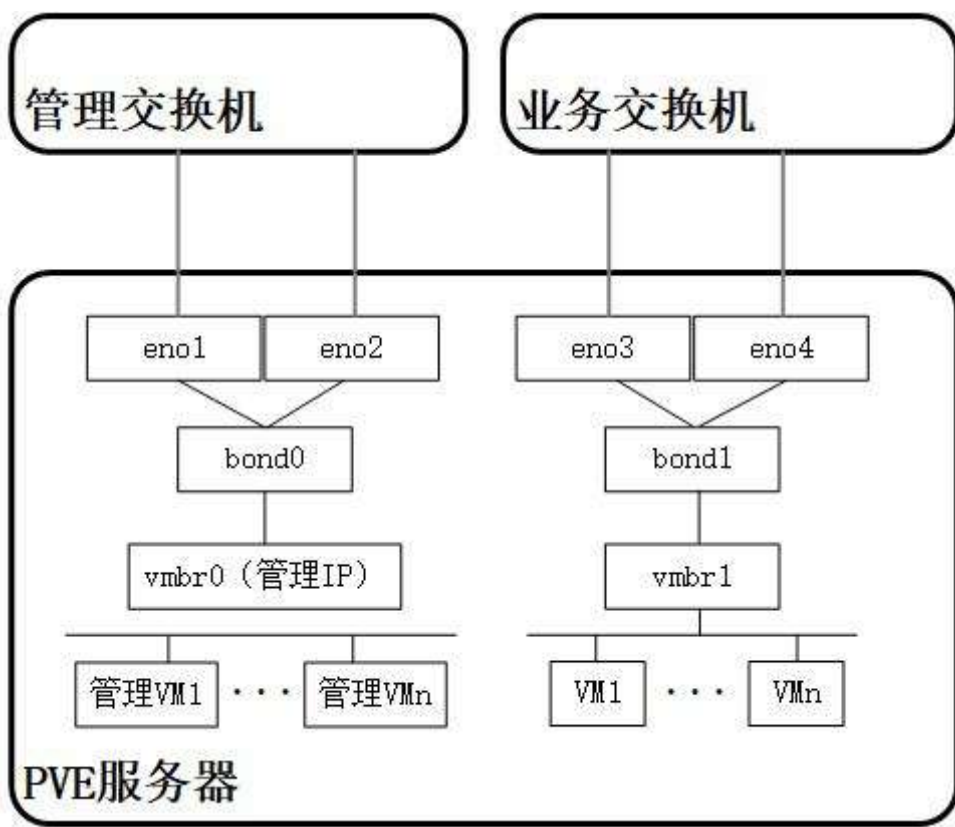
- 如果有条件建议在内网部署一台ntp服务器，所有节点与内网ntp同步时钟。
- 如果没有内网NTP服务器，可以采用下列NTP服务器。

```
阿里云：ntp.aliyun.com
国家授时中心：ntp.ntsc.ac.cn
```

十、PVE网络配置（三个节点都需要配置）

10.1 PVE服务器网络拓扑

放一张图有助于理解配置，可以结合5.1.1章节拓扑图理解。



10.2 管理网络

10.2.1 编辑vmbr0

移除eno1，以便于将eno1加入到bond0中，删除桥接端口中的eno1，然后点击OK。

编辑: Linux Bridge

名称: vmbr0

IPv4/CIDR: 10.1.2.101/24

网关 (IPv4): 10.1.2.254

IPv6/CIDR:

网关 (IPv6):

自动启动: ☒

VLAN感知: ☒

桥接端口: eno0

备注:

MTU: 1500

高级 ☒ OK Reset

10.2.2 创建bond0

点击创建Linux Bond。
名称: bond0
slaves: eno1 eno2
模式: LACP (802.3ad)
Hash策略: layer2+3
按以上参数配置，然后点击OK。

编辑: Linux Bond

名称: bond0

IPv4/CIDR:

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

自动启动: ☒

Slaves: eno1 eno2

模式: LACP (802.3ad)

Hash策略: layer2+3

主bond:

备注:

MTU: 1500

高级 ☒ OK Reset

10.2.3 编辑vmbr0

桥接端口配置: bond0。
选中“vlan感知”选项，然后点击OK。

编辑: Linux Bridge

名称:vmbr0

IPv4/CIDR:10.1.2.101/24

网关 (IPv4):10.1.2.254

IPv6/CIDR:

网关 (IPv6):

自动启动:☒

VLAN感知:☒

桥接端口:bond0

备注:

MTU:1500

高级☒

OK

Reset

10.2.4 应用配置

如果感觉配置有问题可以点击还原按钮，还原为配置前状态。
如果没有问题，点击应用配置使配置生效。
网络配置不正确可能会导致无法管理节点，应用配置前请确认配置正确，如果出现无法管理的情况需要进入shell手动修改配置。

10.3 业务网络

业务网络配置与管理网络类似。

10.3.1 创建bond1

点击创建Linux Bond。
名称：bond1
slaves：eno3 eno4
模式：LACP (802.3ad)
Hash策略：layer2+3
按以上参数配置，然后点击OK。

编辑: Linux Bond

名称:bond1

IPv4/CIDR:

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

自动启动:☒

Slaves:eno3 eno4

模式:LACP (802.3ad)

Hash策略:layer2+3

主bond:

备注:

MTU:1500

高级☒

OK

Reset

10.3.2 创建vmbr1

桥接端口配置：bond1。
选中“vlan感知”选项，然后点击OK。

编辑: Linux Bridge

名称:vmbr1

IPv4/CIDR:

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

自动启动:☒

VLAN感知:☒

桥接端口:bond1

备注:

MTU:1500

高级☒

OK

Reset

10.3.3 应用配置

点击应用配置使配置生效。

10.4 网络配置检查

在交换机中检查聚合接口状态是否正常。

```
1 display eth-trunk 101
2 display eth-trunk 102
3 display eth-trunk 103
4 display eth-trunk 104
5 display eth-trunk 105
6 display eth-trunk 106
```

查看所有eth-trunk接口状态，Status为Selected，则聚合接口正常。

```
[TOR-4]display eth-trunk 101
Eth-Trunk101's state information is:
Local:
LAG ID: 101                WorkingMode: LACP
Preempt Delay: Disabled    Hash arithmetic: According to SIP-XOR-DIP
System Priority: 32768     System ID: 488e-efb5-3d70
Least Active-linknumber: 1 Max Active-linknumber: 8
Operate status: up        Number Of Up Port In Trunk: 2

-----
ActorPortName      Status  PortType PortPri PortNo PortKey PortState Weight
GigabitEthernet0/0/10 Selected 1GE      32768  15    25905  10111100  1
GigabitEthernet0/0/11 Selected 1GE      32768  16    25905  10111100  1

Partner:
-----
ActorPortName      SysPri  SystemID      PortPri PortNo PortKey PortState
GigabitEthernet0/0/10 65535   2004-0fef-b428 255    1     9      10111100
GigabitEthernet0/0/11 65535   2004-0fef-b428 255    2     9      10111100
```

十一、集群组建

创建集群，将三台节点加入同一集群。

11.1 在其中一个节点上点击，“数据中心” — “集群” — “创建集群”。

11.2 点击加入信息，复制加入信息。



11.3 在其他两个节点上，点击加入集群，粘贴加入信息，然后输入第一台节点的root密码，即可完成加入。



11.4 加入集群完成后需要刷新页面才能正常使用，全部加入完成后，就可以在一个节点上管理集群中所有节点了。

数据中心				
Q 搜索 🏠 概览 📁 集群 🌐 Ceph ⚙️ 选项 💾 存储 📁 备份 🔄 复制 🔒 权限	集群信息			
	创建集群加入信息加入集群			
	集群名称: 配置版本: 3节点数: 3			
	集群节点			
	节点名称	ID ↑	表决	链接0
	pve2	1	1	10.1.2.102
	pve1	2	1	10.1.2.101
	pve3	3	1	10.1.2.103

十二、ceph

Ceph是一种为优秀的性能、可靠性和可扩展性而设计的统一的、分布式文件系统。它可以将多台服务器中的硬盘聚合成一个集群，对外可提供文件存储、块存储、对象存储等多种形式，ceph的另一个特点是数据采用多副本方式存储，避免由于单点故障造成的业务中断和数据丢失。

PVE自带了ceph的图形化安装、管理工具， 集群搭建完成后，可以快速完成ceph的安装、配置。同时pve也提供命令行工具，可以使用命令行对ceph进行自定义配置。

12.1 ceph磁盘规划

我的环境中，sda到sdh为8块HDD硬盘，sdi到sdl为4块SSD硬盘。

为提高hdd的性能，我们将sdi作为所有hdd硬盘的DB盘，将sdj作为所有hdd硬盘的WAL磁盘。

请注意DB盘和WAL盘不能相同，如果只有一块SSD可以将SSD分为两个区，分别作为DB盘和WAL盘。
分区建议：根据PVE文档建议，DB盘容量=数据盘总容量10%，WAL盘容量=数据盘总容量1%，假设数据盘共8T，则DB分区容量需>800G，WAL分区容量需>80G，考虑到单位换算及损耗，需要适当增加容量，请根据实际情况规划。

- 详细规划如下：

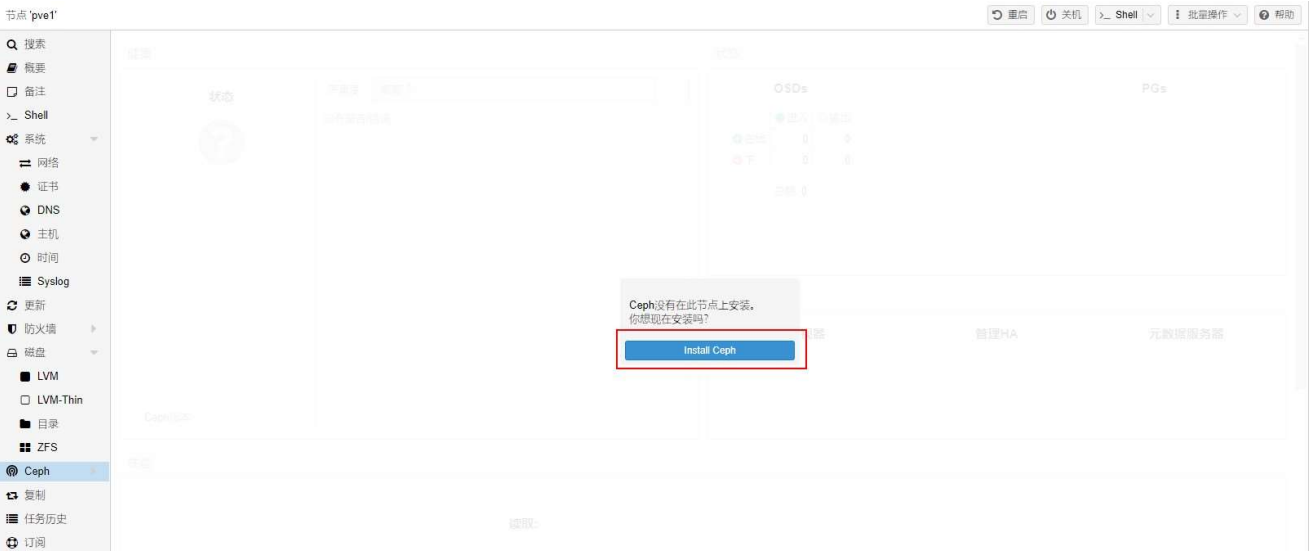
物理路径	磁盘类型	DB磁盘	WAL磁盘	角色
/dev/sda	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdb	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdc	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdd	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sde	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdf	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdg	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdh	HDD	/dev/sdi	/dev/sdj	数据盘
/dev/sdi	SSD	不适用	不适用	DB盘
/dev/sdj	SSD	不适用	不适用	WAL盘
/dev/sdk	SSD	/dev/sdk	/dev/sdk	数据盘
/dev/sdl	SSD	/dev/sdl	/dev/sdl	数据盘

12.2 ceph安装

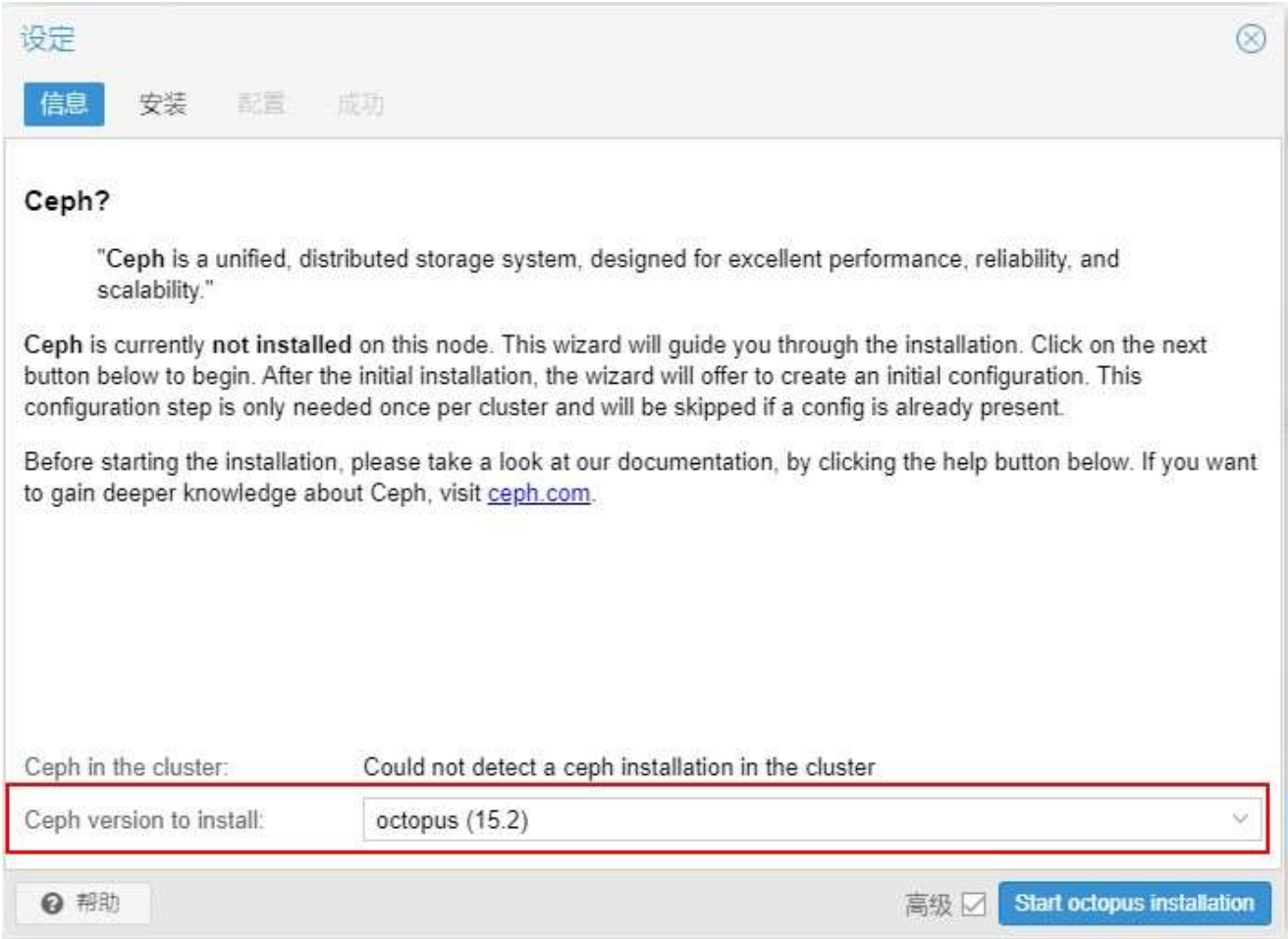
在全部节点中，安装ceph组件。

注意，实际操作中此步骤会在 /etc/apt/sources.list.d/ceph.list 文件中配置官方源。

12.2.1 选择节点 — “ceph” 点击 “Install ceph ”

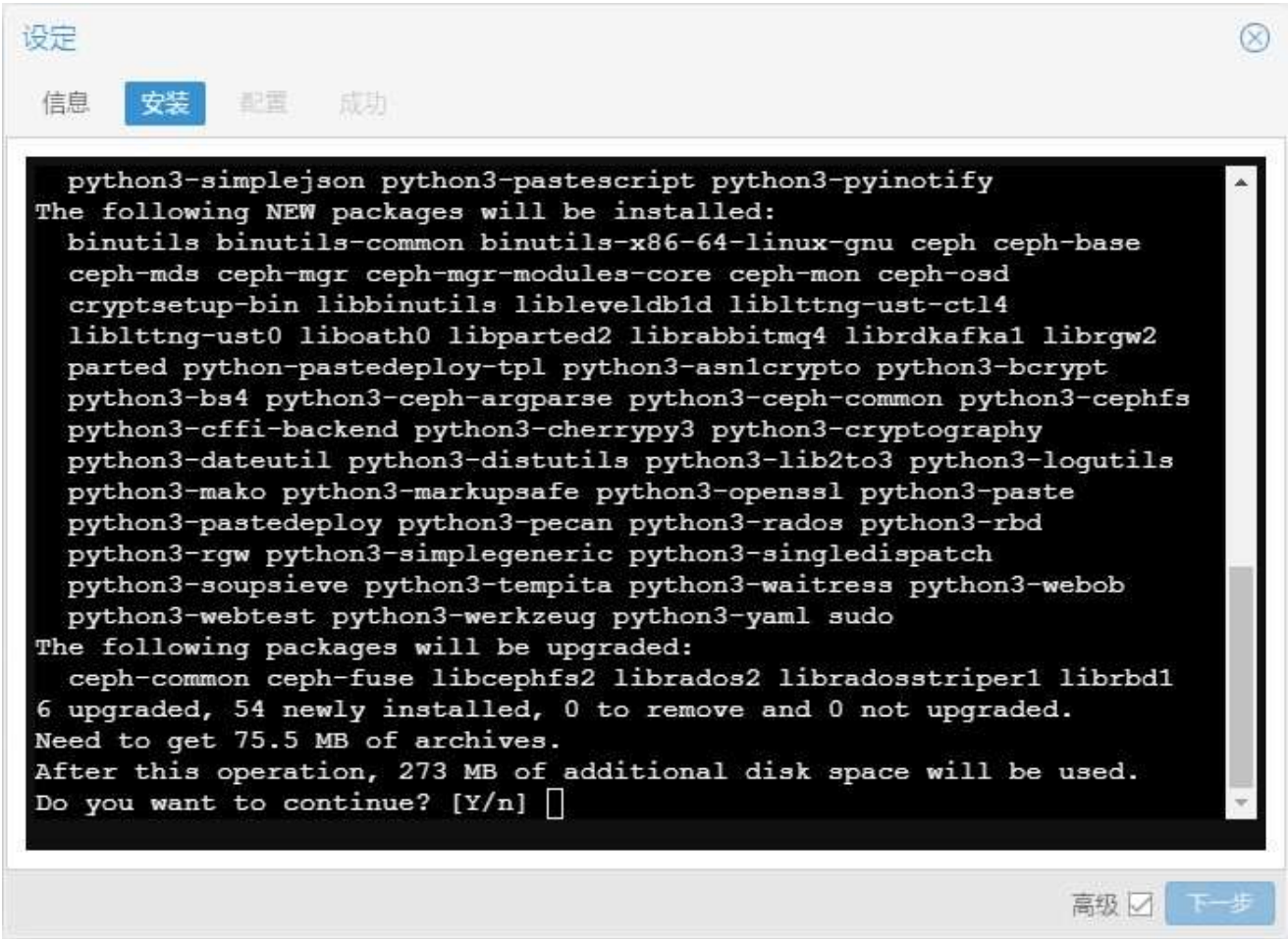


12.2.2 选择版本，此处选择 “ 15.2 ”



12.2.3 输入 “ Y ” 开始安装

如果软件包下载速度较慢，可以先安装一台，然后将安装包拷贝到其他节点相同路径中。
安装包保存路径： /var/cache/apt/archives



12.3 ceph配置

此配置是将ssd硬盘和hdd硬盘分别创建pool，分别为sdd_data和hdd_data。

12.3.1 创建osd

在 “ceph” — “osd” 选项卡创建OSD。

正常情况下磁盘类型会自动识别，如果识别的不正确，注意手动调整。

创建方式如下：

- HDD盘

创建hdd时注意选择DB和WAL磁盘。

创建: Ceph OSD

磁盘: 没有未使用的磁盘

选择HDD盘

数据库磁盘: /dev/sdi

数据库大小 (GiB): 自动

加密OSD: ☐

WAL磁盘: /dev/sdj

Device Class: HDD

WAL大小 (GiB): 自动

Note: Ceph is not compatible with disks backed by a hardware RAID controller. For details see [the reference documentation](#).

帮助

高级 ☒

创建

- SSD盘

创建: Ceph OSD

磁盘: 没有未使用的磁盘

选择SSD盘

数据库磁盘: use OSD disk

数据库大小 (GiB): 自动

加密OSD: ☐

WAL磁盘: use OSD/DB disk

Device Class: SSD

WAL大小 (GiB): 自动

Note: Ceph is not compatible with disks backed by a hardware RAID controller. For details see [the reference documentation](#).

帮助

高级 ☒

创建

- 如果创建时类型选择错误，可以用命令行修改：

假设osd0需要更改。

删除osd类别信息：

```
ceph osd crush rm-device-class osd.0
```

设置osd类别信息

```
ceph osd crush set-device-class ssd osd.0
```

12.3.2 创建class rule

- 创建HDD rule

```
ceph osd crush rule create-replicated hdd_rule default host hdd
```

- 创建SSD rule

```
ceph osd crush rule create-replicated ssd_rule default host ssd
```

12.3.3 创建存储池

- 创建HDD pool

```
ceph osd pool create hdd_data hdd_rule
ceph osd pool application enable hdd_data hdd_data
```

- 创建SSD pool

```
ceph osd pool create ssd_data ssd_rule
ceph osd pool application enable ssd_data ssd_data
```

注意，如果是使用单个节点的测试环境，创建存储池后集群会出现以下警告信息。

```
root@pvel1:~# ceph -s
cluster:
  id:      e78680bc-550b-49e4-baa5-3a6d8e883329
  health: HEALTH_WARN
           1 MDSs report slow metadata IOs
           Reduced data availability: 193 pgs inactive
```

```
Degraded data redundancy: 193 pgs undersized

services:
  mon: 1 daemons, quorum pve1 (age 45m)
  mgr: pve1(active, since 45m)
  mds: 1/1 daemons up
  osd: 3 osds: 3 up (since 44m), 3 in (since 44m)

data:
  volumes: 1/1 healthy
  pools: 4 pools, 193 pgs
  objects: 0 objects, 0 B
  usage: 17 MiB used, 400 GiB / 400 GiB avail
  pgs: 100.000% pgs not active
      193 undersized+peered

progress:
  Global Recovery Event (7m)
  [.....]
```

原因是集群默认的 crushmap 规则是按 host 级别，可使用以下方法修改为 osd 级别：

```
ceph osd getcrushmap -o /etc/ceph/crushmap
crushtool -d /etc/ceph/crushmap -o /etc/ceph/crushmap.txt
sed -i 's/step chooseleaf firstn 0 type host/step chooseleaf firstn 0 type osd/'
/etc/ceph/crushmap.txt
crushtool -c /etc/ceph/crushmap.txt -o /etc/ceph/crushmap-new
ceph osd setcrushmap -i /etc/ceph/crushmap-new
```

12.3.4 添加存储池至PVE

“数据中心” — “存储” — “添加” — “RBD”。

- 添加hdd存储池：



- 添加ssd存储池：



十三、安全加固

13.1 配置公钥登录

13.1.1 生成密钥对

以xshell为例。

- 生成密钥

新建用户密钥生成向导

生产密钥参数

密钥类型的可用性由SSH协议版本确定。SSH1只能使用RSA算法。SSH2可以使用所有密钥类型。

密钥类型(T): RSA

密钥的长度越长安全性越高。普通应用程序中密钥长度的最佳大小为1024位。

密钥长度(L): 2048 位

单击"下一步"生成公钥。

< 上一步(B)

下一步(N) >

完成

取消

- 设置名称及密码

设置密码为可选项，为保证密钥安全，建议设置。

新建用户密钥生成向导

用户密钥信息

请输入用户密钥的名称。

密钥名称(K): TestKey

请输入给用户密钥加密的密码。

密码(P):

确认(C): (重新键入密码)

单击“下一步”以获取有关如何在SSH服务器上注册此公钥的说明。

< 上一步(B)

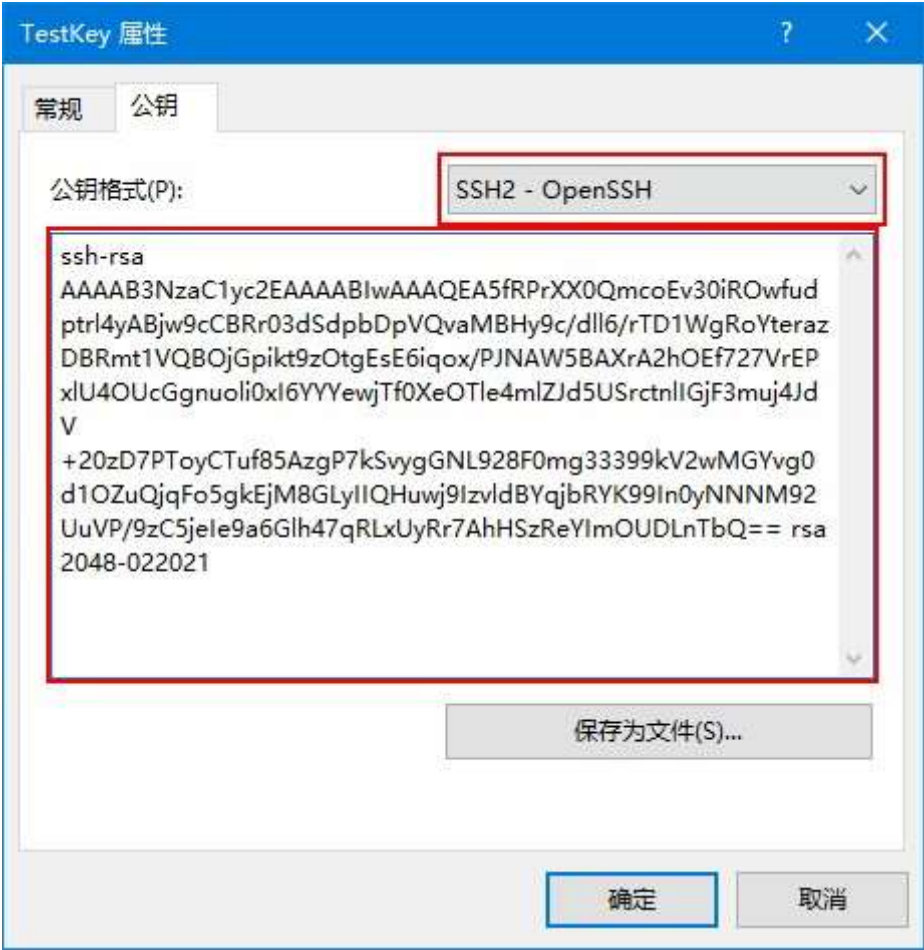
下一步(N) >

完成

取消

13.1.2 将公钥复制至服务器"/root/.ssh/authorized_keys"文件中，选择"SSH2 - OpenSSH"格式。

19



```
vim /root/.ssh/authorized_keys
```

13.1.3 禁止ssh使用密码登录，禁止前先验证一下公钥方式能否登录。

```
# 编辑sshd配置文件
vim /etc/ssh/sshd_config
# 将下列配置添加到最后一行
PasswordAuthentication no
# 重启sshd服务
systemctl restart sshd
```

作者: [Varden](#)
出处: <http://www.cnblogs.com/varden/>
本文内容如有雷同，请联系作者！
本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类: [服务器虚拟化](#), [Ceph](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

- << 上一篇: [Proxmox Mail Gateway 介绍](#)
- >> 下一篇: [\(案例二\) Proxmox VE 备份组件 PBS \(Proxmox Backup Server\) 部署及使用教程](#)

posted @ 2021-08-28 16:57 [Varden](#) 阅读(2297) 评论(0) [编辑](#) [收藏](#) [举报](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

- 编辑推荐:**
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
 - [分布式事务 | 使用 DTM 的 Saga 模式](#)
 - [SQLSERVER 阻塞之 PFS 页到底是什么?](#)
 - [带团队后的日常思考 \(十一\)](#)
 - [记一次使用 gdb 诊断 gc 问题全过程](#)
- 阅读排行:**
- [chatGPT vscode 体验 \(密钥已更新当天有效\)](#)
 - [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
 - [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
 - [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
 - [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
[-取消关注](#)

(案例二) Proxmox VE 备份组件 PBS (Proxmox Backup Server) 部署及使用教程

参考: <https://blog.csdn.net/hanziqing0630/article/details/118678155>

一、Proxmox Backup Server (pbs) 介绍

Proxmox Backup Server (pbs) 是与pve配套的备份解决方案，用于备份和恢复虚拟机、容器和物理主机。通过支持增量、完全重复数据删除的备份，Proxmox Backup Server (pbs) 显着降低了网络负载并节省了宝贵的存储空间。凭借强大的加密和确保数据完整性的方法，确保数据安全。提供web界面，可以极其方便的管理备份文件。

二、环境介绍

2.1 软件版本

本次部署采用当前最新版本：Proxmox Backup Server 1.1-1，可根据当时情况选择最新版本。

2.2 服务器配置

服务器采用一台利旧服务器。

型号	配置	数量
HP ProLiant DL580 G7	CPU : Intel® Xeon® CPU E7- 4807 @ 1.87GHz 内存 : 16GB SSD : 800GB 1块 HDD : 1.2TB 3块	1

三、软件获取

访问官网下载ISO镜像文件，根据需要选择版本：

<https://www.proxmox.com/en/downloads/category/iso-images-pbs>

四、磁盘规划

本案例中采用1块800G SSD作为系统盘，3块 1.2TB HDD作为数据盘使用。

3块数据盘组成RAID0。

五、网络规划

由于备份流量较大，不能备份流量与业务流量放在同一网络中，建议将备份网络与管理网络合并部署或单独规划备份网络。

5.1 物理网络拓扑

此处将pbs服务器部署在管理网络vlan 1当中。

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

常用链接

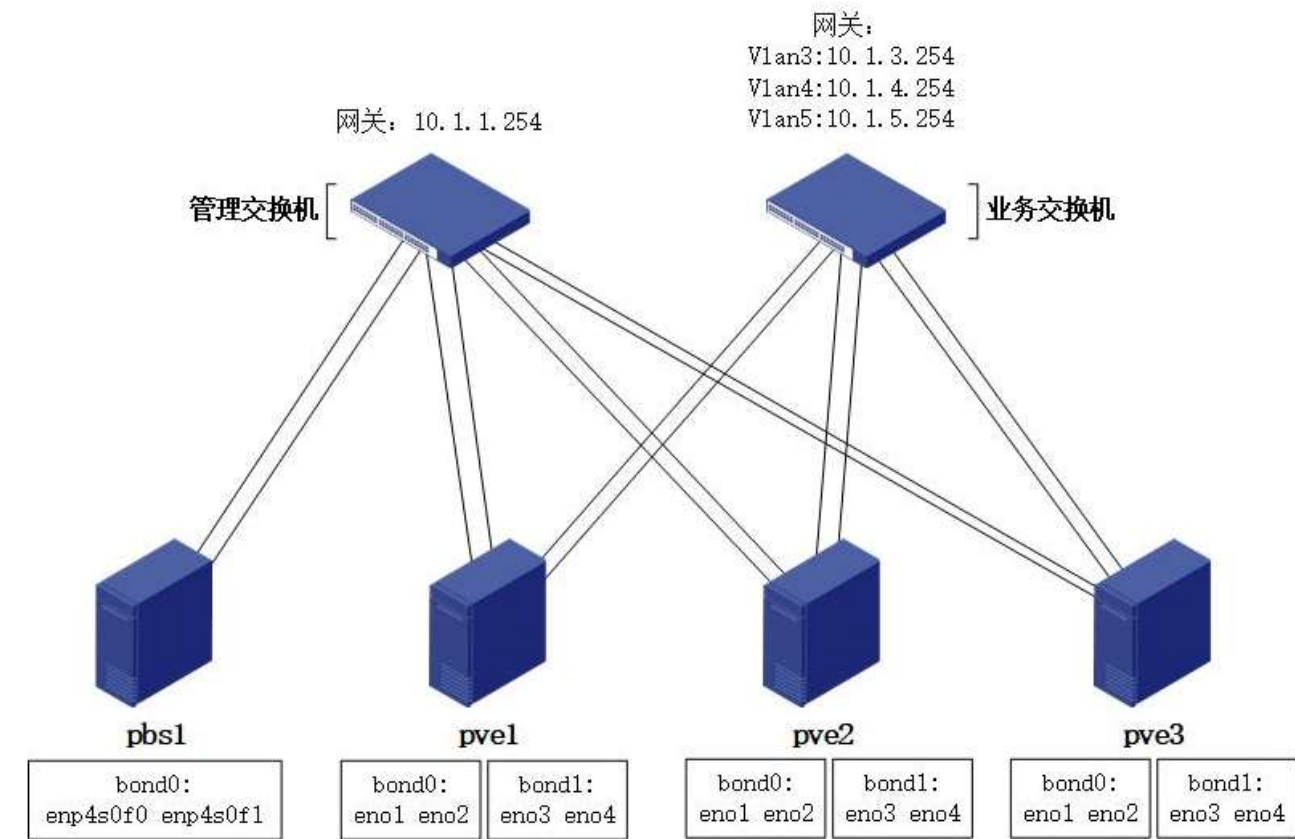
[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

最新随笔

- 1.Ceph 故障域配置 (修改CRUSH MAP)
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类 (Octopus版本)

随笔分类

Alpine(2)
Ansible(23)
CentOS(23)
Ceph(118)
Debian(19)
Docker(28)
ELK日志分析(15)
Helm(3)
Java(1)
Jenkins(10)
K8s(209)
KubectI(4)
Linux集群技术(3)
Linux企业常用服务(9)
Linux企业级安全(5)
Linux系统调优和监控(8)
Linux系统基础及管理(38)
MongoDB(12)
MySQL/MariaDB(6)
NFS(3)
PostgreSQL(1)

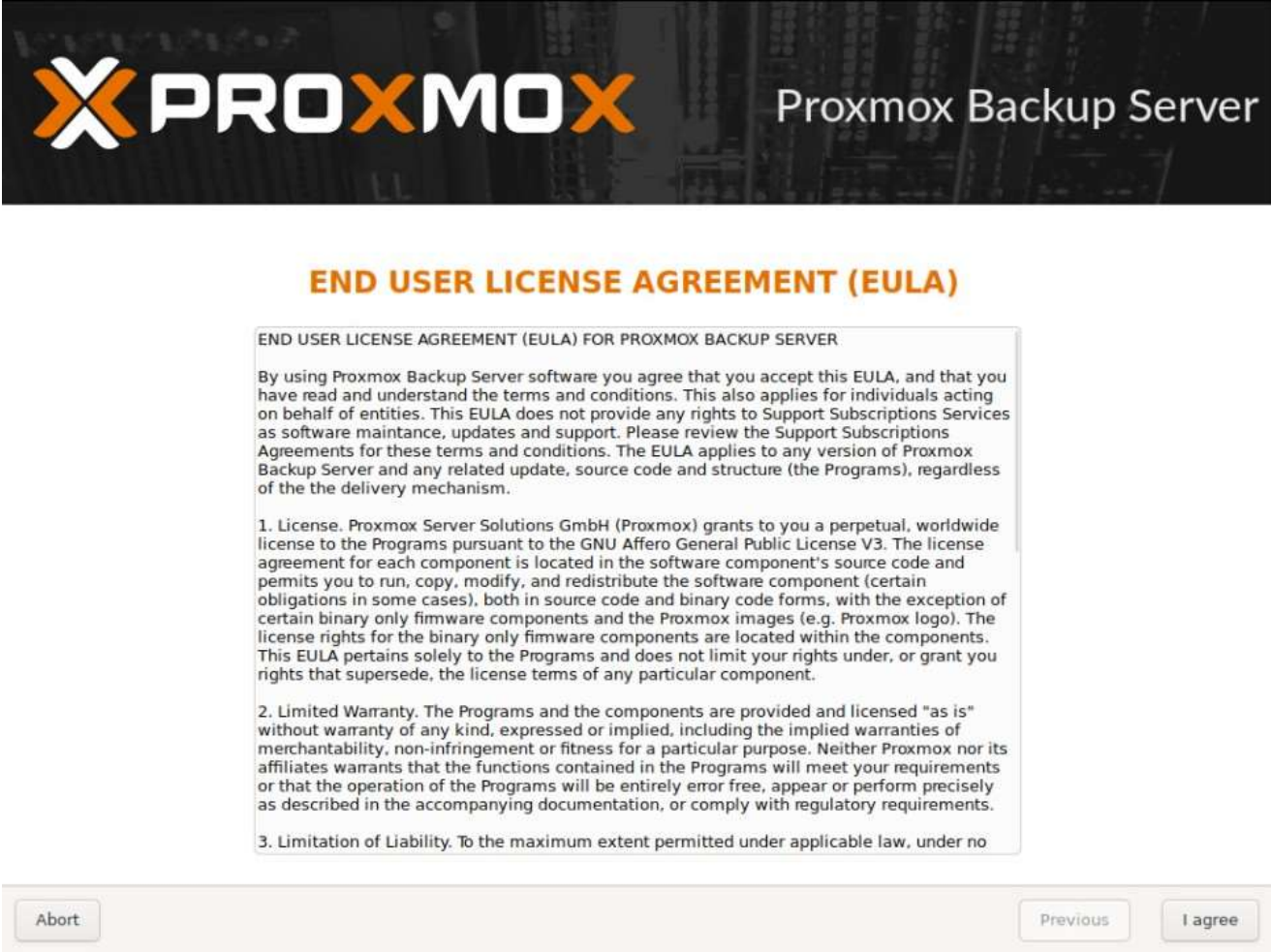


六、pbs安装

6.1 选择第一个，安装“pbs”



6.2 同意协议



Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

1. CentOS7安装源配置详解(9876)
2. leaf-spine（叶脊）网络架构介绍(7105)
3. PVE 备份和恢复介绍(5386)
4. Proxmox Virtual Environment 介绍(4947)
5. Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
6. Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
7. PVE 虚拟机模板和克隆介绍(3936)
8. 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
9. CentOS7安装配置远程桌面详解(3631)
10. Debian安装源配置详解(3457)
11. Apollo配置中心使用详解(3356)
12. Proxmox VE 防火墙介绍(3323)
13. Gitea 简单使用说明(3313)
14. Linux系统路由配置说明(3270)
15. Chrony时间同步服务器安装配置说明(3256)
16. Centos7防火墙配置详解(3231)
17. 【Rancher】简单介绍(3230)
18. Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
19. PVE 超融合架构介绍(3096)
20. 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

1. Proxmox VE 存储介绍(1)
2. Proxmox Virtual Environment 介绍(1)
3. Debian Jessie Preseed自动化安装示例(1)
4. Debian live-build官方教程学习笔记(1)
5. Ceph存储池pg_num配置详解(1)

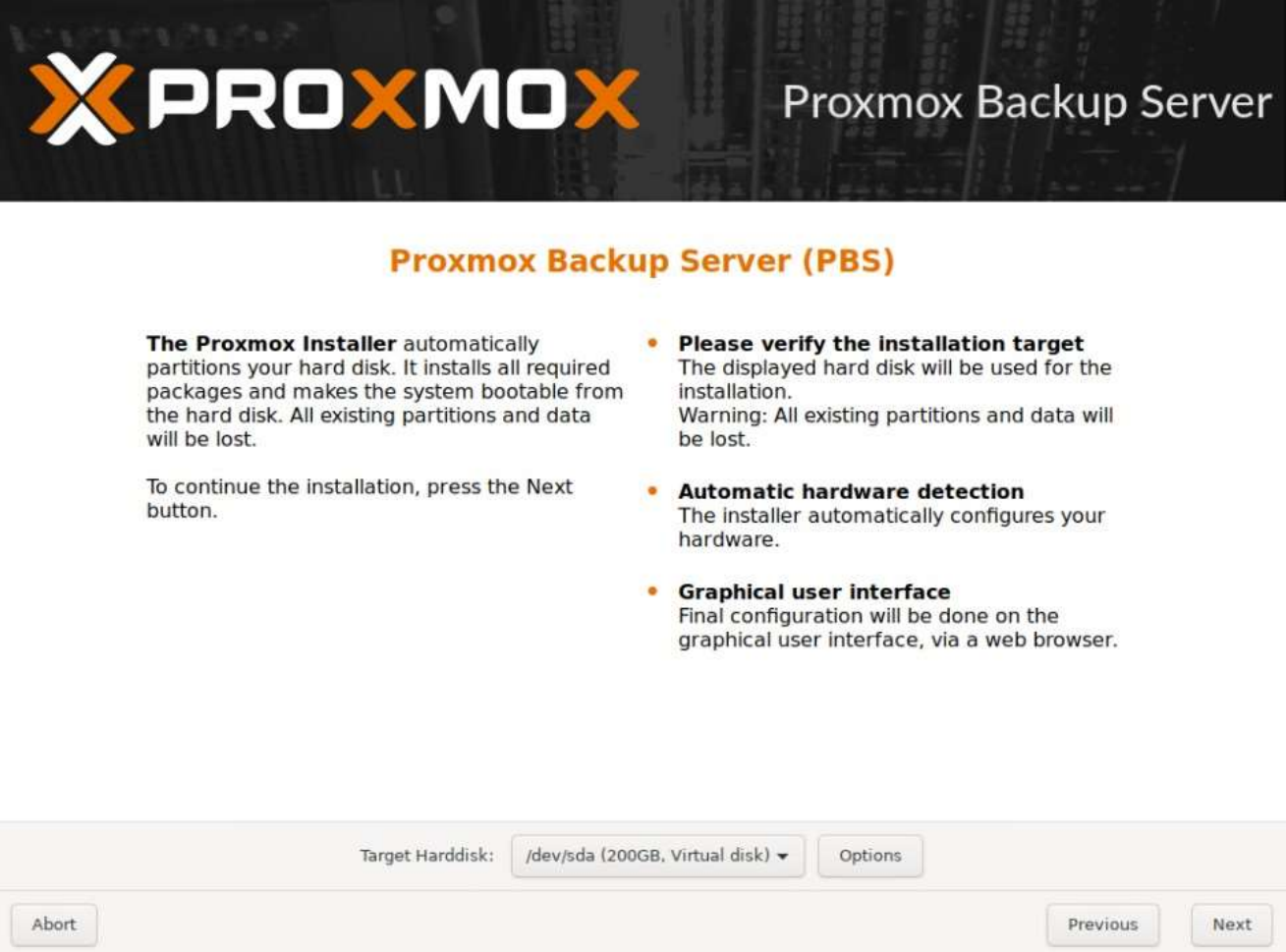
推荐排行榜

1. leaf-spine（叶脊）网络架构介绍(1)
2. Proxmox VE 防火墙介绍(1)
3. Proxmox Virtual Environment 介绍(1)
4. VIM 常用快捷键(1)
5. Debian7 bonding配置详解(1)

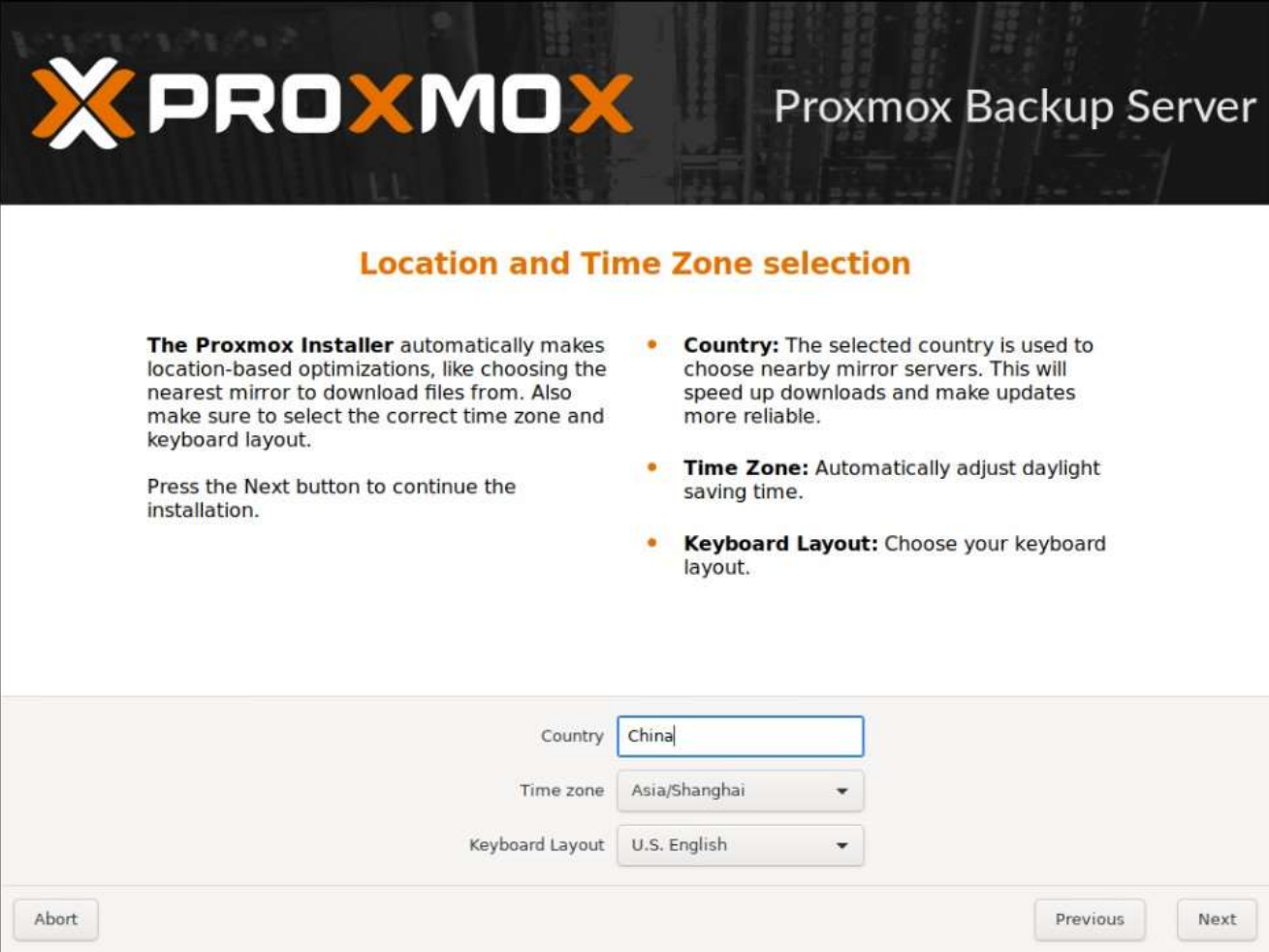
最新评论

1. Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从“Default (No cache)”修改为“Write back (unsafe)”，对磁盘数据有影响么？
--大熊猫子
2. Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的PXC EMC FC-SAN存储（不支持ISCSI），如果改用Proxmox做集群，FC-SAN应该以什么方式挂载给Proxmox好呢？

6.3 选择系统盘



6.4 配置国家、地区



6.5 配置密码、邮箱

如果为生产环境务必配置为强口令。邮箱建议配置为真实邮箱。

- [3. Re:Proxmox Virtual Environment 介绍](#) --9394952
很全面!
- [4. Re:Ceph存储池pg_num配置详解](#) --willblog
mark
- [5. Re:Debian live-build官方教程学习笔记](#) --穷死了
执行成功之后为什么没找看iso
- [--abraham_liuhai](#)



Administration Password and Email Address

Proxmox Backup Server is a full-featured, highly secure system, based on Debian GNU/Linux.

In this step, please provide the root password.

- Password:** Please use a strong password. It should be at least 8 characters long, and contain a combination of letters, numbers, and symbols.
 - Email:** Enter a valid email address. Your Proxmox Backup Server will send important alert notifications to this email account (all emails for 'root').
- To continue the installation, press the Next button.

Password

••••••••

Confirm

••••••••

Email

735273287@qq.com

Abort

Previous

Next

6.6 配置网卡、主机名、IP地址、DNS等信息，配置到第一块网卡



Management Network Configuration

Please verify the displayed network configuration. You will need a valid network configuration to access the management interface after installing.

After you have finished, press the Next button. You will be shown a list of the options that you chose during the previous steps.

- IP address (CIDR):** Set the main IP address and netmask for your server in CIDR notation.
- Gateway:** IP address of your gateway or firewall.
- DNS Server:** IP address of your DNS server.

Management Interface:

ens192 - 00:50:56:b0:7e:5c (vmxnet3)

Hostname (FQDN):

pbs2.cnaec.local

IP Address (CIDR)

10.1.2.122 / 24

Gateway:

10.1.2.1

DNS Server:

114.114.114.114

Abort

Previous

Next

6.7 检查无误后点击安装



Summary

Please confirm the displayed information. Once you press the **Install** button, the installer will begin to partition your drive(s) and extract the required files.

Option	Value
Filesystem:	ext4
Disk(s):	/dev/sda
Country:	China
Timezone:	Asia/Shanghai
Keymap:	en-us
Email:	735273287@qq.com
Management Interface:	ens192
Hostname:	pbs2
IP CIDR:	10.1.2.122/24
Gateway:	10.1.2.1
DNS:	114.114.114.114

☒ Automatically reboot after successful installation

AbortPreviousInstall

6.8 等待安装过程



Proxmox Backup Server

- Data Integrity**
The built in SHA-256 checksum algorithm ensures the accuracy and consistency of your backups.

Compression
The ultra-fast Zstandard compression is able to compress several gigabytes of data per second.
- Remote Sync**
Backup data to dedicated hardware or efficiently synchronize them to remote sites. Only deltas containing new data are transferred.
 - Technical Support**
Enterprise support options are available from Proxmox.

extracting pve-kernel-5.4.106-1-pve_5.4.106-1_amd64.deb68%

AbortInstall

6.9 安装成功

```
-----  
Welcome to the Proxmox Backup Server. Please use your web browser to  
configure this server - connect to:  
  
https://10.1.2.122:8007/  
  
-----  
  
pbs2 login:
```

6.10 安装结束后重启，通过浏览器访问，出现以下页面证明安装成功。

地址：<https://你的IP地址:8007>
用户名：root
密码：为安装时配置的密码。
域：选择Linux PAM



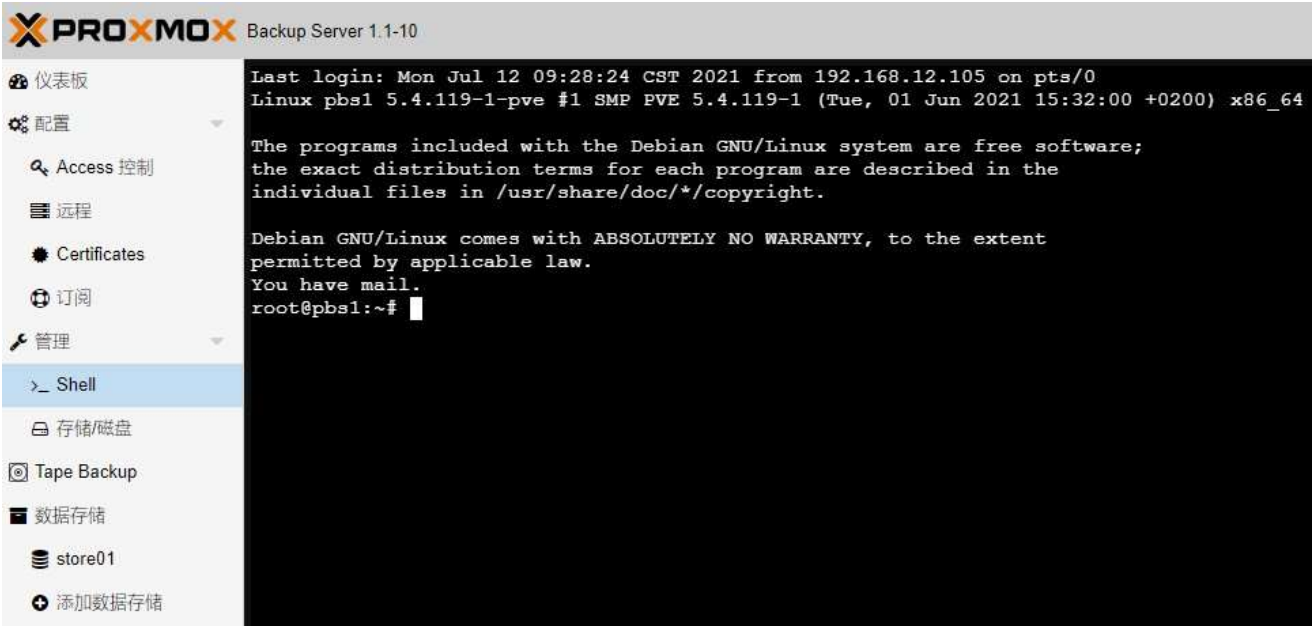
七、更换软件源

由于自带的软件源速度较慢，此处选择科大源：<http://mirrors.ustc.edu.cn>

涉及的软件源有二个，分别为debain、pbs，需要分别修改下列文件。

名称	source文件路径
debain	/etc/apt/sources.list
pbs	/etc/apt/sources.list.d/pbs-enterprise.list

7.1 通过web管理页面中的shell或ssh登录后台。



7.2 更换debian源

```
vi /etc/apt/sources.list
```

删除默认的源，替换为：

```
deb http://mirrors.ustc.edu.cn/debian buster main contrib non-free
# deb-src http://mirrors.ustc.edu.cn/debian buster main contrib non-free
deb http://mirrors.ustc.edu.cn/debian buster-updates main contrib non-free
# deb-src http://mirrors.ustc.edu.cn/debian buster-updates main contrib non-free

deb http://mirrors.ustc.edu.cn/debian-security/ buster-security main non-free contrib
# deb-src http://mirrors.ustc.edu.cn/debian-security/ buster-security main non-free
contrib
```

7.3 更换pbs源为非订阅源

```
mv /etc/apt/sources.list.d/pve-enterprise.list /etc/apt/sources.list.d/pve-enterprise.list.bak
CODENAME=`cat /etc/os-release |grep CODENAME |cut -f 2 -d "="`
echo "deb http://mirrors.ustc.edu.cn/proxmox/debian $CODENAME pve-no-subscription" > /etc/apt/sources.list.d/pve-no-subscription.list
```

7.4 更新软件列表

```
apt update
```

八、更新pbs至最新版本（可选）

```
apt upgrade
```

九、基础配置

9.1 安装vim

```
apt install vim
```

9.2 配置ntp服务

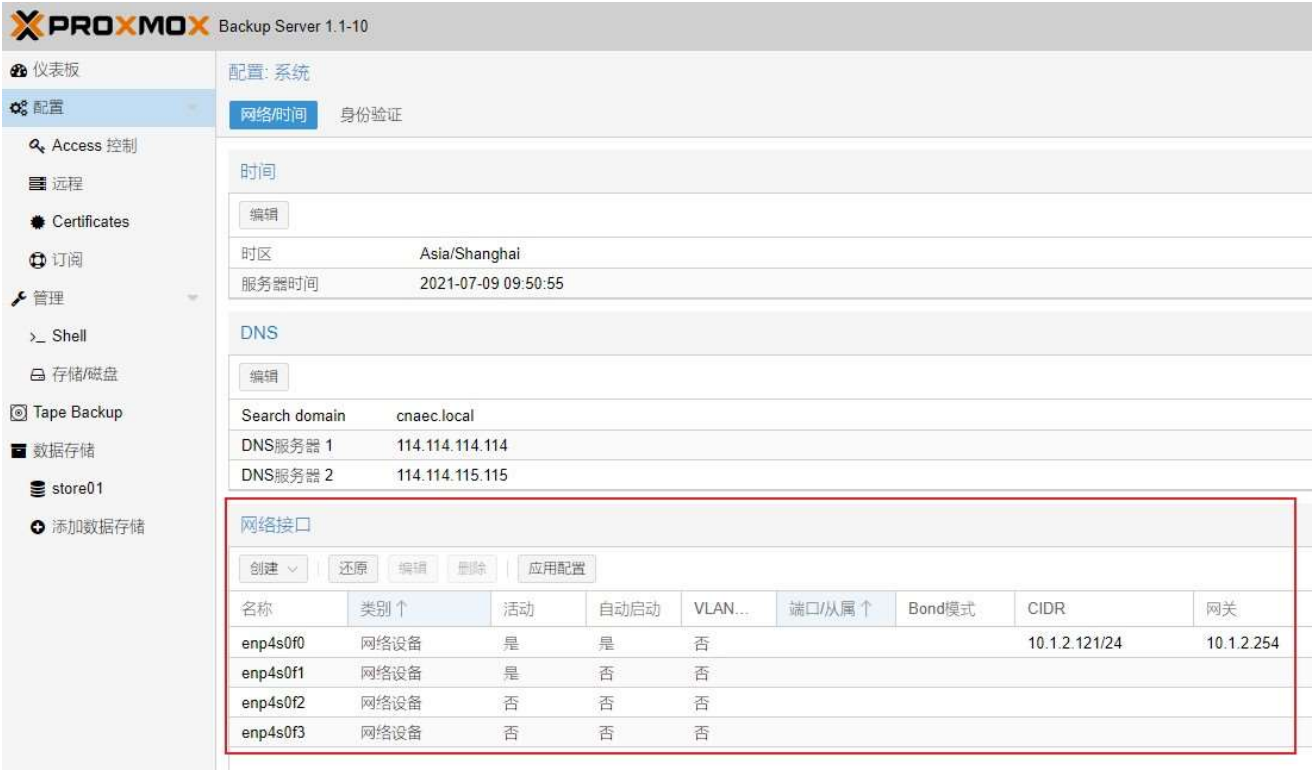
配置统一的时钟源

- 如果有条件建议在网内部署一台ntp服务器，所有节点与网内ntp同步时钟。
- 如果没有网内NTP服务器，可以采用下列NTP服务器。

```
阿里云：ntp.aliyun.com
国家授时中心：ntp.ntsc.ac.cn
```

十、pbs网络配置

10.1 打开"配置" - "网络/时间"选项卡。



10.2 编辑enp4s0f0，将其中"IP"、"网关"删除，点击确定。

网卡名称根据实际情况。

编辑: 网络设备

名称: enp4s0f0

自动启动: ☒

IPv4/CIDR:

备注:

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

MTU:

高级 ☒

确定

Reset

10.3 创建bond0

本案例bond为主备模式，如需配置LACP模式请参考：（案例一）Proxmox VE + Ceph + 物理网络 超融合生产环境部署案例

点击创建Linux Bond。
名称: bond0
IPv4: 按需
网关: 按需
slaves: enp4s0f0 enp4s0f1，网卡名称根据实际情况。
模式: active-backup
bond-primary: enp4s0f0，网卡名称根据实际情况。
按以上参数配置，然后点击OK。

编辑: Linux Bond

名称: bond0

自动启动: ☒

IPv4/CIDR: 10.1.2.122/24

Slaves: enp4s0f0,enp4s0f1

网关 (IPv4): 10.1.2.254

模式: active-backup

IPv6/CIDR:

Hash策略:

网关 (IPv6):

bond-primary: enp4s0f0

备注:

MTU:

高级 ☒

确定

Reset

10.4 应用配置

如果感觉配置有问题可以点击还原按钮，还原为配置前状态。如果没有问题，点击应用配置使配置生效。

网络配置不正确可能会导致无法管理节点，应用配置前请确认配置正确，如果出现无法管理的情况需要进入shell手动修改配置。

网络接口								
创建 还原 编辑 删除 应用配置								
名称	类别 ↑	活动	自动启动	VLAN...	端口/从属 ↑	Bond模式	CIDR	网关
bond0	Linux Bond	是	是	否	enp4s0f0,e...	LACP (802....	10.1.2.122/24	10.1.2.254
enp4s0f0	网络设备	是	是	否				
enp4s0f1	网络设备	是	是	否				
enp4s0f2	网络设备	否	否	否				
enp4s0f3	网络设备	否	否	否				

十一、磁盘初始化（硬件RAID或ZFS）

两种方式二选一即可，如果服务器有RAID卡，就采用硬件RAID，如没有硬件RAID卡，就采用软件RAID。如下是配置方式，也可以采用自己的方式。

11.1 硬件RAID与ZFS RAID

- 常用模式：如果不考虑数据保护推荐采用RAID0，考虑数据保护推荐采用RAIDZ-1（RAID-5）。

硬件RAID	ZFS RAID	最少需要硬盘数量
raid-0	raid-0	1
raid-1	raid-1	2
raid-10	raid-10	4
raid-5	raidz-1	3
raid-6	raidz-2	4

11.2 硬件RAID方式

11.2.1 在服务器RAID卡中配置磁盘，将磁盘按需组成RAID。

11.2.2 进入shell，将RAID磁盘格式化后，挂载到服务器目录下（如：/data）。

```
1 #假设RAID 磁盘：sdk，挂载点：/data，文件系统：ext4
2 #创建挂载目录：
3 mkdir /data
4 #格式化磁盘：
5 mkfs.ext4 /dev/sdk
6 #挂载磁盘：
7 mount /dev/sdk /data/
8 #配置自动挂载：
9 vim /etc/fstab
10 #在/etc/fstab文件最后添加以下内容：
11 /dev/sdk /data ext4 defaults 0 0
```

11.2.3 添加数据存储时选择相应挂载路径。



11.3 ZFS方式（软RAID）

适用于没有硬件RAID卡的情况。
可以使用web页面或shell命令创建，注意：raid0只能用shell创建。
注意：创建zfs时pool名称必须是字母开头，否则会创建失败。

11.3.1 web创建

按需选择RAID级别。

编辑: ZFS

名称: zfs01

RAID级别: 单磁盘

添加为数据存储: ☒

压缩:

ashift:

☒	设备	模型	串行
☒	/dev/sde	Virtual_disk	
☒	/dev/sdf	Virtual_disk	
☒	/dev/sdg	Virtual_disk	20.00 GiB

Note: ZFS is not compatible with disks backed by a hardware RAID controller. For details see [the reference documentation](#).

帮助

确定

Reset

11.3.2 shell创建

登录后复制

```
1 #<pool>字段: 替换为pool名称, <device*>: 替换为成员磁盘的路径.
2 #raid-0: 条带, 最少需要一块硬盘.
3 zpool create -f -o ashift=12 <pool> <device1> <device2>
4 #raid-1: 镜像, 最少需要两块硬盘.
5 zpool create -f -o ashift=12 <pool> mirror <device1> <device2>
6 #raid-10: 先镜像再条带, 最少需要四块硬盘.
7 zpool create -f -o ashift=12 <pool> mirror <device1> <device2> mirror <device3> <device4>
8 #raidz-1: 类似RAID5, 最少需要三块硬盘.
9 zpool create -f -o ashift=12 <pool> raidz1 <device1> <device2> <device3>
10 #raidz-2: 类似RAID6, 最少需要四块硬盘.
11 zpool create -f -o ashift=12 <pool> raidz2 <device1> <device2> <device3> <device4>
```

11.4 添加数据存储

zfs配置成功后, 会自动添加数据存储, 看到数据存储下出现刚才创建的zfs池, 说明配置成功。

十二、pve与pbs对接配置

12.1 在pbs服务器shell中输入以下命令, 复制“指纹”字段。

1 | proxmox-backup-manager cert info

root@pbs2:~# proxmox-backup-manager cert info
Subject: O = Proxmox Backup Server, OU = 06529A4D-2BEF-4EDB-824C-5281C86920E8, CN = pbs2.cnaec.local
IP:[127, 0, 0, 1]
IP:[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
DNS:localhost
DNS:pbs2
DNS:pbs2.cnaec.local
Issuer: O = Proxmox Backup Server, OU = 06529A4D-2BEF-4EDB-824C-5281C86920E8, CN = pbs2.cnaec.local
Validity:
Not Before: Apr 28 08:13:24 2021 GMT
Not After : Aug 29 08:13:24 3020 GMT
Fingerprint (sha256): 78:70:a9:19:6d:cf:45:6c:79:c8:0f:46:c9:b7:ec:62:59:0f:3e:cc:47:52:5b:b6:7c:2d:65:1e:a8:37:f0:31
Public key type: rsaEncryption
Public key bits: 4096

12.2 在pve服务器以下位置点击添加, 选择“Proxmox Backup Server”。



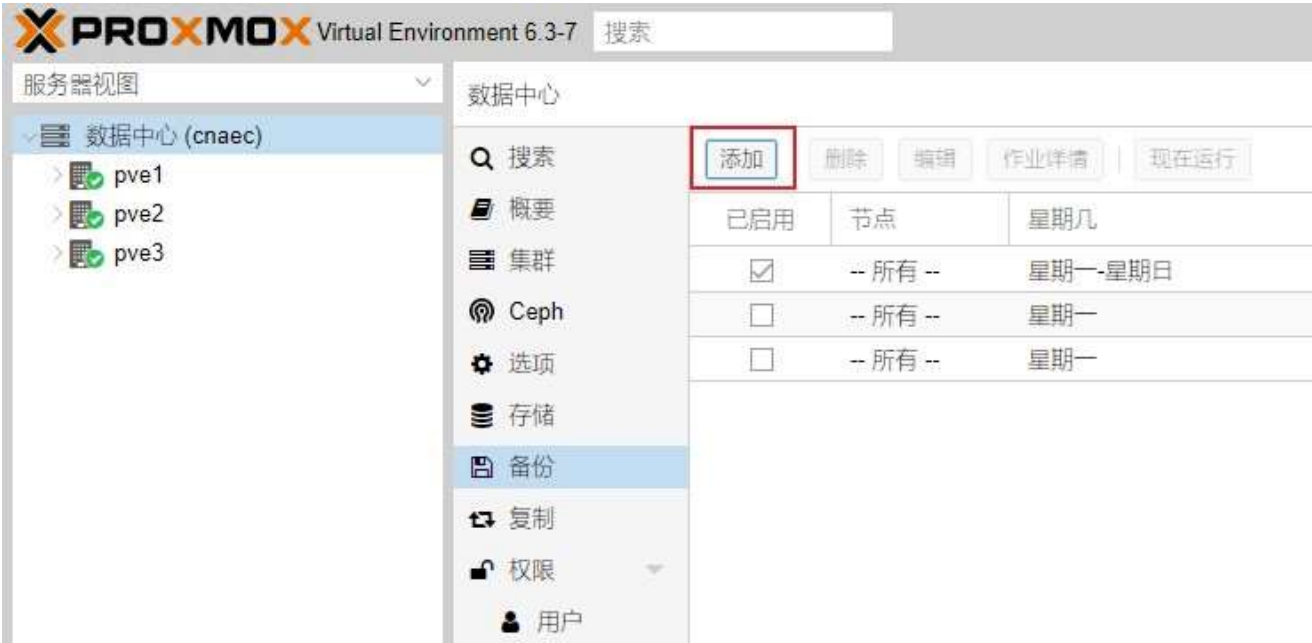
12.3 根据实际情况填写参数。

ID: 服务器名称, 按需填写。
服务器: 填写pbs服务器IP地址。
用户名: 填写root@pam
密码: 按需
Datastore: 填写pbs中数据存储的名称。
指纹: 填写第一步中复制的指纹。



十三、备份策略配置

13.1 在pve服务器以下位置点击添加。



13.2 根据实际情况填写参数。

存储: 选择第12章添加的pbs服务器。
邮件、通知: 按需

备份开始时间：按需
最下方选择需要备份的虚拟机即可。

创建: 备份作业

节点:

-- 所有 --

发送邮件至:

735273287@qq.com

存储:

pbs01

Email通知:

总是

星期几:

星期六, 星期日

压缩:

ZSTD (又快又好)

开始时间:

01:15

模式:

快照

选择模式:

包括选中的VMs

启用:

☒

☐	ID ↑	节点	状态	名称	类别
☒	100	pve1	运行中	AD-Slave	qemu
☒	101	pve3	运行中	AD	qemu
☒	102	pve3	运行中	centos-yum	qemu

十四、“精简/GC”策略配置

通过“精简/GC”配置，可以配置备份数据的保留时间，以及定期进行垃圾回收和精简，提高磁盘利用率。

下图配置的含义是：每周六凌晨5点进行垃圾回收，每周日5点进行精简，备份数据保留62天。

仪表板

配置

Access 控制

远程

Certificates

订阅

管理

Shell

存储/磁盘

Tape Backup

数据存储

store01

添加数据存储

数据存储: store01

概要

内容

精简 & GC

同步作业

验证作业

选项

权限

编辑

启动垃圾回收

垃圾回收计划	sat 05:00
精简时间表	sun 05:00
保留上次	
保留每小时	
保留每天	62
保留每周	
保留每月	
保留每年	

“精简/GC”模拟工具：使用这个工具可以直观看到策略的实际效果。
<https://你的pbs服务器IP:8007/docs/prune-simulator/index.html>

十五、pbs日常维护

15.1 pbs日常维护

在数据存储“内容”选项卡中可以对备份文件进行验证、更改所有者、精简、删除等操作。

PROXMOX Backup Server 1.1-10

文档 任务 0 root@pam

仪表板

配置

Access 控制

远程

Certificates

订阅

管理

Shell

存储/磁盘

Tape Backup

数据存储

store01

添加数据存储

数据存储: store01

概要

内容

精简 & GC

同步作业

验证作业

选项

权限

重载

全部验证

搜索 组、日期或所有者

备份组 ↑	备注	操作 ↑	备份时间 ↑	大小	计数	所有者	加密的	验证状态
vm/100		V. 👤 ⚙️ 🗑️	2021-07-12 01:15:02	100.00 GiB	62	root@pam	否	🟡 无
vm/101		V. 👤 ⚙️ 🗑️	2021-07-12 01:15:02	100.00 GiB	62	root@pam	否	🟡 无
vm/102		V. 👤 ⚙️ 🗑️	2021-07-12 01:15:02	50.90 GiB	8	root@pam	否	🟡 无
vm/103		V. 👤 ⚙️ 🗑️	2021-07-12 01:15:54	700.00 GiB	62	root@pam	否	🟡 无
vm/109		V. 👤 ⚙️ 🗑️	2021-07-12 01:16:34	500.00 GiB	62	root@pam	否	🟡 无
vm/112		V. 👤 ⚙️ 🗑️	2021-03-28 22:10:53	40.00 GiB	7	root@pam	否	🟡 无
vm/8010		V. 👤 ⚙️ 🗑️	2021-03-09 10:36:30	40.00 GiB	1	root@pam	否	🟡 无
vm/8011		V. 👤 ⚙️ 🗑️	2021-03-09 10:38:59	40.20 GiB	1	root@pam	否	🟡 无

十六、安全加固

16.1 配置公钥登录

16.1.1 生成密钥对

以xshell为例。

19

- 生成密钥

新建用户密钥生成向导

生产密钥参数

密钥类型的可用性由SSH协议版本确定。SSH1只能使用RSA算法。SSH2可以使用所有密钥类型。

密钥类型(T):

RSA

密钥的长度越长安全性越高。普通应用程序中密钥长度的最佳大小为1024位。

密钥长度(L):

2048

位

单击"下一步"生成公钥。

< 上一步(B)

下一步(N) >

完成

取消

- 设置名称及密码

设置密码为可选项，为保证密钥安全，建议设置。

新建用户密钥生成向导

用户密钥信息

请输入用户密钥的名称。

密钥名称(K):

TestKey

请输入给用户密钥加密的密码。

密码(P):

●●●●●●

确认(C):

●●●●●●

(重新键入密码)

单击“下一步”以获取有关如何在SSH服务器上注册此公钥的说明。

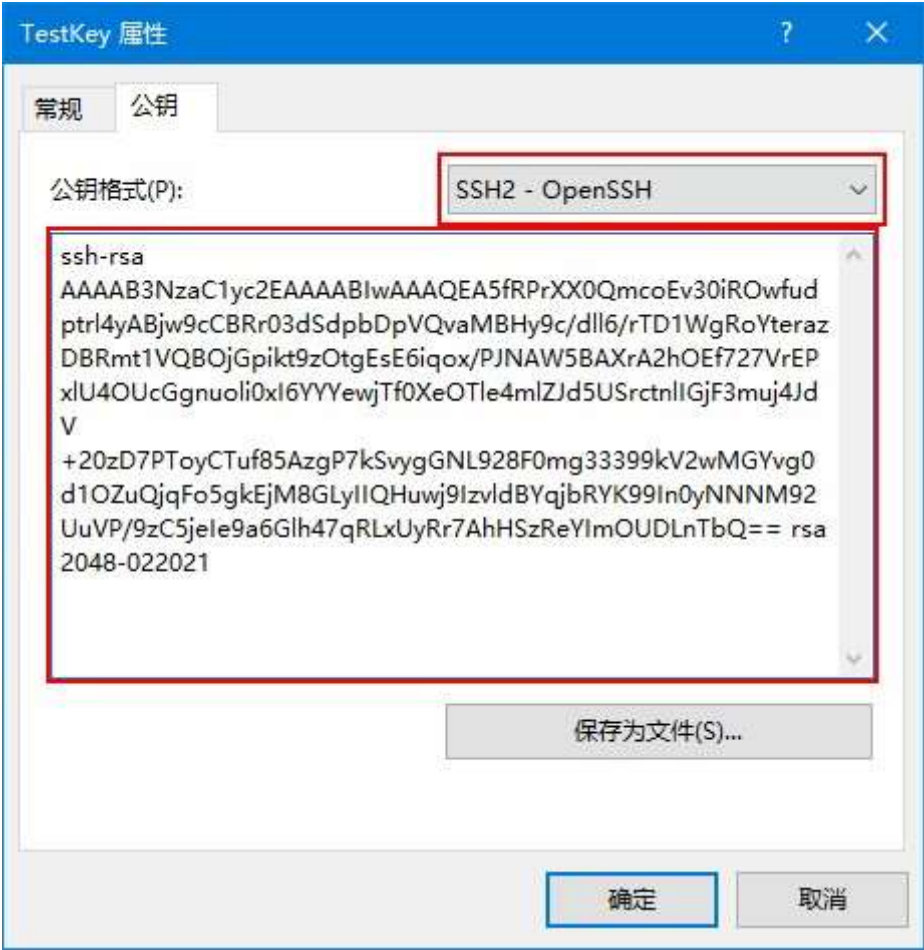
< 上一步(B)

下一步(N) >

完成

取消

16.1.2 将公钥复制至服务器"/root/.ssh/authorized_keys"文件中，选择"SSH2 - OpenSSH"格式。



```
1 | vim /root/.ssh/authorized_keys
```

登录后复制

16.1.3 禁止ssh使用密码登录，禁止前先验证一下公钥方式能否登录。

```
1 | #编辑sshd配置文件
2 | vim /etc/ssh/sshd_config
3 | #将下列配置添加到最后一行
4 | PasswordAuthentication no
5 | #重启sshd服务
6 | systemctl restart sshd
```

登录后复制

作者: [Varden](#)
出处: <http://www.cnblogs.com/varden/>
本文内容如有雷同，请联系作者！
本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类: [服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

« 上一篇: [\(案例一\) Proxmox VE + Ceph + 物理网络 超融合生产环境部署案例](#)
» 下一篇: [PVE 超融合架构介绍](#)

posted @ 2021-08-30 10:18 [Varden](#) 阅读(1428) 评论(0) [编辑](#) [收藏](#) [举报](#)
[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

- 编辑推荐:**
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
 - [分布式事务 | 使用 DTM 的 Saga 模式](#)
 - [SQLSERVER 阻塞之 PFS 页到底是什么？](#)
 - [带团队后的日常思考（十一）](#)
 - [记一次使用 gdb 诊断 gc 问题全过程](#)
- 阅读排行:**
- [chatGPT vscode 体验（密钥已更新当天有效）](#)
 - [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
 - [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
 - [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
 - [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
[-取消关注](#)

PVE 超融合架构介绍

参考: https://blog.51cto.com/u_14742102/3606723

一、什么是超融合？超融合具体包括哪些组件？

所谓“超融合”架构，其实就是“服务器虚拟化+分布式存储”架构，也就是说超融合技术包含了服务器虚拟化技术和分布式存储技术。

服务器虚拟化技术目前在企业数据中心已有广泛的应用，其核心是将服务器的计算资源虚拟化，然后使用虚拟主机代替物理服务器去承载业务负载。常见的服务器虚拟化软件有VMware ESXi、Citrix Xen、Microsoft Hyper-V、KVM。但是服务器虚拟化软件本身不具备数据存储的功能，因此服务器虚拟化需要与本地存储、或集中共享存储、或分布式存储等一起配套使用。

分布式存储也称为软件定义存储，它通过网络将数据分散存储在多台独立的通用服务器上，利用多台通用服务器分担存储负荷，它不但提高了系统的可靠性、可用性和存取效率，还易于扩展。常见的分布式存储软件有Ceph、GlusterFS等。传统存储是采用单一设备，集中式的架构，而分布式存储技术则采用多设备，集群的架构，这是两者最大的区别，如表1所示。

表1. 传统IT架构与超融合架构

IT 架构	方案架构	方案组件
传统 IT 架构	服务器虚拟化+本地存储	虚拟化软件+单节点服务器
	服务器虚拟化+集中共享存储	虚拟化软件+多节点服务器+专用存储设备
超融合架构	服务器虚拟化+分布式存储	虚拟化软件+分布式存储软件+通用服务器

二、超融合为什么要3个节点起步，2个节点或1个节点起步不行吗？

2.1 1个节点服务器只能采用传统IT架构，无法实现超融合

超融合架构的核心组件是分布式存储软件，所谓分布式存储，必需是在多个节点上进行存储才有意义，而1个节点服务器明显不具备“超融合”特性。

- 1个节点方案架构

1个节点服务器方案只能采用传统IT架构，通过在单节点服务器上部署虚拟化软件，然后使用虚拟主机代替物理服务器去承载业务负载，如图1所示。

图1. 1节点服务器方案架构



- 1个节点数据可靠性方案 —— RAID机制

在1个节点的方案中，由于服务器虚拟化软件本身不具备数据存储的功能，需要与本地存储（服务器自带的硬盘）一起使用，因此虚拟化软件本身不具备有数据冗余、灾备功能，所以只能通过硬件服务器本身的RAID机制去实现。

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

最新随笔

- 1.Ceph 故障域配置 (修改CRUSH MAP)
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类 (Octopus版本)

随笔分类

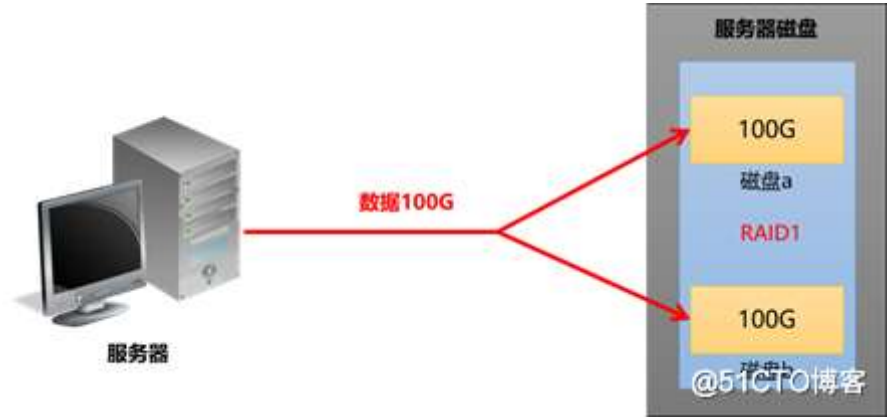
Alpine(2)
Ansible(23)
CentOS(23)
Ceph(118)
Debian(19)
Docker(28)
ELK日志分析(15)
Helm(3)
Java(1)
Jenkins(10)
K8s(209)
KubectI(4)
Linux集群技术(3)
Linux企业常用服务(9)
Linux企业级安全(5)
Linux系统调优和监控(8)
Linux系统基础及管理(38)
MongoDB(12)
MySQL/MariaDB(6)
NFS(3)
PostgreSQL(1)

RAID1是将一个两块硬盘构成RAID磁盘阵列，其容量仅等于一块硬盘的容量，因为另一块只是当作数据“镜像”。RAID1磁盘阵列显然是最可靠的一种阵列，因为它总是保持一份完整的数据备份。

RAID1至少需要两块硬盘，RAID大小等于两个RAID分区中最小的容量（最好将分区大小分为一样），数据有冗余，在存储时同时写入两块硬盘，实现了数据备份。RAID1磁盘利用率为50%，即2块100G的磁盘构成RAID1只能提供100G的可用空间。

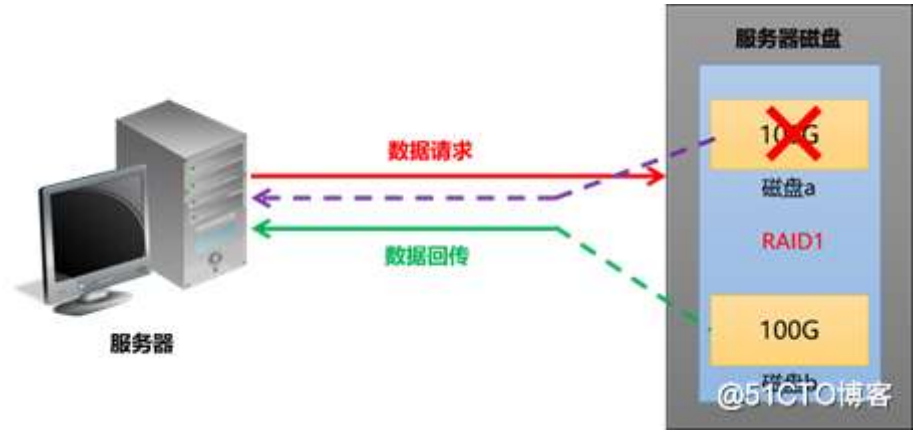
RAID 1机制下，写入数据的工作原理如图2所示。

图2. RAID1写入数据过程



在RAID1机制下，磁盘损坏时数据读取的原理如图3所示。

图3. RAID1读取数据过程



2.2 2个节点服务器采用超融合架构容易发生脑裂，最好采用传统IT架构

- 2个节点方案架构

超融合产品集群的最小规模取决于副本存放策略，如果采用两副本的存储策略，最少3个节点起步。采用2节点也不是不可以，只不过2节点集群在数据安全性方案会有一定隐患。因此即便使用2节点配置，也是在一些开发测试环境下使用，正常的业务系统一般都是2副本3节点起，或者3副本5节点起。具体原因有以下两点：

- 1) 一般2节点的集群部署虚拟机，平均资源占用达到了每个节点50%以上，一旦其中一个节点出现故障宕机，正常的节点的资源不能完全接管集群全部虚拟机，虚拟机会出现非正常关闭的情况；
- 2) 2节点的架构下，集群出现脑裂情况的话，缺少另外一个节点进行仲裁，那么整个集群的虚拟机会出现异常，集群会出现崩溃。

所以，在2个节点的情况之下，也不建议使用超融合架构，最好还是采用传统的IT架构，即“服务器虚拟化+RAID1机制”，如图4所示。

图4. 2节点服务器方案架构



- 2个节点数据可靠性方案 —— RAID机制

注意，在2个节点服务器方案下，RAID1机制只能实现硬盘级冗余保护，无法实现服务器级冗余保护，也就是说在其中一个节点发生宕机的情况之下，无法挽救宕机的服务器的数据。

2.3 3个节点服务器起步采用超融合架构

- 超融合方案架构

“超融合”架构方案与1个节点或2个节点方案比起来，多了一个“分布式存储”软件，它其实就是“服务器虚拟化+分布式存储”架构，方案架构如图5所示。

[Prometheus\(52\)](#)
[Rancher\(44\)](#)
[Redis\(1\)](#)
[ServiceMesh\(1\)](#)
[Ubuntu\(18\)](#)
[Windows\(2\)](#)
[Zabbix\(1\)](#)
[常用工具\(24\)](#)
[代码管理\(5\)](#)
[更多](#)

随笔档案

[2022年10月\(4\)](#)
[2022年6月\(3\)](#)
[2022年5月\(17\)](#)
[2022年4月\(25\)](#)
[2022年3月\(68\)](#)
[2022年2月\(18\)](#)
[2021年12月\(1\)](#)
[2021年11月\(1\)](#)
[2021年10月\(2\)](#)
[2021年9月\(50\)](#)
[2021年8月\(273\)](#)
[2021年7月\(79\)](#)
[2020年11月\(58\)](#)
[2020年10月\(58\)](#)

阅读排行榜

1. [CentOS7安装源配置详解\(9876\)](#)
2. [leaf-spine（叶脊）网络架构介绍\(7105\)](#)
3. [PVE 备份和恢复介绍\(5386\)](#)
4. [Proxmox Virtual Environment 介绍\(4947\)](#)
5. [Proxmox VE Qemu/KVM 虚拟机设置要求\(4533\)](#)
6. [Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）\(4297\)](#)
7. [PVE 虚拟机模板和克隆介绍\(3936\)](#)
8. [【K8s教程】Nginx Ingress控制器Ingress Path匹配\(3843\)](#)
9. [CentOS7安装配置远程桌面详解\(3631\)](#)
10. [Debian安装源配置详解\(3457\)](#)
11. [Apollo配置中心使用详解\(3356\)](#)
12. [Proxmox VE 防火墙介绍\(3323\)](#)
13. [Gitea 简单使用说明\(3313\)](#)
14. [Linux系统路由配置说明\(3270\)](#)
15. [Chrony时间同步服务器安装配置说明\(3256\)](#)
16. [Centos7防火墙配置详解\(3231\)](#)
17. [【Rancher】简单介绍\(3230\)](#)
18. [Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）\(3098\)](#)
19. [PVE 超融合架构介绍\(3096\)](#)
20. [【K8s任务】为容器设置启动时要执行的命令和参数\(3071\)](#)

评论排行榜

1. [Proxmox VE 存储介绍\(1\)](#)
2. [Proxmox Virtual Environment 介绍\(1\)](#)
3. [Debian Jessie Preseed自动化安装示例\(1\)](#)
4. [Debian live-build官方教程学习笔记\(1\)](#)
5. [Ceph存储池pg_num配置详解\(1\)](#)

推荐排行榜

1. [leaf-spine（叶脊）网络架构介绍\(1\)](#)
2. [Proxmox VE 防火墙介绍\(1\)](#)
3. [Proxmox Virtual Environment 介绍\(1\)](#)
4. [VIM 常用快捷键\(1\)](#)
5. [Debian7 bonding配置详解\(1\)](#)

最新评论

1. [Re:KVM虚拟化Disk Cache模式对性能的影响分析](#)
如果从“Default (No cache)”修改为“Write back (unsafe)”，对磁盘数据有影响么？
--大熊子
2. [Re:Proxmox VE 存储介绍](#)
请教一下：原来用在VMWARE集群的EMC FC-SAN存储（不支持ISCSI），如果改用X做集群，FC-SAN应该以什么方式挂载给PRC...人好呢？

图5. 超融合方案架构



超融合产品集群的最小规模取决于副本存放策略，如果采用两副本的存储策略，最少3个节点起步，因为要确保集群多数票机制生效。

- 超融合数据可靠性方案 —— 数据副本方式

数据副本方式实现数据可靠性。Ceph分布式存储采用数据副本方式实现数据可靠性，默认情况下，Ceph采用2副本3节点起步，在一个或者多个磁盘故障的情况之下，该磁盘上所有的数据马上开始从对等磁盘上进行恢复。因为Ceph是分布式系统，所以数据恢复时所有的初始副本和复制副本可以分布到集群中所有的磁盘上，使得不会有初始副本和复制副本位于同一个磁盘的情况。超融合数据副本方式原来如图6、图7、图8、图9、图10所示。

图6. 数据初始状态



如图6所示，节点服务器为3节点，数据D1、D2、D3为双副本，分别存储于三台超融合服务器1、2、3上。

图7.节点3宕机故障



如图7所示，在某一时刻，超融合服务器3发生故障，导致超融合服务器3不可用，这个时候数据D2、D3可靠性降低了；

图8. 重建D2和D3



如图8所示，超融合的集群发现了D2、D3副本丢失，由于是双副本机制，所以集群在超融合服务器1和超融合服务器2上重建D2、D3；

图9. 节点3故障回复



如图9所示，若超融合服务器3故障恢复并加入到集群中，D2、D3会被重新迁移回超融合服务器3上。

图10. 集群恢复正常状态



如图10所示，当D2、D3重新迁移回超融合服务器3之后，整个集群恢复为正常状态。

- 超融合架构的优势

--9394952
3. Re:Proxmox Virtual Environment 介绍
很全面！
--willblog
4. Re:Ceph存储池pg_num配置详解
mark
--穷死了
5. Re:Debian live-build官方教程学习笔记
执行成功之后为什么没找看iso
--abraham_liuhai

- 1) 数据副本方式实现数据可靠性，无需任何特殊硬件来提供数据副本功能。而RAID机制的话，需要独立的价格昂贵的RAID硬件卡，这就增加了系统总成本。
- 2) 硬盘利用率达到100%，由于超融合架构不采用RAID机制，所以硬盘利用率达100%。如果采用RAID机制，RAID需要额外的磁盘来充当备用盘，这些备用品只能用来备份数据。例如，RAID机制下，硬盘利用率最高的是RAID5= (n-1) /n的总磁盘容量（n为磁盘数），可靠性最高的RAID1的硬盘利用率只有50%。而在超融合架构下，一个磁盘既可以是某些数据的初始副本盘，也可以是另外一些数据的复制副本盘，这就大大提高了磁盘利用率。
- 3) 故障恢复过程透明，恢复时间快捷。在超融合架构下，数据副本方式恢复过程对客户来说是透明的，就是在不影响业务的情况之下，进行恢复，恢复时间一般为分钟级别。而RAID机制不同，当1个30多TB的节点做了RAID1机制之后，故障恢复时间一般为小时级别的。
- 4) 集群高可用，当1个节点宕机之后，集群仍可正常工作。在超融合架构下，集群高可用至少3个节点起步，当其中1个节点宕机之后，集群高可用可以将故障节点服务器的虚拟机迁移到无故障的节点服务器，直到故障节点恢复，这一切无需人工干预。而RAID机制下，无法实现集群高可用，分为各个节点的存储是独立的，不像分布式存储那样可以跨节点统一管理存储，一旦某个节点发生宕机，那么这个故障节点的虚拟机、数据等都不会迁移到正常的节点服务器中。

作者: [Varden](#)
出处: <http://www.cnblogs.com/varden/>
本文内容如有雷同，请联系作者！
本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类: [服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

- « 上一篇: [\(案例二\) Proxmox VE 备份组件 PBS \(Proxmox Backup Server\) 部署及使用教程](#)
- » 下一篇: [Proxmox VE + Ceph 超融合项目实战 \(第一、二部分: 网络拓扑和服务器配置\)](#)

posted @ 2021-08-30 15:19 Varden 阅读(3097) 评论(0) 编辑 收藏 举报

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

- 编辑推荐:**
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
 - [分布式事务 | 使用 DTM 的 Saga 模式](#)
 - [SQLSERVER 阻塞之 PFS 页到底是什么?](#)
 - [带团队后的日常思考 \(十一\)](#)
 - [记一次使用 gdb 诊断 gc 问题全过程](#)
- 阅读排行:**
- [chatGPT vscode 体验 \(密钥已更新当天有效\)](#)
 - [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
 - [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
 - [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
 - [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

图2. 白牌服务器Baixxxx-Cloud-X10



白牌服务器Baixxxx-Cloud-X10的技术规格如表1所示：

表1. Baixxxx-Cloud-X10的技术规格

选项	技术规格	数量
CPU	XEON E5-2683V4 2.1G 16 核心 32 线程	2
CPU 散热	2U 2011 针 主动式铜管散热器	2
主板	X10DRI C612 双路 2011 针服务器主板	1
内存	三星 32G 2400MHZ RECC 服务器内存	4
闪存	INTEL 480G 2.5 寸 SSD	2
闪存	INTEL 1.92TB 2.5 寸 SSD	2
机械盘	希捷 8TB ES 企业级行货	4
网卡	UNICACA INTEL X540 芯片双千兆	1
阵列卡	DDELL H730P 2G 带电容 8643 数据线	1
机箱	定制 4U 四路 GPU 服务器机箱	1
导轨	定制 4U 专用服务器导轨	1
电源	康舒 额定功率 1600W 1+1 双电冗余 服务器电源	1

2.1.RAID规划

本项目中，白牌服务器Baixxxx-Cloud-X10是八盘位，如图3所示。

图3. 白牌服务器Baixxxx-Cloud-X10的盘位



白牌服务器Baixxxx-Cloud-X10有八个盘位，一共有2 块1.9TB SSD盘，2块480GB SSD盘，4块8TB HDD 盘。

2.1.1 系统盘

在每个节点服务器上，2块480GB SSD盘组成RAID1，作为系统盘。做RAID1之后，2块480GB SSD盘相互备份，容量只有一半，系统文件存放在RAID1盘中。除了2块480GB SSD盘做RAID1之外，其余SSD盘、HDD盘均不做RAID。

2.1.2 Ceph盘

本项目中，后端存储采用Ceph分布式存储搭建，为了保证可靠性，Ceph采用了数据多副本的存储方式，这意味着不再需要RAID，也就克服了 RAID 存在的诸多问题。换句话说，Ceph已经解决了容错问题，无需使用RAID。

红帽官网文档有关于构建Ceph存储集群时，不建议基于RAID盘创建OSD，除了系统盘之外，其余磁盘全部设置为非RAID模式。也就是说，Ceph不要与RAID混合使用，如果已经使用了Ceph的话，应该禁用RAID，将RADI 控制器配置为RAID0，如图4所示。

Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

- CentOS7安装源配置详解(9876)
- leaf-spine （叶脊）网络架构介绍(7105)
- PVE 备份和恢复介绍(5386)
- Proxmox Virtual Environment 介绍(4947)
- Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
- Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
- PVE 虚拟机模板和克隆介绍(3936)
- 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
- CentOS7安装配置远程桌面详解(3631)
- Debian安装源配置详解(3457)
- Apollo配置中心使用详解(3356)
- Proxmox VE 防火墙介绍(3323)
- Gitea 简单使用说明(3313)
- Linux系统路由配置说明(3270)
- Chrony时间同步服务器安装配置说明(3256)
- Centos7防火墙配置详解(3231)
- 【Rancher】简单介绍(3230)
- Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
- PVE 超融合架构介绍(3096)
- 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

- Proxmox VE 存储介绍(1)
- Proxmox Virtual Environment 介绍(1)
- Debian Jessie Preseed自动化安装示例(1)
- Debian live-build官方教程学习笔记(1)
- Ceph存储池pg_num配置详解(1)

推荐排行榜

- leaf-spine （叶脊）网络架构介绍(1)
- Proxmox VE 防火墙介绍(1)
- Proxmox Virtual Environment 介绍(1)
- VIM 常用快捷键(1)
- Debian7 bonding配置详解(1)

最新评论

- Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从"Default (No cache)"修改为"Write back (unsafe)", 对磁盘数据有影响么?

--大熊猫子

- Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的存储是EMC FC-SAN存储（不支持ISCSI），如果改用Ceph做集群，FC-SAN应该以什么方式挂载给Proxmox好呢?

图4. 红帽官网Ceph章节对RAID使用建议



2.2 服务器网卡规划

本项目中，白牌Baixxxx-Cloud-X10服务器一共有5张物理网卡，物理网卡的IP 地址规划如图5所示。

图5. Baixxxx-Cloud-X10服务器网卡规划



2.3 服务器IPMI接口

IPMI是指一组用于带外管理的计算机接口规范。带外是指无需与系统的物理资产位于同一区域即可访问计算机系统。IPMI 支持远程监视，不需要计算机操作系统的许可。IPMI 在连接到主板或服务器的单独硬件上运行。此单独的硬件是底板管理控制器（BMC）。

通常情况下，白牌Baixxxx-Cloud-X10服务器的IPMI 接口的默认用户名为：ADMIN，密码则贴在机箱前面板上，或者是机箱中的主板上。IPMI 主板集成管理芯片BMC的IP的首次设置，需要在服务器开启时按“DEL”键进入BIOS 设置，具体步骤略。如对该步骤设置感兴趣，请参见《GPU直通+单节点单业务vlan的虚拟网络项目实战》博文。

当IPMI接口的IP地址设置好之后，可通过WEB UI进入查看服务器监控系统，如节点1服务器的IPMI接口的IP地址是10.129.255.21，使用https登录进入IPMI监控系统，如图6所示。

图6. 服务器IPMI监控系统界面



2.4 服务器数据接口

有关服务器数据接口的IP 地址规划，详见表2。具体配置详见《5.3.2.业务网络配置》的章节，这里不再赘述。

2.5 服务器管理接口

有关服务器管理接口的IP 地址规划，详见表2。具体配置详见《5.3.1.管理网络配置》的章节，这里不再赘述。

作者：[Varden](#)

出处：<http://www.cnblogs.com/varden/>

本文内容如有雷同，请联系作者！

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类：[服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

我在关注他 [取消关注](#)

0

0

--9394952

[3. Re:Proxmox Virtual Environment 介绍](#)
很全面！

--willblog

[4. Re:Ceph存储池pg_num配置详解](#)
mark

--穷死了

[5. Re:Debian live-build官方教程学习笔记](#)
执行成功之后为什么没找看iso

--abraham_liuhai

(评论功能已被禁用)

编辑推荐:

- [如何避免让线程摸鱼, 请用异步技术 async await 拿捏他](#)
- [分布式事务 | 使用 DTM 的 Saga 模式](#)
- [SQLSERVER 阻塞之 PFS 页到底是什么?](#)
- [带团队后的日常思考 \(十一\)](#)
- [记一次使用 gdb 诊断 gc 问题全过程](#)

阅读排行:

- [chatGPT vscode 体验 \(密钥已更新当天有效\)](#)
- [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
- [如何避免让线程摸鱼, 请用异步技术 async await 拿捏他~](#)
- [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
- [不用Blazor WebAssembly, 开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 RML

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
-取消关注

2023年2月						
日	一	二	三	四	五	六
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	1	2	3	4
5	6	7	8	9	10	11

Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）

参考: https://blog.51cto.com/u_14742102/3425275

三、PVE安装与基础配置

3.1 PVE安装七步曲

本项目中，3个节点服务器都需要安装PVE虚拟化平台，限于篇幅，这里只演示节点1服务器的PVE安装过程。

- 第一步：选择红框中的选项，敲回车键，安装PVE，如图7所示。

图7. 选择安装PVE



- 第二步：选择红框中的选项，同意协议，如图8所示。

图8. 同意安装协议



- 第三步：在红框④中选择系统盘，也就是在RAID规划中创建的RAID1系统盘，接着选择下一步，如图9所示。注意，由于在做项目的时候，PVE安装的过程忘记截图，《3.1.PVE安装七步曲》采用的是虚拟机来安装PVE进行项目还原，红框④中的参数与实际有所不同，大家知道这个意思就行了，下面就不再重复，见谅。

搜索

找找看

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

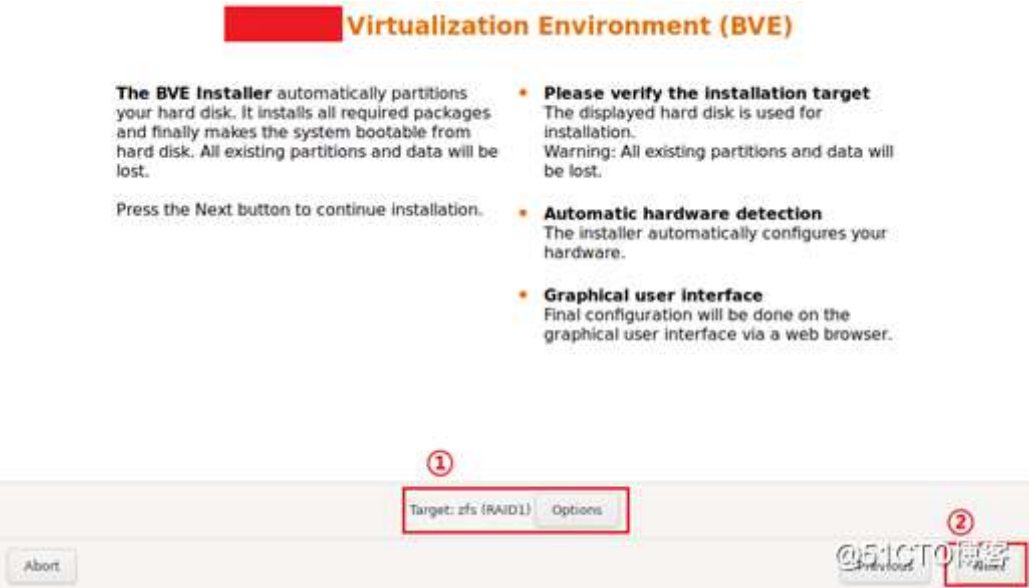
最新随笔

- 1.Ceph 故障域配置（修改CRUSH MAP）
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类（Octopus版本）

随笔分类

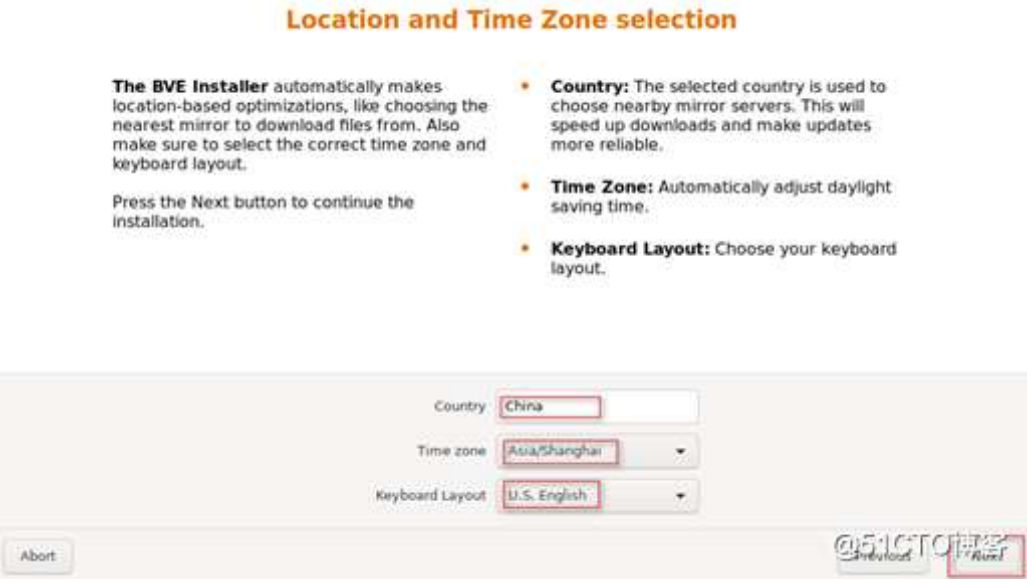
Alpine(2)
Ansible(23)
CentOS(23)
Ceph(118)
Debian(19)
Docker(28)
ELK日志分析(15)
Helm(3)
Java(1)
Jenkins(10)
K8s(209)
KubectI(4)
Linux集群技术(3)
Linux企业常用服务(9)
Linux企业级安全(5)
Linux系统调优和监控(8)
Linux系统基础及管理(38)
MongoDB(12)
MySQL/MariaDB(6)
NFS(3)
PostgreSQL(1)

图9. 选择安装盘



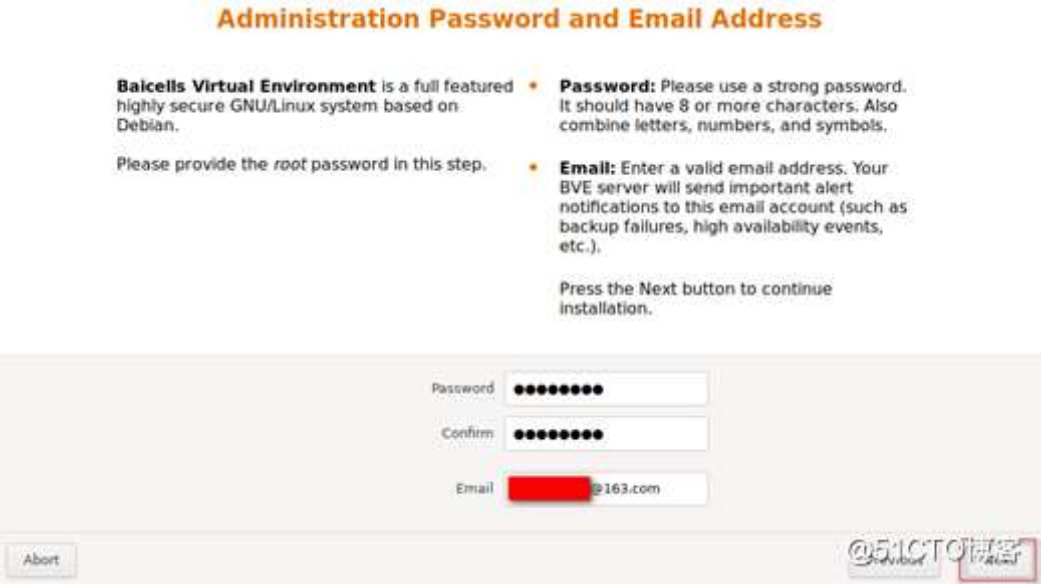
- 第四步：配置国家和地区信息，如图10所示。国家选择“China”，时区选择“亚洲/上海”，键盘布局选择“美式键盘”。

图10. 选择国家和地区



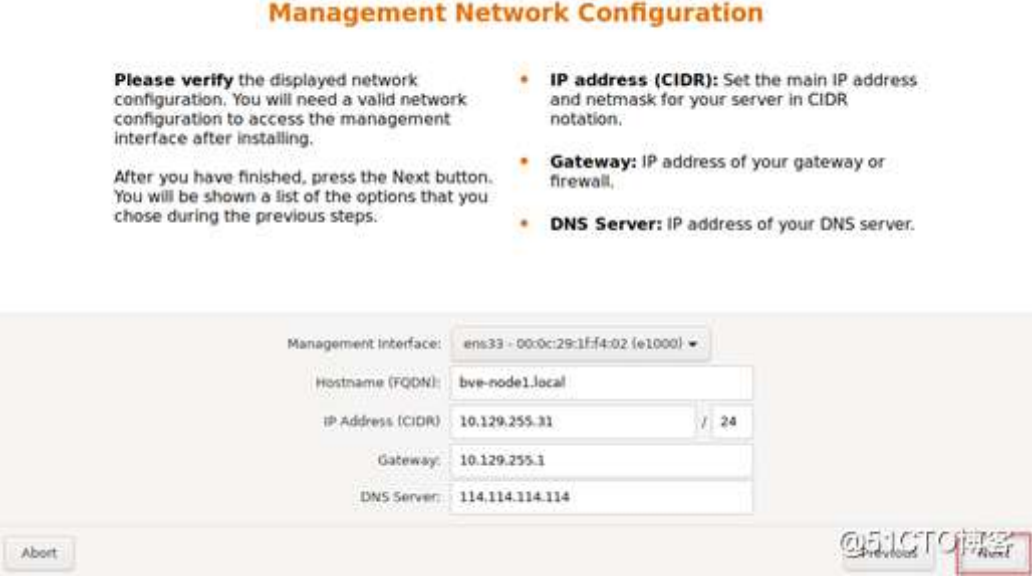
- 第五步：配置密码和邮箱信息，如图11所示。这里的密码是“root”权限的密码，用户名也是“root”。邮箱只需要是合法的都没问题，不需要验证邮箱。

图11. 配置密码和邮箱



- 第六步：配置网卡、主机名、IP地址、DNS等信息，如图12所示。其中，选择第一块网卡用于网络管理，按照我们的规划应该选择“eno2”网卡。主机名称“bve-node1”。IP地址和DNS等信息可配置也可不配置，留待稍后配置也可以，但是最好是配置IP地址，如果需要修改，可以使用Web UI进行修改，这样比较简单。

图12. 配置网卡、主机名、IP地址



- 第七步：检查核对信息，确定无误后点击安装，如图13所示。

Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

1. CentOS7安装源配置详解(9876)
2. leaf-spine （叶脊）网络架构介绍(7105)
3. PVE 备份和恢复介绍(5386)
4. Proxmox Virtual Environment 介绍(4947)
5. Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
6. Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
7. PVE 虚拟机模板和克隆介绍(3936)
8. 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
9. CentOS7安装配置远程桌面详解(3631)
10. Debian安装源配置详解(3457)
11. Apollo配置中心使用详解(3356)
12. Proxmox VE 防火墙介绍(3323)
13. Gitea 简单使用说明(3313)
14. Linux系统路由配置说明(3270)
15. Chrony时间同步服务器安装配置说明(3256)
16. Centos7防火墙配置详解(3231)
17. 【Rancher】简单介绍(3230)
18. Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
19. PVE 超融合架构介绍(3096)
20. 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

1. Proxmox VE 存储介绍(1)
2. Proxmox Virtual Environment 介绍(1)
3. Debian Jessie Preseed自动化安装示例(1)
4. Debian live-build官方教程学习笔记(1)
5. Ceph存储池pg_num配置详解(1)

推荐排行榜

1. leaf-spine （叶脊）网络架构介绍(1)
2. Proxmox VE 防火墙介绍(1)
3. Proxmox Virtual Environment 介绍(1)
4. VIM 常用快捷键(1)
5. Debian7 bonding配置详解(1)

最新评论

1. Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从“Default (No cache)”修改为“Write back (unsafe)”，对磁盘数据有影响么？
--大熊猫子
2. Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的以EMC FC-SAN存储（不支持ISCSI），如果改用Ceph做集群，FC-SAN应该以什么方式挂载给Proxmox好呢？

图13. 检查核对信息



安装结束之后，出现图14界面证明安装成功。默认用户名为root，密码为安装时设置的密码。

图14. PVE操作系统命令行登录



安装结束之后自动重启，如果通过浏览器访问https://10.129.255.31:8006，出现图15界面证明安装成功。

图15. PVE操作系统Web登录



其他两个节点服务器也是用PVE安装七步曲的方法来完成PVE安装。

3.2 PVE基础配置

3.2.1 安装ifconfig

- 第一步：更新apt资源

```
root@pve-node1:~# apt-get update
root@pve-node1:~# apt-get upgrade
```

- 第二步：安装net-loos网络工具

```
root@pve-node1:~# apt install net-tools
```

- 第三步：使用ifconfig

```
root@pve-node1:~# ifconfig          #查看主机当前生效的网卡
root@pve-node1:~# ifconfig -a      #查看主机所有的网卡
```

3.2.2 安装vim

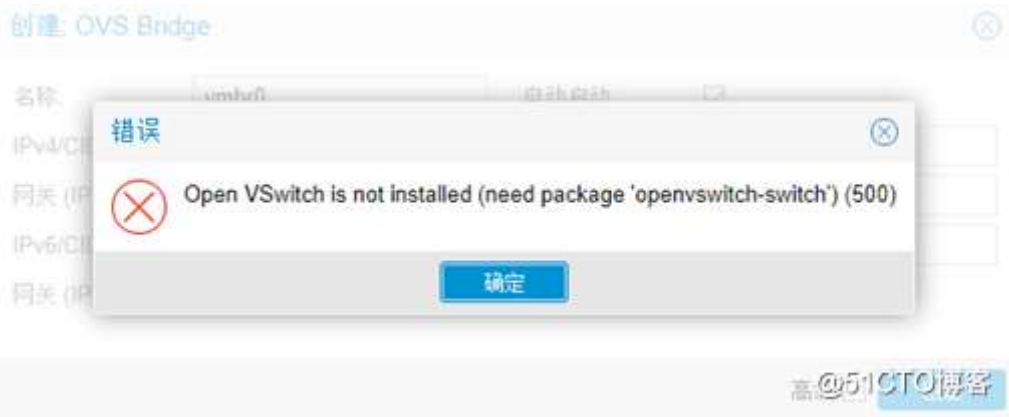
```
root@pve-node1:~# apt install vim
```

3.2.3 安装Open vSwitch虚拟交换机

Proxmox VE默认情况下，并没有安装Open vSwitch虚拟交换机，如图16所示。

- 9394952
- 3. Re:Proxmox Virtual Environment 介绍很全面！
- willblog
- 4. Re:Ceph存储池pg_num配置详解mark
- 穷死了
- 5. Re:Debian live-build官方教程学习笔记执行成功之后为什么没找看iso
- abraham_liuhai

图16. Proxmox VE默认没有安装Open vSwitch



后续配置虚拟网络需要使用到Open vSwitch虚拟交换机，需要安装，安装操作如下所示：

```
root@pve-node1:~# apt-get update      # 3.2.1中已执行
root@pve-node1:~# apt install openvswitch-switch
```

3.2.4 配置ntp服务

如果组建多机Proxmox集群，如果使用Ceph分布式存储，请务必确保所有服务器的时间同步。如果NTP时钟不同步，Ceph服务可能会起不来，登录认证失败，虚拟机迁移失败等等问题。

- 第一步：配置国家统一的时钟源

阿里云：ntp.aliyun.com
国家授时中心：ntp.ntsc.ac.cn

```
root@pve-node1:~# vim /etc/systemd/timesyncd.conf
```

图17. 配置时钟源



- 第二步：重启NTP服务

```
root@pve-node1:~# systemctl restart systemd-timesyncd
```

作者：[Varden](#)
出处：<http://www.cnblogs.com/varden/>
本文内容如有雷同，请联系作者！
本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类：[服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

« 上一篇：[Proxmox VE + Ceph 超融合项目实战（第一、二部分：网络拓扑和服务器配置）](#)
» 下一篇：[Proxmox VE + Ceph 超融合项目实战（第四部分：物理网络）](#)

posted @ 2021-08-30 16:10 [Varden](#) 阅读(3098) 评论(0) [编辑](#) [收藏](#) [举报](#)
[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

- 编辑推荐：**
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
 - [分布式事务 | 使用 DTM 的 Saga 模式](#)
 - [SQLSERVER 阻塞之 PFS 页到底是什么？](#)
 - [带团队后的日常思考（十一）](#)
 - [记一次使用 gdb 诊断 gc 问题全过程](#)
- 阅读排行：**
- [chatGPT vscode 体验（密钥已更新当天有效）](#)
 - [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
 - [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
 - [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
 - [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

[首页](#)
[新随笔](#)
[联系](#)
[管理](#)
[订阅](#)
[RML](#)

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: [Varden](#)
 园龄: [2年3个月](#)
 粉丝: [42](#)
 关注: [2](#)
[-取消关注](#)

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

最新随笔

- 1.Ceph 故障域配置 (修改CRUSH MAP)
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类 (Octopus版本)

随笔分类

Alpine(2)
 Ansible(23)
 CentOS(23)
 Ceph(118)
 Debian(19)
 Docker(28)
 ELK日志分析(15)
 Helm(3)
 Java(1)
 Jenkins(10)
 K8s(209)
 Kubectl(4)
 Linux集群技术(3)
 Linux企业常用服务(9)
 Linux企业级安全(5)
 Linux系统调优和监控(8)
 Linux系统基础及管理(38)
 MongoDB(12)
 MySQL/MariaDB(6)
 NFS(3)
 PostgreSQL(1)

Proxmox VE + Ceph 超融合项目实战（第四部分：物理网络）

参考: https://blog.51cto.com/u_14742102/3425275

四、物理网络

本项目超融合平台主要由3个节点服务器集群组建，且使用Ceph搭建分布式存储，为保证平台可靠性要求，为保证性能要求，为保证虚拟机迁移速率等，强烈建立采用全万兆交换机部署。同时，IPMI网络单独部署交换机、业务网络单独部署交换机、管理网络单独部署交换机、存储网络单独部署交换机，但是由于客户预算有限，本项目只采用了1台三层全千兆交换机，同时在千兆交换机上划分不同VLAN以满足IPMI网络、业务网络、管理网络、存储网络的需求。

本项目主要是面向研发测试人员，除了对PVE超融合平台可靠性要求高之外，研发测试人员往往也比较关注PVE超融合平台的弹性化、自动化、分布式以及监控等问题。

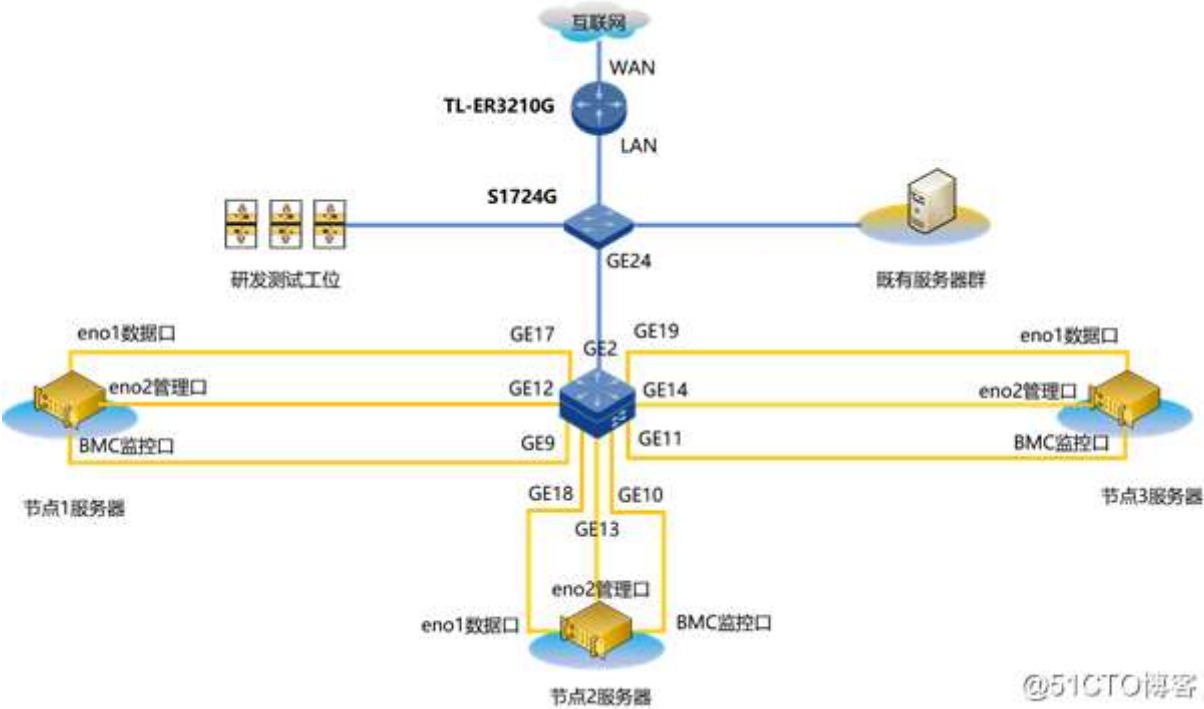
4.1 物理网络拓扑

某某公司超融合项目中的物理网络设备主要包括S1724G交换机若干台，普联TL-ER3210G网关1台，既有服务器6台，PVE节点服务器3台、研发测试工位约40个。

本项目新建的研发测试超融合平台通过新增三层交换机S5110V2-28P-SI，直接接入到既有网络S1724G交换机，这样做是为了保持既有网络配置不变，而不是为了达到最优的网络性能。

由于客户现场网络环境受限，新增的三层交换机S5110V2-28P-SI只能接入到既有二层交换机S1724G，这种“核心”与“接入”倒挂的网络架构本身是不合理的，但是没有办法，客户现场网络环境只能如此组网。某某公司物理网络拓扑如图18所示。

图18. 某某公司物理网络拓扑



4.2 VLAN及IP规划

- 华三的三层交换机S5110V2-28P-SI以及既有网络设备、服务器主机以及虚拟机等的VLAN划分以及IP地址规划如表2所示。

表2. 网络设备的VLAN划分及IP分配

选项	端口	IP 地址	VLAN ID	对应设备及端口
网关 TL-ER3210G	WAN		/	TL-ER3210G WAN→移动 OUN 设备
	LAN	192.168.10.1/24	Vlan1	TL-ER3210G LAN→S1724G GXX 口
交换机 S1724G	GEXX	/	Vlan1	S1724G GXX 口→TL-ER3210G LAN
	GE24	/	Vlan1	S1724G G24 口→S5110V2-28P-SI G1/0/2
交换机 S5110V2-28P-SI	int-vlan1	192.168.10.253/24	Vlan1	虚拟接口，路由 VLAN
	int-vlan250	172.16.21.1/24	Vlan250	虚拟接口，业务 VLAN
	int-vlan251	172.16.20.1/24	Vlan251	虚拟接口，业务 VLAN
	int-vlan255	10.129.255.1/24	Vlan255	虚拟接口，管理 VLAN
	G1/0/2	/	Vlan1	S5110V2-28P-SI G1/0/2→S1724G G24 口
	G1/0/9	/	Vlan255	S5110V2-28P-SI G1/0/9→节点 1 BMC 口
	G1/0/10	/	Vlan255	S5110V2-28P-SI G1/0/10→节点 2 BMC 口
	G1/0/11	/	Vlan255	S5110V2-28P-SI G1/0/11→节点 3 BMC 口
	G1/0/12	/	trunk=252~255	S5110V2-28P-SI G1/0/12→节点 1 eno2 口
	G1/0/13	/	trunk=252~255	S5110V2-28P-SI G1/0/13→节点 2 eno2 口
	G1/0/14	/	trunk=252~255	S5110V2-28P-SI G1/0/14→节点 3 eno2 口
	G1/0/17	/	trunk=250,251	S5110V2-28P-SI G1/0/17→节点 1 eno1 口
	G1/0/18	/	trunk=250,251	S5110V2-28P-SI G1/0/18→节点 2 eno1 口
	G1/0/19	/	trunk=250,251	S5110V2-28P-SI G1/0/19→节点 3 eno1 口

- 节点1~3服务器集群的VLAN划分以及IP地址规划如表3所示。

表3. 节点1~3服务器的VLAN划分及IP分配

选项	端口	IP 地址	VLAN ID	对应设备及端口
节点 1 服务器	brlceph	10.129.252.31/24	Vlan252	with ceph ip
	brlceph_cluster	10.129.253.31/24	Vlan253	with ceph_cluster ip
	brlceph_public	10.129.254.31/24	Vlan254	with ceph_public ip
	brlmgnt	10.129.255.31/24	Vlan255	with mgnt ip
	eno1 口	/	trunk=250,251	节点 1 eno1 口→S5110V2-28P-SI G1/0/17
	eno2 口	/	trunk=252,253,254,255	节点 1 eno2 口→S5110V2-28P-SI G1/0/12
	BMC 口	10.129.255.21/24	Vlan255	节点 1 BMC 口→S5110V2-28P-SI G1/0/9
节点 2 服务器	brlceph	10.129.252.32/24	Vlan252	with ceph ip
	brlceph_cluster	10.129.253.32/24	Vlan253	with ceph_cluster ip
	brlceph_public	10.129.254.32/24	Vlan254	with ceph_public ip
	brlmgnt	10.129.255.32/24	Vlan255	with mgnt ip
	eno1 口	/	trunk=250,251	节点 2 eno1 口→S5110V2-28P-SI G1/0/18
	eno2 口	/	trunk=252,253,254,255	节点 2 eno2 口→S5110V2-28P-SI G1/0/13
	BMC 口	10.129.255.22/24	Vlan255	节点 2 BMC 口→S5110V2-28P-SI G1/0/10
节点 3 服务器	brlceph	10.129.252.33/24	Vlan252	with ceph ip
	brlceph_cluster	10.129.253.33/24	Vlan253	with ceph_cluster ip
	brlceph_public	10.129.254.33/24	Vlan254	with ceph_public ip
	brlmgnt	10.129.255.33/24	Vlan255	with mgnt ip
	eno1 口	/	trunk=250,251	节点 3 eno1 口→S5110V2-28P-SI G1/0/19
	eno2 口	/	trunk=252,253,254,255	节点 3 eno2 口→S5110V2-28P-SI G1/0/14
	BMC 口	10.129.255.23	Vlan255	节点 3 BMC 口→S5110V2-28P-SI G1/0/11

- 子系统虚拟机的VLAN划分以及IP地址规划如表4所示。

Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

- CentOS7安装源配置详解(9876)
- leaf-spine （叶脊）网络架构介绍(7105)
- PVE 备份和恢复介绍(5386)
- Proxmox Virtual Environment 介绍(4947)
- Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
- Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
- PVE 虚拟机模板和克隆介绍(3936)
- 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
- CentOS7安装配置远程桌面详解(3631)
- Debian安装源配置详解(3457)
- Apollo配置中心使用详解(3356)
- Proxmox VE 防火墙介绍(3323)
- Gitea 简单使用说明(3313)
- Linux系统路由配置说明(3270)
- Chrony时间同步服务器安装配置说明(3256)
- Centos7防火墙配置详解(3231)
- 【Rancher】简单介绍(3230)
- Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
- PVE 超融合架构介绍(3096)
- 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

- Proxmox VE 存储介绍(1)
- Proxmox Virtual Environment 介绍(1)
- Debian Jessie Preseed自动化安装示例(1)
- Debian live-build官方教程学习笔记(1)
- Ceph存储池pg_num配置详解(1)

推荐排行榜

- leaf-spine （叶脊）网络架构介绍(1)
- Proxmox VE 防火墙介绍(1)
- Proxmox Virtual Environment 介绍(1)
- VIM 常用快捷键(1)
- Debian7 bonding配置详解(1)

最新评论

- Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从"Default (No cache)"修改为"Write back (unsafe)"，对磁盘数据有影响么？

--大熊猫子

- Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的PXC做FC-SAN存储（不支持ISCSI），如果改用PXC做集群，FC-SAN应该以什么方式挂载给PRC用户好呢？

表4. 虚拟机的VLAN划分及IP分配

选项	虚拟机	IP 地址	VLAN ID	对应设备及端口
节点1服务器	VM1	172.16.20.101	Vlan251	VM1 eth0 → 节点1 eno1 口
	VM2	172.16.20.102	Vlan251	VM2 eth0 → 节点1 eno1 口
	VM3	172.16.20.103	Vlan251	VM3 eth0 → 节点1 eno1 口
	VM4	172.16.20.104	Vlan251	VM4 eth0 → 节点1 eno1 口
	VM5	172.16.20.105	Vlan251	VM5 eth0 → 节点1 eno1 口
	VM6	172.16.20.106	Vlan251	VM6 eth0 → 节点1 eno1 口
	VM7	172.16.20.107	Vlan251	VM7 eth0 → 节点1 eno1 口
节点2服务器	VM8	172.16.21.108	Vlan250	VM8 eth0 → 节点2 eno1 口
	VM9	172.16.21.109	Vlan250	VM9 eth0 → 节点2 eno1 口
	VM10	172.16.21.110	Vlan250	VM10 eth0 → 节点2 eno1 口
	VM11	172.16.21.111	Vlan250	VM11 eth0 → 节点2 eno1 口
	VM12	172.16.21.112	Vlan250	VM12 eth0 → 节点2 eno1 口
	VM13	172.16.21.113	Vlan250	VM13 eth0 → 节点2 eno1 口
	VM14	172.16.21.114	Vlan250	VM14 eth0 → 节点2 eno1 口
节点3服务器	VM15	172.16.21.115	Vlan250	VM15 eth0 → 节点3 eno1 口
	VM16	172.16.21.116	Vlan250	VM16 eth0 → 节点3 eno1 口
	VM17	172.16.21.117	Vlan250	VM17 eth0 → 节点3 eno1 口
	VM18	172.16.21.118	Vlan250	VM18 eth0 → 节点3 eno1 口
	VM19	172.16.21.119	Vlan250	VM19 eth0 → 节点3 eno1 口
	VM20	172.16.21.120	Vlan250	VM20 eth0 → 节点3 eno1 口
	VM21	172.16.21.121	Vlan250	VM21 eth0 → 节点3 eno1 口

4.3.网络配置

4.3.1 网关TL-ER3210G配置

- 某某公司既有网关TL-ER3210G的WAN口配置如图19所示：

图19. 网关TL-ER3210G的WAN口配置



- 某某公司既有网关TL-ER3210G的LAN口配置如图20所示：

图20. 网关TL-ER3210G的LAN口配置



- 网关TL-ER3210G的北向路由配置一条静态默认路由，实现与互联网互通，如图21所示。

图21. 网关TL-ER3210G的北向路由



--9394952

3. Re:Proxmox Virtual Environment 介绍
很全面！

--willblog

4. Re:Ceph存储池pg_num配置详解
mark

--穷死了

5. Re:Debian live-build官方教程学习笔记
执行成功之后为什么没找看iso

--abraham_liuhai

- 网关TL-ER3210G的南向路由配置两条静态路由指向业务VLAN250、VLAN251，如图22所示。

图22. 网关TL-ER3210G的南向路由

策略路由

静态路由

系统路由

静态路由

新增

删除

☐	序号	规则名称	目的地址	子网掩码	下一跳	出接口	Metric	可达性	状态	设置
☐	1	大数据业务系统	172.16.21.0	255.255.255.0	192.168.10.253	LAN	0	可达	已启用	 
☐	2	大数据中心集群	172.16.20.0	255.255.255.0	192.168.10.253	LAN	0	可达	已启用	 

2025

CTO博客

S5110V2-28P-SI 的北向路由配置一条静态默认路由，实现业务VLAN250、VLAN251与互联网互通，如下所示：

```
[H3C] ip route-static 0.0.0.0 0 192.168.10.1
```

- S5110V2-28P-SI 的南向路由配置：

S5110V2-28P-SI 的南向路由都是直连路由，业务VLAN250、VLAN251的终节点都在S5110V2-28P-SI，所以无需再配置南向静态路由。

4.3.3 研发测试终端路由设置

研发测试终端使用默认网关“192.168.10.1”上网，除此之外还需要访问VLAN250和VLAN251的虚拟机，需要增加三条路由：

```
route -p add 10.129.255.0 mask 255.255.255.0 192.168.10.253
route -p add 172.16.20.0 mask 255.255.255.0 192.168.10.253
route -p add 172.16.21.0 mask 255.255.255.0 192.168.10.253
```

作者：[Varden](#)
出处：<http://www.cnblogs.com/varden/>
本文内容如有雷同，请联系作者！
本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类：[服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

- « 上一篇：[Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）](#)
- » 下一篇：[Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）](#)

posted @ 2021-08-30 17:11 [Varden](#) 阅读(1818) 评论(0) [编辑](#) [收藏](#) [举报](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

- 编辑推荐：
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
 - [分布式事务 | 使用 DTM 的 Saga 模式](#)
 - [SQLSERVER 阻塞之 PFS 页到底是什么？](#)
 - [带团队后的日常思考（十一）](#)
 - [记一次使用 gdb 诊断 gc 问题全过程](#)
- 阅读排行：
- [chatGPT vscode 体验（密钥已更新当天有效）](#)
 - [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
 - [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
 - [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
 - [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 HTML

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
-取消关注

Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）

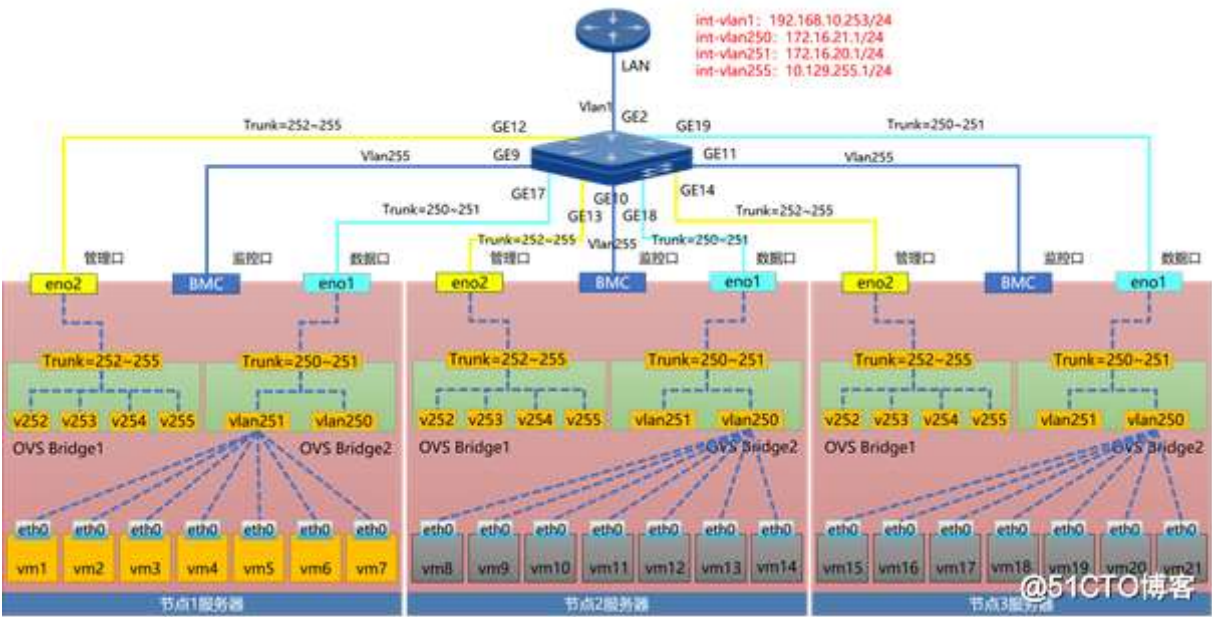
参考: https://blog.51cto.com/u_14742102/3425275

五、虚拟网络

5.1 PVE虚拟网络拓扑

某某公司超融合项目的虚拟网络的核心架构是在PVE虚拟化平台上创建虚拟交换机（OVS类型）、创建虚拟主机，并与物理交换机、物理路由器一起，组建管理网以及业务网。虚拟主机通过接入到虚拟交换机中实现相互通信，虚拟交换机接入到物理交换机中，物理交换机接入到物理路由器中，组成虚拟网络与物理网络互联互通的混合网络，如图23所示。

图23. 虚拟网络与物理网络混合架构



5.2 PVE虚拟网络简介

Proxmox VE基于Debian Linux操作系统，Debian Linux操作系统支持原生的虚拟网络有Linux Bridge、Linux Bond及Linux VLAN三种类型。Proxmox VE同时集成了Open vSwitch（OVS）开放虚拟交换机组件，所以Proxmox VE的原生虚拟网络除了Linux Bridge、Linux Bond及Linux VLAN这三种类型之外，还包括Open vSwitch类型，如图24所示。

图24. Proxmox VE原生虚拟网络模型



根据Proxmox VE官网的wiki文档《Open vSwitch》介绍，Open vSwitch和Linux Bridge、Linux Bond或Linux VLAN不能混合使用。例如，不要尝试将Linux VLAN添加到OVS Bond，或将Linux Bond添加到OVS Bridge，反之亦然。Open vSwitch是专门为在虚拟化环境中运行而定制的，无需使用主机自带的Linux Bridge、Linux Bond或Linux VLAN等功能，如图25所示。

2023年2月						
日	一	二	三	四	五	六
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	1	2	3	4
5	6	7	8	9	10	11

搜索

找找看

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

最新随笔

- 1.Ceph 故障域配置（修改CRUSH MAP）
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类（Octopus版本）

随笔分类

Alpine(2)
Ansible(23)
CentOS(23)
Ceph(118)
Debian(19)
Docker(28)
ELK日志分析(15)
Helm(3)
Java(1)
Jenkins(10)
K8s(209)
KubectI(4)
Linux集群技术(3)
Linux企业常用服务(9)
Linux企业级安全(5)
Linux系统调优和监控(8)
Linux系统基础及管理(38)
MongoDB(12)
MySQL/MariaDB(6)
NFS(3)
PostgreSQL(1)

图25. Open vSwitch不能与Linux原生虚拟网络混用

Installation

Update the package index and then install the Open vSwitch packages by executing:

```
apt update
apt install openvswitch-switch
```

Configuration

Official reference here, though a bit bare: <https://github.com/openvswitch/ovs/blob/master/debian/openvswitch-switch README Debian>

Overview

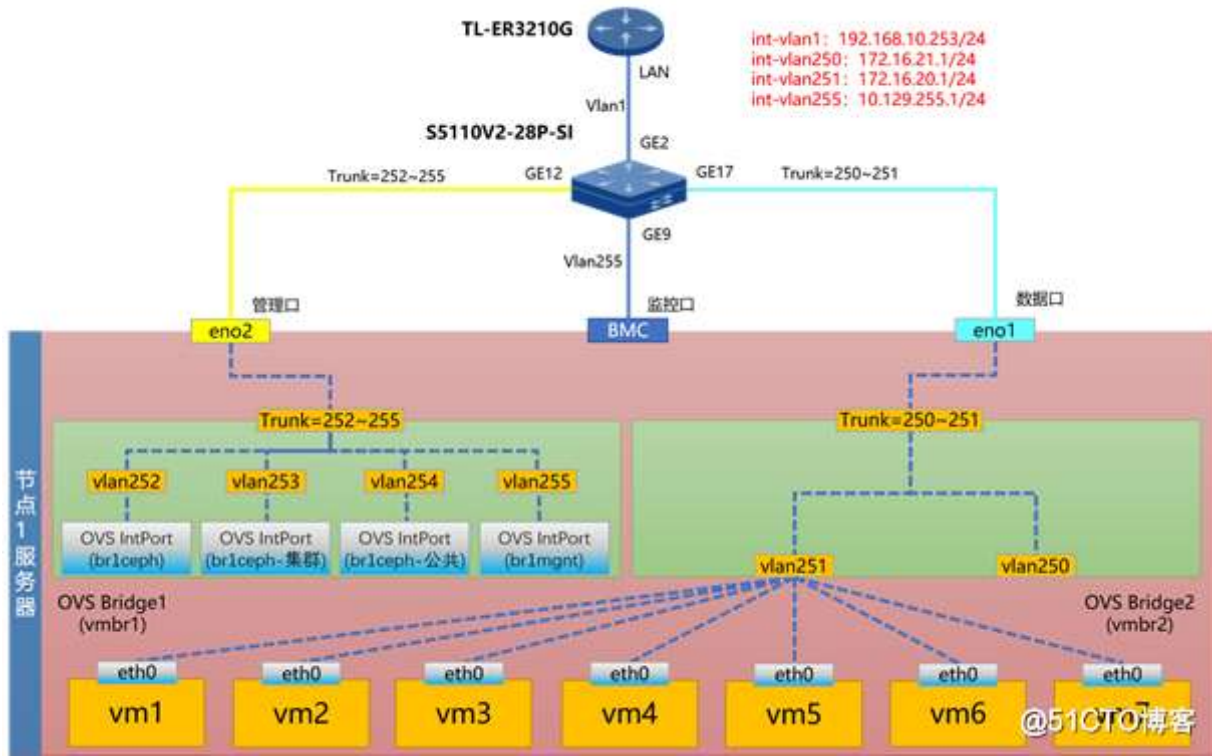
Open vSwitch and Linux bonding and bridging or vlans MUST NOT be mixed. For instance, do not attempt to add a vlan to an OVS Bond, or add a Linux Bond to an OVS Bridge or vice-versa. Open vSwitch is specifically tailored to function within virtualized environments, there is no reason to use the native linux functionality.

在本项目中，Proxmox VE的虚拟网络使用Open vSwitch，不使用Linux原生虚拟网络，后续不再做声明。

5.3 节点1的PVE虚拟网络配置

如图20中的虚拟网络与物理网络的混合架构看起来非常复杂，以致于不得不删减了一些跟集群相关的虚拟网，为了方便进行配置，有必要对图22的网络架构进行解构，进行还原，节点1的PVE虚拟网络如图26所示。

图26. 节点1的PVE虚拟网络架构



5.3.1 管理网络配置

根据表2和表3，管理网络的VLAN ID是VLAN 255，包括节点1的PVE系统的管理以及节点1服务器的BMC接口都统一划分到VLAN 255，其中VLAN 255的虚拟IP地址在S5110V2-28P-SI上。注意，在Proxmox VE语境下，Bridge（网桥）等同于Switch（交换机），后续不再做声明。在做配置演示之前，我先把节点1的虚拟网络配置截图出来，大家有个大概了解，不至于蒙圈，如图27所示。

图27. 节点1的虚拟网络配置

名称	类别	活动	自动	VLA	端口/位置	Bond	CIDR	网关	备注
br1ceph	OVS IntPort	是	否	否			10.129.252.31/24		with ceph ip
br1ceph_clus	OVS IntPort	是	否	否			10.129.253.31/24		with ceph_cluster ip
br1ceph_public	OVS IntPort	是	否	否			10.129.254.31/24		with ceph_public ip
br1mngt	OVS IntPort	是	否	否			10.129.255.31/24	10.12	with mngt ip
eno1	OVS Port	是	否	否					service
eno2	OVS Port	是	否	否					management and ceph
ens40	网络设备	否	否	否					
ens41	网络设备	否	否	否					
vmbr1	OVS Bridge	是	是	否	eno2 br1c...				
vmbr2	OVS Bridge	是	是	否	eno1				service, vlan250,251

好了，下面我们开始演示节点1的管理网络的配置，有了图27，相信大家很容易能理解下面这些步骤。

- 第一步：创建名称为“vmbr1”的OVS Bridge（虚拟交换机），并将“eno2”物理接口捆绑到虚拟交换机，如图28所示。

图28. 创建名称为vmbr1的虚拟交换机

编辑: OVS Bridge

名称: vmbr1

IPv4/CIDR:

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

自动启动: ☒

桥接端口: eno2

OVS选项:

备注:

高级 ☐

OK

取消

注意：eno1和eno2的类型是“网络设备”，这表明是物理接口。当物理接口eno1和eno2被捆绑到虚拟交换机之后，eno1和eno2的类型就会变成“OVS Port”，也就是成为虚拟交换机上的接口，这个接口是可以使用物理网线直接连接的，仍然属于物理接口，只不过在虚拟交换机上，变了个称呼，叫“OVS Port”。

Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

- CentOS7安装源配置详解(9876)
- leaf-spine （叶脊）网络架构介绍(7105)
- PVE 备份和恢复介绍(5386)
- Proxmox Virtual Environment 介绍(4947)
- Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
- Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
- PVE 虚拟机模板和克隆介绍(3936)
- 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
- CentOS7安装配置远程桌面详解(3631)
- Debian安装源配置详解(3457)
- Apollo配置中心使用详解(3356)
- Proxmox VE 防火墙介绍(3323)
- Gitea 简单使用说明(3313)
- Linux系统路由配置说明(3270)
- Chrony时间同步服务器安装配置说明(3256)
- Centos7防火墙配置详解(3231)
- 【Rancher】简单介绍(3230)
- Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
- PVE 超融合架构介绍(3096)
- 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

- Proxmox VE 存储介绍(1)
- Proxmox Virtual Environment 介绍(1)
- Debian Jessie Preseed自动化安装示例(1)
- Debian live-build官方教程学习笔记(1)
- Ceph存储池pg_num配置详解(1)

推荐排行榜

- leaf-spine （叶脊）网络架构介绍(1)
- Proxmox VE 防火墙介绍(1)
- Proxmox Virtual Environment 介绍(1)
- VIM 常用快捷键(1)
- Debian7 bonding配置详解(1)

最新评论

1. Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从“Default (No cache)”修改为“Write back (unsafe)”，对磁盘数据有影响么？

--大熊子

2. Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的存储是EMC FC-SAN存储（不支持ISCSI），如果改用NFS做集群，FC-SAN应该以什么方式挂载给Proxmox好呢？


```
--abraham_liuhai
```



- OVS Port是物理接口“eno1”在虚拟交换机上的接口属性，本质上是一个物理接口，可以使用网线将“eno1”与物理交换机S5110V2-28P-SI的G1/0/17接口连接起来，所以“eno1”的VLAN属性与G1/0/17的VLAN属性要一致，才能进行通信。

图32. “eno1”的VLAN属性

编辑: OVS Port

名称: eno1

IPv4/CIDR:

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

OVS Bridge: vmbr2

VLAN标签: no VLAN

OVS选项: trunks=250,251

备注: service

高级

OK @51CTO博客

本项目中，PVE虚拟网络只有二层交换，不涉及到IP路由，无需创建虚拟路由器，IP路由算法由物理网络设备负责。

5.3.3 存储网络配置

根据表2和表3，存储网络的VLAN ID是VLAN 252，各个节点的存储采用VLAN 252进行通信。由于节点1服务器上物理网卡不多，节点1的存储通信采用IntPort虚拟接口。

第一步：创建名称为“br1ceph”的OVS IntPort（虚拟接口），并将“br1ceph”虚拟接口捆绑到虚拟交换机vmbr1上，如图33所示。

图33. 创建名称为br1ceph的虚拟接口

创建: OVS IntPort

名称: br1ceph

IPv4/CIDR: 10.129.252.31/24

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

OVS Bridge: vmbr1

VLAN标签: 252

OVS选项:

备注: with ceph ip

高级

OK @51CTO博客

“br1ceph”虚拟接口捆绑到虚拟交换机vmbr1上，使用物理接口“eno2”进行通信，因此物理接口“eno2”要允许VLAN 252透传。

5.3.4 集群网络配置

根据表2和表3，集群网络的VLAN ID是VLAN 253，各个节点的集群采用VLAN 253进行通信。由于节点1服务器上物理网卡不多，节点1的集群通信采用IntPort虚拟接口。

- 第一步：创建名称为“br1ceph_cluster”的OVS IntPort（虚拟接口），并将“br1ceph_cluster”虚拟接口捆绑到虚拟交换机vmbr1上，如图34所示。

图34. 创建名称为br1ceph_cluster的虚拟接口

创建: OVS IntPort

名称: br1ceph_cluster

IPv4/CIDR: 10.129.253.31/24

网关 (IPv4):

IPv6/CIDR:

网关 (IPv6):

OVS Bridge: vmbr1

VLAN标签: 253

OVS选项:

备注: with ceph_cluster ip

高级

OK @51CTO博客

“br1ceph_cluster”虚拟接口捆绑到虚拟交换机vmbr1上，使用物理接口“eno2”进行通信，因此物理接口“eno2”要允许VLAN253透传。

5.3.5 公共网络配置

根据表2和表3，公共网络的VLAN ID是VLAN 254，各个节点的公共网络采用VLAN 254进行通信。由于节点1服务器上物理网卡不多，节点1的公共网络通信采用IntPort虚拟接口。

- 第一步：创建名称为“br1ceph_public”的OVS IntPort（虚拟接口），并将“br1ceph_public”虚拟接口捆绑到虚拟交换机vmbr1上，如图35所示。

图35. 创建名称为br1ceph_public的虚拟接口

创建 OVS IntPort

名称:br1ceph_public

OVS Bridge:vmbr1

IPv4/CIDR:10.129.254.31/24

VLAN标签:254

网关 (IPv4):

OVS选项:

IPv6/CIDR:

备注:with ceph_public ip

网关 (IPv6):

高级CTO博客

“br1ceph_public”虚拟接口捆绑到虚拟交换机vmbr1上，使用物理接口“eno2”进行通信，因此物理接口“eno2”要允许VLAN 254透传。

5.4 节点2的PVE虚拟网络配置

节点2的PVE虚拟网络配置与节点1的类似，这里就不演示节点2了，只给出虚拟节点2的PVE虚拟网络架构图和节点2的虚拟网络配置图，如图36和图37所示。

图36. 节点2的PVE虚拟网络架构图

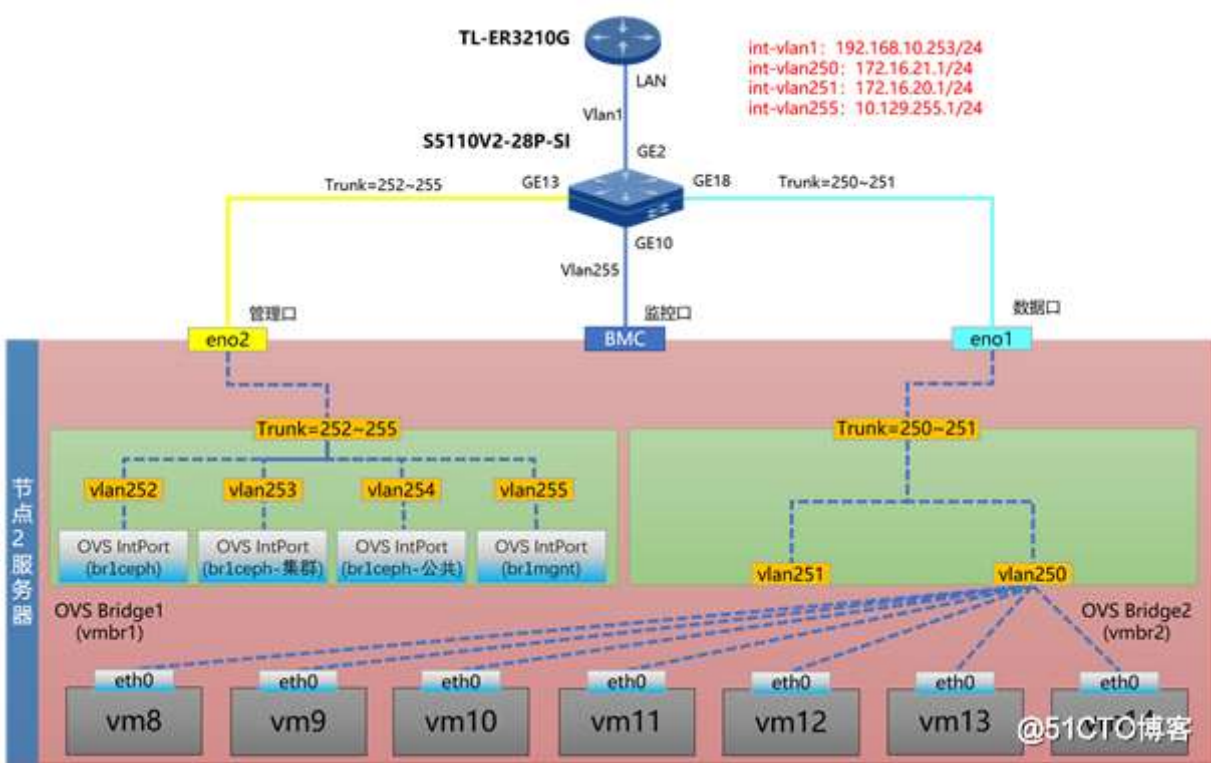


图37. 节点2的虚拟网络配置图

节点 lve-node2

创建 还原 编辑 删除 应用配置

名称	类别	活动	自动	VLA	端口/从属	Bond	CIDR	网关	备注
br1ceph	OVS IntPort	是	否	否			10.129.252.32/24		with ceph ip
br1ceph_cluster	OVS IntPort	是	否	否			10.129.253.32/24		with ceph_
br1ceph_public	OVS IntPort	是	否	否			10.129.254.32/24		with ceph_
br1mgt	OVS IntPort	是	否	否			10.129.255.32/24	10.129.255.1	with mgt ip
eno1	OVS Port	是	否	否					service
eno2	OVS Port	是	否	否					mgmt and c...
ens4f0	网络设备	否	否	否					
ens4f1	网络设备	否	否	否					
vmbr1	OVS Bridge	是	是	否	eno2 br1c...				mgmt and c...
vmbr2	OVS Bridge	是	是	否	eno1				service, vl...

高级CTO博客

5.5 节点3的PVE虚拟网络配置

节点3的PVE虚拟网络配置与节点1的类似，这里就不演示节点3了，只给出虚拟节点3的PVE虚拟网络架构图和节点3的虚拟网络配置图，如图38和图39所示。

图38. 节点3的PVE虚拟网络架构图

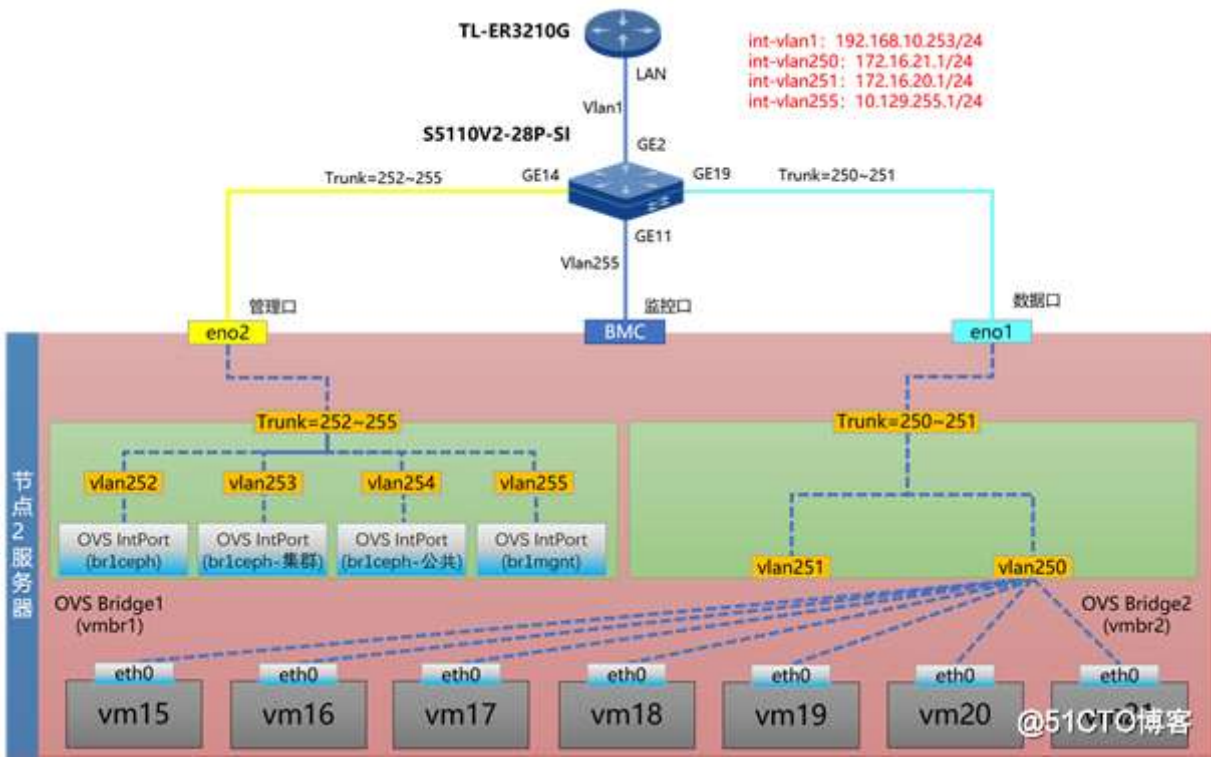


图39. 节点3的虚拟网络配置图

名称	类型	活动	自动	VLA	接口/从属	Bond	CIDR	网关	备注
br1ceph	OVS IntPort	是	否	否			10.129.252.33/24		with ceph ip
br1ceph_cluster	OVS IntPort	是	否	否			10.129.253.33/24		with ceph _
br1ceph_public	OVS IntPort	是	否	否			10.129.254.33/24		with ceph _
br1mngt	OVS IntPort	是	否	否			10.129.255.33/24	10.129.255.1	with mngt ip
eno1	OVS Port	是	否	否					service
eno2	OVS Port	是	否	否					mngt and c...
ens4f0	网络设备	否	否	否					
ens4f1	网络设备	否	否	否					
vibr1	OVS Bridge	是	是	否	eno2 br1c...				mngt and c...
vibr2	OVS Bridge	是	是	否	eno1				service, vl...

5.6 各节点虚拟主机网络配置

根据图26、图36以及图38所示，各节点的虚拟主机的网络配置就很简单了，这里只演示节点1的VM1虚拟机网络配置，其他节点不再演示。

第一步：在节点1的PVE上将VM1的虚拟网卡捆绑到虚拟交换机vibr2上，vlan标签为251

选择“VM1→硬件→网络设备（net0）→编辑”，如图40所示。

图40. 虚拟机VM1网卡属性配置

第二步：根据表4的IP地址规划，在VM1上配置相对于的IP地址和网关，如图41所示。

图41. 虚拟机IP地址配置

作者：[Varden](#)

出处：<http://www.cnblogs.com/varden/>

本文内容如有雷同，请联系作者！

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类：[服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

- « 上一篇: [Proxmox VE + Ceph 超融合项目实战 \(第四部分: 物理网络\)](#)
- » 下一篇: [Proxmox VE + Ceph 超融合项目实战 \(第六部分: 集群组建\)](#)

posted @ 2021-08-31 08:36 [Varden](#) 阅读(4299) 评论(0) [编辑](#) [收藏](#) [举报](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

编辑推荐:

- [如何避免让线程摸鱼, 请用异步技术 async await 拿捏他](#)
- [分布式事务 | 使用 DTM 的 Saga 模式](#)
- [SQLSERVER 阻塞之 PFS 页到底是什么?](#)
- [带团队后的日常思考 \(十一\)](#)
- [记一次使用 gdb 诊断 gc 问题全过程](#)

阅读排行:

- [chatGPT vscode 体验 \(密钥已更新当天有效\)](#)
- [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
- [如何避免让线程摸鱼, 请用异步技术 async await 拿捏他~](#)
- [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
- [不用Blazor WebAssembly, 开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

[首页](#)
[新随笔](#)
[联系](#)
[管理](#)
[订阅](#)
[RML](#)

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: [Varden](#)
 园龄: [2年3个月](#)
 粉丝: [42](#)
 关注: [2](#)
[-取消关注](#)

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

最新随笔

- [1.Ceph 故障域配置 \(修改CRUSH MAP\)](#)
- [2.RGW多站点同步策略配置](#)
- [3.RGW故障转移和灾难恢复](#)
- [4.RGW配置主区域](#)
- [5.Ceph运维记录2022](#)
- [6.Ceph 操作集群](#)
- [7.常见磁盘总线的传输速率](#)
- [8.cephadm 删除主机](#)
- [9.使用s3cmd工具实现RGW对象存储生命周期管理](#)
- [10.RGW池放置和存储类 \(Octopus版本\)](#)

随笔分类

Alpine(2)
 Ansible(23)
 CentOS(23)
 Ceph(118)
 Debian(19)
 Docker(28)
 ELK日志分析(15)
 Helm(3)
 Java(1)
 Jenkins(10)
 K8s(209)
 Kubectl(4)
 Linux集群技术(3)
 Linux企业常用服务(9)
 Linux企业级安全(5)
 Linux系统调优和监控(8)
 Linux系统基础及管理(38)
 MongoDB(12)
 MySQL/MariaDB(6)
 NFS(3)
 PostgreSQL(1)

Proxmox VE + Ceph 超融合项目实战（第六部分：集群组建）

参考: https://blog.51cto.com/u_14742102/3499271

六、集群组建

6.1 PVE集群部署要求

- PVE集群部署有一定的要求，以下是《Proxmox6.2手册》对PVE集群部署的要求，摘录如下：
 - 所有节点必须可以相互访问彼此的UDP 5404和UDP 5405端口，以确保corosync正常工作。
 - 各节点日期和时间需要保持同步。
 - 各节点之间要能够在TCP 22端口建立SSH通信。
 - 如果你需要配置HA，则最少需要3个物理服务器节点，以保证集群多数票机制生效。此外，还需要保证所有节点使用同一版本的Proxmox VE。
 - Proxmox VE集群网络只有在网络延时低于2ms 时（局域网内）才可以正常工作，建议为集群通信分配专用网卡，特别是在配置共享存储的情况下，分配专用网卡能确保集群通信的稳定可靠。
- 注意事项：
 - PVE节点在没有创建虚拟机的情况下，可以加入其他节点创建的集群，也可以自己创建集群，让其他节点加入进来。
 - PVE节点如果创建了虚拟机，那么这台机器就只能通过创建集群 让其他的节点加入进来。
 - 如果把一个节点从集群中删除掉之后，再以相同的主机名和IP地址进行加入。可能会产生错误，需要强制加入。
 - 加入集群后将不允许再修改主机名和IP 地址。
 - 一旦建立集群后，将不允许修改集群名称。

6.2 创建集群

创建集群操作可以在命令行控制台（ssh登录）下进行，也可以通过API调用完成，GUI 界面就是通过调用API来创建集群的（数据中心→集群）。

- 第一步：在节点1上创建名称为“dc-zz”的集群

在节点1中，选择“数据中心→集群→创建集群”，集群名称输入“dc-zz”，集群网络选择“br1ceph接口”的IP地址，如图42所示。

图42. 创建集群



如果要添加第二条链路作为备用链路，选择“添加”之后，添加第二条链路，再点击“创建”，即可创建。

注意1：根据规划，集群网络接口应该是“br1ceph_cluster接口”，但是在交付的时候，由于疏忽，集群网络都使用了“br1ceph接口”，虽然配置没有错，但是名称容易误导。所以在这里做一个说明，后续的配置中，集群网络使用了“br1ceph接口”，也就是“br1ceph接口”代表了集群网络。

注意2：确保为集群通信选择的网络未用于任何高流量负载，如（网络）存储或实时迁移。虽然集群网络本身只产生少量数据，但它对延迟非常敏感。

- 第二步：复制集群密钥，其他节点需要使用集群密钥才能加入集群

在节点1中，点击“加入信息”按钮，弹出界面如图43所示，接着点击“拷贝信息”来复制集群密钥。

图43. 集群密钥



- 第三步：在节点2上，使用节点1的集群密钥，加入集群

在节点2中，点击“加入集群”按钮，粘贴“加入信息”，输入节点1的root密码，选择节点2的集群链路，最后点击“加入dc-zz”集群，如图44所示。

图44. 节点2加入集群



- 第四步：在节点3上，使用节点1的集群密钥，加入集群

在节点3中，点击“加入集群”按钮，粘贴“加入信息”，输入节点1的root密码，选择节点3的集群链路，最后点击“加入dc-zz”集群，如图45所示。

图45. 节点3加入集群



- 第五步：加入集群完成后需要刷新页面才能正常使用，全部加入完成后，就可以在一个节点上管理集群中所有的节点了，如图46所示。

单击“加入dc-zz”按钮之后后，集群加入过程将立即开始。节点加入集群后，其当前节点证书将被集群证书颁发机构(CA)签名的证书替换，这意味着当前会话将在几秒钟后停止工作。然后，可能需要强制重新加载Web 界面，并使用集群凭据重新登录。通俗点说，就是需要刷新下页面才能使用。

图46. dc-zz集群正常使用

Q 搜索

群表

集群

Ceph

选项

存储

备份

复制

集群信息

创建集群

加入信息

加入集群

集群名称dc-zz

配置版本3

节点数3

集群节点

节点名称

ID ↑

状态

链接ID

bve-node11110.129.252.31

bve-node22110.129.252.32

bve-node33110.129.252.33

@51CTO博客

作者：[Varden](#)

出处：<http://www.cnblogs.com/varden/>

本文内容如有雷同，请联系作者！

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类：[服务器虚拟化](#)

好文要顶

已关注

收藏该文

[Prometheus\(52\)](#)
[Rancher\(44\)](#)
[Redis\(1\)](#)
[ServiceMesh\(1\)](#)
[Ubuntu\(18\)](#)
[Windows\(2\)](#)
[Zabbix\(1\)](#)
[常用工具\(24\)](#)
[代码管理\(5\)](#)
[更多](#)

随笔档案

[2022年10月\(4\)](#)
[2022年6月\(3\)](#)
[2022年5月\(17\)](#)
[2022年4月\(25\)](#)
[2022年3月\(68\)](#)
[2022年2月\(18\)](#)
[2021年12月\(1\)](#)
[2021年11月\(1\)](#)
[2021年10月\(2\)](#)
[2021年9月\(50\)](#)
[2021年8月\(273\)](#)
[2021年7月\(79\)](#)
[2020年11月\(58\)](#)
[2020年10月\(58\)](#)

阅读排行榜

1. [CentOS7安装源配置详解\(9876\)](#)
2. [leaf-spine （叶脊）网络架构介绍\(7105\)](#)
3. [PVE 备份和恢复介绍\(5386\)](#)
4. [Proxmox Virtual Environment 介绍\(4947\)](#)
5. [Proxmox VE Qemu/KVM 虚拟机设置要求\(4533\)](#)
6. [Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）\(4297\)](#)
7. [PVE 虚拟机模板和克隆介绍\(3936\)](#)
8. [【K8s教程】Nginx Ingress控制器Ingress Path匹配\(3843\)](#)
9. [CentOS7安装配置远程桌面详解\(3631\)](#)
10. [Debian安装源配置详解\(3457\)](#)
11. [Apollo配置中心使用详解\(3356\)](#)
12. [Proxmox VE 防火墙介绍\(3323\)](#)
13. [Gitea 简单使用说明\(3313\)](#)
14. [Linux系统路由配置说明\(3270\)](#)
15. [Chrony时间同步服务器安装配置说明\(3256\)](#)
16. [Centos7防火墙配置详解\(3231\)](#)
17. [【Rancher】简单介绍\(3230\)](#)
18. [Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）\(3098\)](#)
19. [PVE 超融合架构介绍\(3096\)](#)
20. [【K8s任务】为容器设置启动时要执行的命令和参数\(3071\)](#)

评论排行榜

1. [Proxmox VE 存储介绍\(1\)](#)
2. [Proxmox Virtual Environment 介绍\(1\)](#)
3. [Debian Jessie Preseed自动化安装示例\(1\)](#)
4. [Debian live-build官方教程学习笔记\(1\)](#)
5. [Ceph存储池pg_num配置详解\(1\)](#)

推荐排行榜

1. [leaf-spine （叶脊）网络架构介绍\(1\)](#)
2. [Proxmox VE 防火墙介绍\(1\)](#)
3. [Proxmox Virtual Environment 介绍\(1\)](#)
4. [VIM 常用快捷键\(1\)](#)
5. [Debian7 bonding配置详解\(1\)](#)

最新评论

1. [Re:KVM虚拟化Disk Cache模式对性能的影响分析](#)
如果从“Default (No cache)”修改为“Write back (unsafe)”，对磁盘数据有影响么？

--大熊子

2. [Re:Proxmox VE 存储介绍](#)
请教一下：原来用在VMWARE集群的以EMC FC-SAN存储（不支持ISCSI），如果改用X做集群，FC-SAN应该以什么方式挂载给PRC人好呢？



Varden

粉丝 - 42 关注 - 2

我在关注他 [取消关注](#)

- « 上一篇：[Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）](#)
- » 下一篇：[Proxmox VE + Ceph 超融合项目实战（第七部分：Ceph分布式存储）](#)

posted @ 2021-08-31 11:07 Varden 阅读(2391) 评论(0) 编辑 收藏 举报

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

编辑推荐:

- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
- [分布式事务 | 使用 DTM 的 Saga 模式](#)
- [SQLSERVER 阻塞之 PFS 页到底是什么？](#)
- [带团队后的日常思考（十一）](#)
- [记一次使用 gdb 诊断 gc 问题全过程](#)

阅读排行:

- [chatGPT vscode 体验（密钥已更新当天有效）](#)
- [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
- [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
- [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

- 9394952
3. Re:Proxmox Virtual Environment 介绍很全面！

--willblog
4. Re:Ceph存储池pg_num配置详解mark

--穷死了
5. Re:Debian live-build官方教程学习笔记执行成功之后为什么没找看iso

--abraham_liuhai

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
[-取消关注](#)

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

最新随笔

- 1.Ceph 故障域配置 (修改CRUSH MAP)
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类 (Octopus版本)

随笔分类

Alpine(2)
Ansible(23)
CentOS(23)
Ceph(118)
Debian(19)
Docker(28)
ELK日志分析(15)
Helm(3)
Java(1)
Jenkins(10)
K8s(209)
KubectI(4)
Linux集群技术(3)
Linux企业常用服务(9)
Linux企业级安全(5)
Linux系统调优和监控(8)
Linux系统基础及管理(38)
MongoDB(12)
MySQL/MariaDB(6)
NFS(3)
PostgreSQL(1)

Proxmox VE + Ceph 超融合项目实战（第七部分：Ceph分布式存储）

参考: https://blog.51cto.com/u_14742102/3499271

七、Ceph分布式存储

Ceph以其先进的去中心化设计，成为了分布式块存储领域的佼佼者，它可以将多台服务器中的硬盘聚合成一个集群，对外可提供文件存储、块存储、对象存储等多种形式，Ceph的另一个特点是数据采用多副本方式存储，避免由于单点故障造成的业务中断和数据丢失。

7.1 Ceph的注意事项

Proxmox VE提供了简单易用的Ceph 安装向导，选中集群中的一个节点，然后在菜单树中打开Ceph菜单区，就可以开始安装Ceph 了。

安装向导有多个步骤，每个步骤都需要执行成功才可以完成Ceph 安装。开始安装操作后，向导会自动从 Proxmox VE 的Ceph 软件源下载软件包并完成安装。

完成第一个步骤后，还需要创建配置。对每个集群，生成的配置信息会自动分发到其他节点，所以该操作只需要执行一次即可。

创建的配置包括以下信息：

- 1) Public Network

为避免影响集群通信等其他对网络延迟敏感的服务，也为了提高Ceph 性能，强烈建议为Ceph 准备一个专门的独立网络，将Ceph流量隔离开来。

- 2) Cluster Network

进一步，还可以设置Cluster Network，将OSD复制和心跳流量隔离出来。这将有效降低public network的负载，并有效改善大规模Ceph集群的性能。

7.2 初始化Ceph安装和配置

每个节点都需要安装Ceph，每个节点的Ceph安装基本相同，这里只演示节点1的Ceph安装。

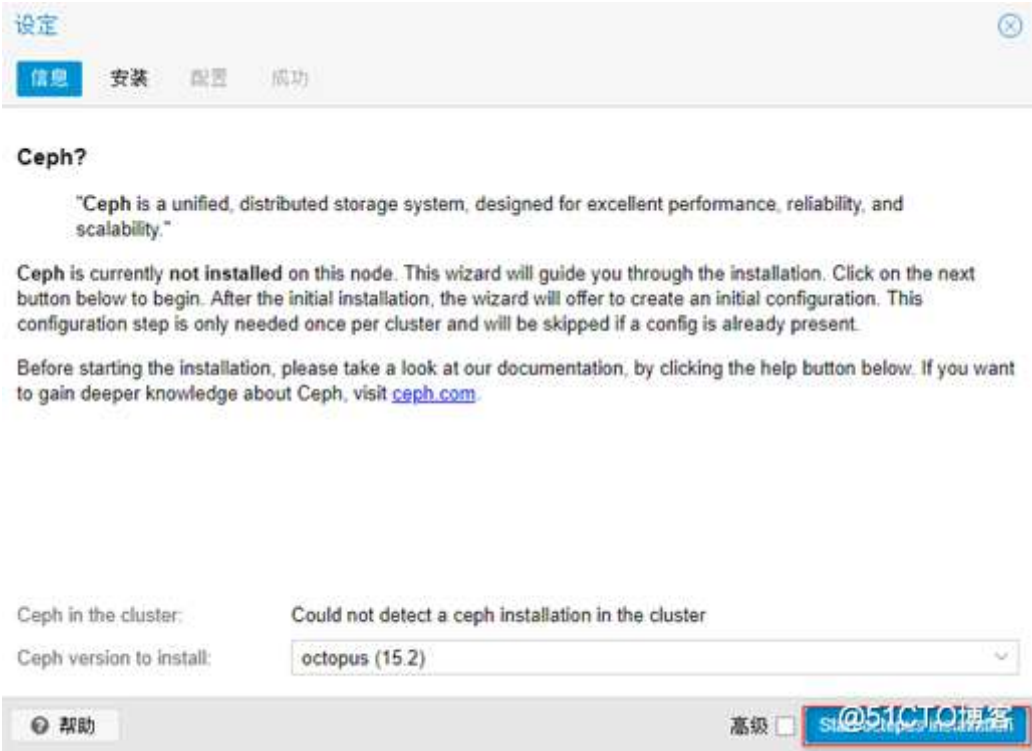
- 第一步：在节点1界面中，选择“Ceph→Install Ceph”，点击“Install Ceph”按钮，如图47所示。

图47. Ceph安装向导



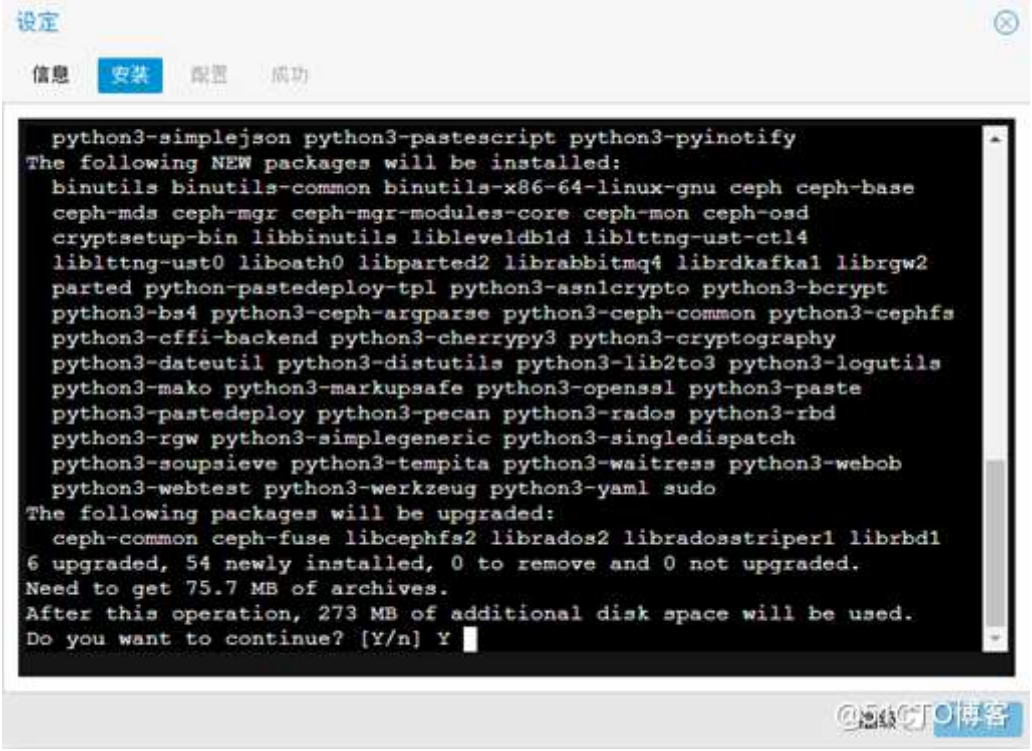
- 第二步：选择“设定→信息→Start octopus installation”，点击“Start octopus installation”按钮，如图48所示。

图48. 选择15.2版本进行安装



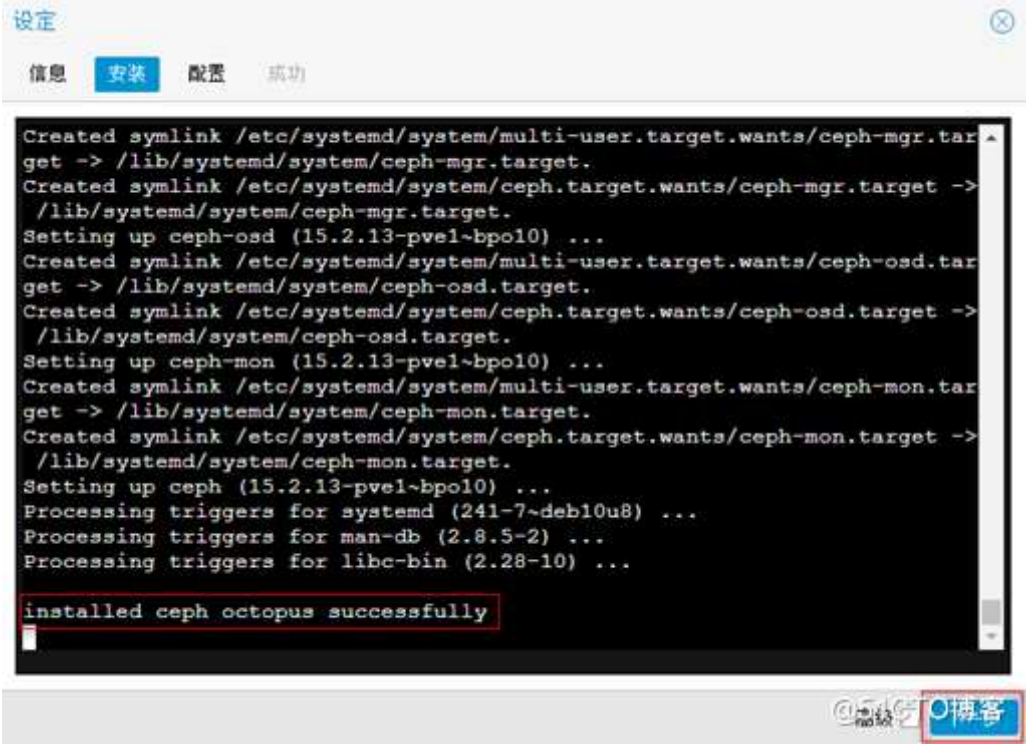
- 第三步：选择“设定→安装”，输入“Y”之后，按回车键，自动安装Ceph，如图49所示。

图49. 自动安装Ceph



- 第四步：安装完成之后，继续点击“下一步”按钮，安装成功之后，会提示安装成功的信息“installed ceph octopus successfully”，如图50所示。

图50. Ceph安装成功



- 第五步：完成Ceph集群配置信息，包括公共网络IP、集群网络IP等信息，如图51所示。

Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

1. CentOS7安装源配置详解(9876)
2. leaf-spine （叶脊）网络架构介绍(7105)
3. PVE 备份和恢复介绍(5386)
4. Proxmox Virtual Environment 介绍(4947)
5. Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
6. Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
7. PVE 虚拟机模板和克隆介绍(3936)
8. 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
9. CentOS7安装配置远程桌面详解(3631)
10. Debian安装源配置详解(3457)
11. Apollo配置中心使用详解(3356)
12. Proxmox VE 防火墙介绍(3323)
13. Gitea 简单使用说明(3313)
14. Linux系统路由配置说明(3270)
15. Chrony时间同步服务器安装配置说明(3256)
16. Centos7防火墙配置详解(3231)
17. 【Rancher】简单介绍(3230)
18. Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
19. PVE 超融合架构介绍(3096)
20. 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

1. Proxmox VE 存储介绍(1)
2. Proxmox Virtual Environment 介绍(1)
3. Debian Jessie Preseed自动化安装示例(1)
4. Debian live-build官方教程学习笔记(1)
5. Ceph存储池pg_num配置详解(1)

推荐排行榜

1. leaf-spine （叶脊）网络架构介绍(1)
2. Proxmox VE 防火墙介绍(1)
3. Proxmox Virtual Environment 介绍(1)
4. VIM 常用快捷键(1)
5. Debian7 bonding配置详解(1)

最新评论

1. Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从“Default (No cache)”修改为“Write back (unsafe)”，对磁盘数据有影响么？

--大熊猫子

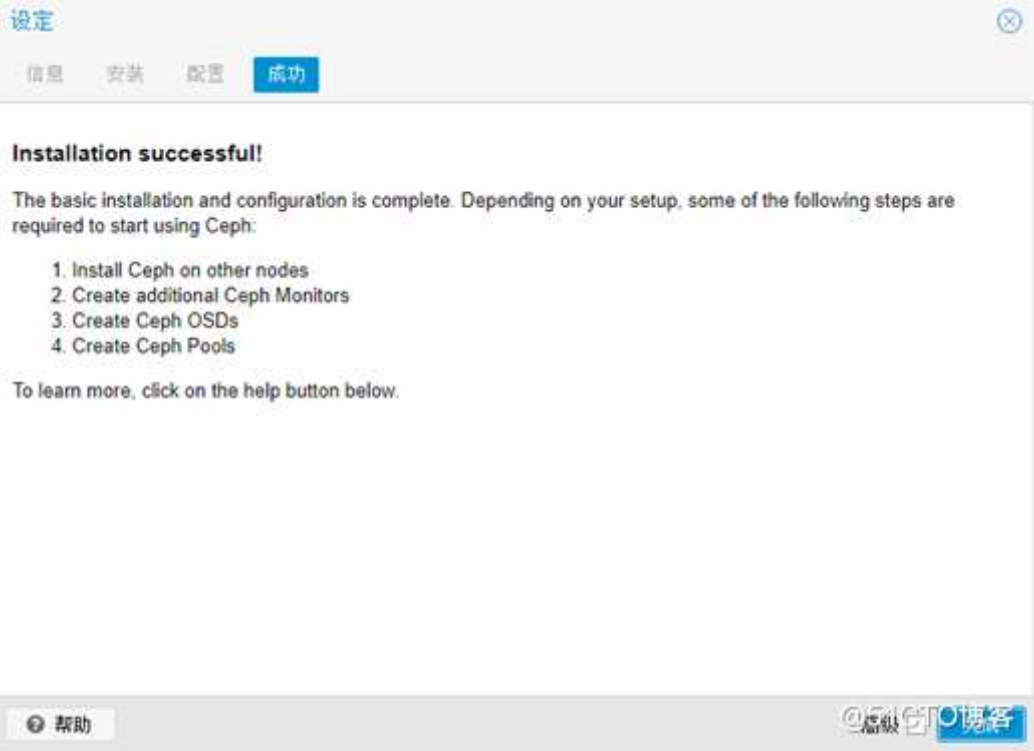
2. Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的比較老的EMC FC-SAN存储（不支持ISCSI），如果改用PROXMOX做集群，FC-SAN应该以什么方式挂载给PROXMOX好呢？

图51. Ceph集群配置



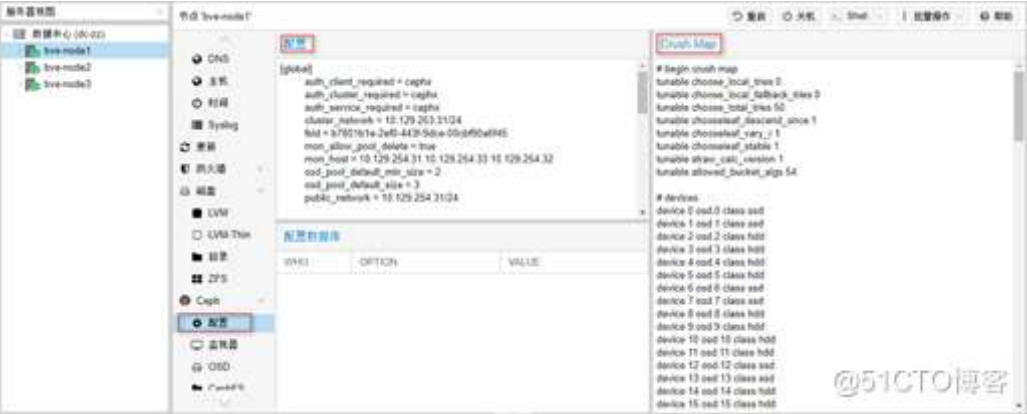
- 第六步：完成配置信息，Ceph安装成功，如图52所示。

图52. Ceph安装成功



- 第七步：查看配置信息，如图53所示。

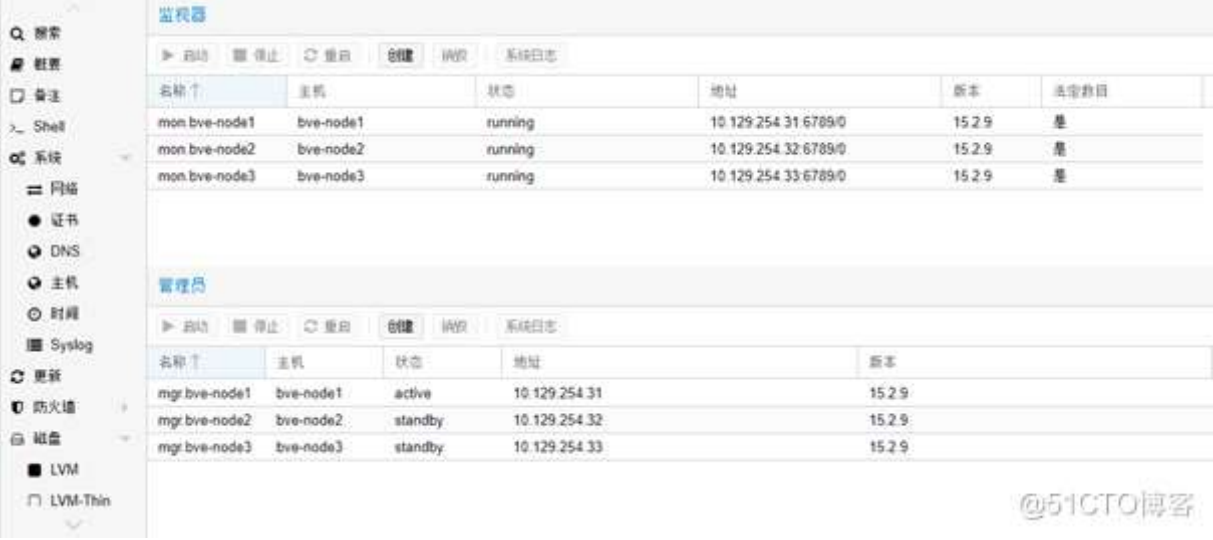
图53. Ceph的配置信息



7.3 配置Ceph监视器

Ceph安装完成之后，节点1出现“监视器”和“管理员”中，还需要把节点2和节点3添加进来。在本项目中，监视器节点和管理员节点信息如图54所示。

图54. Ceph的配置信息



--9394952

3. Re:Proxmox Virtual Environment 介绍
很全面！

--willblog

4. Re:Ceph存储池pg_num配置详解
mark

--穷死了

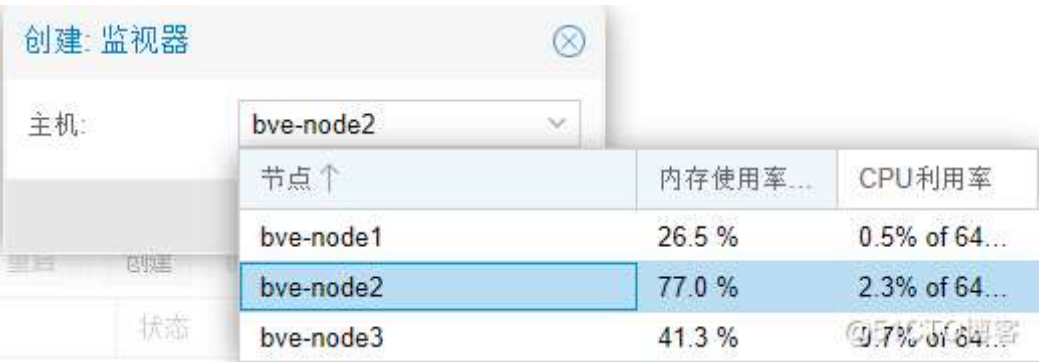
5. Re:Debian live-build官方教程学习笔记
执行成功之后为什么没找看iso

--abraham_liuhai

从图54中可以发现，监视器和管理网络使用的都是公共网络，没有使用集群网络。

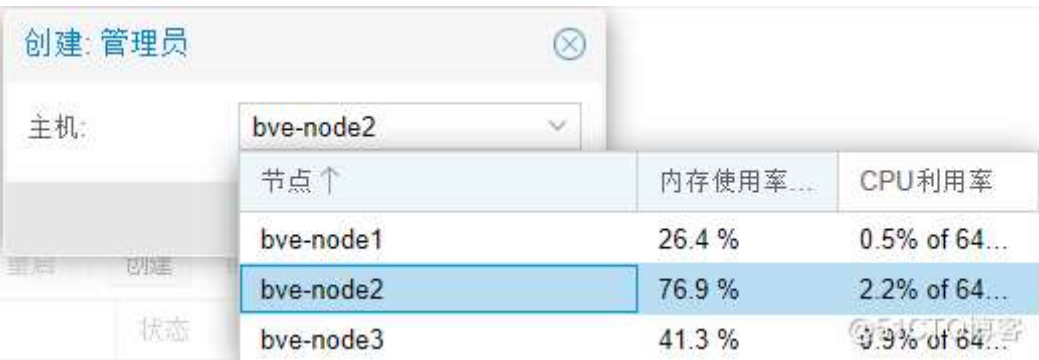
- 第一步：在节点1的监视器中，创建另外两个节点的监视器，点击“创建”按钮，如图55所示。

图55. 创建节点2监视器



- 第二步：在节点1的管理员，创建另外两个节点的管理员，点击“创建”按钮，如图56所示。

图56. 创建节点2管理员



- 第三步：在Ceph监控界面查看节点的监视器和管理HA，选择“bve-node1→Ceph”，如图57所示。

Ceph的监控UI界面中，监视器和管理HA如图57中的红框所示：

图57. Ceph的监视器和管理HA界面



7.4 配置Ceph的OSD

由图3可知，各节点服务器支持八个盘位，其中sdc和sdd组成RAID1，作为系统盘使用，所以Ceph的OSD需要选择未做RAID的硬盘，每个节点的OSD只有6个。

本项目中，三个节点的Ceph的OSD磁盘配置信息如图58、图59所示。

图58. OSD配置信息-1

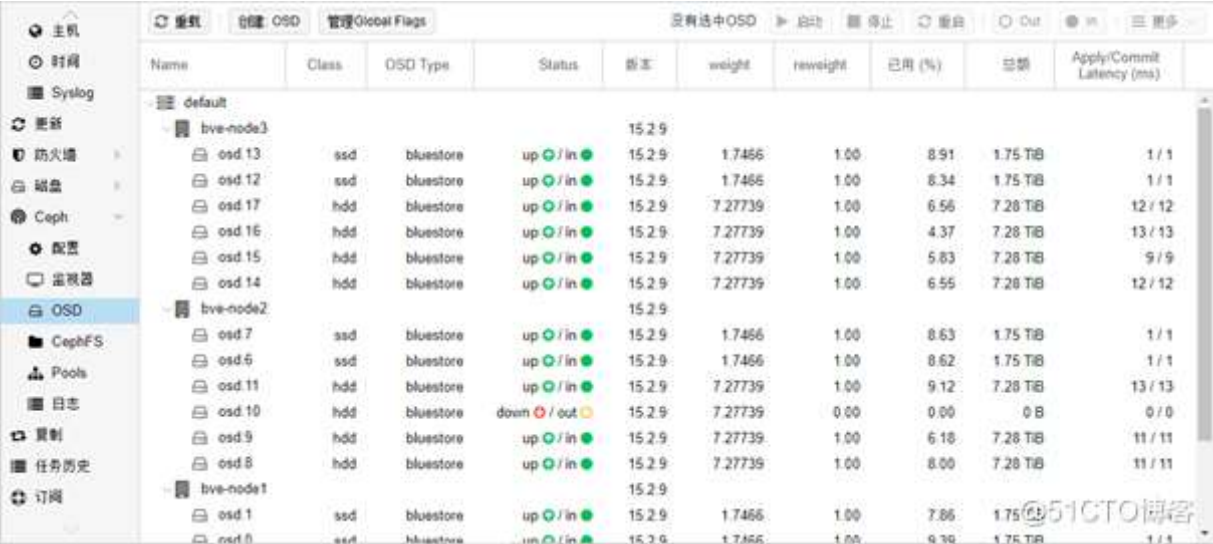


图59. OSD配置信息-2

主机

时间

Syslog

更新

防火墙

磁盘

Ceph

配置

虚拟机

OSD

CephFS

Pools

日志

复制

任务历史

订例

重载

创建 OSD

管理 Global Flags

没有选中 OSD

启动

停止

重启

Out

更多

Name	Class	OSD Type	Status	版本	weight	reweight	已用 (%)	总量	Apply/Commit Latency (ms)
default									
bve-node3				15.2.9					
bve-node2				15.2.9					
osd 7	ssd	bluestore	up / in	15.2.9	1.7466	1.00	8.63	1.75 TiB	1 / 1
osd 6	ssd	bluestore	up / in	15.2.9	1.7466	1.00	8.62	1.75 TiB	1 / 1
osd 11	hdd	bluestore	up / in	15.2.9	7.27739	1.00	9.12	7.28 TiB	13 / 13
osd 10	hdd	bluestore	down / out	15.2.9	7.27739	0.00	0.00	0 B	0 / 0
osd 9	hdd	bluestore	up / in	15.2.9	7.27739	1.00	6.18	7.28 TiB	11 / 11
osd 8	hdd	bluestore	up / in	15.2.9	7.27739	1.00	8.00	7.28 TiB	11 / 11
bve-node1									
osd 1	ssd	bluestore	up / in	15.2.9	1.7466	1.00	7.86	1.75 TiB	1 / 1
osd 0	ssd	bluestore	up / in	15.2.9	1.7466	1.00	9.39	1.75 TiB	1 / 1
osd 5	hdd	bluestore	up / in	15.2.9	7.27739	1.00	6.55	7.28 TiB	14 / 14
osd 4	hdd	bluestore	up / in	15.2.9	7.27739	1.00	6.54	7.28 TiB	15 / 15
osd 3	hdd	bluestore	up / in	15.2.9	7.27739	1.00	6.93	7.28 TiB	14 / 14
osd 2	hdd	bluestore	up / in	15.2.9	7.27739	1.00	3.28	7.28 TiB	9 / 9

- 第一步：创建节点1的OSD

从节点1中登录集群，在节点1中，选择“Ceph→OSD→创建:OSD”，在磁盘中，逐次添加sda、sdb、sde、sdf、sdg、sdh这6块没有被使用的磁盘，数据库磁盘和WAL磁盘保持默认即可，如图60所示。

图60. 创建OSD

创建: Ceph OSD

磁盘: /dev/sda

数据库磁盘: use OSD disk

数据库大小 (GiB): 自动

加密OSD: ☐

WAL磁盘: use OSD/DB disk

Device Class: auto detect

WAL大小 (GiB): 自动

Note: Ceph is not compatible with disks backed by a hardware RAID controller. For details see [the reference documentation](#).

帮助

高级

CTO博客

注意，在本项目中，数据库磁盘（DB Disk）和WAL磁盘（WAL Disk）我都选择了默认方式，其实做一下硬盘规划会更好。

创建其他节点的OSD与节点1类似，这里不做详述。

- 第二步：也可以在配置信息的Crush Map查看集群的OSD配置

选择任意节点登录集群，选择“bve-node1→Ceph→配置→Crush Map”，从Crush Map中可以查看集群的OSD配置信息。

7.5 配置CephFS（可选）

7.5.1 创建CephFS

选择“bve-node1→Ceph→CephFS→创建CephFS”，选择默认值创建CephFS，创建之后如图64所示。

图64. 创建CephFS

节点 bve-node1

创建 CephFS

名称 ?

数据 Pool

Metadata Pool

cephfs

cephfs_data

cephfs_metadata

元数据服务器

名称 ?

主机

状态

IP地址

版本

mds.bve-m...

bve-node1

up standby

10.129.254.33

68013581287248

15.2.9

mds.bve-m...

bve-node2

up standby

10.129.254.32

68013362310565

15.2.9

mds.bve-m...

bve-node3

up active

10.129.254.33

68011072754208

15.2.9

7.5.2 元数据服务器配置

选择“bve-node1→Ceph→CephFS→元数据服务器→创建”，将节点1添加进来，如图65所示。

图65. 添加元数据服务器节点

创建: 元数据服务器

主机: bve-node1

节点 ↑

内存使用率...

CPU利用率

bve-node1

26.6 %

0.5% of 64...

bve-node2

27.4 %

1.6% of 64...

bve-node3

41.8 %

@9% of 64...

@51CTO博客

高级CTO博客

@51CTO博客

@51CTO博客

将节点1添加到元数据服务器之后，再用同样的方法将节点2、节点3添加到元数据服务器之中去。

添加好元数据服务器节点之后，可以在Ceph监控界面看到元数据服务器的监控信息，如图67所示。

图67. 元数据服务器监控界面



7.6 配置Pools资源池

7.6.1 节点1配置Pools资源池

在Pools资源池配置之前，节点1的Pools界面如图68所示。

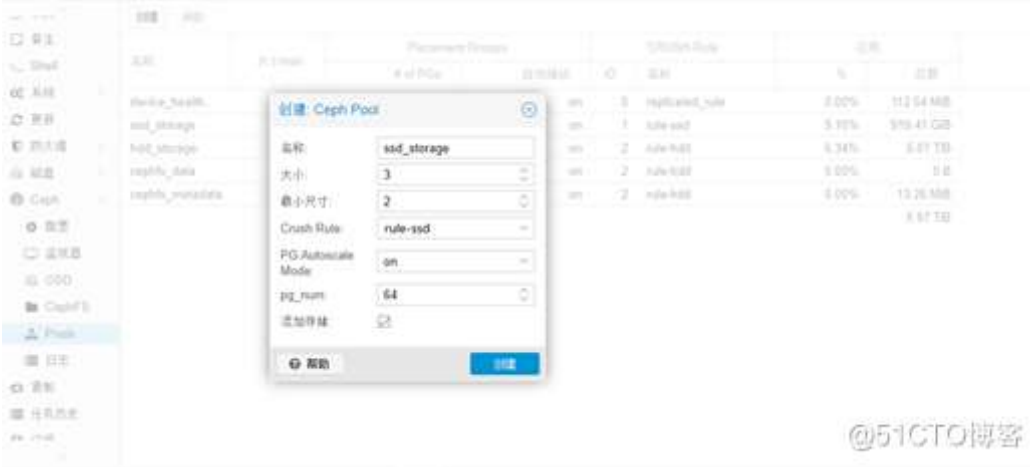
图68. 节点1的Pools界面



- 第一步：创建ssd_storage存储

选择“bve-node1→Ceph→Pools→创建”，点击“创建”按钮，创建名称为“ssd_storage”的存储，创建之后如图69所示。

图69. 创建ssd_storage存储



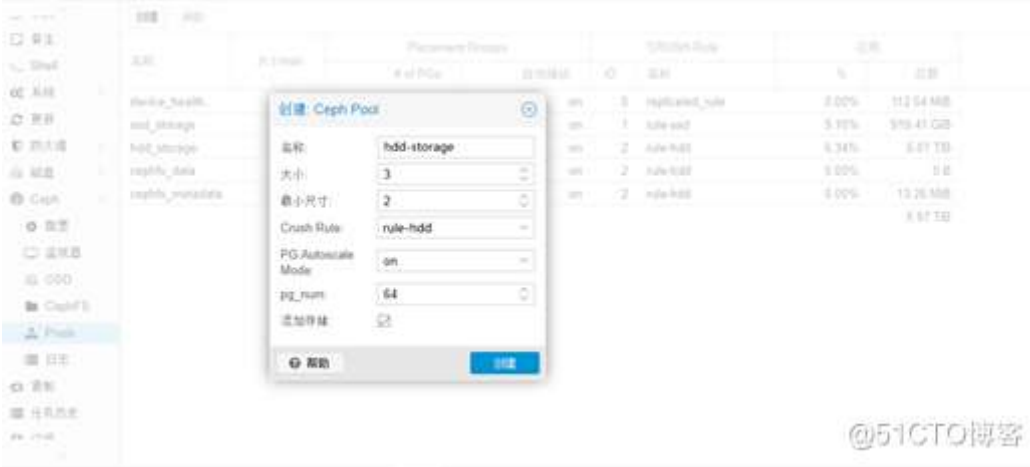
根据官方的文档，pg_num这个数值比较关键，如果不合适的话，Ceph会告警，官方建议：

若少于5个OSD， 设置pg_num为128。
5~10个OSD， 设置pg_num为512。
10~50个OSD， 设置pg_num为4096。

- 第二步：创建hdd_storage存储

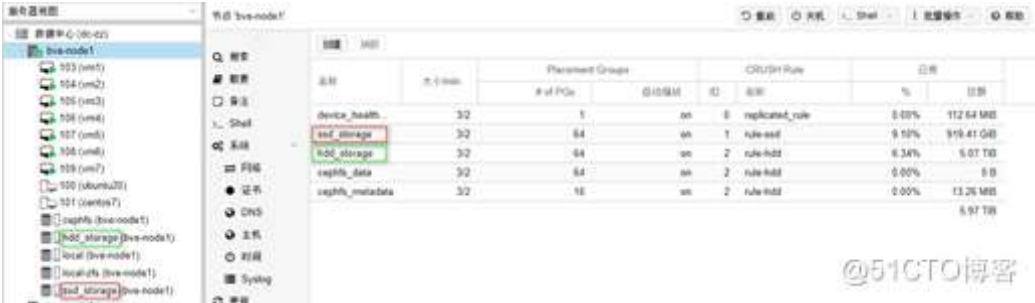
选择“bve-node1→Ceph→Pools→创建”，点击“创建”按钮，创建名称为“hdd_storage”的存储，创建之后如图70所示。

图70. 创建hdd_storage存储



- 第三步：创建完成之后，节点1左侧的存储菜单树中会出现这两个存储名称，如图71所示，正常情况下，就可以在ssd_storage、hdd_storage上安装虚拟机了。

图71. ssd_storage与hdd_storage存储



7.6.2 节点2配置Pools资源池

节点2的Pools资源池配置与节点1的相似，不再赘述。

7.6.3 节点3配置Pools资源池

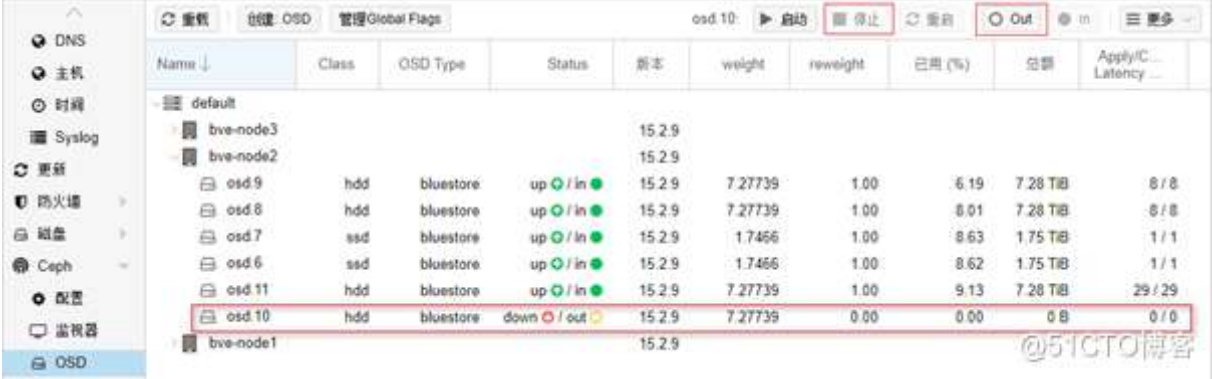
节点3的Pools资源池配置与节点1的相似，不再赘述。

7.7 Ceph磁盘管理

7.7.1 OSD磁盘下线

如果想要把一个Ceph的OSD磁盘更换下来，先选中OSD磁盘，接着按“停止”按钮先停止OSD服务，接着再按“Out”按钮，这个时候OSD.10的状态如图75所示，就可以把该磁盘取出。

图75. OSD磁盘卸载



7.7.2 OSD磁盘上线

OSD磁盘加载与OSD磁盘卸载过程刚好相反，先把磁盘插进服务器的盘位中，接着先按“in”按钮，等“in”图标变为绿色之后，再按“启动”按钮启动OSD服务，过几秒钟之后，OSD便会恢复正常，不再演示截图。

作者：[Varden](#)

出处：<http://www.cnblogs.com/varden/>

本文内容如有雷同，请联系作者！

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类：[服务器虚拟化](#), [Ceph](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

« 上一篇：[Proxmox VE + Ceph 超融合项目实战（第六部分：集群组建）](#)

» 下一篇：[Proxmox VE + Ceph 超融合项目实战（第八部分：OSD磁盘故障排查案例）](#)

posted @ 2021-08-31 13:59 [Varden](#) 阅读(3031) 评论(0) 编辑 收藏 举报

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

编辑推荐：

- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
- [分布式事务 | 使用 DTM 的 Saga 模式](#)
- [SQLSERVER 阻塞之 PFS 页到底是什么？](#)
- [带团队后的日常思考（十一）](#)
- [记一次使用 gdb 诊断 gc 问题全过程](#)

阅读排行：

- [chatGPT vscode 体验（密钥已更新当天有效）](#)
- [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
- [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
- [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 [HTML](#)

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
[-取消关注](#)

Proxmox VE + Ceph 超融合项目实战（第八部分：OSD磁盘故障排查案例）

参考: https://blog.51cto.com/u_14742102/3499271

八、OSD磁盘故障排查案例

- 第一步：Ceph的监控界面发现OSD掉线，但是并没有产生告警

本项目在客户现场稳定运行了半年，基本上没有出现过的故障，这次去客户现场巡查，在Ceph的WEB UI监控界面中，发现了Ceph有一个OSD磁盘掉线了，如图76所示。

图76. OSD磁盘掉线



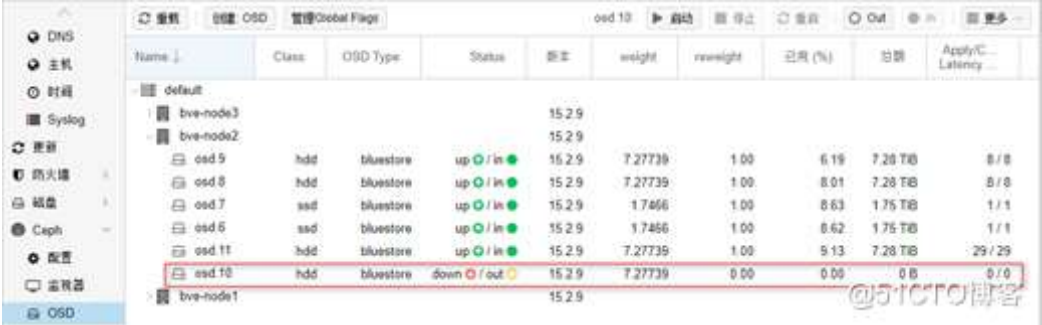
在图76中，OSD总额是18（有18块OSD磁盘），“在线”且是“进入”状态的有17块OSD磁盘，“掉线”且是“输出”状态的有1块OSD磁盘。注意，这里的“进入”和“输出”翻译得不够好，翻译成“加载”和“卸载”还比这个要好一些。

令我不解的是，OSD磁盘掉线也算是一个不大不小的故障了吧，但是在Ceph的WEB UI监控界面的告警栏中，竟然没有任何告警提示！Ceph的自带监控系统缺乏必要的告警功能，不能满足生产级的项目需求，难怪一些大厂在使用Ceph搭建分布式存储时，都使用Zabbix进行监控。

- 第二步：进一步排查，确认掉线的OSD磁盘

进一步直接从Ceph的OSD去查找掉线的OSD磁盘，最后发现在节点2上，Ceph有一个OSD.10的磁盘掉线了，如图77和图78所示。

图77. WEB UI中的Ceph OSD状态



<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

最新随笔

- 1.Ceph 故障域配置（修改CRUSH MAP）
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类（Octopus版本）

随笔分类

Alpine(2)
Ansible(23)
CentOS(23)
Ceph(118)
Debian(19)
Docker(28)
ELK日志分析(15)
Helm(3)
Java(1)
Jenkins(10)
K8s(209)
KubectI(4)
Linux集群技术(3)
Linux企业常用服务(9)
Linux企业级安全(5)
Linux系统调优和监控(8)
Linux系统基础及管理(38)
MongoDB(12)
MySQL/MariaDB(6)
NFS(3)
PostgreSQL(1)

图78. 命令行中的Ceph OSD状态

```
root@ove-node1:~# ceph osd df
ID CLASS WEIGHT REWEIGHT SIZE RAW USE DATA OMAP META AVAIL %USE VAR PGS STATUS
2 hdd 7.27739 1.00000 7.3 TiB 245 GiB 244 GiB 867 KiB 1.1 GiB 7.0 TiB 3.28 0.50 25 up
3 hdd 7.27739 1.00000 7.3 TiB 518 GiB 516 GiB 39 MiB 1.2 GiB 6.8 TiB 6.95 1.05 39 up
4 hdd 7.27739 1.00000 7.3 TiB 488 GiB 487 GiB 3.6 MiB 1.2 GiB 6.8 TiB 6.55 0.99 39 up
5 hdd 7.27739 1.00000 7.3 TiB 489 GiB 488 GiB 3.4 MiB 1.2 GiB 6.8 TiB 6.57 0.99 42 up
0 ssd 1.74660 1.00000 1.7 TiB 168 GiB 167 GiB 7.1 MiB 1017 MiB 1.6 TiB 9.39 1.42 35 up
1 ssd 1.74660 1.00000 1.7 TiB 141 GiB 140 GiB 4.5 MiB 1020 MiB 1.6 TiB 7.86 1.19 29 up
8 hdd 7.27739 1.00000 7.3 TiB 597 GiB 596 GiB 2.7 MiB 1.4 GiB 6.7 TiB 8.02 1.21 41 up
9 hdd 7.27739 1.00000 7.3 TiB 461 GiB 460 GiB 3.1 MiB 1.1 GiB 6.8 TiB 6.19 0.93 46 up
10 hdd 7.27739 0 0 B 0 B 0 B 0 B 0 B 0 0 0 down
11 hdd 7.27739 1.00000 7.3 TiB 681 GiB 679 GiB 39 MiB 1.4 GiB 6.6 TiB 9.13 1.38 58 up
6 ssd 1.74660 1.00000 1.7 TiB 154 GiB 153 GiB 8.3 MiB 1016 MiB 1.6 TiB 8.62 1.30 32 up
7 ssd 1.74660 1.00000 1.7 TiB 154 GiB 153 GiB 4.9 MiB 1019 MiB 1.6 TiB 8.63 1.30 32 up
14 hdd 7.27739 1.00000 7.3 TiB 489 GiB 488 GiB 4.1 MiB 1.2 GiB 6.8 TiB 6.56 0.99 39 up
15 hdd 7.27739 1.00000 7.3 TiB 435 GiB 434 GiB 1.8 MiB 1.1 GiB 6.9 TiB 5.84 0.88 37 up
16 hdd 7.27739 1.00000 7.3 TiB 326 GiB 325 GiB 1.3 MiB 1023 MiB 7.0 TiB 4.38 0.66 31 up
17 hdd 7.27739 1.00000 7.3 TiB 490 GiB 489 GiB 1.9 MiB 1.2 GiB 6.8 TiB 6.57 0.99 37 up
12 ssd 1.74660 1.00000 1.7 TiB 149 GiB 148 GiB 6.3 MiB 1018 MiB 1.6 TiB 8.34 1.26 31 up
13 ssd 1.74660 1.00000 1.7 TiB 159 GiB 158 GiB 41 MiB 994 MiB 1.6 TiB 8.91 1.34 34 up
TOTAL 91 TiB 6.0 TiB 6.0 TiB 172 MiB 19 GiB 85 TiB 6.63
MIN/MAX VAR: 0.50/1.42 STDEV: 1.73
root@ove-node1:~#
```

- 第三步：一顿猛如虎操作，然并卵

在Ceph后台一顿命令操作，包括OSD.10的重启、停止、启动、加载、卸载之后，OSD.10仍然无法上线，又再操作一顿命令，然并卵，直接到机房去看吧，看看哪块硬盘不亮了。

- 第四步：在机房现场找到故障硬盘

到机房之后，发现节点2服务器上有一块机械盘的灯不亮了，故障硬盘找到了，如图79所示。

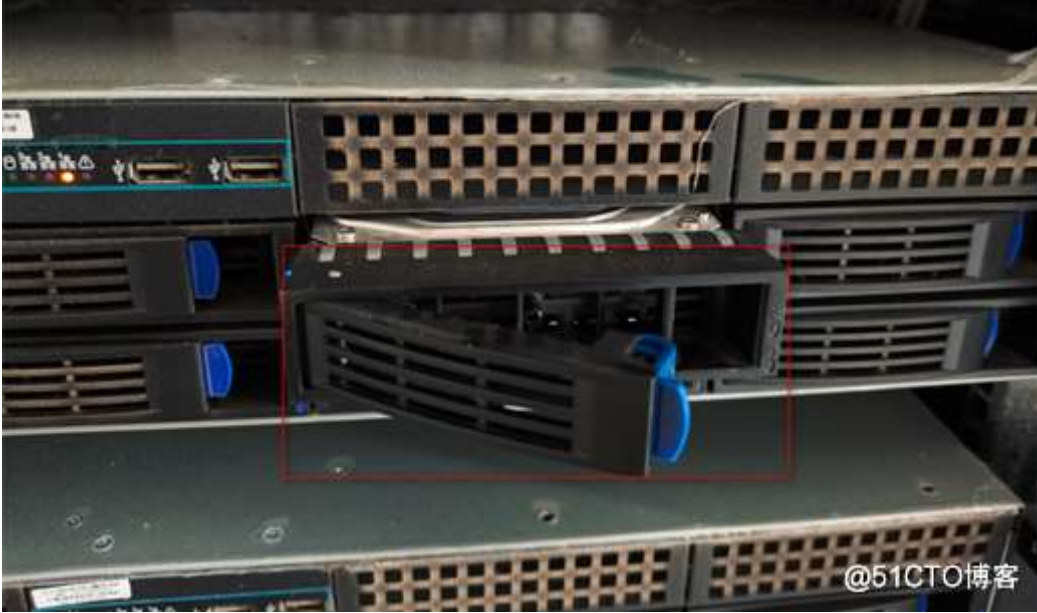
图79. 节点2的机械硬盘灯不亮了



- 第五步：物理拆卸进行检查

将硬盘拆卸下来进行检查，如图80、图81所示。经现场检查，发现硬盘的两个螺丝钉在松动，插槽里边是干净的，不知道是不是跟这螺丝松动有关。

图80. 拆卸故障硬盘



Prometheus(52)
Rancher(44)
Redis(1)
ServiceMesh(1)
Ubuntu(18)
Windows(2)
Zabbix(1)
常用工具(24)
代码管理(5)
更多

随笔档案

2022年10月(4)
2022年6月(3)
2022年5月(17)
2022年4月(25)
2022年3月(68)
2022年2月(18)
2021年12月(1)
2021年11月(1)
2021年10月(2)
2021年9月(50)
2021年8月(273)
2021年7月(79)
2020年11月(58)
2020年10月(58)

阅读排行榜

1. CentOS7安装源配置详解(9876)
2. leaf-spine （叶脊）网络架构介绍(7105)
3. PVE 备份和恢复介绍(5386)
4. Proxmox Virtual Environment 介绍(4947)
5. Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
6. Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
7. PVE 虚拟机模板和克隆介绍(3936)
8. 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
9. CentOS7安装配置远程桌面详解(3631)
10. Debian安装源配置详解(3457)
11. Apollo配置中心使用详解(3356)
12. Proxmox VE 防火墙介绍(3323)
13. Gitea 简单使用说明(3313)
14. Linux系统路由配置说明(3270)
15. Chrony时间同步服务器安装配置说明(3256)
16. Centos7防火墙配置详解(3231)
17. 【Rancher】简单介绍(3230)
18. Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
19. PVE 超融合架构介绍(3096)
20. 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

1. Proxmox VE 存储介绍(1)
2. Proxmox Virtual Environment 介绍(1)
3. Debian Jessie Preseed自动化安装示例(1)
4. Debian live-build官方教程学习笔记(1)
5. Ceph存储池pg_num配置详解(1)

推荐排行榜

1. leaf-spine （叶脊）网络架构介绍(1)
2. Proxmox VE 防火墙介绍(1)
3. Proxmox Virtual Environment 介绍(1)
4. VIM 常用快捷键(1)
5. Debian7 bonding配置详解(1)

最新评论

1. Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从"Default (No cache)"修改为"Write back (unsafe)", 对磁盘数据有影响么?

--大熊子

2. Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的以EMC FC-SAN存储（不支持ISCSI），如果改用Ceph做集群，FC-SAN应该以什么方式挂载给Proxmox好呢?

图81. 检查故障硬盘



- 第六步：重新插入硬盘，对故障OSD执行“启动”、“输入”操作

将硬盘松动的那两个螺丝拧紧之后，重新将故障硬盘插入盘位中，然后执行“启动”、“输入”操作，然并卵，故障硬盘一直掉线中，如图82所示。

图82. 故障硬盘无法上线

osd: 10										
启动 停止 重置 Out In 更多										
Name	Class	OSD Type	Status	版本	weight	reweight	已用 (%)	总量	Apply/C...	Latency ...
default										
bve-node3				15.2.9						
bve-node2				15.2.9						
osd 9	hdd	bluestore	up / in	15.2.9	7.27739	1.00	6.19	7.28 TiB	8 / 8	
osd 8	hdd	bluestore	up / in	15.2.9	7.27739	1.00	8.01	7.28 TiB	8 / 8	
osd 7	ssd	bluestore	up / in	15.2.9	1.7466	1.00	8.63	1.75 TiB	1 / 1	
osd 6	ssd	bluestore	up / in	15.2.9	1.7466	1.00	8.62	1.75 TiB	1 / 1	
osd 11	hdd	bluestore	up / in	15.2.9	7.27739	1.00	9.13	7.28 TiB	29 / 29	
osd 10	hdd	bluestore	down / out	15.2.9	7.27739	0.00	0.00	0 B	0 / 0	
bve-node1				15.2.9						

- 第七步：向服务器厂家报故障，厂家建议故障硬盘返厂检测

折腾一阵子之后，故障硬盘始终无法上线，只好打电话给服务器硬件厂家报故障，厂家反馈白牌服务器的硬盘容易出问题，建议把故障硬盘拆卸下来，返厂进行检测。

作者: [Varden](#)

出处: <http://www.cnblogs.com/varden/>

本文内容如有雷同，请联系作者！

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类: [服务器虚拟化](#), [Ceph](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

0

0

我在关注他 [取消关注](#)

« 上一篇: [Proxmox VE + Ceph 超融合项目实战（第七部分：Ceph分布式存储）](#)

» 下一篇: [Proxmox VE + Ceph 超融合项目实战（第九部分：一些常用的Ceph/OSD命令）](#)

posted @ 2021-08-31 14:43 [Varden](#) 阅读(1058) 评论(0) [编辑](#) [收藏](#) [举报](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

(评论功能已被禁用)

编辑推荐:

- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
- [分布式事务 | 使用 DTM 的 Saga 模式](#)
- [SQLSERVER 阻塞之 PFS 页到底是什么？](#)
- [带团队后的日常思考（十一）](#)
- [记一次使用 gdb 诊断 gc 问题全过程](#)

--9394952

[3. Re:Proxmox Virtual Environment 介绍](#)
很全面！

--willblog

[4. Re:Ceph存储池pg_num配置详解](#)
mark

--穷死了

[5. Re:Debian live-build官方教程学习笔记](#)
执行成功之后为什么没找看iso

--abraham_liuhai

阅读排行:

- [chatGPT vscode 体验 \(密钥已更新当天有效\)](#)
- [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
- [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
- [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

IT 随笔 FROM 2020 – Varden Zhang

首页 新随笔 联系 管理 订阅 

随笔- 657 文章- 0 评论- 5 阅读- 37万

昵称: Varden
园龄: 2年3个月
粉丝: 42
关注: 2
-取消关注

<	2023年2月						>
日	一	二	三	四	五	六	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

常用链接

- 我的随笔
- 我的评论
- 我的参与
- 最新评论
- 我的标签

最新随笔

- 1.Ceph 故障域配置 (修改CRUSH MAP)
- 2.RGW多站点同步策略配置
- 3.RGW故障转移和灾难恢复
- 4.RGW配置主区域
- 5.Ceph运维记录2022
- 6.Ceph 操作集群
- 7.常见磁盘总线的传输速率
- 8.cephadm 删除主机
- 9.使用s3cmd工具实现RGW对象存储生命周期管理
- 10.RGW池放置和存储类 (Octopus版本)

随笔分类

- Alpine(2)
- Ansible(23)
- CentOS(23)
- Ceph(118)
- Debian(19)
- Docker(28)
- ELK日志分析(15)
- Helm(3)
- Java(1)
- Jenkins(10)
- K8s(209)
- KubectI(4)
- Linux集群技术(3)
- Linux企业常用服务(9)
- Linux企业级安全(5)
- Linux系统调优和监控(8)
- Linux系统基础及管理(38)
- MongoDB(12)
- MySQL/MariaDB(6)
- NFS(3)
- PostgreSQL(1)

Proxmox VE + Ceph 超融合项目实战（第九部分：一些常用的Ceph/OSD命令）

参考: https://blog.51cto.com/u_14742102/3499271

九、一些常用的Ceph/OSD命令

1) 常见命令查看Ceph/OSD状态

```
root@pve-node1:~# ceph df
```

```
root@pve-node1:~# ceph df
--- RAW STORAGE ---
CLASS  SIZE      AVAIL    USED      RAW USED  %RAW USED
hdd    80 TiB    75 TiB   5.1 TiB   5.1 TiB    6.37
ssd    10 TiB    9.6 TiB  920 GiB   926 GiB    8.63
TOTAL  91 TiB    85 TiB   6.0 TiB   6.0 TiB    6.63

--- POOLS ---
POOL                                ID  PGS  STORED  OBJECTS  USED      %USED  MAX AVAIL
device_health_metrics              1    1    38 MiB      18    113 MiB      0    28 TiB
ssd_storage                        2   64    306 GiB   88.88k    919 GiB    9.10    3.0 TiB
hdd_storage                       3   64    1.7 TiB  452.93k    5.1 TiB    6.35    25 TiB
cephfs_data                       4   64      0 B        0      0 B        0    25 TiB
cephfs_metadata                   5   16    4.7 MiB     22    16 MiB      0    25 TiB
root@pve-node1:~#
```

```
root@pve-node1:~# ceph osd df
```

```
root@pve-node1:~# ceph osd df
ID CLASS WEIGHT REWEIGHT SIZE      RAW USE  DATA  OMAP  META  AVAIL  %USE  VAR  PGS  STATUS
 2  hdd  7.27739  1.00000  7.3 TiB  245 GiB  244 GiB  867 KiB  1.1 GiB  7.0 TiB  3.28  0.50  25    up
 3  hdd  7.27739  1.00000  7.3 TiB  518 GiB  516 GiB   39 MiB  1.2 GiB  6.8 TiB  6.95  1.05  39    up
 4  hdd  7.27739  1.00000  7.3 TiB  488 GiB  487 GiB   3.6 MiB  1.2 GiB  6.8 TiB  6.55  0.99  39    up
 5  hdd  7.27739  1.00000  7.3 TiB  489 GiB  488 GiB   3.4 MiB  1.2 GiB  6.8 TiB  6.57  0.99  42    up
 0  ssd  1.74660  1.00000  1.7 TiB  168 GiB  167 GiB   7.1 MiB 1017 MiB  1.6 TiB  9.39  1.42  35    up
 1  ssd  1.74660  1.00000  1.7 TiB  141 GiB  140 GiB   4.5 MiB 1020 MiB  1.6 TiB  7.86  1.19  29    up
 8  hdd  7.27739  1.00000  7.3 TiB  597 GiB  596 GiB   2.7 MiB  1.4 GiB  6.7 TiB  8.02  1.21  41    up
 9  hdd  7.27739  1.00000  7.3 TiB  461 GiB  460 GiB   3.1 MiB  1.1 GiB  6.8 TiB  6.19  0.93  46    up
10  hdd  7.27739  0          0 B      0 B      0 B      0 B      0 B      0 B      0      0      0    down
11  hdd  7.27739  1.00000  7.3 TiB  681 GiB  679 GiB   39 MiB  1.4 GiB  6.6 TiB  9.13  1.38  58    up
 6  ssd  1.74660  1.00000  1.7 TiB  154 GiB  153 GiB   8.3 MiB 1016 MiB  1.6 TiB  8.62  1.30  32    up
 7  ssd  1.74660  1.00000  1.7 TiB  154 GiB  153 GiB   4.9 MiB 1019 MiB  1.6 TiB  8.63  1.30  32    up
14  hdd  7.27739  1.00000  7.3 TiB  489 GiB  488 GiB   4.1 MiB  1.2 GiB  6.8 TiB  6.56  0.99  39    up
15  hdd  7.27739  1.00000  7.3 TiB  435 GiB  434 GiB   1.8 MiB  1.1 GiB  6.9 TiB  5.84  0.88  37    up
16  hdd  7.27739  1.00000  7.3 TiB  326 GiB  325 GiB   1.3 MiB 1023 MiB  7.0 TiB  4.38  0.66  31    up
17  hdd  7.27739  1.00000  7.3 TiB  490 GiB  489 GiB   1.9 MiB  1.2 GiB  6.8 TiB  6.57  0.99  37    up
12  ssd  1.74660  1.00000  1.7 TiB  149 GiB  148 GiB   6.3 MiB 1018 MiB  1.6 TiB  8.34  1.26  31    up
13  ssd  1.74660  1.00000  1.7 TiB  159 GiB  158 GiB   41 MiB  994 MiB  1.6 TiB  8.91  1.34  34    up
TOTAL  91 TiB  6.0 TiB  6.0 TiB  172 MiB  19 GiB  85 TiB  6.63
MIN/MAX VAR: 0.50/1.42  STDEV: 1.73
root@pve-node1:~#
```

```
root@pve-node1:~# ceph status
```

```
root@pve-node1:~# ceph status
cluster:
  id:      b7801b1e-2ef0-443f-9dce-00cbf90a6f45
  health: HEALTH_OK

services:
  mon: 3 daemons, quorum bve-node1,bve-node3,bve-node2 (age 3h)
  mgr: bve-node1(active, since 3M), standbys: bve-node2, bve-node3
  mds: cephfs:1 {0=bve-node3=up:active} 2 up:standby
  osd: 18 osds: 17 up (since 3h), 17 in (since 4M)

task status:

data:
  pools:   5 pools, 209 pgs
  objects: 541.85k objects, 2.1 TiB
  usage:   6.0 TiB used, 85 TiB / 91 TiB avail
  pgs:     209 active+clean

io:
  client:  0 B/s rd, 861 KiB/s wr, 26 op/s rd, 37 op/s wr
root@pve-node1:~#
```



```
root@pve-node1:~# ceph osd stat
```

```
root@pve-node1:~# ceph osd stat
18 osds: 17 up (since 3h), 17 in (since 4M); epoch = 3470
```

```
root@pve-node1:~# ceph osd tree
```

```
root@pve-node1:~# ceph osd tree
ID CLASS WEIGHT TYPE NAME STATUS REWEIGHT PRI-AFF
-1 97.80826 root default
-3 32.60275 host bve-node1
2 hdd 7.27739 osd.2 up 1.00000 1.00000
3 hdd 7.27739 osd.3 up 1.00000 1.00000
4 hdd 7.27739 osd.4 up 1.00000 1.00000
5 hdd 7.27739 osd.5 up 1.00000 1.00000
0 ssd 1.74660 osd.0 up 1.00000 1.00000
1 ssd 1.74660 osd.1 up 1.00000 1.00000
-7 32.60275 host bve-node2
8 hdd 7.27739 osd.8 up 1.00000 1.00000
9 hdd 7.27739 osd.9 up 1.00000 1.00000
10 hdd 7.27739 osd.10 down 0 1.00000
11 hdd 7.27739 osd.11 up 1.00000 1.00000
6 ssd 1.74660 osd.6 up 1.00000 1.00000
7 ssd 1.74660 osd.7 up 1.00000 1.00000
-10 32.60275 host bve-node3
14 hdd 7.27739 osd.14 up 1.00000 1.00000
15 hdd 7.27739 osd.15 up 1.00000 1.00000
16 hdd 7.27739 osd.16 up 1.00000 1.00000
17 hdd 7.27739 osd.17 up 1.00000 1.00000
12 ssd 1.74660 osd.12 up 1.00000 1.00000
13 ssd 1.74660 osd.13 up 1.00000 1.00000
```

2) 常见命令lsblk查看节点所有块设备

```
root@pve-node1:~# lsblk
```

```
root@pve-node1:~# lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
sda 8:0 0 1.8T 0 disk
├─ceph--e83bb23b--068a--43bf--Be2f--97457725a86e-osd--block--799ef595--a6c5--4dbe--84d6--ae063a706660 253:5 0 1.8T 0 lvm
├─sdb 8:16 0 1.8T 0 disk
├─ceph--8b33d230--9c46--418f--8a9f--e09a4dd059e6-osd--block--e37da15f--a71f--41e1--9742--baf0555d49ac 253:4 0 1.8T 0 lvm
├─sdc 8:32 0 447.1G 0 disk
├─sdc1 8:33 0 1007K 0 part
├─sdc2 8:34 0 512M 0 part
├─sdc3 8:35 0 446.6G 0 part
├─sdd 8:48 0 447.1G 0 disk
├─sdd1 8:49 0 1007K 0 part
├─sdd2 8:50 0 512M 0 part
├─sdd3 8:51 0 446.6G 0 part
├─sde 8:64 0 7.3T 0 disk
├─ceph--2952d019--94a3--4f3c--9c6e--c88264a9606a-osd--block--6b9d515c--6ba6--4933--b6f6--ab3012e7541a 253:3 0 7.3T 0 lvm
├─sdf 8:80 0 7.3T 0 disk
├─ceph--edc2c2cb--fde1--43a4--B8e0--dd0b0c886e26-osd--block--f0300fff--cf79--4a25--8ae1--538f7f2405f4 253:2 0 7.3T 0 lvm
├─sdg 8:96 0 7.3T 0 disk
├─ceph--75839c63--a32d--40fa--932a--1e9ec19e8a32-osd--block--e83be094--aeff--43b4--887b--11884a46d739 253:1 0 7.3T 0 lvm
├─sdh 8:112 0 7.3T 0 disk
└─ceph--e7c803cb--7e77--4521--82a1--7af5d2d18430-osd--block--2ccd7b90--5e10--40be--baa7--dbec9e265200 253:0 0 7.3T 0 lvm
```

从上图可以发现，sdc和sdd盘不属于Ceph盘，它其实是系统盘，做了RAID1的。

设备	类别	使用率	大小	GPT	模型	串行	S.M.A.R.T.	Wearout
/dev/sda	SSD	LVM	1.75 TiB	否	INTEL_SSDSC2KB019T8	PHYF052207HQ1P9DGN	PASSED	0%
/dev/sdb	SSD	LVM	1.75 TiB	否	INTEL_SSDSC2KB019T8	PHYF052204571P9DGN	PASSED	0%
/dev/sdc	SSD	ZFS	447.13 GiB	是	INTEL_SSDSC2KB480G8	PHYF0051000D400BGN	PASSED	0%
/dev/sdd	SSD	ZFS	447.13 GiB	是	INTEL_SSDSC2KB480G8	PHYF0052037M480BGN	PASSED	0%
/dev/sde	Hard Disk	LVM	7.28 TiB	否	ST8000NM000A-2KE101	WKD3LTG8	PASSED	N/A
/dev/sdf	Hard Disk	LVM	7.28 TiB	否	ST8000NM000A-2KE101	WKD3LA1K	PASSED	N/A
/dev/sdg	Hard Disk	LVM	7.28 TiB	否	ST8000NM000A-2KE101	WRD0CYHS	PASSED	N/A
/dev/sdh	Hard Disk	LVM	7.28 TiB	否	ST8000NM000A-2KE101	WRD0E46Z	PASSED	N/A

作者: [Varden](#)
出处: <http://www.cnblogs.com/varden/>
本文内容如有雷同，请联系作者！
本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，否则保留追究法律责任的权利。

分类: [服务器虚拟化](#), [Ceph](#)

好文要顶

已关注

收藏该文

[Varden](#)
粉丝 - 42 关注 - 2

我在关注他 [取消关注](#)

« 上一篇: [Proxmox VE + Ceph 超融合项目实战（第八部分：OSD磁盘故障排查案例）](#)
» 下一篇: [PVE 集群配置使用外部 Ceph 存储](#)

[Prometheus\(52\)](#)
[Rancher\(44\)](#)
[Redis\(1\)](#)
[ServiceMesh\(1\)](#)
[Ubuntu\(18\)](#)
[Windows\(2\)](#)
[Zabbix\(1\)](#)
[常用工具\(24\)](#)
[代码管理\(5\)](#)
[更多](#)

随笔档案

[2022年10月\(4\)](#)
[2022年6月\(3\)](#)
[2022年5月\(17\)](#)
[2022年4月\(25\)](#)
[2022年3月\(68\)](#)
[2022年2月\(18\)](#)
[2021年12月\(1\)](#)
[2021年11月\(1\)](#)
[2021年10月\(2\)](#)
[2021年9月\(50\)](#)
[2021年8月\(273\)](#)
[2021年7月\(79\)](#)
[2020年11月\(58\)](#)
[2020年10月\(58\)](#)

阅读排行榜

- CentOS7安装源配置详解(9876)
- leaf-spine （叶脊）网络架构介绍(7105)
- PVE 备份和恢复介绍(5386)
- Proxmox Virtual Environment 介绍(4947)
- Proxmox VE Qemu/KVM 虚拟机设置要求(4533)
- Proxmox VE + Ceph 超融合项目实战（第五部分：虚拟网络）(4297)
- PVE 虚拟机模板和克隆介绍(3936)
- 【K8s教程】Nginx Ingress控制器Ingress Path匹配(3843)
- CentOS7安装配置远程桌面详解(3631)
- Debian安装源配置详解(3457)
- Apollo配置中心使用详解(3356)
- Proxmox VE 防火墙介绍(3323)
- Gitea 简单使用说明(3313)
- Linux系统路由配置说明(3270)
- Chrony时间同步服务器安装配置说明(3256)
- Centos7防火墙配置详解(3231)
- 【Rancher】简单介绍(3230)
- Proxmox VE + Ceph 超融合项目实战（第三部分：PVE安装与基础配置）(3098)
- PVE 超融合架构介绍(3096)
- 【K8s任务】为容器设置启动时要执行的命令和参数(3071)

评论排行榜

- Proxmox VE 存储介绍(1)
- Proxmox Virtual Environment 介绍(1)
- Debian Jessie Preseed自动化安装示例(1)
- Debian live-build官方教程学习笔记(1)
- Ceph存储池pg_num配置详解(1)

推荐排行榜

- leaf-spine （叶脊）网络架构介绍(1)
- Proxmox VE 防火墙介绍(1)
- Proxmox Virtual Environment 介绍(1)
- VIM 常用快捷键(1)
- Debian7 bonding配置详解(1)

最新评论

- Re:KVM虚拟化Disk Cache模式对性能的影响分析
如果从"Default (No cache)"修改为"Write back (unsafe)", 对磁盘数据有影响么?
--大熊子
- Re:Proxmox VE 存储介绍
请教一下：原来用在VMWARE集群的以EMC FC-SAN存储（不支持ISCSI），如果改用X做集群，FC-SAN应该以什么方式挂载给PROX做集

(评论功能已被禁用)

编辑推荐:

- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他](#)
- [分布式事务 | 使用 DTM 的 Saga 模式](#)
- [SQLSERVER 阻塞之 PFS 页到底是什么?](#)
- [带团队后的日常思考 \(十一\)](#)
- [记一次使用 gdb 诊断 gc 问题全过程](#)

阅读排行:

- [chatGPT vscode 体验 \(密钥已更新当天有效\)](#)
- [【.NET 8】ASP.NET Core计划 - 支持更完善的AOT发布](#)
- [如何避免让线程摸鱼，请用异步技术 async await 拿捏他~](#)
- [花了半个小时基于 ChatGPT 搭建了一个微信机器人](#)
- [不用Blazor WebAssembly，开发在浏览器端编译和运行C#代码的网站](#)

--9394952

[3. Re:Proxmox Virtual Environment 介绍](#)
很全面!

--willblog

[4. Re:Ceph存储池pg_num配置详解](#)
mark

--穷死了

[5. Re:Debian live-build官方教程学习笔记](#)
执行成功之后为什么没找看iso

--abraham_liuhai