ceph 测试环境搭建

runsisi@hust.edu.cn
http://www.cppblog.com/runsisi

前言

本文只适用于 ceph 测试环境的搭建，实际生产环境下的部署请参考 ceph 官方相关文档。实际上 ceph 官方对测试环境的部署已有比较详细的介绍，本文内容主要参考了官方的文档，只是对部署过程稍作简化，如需参阅官方文档请点击：http://ceph.com/docs/master/start/

环境描述

由于缺少实际硬件环境，我们使用 VMware 模拟所需的硬件环境，对于测试环境而言，这样做无可厚非，理论上任意 VMware 版本应该都是可以的。

Ceph 只能部署在 linux 系统上，其运行环境对发行版理论上无要求，但 ceph 官方提供的 ceph 测试环境部署工具(ceph-deploy)可能对发行版有要求，同时 ceph 客户端对内核以及 glibc 版本也有一定要求，具体环境要求请参阅官方文档：http://ceph.com/docs/master/start/os-recommendations/

Linux 各发行版之间差异很大，官方的部署文档基本上是分 debian 和 redhat 两种类型的发行版分别进行介绍，由于精力有限，本文不对 debian 系列（debian, ubuntu, linux mint, linuxdeepin 等）进行介绍。

本文用到的环境清单如下：

硬件环境：VMware workstation 10.0.2

软件环境：Centos 7.0 x86_64

准备工作

组网环境如图 1 所示，我们用 VMware 虚拟出来四台 guest 机器，主机名分别为 admin-node, node1, node2, node3，其中 admin-node 用于运行 ceph-deploy 测试环境部署工具，它并不属于 ceph 存储集群，node1 用于运行 monitor，node2 和 node3 用于运行 osd，为简单起见，四台机器属于同一个子网，当然这只是一个基础的测试环境，后续还可以做进一步的扩展。

由于 linux 对网络的依赖性，四台 guest 机器都必须保证能够连接 Internet，

当然如果能够手工解决包之间的依赖关系或者有内部架设的源服务器也是不需要 Internet 连接的。由于公司的网络环境无法连接外网，又不想浪费时间去折腾，我的环境是在自己家里的机器上搭建完成然后拷贝到公司电脑上运行的。

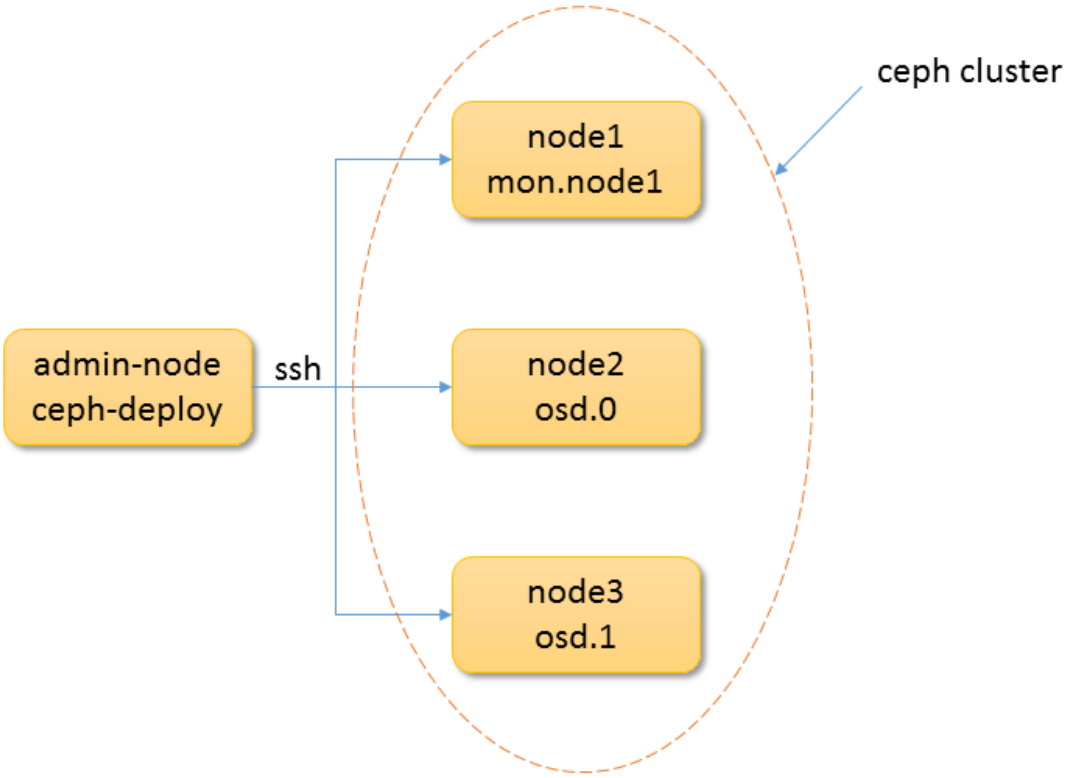

图 1. ceph 测试环境组网

首先使用 VMware 创建第一台 guest 机器并安装 Centos 系统，为了避免重启 guest 机器导致 IP 地址变化，虚拟机的网络选择桥接网络，如图 2 所示，关于 VMware 各种网络连接模式之间的区别请参考网上的相关资料。

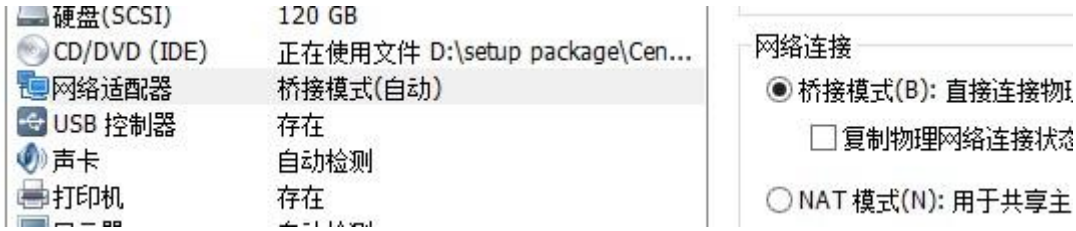

图 2. guest 机器网络连接模式

给第一台 guest 机器装好系统后，可以使用克隆的方式创建另外的三台机器，当然如果不嫌麻烦可以按照第一台的方式单独安装。克隆操作请在 guest 机器关闭的情况下进行，如图 3 所示。

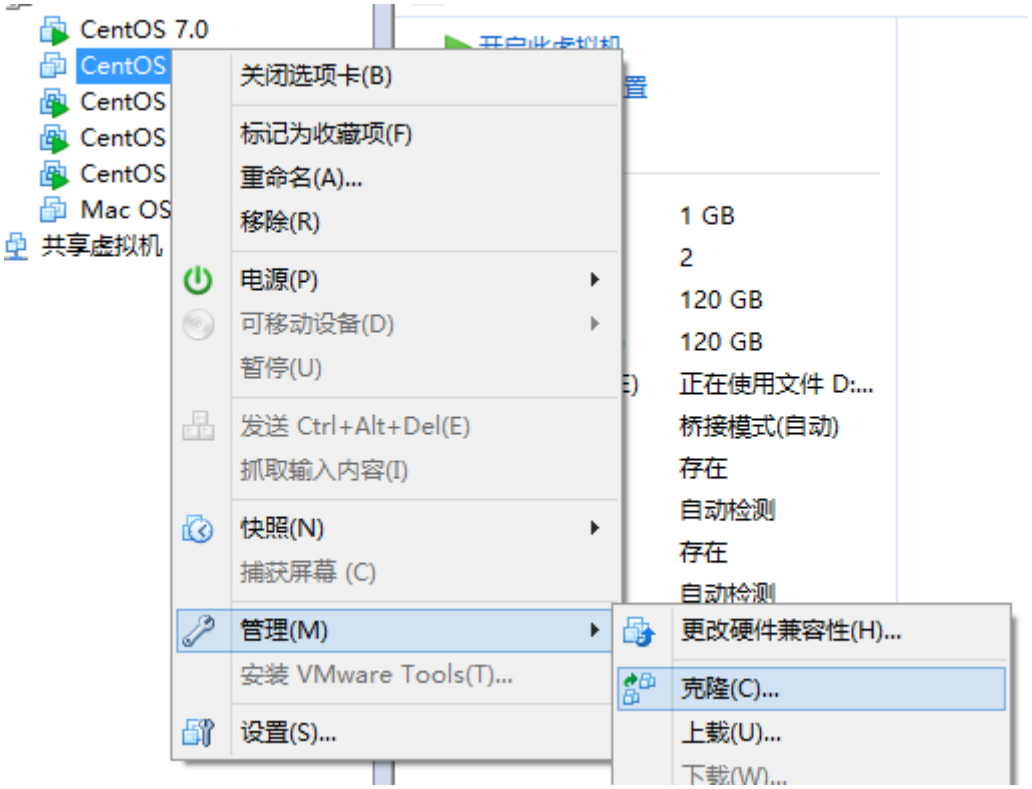

图 3. guest 机器克隆操作

如前所述，guest 机器使用的是桥接模式网络，在四台 guest 机器创建完成后，请将各机器的 IP 地址获取方式修改为手动，并配置相应的 IP 地址和掩码，使它们属于同一个子网。ceph-deploy 工具在使用过程中基本上都需要指定 ceph 节点，如果参数给的是主机名，且没有通过额外的参数显式指定节点的 IP 地址，则 ceph-deploy 会尝试对给的主机名进行名字解析。为了简单起见，我们这里都只指定主机名。

为了支持名字解析，最简单的方式就是修改 guest 机器中的/etc/hosts 文件，如图 4 中 4~7 行为手工添加的主机名与 IP 地址的映射关系，虽然实际上只需要在 admin-node 节点添加这样对应关系，但为了方便管理在这里我给所有的四台 guest 机器都添加了图 4 所示的关系。

然后修改各 guest 机器的主机名与/etc/hosts 文件中的主机名对应，注意 RHEL/Centos 7.x 系列版本修改主机名是修改/etc/hostname 文件，而不是 6.x 系列的/etc/sysconfig/network 文件。

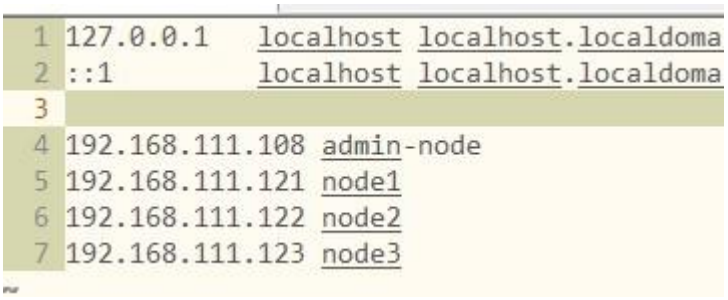

图 4. 修改后的/etc/hosts 文件

ceph-deploy 工具通过 ssh 登录各 ceph 集群节点进行环境部署，因此如果集群节点没有安装 ssh 服务端，则需要在各集群节点执行安装操作。

```
yum install openssh-server
```

作为一个存储集群，ceph 各节点间自然会有数据交互，Centos 7.0 默认会安装并启用防火墙，从而影响 ceph 节点间通信，由于是测试环境我们直接禁用防火墙服务。

```
systemctl stop firewalld
```

```
systemctl disable firewalld
```

注意这里相比 Centos 6.x 系列有两个明显的不同，服务使用 systemctl 控制接口，原有的防火墙服务 iptables 默认不再启用，取而代之的是 firewalld 服务。

现在组网环境已经搭建完成，接下来开始真正的 ceph 测试环境搭建。注意整个环境搭建过程中，我们都直接使用 root 用户进行操作，非 root 用户只是给自己增加麻烦而已，看好多人把普通用户加到了 sudoer 列表中然后又赋予该用户以 root 用户执行所有命令的权限，每个命令都 sudo 一把，呵呵。

1) 修改各 guest 机器允许通过 ssh hostname sudo <cmd>执行命令

修改/etc/sudoers 的第 56 行，改成如图 5 所示，或直接注释掉该行，注意请使用 visudo 修改该文件，实际上只需要修改三个 ceph 集群节点即可。

```
52 #
53 # Disable "ssh hostname sudo <cmd>", because it will show
54 #         You have to run "ssh -t hostname sudo <cmd>".
55 #
56 Defaults    !requiretty
57
```

图 5. 修改后的/etc/sudoers 文件

2) 允许 admin-node 以无密码方式访问 ceph 节点

```
ssh-keygen
```

```
ssh-copy-id node1
ssh-copy-id node2
ssh-copy-id node3
```

3) 在 admin-node 节点安装 ceph-deploy 工具

这里没有使用通过 ceph 官方源直接 yum 安装 ceph-deploy 的方式，而是直接克隆源代码的方式，注意只需要在 admin-node 节点安装即可。

```
yum install git
```

```
git clone https://github.com/ceph/ceph-deploy.git
```

首先切换到 ceph-deploy 目录下，在本文完成时，ceph-deploy 还不支持 Centos 7.0，为避免最新的代码可能存在问题，我们先切换到当前最新的版本：

```
git checkout v1.5.9 v1.5.9
```

对 ceph-deploy 源代码稍作修改以适用于 Centos 7.0，打开：

```
ceph_deploy/hosts/centos/install.py
```

如图 6 所示，修改第 7 行，并增加 27~28 行，之所以需要这样做因为是 ceph 官方暂时还没提供 Centos 7.0 的 ceph 源，但提供了 RHEL 7.0 的，因此我们使用 RHEL 的即可，如果不修改，则在执行 ceph-deploy install 的时候会下载适用于 Centos 6.x 的 ceph 包，安装时会出现包依赖关系无法解决的问题。

安装 ceph-deploy：

```
python setup.py install
```

解决对 python-flask 的依赖问题：

使用 RHEL 的源在安装 ceph 时貌似存在无法解决对 python-flask 依赖的问题，请访问 http://ceph.com/rpm-firefly/rhel7/noarch/ 下载如图 7 所示的 3 个 rpm 包，然后手工安装，注意需要在所有 guest 机器都进行安装。

```
def rpm_dist(distro):
    # start using the el7 prefix now that rhel7 exists.
    if distro.release.startswith('7'):
        return 'el7'
    return 'el6'

def repository_url_part(distro):
    """
    Historically everything CentOS, RHEL, and Scientific ha
    `el6` urls, but as we are adding repositories for `rhel
    map correctly to, say, `rhel6` or `rhel7`.

    This function looks into the `distro` object and determ
    part for the given distro, falling back to `el6` when a

    Specifically to work around the issue of CentOS vs RHEL

        >>> import platform
        >>> platform.linux_distribution()
        ('Red Hat Enterprise Linux Server', '7.0', 'Maipo')

    """
    if distro.release.startswith('7'):
        return 'rhel7'
    if distro.normalized_name == 'redhat':
        if distro.release.startswith('6'):
            return 'rhel6'
        elif distro.release.startswith('7'):
            return 'rhel7'
    return 'el6'
```

图 6. 修改后的 ceph_deploy/hosts/centos/install.py 文件

ceph-deploy-1.5.7-0.noarch.rpm	01-Jul-2014 13:56	207K
ceph-deploy-1.5.8-0.noarch.rpm	09-Jul-2014 08:52	209K
ceph-deploy-1.5.9-0.noarch.rpm	14-Jul-2014 10:13	209K
ceph-release-1-0.el7.noarch.rpm	07-May-2014 12:39	3.8K
python-flask-0.10.1-3.el7.noarch.rpm	07-May-2014 12:13	204K
python-itsdangerous-0.23-1.el7.noarch.rpm	07-May-2014 12:13	23K
python-werkzeug-0.9.1-1.el7.noarch.rpm	07-May-2014 12:13	562K
repodata/	23-Jul-2014 09:39	-

图 7. python-flask 依赖包

4) 创建 ceph 集群

注意：所有 ceph-deploy 命令都在 admin-node 节点执行，同时请保证所有命令执行时当前目录不变。

指定 node1 节点为 monitor，在当前生成集群配置文件：

```
ceph-deploy new node1
```

修改集群配置文件：

```
vi ceph.conf
```

暂时只准备在 node2 和 node3 上建立 osd，在该配置文件末尾加上：

```
osd pool default size = 2
```

登录各集群节点安装 ceph，同时在 admin-node 节点也进行安装：

```
ceph-deploy install node1 node2 node3 admin-node
```

创建初始的 monitor：

```
ceph-deploy mon create-initial
```

准备 osd 所需的磁盘空间，可以使用真实的磁盘分区，也可以使用文件夹进行模拟，这里使用 node2 和 node3 上的文件夹进行模拟。

分别登录 node2 和 node3 创建模拟磁盘分区所需要的空文件夹：

```
ssh node2
mkdir /var/local/osd0
exit
```

```
ssh node3
mkdir /var/local/osd1
exit
```

分区格式化，准备 osd 环境：

```
ceph-deploy osd prepare node2:/var/local/osd0
node3:/var/local/osd1
```

启用 osd：

```
ceph-deploy osd activate node2:/var/local/osd0
node3:/var/local/osd1
```

拷贝/etc/ceph/ceph.conf 以及/etc/ceph/ceph.client.admin.keyring 到各 ceph 节点以及 admin-node 节点以方便使用：

```
ceph-deploy admin node1 node2 node3 admin-node
```

ceph 正常运行需要各节点的时间保持同步，因此最好假设 NTP 服务器，Centos 7.0 相比 Centos 6.x 系统使用 chronyd 替换了原有的 ntpd，这一点需要注意。