

Ceph13.2.8安装手册

目录

[1 文档概要 3](#_Toc48825592)

[1.1 目的 3](#_Toc48825593)

[1.2 适用人员 3](#_Toc48825594)

[1.3 本文约定规定 3](#_Toc48825595)

[1.4 ceph基本知识 3](#_Toc48825596)

[1.5 环境需求 4](#_Toc48825597)

[2 ceph安装 6](#_Toc48825598)

[2.1 下载安装脚本 6](#_Toc48825599)

[2.1.1 解压安装文件 6](#_Toc48825600)

[2.2 编辑安装配置文件 6](#_Toc48825601)

[2.2.1 编辑ceph-install.conf 6](#_Toc48825602)

[2.2.2 编辑ceph-ansible/host-inventory 6](#_Toc48825603)

[2.2.3 编辑ceph-ansible-3.2.35/group\_vars/all.yml 8](#_Toc48825604)

[2.2.4 执行安装脚本 12](#_Toc48825605)

[2.2.5 检查ceph集群状态 12](#_Toc48825606)

[2.2.6 添加其他的OSD 12](#_Toc48825607)

[2.2.7 Cephfs挂载 12](#_Toc48825608)

[2.3 Ceph优化 13](#_Toc48825609)

[2.3.1 创建存储规则 13](#_Toc48825610)

# 文档概要

## 目的

本文档主用于ceph 13.2.8版本安装。

## 适用人员

适用于对ceph原理有一定的了解，熟悉linux 基本操作命令。

## 本文约定规定

1．#开的是需要执行的linux命令。

2．红色字体的是要根据实际情况替换

## ceph基本知识

1. PG数据的计算

推荐每个OSD上的PG数配置为100~150之间,每个存储池的PG数只能调大,不能调小,

PG数必须是最接近2的N次方幂的值

PG总数=(OSD总数 \* 100)/最大副本数

存储池PG总数=(OSD 总数 \* 100/最大副本数)/池数

1. Ceph部署架构



部署说明:

1. 默认mon,mds,rgw分别部署3个,mgr部署2个
2. 生产环境建议业务网络和存储同步网络分开

## 环境需求

1. 操作系统:CentOS 7.4, CentOS 7.5,CentOS 7.6
2. 操作系统基本环境配置(必须)
3. 关闭firewalld
4. 关闭selinux
5. 设置ntp服务
6. 主机名设置
7. 配置主机名解析
8. 设置固定IP

以上配置参考系统安装手册

1. 生产环境硬件建议：
2. 节点：至少三台，配置相同，内存至少32G。
3. 磁盘：系统盘SSD 240G以上或都300G SAS ,RAID1,每台至少两个SSD,用于存储hdd磁盘的元数据和对象存储的索引数据，其他若干个HDD磁盘。
4. 网络：至少一个千兆网络，生产环境建议万兆网络。
5. 其他说明
6. 对象存储默认端口8080,

对象存储的接口地址：http://rgw\_ip:8080

1. mgr端口7000,用户名/密码 admin/edoc2

Ceph\_HealthUrl的配置地址: http://mgr\_ip:7000/api/auth

1. Ceph版本为13.2.8
2. 创建默认用户

access\_key: W6XCJFJX2ETBB387IK4D

secret\_key=yh8wpPjiQ7wNfbGhKYkFHSNa7OBPHaIWjEQgOPlD

默认存储桶: eDoc2

1. 磁盘查看

|  |
| --- |
| #lsblk -d -o name,rota,size |

ROTA字段: 0表示是SSD,1表示为HDD磁盘

|  |
| --- |
| # echo 0 > /sys/block/{sdb}/queue/rotational |

注: 把一个磁盘强制设为ssd,替换命令中标红的参数

# ceph安装

## 下载安装脚本

[root@node1 ~]# curl http://download.edoc2.com:5999/ceph/ceph\_install\_13.2.8.tar.gz -o ceph\_install\_13.2.8.tar.gz

### 解压安装文件

|  |
| --- |
| [root@node1 ~]# tar -zxvf ceph\_install\_13.2.8.tar.gz  [root@node1 ~]# cd ceph\_install\_13.2.8 |

## 编辑安装配置文件

Ceph的安装配置文件有以下三个： ceph\_install.conf，ceph-ansible-3.2.35/ host-inventory，ceph-ansible-3.2.35/group\_vars/all.yml

### 编辑ceph-install.conf

|  |
| --- |
| [root@node1 ~]# vi ceph-install.conf  HOST=  #填写要安装ceph的主机IP，有多个主机时用逗号分䣓开，如：192.168.251.132,192.168.251.133  PASSWORD=  #要保证三台主机密码相同 |

### 编辑ceph-ansible/host-inventory

以下配置以有IP1,IP2,IP3三台ceph节点配置为例，分别在三台节点上安装mon,mgr,mds,rgw,osd服务，在不使用对象存储的环境，不用安装rgw的服务。

|  |
| --- |
| [root@node1 ~]# vi ceph-ansible-3.2.35/host-inventory  [mons]  #IP1  #192.168.251.154 monitor\_address=192.168.251.154  #按照这个格式配置,集群环境配置三个IP  [mgrs]  #IP1  #192.168.251.154  #配置两个节点  [mdss]  #IP1  #192.168.251.154  #集群环境配置3个节点  [rgws]  #IP1  #192.168.251.154 radosgw\_address=192.168.251.154  #按照这个格式配置,集群环境配置三个IP  [osds]  #IP1 devices="[]"  #192.168.251.154 devices="['/dev/sdb','/dev/sdc','/dev/sdd','/dev/sdi']"  以上配置说明：假设现在有IP1，IP2，IP3三台ceph节点，每台服务器有sdb,sdc,sdd三块hdd磁盘，一块sdi的ssd磁盘，用于每个osd元数据存储,如果有多块SSD只填写需要写日志的那块SSD即可。 |

使用实际IP替换，mons，mgrs，mdss，rgws建议不超过三个节点，osds下面可以配置多个，主要用于数据存储。如果不使用对象存储，不用配置rgws 的IP

### 编辑ceph-ansible-3.2.35/group\_vars/all.yml

注意：配置文件中默认为1副本，生产环境建议设置为3副本，即以下配置文件中：

size: "1" 更改为size: "3"

|  |
| --- |
| ---  dummy:  ceph\_origin: distro  ceph\_stable\_release: mimic  configure\_firewall: false  ntp\_service\_enabled: false  fsid: 78bcda49-b1c8-440f-8138-8440b44800a9  # 使用uuidgen生成一个集群ID  generate\_fsid: false  fetch\_directory: fetch/  cephx: true  copy\_admin\_key: false  ip\_version: ipv4  public\_network: 192.168.251.0/24  #替换public的网络地址,24位掩码,最后一位是0  osd\_objectstore: bluestore  osd\_auto\_discovery: false  osd\_memory\_target: 2147483648  osd\_scenario: lvm  os\_tuning\_params:  - { name: vm.min\_free\_kbytes, value: 1048576 }  ceph\_mgr\_modules: [dashboard,status,restful]  mds\_max\_mds: 1  cephfs: cephfs  cephfs\_data: cephfs\_data  cephfs\_metadata: cephfs\_metadata  cephfs\_pools:  - { name: "{{ cephfs\_data }}", pgs: "64", size: "1" }  - { name: "{{ cephfs\_metadata }}", pgs: "16", size: "1" }  radosgw\_num\_instances: 1  radosgw\_civetweb\_port: 8080  radosgw\_civetweb\_num\_threads: 500  rgw\_override\_bucket\_index\_max\_shards: 128  rgw\_multisite: true #需要部署multi-site把此项修改为false，安装完成后参考multisite部署手册  rgw\_realm: macrowing  rgw\_zonegroup: edoc2  rgw\_zone: master  rgw\_zone\_user: admin  rgw\_zonemaster: true  rgw\_multisite\_endpoint\_addr: "{{ radosgw\_address }}"  system\_access\_key: 6kWkikvapSnHyE22P7nO  system\_secret\_key: MGecsMrWtKZgngOHZdrd6d3JxGO5CPWgT2lcnpSt  rgw\_create\_pools:  master.rgw.buckets.data:  pg\_num: 64  size: "1"  master.rgw.buckets.index:  pg\_num: 16  size: "1"  master.rgw.control:  pg\_num: 16  size: "1"  master.rgw.meta:  pg\_num: 16  size: "1"  master.rgw.log:  pg\_num: 16  size: "1"  master.rgw.buckets.non-ec:  pg\_num: 16  size: "1"  ceph\_conf\_overrides:  global:  osd pool default size: 1  osd pool default pg num: 8  osd pool default pgp num: 8  rgw num rados handles: 2  rgw multipart part upload limit: 50000  mon:  mon allow pool delete: true  mon max pg per osd: 500  mds:  mds standby replay: True  mds cache memory limit: 2147483648  osd:  osd scrub begin hour: 23  osd scrub end hour: 6  osd scrub sleep: 3  osd scrub chunk min: 1  osd scrub chunk max: 1 |

### 执行安装脚本

|  |
| --- |
| 执行部署脚本，部署过程时间比较长  [root@node1 ~]# sh ceph-deploy.sh |

### 检查ceph集群状态

|  |
| --- |
| [root@node1 ~]# ceph -s  cluster:  id: 197da3fb-2e86-420b-85ae-7c3f67332c74  health: HEALTH\_OK  出现HEALTH\_OK时说明集群状态正常 |

### 添加其他的OSD

登录需要添加osd的服务器,执行如果命令,

比如现在要添加sdi这个磁盘.设置名称根据实际情况替换.

|  |
| --- |
| # ceph-volume lvm batch /dev/sdi --bluestore --yes |

### Cephfs挂载

|  |
| --- |
| [root@node1 ~]# vi /etc/fstab  192.168.251.176:6789,192.168.251.154:6789:/ /MOUNT/PATH ceph defaults,\_netdev,name=admin,secret=AQAaOERbsf2gHhAAwG2KnYrxdNDx++rYYwULZw== 0 0  选项说明：  192.168.251.176:6789：mon节点的IP和端口号，多个mon节点之间用逗号隔开。  /MOUNT/PATH ： 本地的挂载目录  AQAaOERbsf2gHhAAwG2KnYrxdNDx++rYYwULZw== ： cephx认证的密钥，在mon节点使用cat /etc/ceph/ceph.client.admin.keyring命令可以获取密钥，key等号后面的为密钥 |

挂载ceph

[root@node1 ~]# mount -a

## Ceph优化

### 创建存储规则

|  |
| --- |
| # ceph osd crush rule create-replicated ssd-rule default host ssd  # ceph osd crush rule create-replicated hdd-rule default host hdd |

设置存储池的存储规则

|  |
| --- |
| # ceph osd pool set cephfs\_data crush\_rule ssd-rule  # ceph osd pool set master.rgw.buckets.data crush\_rule ssd-rule |

Ceph默认有以下存储池:

cephfs\_data # cephfs存储的数据,建议使用ssd存储池

cephfs\_metadata #cephfs的元数据,建议使用ssd存储池

.rgw.root #对象存储数据

master.rgw.control #对象存储的数据

master.rgw.meta #对象存储的元数据

master.rgw.log #对象存储的日志

master.rgw.buckets.non-ec #对象存储的临时文件

master.rgw.buckets.data #对象存储的实体数据

master.rgw.buckets.index #对象存储的索引数据,建议使用ssd