誉天教育
YU TIAN EDU

Linux云计算架构师涨薪班

Ceph集群配置

学习目标

- ceph配置文件
- ceph中的元变量
- 临时修改ceph配置
- 持久修改ceph配置
- 管理集中配置数据库

Ceph配置文件说明

- 默认情况下，无论是ceph的服务端还是客户端，配置文件都存储在/etc/ceph/ceph.conf文件中
- 如果修改了配置参数，必须使用/etc/ceph/ceph.conf文件在所有节点（包括客户端）上保持一致。
- ceph.conf 采用基于 INI 的文件格式，包含具有 Ceph 守护进程和客户端相关配置的多个部分。每个部分具有一个使用 [name] 标头定义的名称，以及键值对的一个或多个参数
- 配置文件使用#和;来注释
- 参数名称可以使用空格、下划线、中横线来作为分隔符。如osd journal size 、osd_jounrnal_size 、 osd-journal-size是有效且等同的参数名称
- 通过中括号将特定守护进程的设置分组在一起：

ceph的配置文件

- ceph 全局配置文件 /etc/ceph.conf
- [global] 部分存储所有守护进程或读取配置的任何进程（包括客戶端）所共有的一般配置。
- [mon] 部分存储监控器 (MON) 的配置。
- [osd] 部分存储 OSD 守护进程的配置。
- [mgr] 部分存储管理器 (MGR) 的配置。
- [mds] 部分存储元数据服务器 (MDS) 的配置。
- [client] 部分存储应用到所有 Ceph 客戶客戶端的配置。

元变量

所谓元变量是即Ceph内置的变量。可以用它来简化ceph.conf文件的配置：

- $cluster
  - 红帽 Ceph 存储 5 集群的名称。默认集群名称为 ceph。
- $type
  - 守护进程类型，如监控器的值为 mon。OSD 使用 osd，元数据服务器使用 mds，管理器使用mgr，客戶端软件使用 client。
- $id
  - 守护进程实例 ID。对于此变量，serverc 上监控器的值为 serverc。osd.1 的 $id 值为 1 ，客戶端应用的值为用戶名。
- $name
  - 守护进程名称和实例 ID。此变量是 $type.$id 的简写。
- $host
  - 运行守护进程的主机的名称。

配置文件路径

- $CEPH_CONF（CEPH_CONF环境变量所指示的路径）　优先级较高
- -c path / path（ceph -c）　优先级最高
- /etc/ceph/ceph.conf　　优先级最低　全局的默认配置文件
- ~/.ceph/ceph.conf　　优先级低于当前工作路径
- ./ceph.conf（就是当前所在的工作路径）　优先级低于环境变量和直接指定
- 仅限FreeBSD系统，　/usr/local/etc/ceph/$cluster.conf

ceph运行时修改临时有效

- 使用tell修改
  - ceph tell $type.$id config set mon_allow_pool_delete true 修改某一个服务的配置
  - ceph tell $type.* config set mon_allow_pool_delete true 修改一整个类型
- 使用daemon修改（进入指定容器内进行）
  - ceph daemon <name> config set <option> <value>
  - ceph daemon osd.4 config set debug_osd 20

  差异对比： daemon可以在mon故障时进行对容器的设置，而tell依赖于mon

配置集中式数据库

- 查看集群中所有的配置项
  - ceph config ls 列出所有配置项
  - ceph config help setting 有助于进行特定配置设置
  - ceph config dump 可显示集群配置数据库设置
  - ceph config show $type.$id 可显示特定守护进程的数据库设置
  - ceph config show-with-defaults $type.$id 查看特定守护进程的默认配置项
- 查看集群中当前生效的配置项
  - 使用 ceph config get $type.$id 可获得特定配置设置
  - 使用 ceph config set $type.$id 可设置特定配置设置

从文件中读取ceph的配置项

- 使用 assimilate-conf 子命令可将文件中的配置应用到正在运行的集群。此过程将会识别配置文件中更改的设置并将其应用到集中式数据库
- 注意：只能将集群中默认的选项改为配置项，无法将已经修改的配置再改变回去

example： vim ~/test.conf

```
[mon]
mon_allow_pool_delete = true
```

```
ceph config assimilate-conf -i ~/test.conf
```

管理集中配置数据库

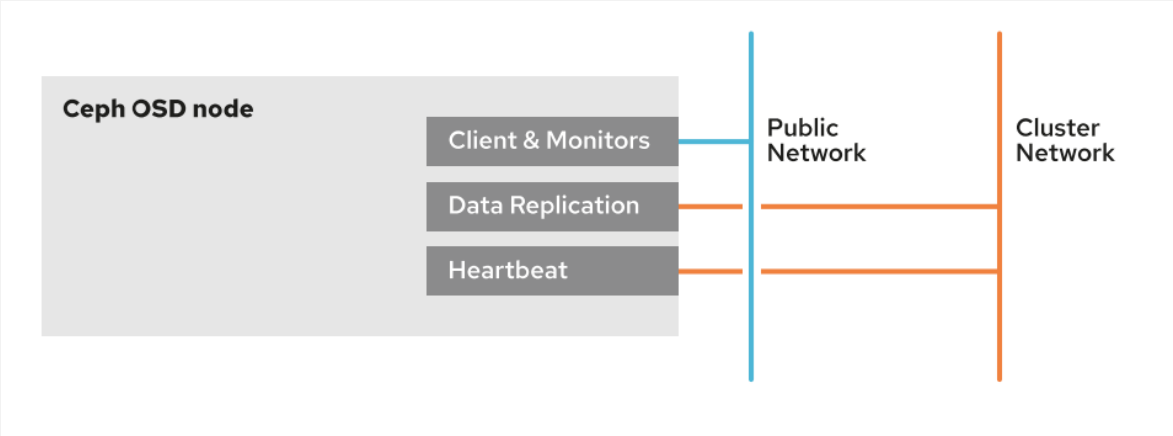
- MON 节点存储和维护集中配置数据库。数据库在每个 MON 节点上的默认位置是 /var/lib/ ceph/$fsid/mon.$host/store.db
- ceph tell mon.$id compact 压缩指定主机进程的数据库
- ceph config set mon mon_compact_on_start true mon每次启动时自动压缩

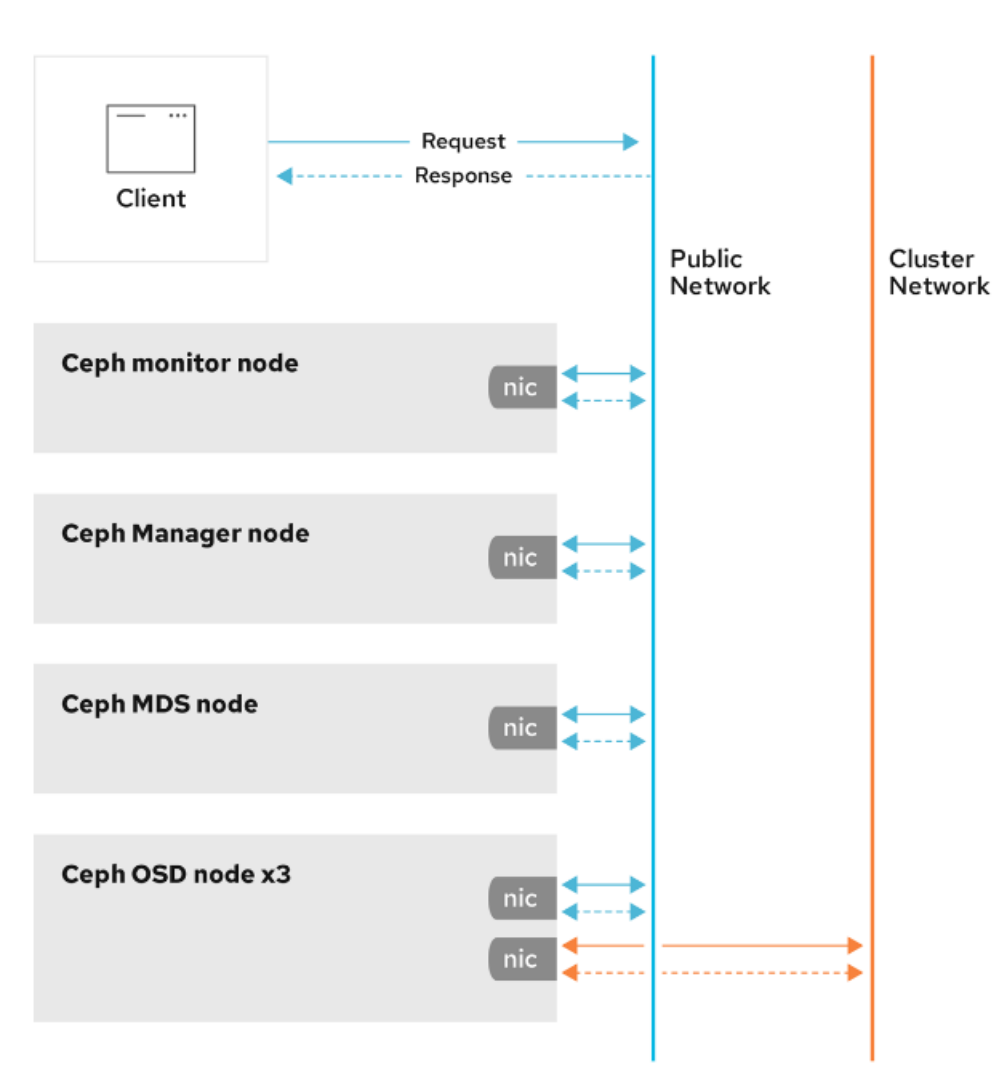
描述	设置	默认值
当配置数据库的大小超过此值时，将集群运行状况更改为 HEALTH_WARN。	mon_data_size_warn	15 (GB)
当包含配置数据库的文件系统剩余容量小于或等于此百分比时，将集群运行状况更改为 HEALTH_WARN。	mon_data_avail_warn	30 (%)
当包含配置数据库的文件系统剩余容量小于或等于此百分比时，将集群运行状况更改为 HEALTH_ERR。	mon_data_avail_crit	5 (%)

ceph 网络架构

- 集群网络类型
  - public 公共网络 供客户端和集群节点之间进行通信
  - cluster 集群网络 供OSD实现副本创建、数据恢复和再平衡以及心跳通信
- public 网络是所有 Ceph 集群通信的默认网络
  - cephadm 工具假定第一个 MON 守护进程 IP 地址 的网络是 public 网络
  - Ceph 客戶端通过集群的 public 网络直接向 OSD 发送请求
  - 如果没有cluster network OSD 复制和恢复流量会使用 public 网络

ceph 网络架构





集群网络优势

- 性能：消除副本创建、数据恢复和再平衡对 public network 的压力；增强 OSD 心跳网络的可靠性
- 安全 ：集群网络和公共网络隔离通过cluster network，防止例如 DDOS 网络攻击带来的影响
- OSD使用6800-7300范围内三个端口进行通信：
  - 一个用于通过公共网络与客戶端和 MON 通信
  - 一个用于通过集群网络或公共网络发送数据到其他 OSD
  - 一个用于通过集群网络或公共网络交换心跳数据包

配置集群网络

- 创建集群时配置
  - cephadm bootstrap --cluster_network
- 修改已经存在的集群
  - ceph config set osd cluster_network 172.25.252.0/24
- 启用IPV6
  - 默认开启ms_bind_ipv4 true 而 ms_bind_ipv6位false
  - ceph config set global ms_bind_ipv6 true

誉天教育
YU TIAN EDU

Thank you