# 案例分析

## glusterfs环境搭建

1. 准备3台机器gfs-223,gfs-224,gfs-226，操作系统为centos7。
2. 对每台机器，关闭防火墙，systemctl stop firewalld.service ，禁用防火墙，systemctl disable firewalld.service；
3. 安装glusterfs

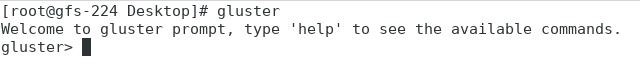
yum install -y centos-release-gluster

yum install -y glusterfs glusterfs-server glusterfs-fuse glusterfs-rdma

systemctl start glusterd.service # 启动gfs

systemctl enable glusterd.service # 开机启动gfs

安装成功如下：



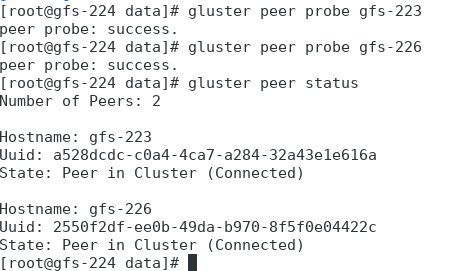
1. 将节点加入集群

在任一台机器上，以gfs-224为例：

gluster peer probe gfs-223

gluster peer probe gfs-226

创建成功如下：



1. 对每台机器，创建存储空间

mkdir –p /opt/data/ gfs

mount /dev/sdb1 /opt/data/gfs/

mkdir –p /opt/data/ gfs /bk1 # gfs brick

1. 创建卷

以创建复制卷为例：

gluster volume create gvol replica 3 gfs-224:/opt/data/gfs/bk1 gfs-223:/opt/data/gfs/bk1 gfs-226:/opt/data/gfs/bk1

成功如下：

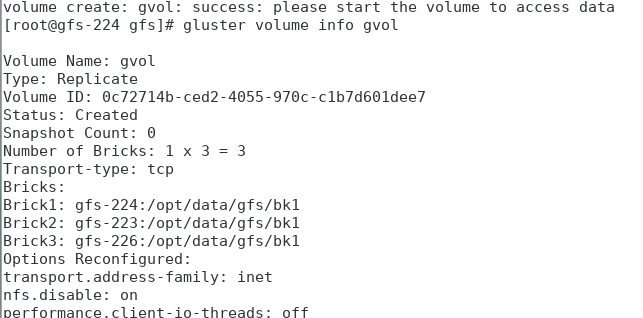


1. 运行卷

查看卷信息

gluster volume info gvol

信息如下：



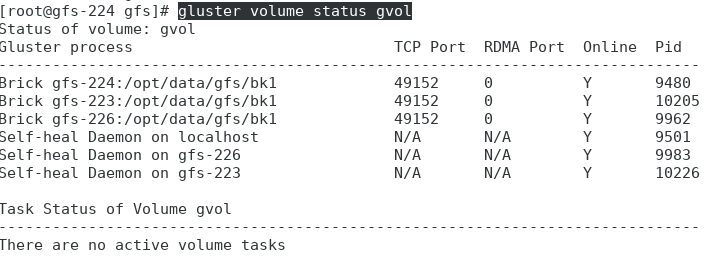
运行卷

gluster volume start gvol



查看卷状态

gluster volume status gvol



1. 挂载客户端

任一机器上

mkdir –p /mnt/gfs\_cli

mount -t glusterfs gfs-224:/gvol /mnt/gfs\_cli/

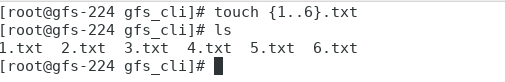


1. 数据写入

客户端：

cd /mnt/gfs\_cli

touch {1..6}.txt



gfs-224:



gfs-223:



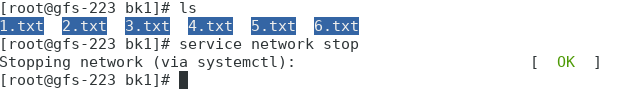
gfs-226:



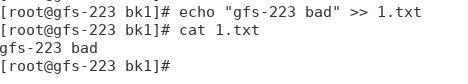
## Split-brain模拟（不通过客户端写入）

gfs-223机器上：

断开网络：service network stop

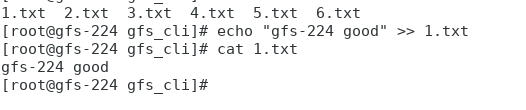


写入数据：cd /opt/data/gfs/bk1 , echo "gfs-223 bad" >> 1.txt

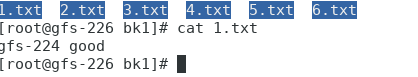


gfs-224机器上：

客户端 /mnt/gfs\_cli写入数据: echo "gfs-224 good" >> 1.txt

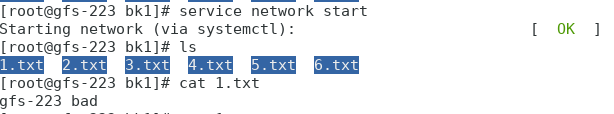


正常状态的gfs-226机器上：



此时处于connected状态的gfs-226,gfs-224文件内容一致。

恢复gfs-223的网络：service network start



gfs-223与 gfs-226,gfs-224文件1.txt内容不一致。

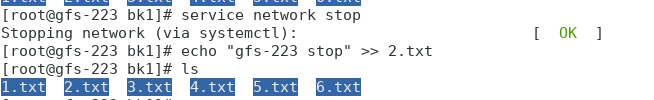
分析：

通过/opt/data/gfs/bk1/1.txt 直接写入不会再集群中，只会在本地存有。

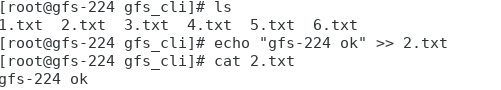
## Split-brain模拟（通过客户端写入）

gfs-224断开网络：service network stop

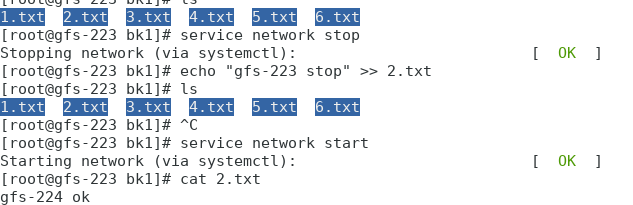
写入文件：echo "gfs-223 stop" >> 2.txt



gfs-224 客户端：写入文件，echo "gfs-224 ok" >> 2.txt



恢复gfs-223的网络：



gfs-223上文件2.txt与集群一致。

总结：gfs文件写入集群需要通过挂载客户端写入。直接通过gfs的brick写入会出现问题。

## gfs常见问题分析与解决

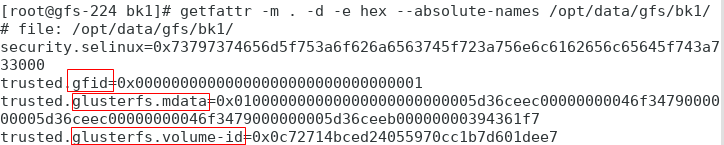
1. p：peer probe: failed: Probe returned with Transport endpoint is not connected

A: gfs peer的防火墙没关闭

2) p: /opt/data/gfs/bk1 is already part of a volume

A: getfattr -m . -d -e hex --absolute-names /opt/data/gfs/bk1

如下：



通过setfattr将该目录的gfs信息清除，然后使用。

setfattr -x trusted.glusterfs.mdata /opt/data/gfs/bk1

setfattr -x trusted.glusterfs.volume-id /opt/data/gfs/bk1

setfattr -x trusted.gfid /opt/data/gfs/bk1

1. p: peer reject

A: 删除/var/lib/glusterd/目录下除glusterd.info文件的其他文件，然后重启gluster再执行

1. p: 集群中的节点异常不能connected， peer probe <gfs-223>失败

A: 确认peer gfs-223环境gluster正常。

gluster peer status 找到gfs-223的uuid

将/var/lib/glusterd/peers/目录下，gfs-223的uuid相关文件删除。

重启gluster

5） p: 磁盘挂载的目录不能直接作为gfs的brick

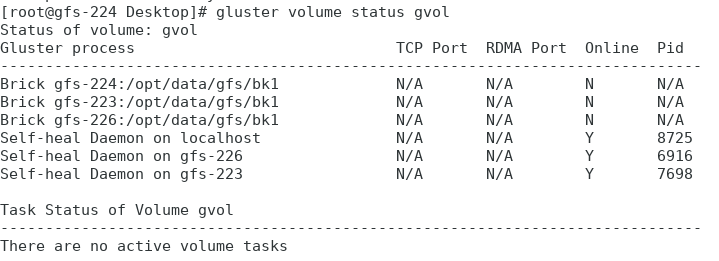
A: mount /dev/sdb1 /opt/data/gfs 则/opt/data/gfs不能作为brick

需要在改目录下新建子目录如bk1，将/opt/data/gfs/bk1作为brick加入集群。

6） p: gfs某个peer异常关机，开机后重新加入gluster集群，该机器上的brick处于N/A状态。重启：gluster volume start <volumne name>之后该状态还是N/A.

A: 重新创建新的brick bk2, mkdir –p /opt/data/gfs/bk2。替换之前异常的bk1

gluster volume replace-brick gvol gfs-223:/opt/data/gfs/bk1 gfs-223:/opt/data/gfs1/bk1 commit force



# 小结

本文只涉及到gfs基本使用与模拟split-brain产生及分析。通过自建环境使用gfs过程中发现的一些问题及找到一个可行的解决方案。需要深入了解的可以参考以下链接：

<https://blog.csdn.net/liuaigui/article/details/6284551>

<https://docs.gluster.org/en/latest/>