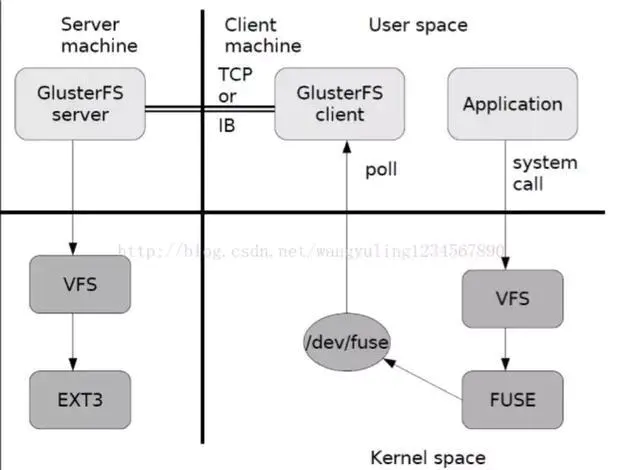
#### 一、Glusterfs集群的五种模式



附件1

如果创建的brick在系统盘，在gluster默认情况下是不允许的，如果需要使用在创建卷时在后面加上force

##### 1.分布卷

文件分布在不同的块服务器（文件1可分布在服务器1或2，但不能两台同时存在，没有冗余），将完整的数据哈希成两份，分布存在于两个brick，以文件为单位的哈希分配

更容易和廉价的扩展卷的大小

单独故障会造成数据丢失

eg:

gluster volume create new-volume server1:/data/brick1/exp2 server2:/data/brick1/exp2

[root@lvs.17.sg1.test disperse]# pwd

/data0/disperse

[root@lvs.17.sg1.test disperse]# touch testfile{1,2,3,4,5,6,7,8,9,10}

[root@lvs.17.sg1.test disperse]# ls

testfile1  testfile10  testfile2  testfile3  testfile4  testfile5  testfile6          testfile7  testfile8  testfile9

我们去192.168.21.18和19的对应的brick上看分别存储的数据。

19：

[root@lvs ~]# ls /data0/gluster/data1/

testfile1  testfile10  testfile2  testfile3  testfile4  testfile6  testfile8

18：

[root@lvs ~]# ls /data0/gluster/data1/

testfile5  testfile7  testfile9

##### 2.复制卷

类似于raid1，可以进行三复制或两复制，所有的块服务器都保持一个精确的副本，即使一个块损坏，仍可以从它的复制块访问数据，具有更高的可靠性和冗余性

eg:

gluster volume create test-volume3 replica 2 transport tcp server1:/data/brick1/test1/ server2:/data/brick1/test1/

gluster volume create test-volume3 replica 3 transport tcp server1:/data/brick1/test1/ server2:/data/brick1/test1/ server3:/data/brick1/test1/

##### 3.分布复制卷

最少需要四台服务器才能创建，块服务器必须是复制的倍数，彼此相邻的块服务器成为彼此的复制

eg：

gluster volume create test-volume replica 2 transport tcp server1:/exp1 server2:/exp2 server3:/exp3 server4:/exp4

18节点：

[root@lvs ~]# ls /data0/gluster/data\_rd\_1/

testfile5  testfile7  testfile9

[root@lvs ~]# ls /data0/gluster/data\_rd\_2

root.tar.gz  testfile1  testfile10  testfile2  testfile3  testfile4  testfile6  testfile8

19节点：

[root@lvs ~]# ls /data0/gluster/data\_rd\_1/

testfile5  testfile7  testfile9

[root@lvs ~]# ls /data0/gluster/data\_rd\_2

root.tar.gz  testfile1  testfile10  testfile2  testfile3  testfile4  testfile6  testfile8

可以看到同一个节点下的两个brick构成哈希卷，文件哈希分配到两个brick上，而两个节点对应的brick上都是对方数据的完成副本，这个应该相当于raid10吧。

##### 4.条带卷

数据被分割成更小块分布到快服务器群中的不同条带区

分布减少了负载且更小的文件加速了存取的速度

没有数据冗余（仅在高并发环境中访问非常大的文件时才使用）

条带卷将一个文件分成两个部分存储在不同的brick下，同时拆分数据，没有冗余，不能坏盘，任何一个节点出现问题，所有文件都不能访问

eg:

gluster volume create test-volume3 stripe 2 transport tcp server1:/data/brick1/test1/ server2:/data/brick1/test1/

[root@lvs.17.sg1.test stripe]# pwd

/data0/stripe

[root@lvs.17.sg1.test stripe]# du -sh /root/root.tar.gz

140M    /root/root.tar.gz

[root@lvs.17.sg1.test stripe]# cp /root/root.tar.gz  .

在192.168.21.17上有一个140M的文件，复制到挂载了条待卷的目录下

18节点：

[root@lvs ~]# ls /data0/gluster/data3/

root.tar.gz

[root@lvs ~]# du -sh /data0/gluster/data3/root.tar.gz

70M     /data0/gluster/data3/root.tar.gz

19节点：

[root@lvs ~]# du -sh /data0/gluster/data3/root.tar.gz

70M     /data0/gluster/data3/root.tar.gz

#### 二、glusterfs服务端部署

初装配置

|  |  |  |
| --- | --- | --- |
| 主机名 | Ip地址 | 盘符 |
| gluster-node1 | 192.168.189.131 | /dev/sdb |
| gluster-node2 | 192.168.189.132 | /dev/sdb |
| localhost | 192.168.189.150 |  |

##### 1.安装常见依赖包

yum install -y vim net-tools ntpdate

##### 2.配置yum源，安装glusterfs相关包

yum install epel-release -y

yum install -y centos-release-gluster6.noarch

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

##### 3.启动glusterd服务并设置开机自启

systemctl start glusterd

systemctl enable glusterd

##### 4.设置主机名并写入hosts文件

hostnamectl set-hostname glusterfs-node1

su

hostnamectl set-hostname glusterfs-node2

su

[root@glusterfs-node1 ~]# cat /etc/hosts

127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4

::1 localhost localhost.localdomain localhost6 localhost6.localdomain6

192.168.189.131 glusterfs-node1

192.168.189.132 glusterfs-node2

scp -rp /etc/hosts glusterfs-node2:/etc/

##### 5.两台分别关闭防火墙

systemctl stop firewalld

systemctl disable firewalld

setenforce 0

sed -i 's/enforcing/disabled/g' /etc/selinux/config

##### 6.配置ssh互信

[root@glusterfs-node1 ~]# ssh-keygen

[root@glusterfs-node1 ~]# ssh-copy-id glusterfs-node2

[root@glusterfs-node2 ~]# ssh-keygen

[root@glusterfs-node2 ~]# ssh-copy-id glusterfs-node1

##### 7.配置时间同步

###### 7.1更改当前时区

[root@glusterfs-node1 ~]# cp /usr/share/zoneinfo/Asia/Shanghai /etc/localtime

[root@glusterfs-node1 ~]# ntpdate cn.pool.ntp.org

[root@glusterfs-node2 ~]# cp /usr/share/zoneinfo/Asia/Shanghai /etc/localtime

[root@glusterfs-node2 ~]# ntpdate cn.pool.ntp.org

###### 7.2 配置时间同步服务端

修改glusterfs-node1的/etc/’chrony.conf文件如下：

[root@glusterfs-node1 ~]# grep -v "#" /etc/chrony.conf

server cn.pool.ntp.org iburst

driftfile /var/lib/chrony/drift

makestep 1.0 3

rtcsync

allow 192.168.189.0/24

local stratum 10

logdir /var/log/chrony

###### 7.3 配置时间同步客户端

修改glusterfs-node2的/etc/’chrony.conf文件如下：

[root@glusterfs-node2 ~]# grep -v "#" /etc/chrony.conf

server 192.168.189.131 iburst

driftfile /var/lib/chrony/drift

makestep 1.0 3

rtcsync

logdir /var/log/chrony

###### 7.4 两台分别重启chrony服务并设置开机自启

Systemctl restart chronyd

Systemctl enable chronyd

###### 7.5 查看当前时间同步状态

[root@glusterfs-node1 ~]# chronyc sources -v

[root@glusterfs-node2 ~]# chronyc sources -v

##### 8. 添加节点

###### 8.1 节点一

[root@glusterfs-node1 ~]# gluster peer probe glusterfs-node2

peer probe: success.

[root@glusterfs-node1 ~]# gluster peer status

Number of Peers: 1

Hostname: glusterfs-node2

Uuid: 1131bc82-350e-4da0-9ec9-a4f25f768dc8

State: Peer in Cluster (Connected)

###### 8.2 节点二

[root@glusterfs-node2 ~]# gluster peer probe glusterfs-node1

##### 9. 创建卷，双复制卷

gluster v create fuzhi transport tcp replica 2 glusterfs-node1:/gluster/brick1/ glusterfs-node2:/gluster/brick1/

启动卷

gluster v start fuzhi

查看卷组状态

gluster v status

##### 10. 将盘符写入/etc/fstab

必须将gluster的brick所在的盘符设置为开机自动挂载

#### 三、glusterfs客户端安装并挂载

##### 1. 安装客户端相关包

yum -y install glusterfs glusterfs-fuse glusterfs-cli glusterfs-libs glusterfs-client-xlator

##### 2. 挂载卷组

**未配置hosts文件时，挂载会报错**

[root@localhost ~]# mount.glusterfs glusterfs-node1:/fuzhi /media/

Mount failed. Please check the log file for more details.

**客户端必须配置hosts文件才可以挂载，该版本如下（不知道是该版本问题还是glusterfs通有问题）：**

[root@localhost ~]# rpm -qa | grep gluster

glusterfs-libs-6.10-1.el7.x86\_64

glusterfs-client-xlators-6.10-1.el7.x86\_64

glusterfs-cli-6.10-1.el7.x86\_64

centos-release-gluster6-1.0-1.el7.centos.noarch

glusterfs-6.10-1.el7.x86\_64

glusterfs-fuse-6.10-1.el7.x86\_64

[root@localhost ~]# cat /etc/hosts

127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4

::1 localhost localhost.localdomain localhost6 localhost6.localdomain6

192.168.189.131 glusterfs-node1

192.168.189.132 glusterfs-node2

配置hosts后再次执行命令挂载正常

[root@localhost ~]# mount.glusterfs glusterfs-node1:/fuzhi /media/

[root@localhost ~]# df -h

Filesystem Size Used Avail Use% Mounted on

devtmpfs 475M 0 475M 0% /dev

tmpfs 487M 0 487M 0% /dev/shm

tmpfs 487M 7.7M 479M 2% /run

tmpfs 487M 0 487M 0% /sys/fs/cgroup

/dev/mapper/centos-root 17G 1.8G 16G 11% /

/dev/sda1 1014M 137M 878M 14% /boot

tmpfs 98M 0 98M 0% /run/user/0

glusterfs-node1:/fuzhi 5.0G 84M 5.0G 2% /media

#### 四、glusterfs常用命令

##### 1. 启停/查看glusterd服务

systemctl start glusterd

systemctl stop glusterd

systemctl status glusterd

##### 2. 开机自启glusterd服务

systemctl enable glusterd

systemctl list-unit-files | grep glusterd

##### 3. 为存储池添加/移除服务器节点

gluster peer probe gluster2 添加服务器节点

gluster peer detach gluster2 移除服务器节点 移除节点前必须将该节点上的brick删除

gluster peer status 查看所有节点的基本状态

##### 创建/启动/停止/删除卷

gluster volume start 卷名<volname>

gluster volume stop 卷名<volname>

gluster volume delete 卷名<volname> 删除卷之前必须先停止卷，最后可清空brick server节点对应目录下的内容

rm -f /brick1/dis\_volume

##### 删除brick

gluster volume remove-brick volume-name server1:/PATH server2:/PATH start

gluster volume remove-brick volume-name server1:/PATH server2:/PATH commit

gluster volume remove-brick volume-name server1:/PATH server2:/PATH status

如果无法删除，执行如下命令,将gluster节点的配置文件删除

rm -rf /var/lib/glusterd/peers/\*

##将gluster配置全部删除重新配置

rm -rf /var/lib/glusterd

gluster peer status

**格式化要做gluster的盘符**

**mkfs.xfs /dev/sdb**

**创建挂载点，mkdir -p /data**

**将盘符挂载 mount /dev/sdb /data**

**在/etc/fstab里加入**

**/dev/sdb的UUID /data xfs defaults 0 0**

##### 客户端挂载gluster

1. 客户端以glusterfs方式挂载

创建挂载点

mkdir /datavol

mount -t glusterfs gluster1:/datavol /datavol

mount -t glusterfs <server>:/<volume> <mountdir>

节点:/卷名 挂载点

1. 客户端以nfs方式挂载

glusterfs默认没有打开nfs挂载方式，需要在服务端打开

服务端：

开放nfs

gluster volume set fuzhi nfs.disable off

卷名

gluster volume stop fuzhi

gluster volume start fuzhi

安装nfs-ganesha包

yum install -y nfs-ganesha nfs-ganesha-gluster

修改配置文件

[root@glusterfs-node1 /]# grep -Ev "#|^$" /etc/ganesha/ganesha.conf

EXPORT

{

Export\_Id=1 ;

Path = "/glu"; //指定nfs共享目录的位置，客户都挂载使用的就是服务端的该目录

Pseudo = "/glu-pseudo";

Disable\_ACL =True;

Protocols = "3","4";

Access\_Type = RW;

Squash = No\_root\_squash;

Sectype="sys";

Transports = "UDP","TCP";

FSAL { //定义的是准备导出的gluster的volume

Name = "GLUSTER"; //是应该导出卷的卷格式GLUSTER

Hostname="glusterfs-node1"; //是主机名 ，不同主机是不同的

Volume = "fuzhi"; //是gluster的卷名

}

}

重启nfs-ganesha服务

systemctl restart nfs-ganesha

客户端：

yum install -y nfs-utils

mkdir /nfs

查看nfs共享目录

showmount -e 192.168.189.131

开始挂载

mount.nfs 192.168.189.131:/glu /nfs/

##### 查看卷信息

* 1. 列出集群中的所有卷

gluster volume list

* 1. 查看集群中的卷信息

gluster volume info

* 1. 查看集群中的卷状态

gluster volume status

gluster volume status <volname>

1. 扩展卷

在gluster2上添加两块磁盘

先格式化mkfs.xfs /dev/sdc mkfs.xfs /dev/sdd

创建挂载点mkdir -p /data1 /data2

mount /dev/sdc /data1

mount /dev/sdd /data2

将挂载信息写入/etc/fstab文件中

云桌面三加一模式，第四个服务器可以使用如下方法添加至gluster集群里，添加的brick必须是replica的整数倍

gluster volume add-brick datavol gluster2:/data1/brick1 gluster2:/data2/brick1 force

**brick**默认不能在同一服务器上，如果处于同一服务器后面必须加上force参数才可以添加成功

##### brick管理

1、添加brick

gluster volume add-brick volume-name server1:/PATH server2:/PATH

2、移除brick

gluster volume remove-brick VOLNAME BRICK [start | status | commit]

3、替换brick

gluster volume replace-brick VOLNAME BRICKNEW-BRICK [start | pause | sbortstatus | commit]

2、手动同步数据

gluster volume volume-name reblance status

gluster volume volume-name reblance start

————————————————

##### 创建仲裁卷

如果是副本为3的仲裁器卷，其中第三个brick 充当仲裁器brick。 该配置具有防止发生裂脑的机制。

你可以使用下面的命令创建它：

# gluster volume create <VOLNAME> replica 3 arbiter 1 host1:brick1 host2:brick2 host3:brick3`

##### 重命名glusterfs集群名

要将群集 OpsClust 重命名为 OpsClust1，请键入：

**cluster opsclust /rename:opsclust1**

##### 防止脑裂

gluster volume set <volname> cluster.server-quorum-type none/server

 #gluster volume set all cluster.server-quorum-ratio <percentage%>

gluster volume set ksvd\_vol cluster.quorum-type auto

#### 五、常见故障处理

##### 1. 从一个brick迁移到另外一个brick

gluster volume replace-brick volume source-brick new-brick commit force

new-brick未被其他gluster使用并且和源brick大小一致

eg: gluster volume replace-brick replica2 192.168.189.132:/gluster1/brick1 192.168.189.132:/gluster/brick1 commit force

迁移完成后，源brick就不会再存储新写入的数据

##### 2. 双复制一个节点的磁盘损坏

节点：gluster1、gluster2

gluster1节点的brick硬盘损坏

移除gluster1的brick

1）初始状态

[root@gluster2 ~]# gluster volume status

Status of volume: fuzhi

Gluster process TCP Port RDMA Port Online Pid

------------------------------------------------------------------------------

Brick gluster1:/data/brick1 49153 0 Y 3047

Brick gluster2:/data/brick1 49155 0 Y 1617

Self-heal Daemon on localhost N/A N/A Y 1634

Self-heal Daemon on gluster1 N/A N/A Y 3064

Task Status of Volume fuzhi

------------------------------------------------------------------------------

There are no active volume tasks

2）移除brick

[root@gluster2 ~]# gluster volume remove-brick fuzhi replica 1 gluster2:/data/brick1/ force

3）查看当前状态

[root@gluster2 ~]# gluster volume status

Status of volume: fuzhi

Gluster process TCP Port RDMA Port Online Pid

------------------------------------------------------------------------------

Brick gluster1:/data/brick1 49153 0 Y 3047

Task Status of Volume fuzhi

------------------------------------------------------------------------------

There are no active volume tasks

4）添加新的brick

[root@gluster2 ~]# gluster volume add-brick fuzhi replica 2 gluster2:/data1/brick1/

Replica 2 volumes are prone to split-brain. Use Arbiter or Replica 3 to avoid this. See: http://docs.gluster.org/en/latest/Administrator%20Guide/Split%20brain%20and%20ways%20to%20deal%20with%20it/.

Do you still want to continue?

(y/n) y

volume add-brick: failed: /data1/brick1 is already part of a volume

brick添加失败，该brick已是卷的一部分，对于硬盘原有做过gluster，该硬盘的brick1下仍存在原有的brick信息，将原有信息删除

5）删除原有brick相关信息

[root@gluster2 ~]# rm -rf /data1/brick1/

6) 重新添加brick

[root@gluster2 data1]# gluster volume add-brick fuzhi replica 2 gluster2:/data1/brick1/

Replica 2 volumes are prone to split-brain. Use Arbiter or Replica 3 to avoid this. See: http://docs.gluster.org/en/latest/Administrator%20Guide/Split%20brain%20and%20ways%20to%20deal%20with%20it/.

Do you still want to continue?

(y/n) y

volume add-brick: success

7）查看当前状态

[root@gluster2 data1]# gluster volume status

Status of volume: fuzhi

Gluster process TCP Port RDMA Port Online Pid

------------------------------------------------------------------------------

Brick gluster1:/data/brick1 49153 0 Y 3047

Brick gluster2:/data1/brick1 49155 0 Y 2063

Self-heal Daemon on localhost N/A N/A Y 2080

Self-heal Daemon on gluster1 N/A N/A Y 3493

Task Status of Volume fuzhi

------------------------------------------------------------------------------

There are no active volume tasks

##### 3. 将二复制转为三复制

当前状态

[root@glusterfs-node2 ~]# gluster v info

Volume Name: replica2

Type: Replicate

Volume ID: 64d9b9e6-e9f1-45d4-8258-3e8bb00c3f00

Status: Started

Snapshot Count: 0

Number of Bricks: 1 x 2 = 2

Transport-type: tcp

Bricks:

Brick1: glusterfs-node1:/gluster/brick1

Brick2: glusterfs-node2:/gluster/brick1

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

performance.client-io-threads: off

节点glusterfs-node2新增一块和已存在的两块brick大小一致的硬盘

Vmware热添加硬盘

echo '- - -' >/sys/class/scsi\_host/host0/scan

echo '- - -' >/sys/class/scsi\_host/host1/scan

echo '- - -' >/sys/class/scsi\_host/host2/scan

lsblk查看，多了一块sdc硬盘

[root@glusterfs-node2 ~]# lsblk

NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT

sda 8:0 0 20G 0 disk

├─sda1 8:1 0 1G 0 part /boot

└─sda2 8:2 0 19G 0 part

├─centos-root 253:0 0 17G 0 lvm /

└─centos-swap 253:1 0 2G 0 lvm [SWAP]

sdb 8:16 0 5G 0 disk /gluster

sdc 8:32 0 5G 0 disk

sr0 11:0 1 973M 0 rom

进行格式化并挂载

mkfs.xfs /dev/sdc

mkdir /gluster1

mount /dev/sdc /gluster1

mkdir -p /gluster1/brick1

添加brick到原有的二复制卷，并由二复制更改为三复制

gluster v add-brick replica2 replica 3 192.168.189.132:/gluster1/brick1/

卷名 三复制

更改完以后的gluster状态

[root@glusterfs-node2 ~]# gluster v info

Volume Name: replica2

Type: Replicate

Volume ID: 64d9b9e6-e9f1-45d4-8258-3e8bb00c3f00

Status: Started

Snapshot Count: 0

Number of Bricks: 1 x 3 = 3

Transport-type: tcp

Bricks:

Brick1: glusterfs-node1:/gluster/brick1

Brick2: glusterfs-node2:/gluster/brick1

Brick3: 192.168.189.132:/gluster1/brick1

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

performance.client-io-threads: off

##### 4. 将三复制转为二复制

[root@glusterfs-node2 ~]# gluster v info

Volume Name: replica2

Type: Replicate

Volume ID: 64d9b9e6-e9f1-45d4-8258-3e8bb00c3f00

Status: Started

Snapshot Count: 0

Number of Bricks: 1 x 3 = 3

Transport-type: tcp

Bricks:

Brick1: glusterfs-node1:/gluster/brick1

Brick2: glusterfs-node2:/gluster/brick1

Brick3: 192.168.189.132:/gluster1/brick1

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

performance.client-io-threads: off

移除一个brick并从三复制更改为二复制

gluster v remove-brick replica2 replica 2 192.168.189.132:/gluster1/brick1 force

卷名 二复制

更改以后的状态

[root@glusterfs-node2 ~]# gluster v info

Volume Name: replica2

Type: Replicate

Volume ID: 64d9b9e6-e9f1-45d4-8258-3e8bb00c3f00

Status: Started

Snapshot Count: 0

Number of Bricks: 1 x 2 = 2

Transport-type: tcp

Bricks:

Brick1: glusterfs-node1:/gluster/brick1

Brick2: glusterfs-node2:/gluster/brick1

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

performance.client-io-threads: off

##### 5. 更改卷模式（从复制卷更改为分布式卷）

原有模式为二复制

[root@glusterfs-node2 brick1]# gluster v info

Volume Name: replica2

Type: Replicate

Volume ID: 64d9b9e6-e9f1-45d4-8258-3e8bb00c3f00

Status: Started

Snapshot Count: 0

Number of Bricks: 1 x 2 = 2

Transport-type: tcp

Bricks:

Brick1: glusterfs-node1:/gluster/brick1

Brick2: 192.168.189.132:/gluster/brick1

Options Reconfigured:

performance.client-io-threads: off

nfs.disable: on

transport.address-family: inet

删除其中一个brick

gluster volume remove-brick replica2 replica 1 glusterfs-node2:/gluster/brick1 force

停止卷volume

gluster volume stop replica2

gluster volume delete replica2

再创建volume

gluster v create replica2 glusterfs-node1:/gluster/brick1/ glusterfs-node2:/gluster/brick1/ force

更改以后的模式为分布式

[root@glusterfs-node2 brick1]# gluster v info

Volume Name: replica2

Type: Distribute

Volume ID: 003cc5fc-d753-4666-8222-b10d12c865cf

Status: Created

Snapshot Count: 0

Number of Bricks: 2

Transport-type: tcp

Bricks:

Brick1: glusterfs-node1:/gluster/brick1

Brick2: glusterfs-node2:/gluster/brick1

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

启动volume

gluster v start replica2

更改为分布式后，所有

##### 6. 对volume扩容

1. 初始状态

[root@gluster2 data1]# gluster volume status

Status of volume: fuzhi

Gluster process TCP Port RDMA Port Online Pid

------------------------------------------------------------------------------

Brick gluster1:/data/brick1 49153 0 Y 3047

Brick gluster2:/data1/brick1 49155 0 Y 2063

Self-heal Daemon on localhost N/A N/A Y 2080

Self-heal Daemon on gluster1 N/A N/A Y 3493

Task Status of Volume fuzhi

------------------------------------------------------------------------------

There are no active volume tasks

1. 扩展volume

[root@gluster2 data1]# gluster volume add-brick fuzhi replica 2 gluster2:/data/brick1 gluster2:/data2/brick1 force

volume add-brick: success

force：因为两个brick属于同一个节点，所有必须加上force

1. 查看当前状态

[root@gluster2 data1]# gluster volume info

Volume Name: fuzhi

Type: Distributed-Replicate

Volume ID: e1ccb13f-8aad-4aaf-897f-2c9915b70686

Status: Started

Snapshot Count: 0

Number of Bricks: 2 x 2 = 4

Transport-type: tcp

Bricks:

Brick1: gluster1:/data/brick1

Brick2: gluster2:/data1/brick1

Brick3: gluster2:/data/brick1

Brick4: gluster2:/data2/brick1

Options Reconfigured:

cluster.granular-entry-heal: on

storage.fips-mode-rchecksum: on

transport.address-family: inet

nfs.disable: on

performance.client-io-threads: off

[root@gluster2 data1]# gluster volume status

Status of volume: fuzhi

Gluster process TCP Port RDMA Port Online Pid

------------------------------------------------------------------------------

Brick gluster1:/data/brick1 49153 0 Y 3047

Brick gluster2:/data1/brick1 49155 0 Y 2063

Brick gluster2:/data/brick1 49156 0 Y 2489

Brick gluster2:/data2/brick1 49158 0 Y 2505

Self-heal Daemon on localhost N/A N/A Y 2080

Self-heal Daemon on gluster1 N/A N/A Y 3493

Task Status of Volume fuzhi

------------------------------------------------------------------------------

There are no active volume tasks

##### 7. 重新均衡卷

不迁移数据：

# gluster volume rebalance <VOLNAME> lay-outstart

# gluster volume rebalance <VOLNAME> start

# gluster volume rebalance <VOLNAME> startforce

# gluster volume rebalance <VOLNAME> status

# gluster volume rebalance <VOLNAME> stop

##### 8. 系统扩展维护

开启/关闭系统配额：

# gluster volume quota <VOLNAME> enable | disable

设置目录配额：

# gluster volume quota <VOLNAME> limit-usage <DIR> <VALUE>

查看配额：

# gluster volume quota <VOLNAME> list [<DIR>]

地域复制（geo-replication）：

# gluster volume geo-replication <MASTER> <SLAVE> start | status | stop

IO信息查看：

# gluster volume profile <VOLNAME> start | info | stop

Top监控：

Top命令允许你查看Brick的性能，例如：read,write, file open calls, file read calls, file write calls, directory opencalls, and directory real calls。所有的查看都可以设置 top数，默认100。

查看打开的 fd：

# gluster volume top <VOLNAME> open[brick <BRICK>] [list-cnt <COUNT>]

其中，open可以替换为read, write, opendir, readdir等。

查看每个 Brick 的读性能：

# gluster volume top <VOLNAME> read-perf [bs <BLOCK-SIZE> count <COUNT>] [brick <BRICK>] [list-cnt <COUNT>]

其中，read-perf可以替换为write-perf等。

##### 9. GlusterFS优化

# 设置 cache 大小, 默认32MB

gluster volume set senyintvolume performance.cache-size 4GB

# 设置 io 线程, 太大会导致进程崩溃

gluster volume set senyintvolume performance.io-thread-count 16

# 设置 网络检测时间, 默认42s

gluster volume set senyintvolume network.ping-timeout 10

# 设置 写缓冲区的大小, 默认1M

gluster volume set senyintvolume performance.write-behind-window-size 1024MB

# 开启 指定 volume 的配额，不使用

gluster volume quota k8s-volume enable

# 限制 指定 volume 的配额，不使用

gluster volume quota k8s-volume limit-usage / 1TB

##### 常见问题

创建gluster卷时报错

volume create: replica2: failed: /gluster/brick1 is already part of a volume

各节点都需要执行

删除对应brick下的.glusterfs目录

setfattr -x trusted.glusterfs.volume-id brick

setfattr -x trusted.gfid brick

再次执行创建操作即可

#### 六、glusterfs相关日志

相关日志，在/var/log/glusterfs/目录下，可根据需要查看；

如/var/log/glusterfs/brick/下是各brick创建的日志；

如/var/log/glusterfs/cmd\_history.log是命令执行记录日志；

如/var/log/glusterfs/glusterd.log是glusterd守护进程日志。