





https://c.lanmit.com/czxt/Linux/5451.html

node1: 192.168.1.1

node2: 192.168.1.2

node3: 192.168.1.3

node4: 192.168.1.4

client: 192.168.1.5

一、部署前准备：

1、在所有节点执行以下操作：根据上表中添加磁盘，通过fdisk命令进行分区，mkfs格式化，创建相应的挂载目录，并将格式化的磁盘挂载到相应的目录中，最后修改/etc/fstab配置文件，使其永久挂载，具体配置可以参考我之前的博文：centos 7.3创建、挂载及卸载（包括自动挂载）文件系统。（我这里主要目的是做个相关笔记，所以磁盘大小不以实际环境为准，可以根据自己的实际环境进行磁盘分区）

2、自己配置防火墙和selinux，我这里为了方便，直接关闭了。

3、下载我提供的本地yum仓库，并上传至各个节点服务器。

二、开始部署：

1、node1配置如下：

[root@node1 ~]# vim /etc/hosts #写入末尾四行，以便添加4个节点的解析

127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4

::1 localhost localhost.localdomain localhost6 localhost6.localdomain6

192.168.1.1 node1

192.168.1.2 node2

192.168.1.3 node3

192.168.1.4 node4

[root@node1 ~]# mount /dev/cdrom /media #挂载我提供的yum仓库

mount: /dev/sr0 写保护，将以只读方式挂载

[root@node1 ~]# rm -rf /etc/yum.repos.d/\* #删除或移走原有yum配置文件

[root@node1 ~]# yum clean all #清除yum缓存

[root@node1 ~]# vim /etc/yum.repos.d/a.repo #编辑yum配置文件，写入以下内容

[fd]

baseurl=file:///media

gpgcheck=0

#写完以上三行，保存退出即可。

[root@node1 ~]# yum -y install glusterfs glusterfs-server glusterfs-fuse glusterfs-rdma

#安装GlusterFS软件

[root@node1 ~]# systemctl start glusterd #启动该服务

[root@node1 ~]# systemctl enable glusterd #设置为开机自启动

2、node2配置如下：

[root@node2 ~]# scp root@192.168.1.1:/etc/hosts /etc/ #将node1的hosts配置文件复制过来

The authenticity of host 192.168.1.1 (192.168.1.1)cat be established.

ECDSA key fingerprint is SHA256:BS+lKMN05pYF3F1XeIYU69VnHjzKBiBiMZ1SDKgsxxs.

ECDSA key fingerprint is MD5:ba:0b:a7:47:55:01:6f:41:41:5f:ee:b8:88:bf:7a:60.

Are you sure you want to continue connecting (yes/no)? yes #输入“yes”

Warning: Permanently added '192.168.1.1' (ECDSA) to the list of known hosts.

root@192.168.1.1s password: #输入对方用户名的密码

hosts 100% 230 286.9KB/s 00:00

[root@node1 ~]# rm -rf /etc/yum.repos.d/\* #删除或移走原有yum配置文件

[root@node1 ~]# yum clean all #清除yum缓存

[root@node2 ~]# scp root@192.168.1.1:/etc/yum.repos.d/a.repo /etc/yum.repos.d/

#将node1的yum文件复制过来

root@192.168.1.1s password:

a.repo 100% 38 31.1KB/s 00:00

[root@node2 ~]# mount /dev/cdrom /media #挂载我提供的yum仓库

mount: /dev/sr0 写保护，将以只读方式挂载

[root@node1 ~]# yum -y install glusterfs glusterfs-server glusterfs-fuse glusterfs-rdma

#安装GlusterFS软件

[root@node1 ~]# systemctl start glusterd #启动该服务

[root@node1 ~]# systemctl enable glusterd #设置为开机自启动

至此，node2配置完成了，node3、node4重复node2的配置即可，我这就不写了，请自行配置。

3、添加节点（接下来的所有配置，在任意一台节点上执行即可，我这里在node1上执行）：

[root@node1 ~]# gluster peer probe node1 #添加node1，所以提示可以不用添加

peer probe: success. Probe on localhost not needed

[root@node1 ~]# gluster peer probe node2 #添加node2

peer probe: success.

[root@node1 ~]# gluster peer probe node3 #添加node3

peer probe: success.

[root@node1 ~]# gluster peer probe node4 #添加node4

[root@node1 ~]# gluster peer status #查看群集状态

Number of Peers: 3

Hostname: node2

Uuid: d733aa7c-5078-43b2-9e74-6673f3aaa16e

State: Peer in Cluster (Connected) #如果某个节点显示Disconnected，请检查hosts配置文件

Hostname: node3

Uuid: dc64b6c6-ce2d-41d3-b78b-56f46038ab52

State: Peer in Cluster (Connected)

Hostname: node4

Uuid: 926b51e9-4599-4fe8-ad2b-11f53a2ffb5a

State: Peer in Cluster (Connected)

4、创建各种类型的卷

（1）创建分布式卷：

[root@node1 ~]# gluster volume create dis-volume node1:/e6 node2:/e6 force

#创建分布式卷，其中“dis-volume”为卷名，没有指定类型，默认创建的是分布式卷。

volume create: dis-volume: success: please start the volume to access data

[root@node1 ~]# gluster volume info dis-volume #查看该卷相关信息

Volume Name: dis-volume

Type: Distribute

Volume ID: 2552ea18-b8f4-4a28-b411-a5b1bd168009

Status: Created

Snapshot Count: 0

Number of Bricks: 2

Transport-type: tcp

Bricks:

Brick1: node1:/e6

Brick2: node2:/e6

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

[root@node1 ~]# gluster volume start dis-volume #启用该卷

volume start: dis-volume: success

（2）创建条带卷：

[root@node1 ~]# gluster volume create stripe-volume stripe 2 node1:/d5 node2:/d5 force

#创建条带卷，指定条带卷的个数为2个。“stripe-volume”为卷名

#指定类型为stripe，数值为2，而且后面跟了2个brick server，所以创建的是条带卷

volume create: stripe-volume: success: please start the volume to access data

[root@node1 ~]# gluster volume info stripe-volume #查看该卷相关信息

Volume Name: stripe-volume

Type: Stripe #卷类型为条带

Volume ID: c38107e9-9d92-4f37-a345-92568c2c9e9a

Status: Created

Snapshot Count: 0

Number of Bricks: 1 x 2 = 2

Transport-type: tcp

Bricks:

Brick1: node1:/d5

Brick2: node2:/d5

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

[root@node1 ~]# gluster volume start stripe-volume #启用该卷

volume start: stripe-volume: success

（3）创建复制卷：

[root@node1 ~]# gluster volume create rep-volume replica 2 node3:/d5 node4:/d5 force

#指定类型为“replica”，数值为“2”，而且后面跟了两个brick server，所以创建的是复制卷

volume create: rep-volume: success: please start the volume to access data

[root@node1 ~]# gluster volume info rep-volume #查看该卷相关信息

Volume Name: rep-volume

Type: Replicate #卷类型为复制

Volume ID: 03553b49-c5fa-4a5f-8d66-8c229e617696

Status: Created

Snapshot Count: 0

Number of Bricks: 1 x 2 = 2

Transport-type: tcp

Bricks:

Brick1: node3:/d5

Brick2: node4:/d5

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

[root@node1 ~]# gluster volume start rep-volume #启用该卷

volume start: rep-volume: success

（4）创建分布式条带卷：

[root@node1 ~]# gluster volume create dis-stripe stripe 2 node1:/b3 node2:/b3 node3:/b3 node4:/b3 force

#指定类型为stripe，数值为2，而且后面跟了4个brick server，所以创建的是分布式条带卷

volume create: dis-stripe: success: please start the volume to access data

[root@node1 ~]# gluster volume info dis-stripe #查看该卷相关信息

Volume Name: dis-stripe

Type: Distributed-Stripe #卷类型为分布式+条带

Volume ID: 059ee6e3-317a-4e47-bf92-47d88e3acf3c

Status: Created

Snapshot Count: 0

Number of Bricks: 2 x 2 = 4

Transport-type: tcp

Bricks:

Brick1: node1:/b3

Brick2: node2:/b3

Brick3: node3:/b3

Brick4: node4:/b3

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

[root@node1 ~]# gluster volume start dis-stripe #启用该卷

volume start: dis-stripe: success

（5）创建分布式复制卷：

[root@node1 ~]# gluster volume create dis-rep replica 2 node1:/c4 node2:/c4 node3:/c4 node4:/c4 force

#指定类型为replica，数值为2，而且后面跟了4个brick server，是2的两倍，所以创建的是分布式复制卷

volume create: dis-rep: success: please start the volume to access data

[root@node1 ~]# gluster volume info dis-rep #查看该卷相关信息

Volume Name: dis-rep

Type: Distributed-Replicate #卷类型为分布式+复制

Volume ID: 9e702694-92c7-4a3a-88d2-dcf9ddad741c

Status: Created

Snapshot Count: 0

Number of Bricks: 2 x 2 = 4

Transport-type: tcp

Bricks:

Brick1: node1:/c4

Brick2: node2:/c4

Brick3: node3:/c4

Brick4: node4:/c4

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

[root@node1 ~]# gluster volume start dis-rep #启用该卷

volume start: dis-rep: success

OK，现在所涉及到的卷已经创建完成，可以在客户端上挂载使用了：

5、部署Gluster客户端：

（1）部署安装：

[root@node1 ~]# rm -rf /etc/yum.repos.d/\* #删除或移走原有yum配置文件

[root@node1 ~]# yum clean all #清除yum缓存

[root@node2 ~]# scp root@192.168.1.1:/etc/yum.repos.d/a.repo /etc/yum.repos.d/

#将node1的yum文件复制过来

root@192.168.1.1s password:

a.repo 100% 38 31.1KB/s 00:00

[root@node2 ~]# mount /dev/cdrom /media #挂载我提供的yum仓库

mount: /dev/sr0 写保护，将以只读方式挂载

[root@client ~]# yum -y install glusterfs glusterfs-fuse #安装客户端所需的GlusterFS软件

[root@client ~]# mkdir -p /test/{dis,stripe,rep,dis\_and\_stripe,dis\_and\_rep} #创建挂载目录

[root@client ~]# ls /test #查看挂载目录是否创建

dis dis\_and\_rep dis\_and\_stripe rep stripe

[root@client ~]# scp root@192.168.1.1:/etc/hosts /etc/

#client也需要解析节点服务器，所以将主机1.1的hosts文件复制过来

root@192.168.1.1 s password: #输入对端服务器用户密码

hosts 100% 230 0.2KB/s 00:00

（2）挂载Gluster文件系统：

[root@client ~]# mount -t glusterfs node1:dis-volume /test/dis

[root@client ~]# mount -t glusterfs node2:stripe-volume /test/stripe

[root@client ~]# mount -t glusterfs node3:rep-volume /test/rep

[root@client ~]# mount -t glusterfs node4:dis-stripe /test/dis\_and\_stripe

[root@client ~]# mount -t glusterfs node1:dis-rep /test/dis\_and\_rep

#如果挂载不成功，请检查hosts文件解析。在挂载时，可以指定逻辑存储卷内的任意主机，

#因为所有GlusterFS配置信息在节点中是共享的，这也避免了node1故障，就无法使用其他

#卷的问题。

（3）修改fstab配置文件，以便实现开机自动挂载：

[root@client ~]# vim /etc/fstab #在文件末尾写入下面几行

node2:stripe-volume /test/stripe glusterfs defaults,\_netdev 0 0

node3:rep-volume /test/rep glusterfs defaults,\_netdev 0 0

node4:dis-stripe /test/dis\_and\_stripe glusterfs defaults,\_netdev 0 0

node1:dis-rep /test/dis\_and\_rep glusterfs defaults,\_netdev 0 0

node1:dis-volume /test/dis glusterfs defaults,\_netdev 0 0

设置自动挂载时，必须当前已经手动挂载了相关目录，才可以实现自动挂载，虽然在手动挂载时，指定哪个节点都可以，但建议在写入/etc/fstab时，当前挂载时指定的哪个节点，就写哪个节点。

（三）GlusterFS维护命令：

[root@node1 ~]# gluster volume list #查看卷列表

dis-rep

dis-stripe

dis-volume

rep-volume

stripe-volume

[root@node1 ~]# gluster volume info #查看所有卷的信息

Volume Name: dis-rep

Type: Distributed-Replicate

Volume ID: 9e702694-92c7-4a3a-88d2-dcf9ddad741c

Status: Started

Snapshot Count: 0

Number of Bricks: 2 x 2 = 4

Transport-type: tcp

Bricks:

Brick1: node1:/c4

Brick2: node2:/c4

Brick3: node3:/c4

Brick4: node4:/c4

Options Reconfigured:

transport.address-family: inet

nfs.disable: on

....................... #省略部分内容

[root@node1 ~]# gluster volume status #查看卷的状态

Status of volume: dis-rep

Gluster process TCP Port RDMA Port Online Pid

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

Brick node1:/c4 49155 0 Y 11838

Brick node2:/c4 49155 0 Y 12397

Brick node3:/c4 49154 0 Y 12707

Brick node4:/c4 49154 0 Y 12978

Self-heal Daemon on localhost N/A N/A Y 11858

Self-heal Daemon on node4 N/A N/A Y 12998

Self-heal Daemon on node2 N/A N/A Y 12417

Self-heal Daemon on node3 N/A N/A Y 12728

............................

[root@node1 ~]# gluster volume stop dis-stripe #停止一个卷

Stopping volume will make its data inaccessible. Do you want to continue? (y/n) y

volume stop: dis-stripe: success

[root@node1 ~]# gluster volume delete dis-stripe #删除一个卷

Deleting volume will erase all information abe. Do you want to continue? (y/n) y

volume delete: dis-stripe: success

[root@node1 ~]# gluster volume set dis-rep auth.allow 192.168.1.\*,10.1.1.\*

#设置只允许特定网段的客户端访问卷dis-rep

volume set: success

GlusterFS是一个开源的分布式文件系统，同时也是Scale-Out存储解决方案Gluster的核心，在存储数据方面有强大的横向扩展能力。GlusterFS主要由存储服务器、客户端及NFS/Samba存储网关（可选组件）组成。GlusterFS架构中最大的设计特点是没有元数据服务器组件，也就是说没有主/从服务器之分，每一个节点都可以是主服务器。

1）Gluster相关参考文档如下（我下面的配置是基于本地yum配置的，若需要搭建最新版本，可直接按照下面的文档链接进行配置）：

Gluster官网 ，基于centos7/Redhat安装Gluster官方文档

2） GlusterFS相关术语：

Brick（存储块）：指可信主机池中由主机提供的用于物理存储的专用分区。 Volume（逻辑卷）：一个逻辑卷是一组Brick的集合。卷是数据存储的逻辑设备。 FUSE：是一个内核模块，允许用户自己创建文件系统，无须修改内核代码。 Glusterd（后台管理进程）：在存储群集中的每个节点上都要运行。 VFS：内核空间对用户空间提供的访问磁盘的接口。

3）GlusterFS的卷类型：

分布式卷：相当于Windows中的跨区卷，只是扩大了磁盘空间，不具有容错能力； 条带卷：相当于Windows中的带区卷，属于RAID 0级别，一份文件会在多个磁盘上进行读写，文件越大，读写效率越高，但不具有容错能力； 复制卷：相当于Windows中的镜像卷，属于RAID 1级别，具有容错能力，读性能高，但写性能下降，因为要将同一份文件同步写入在多个Brick上。 分布式条带卷：brick server数量是条带数（数据块分布的brick数量）的倍数，兼具分布式卷和条带卷的特点。 分布式复制卷：brick server数量是镜像数（数据副本数量）的倍数，兼具分布式卷和复制卷的特点。 条带复制卷：类似于RAID 10，同时具备条带卷和复制卷的特点。 分布式条带复制卷：三种基本卷的复合卷，通常用于map reduce应用。

在以上几种卷类型中，有些可能无法理解透彻，但是没关系，在生产环境中，大多数公司考虑到磁盘的利用率，会使用RAID5，或者RAID 10，关于RAID 5卷的配置可以参考：GlusterFS Dispersed Volume(纠错卷)总结。

4）下面介绍GlusterFS一些卷类型的特点（不包括RAID5）：

1、分布式卷（类似Windows中的跨区卷）：

分布式卷是GlusterFS的默认卷，在创建卷时，默认选项是创建分布式卷。在该模式下，并没有对文件进行分块处理，文件直接存储在某个server节点上。

分布式卷具有如下特点：

文件分布在不同的服务器，不具备冗余性。 更容易且廉价地扩展卷的大小。 单点故障会造成数据丢失。 依赖底层的数据保护。

2、条带卷（类似Windows中的带区卷，也称为RAID 0）：

stripe模式相当于RAID 0，在该模式下，根据偏移量将文件分成N块（N个条带节点），轮询地存储在每个Brick Server节点。节点把每个数据块都作为普通文件存入本地文件系统，通过扩展属性记录总块数和每块的序号。在配置时指定条带数必须等于卷中Brick 所包含的存储服务器数，在存储大文件时，性能尤为突出，但是不具备冗余性。

条带卷具有如下特点：

数据被分割成更小块分布到块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。

3、复制卷（类似于Windows中的镜像卷，也称为RAID 1）

复制模式，即同一文件保存一份或多份副本，每个节点上保存相同的内容和目录结构。复制模式因为要保存副本，所以磁盘利用率较低，如果多个节点上的存储空间不一致，那么将按照木桶效应取最低节点的容量作为该卷的总容量。复制卷具有冗余性，即使一个节点损坏，也不影响数据的正常使用。

复制卷具有如下特点：

卷中所有的服务器均保存一个完整的副本。 卷的副本数量可由客户创建的时候决定。 至少有两个块服务器或更多服务器。 具备冗余性。

4、分布式复制卷（也称为RAID 10）：

分布式复制卷兼顾分布式卷和复制卷的功能，主要用于需要冗余的情况下。