#### Beegfs安装与基本操作

目录

**[1. 系统环境 1](#_Toc22169)**

**[2. Beegfs安装 2](#_Toc3391)**

**[3. 部署beegfs集群准备工作 2](#_Toc24938)**

[（1）编辑每个节点的/etc/hosts文件 2](#_Toc29125)

[（2）对每个节点进行时间同步 2](#_Toc31372)

[（3）关闭防火墙 2](#_Toc20831)

[（4）关掉selinux 2](#_Toc23903)

[（5）集群分布情况： 3](#_Toc7208)

[（6）文件系统选择： 3](#_Toc15072)

[（7）格式化并挂载 3](#_Toc818)

**[4. 配置beegfs集群 3](#_Toc23484)**

[（1）Mgmtd配置 3](#_Toc8398)

[（2）meta配置 4](#_Toc31728)

[（3）Storage节点配置 4](#_Toc26677)

[（4） Client节点配置 4](#_Toc29784)

**[5. 启动集群服务 4](#_Toc16198)**

**[6. Storage-pool的管理 5](#_Toc10727)**

[（1）查看pool 5](#_Toc31610)

[（2）查看node 5](#_Toc14163)

[（3）新增storage 5](#_Toc831)

[（4）移除storage-target 5](#_Toc11650)

[（5）移除nodes 5](#_Toc8021)

**[7. 配置冗余 6](#_Toc26617)**

[（1）元数据池冗余 6](#_Toc9187)

[（2）数据池冗余 6](#_Toc7703)

**[8. 用户配额 6](#_Toc5977)**

[（1）设置配额 6](#_Toc24297)

[（2）查看配额 7](#_Toc28006)

#### 系统环境

硬件：X86平台

操作系统：centos7.6

默认内核：3.10.0-957

高版本内核：4.14.113

#### Beegfs安装

##### （1）3.10内核选用beegfs-7.0版本

##### （2）使用主机提供的安装包，rpm -ivh \*

#### 部署beegfs集群准备工作

##### （1）编辑每个节点的/etc/hosts文件

#vi /etc/hosts

添加集群中每个节点的主机名和主机的IP地址的对应关系，如下：

192.168.15.129 node1

192.168.15.130 node2

192.168.15.131 node3

192.168.15.132 node4

##### （2）对每个节点进行时间同步

A．手动设置时间

# date //查看系统时间是否正确，正确的话则忽略下面两步

# date -s "2019-11-6 19:05:05" //设置系统时间

# hwclock -w //写入硬件时间

B．使用ntpdate服务同步时间

# yum install -y ntp ntpdate //安装ntp服务

# systemctl start ntpd //主节点启动ntpd服务

# ntpdate server-ip //其他节点同步主节点时间

##### （3）关闭防火墙

如果安装了防火墙，则需要关闭它，命令如下：

# systemctl stop firewalld.service //停止firewall

# systemctl disable firewalld.service //禁止firewall开机启动

##### （4）关掉selinux

命令查看出selinux的状态，命令# sestatus -v或者# getenforce

目前 SELinux 支持三种模式，分别如下：

•enforcing：强制模式，代表 SELinux 运作中，且已经正确的开始限制 domain/type 了；

•permissive：宽容模式：代表 SELinux 运作中，不过仅会有警告讯息并不会实际限制 domain/type 的存取。这种模式可以运来作为 SELinux 的 debug 之用；

•disabled：关闭，SELinux 并没有实际运作。

A．永久方法 – 需要重启服务器（推荐）

编辑# vi /etc/selinux/config，改为SELINUX=disabled，然后重启服务器# reboot。

B．临时方法 – 设置系统参数，服务器重启后失效

使用命令# setenforce 0

附：

# setenforce 1 设置SELinux 成为enforcing模式

# setenforce 0 设置SELinux 成为permissive搜索

##### （5）集群分布情况：

Node1:management server

Node2:metadate server

Node3:storage server

Node4:client

##### （6）文件系统选择：

Storage server/management server : xfs

Metadate server: ext4

##### （7）格式化并挂载

mkfs.xfs /dev/sda ; mount /dev/sda /mgmtd

mkfs.ext4 /dev/sdb ; mount /dev/sdb /meta ;

打开meta的扩展属性 tune2fs -o user\_xattr /dev/sdb

mkfs.xfs /dev/sdc ; mount /dev/sdc /storage

#### 配置beegfs集群

##### （1）Mgmtd配置

management节点配置management service,管理服务需要知道它可以在哪存储数据，它只存储一些节点信息，比如连接性数据，因此不需要太多的存储空间，而且它的数据访问不是性能关键

/opt/beegfs/sbin/beegfs-setup-mgmtd -p /mgmtd/beegfs/beegfs\_mgmtd

##### （2）met**a**配置

metadata service,元数据服务器需要知道它可以在哪存储数据，以及管理服务在哪里运行，选择定义一个确定的数字作为元数据服务器ID（1~65535）由于是第二个节点可设置为2；一个节点配一个元数据；配置多个节点元数据不是分布式，而是只有第一个启动的元数据服务器工作，除非配置为镜像模式。

/opt/beegfs/sbin/beegfs-setup-meta -p /meta/beegfs/beegfs\_meta -s 2 -m node1

-s 表示元数据服务器ID

-m mgmgt服务器节点

Beegfs默认是raid0模式，当一个pool有2个storage时，用户写入的文件会被分切为2个文件分别写入，当一个pool有3个storage时，用户写入的文件会被切分为3个文件分别写入。切片最小默认为512kB。

##### （3）Storage节点配置

Storage Service，存储服务需要知道它可以在哪里存储数据，以及如何到达管理服务器。通常，每个存储服务将在不同的机器上运行多个存储服务和/或多个存储目标(例如多个RAID卷)。选择定义自定义数字存储服务ID和数字存储目标ID(范围1~65535)。这里我们的Storage节点是第3个节点，因此我们将选择编号“3”作为此存储服务的ID，并使用“301”作为存储目标ID，以表明这是存储服务“1”的第一个目标(“01”)。

同一节点的多个storage为切片；不同节点的storage为分布式；

/opt/beegfs/sbin/beegfs-setup-storage -p /storage/beegfs/beegfs\_storage -s 3 -i 301 -m node1

-s 表示存储节点ID

-i 表示本节点的目标ID

-m 表示mgmtd服务器

##### Client节点配置

Client（BeeGFS默认会挂载到/mnt/beegfs，可以自行在配置文件/etc/beegfs/beegfs-mounts.conf中修改）

/opt/beegfs/sbin/beegfs-setup-client -m node1

-m 表示mgmtd服务器

#### 启动集群服务

Management节点:  
 systemctl start beegfs-mgmtd  
 systemctl enable beegfs-mgmtd  
Metadata节点:  
 systemctl start beegfs-meta  
 systemctl enable beegfs-meta  
Storage节点:  
 systemctl start beegfs-storage   
 systemctl enable beegfs-storage   
Client节点:  
 systemctl start beegfs-helperd  
 systemctl enable beegfs-helperd  
 systemctl start beegfs-client  
 systemctl enable beegfs-client

#### Storage-pool的管理

##### （1）查看pool

beegfs-ctl --listtargets --nodetype=meta --state

beegfs-ctl --listtargets --nodetype=storage --state

##### （2）查看node

beegfs-ctl --listnodes --nodetype=meta

beegfs-ctl --listnodes --nodetype=storage

beegfs-ctl --listnodes --nodetype=client

##### （3）查看storage-pool工作模式

beegfs-ctl --getentryinfo /mnt/beegfs/

##### （3）新增storage

/opt/beegfs/sbin/beegfs-setup-storage -p /storage/beegfs/beegfs\_storage -s 3 -i 302 -m node1

重启beegfs-storage生效

##### （4）移除storage-target

查看targetid：beegfs-ctl --liststoragepools

beegfs-ctl --removetarget 302

重启beegfs-client生效

##### （5）移除nodes

beegfs-ctl --removenode --ndoetype=meta $number

beegfs-ctl --removenode --ndoetype=storage $number

#### 配置冗余

##### （1）数据池冗余

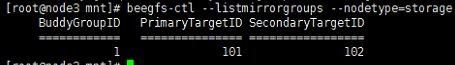
###### 1) 在client节点执行

例：将id 101和102的数据池配成id为1的镜像组(推荐手动配置)，group可以不指定，默认值会从1开始。

beegfs-ctl --addmirrorgroup --nodetype=storage --primary=101 --secondary=102 [--groupid=1]



查看beegfs-ctl --listmirrorgroups --nodetype=storage



###### 参数说明

--numtargets：是切片的个数（默认是4），如果设置的numtargets大于等于镜像组的个数，则表示是RAID0+1；如果镜像组有多个但numtargets设置为1，则表示不切片，数据池是分布式RIAD1。当镜像组不超过4组时保持默认值。

【注意：如果有3个镜像组，设置numtargets为2时，会随机两两组合，即每个文件会随机存在某2个镜像组中，不建议这样配置。】

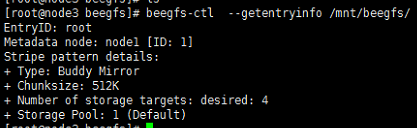
--chunksize：是切片的大小，默认是512K，可以通过k/m（KB/MB）调整，建议默认值

###### 设置

beegfs-ctl --setpattern --pattern=buddymirror /mnt/beegfs/

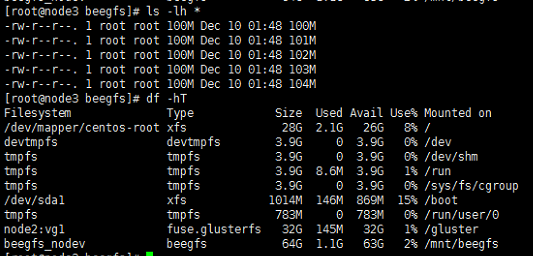
###### 查看工作模式：

beegfs-ctl --getentryinfo /mnt/beegfs/



###### 写数据结果：

写入500M数据，实际已使用1G空间。



##### （2）元数据池冗余

###### 1) 在client节点执行

例：将id 1和2的元数据池配成id为1的镜像组(与数据池id不冲突)group可以不指定，默认值会从1开始。

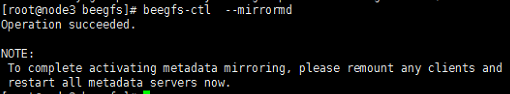
###### 设置

beegfs-ctl --addmirrorgroup --nodetype=meta --primary=1 --secondary=2 [--groupid=1]



###### 激活

beegfs-ctl --mirrormd



按要求重启client服务和meta服务

systemctl restart beegfs-client

systemctl restart beegfs-meta

###### 查看状态

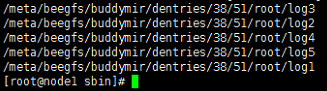
beegfs-ctl --listmirrorgroups --nodetype=meta



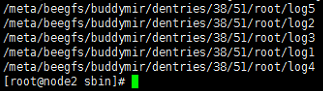
###### 写数据结果:



Target1-meta上元数据：



Target2-meta上元数据：



##### 同步数据

配置了镜像冗余后，丢失数据可以手动同步，但是不论是数据池还是元数据池，只能是次节点2丢数据才能恢复，恢复的原理类似于备份，主节点1不能反过来从次节点2同步数据，如果主节点1丢数据，那么启动同步数据会导致次节点2也丢失数据。

###### 数据池同步：

数据池为了避免不必要的数据传输，主target记录了最后一次同步数据的时间，超过这个时间戳的数据才会同步，但是考虑到缓存问题会加一个时间段（由配置文件中参数sysResyncSafetyThresholdMins确定，默认10分钟）

A 同步数据

beegfs-ctl --startresync --nodetype=storage --mirrorgroupid=1



B 全量同步

如果需要同步时间点以前的数据，我们可以自己给一个时间段：

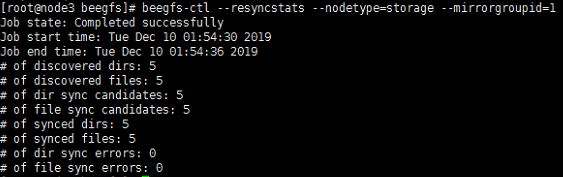
d表示天，h表示小时；

重新同步过去5天的镜像组中的更改

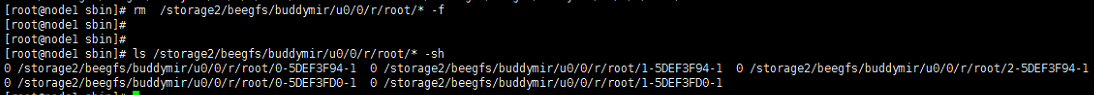
beegfs-ctl --startresync --nodetype=storage --mirrorgroupid=1 --timespan=5d

C 查看同步状态

beegfs-ctl --resyncstats --nodetype=storage --mirrorgroupid=1



D 同步结果



###### 元数据池同步：

元数据比较小，所以没有时间戳的限制，同步时会将全部数据发给伙伴。

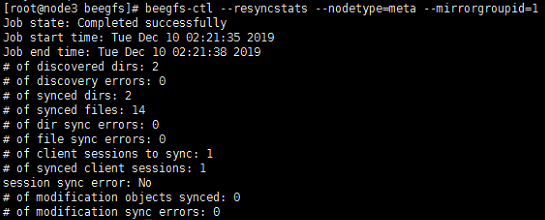
A 同步

beegfs-ctl --startresync --nodetype=meta --mirrorgroupid=1

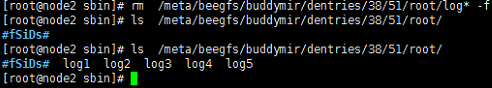


B查看状态

beegfs-ctl --resyncstats --nodetype=meta --mirrorgroupid=1



C 同步结果:



#### 用户配额

BeeGFS允许按用户或组定义系统范围内的磁盘空间分配配额和块文件数。这可以用于将用户组织到具有不同限制级别的不同访问层中，还可以防止个人单独使用文件系统的所有资源。BeeGFS配额管理机制由两个功能组成：配额跟踪和配额执行。配额跟踪允许查询用户和组在系统中使用的数据量和块文件数。

##### （1）设置配额（针对已存在的集群）

###### A.关闭已启动的所有storage和client服务

Systemctl stop beegfs-storage

Systemctl stop beegfs-client

###### B.重新挂载storage磁盘目录(xfs)

mount /dev/sdb /data -t xfs -orw,uqnoenforce,gqnoenforce

...

###### C.在所有client节点上设置

/etc/beegfs/beegfs-client.conf

quotaEnabled = true

###### D.在所有storage节点设置

/etc/beegfs/beegfs-storage.conf

quotaEnableEnforcement = true

###### E.启动所有的beegfs-client， beegfs-storage服务

Systemctl start beegfs-client

Systemctl start beegfs-storage

###### F.在mgmtd节点上设置

/etc/beegfs/beegfs-mgmtd.conf

quotaEnableEnforcement = true

重启beegfs-mgmtd服务systemctl restart beegfs-mgmtd

###### G.设置用户配额

0表示root用户，对应uid可通过/etc/passwd查看,设置大小限制为1G，inode数不限制：

beegfs-ctl --setquota --uid 0 --sizelimit=1G --inodelimit=unlimited --storagepoolid=1

##### （2）查看配额

默认是10分钟mgmtd去查询一次，考虑到可以及时发现用户容量的使用情况可以减少时间间隔，但是同样也会加大性能的消耗，影响io性能；可以调整quotaUpdateIntervalMin。查询范围默认是全部的uid和gid，可以调整quotaQueryUIDRange&quotaQueryGIDRange。

beegfs-ctl --getquota --uid 0 //查看root用户的配额信息