CentOS7安装配置BeeGFS并行文件系统

[发表评论](https://www.kclouder.cn/centos7-beegfs/#respond)

BeeGFS原名为FhGFS，是由Fraunhofer Institute为工业数学计算而设计开发，由于在欧洲和美国的中小  
型HPC系统性能表现良好，在2014年改名注册为BeeGFS并受到科研和商业的广泛应用。BeeGFS既是一个网络  
文件系统也是一个并行文件系统。客户端通过网络与存储服务器进行通信(具有TCP/IP或任何具有RDMA功能  
的互连，如InfiniBand，RoCE或Omni-Path，支持native verbs 接口)。通过BeeGFS添加更多的服务器，  
其容量和性能被聚合在单个命名空间中。BeeGFS是遵循GPL的“免费开源”产品，文件系统没有许可证费用。  
由ThinkParQ提供专业支持，系统集成商可以为客户构建使用BeeGFS的解决方案。

**环境准备**  
1，本次安装采用虚拟机的形式，所有节点安装在VMware ESXi上。  
**Software:** CentOS 7.5 on all nodes   
**Host Services:**    
 node01: Management Server (IP:172.16.80.111)  
 node02: Metadata Server (IP:172.16.80.112)   
 node03: Storage Server (IP:172.16.80.113)   
 node04: Client (IP:172.16.80.114)   
 node05: Admon Server (IP:172.16.80.115) (可选项)  
**Storage:**  
 Storage servers with RAID-6, xfs, mounted to "/data".  
 Metadata servers with RAID-1, ext4, mounted to "/mnt/md0".

参考文档：  
[**BeeGFS Documentation**](https://www.beegfs.io/wiki/TableOfContents#hn_59ca4f8bbb)

2， 所有节点 配置/etc/hosts

3，所有节点关闭防火墙和SELINUX  
 systemctl stop firewalld  
 systemctl disable firewalld  
 sed -i 's/SELINUX=enforcing/SELINUX=disabled/g'/etc/sysconfig/selinux

**RAID配置**  
由于采用虚拟机的形式进行安装，所以RAID采用软RAID的方式实现，分别在Metadata节点和Storage节点通  
过mdadm创建RAID。  
1，在Metadata节点创建RAID1  
 fdisk /dev/sdb  
 fdisk /dev/sdc  
 mdadm -C /dev/md0 -ayes -l1 -n2 /dev/sd[b,c]1

2，格式化md0并挂载到/data，同时写入fstab  
 mkfs.ext4 /dev/md0  
 mkdir -p /data  
 mount /dev/md0 /data  
 echo /dev/md0 /data ext4 defaults 0 0 | sudo tee -a /etc/fstab

3，打开Metadata Server的扩展属性功能：  
 tune2fs -o user\_xattr /dev/md0

4，以同样的方式在Storage节点创建RAID6，并挂载到/mnt/md0  
 fdisk /dev/sdb  
 fdisk /dev/sdc  
 fdisk /dev/sdd  
 fdisk /dev/sde  
 mdadm -Cv /dev/md0 -ayes -l6 -n4 /dev/sd[b,c,d,e]1  
 mkfs.xfs /dev/md0  
 mkdir -p /mnt/md0  
 mount /dev/md0 /mnt/md0  
 echo /dev/md0 /mnt/md0 xfs defaults 0 0 | sudo tee -a /etc/fstab

5，RAID配置完成后，系统分区如下图，如果Storage节点配置有多个RAID6，以同样的方式进行操作。

**节点安装（Package Download and Installation）**  
1，在所有节点下载BeeGFS的repo文件到/etc/yum.repos.d/  
 wget -O /etc/yum.repos.d/beegfs-rhel7.repo https://www.beegfs.io/release/latest-stable/dists/beegfs-rhel7.repo

2，在管理节点安装Management Service  
 yum install beegfs-mgmtd   
3，在Metadata节点安装Metadata Service  
 yum install -y beegfs-meta   
4，在Storage节点安装Storage Service   
 yum install -y beegfs-storage   
5，在Client节点安装Client and Command-line Utils   
 yum install -y beegfs-client beegfs-helperd beegfs-utils   
6，在监控节点（Admon）安装Admon Service   
 yum install beegfs-admon  
7，如果需要使用Infiniband RDMA功能，还需要在Metadata和Storage节点安装libbeegfs-ib  
 yum install libbeegfs-ib

**节点配置**  
1，Management节点配置Management Service，管理服务需要知道它可以在哪里存储数据。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。因此，此服务  
通常可以不在专用机器上运行（可以跑在虚拟机上）。  
 /opt/beegfs/sbin/beegfs-setup-mgmtd -p /data/beegfs/beegfs\_mgmtd

2，Meta节点配置Metadata Service，元数据服务需要知道它可以在哪里存储数据，以及管理服务在哪里运  
行。选择定义一个定制的数字元数据服务ID(范围1~65535)。这里我们的Metadata节点是第二个节点，所以  
这里我们选择数字“2”作为元数据服务ID。  
/opt/beegfs/sbin/beegfs-setup-meta -p /data/beegfs/beegfs\_meta -s 2 -m bgfs01

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节点是第三个节点，  
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“3”的第  
一个目标(“01”)。  
/opt/beegfs/sbin/beegfs-setup-storage -p /mnt/md0/beegfs\_storage -s 3 -i 301 -m bgfs01

4，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 bgfs01

5，监控节点修改配置文件：  
 vi /etc/beegfs/beegfs-admon.conf  
 sysMgmtdHost=bgfs01

6，以上配置完成后，在所有节点启动服务并设置开机自动启动  
 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 start beegfs-client  
 systemctl start beegfs-helperd  
 systemctl enable beegfs-helperd  
 systemctl start beegfs-client  
 systemctl enable beegfs-client  
 Admon节点:  
 systemctl start beegfs-admon  
 systemctl enable beegfs-admon

7，执行完以上配置后，重启所有节点，并确认重启后所有节点上的服务均已经正常启动，到这里BeeGFS的基  
本配置就完成了。

8，检查节点状态，在Client节点执行以下命令：  
 beegfs-ctl --listnodes --nodetype=meta --nicdetails  
 beegfs-ctl --listnodes --nodetype=storage --nicdetails  
 beegfs-ctl --listnodes --nodetype=client --nicdetails  
 beegfs-net #显示Client实际使用的连接  
 beegfs-check-servers #显示服务的连接性  
 beegfs-df #显示存储和元数据目标的空闲空间和索引节点

**读写测试**  
1，使用dd进行测试，在Client节点执行以下操作进行写入测试  
 mkdir -p /mnt/beegfs/dd-test  
 dd if=/dev/zero of=/mnt/beegfs/dd-test/test-file bs=1MB count=50000  
 50000+0 records in  
 50000+0 records out  
 50000000000 bytes (50 GB) copied, 545.962 s, 91.6 MB/s  
 df -hT  
 Filesystem Type Size Used Avail Use% Mounted on  
 beegfs\_nodev beegfs 100G 47G 54G 47% /mnt/beegfs

2，在Storage节点查看写入情况，确认上一步测试中写入的数据写入了/mnt/md0

3，使用sysbench进行测试，在Client节点执行以下操作进行写入测试  
 yum install -y epel-release  
 yum install -y sysbench  
 mkdir -p /mnt/beegfs/sysbench-test  
 cd /mnt/beegfs/sysbench-test  
 sysbench --test=fileio --threads=20 --file-total-size=1G --file-test-mode=rndrw prepare   
 sysbench --test=fileio --threads=20 --file-total-size=1G --file-test-mode=rndrw run

4，在Storage节点查看写入情况，确认上一步测试中写入的数据写入了/mnt/md0

**Storage节点配置多个Storage Target**  
1，以相同的方式在Storage节点创建两个RAID6，md1和md2  
2，挂载md1和md2  
 mkdir -p /mnt/md1  
 mkdir -p /mnt/md2  
 mount /dev/md1 /mnt/md1  
 mount /dev/md2 /mnt/md2  
 echo /dev/md1 /mnt/md1 xfs defaults 0 0 | sudo tee -a /etc/fstab  
 echo /dev/md2 /mnt/md2  xfs   defaults   0   0 | sudo tee -a /etc/fstab   
3，添加Storage Target，配置完成后如下图  
 /opt/beegfs/sbin/beegfs-setup-storage -p /mnt/md1/beegfs\_storage -s 3 -i 302  
 /opt/beegfs/sbin/beegfs-setup-storage -p /mnt/md2/beegfs\_storage -s 3 -i 303

4，在Client节点执行dd写入测试  
 dd if=/dev/zero of=/mnt/beegfs/dd-test/test-file bs=1MB count=100000  
 100000+0 records in  
 100000+0 records out  
 100000000000 bytes (100 GB) copied, 954.976 s, 105 MB/s  
5，dd写入完成后，在Storage节点查看写入情况，如下图100G的数据被并行平均写入到三个md中。

**多Storage节点**  
1，我们再安装一台虚拟机VM，并配置为Storage节点，创建1个Storage Targate为601，因为是第六个  
节点的第一个Storage Targate，所以ID配置为601方便管理。

2，使用dd执行测试写入10G数据。  
 dd if=/dev/zero of=/mnt/beegfs/dd-test/4node bs=10M count=1000  
 1000+0 records in  
 1000+0 records out  
 10485760000 bytes (10 GB) copied, 182.301 s, 57.5 MB/s

3，dd写入完成以后，查看Storage节点空间占有用情况，与写入前对比，可以看到10GB数据被平均写入到  
4个Storage Target中。这里我们一共配置了两个存储节点，4个Storage Target。

**Admon GUI控制台**  
1，在Admon节点上执行以下命令即可启动web管理页面，首次运行时系统会提示运行Admon守护进程的主机的  
主机名和端口(默认值:8000)。还可以更改GUI的内部桌面的分辨率和GUI的默认日志级别。  
 java -jar /opt/beegfs/beegfs-admon-gui/beegfs-admon-gui.jar

2，登录机制基于两个预定义的用户。  
用户“Information”(初始密码“Information”)只能查看统计信息，而用户“Administrator”(初始密码“  
admin”)可以执行管理任务。

3，登录后，即可以图形界面方式监控和管理BeeGFS。

到这里BeeGFS的安装测试就完成了，由于是在VM虚拟机中进行的演示，所有节点都是在同一个LUN上面，并  
且网络为千兆网卡，所以这里的演示并不能展示出BeeGFS的实际性能。如果Storage节点采用物理机进行安  
装，并采用10G或目前比较流行的25G网络，同时采用RDMA功能，所有Storage建立在硬件RAID上，其并行  
数据读写性能才会得到充分发挥。这里只是对BeeGFS的安装配置进行一个基本的演示。