一体机存储布局与挂载设置

v1.1

2016-06-16

# Glusterfs分布式存储后端布局

一体机中，由于采用了RAID卡来聚合磁盘容量，因此glusterfs卷的布局为：在一个brick中包含多个volume。将同一个brick上的多个volume视为一个虚拟的卷组，在一体机中，为HDD盘阵和PCIE-FLASH卡预设了两个虚拟卷组，分别为vg-raid和vg-flash。

在vg-raid卷组，预设gv0, gv1, gv2, gv3, sv0, sv1, sv2, sv3, tv0, tv1, tv2, tv3十二个卷，配置和用途如下

|  |  |  |
| --- | --- | --- |
| 卷名称 | 卷配置 | 预设用途 |
| gv0 | Distribute | 用户数据（临时数据等） |
| gv1 | Distribute+2副本 | 用户数据（普通数据） |
| gv2 | Distribute+3副本 | 用户数据 |
| gv3 | Distribute+4副本 | 用户数据（重要数据） |
| sv0 | Distribute | 系统用途（虚拟机镜像等） |
| sv1 | Distribute+2副本 | 系统用途（虚拟机镜像等） |
| sv2 | Distribute+3副本 | 系统用途（虚拟机镜像等） |
| sv3 | Distribute+4副本 | 系统用途（虚拟机镜像等） |
| tv0 | Distribute | 测试用 |
| tv1 | Distribute+2副本 | 测试用 |
| tv2 | Distribute+3副本 | 测试用 |
| tv3 | Distribute+4副本 | 测试用 |

在vg-flash卷组，预设fv0, fv1, fv2, fv3四个卷，配置和用途如下

|  |  |  |
| --- | --- | --- |
| 卷名称 | 卷配置 | 预设用途 |
| fv0 | Distribute | 需要高IOPS的中间数据 |
| fv1 | Distribute+2副本 | 需要高IOPS的中间数据 |
| fv2 | Distribute+3副本 | 需要高IOPS的中间数据 |
| fv3 | Distribute+4副本 | 需要高IOPS的中间数据 |

按照gulsterfs推荐的命名规范配置brick路径如下

HDD盘阵存储设备挂载路径为

/data/glusterfs/vg-raid/brick1/，在设备上建立gv0, gv1, gv2, gv3, sv0, sv1, sv2, sv3, tv0, tv1, tv2, tv3目录作为逻辑卷。

PCIE-FLASH卡设备挂载路径为

/data/glusterfs/vg-flash/brick1/，在设备上建立fv0, fv1, fv2, fv3目录作为逻辑卷。

# 存储系统前端布局

存储系统前端目录结构

|  |  |  |
| --- | --- | --- |
| 路径 | 预设用途 |  |
| /storage/ | 存储系统目录 |  |
| /storage/.system/ | 隐藏的系统目录 |  |
| /storage/.system/cfg/ | 系统配置 | 包含必要的配置、脚本等 |
| /storage/.system/vol/ | 系统预设卷 | gv0, gv1, gv2, gv3, sv0, sv1, sv2, sv3, tv0, tv1, tv2, tv3, fv0, fv1, fv2, fv3等预设卷会在此目录下挂载 |
| /storage/data/ | 默认用户数据入口 | 根据用户对副本的需求，通过mount –bind绑定到某个系统默认卷（由配置脚本操作实现） |
| /storage/其它目录 | 其它用户数据入口，根据需要增添 |  |

# 存储系统配置文件

前端在挂载glusterfs卷时，需要用主机名”storage-server”来指定服务器，该主机名称为存储系统保留的字符。在/etc/hosts文件中指定storage-server的ip地址。对于一体机中的物理机系统，storage-server设定为127.0.0.1；对于虚拟机系统，sotrage-server设定为所在宿主的存储访问接口ip。

/etc/fstab :

包含后端brick挂载、前端预设卷挂载。

|  |
| --- |
| #  # Gluster Storage server bricks  #  UUID=124e54c5-d428-4099-84c2-8252b3a3a411 /data/glusterfs/vg-raid/brick1 xfs noatime,nodiratime,logbufs=8,logbsize=256k,largeio,inode64,swalloc,allocsize=131072k,nobarrier 1 2  #  # Gluster Storage client volumes  #  #storage-server:/fv0 /storage/.system/vol/fv0 glusterfs defaults,\_netdev 0 0  #storage-server:/fv1 /storage/.system/vol/fv1 glusterfs defaults,\_netdev 0 0  #storage-server:/fv2 /storage/.system/vol/fv2 glusterfs defaults,\_netdev 0 0  #storage-server:/fv3 /storage/.system/vol/fv3 glusterfs defaults,\_netdev 0 0  storage-server:/gv0 /storage/.system/vol/gv0 glusterfs defaults,\_netdev 0 0  #storage-server:/gv1 /storage/.system/vol/gv1 glusterfs defaults,\_netdev 0 0  #storage-server:/gv2 /storage/.system/vol/gv2 glusterfs defaults,\_netdev 0 0  #storage-server:/gv3 /storage/.system/vol/gv3 glusterfs defaults,\_netdev 0 0  #storage-server:/sv0 /storage/.system/vol/sv0 glusterfs defaults,\_netdev 0 0  #storage-server:/sv1 /storage/.system/vol/sv1 glusterfs defaults,\_netdev 0 0  #storage-server:/sv2 /storage/.system/vol/sv2 glusterfs defaults,\_netdev 0 0  #storage-server:/sv3 /storage/.system/vol/sv3 glusterfs defaults,\_netdev 0 0  storage-server:/tv0 /storage/.system/vol/tv0 glusterfs defaults,\_netdev 0 0  storage-server:/tv1 /storage/.system/vol/tv1 glusterfs defaults,\_netdev 0 0  #storage-server:/tv2 /storage/.system/vol/tv2 glusterfs defaults,\_netdev 0 0  #storage-server:/tv3 /storage/.system/vol/tv3 glusterfs defaults,\_netdev 0 0 |

/etc/fstab\_sotrage:

包含两部分

1. 要挂载的glusterfs volume，即前述的预设的后端卷名称，从中导出需要挂载的卷名称
2. 用户数据入口的mount绑定。

示例如下：

|  |
| --- |
| #  # Gluster Storage server volume entries  # line start with #export, then the pre-defined  # system volume entry name, each entry one line.  # by Liu Tao <lt@ncic.ac.cn>  #  #export gv0  #export tv1  #  # Gluster Storage user entries  # /etc/fstab format  #  /storage/.system/vol/gv0 /storage/data none rw,bind 0 0  /storage/.system/vol/tv1 /storage/backup none rw,bind 0 0 |

/storage/.system/cfg/vol\_start脚本:

该脚本需在/etc/rc.local文件中被调用。（需要注意，centos7中需要为/etc/rc.local文件开启可执行权限。）

执行两部分功能

1. 确保/etc/fstab\_storage脚本中所导出的后端卷已经全部被挂载
2. 执行/etc/fstab\_sotrage脚本第二部分的用户数据入口绑定挂载

由于glusterfs为网络文件系统，在系统初始化过程中有可能mount不成功，因此vol\_start脚本检查并确保在/etc/fstab\_storage文件第一部分导出的后端卷全部挂载在前端预设卷上，之后根据/etc/fstab\_storage文件的第二部分，将用户数据入口绑定到指定的前端预设卷上。（确保先挂载前端预设卷，再绑定用户数据入口。这样用户数据入口umount之后，不影响前端预设卷的挂载状态。）

所以/etc/fstab\_sotrage文件中导出的后端卷必须写正确，否则前端用户数据入口无法完成挂载。