LVM Snapshot简介

2014/9/3

renyl

# 1 基本原理

1. [Logical Volume Manager (LVM)](http://en.wikipedia.org/wiki/Logical_Volume_Manager_(Linux))提供了对任意一个Logical Volume(LV)做“快照”(snapshot)的功能，以此来获得一个分区一致性备份。
2. LVM的snapshot是通过“写时复制”(copy on write) 的方法实现：
3. 当一个snapshot创建的时候，仅拷贝原始卷里数据的元数据(meta-data)。创建的时候，并不会有数据的物理拷贝，因此snapshot的创建几乎是实时的。
4. 当原始卷上有写操作执行时，snapshot跟踪原始卷块的改变，这个时候原始卷上将要改变的数据在改变之前被拷贝到snapshot预留的空间里，因此这个原理的实现叫做写时复制(copy-on-write)。
5. 在写操作写入块之前，CoW会将原始数据移动到snapshot空间里，这样就保证了所有的数据在snapshot创建时保持一致。而对于snapshot的读操作，如果是读取数据块是没有修改过的，那么会将读操作直接重定向到原始卷上，如果是要读取已经修改过的块，那么就读取拷贝到snapshot中的块。
6. 这样，通常的文件I/0流程有一个改变，那就是在文件系统和设备驱动之间增加了一个cow层，变成了下面这个样子：

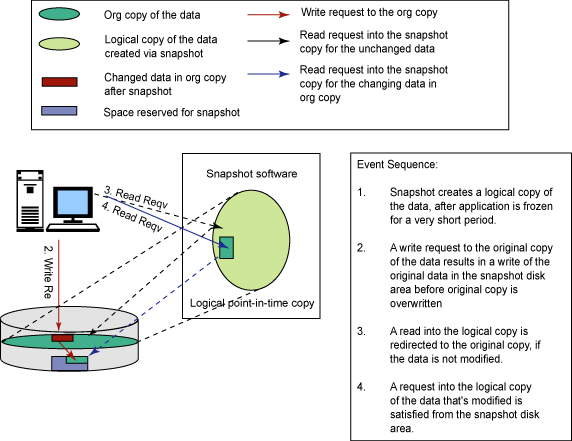
file I/0

filesystem

CoW

block I /O

下图描述了COW的实现原理：



说明：

1. 采取CoW实现方式时，snapshot空间的大小并不需要和原始卷一样大，其大小仅仅只需要考虑，从shapshot创建到释放这段时间内，估计块的改变量有多大。
2. 如果snapshot的空间记录满了原始卷块变换的信息，那么这个snapshot立刻被释放，从而无法使用，从而导致这个snapshot无效。
3. 因此，一定要snapshot的生命周期里，做完需要做得事情。否则，当原始卷的改变量大于snapshot空间大小时，就无法恢复到分区原始状态了。

# LVM基本命令

## 2.1 修改分区System ID

|  |
| --- |
| [root@localhost /]# parted /dev/sde  GNU Parted 3.1  Using /dev/sde  Welcome to GNU Parted! Type 'help' to view a list of commands.  (parted) p  Model: LSI RAID 5/6 SAS 6G (scsi)  Disk /dev/sde: 146GB  Sector size (logical/physical): 512B/512B  Partition Table: msdos  Disk Flags:  Number Start End Size Type File system Flags  1 512B 20.0GB 20.0GB primary xfs lvm  2 20.0GB 40.0GB 20.0GB primary xfs  (parted) set 2 lvm  New state? [on]/off? on  (parted) p  Model: LSI RAID 5/6 SAS 6G (scsi)  Disk /dev/sde: 146GB  Sector size (logical/physical): 512B/512B  Partition Table: msdos  Disk Flags:  Number Start End Size Type File system Flags  1 512B 20.0GB 20.0GB primary xfs lvm  2 20.0GB 40.0GB 20.0GB primary xfs lvm  (parted) q  Information: You may need to update /etc/fstab.  [root@localhost /]# |

## 建立PV（Physical Volume）

|  |
| --- |
| [root@localhost /]# pvcreate /dev/sde1  WARNING: xfs signature detected on /dev/sde1 at offset 0. Wipe it? [y/n] y  Wiping xfs signature on /dev/sde1.  Physical volume "/dev/sde1" successfully created  [root@localhost /]# pvcreate /dev/sde2  WARNING: xfs signature detected on /dev/sde2 at offset 0. Wipe it? [y/n] y  Wiping xfs signature on /dev/sde2.  Physical volume "/dev/sde2" successfully created |

## 建立VG（Volume Group）

|  |
| --- |
| [root@localhost ~]# vgcreate my\_vg /dev/sde1 /dev/sde2  Volume group "my\_vg" successfully created  [root@localhost ~]# |

## 建立LV（Logical Volume）

|  |
| --- |
| [root@localhost ~]# lvcreate -L 10GB -n my\_lv my\_vg  Logical volume "my\_lv" created  [root@localhost ~]# |

注：

1）-L：后面接容量，容量的单位可以是KB、MB、GB等。

2）-n：后面接LV的名称。

## 格式化分区

|  |
| --- |
| [root@localhost ~]# mkfs.xfs /dev/my\_vg/my\_lv  meta-data=/dev/my\_vg/my\_lv isize=256 agcount=4, agsize=655360 blks  = sectsz=512 attr=2, projid32bit=1  = crc=0  data = bsize=4096 blocks=2621440, imaxpct=25  = sunit=0 swidth=0 blks  naming =version 2 bsize=4096 ascii-ci=0 ftype=0  log =internal log bsize=4096 blocks=2560, version=2  = sectsz=512 sunit=0 blks, lazy-count=1  realtime =none extsz=4096 blocks=0, rtextents=0  [root@localhost /]# mount /dev/my\_vg/my\_lv /mnt/  [root@localhost /]# df -hT  Filesystem Type Size Used Avail Use% Mounted on  /dev/mapper/vgrhel-root xfs 20G 5.7G 14G 30% /  devtmpfs devtmpfs 16G 0 16G 0% /dev  tmpfs tmpfs 16G 80K 16G 1% /dev/shm  tmpfs tmpfs 16G 9.3M 16G 1% /run  tmpfs tmpfs 16G 0 16G 0% /sys/fs/cgroup  /dev/sda1 xfs 509M 121M 388M 24% /boot  /dev/mapper/my\_vg-my\_lv xfs 10G 33M 10G 1% /mnt |

## 扩展VG和LV大小

|  |
| --- |
| [root@localhost /]# parted /dev/sde  GNU Parted 3.1  Using /dev/sde  Welcome to GNU Parted! Type 'help' to view a list of commands.  (parted) p  Model: LSI RAID 5/6 SAS 6G (scsi)  Disk /dev/sde: 146GB  Sector size (logical/physical): 512B/512B  Partition Table: msdos  Disk Flags:  Number Start End Size Type File system Flags  1 512B 20.0GB 20.0GB primary lvm  2 20.0GB 40.0GB 20.0GB primary lvm  3 40.0GB 60.0GB 20.0GB primary xfs lvm  [root@localhost /]# umount /mnt/  [root@localhost /]# pvcreate /dev/sde3  WARNING: xfs signature detected on /dev/sde3 at offset 0. Wipe it? [y/n] y  Wiping xfs signature on /dev/sde3.  Physical volume "/dev/sde3" successfully created  [root@localhost /]# vgextend my\_vg /dev/sde3  Volume group "my\_vg" successfully extended  [root@localhost /]# vgdisplay  --- Volume group ---  VG Name my\_vg  System ID  Format lvm2  Metadata Areas 3  Metadata Sequence No 10  VG Access read/write  VG Status resizable  MAX LV 0  Cur LV 1  Open LV 0  Max PV 0  Cur PV 3  Act PV 3  VG Size 55.88 GiB  PE Size 4.00 MiB  Total PE 14304  Alloc PE / Size 2560 / 10.00 GiB  Free PE / Size 11744 / 45.88 GiB  VG UUID YecA1m-RiT2-cq5n-ownL-Ej2i-2eSe-J1RfAg  [root@localhost /]# lvextend -L +10GB /dev/my\_vg/my\_lv  Extending logical volume my\_lv to 20.00 GiB  Logical volume my\_lv successfully resized  [root@localhost /]# mount /dev/my\_vg/my\_lv /mnt/  [root@localhost /]# df -hT  Filesystem Type Size Used Avail Use% Mounted on  /dev/mapper/vgrhel-root xfs 20G 5.7G 14G 30% /  devtmpfs devtmpfs 16G 0 16G 0% /dev  tmpfs tmpfs 16G 80K 16G 1% /dev/shm  tmpfs tmpfs 16G 9.3M 16G 1% /run  tmpfs tmpfs 16G 0 16G 0% /sys/fs/cgroup  /dev/sda1 xfs 509M 121M 388M 24% /boot  /dev/mapper/my\_vg-my\_lv xfs 10G 33M 10G 1% /mnt  [root@localhost /]# xfs\_growfs /mnt/  meta-data=/dev/mapper/my\_vg-my\_lv isize=256 agcount=4, agsize=655360 blks  = sectsz=512 attr=2, projid32bit=1  = crc=0  data = bsize=4096 blocks=2621440, imaxpct=25  = sunit=0 swidth=0 blks  naming =version 2 bsize=4096 ascii-ci=0 ftype=0  log =internal bsize=4096 blocks=2560, version=2  = sectsz=512 sunit=0 blks, lazy-count=1  realtime =none extsz=4096 blocks=0, rtextents=0  data blocks changed from 2621440 to 5242880  [root@localhost /]# df -hT  Filesystem Type Size Used Avail Use% Mounted on  /dev/mapper/vgrhel-root xfs 20G 5.7G 14G 30% /  devtmpfs devtmpfs 16G 0 16G 0% /dev  tmpfs tmpfs 16G 0 16G 0% /sys/fs/cgroup  /dev/sda1 xfs 509M 121M 388M 24% /boot  /dev/mapper/my\_vg-my\_lv xfs 20G 33M 20G 1% /mnt  [root@localhost /]# |

## LVM删除

|  |
| --- |
| [root@localhost /]# umount /mnt/  [root@localhost /]# lvremove /dev/my\_vg/my\_lv  Do you really want to remove active logical volume my\_lv? [y/n]: y  Logical volume "my\_lv" successfully removed  [root@localhost /]# vgchange -a n my\_vg //让这个vg不具有Active的标志。  0 logical volume(s) in volume group "my\_vg" now active  [root@localhost /]# vgremove my\_vg  Volume group "my\_vg" successfully removed  [root@localhost /]# pvremove /dev/sde1  Labels on physical volume "/dev/sde1" successfully wiped  [root@localhost /]# pvremove /dev/sde2  Labels on physical volume "/dev/sde2" successfully wiped  [root@localhost /]# pvremove /dev/sde3  Labels on physical volume "/dev/sde3" successfully wiped  [root@localhost /]# |

## 相关命令

|  |  |  |  |
| --- | --- | --- | --- |
| 任务 | PV | VG | LV |
| 搜索（scan） | pvscan | vgscan | lvscan |
| 建立（create） | pvcreate | vgcreate | lvcreate |
| 列出（display） | pvdisplay | vgdisplay | lvdisplay |
| 增加（extend） | - | vgextend | lvextend |
| 减少（reduce） | - | vgreduce | lvreduce |
| 删除（remove） | pvremove | vgremove | lvremove |
| 改变容量（resize） | - | lvresize | - |

# LVM Snaphost

## 3.1 Backup

|  |
| --- |
| [root@localhost /]# lvcreate -L 5GB -s -n my\_snapshot /dev/my\_vg/my\_lv  Logical volume "my\_snapshot" created  [root@localhost /]# |

## Restore

|  |
| --- |
| [root@localhost /]#lvconvert --merge /dev/my\_vg/my\_snapshot  Logical volume my\_vg/my\_lv contains a filesystem in use.  Can't merge over open origin volume.  Merging of snapshot my\_snapshot will start next activation.  [root@localhost /]#reboot //重启后生效 |

# 注意事项

1. 当Snapshot的空间记录满了原始卷块变换的信息，那么这个Snapshot将立刻被释放，从而导致无法使用这个Snapshot。因此，在建立Snapshot时，需要预估计原始卷块需要做多大的修改量。
2. 在建立Snapshot之前确保被备份的文件都在磁盘上，因此需要umount这个分区或者执行命令“echo 3 > /proc/sys/vm/drop\_caches”。
3. 在使用Snapshot进行恢复时，如果原始卷块被umount的话，恢复立刻生效。如果原始卷块正在mount被使用中，那么系统重启后将生效。
4. 系统的boot分区不能使用LVM进行管理。

# 附录

自动备份与恢复脚本：

|  |
| --- |
| [root@localhost renyl]# cat lvm\_snapshot.sh  #!/bin/bash    #Program:  # backup and restore with filesystem snapshots  #  #Histroy:  # renyl 2014/7/1 0.1version  #  # renyl 2014/7/2 0.2version  # add:auto backup after restore snapshot      help()  {  set +x  echo "Parameter is wrong."  echo "Usage: $0 -check"  echo "Usage: $0 -backup <size> <snapshot\_name> <full\_backup\_lv>"  echo "Usage: $0 -restore <full\_snapshot\_name>"  echo "Example: $0 -backup 5GB renyl\_snap /dev/vgsnap/lvroot"  echo "Example: $0 -restore /dev/vgsnap/renyl\_snap"  exit 1  }    option=$1    case ${option} in    "-check")  vgdisplay  lvdisplay  exit 0  ;;    "-backup")  :  ;;    "-restore")  full\_snapshot\_name=$2    if [ -e "${full\_snapshot\_name}" ];then    snap\_path\_line=`lvdisplay | grep -n "${full\_snapshot\_name}" | awk 'BEGIN {FS=":"}; {print $1}'`  backup\_vg\_line=`expr ${snap\_path\_line} + 2` #magin number  backup\_lv\_line=`expr ${snap\_path\_line} + 6` #magic number  snap\_size\_line=`expr ${snap\_path\_line} + 11` #magin number  snap\_lv\_line=`expr ${snap\_path\_line} + 1` #magin number    backup\_vg\_name=`lvdisplay | sed -n "${backup\_vg\_line}p" | awk '{print $3}'` #magic number  backup\_lv\_name=`lvdisplay | sed -n "${backup\_lv\_line}p" | awk '{print $7}'` #magic number  snapshot\_size=`lvdisplay | sed -n "${snap\_size\_line}p" | awk '{print $3}'` #magic number  snapshot\_name=`lvdisplay | sed -n "${snap\_lv\_line}p" | awk '{print $3}'` #magic number    PWD=`pwd` echo "${PWD}/lvm\_snapshot.sh -backup ${snapshot\_size} ${snapshot\_name} /dev/${backup\_vg\_name}/${backup\_lv\_name}" >>/etc/rc.d/rc.local    chmod +x /etc/rc.d/rc.local    lvconvert --merge ${full\_snapshot\_name}    echo "------------------------------------------------"  echo "Warning: It will be effective after system reboot."    else  echo "File ${full\_snapshot\_name} is not exsit!"  echo "------------------------------------------------"  help  fi    exit 0  ;;    \*)  help  ;;  esac    parameter\_num=$#    if [ ${parameter\_num} -ne "4" ];then    echo "Run backup need 4 parameter."  echo "------------------------------------------------"  help  fi  snap\_size=$2  snap\_name=$3  full\_backup\_lv=$4  lvcreate -L ${snap\_size} -s -n ${snap\_name} ${full\_backup\_lv}  remove\_line=`cat /etc/rc.d/rc.local | grep -n "lvm\_snapshot.sh" | awk 'BEGIN {FS=":"}; {print $1}'`  if [ -n "${remove\_line}" ];then  sed -i "${remove\_line}d" /etc/rc.d/rc.local  fi |