Le Logical Volume Manager dans Linux

Imed Chihi, imed.chihi@gmail.com
Séminaire CLLFST
Faculté des Sciences de Tunis, Novembre 2008

C'est quoi “LVM”?

- Couche d'abstraction en dessus des disques physiques
- Composante de base: Physical Volume (volume physique)
- Un Physical Volume est typiquement une partition ou un disque entier avec un label dans les premiers secteurs
- Des Physical Volumes sont regroupés en un Volume Group

C'est quoi “LVM”?

- Un Volume Group est partitionné en des Logical Volumes
- Un Logical Volume dans un Volume Group est comme une partition sur un disque
- Les Logical Volumes peuvent être utilisés pour créer des systemes de fichiers
- L'unité d'allocation est l'extent: Logical Extent et Physical Extent
- 4 MB par extent par défaut

Implémentation

- Composante noyau: module `dm_mod`, `dm_snapshot` et autres
- Composante userspace: `libdevmapper` et suite d'outils pv*, lv*, vg*, etc.
- Version statique dans `/sbin/lvm.static`
- Fichiers de configuration sous `/etc/lvm/`

Historique

- LVM a existé depuis longtemps sur les plateformes Unix
- Introduit dans Linux avec le noyau 2.4, implémentation LVM1
- Re-écrit en utilisant le driver Device Mapper dans le noyau 2.6, implémentation LVM2
- Alasdair Kergon est le developpeur principal actuellement, agk@redhat.com

Avantages

- Créer des systèmes de fichiers plus gros que le plus grand disque
- Cache les détails du système de stockage rendant possible des modifications transparentes aux applications
- Les données peuvent être re-arrangées sur les disques à chaud
- Prendre des snapshots de l'etat des systèmes de fichiers à un moment donné

Fonctionnalités avancées

- Clustered LVM
- High-Availability LVM (HA-LVM)
- Mirroring

Pratique: créer les volumes physiques

- Créer des Physical Volumes
  - `pvcreate /dev/sda1 /dev/sdb1 /dev/sdc1`
- Lister les Physical Volumes
  - `pvs`, `pvdisplay` et `pvscan`

Pratique: créer un groupe de volumes

- Créer un Volume Group
  - `vgcreate vg1 /dev/sda1 /dev/sdb1`
- Lister les Volume Groups
  - `vgs`, `vgdisplay` et `vgscan`

Pratique: créer un volume logique

- Créer un Logical Volume
  - lvcreate -L 1G vg1
- Lister les Logical Volumes
  - `lvs`, `lvdisplay` et `lvscan`
- Créer un système de fichiers
  - `mkfs.ext3 /dev/vg1/lv0`

Pratique: étendre un système de fichiers

- Etendre le Volume Group tout d'abord
  - `vgextend vg1 /dev/loop2`
- Etendre le Logical Volume
  - `lvextend -l 146 /dev/vg1/lv0`
- Et finalement, étendre le système de fichiers
  - `resize2fs /dev/vg1/lv0`

Pratique: remplacer un disque

- Ajouter un nouveau disque
  - `pvcreate, vgextend`
- Libérer les extents sur l'ancien disque
  - `pvmove -i1 /dev/sdc1`
- Enlever le disque du groupe de volumes
  - `vgreduce vg1 /dev/sdc1`
- Détruire le label LVM
  - `pvremove /dev/sdc1`

Pratique: utiliser des snapshots

- Un snapshot permet de maintenir l'etat d'un volume logique au moment où il est pris

```
# lvcreate -s -l50 -n slv0 /dev/vg1/lv0
```

- Les changements subsequents du volume logique original seront copiés dans le volume supportant le snapshot (ici 50 extents)
- Les snapshots utilisent un mécanisme Copy-On-Write (COW)

Pratique: tout nettoyer

- Démonter les systèmes de fichiers
- Désactiver les logical volumes
- Détruire les logical volumes
- Détruire le volume group
- Détruire les physical volumes
- Effacer les données des disques

Le Logical Volume Manager dans Linux

Imed Chihi, imed.chihi@gmail.com
Séminaire CLLFST
Faculté des Sciences de Tunis, Novembre 2008

2

C'est quoi “LVM”?

- Couche d'abstraction en dessus des disques physiques
- Composante de base: Physical Volume (volume physique)
- Un Physical Volume est typiquement une partition ou un disque entier avec un label dans les premiers secteurs
- Des Physical Volumes sont regroupés en un Volume Group

LVM veut dire Logical Volume Manager.

Il y a des technologies de stockage differentes comme: IDE, SATA, SCSI, Fibre Channel et iSCSI. Ces technologies necessitent une gestion et des traitements differents.

Utiliser le concept de Physical Volume permet un usage uniforme de toutes les ressources de stockage avec une nomenclature unique.

3

C'est quoi “LVM”?

- Un Volume Group est partitionné en des Logical Volumes
- Un Logical Volume dans un Volume Group est comme une partition sur un disque
- Les Logical Volumes peuvent être utilisés pour créer des systemes de fichiers
- L'unité d'allocation est l'extent: Logical Extent et Physical Extent
- 4 MB par extent par défaut

Le physical extent (PE) est l'unité d'allocation pour les physical volumes et les volume groups. Le logical extent (LE) est l'unité d'allocation des volumes logiques. La taille du logical extent doit correspondre à celle du physical extent. Par analogie, le secteur (512 B) est l'unité d'allocation des disques et des partitions physiques.

LVM gère donc *m* logical volumes distribués de manière transparente sur *n* physical volumes.

Implémentation

- Composante noyau: module `dm_mod`, `dm_snapshot` et autres
- Composante userspace: `libdevmapper` et suite d'outils pv*, lv*, vg*, etc.
- Version statique dans `/sbin/lvm.static`
- Fichiers de configuration sous `/etc/lvm/`

5

Historique

- LVM a existé depuis longtemps sur les plateformes Unix
- Introduit dans Linux avec le noyau 2.4, implémentation LVM1
- Re-écrit en utilisant le driver Device Mapper dans le noyau 2.6, implémentation LVM2
- Alasdair Kergon est le developpeur principal actuellement, agk@redhat.com

LVM peut ressembler au systeme RAID dans certains aspects mais l'objective de LVM est plutot la flexibilite alors que RAID a pour objectif la fiabilite et la performance.

Les grandes installations utilisent souvent LVM en dessus de disques en RAID pour avoir et la performance/disponibilite de RAID et la flexibilite de gestion de LVM.

Avantages

- Créer des systèmes de fichiers plus gros que le plus grand disque
- Cache les détails du système de stockage rendant possible des modifications transparentes aux applications
- Les données peuvent être re-arrangées sur les disques à chaud
- Prendre des snapshots de l'etat des systèmes de fichiers à un moment donné

Fonctionnalités avancées

- Clustered LVM
- High-Availability LVM (HA-LVM)
- Mirroring

8

Pratique: créer les volumes physiques

- Créer des Physical Volumes
  - `pvcreate /dev/sda1 /dev/sdb1 /dev/sdc1`
- Lister les Physical Volumes
  - `pvs`, `pvdisplay` et `pvscan`

8

Vous pouvez tester LVM sans avoir de veritables disques physiques. Vous pouvez utiliser les “loop devices” de Linux pour tester les outils LVM. Un loop device est etabli en dessus d'un fichier ordinaire. Il est presente aux applications comme un peripherique de type block souvent appele /dev/loop0, /dev/loop1, etc.

Creer le fichiers qui vont servir de disques virtuels

```
# dd if=/dev/zero of=pv0 bs=1M count=200
# dd if=/dev/zero of=pv1 bs=1M count=200
# dd if=/dev/zero of=pv2 bs=1M count=200
# dd if=/dev/zero of=pv3 bs=1M count=200
```

Associer les fichiers a des "loop devices"

```
# losetup /dev/loop0 pv0
# losetup /dev/loop1 pv1
# losetup /dev/loop2 pv2
# losetup /dev/loop3 pv3
# losetup -a
/dev/loop0: [0806]:49153 (pv0)
/dev/loop1: [0806]:49154 (pv1)
/dev/loop2: [0806]:49155 (pv2)
/dev/loop3: [0806]:49156 (pv3)
```

Creer des volumes physiques

```
# pvcreate /dev/loop0
  Physical volume "/dev/loop0" successfully created
# pvcreate /dev/loop1
  Physical volume "/dev/loop1" successfully created
# pvcreate /dev/loop2
  Physical volume "/dev/loop2" successfully created
# pvs
  PV         VG   Fmt  Attr PSize   PFree
  /dev/loop0      lvm2 --   200.00M 200.00M
  /dev/loop1      lvm2 --   200.00M 200.00M
  /dev/loop2      lvm2 --   200.00M 200.00M
```

9

Pratique: créer un groupe de volumes

- Créer un Volume Group
  - `vgcreate vg1 /dev/sda1 /dev/sdb1`
- Lister les Volume Groups
  - `vgs`, `vgdisplay` et `vgscan`

Creer un groupe de volumes sur 2 volumes physiques uniquement

```
# vgcreate vg1 /dev/loop0 /dev/loop1
  Volume group "vg1" successfully created
```

Afficher l'etat du groupe de volumes

```
# vgs
  VG   #PV #LV #SN Attr   VSize   VFree
  vg1    2   0   0 wz--n- 392.00M 392.00M
```

Voir la page de manuel de vgs(8) pour une explication de la signification des attributs w, z et n.

```
# vgdisplay
  --- Volume group ---
  VG Name               vg1
  System ID
  Format                lvm2
  Metadata Areas        2
  Metadata Sequence No  1
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                0
  Open LV               0
  Max PV                0
  Cur PV                2
  Act PV                2
  VG Size               392.00 MB
  PE Size               4.00 MB
  Total PE              98
  Alloc PE / Size       0 / 0
  Free  PE / Size       98 / 392.00 MB
  VG UUID               FZvILm-ahFd-9bVt-XXtv-mXVp-Q5jh-SFYbV3
```

10

Pratique: créer un volume logique

- Créer un Logical Volume
  - lvcreate -L 1G vg1
- Lister les Logical Volumes
  - `lvs`, `lvdisplay` et `lvscan`
- Créer un système de fichiers
  - `mkfs.ext3 /dev/vg1/lv0`

10

vgdisplay nous dit que le groupe de volumes contient 98 extents physiques. Creer un volume logique de 98 extents logiques:

```
# lvcreate -l 98 -n lv0 vg1
  Logical volume "lv0" created
# lvs
  LV   VG   Attr   LSize   Origin Snap%  Move Log Copy%
  lv0  vg1  -wi-a- 392.00M
# lvdisplay
  --- Logical volume ---
  LV Name                /dev/vg1/lv0
  VG Name                vg1
  LV UUID                0Cklyq-YvUS-0dnL-rkQJ-LPAj-Ikyx-JM7w5p
  LV Write Access        read/write
  LV Status              available
  # open                 0
  LV Size                392.00 MB
  Current LE             98
  Segments               2
  Allocation             inherit
  Read ahead sectors     0
  Block device           253:0
```

Ce volume logique peut etre utilise pour mettre dessus un systeme de fichiers par exemple:

```
# mkfs.ext3 /dev/vg1/lv0
# mkdir /mnt/lvm
# mount /dev/vg1/lv0 /mnt/lvm
```

Le nom `/dev/vg1/lv0` est en fait un lien symbolique vers le veritable nom Device Mapper qui est ici `/dev/mapper/vg1-lv0`

11

Pratique: étendre un système de fichiers

- Etendre le Volume Group tout d'abord
  - `vgextend vg1 /dev/loop2`
- Etendre le Logical Volume
  - `lvextend -l 146 /dev/vg1/lv0`
- Et finalement, étendre le système de fichiers
  - `resize2fs /dev/vg1/lv0`


```
# vgextend vg1 /dev/loop2
  Volume group "vg1" successfully extended

# vgdisplay | grep PE
  PE Size               4.00 MB
  Total PE              147
  Alloc PE / Size       98 / 392.00 MB
  Free  PE / Size       49 / 196.00 MB

# lvextend -l 146 /dev/vg1/lv0
  Extending logical volume lv0 to 584.00 MB
  Logical volume lv0 successfully resized

# df -h /mnt/lvm
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/vg1-lv0   380M   11M  350M   3% /mnt/lvm

# resize2fs /dev/vg1/lv0
resize2fs 1.40.4 (31-Dec-2007)
Filesystem at /dev/vg1/lv0 is mounted on /mnt/lvm; on-line resizing required
old desc_blocks = 2, new_desc_blocks = 3
Performing an on-line resize of /dev/vg1/lv0 to 598016 (1k) blocks.
The filesystem on /dev/vg1/lv0 is now 598016 blocks long.

# df -h /mnt/lvm/
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/vg1-lv0   566M   11M  527M   2% /mnt/lvm
```

Les reductions se font dans l'ordre inverse mais elles ne sont pas possibles a chaud. Il faut demonter le systeme de fichiers avant de le retrecir.

12

Pratique: remplacer un disque

- Ajouter un nouveau disque
  - `pvcreate, vgextend`
- Libérer les extents sur l'ancien disque
  - `pvmove -i1 /dev/sdc1`
- Enlever le disque du groupe de volumes
  - `vgreduce vg1 /dev/sdc1`
- Détruire le label LVM
  - `pvremove /dev/sdc1`

12

On peut avoir besoin de remplacer un disque dans une matrice LVM pour maintenance ou pour remplacement.

Preparer le nouveau disque de remplacement
```
# pvcreate /dev/loop3
```

L'ajouter au groupe de volumes
```
# vgextend vg1 /dev/loop3
```

Liberer les extents sur le disque a remplacer
```
# pvmove -i1 /dev/loop2
  /dev/loop2: Moved: 14.6%
  /dev/loop2: Moved: 22.9%
  /dev/loop2: Moved: 25.0%
  /dev/loop2: Moved: 31.2%
  /dev/loop2: Moved: 41.7%
  /dev/loop2: Moved: 43.8%
  /dev/loop2: Moved: 50.0%
  /dev/loop2: Moved: 60.4%
  /dev/loop2: Moved: 70.8%
  /dev/loop2: Moved: 79.2%
  /dev/loop2: Moved: 83.3%
  /dev/loop2: Moved: 95.8%
  /dev/loop2: Moved: 100.0%
```

Enlever le disque du groupe de volumes
```
# vgreduce vg1 /dev/loop2
  Removed "/dev/loop2" from volume group "vg1"
# pvremove /dev/loop2
  Labels on physical volume "/dev/loop2" successfully wiped
```

Detruire les donnees sur le disque avant de le jeter:
```
# shred -v /dev/loop2
```

13

Pratique: utiliser des snapshots

- Un snapshot permet de maintenir l'etat d'un volume logique au moment où il est pris

```
# lvcreate -s -l50 -n slv0 /dev/vg1/lv0
```

- Les changements subsequents du volume logique original seront copiés dans le volume supportant le snapshot (ici 50 extents)
- Les snapshots utilisent un mécanisme Copy-On-Write (COW)

13

Nous avons maintenant notre groupe de volumes avec 3 volumes physiques. Nous allons re-inserer /dev/loop2 pour tester les snapshots maintenant:

```
# pvcreate /dev/loop2
# vgextend vg1 /dev/loop2
```

Creer un snapshot utilisant un espace de copie de 50 extents (200MB). Cet espace doit etre capable de contenir le volumes des differences entre lv0 et sont snapshot slv0:

```
# modprobe dm_snapshot
# lvcreate -s -l50 -n slv0 /dev/vg1/lv0
# mkdir /mnt/slvm
# mount /dev/vg1/slv0 /mnt/slvm
```

Essayez maintenant de changer le contenu de /mnt/lvm en ajoutant et/ou en effacant des fichiers par exemple, tout en gardant un oeil sur l'occupation du snapshot (champs “Allocated to snapshot”) avec:

```
# lvdisplay
```

Notez que /mnt/slvm garde le meme contenu independament des changements apportees dans /mnt/lvm, et vice versa. Ceci est tres utile pour prendre des backups consistents: un administrateur peut momentanement (quelques secondes) demonter le systeme de fichier, prendre un snapshot du volume logique et remonter le systeme de fichiers. Les utilisateurs peuvent continuer a utiliser le systeme de fichiers pendant que le backup se fait depuis le snapshot.

Notez aussi que si le volume supportant le snapshot (les 50 extents ici) devient plein alors LVM va automatiquement disactiver le snapshot, puisqu'il devient inutile.

Puisqu'il est toujours possible de modifier le contenu du snapshot sans affecter le volume original, ceci peut servir a faire des tests sur un jeu de donnees avec la possibilite de les restituer dans un etant initial connu comme correct.

Pratique: tout nettoyer

- Démonter les systèmes de fichiers
- Désactiver les logical volumes
- Détruire les logical volumes
- Détruire le volume group
- Détruire les physical volumes
- Effacer les données des disques

Demonter les systemes de fichiers
```
# umount /mnt/lvm
# umount /mnt/slvm
```

Desactiver les volumes logiques
```
# lvchange -a n /dev/vg1/slv0
# lvchange -a n /dev/vg1/lv0
```

Detruire les volumes logiques
```
# lvremove /dev/vg1/slv0
# lvremove /dev/vg1/lv0
```

Detruire le groupe de volumes
```
# vgremove vg1
```

Detruire les volumes physiques
```
# pvremove /dev/loop0 /dev/loop1 /dev/loop2 /dev/loop3
```

Detruire le contenu des disques de maniere sure
```
# shred -v /dev/loop0
# shred -v /dev/loop1
# shred -v /dev/loop2
# shred -v /dev/loop3
```

Dissocier les loop devices
```
# losetup -d /dev/loop0
# losetup -d /dev/loop1
# losetup -d /dev/loop2
# losetup -d /dev/loop3
```

Effacer les fichiers
```
# rm pv0 pv1 pv2 pv3
```