debian, système de fichiers, archive

Raison de l'abandon
J'ai plusiers fois rencontré des problèmes à l'occasion du remplacement d'un disque (pour finir par devoir recommencer à zéro et restaurer les données). Le système LVM me semble toutefois mature et efficace mais réservé à ceux qui passent leur vie sur Linux alors que ma vie professionnelle est orientée sur un autre système d'exploitation... Si l'on maîtrise LVM, même les groupes de volumes ou les volumes logiques peuvent être restaurés ce qui permet de limiter au maximum le risque d'y laisser des plumes 😉.

Késako ?

Logical Volume Manager (ou *LVM* pour les intimes) permet de regrouper des partitions ou *volumes physiques* (qui peuvent ou non être sur le même disque) dans un *volume logique* ou *groupe de volumes*.

Mise en place

Préparation du système

```
# aptitude install lvm2
```

Pffft... Purée c'est compliqué... 🤣

Préparation des disques

Le nouveau volume logique dont l'installation est décrite ici va utiliser 2 disques de 3To Western Digital Red « NAS Hard Drive » SATA III WD30EFRX utilisant des secteurs de 4Ko ce qui obligeait à une époque à aligner manuellement le premier secteur logique de chaque partition sur un secteur physique ce qui était loin d'être simple.

Partitionnement

Sous Debian 7 « *Wheezy* », LVM n'a pas besoin que les partitions possèdent un système de fichier (c'était peut être possible avant mais c'était méconnu et je suis passé à côté). La seule contrainte dans ce cas étant que les disques ne peuvent être attribués à un volume logique qu'en entier et pas être partagés entre plusieurs volumes logiques. Comme c'est l'objectif ici pas de soucis.

L'avantage de ne pas attribuer de système de fichier aux partitions est de ne pas avoir à prendre le moindre risque quand au choix du type de partition (GPT indispensable pour les disques de plus de 2To alors que LVM requiert théoriquement des partitions au format éponyme dédié).

Donc il suffit d'écrire sur les disque une table de partition :

```
# fdisk /dev/sdb
```

```
Attention : la taille du disque est 3.0 To (3000592982016 octets).
Ce format de table de partitions DOS ne peut pas être utilisé
sur des disques pour des volumes plus grand (2199023255040 octets) et
secteurs de 512 octets. Utiliser parted(1) et le format de table
de partitions à GUID (GPT).


Le périphérique dispose d'une taille de secteur logique plus
petite que la taille de secteur physique. Ajuster la limite de
taille à celle de secteur physique (ou optimale en terme d'E/S)
est conseillé, sinon les performances risquent d'être affectées.
```

```
Commande (m pour l'aide): m
Commande d'action
   a   basculer l'indicateur d'amorçage
   b   éditer l'étiquette BSD du disque
   c   basculer l'indicateur de compatibilité DOS
   d   supprimer la partition
   l   lister les types de partitions connues
   m   afficher ce menu
   n   ajouter une nouvelle partition
   o   créer une nouvelle table vide de partitions DOS
   p   afficher la table de partitions
   q   quitter sans enregistrer les changements
   s   créer une nouvelle étiquette vide pour disque de type Sun
   t   modifier l'identifiant de système de fichiers d'une partition
   u   modifier les unités d'affichage/saisie
   v   vérifier la table de partitions
   w   écrire la table sur le disque et quitter
   x   fonctions avancées (pour experts seulement)

Commande (m pour l'aide): n
Type de partition :
   p   primaire (0 primaire(s), 0 étendue(s), 4 libre(s))
  e   étendue
Sélection (p par défaut) : p
Numéro de partition (1-4, par défaut 1):
Utilisation de la valeur par défaut 1
Premier secteur (2048-4294967295, par défaut 2048):
Utilisation de la valeur par défaut 2048
Dernier secteur, +secteurs or +taille{K,M,G} (2048-4294967294, par défaut 4294967294):
Utilisation de la valeur par défaut 4294967294

Commande (m pour l'aide): w
La table de partitions a été altérée.

Appel de ioctl() pour relire la table de partitions.
Synchronisation des disques.
```

Il faut ensuite procéder à la même opération sur le second disque qui va rejoindre le volume logique (*/dev/sdc* ici).

Initialisation des volumes physiques

Pour que LVM puisse utiliser une partition, il faut l'initialiser en tant que volume physique LVM. Puisque qu'ici il n'y a qu'une partition par disque :

```
# pvcreate /dev/sdb1
```

Opération a répéter là aussi sur le second disque.

Groupe de volumes

L'opération est très simple et consiste donc a créer un groupe de volumes en lui attribuant un nom (ici *vgdata*) et en y intégrant un volume physique précédemment initialisé :

```
# vgcreate vgdata /dev/sdb1
```

```
Volume group "vgdata" successfully created
```

Ajout d'espace

```
vgextend vgdata /dev/sdc1
```

Volume logique

Il faut commencer par créer le volume logique (*lvdata*) et lui attribuer une partie plus ou moins importante de l'espace disponible dans le groupe de volumes (*vgdata*) :

```
lvcreate --name lvdata -l 100%FREE vgdata
```

```
Logical volume "lvdata" created
```

lvcreate propose de nombreuses syntaxes différentes en ce qui concerne la portion de l'espace à allouer...

Puis il faut ensuite formater ce volume, par exemple :

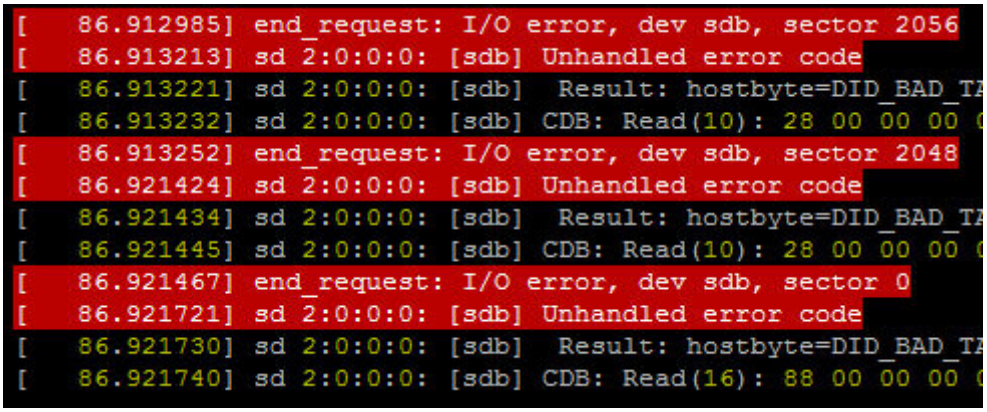
```
mkfs.ext4 /dev/vgdata/lvdata
```

En cas de corruption

Symptômes

Dans mon cas, le premier symptôme a été un problème d'accès aux partages Samba après un reboot qui cachait tout simplement un problème d'accès au volume logique contenant les données.

J'ai évidement commencé par chercher du côté de Samba puis à force de chercher dans les logs, j'ai trouvé un paquets d'erreurs au niveau du système de fichiers de l'un des disques du volume logique (dans */var/log/kern.log*) :



L'étape suivante a été de tenter de forcer une vérification au démarrage avec les commandes suivantes :

```
# touch /forcefsck
# reboot
```

N'ayant pas d'écran sur cette machine j'ignore si un éventuel message a donné des indications sur le déroulement (ou plutôt le non-déroulement de l'opération) et la marche à suivre, mais, après le démarrage, le volume logique était à nouveau monté et les partages étaient à nouveau accessibles au bout de 3 ou 4 minutes (curieuse aussi cette latence mais je n'ai pas eu le temps sur le coup de chercher l'explication et je ne l'ai pas trouvée par la suite). Avec le recul, le temps très court du redémarrage aurait dû me mettre la puce à l'oreille

Forcer une vérification complète

Il faut d'abord retrouver le nom du (ou des) groupes de volumes (*VG*) et du (ou des) volumes logiques (*LV*) :

```
# lvm lvs
```

Ce qui va renvoyer quelque chose comme ceci :

```
LV     VG     Attr     LSize Pool Origin Data%  Move Log Copy%  Convert
lvdata vgdata -wi-ao-- 5,46t
```

Comme il va falloir démonter le volume en question, il est préférable d'arrêter les service correspondants (uniquement Samba dans mon cas) :

```
# /etc/init.d/samba stop
```

On peut ensuite démonter le volume et lancer la vérification complète (le paramètre de chacune des commandes dépends évidement de ce qu'a retourné la commande *lvm lvs* précédemment) :

```
umount /dev/vgdata/lvdata
e2fsck -y /dev/vgdata/lvdata
```

Et le retour très rapide a été le suivant :

```
/dev/vgdata/lvdata : propre, 577829/183144448 fichiers, 701525008/1465131008 blocs
```

Mouais... Mon œil!
Donc même si le commutateur *-y* est supposé tenter de détecter et réparer automatiquement toute corruption du système de fichier, il est manifestement insuffisant (et c'est probablement le même genre de vérification que la commande *touch /forcefsck* déclenche). Donc on insiste :

```
e2fsck -yfv /dev/vgdata/lvdata
```

Les commutateurs utilisés :

- y : est donc supposé tenter de détecter et réparer automatiquement toute corruption du système de fichier
- f : force une vérification approfondie même si le système semble propre
- v : force le mode verbeux afin d'obtenir un maximum d'infos en retour

On peut aussi ajouter le commutateur *c* pour forcer une vérification de chaque secteur mais c'est extrêmement long.

Et voici le résultat obtenu, cette fois après un long moment :

```
e2fsck 1.42.5 (29-Jul-2012)
Passe 1 : vérification des i-noeuds, des blocs et des tailles
Passe 2 : vérification de la structure des répertoires
Passe 3 : vérification de la connectivité des répertoires
/lost+found n'a pas été trouvé. Créer ? oui

Passe 4 : vérification des compteurs de référence
Passe 5 : vérification de l'information du sommaire de groupe

/dev/vgdata/lvdata: ***** LE SYSTÈME DE FICHIERS A ÉTÉ MODIFIÉ *****

      577830 inodes used (0.32%, out of 183144448)
        3916 non-contiguous files (0.7%)
         166 non-contiguous directories (0.0%)
             # of inodes with ind/dind/tind blocks: 0/0/0
             Extent depth histogram: 569736/1913/7
   701525009 blocks used (47.88%, out of 1465131008)
           0 bad blocks
          84 large files

      511618 regular files
       59887 directories
           0 character device files
           0 block device files
           1 fifo
           0 links
        6308 symbolic links (6158 liens symboliques rapides)
           6 sockets
------------
      577820 files
```

Le mode verbeux n'est pas si causant que çà, il y avait vraiment un paquet d'erreurs dans le log. Qu'a cela ne tienne...

```
reboot
```

Bingo ! Les partages Samba sont instantanément dispos et surtout, plus aucune erreur dans le log.

Seule chose qui me chagrine : rien dans *lost+found* et à priori pas de données perdues si je compare aux sauvegardes (ou aux dossiers d'origine selon les cas). Voir ci-dessous pour la suite des aventures de ce disque (bon ça ne fait jamais plaisir mais au moins il y a une certaine logique).

Et paf ! le disque

Donc au premier reboot après la “réparation” tout semblait nickel, sauf que les erreurs sont revenues au reboot suivant .

Il faut donc vite passer au remplacement du disque tant qu'il marche encore sur 3 pattes...

Donc pour faire fonctionner le disque (même partiellement) donc j'ai réutilisé ces commandes:

```
# touch /forcefsck
# reboot
```

Quand on a plusieurs disques identiques et qu'il faut en identifier un, il faut récupérer son numéro de série :

```
# smartctl -i /dev/sdb
```

```
smartctl 5.41 2011-06-09 r3365 [i686-linux-3.2.0-4-686-pae] (local build)
Copyright (C) 2002-11 by Bruce Allen, http://smartmontools.sourceforge.net

=== START OF INFORMATION SECTION ===
Device Model:     WDC WD30EFRX-68AX9N0
Serial Number:    WD-WCC1T0763637
LU WWN Device Id: 5 0014ee 2b30390ae
Firmware Version: 80.00A80
User Capacity:    3 000 592 982 016 bytes [3,00 TB]
Sector Sizes:     512 bytes logical, 4096 bytes physical
Device is:        Not in smartctl database [for details use: -P showall]
ATA Version is:   9
ATA Standard is:  Exact ATA specification draft version not indicated
Local Time is:    Fri May 15 03:04:30 2015 CEST
SMART support is: Available - device has SMART capability.
SMART support is: Enabled
```

Pour avoir accès à la commande *smartctl*, il faudra que le paquet *smartmontools* soit installé.

Et il faut aussi connaître le nom du volume logique :

```
# lvm lvs
```

```
LV     VG     Attr     LSize Pool Origin Data%  Move Log Copy%  Convert
lvdata vgdata -wi-ao-- 5,46t
```

J'ai donc ensuite ajouté mon disque (à chaud dans mon cas) et vérifié qu'il était bien détecté par le système en jetant simplement un œil au contenu de */dev* (il faut évidement avoir une idée de son contenu ou avoir pensé à regarder avant 😛).

Il ne reste plus qu'à préparer le disque pour l'utilisation avec LVM, à l'ajouter au volume logique et y déplacer les données contenues sur le disque défectueux et voici le scénario :

```
# pvcreate /dev/sdd
# vgextend vgdata /dev/sdd
# pvmove /dev/sdb /dev/sdd
# vgreduce vgdata /dev/sdb
# pvremove /dev/sdb
```

Sauf que ça coince dès la commande *vgextend* :

```
/dev/vgdata/lvdata: read failed after 0 of 4096 at 0: Erreur d'entrée/sortie
/dev/vgdata/lvdata: read failed after 0 of 4096 at 4096: Erreur d'entrée/sortie
Couldn't find device with uuid dbUOOk-J3bJ-3g6l-9fyx-a3Ry-7GT2-6gATBQ.
Cannot change VG vgdata while PVs are missing.
Consider vgreduce --removemissing.
```

Le but étant de récupérer ce qui est à récupérer sur le disque en question, il faut creuser la question en regardant l'état des volumes physiques :

```
# pvs
```

```
/dev/vgdata/lvdata: read failed after 0 of 4096 at 0: Erreur d'entrée/sortie
/dev/vgdata/lvdata: read failed after 0 of 4096 at 4096: Erreur d'entrée/sortie
Couldn't find device with uuid dbUOOk-J3bJ-3g6l-9fyx-a3Ry-7GT2-6gATBQ.
PV             VG     Fmt  Attr PSize PFree
/dev/sdc1      vgdata lvm2 a--  2,73t    0
/dev/sdd              lvm2 a--  2,73t 2,73t
unknown device vgdata lvm2 a-m  2,73t    0
```

C'est l'attribut *m* (pour *missing* qui coince)...

Il est théoriquement possible de réinitialiser cet attribut avec cette commande :

```
# vgextend --restoremissing vgdata /dev/sdb1
```

```
/dev/vgdata/lvdata: read failed after 0 of 4096 at 0: Erreur d'entrée/sortie
/dev/vgdata/lvdata: read failed after 0 of 4096 at 4096: Erreur d'entrée/sortie
Couldn't find device with uuid dbUOOk-J3bJ-3g6l-9fyx-a3Ry-7GT2-6gATBQ.
WARNING: The PV /dev/sdb1 is still missing.
No PV has been restored.
```

Bon ben voilà... La messe est dite pour ce disque 😕. Malgré mes recherches, je n'ai pas trouvé comment passer outre donc il est temps de retirer le disque du volume logique :

```
# vgreduce --removemissing vgdata
```

```
/dev/vgdata/lvdata: read failed after 0 of 4096 at 0: Erreur d'entrée/sortie
/dev/vgdata/lvdata: read failed after 0 of 4096 at 4096: Erreur d'entrée/sortie
Couldn't find device with uuid dbUOOk-J3bJ-3g6l-9fyx-a3Ry-7GT2-6gATBQ.
WARNING: Partial LV lvdata needs to be repaired or removed.
There are still partial LVs in VG vgdata.
To remove them unconditionally use: vgreduce --removemissing --force.
Proceeding to remove empty missing PVs.
```

Pas le choix, il faut utiliser le commutatuer *–force* pour retirer ce fichu disque:

```
# vgreduce --removemissing vgdata --force
```

```
/dev/vgdata/lvdata: read failed after 0 of 4096 at 0: Erreur d'entrée/sortie
/dev/vgdata/lvdata: read failed after 0 of 4096 at 4096: Erreur d'entrée/sortie
Couldn't find device with uuid dbUOOk-J3bJ-3g6l-9fyx-a3Ry-7GT2-6gATBQ.
Removing partial LV lvdata.
Logical volume "lvdata" successfully removed
Wrote out consistent volume group vgdata
```

Comment ça « *Logical volume “lvdata” successfully removed* » ?! 🤐

Bon ben voilà, ça c'est fait, y'a plus qu'à recréer un volume logique tout neuf et restaurer les données. 😕 <blockquote>Braves gens ! Notre enchanteur m’informe, que d’habitude il y arrive très bien mais que là, visiblement, ça marche pas ! <cite>Kaamelott (Livre II - L'Invincible)</cite> </blockquote>

Ce fameux volume logique dont la fin funeste est décrite ci-dessus utilisait 2 disques de 3To chacun, or, pour utiliser une partition de plus de 2To, il faut une partition de type GPT qui n'est pas géré par l'outil intégré à Debian *fdisk*. Par ailleurs l'alignement des partitions sur ces disques utilisant des secteurs de 4Ko était très délicat et j'ai fait le choix de passer outre. Il est possible que la combinaison GPT/LVM ou le mauvais alignement des partition soit responsable de la suppression non désirée du volume logique... Espérons en tout cas...