

Chapitre 3

TECHNOLOGIES DE GESTION DE STOCKAGE
(DISQUES, RAID, LVM)

Gestion des disques

La gestion des disques est une fonctionnalités très importante dans l'administration système et réseau

Un bon administrateur doit pouvoir gérer l'évolution de ses besoins en espace de stockage sans devoir recourir à une coûteuse réinstallation.

Il convient donc de maîtriser les outils qui servent à manipuler des volumes RAID et LVM

RAID: pour sécuriser les données face aux pannes matérielles en dupliquant les informations

LVM: pour gérer ses données à sa guise en faisant abstraction de la taille réelle de ses disques

➔ De nouveaux périphériques de type bloc sur lesquels on pourra créer des systèmes de fichiers, ou des espaces de mémoire virtuelle, qui ne correspondent pas directement à un seul disque dur réel.

RAID

Un volume RAID (Redudent Array of Independent Disks) est constitué d'un ensemble de disques ou de partitions de disques. L'objectif est d'assurer une tolérance aux pannes, en dupliquant les données sur plusieurs disques. La performance est améliorée en permettant la lecture ou l'écriture en parallèle des données depuis ou vers les disques qui constituent le volume RAID.

Le RAID peut être matériel ou logiciel. Le RAID matériel utilise un contrôleur matériel permettant de gérer le volume RAID. Ce contrôleur RAID est doté d'un processeur spécifique, une mémoire et un logiciel embarqué (RAID firmware). Le contrôleur RAID matériel cache les caractéristiques du volume RAID au système d'exploitation. Ce dernier perçoit le volume RAID comme un disque dur classique.

En RAID logiciel, le contrôleur RAID est un composant du système d'exploitation. Le module MD (Multiple Disk) du noyau Linux est un contrôleur RAID logiciel permettant de gérer le volume RAID et offrant aux applications un seul disque dur virtuel.

LVM (Logical Volume Manager) est une technique qui crée des systèmes de fichiers sur des volumes logiques, ce qui permet par exemple de redimensionner la taille des partitions en toute transparence du point de vue des applications et sans avoir besoin de redémarrer un serveur en production.

RAID: Concept Général

Selon le type d'architecture mis en place, la technique RAID permet d'améliorer :

- soit la performance de lecture et d'écriture, en distribuant les données sur plusieurs disques, ce qui permet au contrôleur de travailler sur plusieurs disques simultanément ;
- soit la tolérance aux pannes, en dupliquant des données sur plusieurs disques, ce qui diminue les risques en cas de défaillance de l'un d'eux ;
- soit les deux.

RAID Concept Général (suite)

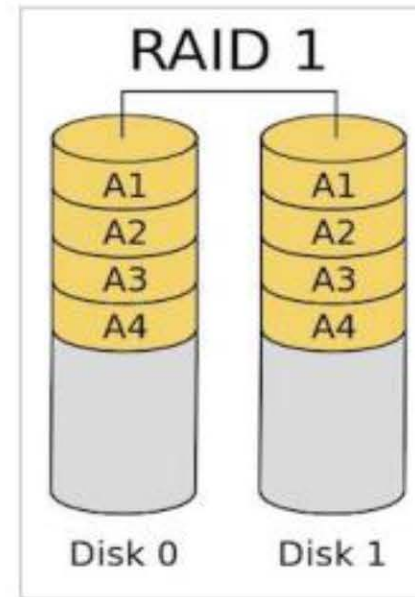
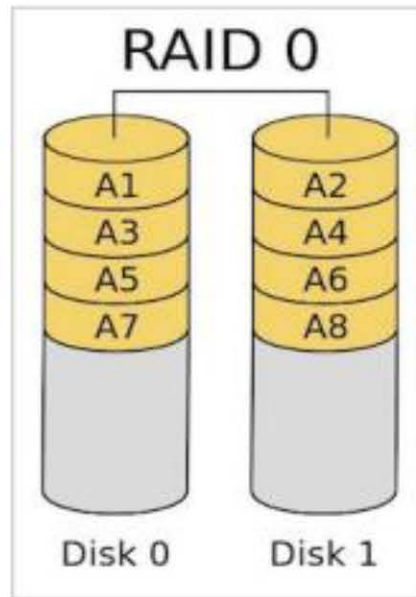
Il existe plusieurs types de RAID, appelés *niveaux*. Les plus utilisés sont *RAID 0*, *RAID 1*, *RAID 5* et *RAID 10*.

RAID 0 : il est utilisé uniquement pour améliorer les performances. Les données sont découpées en blocs (*chunk*) et ces blocs sont répartis sur plusieurs disques, ce qui diminue les temps de lecture et d'écriture.

RAID 1 : il offre une redondance des données. En effet elles sont dupliquées sur deux ou plusieurs disques. La performance de lecture augmente avec le nombre de disques du volume RAID. Ceci peut assurer une meilleure tolérance aux pannes, mais peut nuire à la performance d'écriture, car l'information doit être écrite plusieurs fois ;

RAID

Concept Général (3)



RAID

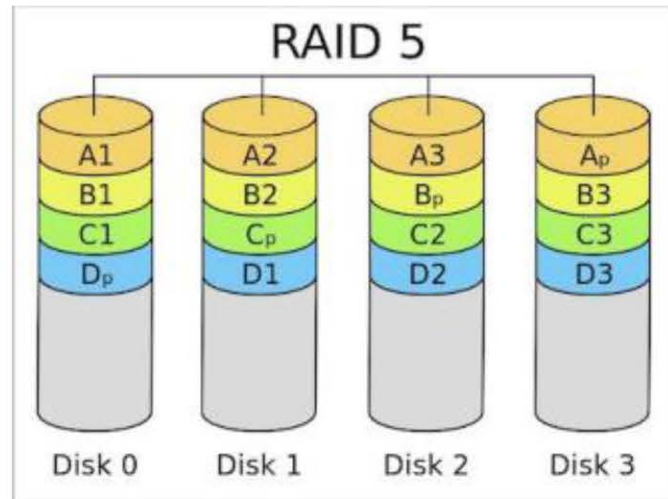
Concept Général (4)

RAID 4 : les données sont distribuées en petits blocs sur les différents disques durs. Le contrôle de parité est inscrit sur un disque spécifique. Ceci permet de créer une architecture redondante qui améliore le temps d'écriture. Si un disque tombe en panne, les données manquantes seront recalculées à partir des données utiles et des parités. RAID 4 n'est plus utilisé, il est remplacé par RAID 5.

RAID 5 : à la différence de RAID 4, les données et la parité sont elles aussi distribuées sur l'ensemble des disques

RAID

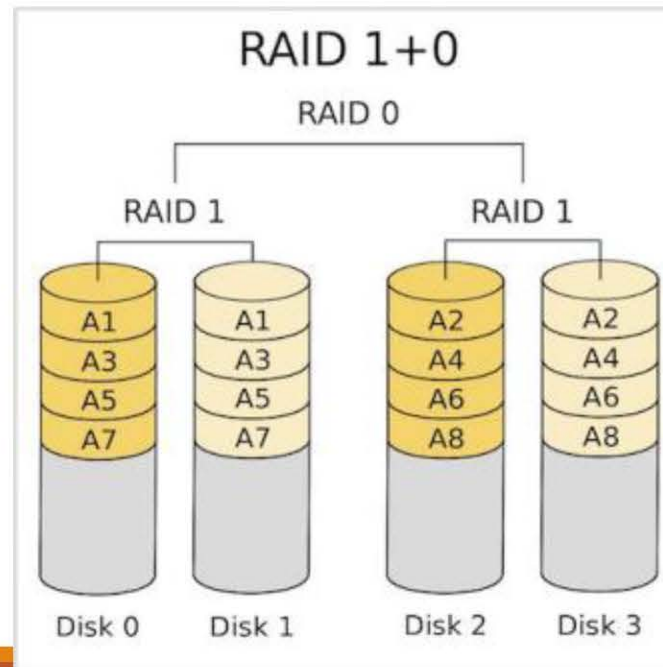
Concept Général (5)



RAID

Concept Général (6)

RAID 10 ou (1+0) : combinaison de RAID 1 et de RAID 0. C'est un RAID 0 composé de deux volumes RAID 1 (*figure 10*), ce qui offre les avantages simultanés du RAID 1 et RAID 0 : augmentation de la taille de stockage et des performances de lecture. Il faut au minimum quatre disques de stockage pour créer un volume RAID 10.



Gestion du RAID logiciel sous Linux

La plupart des distributions Linux utilisent la commande **mdadm** pour générer un volume RAID.

La commande mdadm permet de créer, contrôler et administrer le volume RAID. Elle possède sept modes de fonctionnement différents, chacun avec ses propres options.

Les volumes RAID sont configurés dans le fichier `/etc/mdadm.conf`. Ce fichier est utilisé par la commande mdadm dans le mode create pour créer le volume RAID ou pour l'initialiser après chaque démarrage. Les principaux paramètres de ce fichier sont :

- **DEVICE** : liste des disques et partitions susceptibles d'être utilisés dans le volume RAID ;
- **ARRAY** : nom du volume RAID, par exemple `/dev/md0` ou `/dev/md/00`
- **Level** : niveau RAID, par exemple `raid0`, `raid1`, `raid4` ou `raid5`
- **Devices** : liste des disques ou partitions (séparés par des virgules) qui seront utilisés pour assembler le volume RAID
- **num-devices** : nombre de disques ou partitions sous-jacents dans le volume RAID
- **spare-group** : nom du groupe de secours. Les disques ou partitions appartenant à un groupe de secours peuvent être utilisés par plusieurs volumes RAID

Gestion du RAID logiciel sous Linux

Les options du mode create sont :

- -n ou --raid-devices : nombre de disques du volume RAID ;
- -l ou --level : niveau RAID ;
- -c ou --chunk : taille en kilo-octets du bloc (*chunk*). *La valeur indiquée doit être une puissance de 2. La valeur par défaut est 64 ko ;*
- -x ou --spare-devices : nombre de disques de secours ;
- -z ou --size : taille en kilo-octets de l'espace à utiliser dans chaque disque pour les niveaux RAID 1, 4, 5 et 6. Cette taille doit être multiple de la taille d'un bloc (*chunk size*). *Par défaut elle correspond à la taille du plus petit disque (ou partition) sous-jacent du volume RAID*
- -p ou --parity : algorithme de parité utilisé. Par défaut c'est l'algorithme *left-symmetric*.

Les différents modes de fonctionnement du RAID

Mode	Option pour sélectionner le mode	Description
assemble	--assemble ou --A	Assemble les composants d'un volume RAID déjà créé. Ce mode est utile pour le dépannage ou l'activation d'un volume RAID transféré depuis un autre ordinateur. L'option <code>-s</code> ou <code>--scan</code> récupère les informations sur le volume RAID et les disques sous-jacents.
build	--build ou -B	Crée un volume RAID à partir des disques qui n'ont pas de superblocs (des métadonnées).
create	--create ou -C	Crée un nouveau volume RAID avec des superblocs (des métadonnées) dans chaque disque
manage	(par défaut)	Peut ajouter ou supprimer des disques à un volume RAID en cours d'exécution. Ceci est utile pour enlever des disques durs en panne, ajouter des disques de rechange ou remplacer des disques.
misc	(par défaut)	Effectue des opérations spécifiques sur un disque sous-jacent, telles que l'effacement de superblocs ou du paramètre en lecture seule.
monitor	--follow, --monitor ou -F	Surveille un volume RAID. Peut par exemple automatiser l'envoi des alertes aux administrateurs lorsque le volume RAID rencontre des erreurs. Peut aussi exécuter automatiquement des commandes, comme enlever et réinsérer un disque pour tenter de corriger une défaillance non fatale.
grow	--grow ou -G	Change le nombre ou la taille des blocs ou des disques sous-jacents.

EXEMPLES

Créer un volume RAID niveau 5, /dev/md1, composé des partitions /dev/sda1, /dev/sdb1 et /dev/sdc1

```
# mdadm --create /dev/md1 --raid-devices=3 /dev/sda1 /dev/sdb1 /dev/sdc1 --level=5
```

Commande identique à la précédente avec des options courtes :

```
# mdadm -C /dev/md1 -n3 /dev/sda1 /dev/sdb1 /dev/sdc1 -l5
```

Créer un volume RAID à partir des informations de configuration du fichier /etc/mdadm.conf. La commande mdadm prend comme argument le nom du volume RAID :

```
# mdadm --create /dev/md0
```

Créer deux volumes RAID, /dev/md0 et /dev/md1. Le premier volume RAID est créé avec un disque de secours /dev/sdd1. Les deux volumes RAID utilisent le même groupe de secours mongroupe (défini par l'option --spare-group), ce qui leur permet de partager les disques de secours. Cela signifie que si un disque du volume RAID /dev/md1 tombe en panne, il sera remplacé automatiquement par le disque /dev/sdd1 :

```
# mdadm --create /dev/md0 --raid-devices=3 /dev/sda1 /dev/sdc1 -x /dev/sdd1 --level=1 --spare-group= mongroupe
```

```
# mdadm --create /dev/md1 --raid-devices=2 /dev/sda2 /dev/sdc2 --level=1 --spare-group= mongroupe
```


LVM

Concept Général

LVM, ou *Logical Volume Manager*, consiste à *abstraire les volumes logiques* des disques physiques.

LVM permet de modifier dynamiquement un volume logique, en toute transparence du point de vue des applications.

On peut ainsi par exemple ajouter de nouveaux disques, migrer les données dessus et récupérer les anciens disques ainsi libérés, sans démonter le volume.

LVM manipule trois types de volumes pour atteindre cette flexibilité:

- Volumes physiques
- Volumes groupes
- Volumes logiques

LVM offre plusieurs avantages :

- avoir une utilisation et une allocation efficaces de l'espace de stockage, puisque les volumes logiques sont répartis sur plusieurs disques physiques ;
- augmenter et réduire la taille des volumes logiques sans risque d'interrompre des services du système ni de perdre des données ;
- on peut prendre des instantanés (*snapshots*) sur le système de fichiers. Ces instantanés servent à sauvegarder et restaurer les données.

Gestion du LVM sous Linux

La configuration LVM sous Linux est faite en trois étapes :

- création et initialisation des volumes physiques ;
- ajout des volumes physiques à un groupe de volumes ;
- création des volumes logiques au sein du groupe de volumes.

Les commandes LVM commencent par deux lettres qui reflètent le niveau d'abstraction LVM :

- les commandes pv manipulent les volumes physiques ;
- les commandes vg manipulent les groupes de volumes ;
- les commandes lv manipulent les volumes logiques.

Les principales commandes de gestion des volumes physiques sont :

- pvcreate : initialise un périphérique (partition, disque ou volume RAID) comme un volume physique pour une utilisation par LVM ;
- pvdisplay : affiche des informations détaillées sur un volume physique, y compris le nom du groupe de volumes auquel il appartient et sa taille ;
- pvscan : analyse les partitions de disque à la recherche des périphériques de blocs contenant des volumes physiques ;
- pvck : contrôle la cohérence du volume physique ;
- pvs : affiche des informations sommaires sur les volumes physiques.

Gestion du LVM sous Linux

Les principales commandes de gestion des groupes de volumes sont :

- `vgcreate` : crée un groupe de volumes ;
- `vgchange` : modifie certains attributs d'un groupe de volumes, par exemple pour l'activer ou le désactiver ;
- `vgdisplay` : affiche les caractéristiques détaillées d'un volume logique. L'option `-v` permet de visualiser la liste des volumes logiques et des volumes physiques de chaque groupe de volumes ;
- `vgscan` : analyse le système pour rechercher des groupes de volumes ;
- `vgextend` : ajoute des volumes physiques à un groupe de volumes existant ;
- `vgreduce` : supprime un ou plusieurs volumes physiques d'un groupe de volumes ;
- `vgremove` : supprime un groupe de volumes ;
- `vgrename` : renomme un groupe de volumes ;
- `vgs` : affiche des informations sommaires sur les groupes de volumes

Les principales commandes de gestion des volumes logiques :

- `lvdisplay` : affiche des informations détaillées sur les volumes logiques ;
- `lvcreate` : crée un volume logique ;

Les paramètres d'accès aux disques

Les serveurs connectent généralement leurs disques à travers une interface SCSI (*Small Computer Systems Interface*). Sur les ordinateurs de bureau et les ordinateurs portables c'est l'interface IDE (*Integrated Drive Electronics*) qui est utilisée. Il existe plusieurs pilotes pour ces différents types de disques et d'interfaces.

L'interface du disque dur sert à transmettre les données entre le disque et son contrôleur qui les transmet alors au système. Il existe plusieurs interfaces qui se distinguent par des contrôleurs, des connecteurs, des débits et par le nombre de disques gérés.

PATA

L'interface PATA (*Parallel Advanced Technology Attachment*), permet de relier les périphériques de masse (disques, lecteurs de cédéroms ...) à la carte mère par des câbles plats, souples et composés de 40 ou 80 broches.

On peut connecter deux disques (un maître et un esclave) sur le même câble. L'interface PATA est aussi connue sous le nom IDE (IDE) ou *Enhanced IDE (eIDE)*.

SATA

L'interface SATA (*Serial ATA*) succède à l'interface PATA. Elle permet un débit de transfert plus élevé. La méthode de transfert des données est en série. Chaque disque dur est relié à son propre contrôleur via son propre câble ainsi chaque disque bénéficie de la totalité de la bande passante. Les câbles utilisés sont beaucoup plus minces que les câbles des disques PATA, ils ne comportent que 7 fils, ce qui encombre moins les boîtiers des ordinateurs.

L'interface SATA supporte le branchement à chaud (*hot plug*). Les périphériques SATA sont internes à l'ordinateur. Une variante de l'interface, connue sous le nom de eSATA, est utilisée pour les disques durs externes.

SCSI

L'interface SCSI (*Small Computer System Interface*) permet de relier, simultanément et en série, plusieurs périphériques tels que les disques durs, les lecteurs de cédéroms, les graveurs, les scanners. etc. Selon sa version, l'interface SCSI peut prendre en charge de 8 à 16 périphériques par câble. Le contrôleur SCSI est considéré comme un périphérique à part entière, de sorte qu'on peut brancher réellement 7 ou 15 périphériques.

L'interface SCSI se présente sous la forme d'une petite carte comportant un microcontrôleur indépendant du processeur, qui permet de bien le soulager et d'augmenter les performances systèmes.

Généralement on trouve l'interface SCSI dans les serveurs et les stations de travail haut de gamme. Elle est moins utilisée sur les ordinateurs du bureau et les ordinateurs portables à cause de son coût élevé.

USB

L'interface USB (*Universal Serial Bus*) permet de relier plusieurs types de périphériques externes à un ordinateur, y compris les disques durs et les lecteurs flash. La première et la deuxième génération de l'USB sont peu performantes, mais l'USB 3.0 est nettement plus rapide.