

Création d'un volume RAID 5 logiciel avec mdadm

Author : pierrehirel

Le RAID 5 est un moyen de sécuriser ses données. Seulement, les cartes contrôleur permettant le RAID matériel ont tendance à être chères. Heureusement il existe un outil libre permettant de faire du RAID5 logiciel : mdadm.

0. Introduction

Le principe du RAID est de réunir plusieurs disques physiques en un seul volume (en anglais *array*). Il existe plusieurs niveaux de RAID, et chaque niveau permet un gain en performances et/ou en capacité de stockage par rapport à des disques durs isolés (voir [l'article de Wikipedia](#)). Dans le cas du RAID 5, la redondance des données est assurée par un contrôle de parité par l'opération **OU exclusif** (XOR) : si par exemple on souhaite écrire un 0 puis un 1, alors le bit de parité sera $0 \text{ XOR } 1 = 1$. Ainsi dans un RAID 5 pour sauvegarder deux bits d'information, il faut en stocker trois ; un volume RAID 5 doit donc être constitué d'au moins trois disques physiques (en réalité, de trois partitions, comme nous le verrons plus bas). Avec trois disques d'1 To, la capacité « utile » du volume sera de 2 To, le To restant servant à stocker les données de parité. Notons dès à présent qu'il est possible d'utiliser davantage de disques : avec quatre disques d'1 To la capacité utile est d'environ 2,66 To, et ainsi de suite. L'avantage évident d'un tel système est que si un disque vient à perdre ses données, il est possible de remplacer le disque défaillant puis de re-calculer les données manquantes grâce aux données des disques sains.

Le fonctionnement du RAID peut être assuré de trois façons différentes. Dans le cas d'un **RAID matériel**, c'est un contrôleur dédié qui effectue tous les calculs de parité et gère les volumes (y compris en cas de défaillance d'un disque, de reconstruction du volume, etc.) ; cette méthode est néanmoins coûteuse puisqu'elle requiert généralement l'achat d'une carte contrôleur RAID dédiée. Les cartes mères grand public n'offrent généralement pas de réel RAID matériel, il s'agit le plus souvent de « **fake-RAID** » : le contrôleur intégré supporte le RAID, mais c'est le CPU qui effectue tous les calculs. Enfin, il est possible d'utiliser une solution **RAID logicielle** : c'est un logiciel qui va permettre de gérer les volumes RAID.

C'est à cette dernière solution que nous allons nous intéresser ici. Plus précisément, nous allons utiliser l'outil libre *mdadm* pour créer un volume RAID 5 logiciel à partir de plusieurs disques physiques.

1. Partitionnement les disques physiques

Après cette longue introduction, passons à la pratique. Pour l'exemple, admettons que vous ayez quatre disques qui sont reconnus par votre système GNU/Linux comme *sdC*, *sdd*, *sde* et *sdf*. Il faut commencer par partitionner ces disques : en effet, *mdadm* travaille avec des **partitions** et non des disques physiques. Pour créer les partitions, les disques ne doivent pas être montés par votre distribution (et donc pas en cours d'utilisation). Les partitions peuvent être créées à l'aide de l'outil [fdisk](#) si vous en connaissez bien les commandes ; ou bien vous pouvez utiliser un outil graphique tel que [Gparted](#). Une fois les partitions créées, il est inutile de les formater individuellement : c'est le volume RAID 5 que l'on formatera dans son intégralité par la suite.

La suite de l'article continue avec cette configuration à quatre disques comme exemple -adaptez donc la suite à votre cas particulier.

2. Création du volume RAID 5 : utilisation de mdadm

Par défaut l'outil *mdadm* n'est pas installé sur la plupart des distributions, installez-le donc à l'aide de votre

gestionnaire de paquets. Je ne détaille pas cette étape qui dépend de la distribution que vous utilisez, mais ne devrait pas poser problème...

Le volume RAID 5 se crée grâce à *mdadm*, appelé en tant que super-utilisateur et avec les paramètres suivants :

```
server:~# mdadm --create --verbose /dev/md0 --level=5 --raid-devices=4 /dev/sdb1 /dev/sdc1 /dev/sdd1 /dev/sde1
```

Détaillons cette commande :

- **--create** indique que le volume RAID était inexistant auparavant et qu'on le crée ;
- **--verbose** indique à *mdadm* d'être "bavard", autrement dit d'afficher à l'écran la progression ;
- **/dev/md0** est le point de montage du volume RAID; vous pouvez bien sûr le modifier à votre convenance ;
- **--level=5** est le niveau de RAID que l'on crée, ici il s'agit bien de RAID 5 ;
- **--raid-devices=4** est le nombre de partitions constituant le volume RAID ;
- viennent ensuite les points de montage des 4 partitions qui vont constituer le volume RAID.

Il faut ensuite enregistrer ce nouveau volume auprès de *mdadm*, afin qu'il soit activé à chaque démarrage. Pour cela on appelle *mdadm* une nouvelle fois pour scanner les volumes RAID existants et les inscrire dans ce fichier :

```
server:~# mdadm --detail --scan >> /etc/mdadm/mdadm.conf
```

Le volume RAID 5 est désormais créé, mais il n'est pas encore utilisable par le système : il va falloir y créer un système de fichiers puis le monter.

3. Création du système de fichiers du volume RAID

Le volume se formate grâce à l'outil *mke2fs* :

```
server:~# mke2fs -j /dev/md0
```

L'option **-j** indique de formater la partition (ici **/dev/md0**) avec le système de fichiers ext3 ; si cette option n'est pas utilisée, alors elle sera formatée en ext2 (le défaut de *mke2fs*).

Puis pour monter le volume dans, par exemple, le dossier */mount/raid5* :

```
server:~# mount /dev/md0 /mount/raid5
```

Vous devriez avoir accès à votre volume RAID 5, sans même avoir à redémarrer la machine. Il peut être pratique de toujours monter le volume RAID 5 au démarrage du système, pour cela ajoutez dans le fichier */etc/fstab* la ligne :

```
/etc/fstab
```

```
/dev/md0 /mount/raid5 auto defaults 0 3
```

Performances du volume RAID 5

Lorsque des données doivent être écrites sur un volume RAID 5 logiciel, les bits de parité doivent être calculés par le processeur puis écrits sur le volume. Les performances en écriture sont donc limitées non seulement par les disques constituant le volume, mais également par la puissance de calcul du processeur. Lorsque des données sont lues en revanche, les performances sont généralement bonnes car les données constituant un fichier sont lues depuis plusieurs disques simultanément.

Vérifier l'intégrité du volume RAID

Vous pouvez à tout moment vérifier l'intégrité du volume en lisant le fichier `/proc/mdstat` (ceci peut se faire en tant que simple utilisateur), qui ressemblera à ça :

```
/proc/mdstat
```

```
Personalities : [raid1] [raid6] [raid5] [raid4]
md0 : active raid5 sdc1[0] sdf1[3] sde1[2] sdd1[1]
2895866452 blocks level 5, 64k chunk, algorithm 2 [4/4] [UUUU]
```

La ligne « personalities » indique juste quels types de RAID sont supportés par le système. La ligne suivante indique le nom du volume (ici **md0**), son statut (ici actif), son type (raid5), et les partitions le constituant (l'ordre peut être aléatoire et n'est pas significatif). La dernière ligne indique la taille de ce volume, et tout au bout l'intégrité des partitions, représentée par une lettre pour chaque partition. La lettre U signifie que la partition est en bon état, et le signe underscore (`_`) que la partition est défectueuse ou inaccessible, qui peut être le signe d'un disque défectueux. Ici dans notre volume constitué de quatre partitions, le sigle [UUUU] indique que toutes les partitions sont en bon état.

Sur les distributions GNU/Linux basées sur Debian, l'intégrité du volume RAID est vérifiée régulièrement par `mdadm`, par défaut le premier dimanche de chaque mois à 1h06 du matin (voyez `/etc/cron.d/mdadm`). Ce processus vérifie que tous les bits de parité sont corrects, aussi peut-il utiliser beaucoup de ressources, prendre du temps selon la taille du volume, et diminuer les performances du volume voire de toute la machine le temps de la vérification. La progression de cette tâche est inscrite en temps réel dans le fichier `/proc/mdstat`, il suffit donc d'y jeter un œil pour la suivre.

Ne jamais faire confiance à un seul système...

Le RAID 5 permet de sécuriser ses données, certes, cependant il est faux de croire qu'il est infailible. Ce système permet de récupérer les données si **un** disque est défectueux, mais dans le cas où **deux** disques lâchent en même temps alors toutes les données sont perdues. Ce scénario semble peu probable, cependant il faut considérer que souvent plusieurs disques identiques sont utilisés pour construire un volume RAID. « Identiques » signifie que ces disques appartiennent à la même marque, sont de la même série, et donc... ont des durées de vie similaires. Dans de telles conditions, la probabilité pour que plusieurs disques flanchent à peu près au même moment se trouve augmentée. Un autre scénario possible serait un accident électrique : foudre, sautes de tension, bloc d'alimentation du PC défectueux... Un tel accident pourrait endommager physiquement plusieurs disques à la fois, perdant du coup

toutes les données.

En résumé, le RAID 5 permet de sécuriser ses données dans une certaine mesure ; mais il ne dispense pas de faire des sauvegardes sur d'autres supports : CD/DVD, disques externes, voire si les données sont vraiment sensibles (en entreprise par exemple), redondance sur un autre serveur placé sur un site différent et/ou archivage sur bandes magnétiques. La règle d'or de la sauvegarde de données : ne faites jamais, **JAMAIS** confiance à un seul et unique moyen de stockage...