

Informatique et réseaux



Les systèmes RAID

Architecture des ordinateurs

LES DIFFERENTS SYSTEMES RAID

SOMMAIRE

INTRODUCTION

I LES DIFFERENTS RAID

- I.1 Le RAID 0
- I.2 Le RAID 1
- I.3 Le RAID 2
- I.4 Le RAID 3
- I.5 Le RAID 4
- I.6 Le RAID 5
- I.7 Le RAID 6
- I.8 Le RAID 7

II PRINCIPALES COMBINAISONS DE RAID

- II.1 La combinaison RAID 01
- II.2 La combinaison RAID 10
- II.3 La combinaison RAID 53

CONCLUSION

INTRODUCTION

RAID signifie «Redundant Array of Inexpensive / Independant Disks ». Ce système a été inventé en 1987 par trois chercheurs de l'université californienne de Berkeley, dans le but de pouvoir utiliser des disques de faibles capacités, donc peu coûteux (Inexpensive) de telles façons qu'ils soient vus comme un disque unique. De cette manière, le RAID peut être décrit comme étant l'opposé du partitionnement : dans un cas, on crée plusieurs unités logiques à partir d'un seul disque, dans l'autre, on crée une seule unité logique à partir de plusieurs disques physiques.

La définition officielle de RAID est la suivante :

une matrice de disques dans laquelle une partie de la capacité physique est utilisée pour stocker de l'information redondante concernant les données utilisateurs. Cette information redondante permet la régénération des données d'utilisateurs perdues lorsqu'une unité ou un chemin de données à l'intérieur d'une matrice est défaillant.

Un système RAID organise les données parmi plusieurs disques durs et utilise un processus de correction d'erreurs afin d'en assurer la fiabilité des archives. Avec l'évolution de la capacité des disques durs et donc de la quantité de données stockées, la fiabilité est de plus en plus un élément primordial. En 1998, c'est à dire à l'origine du système RAID, il existait 5 niveaux de RAID, numérotés de 1 à 5, correspondant aux différents niveaux de fiabilité et de performance d'un système. Depuis, d'autres types de RAID sont apparus. Ils sont soit l'évolution de RAID déjà existants, soit un combinaison de RAID de base.

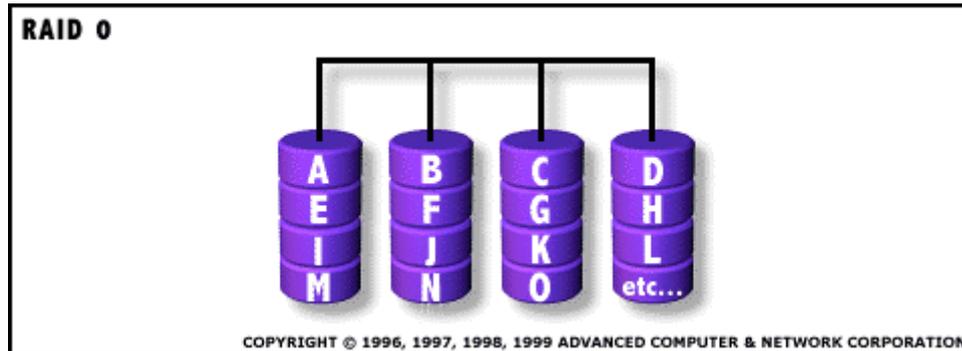
Le système RAID a été développé pour :

- Augmenter la capacité : le système RAID permet de mettre bout à bout des disques durs, ce qui permet d'accroître la taille du volume.
- Améliorer les performances : Les données sont écrites sur plusieurs disques à la fois. Ainsi, chacun des disques n'a qu'une partie des données à inscrire.
- Apporter la tolérance de panne : Certaines configurations RAID permettent de se prémunir contre les défaillances d'un disque. Cette fonctionnalité est très importante, car sinon, la panne d'un seul des disques d'un ensemble RAID entraîne la perte des données de tous les disques. C'est d'ailleurs ce qui arrive au niveau de RAID 0.

I LES DIFFERENTS RAID.

I.1 Le RAID 0 : striping.

Le RAID 0 n'est pas réellement un RAID dans la mesure où il ne répond pas exactement à la définition d'un RAID. Dans ce système, il n'y a aucune répétition d'informations, donc aucune redondance, ce qui implique qu'il n'y a aucune sécurité face à la panne. Ce mode consiste juste à unifier plusieurs disques durs les uns aux autres sans aucun contrôle : on répartit les données sur plusieurs disques.



Ce système permet un gain de temps lors de l'écriture et de la lecture des données car les disques sont montés en parallèles bien que, pour le système d'exploitation, la mise bout à bout des disques est transparente. De plus, on utilise toute la capacité de stockage des différents disques. Il n'y a pas d'espace perdu.

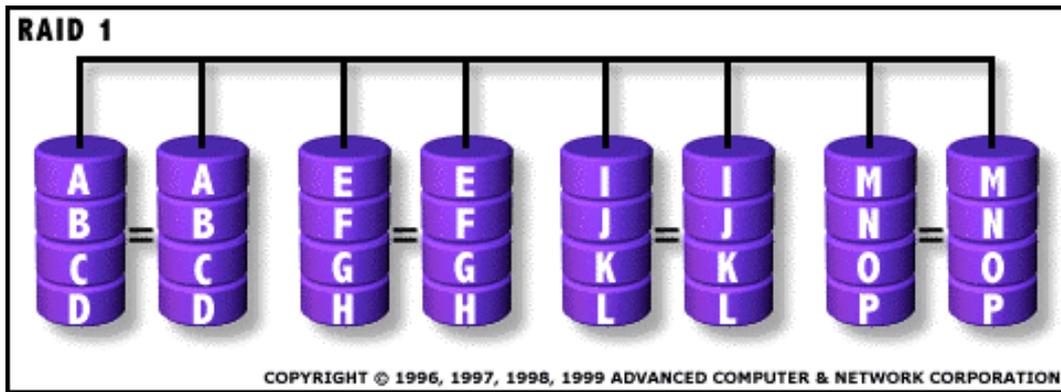
Cependant, si un disque tombe en panne, alors l'information contenue sur ce disque est perdue. C'est là le principal défaut de ce pseudo RAID.

Ce type de RAID est particulièrement intéressant pour les applications nécessitant une vitesse d'enregistrement très élevée.

I.1 Le RAID 1 : mirroring.

Cette technique est utilisée depuis les années 60. Elle est souvent appelée «disk mirroring», «dual copy» ou encore «disk shadowing». Le principe est relativement simple. Tout fonctionne par paires de disques durs. Lorsque l'on écrit une donnée, on l'écrit sur deux disques durs. Ainsi on obtient deux disques identiques au niveau des données. Lorsque le plus petit disque de la première paire est plein, on passe à la paire de disques suivante.

Si jamais un des deux disques durs tombe en panne, l'autre contient toutes les informations, d'où le très bon niveau de sécurité. Tant que le deuxième disque est en panne, le RAID fonctionne en mode dégradé et dès que l'on remplace le disque défectueux, les données sont systématiquement recopiées.



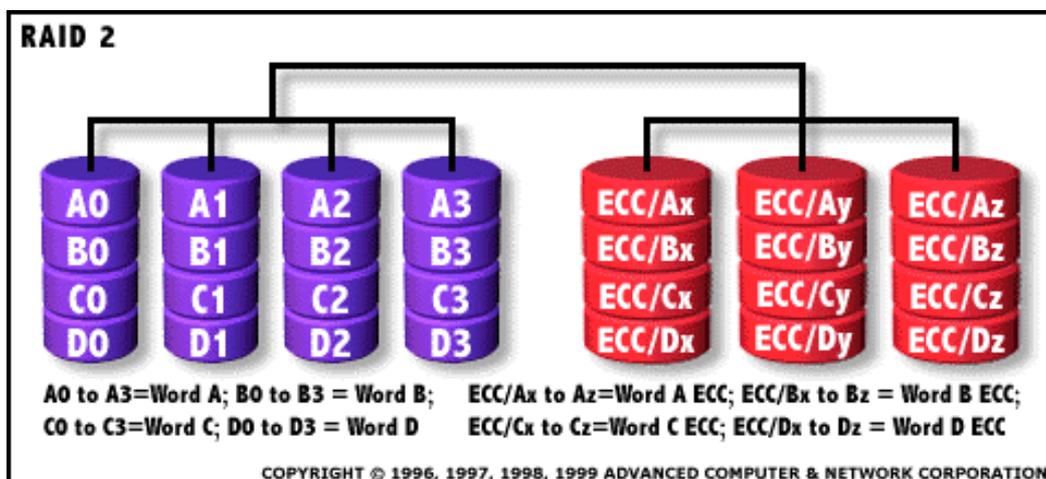
Les performances en écritures sont identiques à celle d'un seul disque puisque les deux disques doivent sauvegarder les mêmes données au même moment. Cependant lors de la phase de lecture, les performances sont améliorées. En effet, la lecture s'effectuant sur deux disques en parallèle, on double alors la vitesse de lecture.

Le problème principal du RAID 1 vient du fait que les deux disques miroirs sont identiques et que l'on perd la moitié de l'espace mémoire. En effet, avec deux disques de 10 Go, on ne peut stocker que 10 Go. Ce système devient donc assez coûteux.

I.3 Le RAID 2.

Les systèmes de types RAID 2 ne sont que très peu utilisés. Ils utilisent des disques pour les données ainsi que des disques pour des codes de correction d'erreurs (ECC). Ils effectuent une segmentation au niveau des bits pour les données, puis à l'aide d'un algorithme de calcul, fournit des codes de correction d'erreurs, dits codes de Hamming.

Cette technique permet de vérifier l'intégrité des données à l'aide de ces codes de Hamming. Chaque écriture de données nécessite le calcul et l'écriture du code de vérification d'erreurs associé. De même chaque lecture de données nécessite la lecture du code d'erreurs associé afin de vérifier que les données sont bonnes, voire de faire des corrections d'erreurs à la volée.



Les systèmes RAID

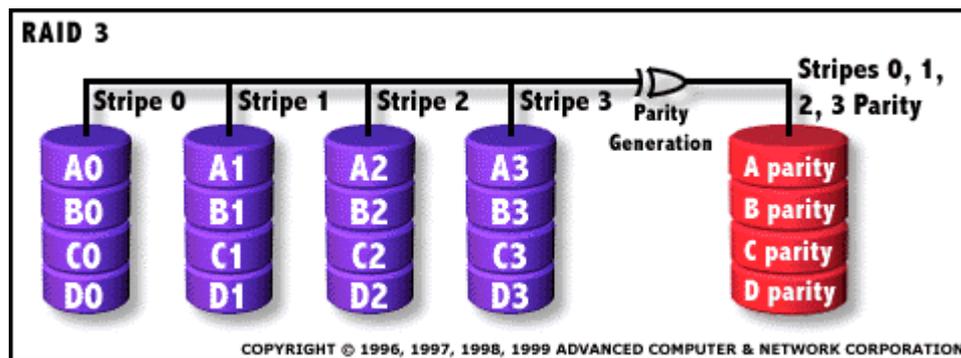
Ce système reste très peu utilisé. En effet son intérêt a grandement diminué puisque tous les disques durs actuels sont pourvus d'un dispositif de détection d'erreur basé sur le même principe. De plus les systèmes de RAID 2 nécessitent un contrôleur complexe très cher.

Cette solution peut donc être intéressante seulement si on a à manipuler des dispositifs dépourvus de tout dispositif de correction d'erreurs.

I.4 Le RAID 3.

La technique du RAID 3 utilise la principe du RAID 0 auquel on associe un disque supplémentaire afin de stocker des bits de parité. Ces bits de parité correspondent alors à des codes d'erreurs.

Si un disque tombe en panne, il est possible, à partir du disque de parité, de reconstituer l'information présente sur le disque défectueux. Le principe des bits de parité est aussi simple que celui d'une addition. Si on sait que $A + B = C$, on peut aisément retrouver B à partir de A et de C. C'est le même principe pour les bits de parité. Si le disque B tombe en panne, à partir des bits de parité et des bits du disque A, on reconstitue l'information sur B. Ainsi tout système utilisant le calcul de parité nécessite au minimum trois disques.



Afin que ces calculs de parité puissent être fiables, il faut que tous les disques soient synchronisés. Ceci affecte les performances en écriture du système. En effet, chaque écriture sur un des disques nécessite une écriture sur le disque de parité. Ainsi il se forme un goulet d'étranglement au niveau du disque de parité. De ce fait, le nombre de disques à mettre en parallèle se voit limité.

Cependant les performances en lecture ont augmentées. La lecture se fait simultanément sur tous les disques comme pour un RAID 0.

Si le disque de parité tombe en panne, alors on retombe sur le cas d'un simple système de type RAID 0 perdant ainsi toute sécurité. Le disque de parité est donc extrêmement important et nécessite un niveau de fiabilité particulièrement élevé.

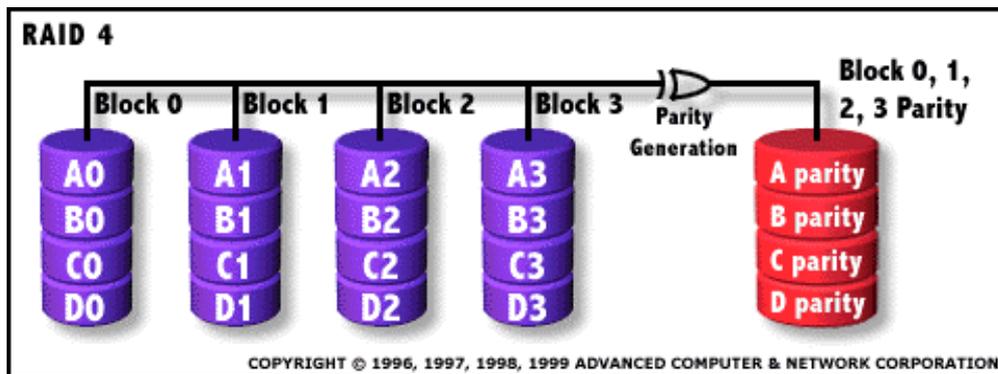
I.5 Le RAID 4.

La structure physique d'un système RAID 4 est identique à celle d'un RAID 3. En effet, en plus des différents disques de données, on utilise un disque de parité. La différence s'effectue au niveau de la segmentation des données. Dans un RAID 3, la segmentation des

données s'effectue au niveau des octets, tandis que dans un RAID 4, on effectue une segmentation par bloc. Il existe une autre différence avec le RAID 3 : les accès disque ne sont plus synchronisés.

La lecture et l'écriture s'effectuent simultanément sur tous les disques. Les données sont décomposées en segments de petites tailles variables, d'un ou plusieurs secteurs. Cela permet d'assurer le traitement de demandes de lecture de taille inférieure au segment en parallèle sur plusieurs disques.

En cas de défaillance d'un disque, le système continue toujours de fonctionner en mode dit dégradé. Ces données perdues sont reconstituées toujours à l'aide du disque de parité.



Lors d'écritures de taille inférieure à la taille des segments, le contrôleur doit mettre à jour les données de parité. Cette opération nécessite entre autre une lecture de l'ancienne parité suivie de l'écriture de la nouvelle. Ces deux opérations impliquent un unique disque de parité mais le transforment en goulot d'étranglement et dégradent les performances de la grappe RAID 4. Ce problème de disque de parité est une bonne sécurité mais il limite les performance en écriture et en lecture.

Le RAID 4 n'est que peu utilisé. En fait, il est à mi-chemin entre le RAID 3 et le RAID 5 qui restent les plus utilisés. Il utilise le un disque de parité comme le RAID 3 et une segmentation par bloc comme le RAID 5.

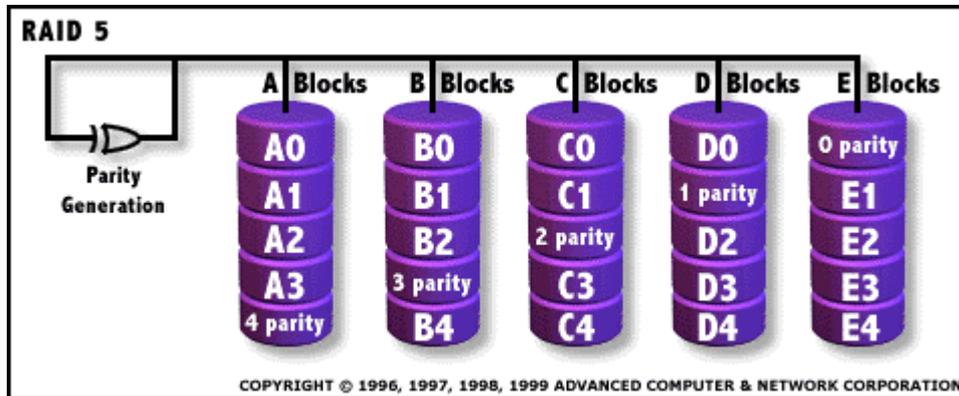
1.6 Le RAID 5.

Le RAID 5 est le plus utilisé des niveaux de RAID. La taille des segments de données est fixée à celle d'un cluster. Les disques de la matrice RAID 5 travaillent de manière asynchrone et, plutôt que d'utiliser un disque spécifique pour la parité, chaque disque de données contient une partie réservée à celle-ci. Le disque de parité est reparti sur les différents disques de données. Ceci permet de supprimer le problème majeur du disque de parité, c'est à dire celui du goulet d'étranglement à son niveau. Ce problème est lui aussi reparti sur les différents disques.

Si on a N disques de données, les performances en lecture et écriture seront accrues de N-1. La capacité de stockage, comme pour les RAID 3 et 4, est de N-1 fois la capacité d'un disque. En effet l'équivalent d'un disque est perdu pour les données de parité.

Les systèmes RAID

En cas de panne d'un disque, il est possible de le changer, les données étant alors reconstruite à partir de celles des autres disques. Cette opération peut bien évidemment prendre un certain temps, pendant lequel les performances sont dégradées. A noter que pour le RAID 5, comme pour le RAID 3 et le RAID 4, le changement de disque dur à chaud (*hot spare* et *hot plug*) est possible, ce qui évite d'avoir à rebooter un serveur (la reconstruction des données démarre aussitôt le nouveau disque en place).

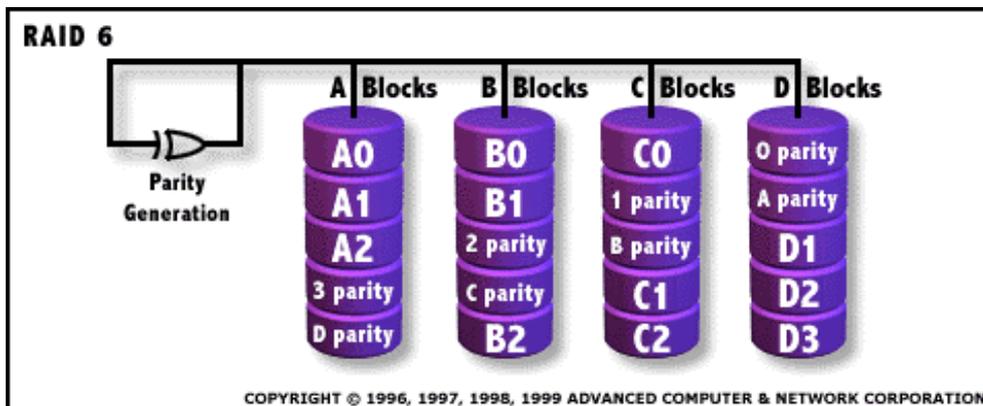


Ce système RAID est donc particulièrement utile pour sécuriser et accélérer les performances d'un serveur de base de données.

I.7 Le RAID 6.

Le RAID 6 est une évolution du RAID 5. Les systèmes RAID 3 à 5 utilisent un système de calcul de bits de parité. Il permet de reconstituer le contenu d'un disque si ce dernier est défaillant. Cependant si un deuxième disque tombe en panne, alors on perd les données contenues sur les deux disques défectueux. C'est pourquoi le système RAID 6 utilise un deuxième code de redondance.

Ces deux codes de redondances sont écrits sur les disques de données comme pour le RAID 5. L'espace disque réservé à ces codes est donc équivalent à deux disques, ce qui augmente le coût de la solution.



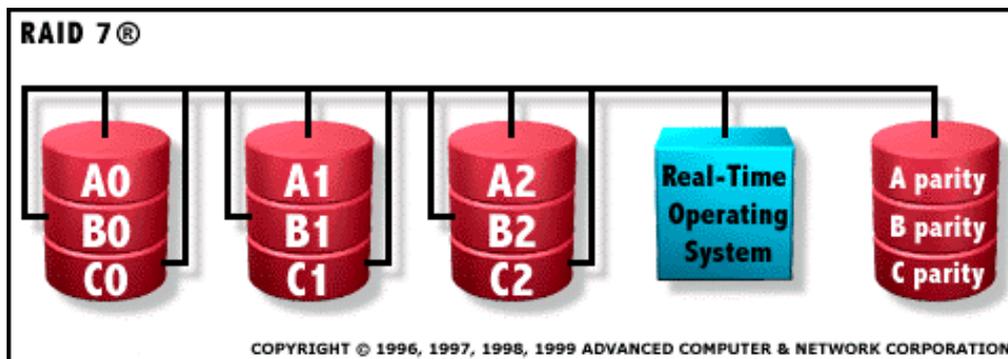
Les performances en écriture sont diminuées par rapport au RAID 5. En effet étant donné que l'on a deux codes de redondance, lors d'une écriture, il faut écrire la donnée ainsi que les deux codes de redondance.

Le RAID 6 est en réalité peu utilisé. Il est plus coûteux. Mais ce n'est pas la seule raison : le niveau de sécurité supplémentaire apporté par le deuxième code de redondance paraît inutile. Il est très rare que deux disques d'une même matrice RAID tombent en panne simultanément. Si un tel cas se produit, il y a une forte possibilité que le problème soit plus grave et que toute la matrice RAID soit défaillante.

Le RAID 6 convient parfaitement aux applications critiques par son haut niveau de sécurité.

1.8 Le RAID 7.

Le RAID 7 est le système RAID le plus performant. Il est annoncé comme étant de 1,5 à 6 fois plus performant que n'importe lequel des autres niveaux RAID. Ce système se rapproche du RAID 3 par l'utilisation d'un disque de parité. Mais le problème du goulet d'étranglement, dû au fait que les disques devaient être synchrones, a été résolu. Une carte microprocesseur, fonctionnant sous un noyau temps réel, contrôle et calcule la parité, gère le cache sur plusieurs niveaux et surveille les disques.



C'est cette gestion du cache qui permet au RAID 7 de travailler de manière asynchrone aussi bien en lecture qu'en écriture, et donc d'améliorer les performances. En réalité, un RAID 7 n'est qu'un RAID 3 asynchrone.

Le principal problème de cette matrice RAID est qu'elle est très coûteuse à mettre en œuvre et encore très peu répandue.

II PRINCIPALES COMBINAISONS DE RAID.

II.1 Présentation des combinaisons de RAID.

Chaque système RAID présente des avantages mais aussi des défauts. Il est possible de palier à ces défauts en associant deux systèmes RAID différents à plusieurs disques ou au contraire de profiter de leurs avantages.

Le système le plus utilisé dans les combinaisons est le RAID 0. S'il est associé avec un système RAID de niveau 1, 3 ou 5, le système obtenu possède alors un niveau de performance très élevé ainsi qu'une tolérance à la panne.

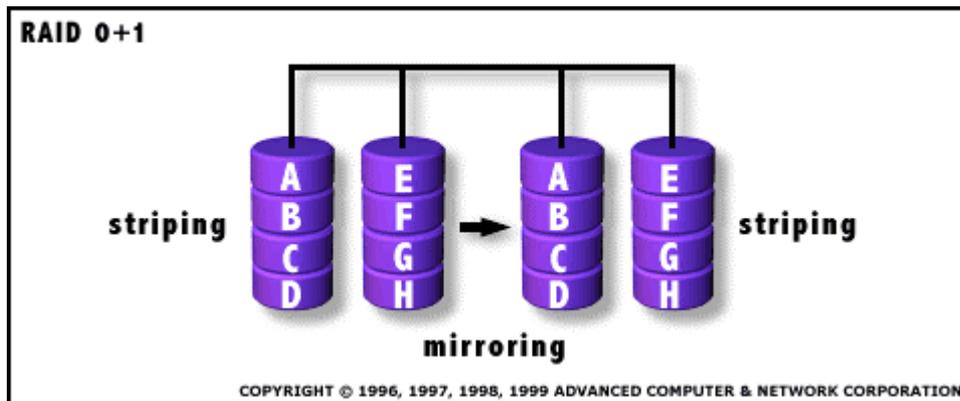
Le nombre de combinaisons de RAID est relativement élevé car la combinaison de deux RAID permet d'obtenir deux systèmes différents suivant le sens d'implémentation des

RAID . Par exemple, la combinaison de RAID 0 et 1 donne des résultat totalement différents suivant que l'on implique d'abord le RAID 0 ou le RAID 1.

II.2 La combinaison RAID 0+1.

Un système RAID 0+1 associe un RAID 0 et un RAID 1. Le RAID 0 est implémenté en premier et le RAID 1 en second.

Si on a une matrice de 10 disques, on divise les disques en deux groupes de 5, chaque groupe étant configuré en RAID 0. Les deux groupes sont alors reliés entre eux par un système RAID 1. On applique donc la technique du mirroring sur un système de RAID 0.



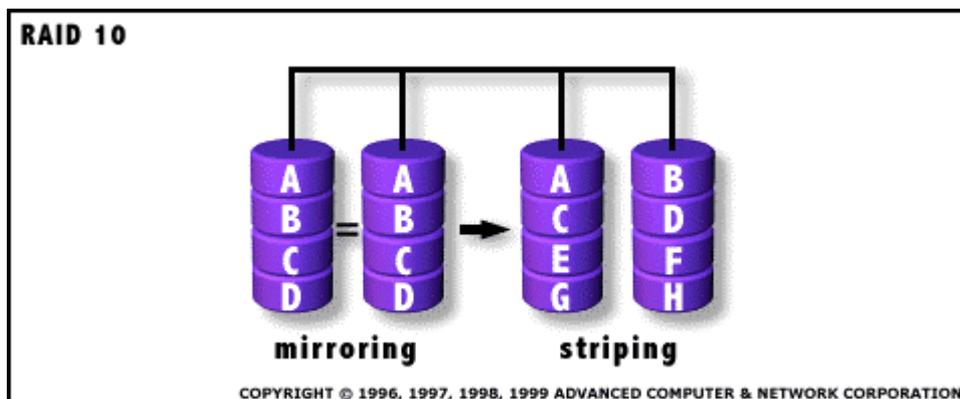
Si un disque tombe en panne, alors on perd le RAID 1, et on se retrouve avec un simple système RAID 0. De plus, si un disque de cette matrice tombe en panne, alors l'ensemble des données est perdu.

Dans le cas d'une panne, temps de reconstitution des données est élevé. En effet, il faut rétablir les données des cinq disques. C'est pourquoi le RAID 0+1 n'est que peu utilisé.

II.3 La combinaison RAID 10.

Un système RAID 10 associe lui aussi un RAID 0 et un RAID 1. Cette fois-ci, on implémente d'abord le RAID 1, puis le RAID 0.

Pour une matrice de 10 disques, on regroupe les disques deux par deux afin de créer cinq systèmes RAID 1. Puis on configure les 5 paires de disques en RAID 0.



Les systèmes RAID

En cas de panne d'un disque, il n'y a pas de problème. En effet, il y a juste une paire de disque qui fonctionne en mode dégradé. Si un deuxième disque tombe en panne, mis à part le cas où il s'agit du second disque de la même paire, il n'y a toujours pas de problème. La probabilité d'avoir les deux disques de la même paire défectueux est très faible.

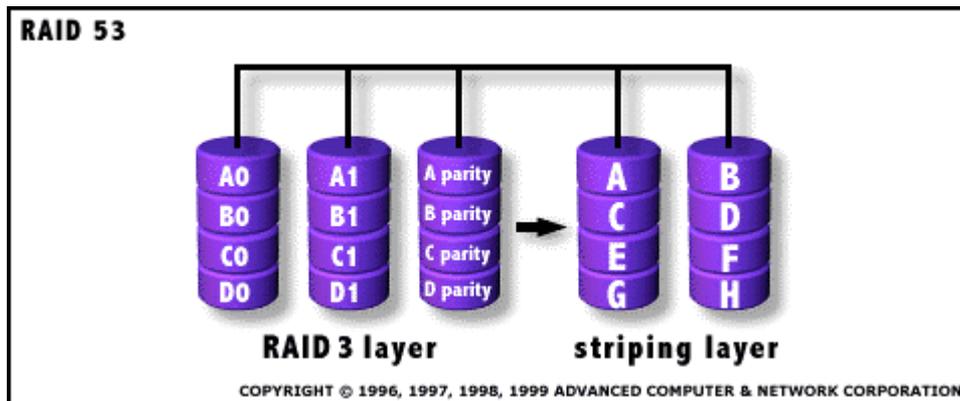
Dans ce cas toutes les données seraient perdues comme pour un simple RAID 0. Dans le cas d'une panne, il n'y a alors qu'un seul disque à reconstituer. Ce système offre les bonnes performances du RAID 0 tout en offrant la tolérance aux pannes du RAID 1.

Ce mode de RAID a toujours le même inconvénient que le RAID 1, c'est à dire qu'il est coûteux en capacité de stockage à cause du *mirroring*. Il nécessite au minimum 4 disques durs : le RAID 0 nécessite au moins 2 disques qui dans ce cas sont les 2 paires de disques du RAID 1.

Etant donné la baisse actuelle des prix des disques, cette solution est de plus en plus envisagée.

II.4 La combinaison RAID 53.

Le RAID 53 n'a strictement rien à voir avec une combinaison RAID 5 et 3. Il s'agit en réalité d'un RAID 03.



Le RAID 53 est très complexe dans son fonctionnement. Il permet d'obtenir des performances supérieures au RAID 3. Cependant, le coût de mise en œuvre d'un tel système est bien évidemment plus élevé.

CONCLUSION

Chaque système RAID correspond à des performances différentes. Il est nécessaire de bien définir l'objectif du futur système afin d'évaluer le niveau de RAID.

Il existe deux niveaux pour mettre en œuvre un système RAID.

Le premier se situe au niveau software. Certains systèmes d'exploitation permettent de gérer les RAID. C'est le cas de Windows NT et Windows 2000, qui permettent de gérer des RAID 0,1,5.

Le second niveau est le hardware. On peut soit brancher une carte contrôleur sur un port PCI et les disques ne sont plus branchés sur les contrôleurs SCSI ou IDE mais sur la carte contrôleur elle-même, soit utiliser un contrôleur RAID externe. Le contrôleur RAID et les disques de la matrice RAID se situent dans une tour externe. L'interface avec le reste du système se fait par une carte SCSI situé dans le PC. Un système RAID en hardware obtiendra des performances élevées au dépend du prix.

Les systèmes RAID possèdent des atouts très important, c'est à dire performance et fiabilité. De plus, les coûts sont de plus en plus faible du fait de la baisse constante des prix des disques durs.

Un système RAID ne remplace en aucun cas à un système de sauvegarde. Il ne protège en rien d'une erreur humaine.