La Technologie

RAID

Version 2006

SOMMAIRE

Introduction générale

La technologie RAID (Redundant Array of Independent Disks) a été crée en 1987 par des chercheurs de l'université de Berkeley (Californie) ; Le but de cette technologie est de remplacer les coûteux disques de 6.5 pouces et de 9.5 pouces (ces disques sont aussi appelés « disques grands systèmes ») qui équipent les gros serveurs (comme les serveurs de fichiers) par des ensembles formé de disques de 3.5 pouces plus répandus et beaucoup moins chers. Cela explique pourquoi l'acronyme RAID signifie aussi Redundant Array of Inexpensive Disks.

Cette technologie est principalement utilisée dans le domaine de l'entreprise pour sécuriser ou bien pour augmenter les performances des serveurs. Elle est parfois mise en place avec d'autres technologies telles que le clustering afin d'obtenir des performances encore plus grande ou bien d'avoir un sécurité maximale.
Cependant, le RAID se popularise rapidement par le biais des cartes d'extension (par exemple les cartes au format PCI 32 bits) et des chipsets intégrés au sein des cartes mères grand public (ex. : le chipset sud Intel ICH5R). Généralement les contrôleurs RAID que l'on trouve sur le marché du matériel « grand public » sont très limités par rapport aux contrôleurs RAID haut de gamme (niveaux RAID supportés réduit au RAID 0 et 1, pas de mémoire cache intégrée au contrôleur …).

Le but du RAID est de rassembler plusieurs disques durs physiques en seule unité logique. C'est-à-dire que pour plusieurs disques utilisant la technologie RAID, le système d'exploitation ne verra qu'un seul disque que l'on pourra qualifier de « disque virtuel ». Le RAID fonctionne donc grâce à un ensemble (ou agrégat) de disques. Il existe plusieurs niveaux (ou configuration) de RAID qui rassemblent différemment les données sur l'ensemble des disques. En fonction du niveau de RAID sélectionné on peut :

- additionner la capacité des disques durs pour former un disque virtuel plus grand
- obtenir un disque virtuel plus performant
- mettre en place la tolérance de panne (c'est-à-dire sécuriser les données)
- augmenter la capacité, les performances et la sécurité.

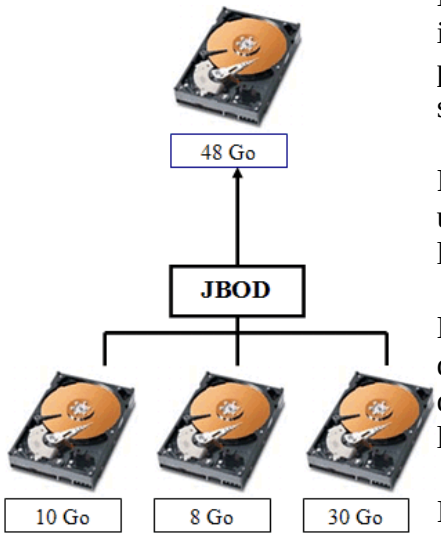
La technologie RAID peut être mise en place au niveau matériel ou niveau logiciel. Elle fonctionne avec des disques dur utilisant l'interface PATA (Parallel Advanced Technology Attachment), SATA (Serial Advanced Technology Attachment) ou bien SCSI (Small Computer System Interface).

I Les Niveaux Simples

I.A Introduction

En 1987, les chercheurs de l'université de Berkeley ont proposés cinq niveaux de RAID (les niveaux 1, 2, 3, 4 et 5). Depuis d'autres niveaux ont été implémentés comme le RAID 6 et le RAID 7. Les niveaux JBOD et 0 sont aussi considérés comme des modes de RAID bien qu'ils n'implémentent pas de redondance des données (le 'R' de RAID signifie « Redundant » c'est-à-dire redondance). Le RAID 0 et le JBOD sont d'ailleurs présents en standard sur toutes les cartes RAID du marché. Au niveau logiciel, les niveaux 0, 1, 5 et JBOD sont inclus dans le système d'exploitation de Microsoft depuis Windows NT4. Ces modes (niveaux) de RAID sont considérés comme des modes simples. En, effet nous verrons plus tard qu'il est possible de créer des niveaux de RAID combinés.

I.B Le JBOD



Le mode JBOD (Just a Bunch Of Disks) ou RAID linear, est implémenté dans quasiment toutes les cartes RAID du marché. Il permet tout simplement de rassembler plusieurs disques durs en une seule unité logique de plus grosse capacité.

Par exemple si l'on possède un disque de 8Go, un disque de 10Go et un disque de 30Go, on peut les rassembler en un seul disque dur logique de 48Go.

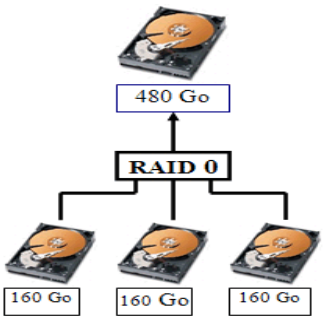
Les données sont ensuite écrites disque par disque, c'est-à-dire que dès qu'un disque est plein, la suite des données est copiée sur un autre disque. Tout cela reste bien sur complètement transparent pour l'utilisateur.

Il n'implémente aucune tolérance de panne et n'augmente pas les performances. Il sert uniquement à étendre la capacité d'un disque.

Ce mode est intéressant si l'on souhaite stocker de gros fichiers ou bien si l'on dispose d'un grand nombre de petits disques durs.

I.C Le RAID 0 ou « stripping »

Le RAID 0 est un agrégat (ensemble) de disques formé d'au moins deux disques durs. Avec le RAID 0, les données sont réparties sur l'ensemble des disques de l'agrégat. Cela améliore les performances en lecture et en écriture. Par exemple si on copie un fichier de 99 Mo sur un agrégat composé de 3 disques, alors le contrôleur RAID va diviser de façon équitable le travail entre les trois disques qui écriront chacun 33 Mo. Cela va donc multiplier le débit par 3. De plus la capacité des disques s'additionne pour former un disque virtuel de plus grosse capacité. Attention cette partie de l'article définie l'ensemble des niveaux RAID dit "simples". Toutes les caractéristiques (débit, capacité, ...) sont données à titre théorique et ne reflètent pas forcément les perfomances réelles.



Supposons que l'on crée un agrégat (en RAID 0) avec 3 disques durs ayant les mêmes caractéristiques :

- capacité de 160Go
- débit de 100Mo/s en lecture
- débit de 80Mo/s en écriture

Cet agrégat se comportera alors pour le système d'exploitation comme un disque dur virtuel de 480Go (3*160 Go = 480Go) avec :

- un débit en lecture de 300 Mo/s (3*100Mo/s)
- un débit en écriture de 240 Mo/s. (3*80Mo/s)

Si on crée un agrégat avec des disques possédant des caractéristiques différentes alors les caractéristiques du disque le moins performant sont utilisées.

Par exemple, si l'on met en place un niveau RAID 0 avec un disque d'une capacité de 60Go et un disque d'une capacité de 2Go, alors seulement deux giga-octets seront utilisés sur chaque disque et le système d'exploitation verra un disque de 4Go alors que la capacité réelle des deux disques réunis s'élève en réalité à 2Go + 60Go = 62Go. Il y aura donc un espace disque de 58Go (60Go – 2Go = 58Go) qui sera inutilisé sur le premier disque. Lorsque l'on implémente un niveau RAID 0, il est donc préférable de choisir des disques de même capacité.

De plus si l'on crée un système RAID 0 avec un disque rapide avec un disque lent, on obtiendra un disque virtuel possédant environ le double des performances du disque lent. Le disque lent va donc brider le disque rapide. Il faut donc veiller à utiliser des disques proposant des caractéristiques similaires afin d'optimiser les performances. Par exemple associer un disque fonctionnant à 10 000 tr/min et possédant 8Mo de mémoire cache et un disque fonctionnant à 5400 tr/min et possédant 2Mo de mémoire cache) serait un véritable gâchis.

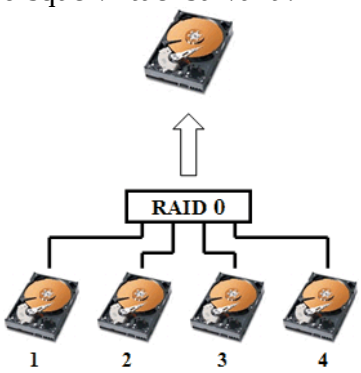
Pour résumer, un niveau RAID 0 doit utiliser des disques durs ayant des caractéristiques très proches (notamment en capacité et en performances).

Pour bien illustrer ces propos voici un petit exemple : on met en place un niveau RAID 0 au moyen de 4 disques avec les caractéristiques suivantes :

	Disque 1	**Disque 2**	**Disque 3**	**Disque 4**
Capacité	300 Go	18Go	20Go	45Go
Débit en lecture	90Mo/s	160Mo/s	20Mo/s	35Mo/s
Débit en écriture	82Mo/s	125Mo/s	20Mo/s	15Mo/s

Les valeurs les plus faibles pour chacune des caractéristiques ont été misent en rouge sur le tableau ci-dessus.

Lorsque l'on branche ces disques sur une carte RAID et que l'on défini un niveau RAID 0, on obtient le disque virtuel suivant :



- capacité : 4*18Go = **72Go**
- débit en lecture : 4*20Mo/s = **80Mo/s**
- débit en écriture : 4*15Mo/s = **60Mo/s**

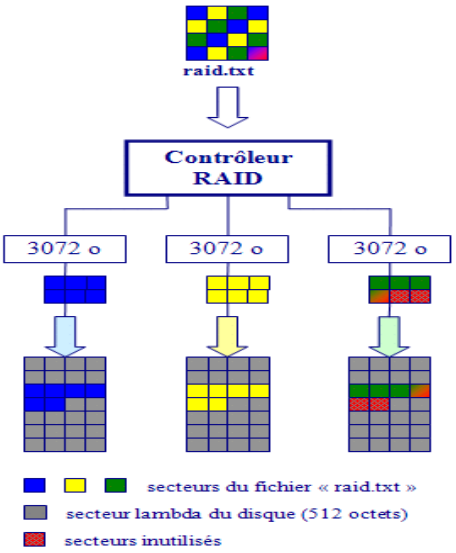
Le disque virtuel composé des quatre disques physiques s'avère donc plus lent et plus petit que le disque dur N°1 utilisé seul.
Cet exemple est donc une utilisation catastrophique du niveau RAID 0.

Lorsque l'on utilise un niveau RAID 0, nous avons vu que les données sont réparties entre les différents disques, ce qui permet d'augmenter les performances. Pour répartir les données sur l'ensemble des disques de la façon la plus optimisée, la carte RAID (ou bien le système d'exploitation dans le cas d'un RAID logiciel) doit les découper en petites unités appelées segments. Lorsque l'on met en place un niveau RAID 0, on doit spécifier la taille de ces segments (ou block size, et parfois nommé chunk size). Cette taille est un multiple de la capacité d'un secteur de disque dur (soit 512octets). Ainsi, la plus petite taille possible pour un segment est de 512octets et peut attendre plusieurs Méga-octets (parfois plus de 4Mo).

La taille sélectionnée va influencer sur les performances en lecture et en écriture. Pour le stockage de gros fichiers, il est recommandé de choisir une taille élevée (256Ko, ou plus) alors que pour les petits fichiers une petite taille est à privilégier (ex. : 512octets, 1024 octets, …). Mais il ne faut surtout pas mettre une taille de 2048 Ko pour stocker des petits fichiers de 4Ko chacun car cela occuperait énormément d'espace disque inutilement et les performances seraient désastreuses.

Voici une petite représentation logique du travail de segmentation effectué par le contrôleur RAID.

Dans l'exemple ci-contre, on a crée un agrégat en mode RAID 0 à l'aide de trois disques durs. On a sélectionné une taille de segment de 3072 octets. Un segment occupe donc 6 secteurs de disque dur (car 3072 octets = 6 * 512 octets).



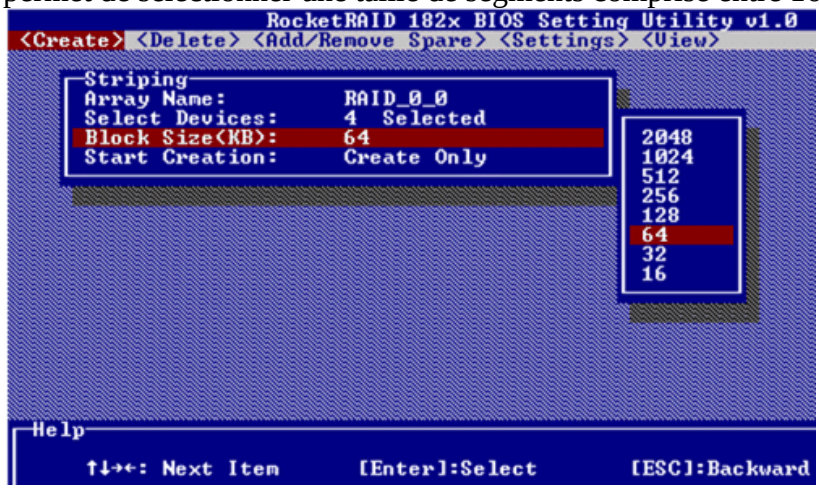
On vient de créer un nouveau document texte nommé « raid.txt » et on souhaite le sauvegarder sur le disque. Ce fichier a une taille de 8Ko (soit 8000 octets).

Le fichier est envoyé au contrôleur RAID qui le découpe en 3 segments de 3072 octets. Une fois le découpage des données terminé, le contrôleur réparti les segments entre les trois disques qui les écrivent en même temps.

Le troisième segment n'est pas utilisé au maximum car il contient uniquement 1856 octets (8000 octets – 2 * 3072 octets = 1856 octets) sur les 3072.
On perd donc un espace de stockage de 3072 – 1856 = 1216 octets.
C'est pourquoi il ne faut pas choisir une taille de blocks trop élevée.

Pour avoir un ordre d'idée des tailles de segments que l'on rencontre généralement, voici une capture d'écran du BIOS de la carte RocketRAID 1820 (carte RAID de marque HightPoint qui gère les niveaux RAID 0, 1, 5, 10 et JBOD et qui possède 8 canaux SATA, une carte récente donc).

On remarque qu'elle permet de sélectionner une taille de segments comprise entre 16Ko et 2048Ko.



Une fois que le niveau de RAID est crée, on doit ensuite partitionner et formater avec un système de fichier (ex. : NTFS), le disque virtuel de manière logicielle (à l'aide de l'utilitaire *fdisk.exe* par exemple). Selon le système de fichiers sélectionné, on peut paramétrer diverses options dont la tailles des unités d'allocation (ou clusters). Pour maximiser les performances et éviter de perdre trop d'espace disque, il est fortement recommandé de mettre une taille de segments (ou blocks) égale à celle des unités d'allocation (ou clusters). C'est d'ailleurs ce que propose Windows lorsque l'on crée un niveau RAID 0 de manière logicielle.

Pour conclure, l'avantage du RAID 0 est une augmentation des performances proportionnelle au nombre de disque et ce sans perte de capacité. Cependant ce mode ne tolère pas les pannes (si un des disques tombe en panne alors toutes les données stockées sur l'agrégat sont perdues) et induit une augmentation de la probabilité d'une panne (plus il y a de disques plus il y a de chances pour que l'un d'eux tombe en panne).

I.D Le RAID 1

Un système RAID 1 duplique les données sur tous les disques. Tout comme le RAID 0, le RAID 1 nécessite un minimum de deux disques durs pour fonctionner. Les disques ont un contenu identique ce qui implique une perte d'espace proportionnelle au nombre de disques rajoutés. En effet, si on met 3 disques de 40 Go en RAID 1 alors les 40 Go écris sur le disque 1 seront copiés sur les deux autres disques. On perdra donc 2 * 40 = 80 Go d'espace disque sur la capacité totale des 3 disques (120 Go).

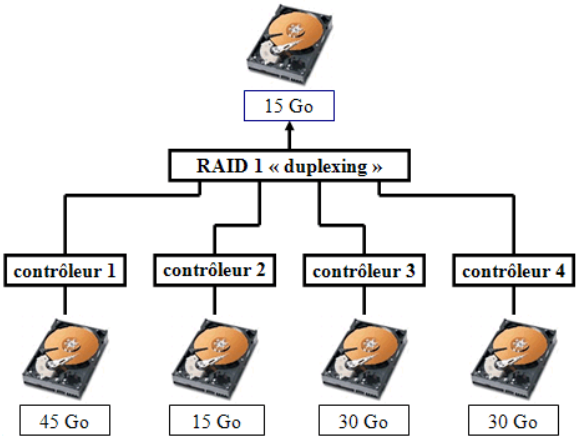
En cas de dysfonctionnement de l'un des disque dur (cela arrive fréquemment sur des serveurs recevant beaucoup de requêtes par secondes), le contrôleur RAID utilisera alors uniquement les disques en état de marche. Le RAID 1 apporte donc la tolérance de pannes. L'utilisation du RAID 1 est particulièrement recommandée sur des serveurs stockant des données sensibles qui doivent rester fonctionnels 24 heures sur 24.

Lorsque l'on met en place un RAID 1 avec des disques dur de capacités différentes, alors la capacité de stockage de l'ensemble en RAID 1 sera celle du disque de plus faible capacité.

Lorsqu'un accès en lecture est réalisé sur un ensemble de disques en mode RAID 1, alors tous les disques lisent chacun une partie de la données ce qui améliore grandement les performances en lecture. En revanche lors de l'écriture sur un ensemble de disques RAID 1, les performances ne sont pas améliorées.

En effet, les disques vont écrire les données de façon synchronisée et ce au fur et à mesure que les données vont arriver au contrôleur. Le travail lors d'un accès en écriture n'est donc pas divisé entre les disques mais est commun à tous les disques. Le débit en écriture d'un ensemble RAID 1 est donc le même que celui d'un disque seul. C'est pourquoi dans un ensemble RAID 1, contenant 5 disques les accès en lecture seront grandement améliorés alors que les accès en écriture stagneront.

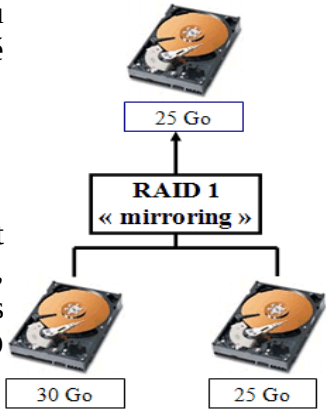
Le but du RAID 1 est d'offrir une sécurité maximale aux données stockées sur une machine. C'est pourquoi certaines cartes RAID incluent plusieurs contrôleurs RAID (un par disque) afin que malgré la perte d'un contrôleur les données soient toujours accessibles. Dans ce cas on parle de " duplexing ".



Le schéma à gauche présente un RAID 1 en « duplexing ».

Ici seuls 15 Go sont exploitables car le plus petit des disques utilisés à une capacité de 15 Go. Étant donné que 4 disques sont utilisés, les données ne seront pas perdues même si 3 des 4 disques tombent en panne. De plus grâce au « duplexing », le système supporte la perte d'un contrôleur ou plus.

Dans la majorité des cas, la carte RAID ou le circuit intégré (sur certaine cartes mères d'entrée de gamme, le contrôleur IDE intégré gère le RAID) ne possède qu'un seul contrôleur. On parle alors de « mirroring ».
Le schéma (à droite) présente un RAID 1 en « mirroring ». Seul 25 Go sont exploitables et les performances en lecture sont multipliées par deux.



Il existe donc bien deux types de RAID 1 différents. Le « duplexing » étant méconnu et très peu utilisé, une certaine confusion s'est crée à propos du RAID 1. En effet, « mirroring » n'est pas synonyme de RAID 1 (contrairement à ce qu'on peut voir dans des forums sur la toile et même dans les manuels de certains constructeurs de carte RAID !).

I.E Le RAID 2

Le RAID 2 est une alternative au RAID 1. En effet, le RAID 2 est identique au RAID 1 sauf sur un point : lors des opérations de lecture, un seul des disques travaille ce qui permet aux autres disques de se « reposer ».

Les autres disques étant moins sollicités, cela augmente leur durée de vie. En outre le RAID 2 implémente un système de corrections des erreurs. Cependant, cela est devenu inutile car tous les disques durs SCSI intègrent dorénavant leur propre système de correction d'erreurs.

Le RAID 2 augmente donc la fiabilité du RAID 1 mais induit une baisse des performances. Il a peu été utilisé et n'est plus utilisé de nos jours à cause de son obsolescence.

I.F Le RAID 3

Le RAID 3 apporte les avantages du RAID 0 (amélioration des performances) et du RAID 1 (tolérance de panne). Le mode RAID 3 fonctionne avec un minimum de trois disques. A l'instar du RAID 0, tous les

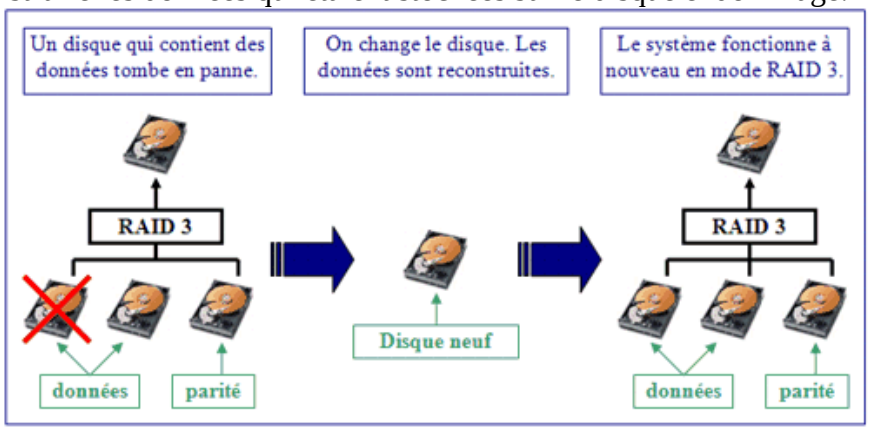
disques sauf un qui a une tâche spéciale se divisent les opérations de lecture et d'écriture ce qui augmente les performances. En RAID 3, la taille des segments n'est pas modifiable et est fixée à 512 octets (en RAID 3 : un segment = un secteur de disque dur = 512 octets). L'un des disques est un disque de parité.

I.F.1 La tolérance de panne grâce à la parité

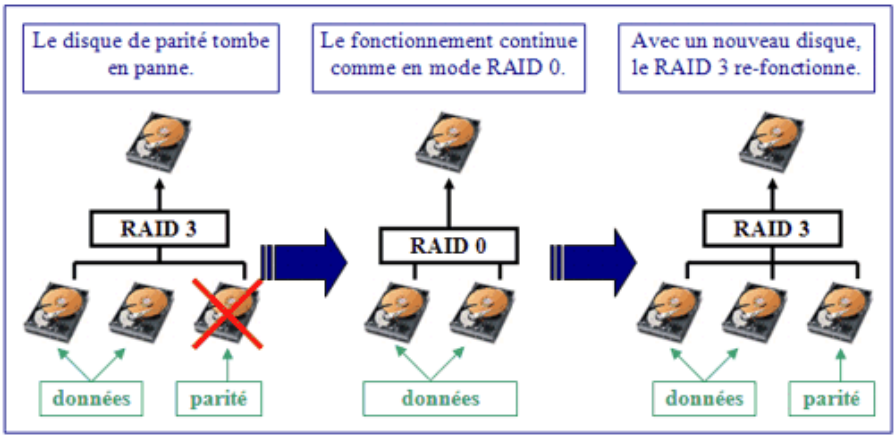
La parité est un processus algorithmique utilisant l'intégrale de parité inventée par Gérard Langlet. Ses applications sont très nombreuses en informatique et en particulier dans le domaine du RAID. Avec un ensemble RAID3, lors de chaque opération d'écriture les données vont êtres réparties entres tous les disques (sauf le disque de parité) de façon équitable.

Au moment où les données sont écrites, des informations indiquant sur quel disque a été stocké tel ou tel fragment des données sont enregistrées dans le disque de parité. Ce sont les informations de parité. Ainsi, si un fragment d'une donnée est défectueux ou manquant, la comparaison entre les informations de parité et les autres fragments de la donnée (qui sont stockés sur les autres disques durs) vont permettre de reconstituer le fragment.

Si l'un des disques contenant des données tombe en panne, alors le disque de parité et les disques restant permettent de reconstruire les données qui étaient stockées sur le disque endommagé.

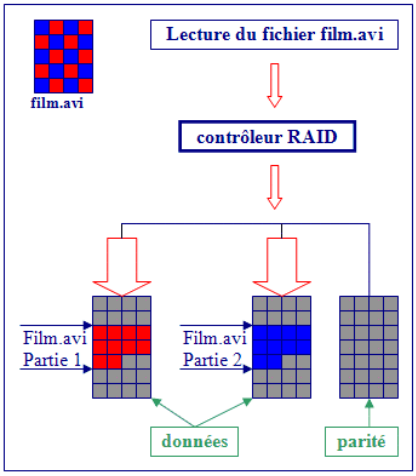


Lorsque le disque de parité tombe en panne le système continue de fonctionner en mode RAID 0 avec les disques restants qui contiennent les données. Lorsque l'on rajoute un nouveau disque de parité, les informations de parité sont reconstruites et le RAID 3 redeviens actif.



Le RAID 3 apporte donc la tolérance de panne car le système supporte la perte d'un disque quel qu'il soit (données ou parité). En revanche si deux disques durs tombent en panne simultanément, les données sont perdues.

I.F.2 Fonctionnement du RAID 3 pour un accès en lecture



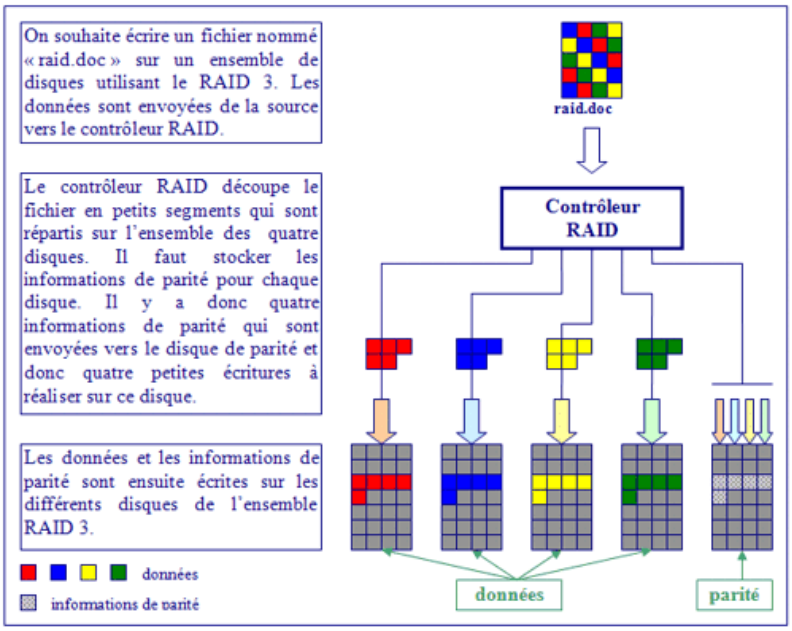
A chaque fois qu'il y a une lecture sur un système RAID 3, tous les disques (sauf le disque de parité) travaillent de concert ce qui augmente fortement les performances.

Dans l'exemple ci-contre, on veut lire un fichier nommé « film.avi » situé sur un ensemble RAID 3 composé de trois disques. Les données du fichier sont été réparties sur les deux premiers disques. Le disque de parité n'est pas sollicité lors des opérations de lecture.

Donc dans le cas d'un accès en lecture un ensemble RAID 3 composé de trois disques se comporte comme un ensemble RAID 0 composé de 2 disques. Ainsi lors de la lecture du film les deux disques travaillerons en parallèle : chacun lisant une partie des données, ce qui va quasiment multiplier les performances par deux.

Cependant, ce mode a un gros défaut. En effet comme nous l'avons vu précédemment, à chaque fois qu'on écrit un fichier sur le disque, il est découpé en morceaux qui sont copiés en même temps sur chacun des disques (sauf le disque de parité). Et à chaque écriture sur l'un des disques, il y a une écriture sur le disque de parité.

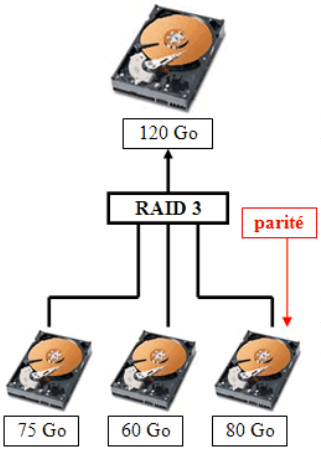
I.F.3 Fonctionnement du RAID 3 pour un accès en écriture



Cela signifie que le disque de parité travaille plus que les autres disques. Par exemple, dans un RAID 3 composé de 3 disques, le travail du disque de parité est deux fois plus important que celui des autres disques. De même dans un système RAID 3 composé de 6 disques, les accès en écriture au disque de parité seront cinq fois plus importants que ceux des cinq premiers disques. Cela est généralisable, ainsi lorsqu'on a un ensemble composé de n disques, les accès au disque de parité sont (n-1) fois plus nombreux que pour les autres disques.

On peut donc dire que le disque de parité se révèle être un véritable goulot d'étranglement au niveau des performances en écriture. De plus comme sa charge de travail est plus élevée, le disque de parité à tendance à tomber en panne très fréquemment.

I.F.4 Un petit exemple



Dans notre exemple, on met en place un RAID 3 composé de 3 disques durs. Le système d'exploitation verra un seul disque dur d'une capacité de 120 Go. Cela peut sembler étrange car la capacité physique totale des trois disques durs est de 215 Go (75 Go+ 60 Go + 80 Go = 215 Go).

Pour calculer la capacité utilisable d'un ensemble RAID 3, il faut prendre la capacité du plus petit des disques (soit 60Go dans notre exemple) et la multiplier par le nombre de disques – 1 (car l'un des disques est utilisé pour stocker les informations de parité).

On obtient donc un espace disque utilisable de 60 Go * (3 – 1) disques = 60 Go * 2 disques = 120 Go.

Au niveau des performances, l'accès en lecture sera quasiment multiplié par deux par rapport à un disque seul. En outre, si l'un des 3 disques tombe en panne alors aucune donnée ne sera perdue.

Le RAID 3 permet donc d'augmenter les performances tout en gérant la tolérance de panne. Les performances en lecture sont améliorées proportionnellement au nombre de disques (plus il y a de disques, plus les performances en lecture augmentent). Malgré une légère amélioration des performances en écriture (elle sont bridées par le disque de parité) ce système n'est pas plus intéressant que le RAID 1. En effet, sa fiabilité est moindre car seul la perte d'un seul disque est possible et car le disque de parité s'use plus vite.

I.G Le RAID 4

Ce mode nécessite au moins trois disques et est très proche du RAID 3. La seule différence notable avec le RAID 3 est l'organisation des données. En effet, avec le RAID 4, la taille des segments est variable et se modifie en temps réel. Cela implique que les informations de parité doivent être mise à jour à chaque écriture afin de vérifier si la taille des segments a été modifiée.

Dans un système RAID 4, le disque de parité devient donc un facteur encore plus limitant lors des opérations d'écriture. Cependant, les performances en lecture sont les même qu'avec le RAID 3 (c'est-à-dire excellentes). Le mode RAID 4 gère donc la tolérance de panne et apporte un gain en lecture mais pas en écriture. Généralement on lui préfère le RAID 5.

I.H Le RAID 5

Ce mode est très proche du RAID 4 car il utilise au moins trois disques durs et car il procure une augmentation des performances et gère la perte d'un disque dur.

Par rapport au RAID 4, la configuration RAID 5 répartie la parité sur l'ensemble des disques ce qui élimine le goulot d'étranglement qu'est le disque de parité en RAID 3 et en RAID 4. Un autre avantage de répartir la parité sur chacun des disques est que les disques travaillent tous autant. Cela empêche l'usure prématurée de l'un des disques (comme c'est le cas avec le disque de parité en RAID 3 et surtout en RAID 4).

Le RAID 5 écrit donc simultanément les données sur plusieurs disques ce qui améliore les performances en lecture et en écriture. Ainsi par rapport à un disque seul, les performances d'une grappe RAID 5 utilisant n disques seront (n-1) fois plus élevées (en effet, l'un des disques doit écrire la parité, la donnée écrite est donc répartie sur les n-1 disques restants).

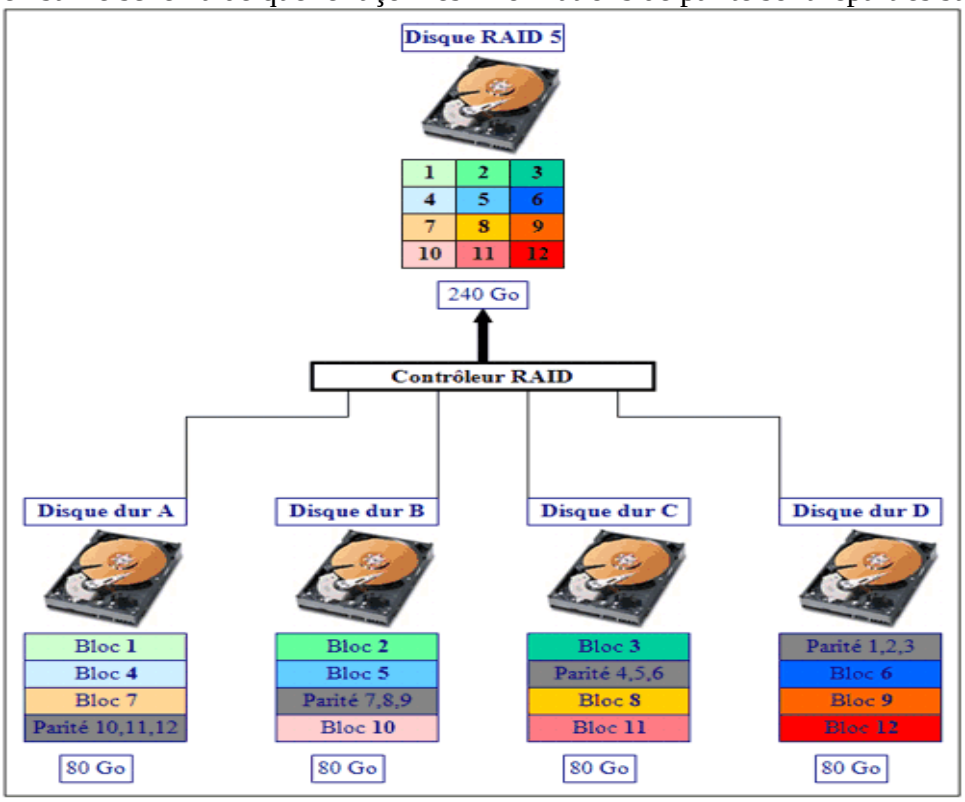
Tout comme pour le mode RAID 0, on doit sélectionner une taille de segment adaptée lorsque l'on met en place un niveau RAID 5.

Dans l'exemple suivant on a crée un grappe RAID 5 avec quatre disques durs identiques ayant une capacité de 80 Go chacun. Au final seulement 240 Go sont exploitables, ce qui représente les ¾ de l'espace disque total (320 Go). En fait l'espace d'un disque entier soit 80 Go est nécessaire afin de stocker les informations de parité. En RAID 5, on peut calculer la capacité utilisable de la manière suivante :

Capacité utilisable = (taille du disque le plus petit) * (nombre de disques – 1)

Ainsi la capacité utilisable est bien de 80 Go * (4-1) disques = 240 Go.

On remarque bien sur le schéma de quelle façon les informations de parité sont réparties sur les 4 disques.



De par ses nombreux avantages le RAID 5 est très populaire dans le monde professionnel car il apporte la tolérance de panne tout en conservant d'excellentes performances en lecture et en écriture. De plus si on utilise une interface qui supporte le branchement à chaud ou "hot plug" (SCSI ou Sata), il n'est même plus nécessaire d'éteindre le serveur. Bien entendu la mise en place d'un ensemble RAID 5 fait perdre de l'espace disque (en fait l'espace réservé à la parité qui représente la taille d'un disque). De plus la répartition de la parité sur l'ensemble des disques entraîne une remise en ordre un peu lente lorsqu'un disque dur est échangé suite à une panne.

I.I L'orthogonal RAID 5

L'Orthogonal RAID 5 est une technique logicielle crée par IBM. Ce mode est identique au RAID 5 dans toutes ses fonctionnalités.

Sa seule différence notable est l'utilisation d'un contrôleur par disque. Cela permet d'augmenter la sécurité des données. En effet, même si l'un des contrôleurs tombe en panne les données sont sauves.

I.J Le RAID 6

Le RAID 6 met en place une double redondance des données de parité. Cela signifie que les informations de parité sont stockées en doubles exemplaires. Tout comme pour le RAID 5, les informations de parité sont réparties équitablement sur l'ensemble des disques. La double redondance des données permet la perte de deux disques sans qu'aucune donnée ne soit perdue.

Le RAID 6 nécessite au moins quatre disques durs pour fonctionner. En ce qui concerne la capacité utilisable de stockage, on perd l'espace de deux disques. Donc dans un système utilisant n disques, la capacité utilisable est définie par la relation suivante :

Capacité utilisable = (taille du disque le plus petit) * (nombre de disques – 2)

Les performances en lecture comme en écriture sont augmentées. Avec un nombre de disques équivalents, un ensemble RAID 6 sera moins performant qu'un ensemble RAID 5 (car un disque de plus est utilisé pour la parité).

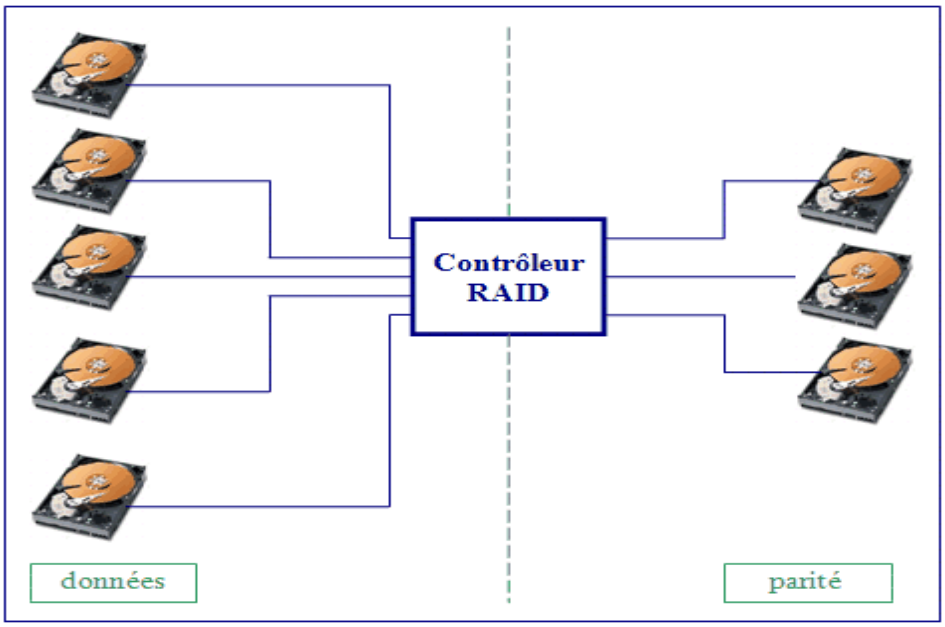
I.K Le RAID 7

Ce mode est utilisé pour avoir un niveau de sécurité ou des performances très supérieures aux autres modes RAID. Il utilise un grand nombre de disques (avec un maximum de 48 disques).

On peut définir manuellement le nombre de disques dédiés au stockage de la parité et au stockage des données. Le nombre de disques que le système peut perdre est proportionnel au nombre de disques dédiés au stockage des informations de parité.

Le RAID 7 fait appel à une carte microprocesseur qui calcule la parité, la gestion du disque et qui gère la surveillance des disques en temps réel.

Tous les transferts de données se font en mode asynchrone ce qui augmente de 1,5 à 6 fois les performances en écriture.



Il est peu utilisé car il très coûteux en équipement et difficile à mettre en place.

I.L Conclusion

Pour conclure, les modes les plus intéressant sont les modes 0, 1 et 5. Les autres modes sont peu ou pas utilisés car :

- le RAID 2 est obsolète
- le RAID 3 et 4 utilisent un disque de parité ce qui crée un goulot d'étranglement.
- le RAID 6 est onéreux
- le RAID 7 est très onéreux et difficile à mettre en place

Voici un petit tableau comparatif des trois modes les plus intéressants fonctionnant avec 10 disques durs d'une capacité unitaire de 120 Go et ayant un débit réel en lecture/écriture de 70/50 Mo/s.

	RAID 0	RAID 1	RAID 5
capacité	1200 Go (100 %)	120 Go (10 %)	1080 Go (90%)
débit en lecture	700 Mo/s	700 Mo/s	620 Mo/s
débit en écriture	500 Mo/s	50 Mo/s	450 Mo/s
Tolérance de panne (nombre de disques)	aucun	jusqu'à 9 disques	un disque

Le système ayant le meilleur rapport capacité/performances/sécurité est sans conteste le RAID 5. C'est d'ailleurs lui le plus utilisé en Entreprise.

Cependant, il est possible de combiner différents niveaux de RAID, ce qui permet d'obtenir des alternatives très intéressantes au RAID 5. Dans la seconde partie nous allons voir ces niveaux combinés.

II Les Niveaux Combinés

II.A Introduction

Les niveaux de RAID offrent un large choix d'alternatives qui permettent de choisir un rapport performance/sécurité différent. Il est aussi possible de combiner deux niveaux de RAID entre eux. On parle alors de niveaux combinés.

Certaines cartes permettent de combiner différents modes (ou niveaux) RAID en hardware mais elle sont chères mis à part les cartes RAID IDE d'entrées de gamme qui gèrent toutes le RAID 01 ou 10 selon les cas. Cependant ces dernières ne sont pas intéressantes pour réaliser des combinaisons étant donné que ces « cartes bas de gamme » n'intègrent qu'un seul contrôleur (donc pas de tolérance de panne pour le contrôleur) et que le nombre de disques utilisables est souvent limité à quatre.

Généralement la manière la moins onéreuse pour réaliser une combinaison entre deux modes RAID est :

1) d'utiliser plusieurs contrôleurs au niveau matériel pour mettre en place plusieurs ensemble en mode RAID X.
2) de créer un RAID Y au niveau logiciel à partir de tous les ensembles RAID X.

L'ensemble obtenu sera un ensemble fonctionnant en un mode RAID nommé XY. Cela signifie que l'on met d'abord en place plusieurs RAID X, puis qu'on les rassemble avec un mode RAID Y. Un RAID XY est donc différent d'un RAID YX.

Voici un exemple : admettons que l'on dispose de 9 disques durs de 20 Go chacun, de trois cartes RAID gérant le RAID 0 et le RAID 1 ainsi que d'un 7è disque dur (branché sur le contrôleur intégré à la carte mère) où est installé Windows Server 2003. On peut réaliser deux combinaisons intéressantes :

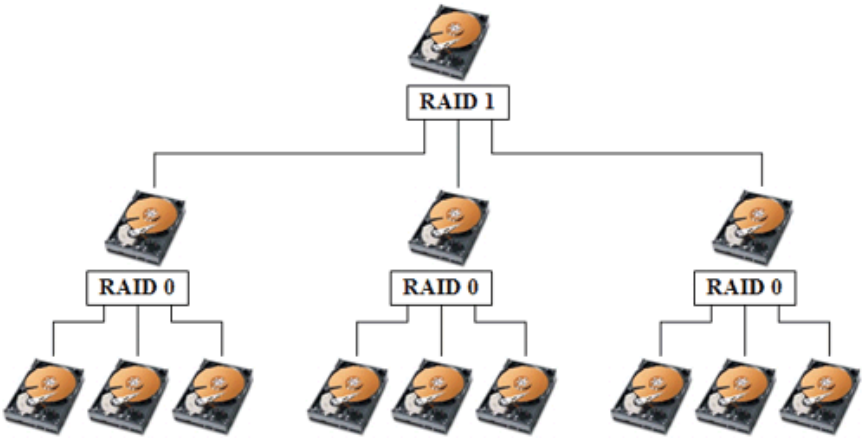
- un RAID 01
- un RAID 10

II.B Le RAID 01

On divise les disques en trois groupes de trois. On met trois disques sur chacune des cartes RAID et on crée trois ensembles RAID 0.

Chaque ensemble RAID 0 comporte 3 disques de 20 Go chacun, ce qui donne une capacité de 60Go par ensemble.

Ensuite, on crée un ensemble RAID 1 sous Windows Server 2003 avec les trois ensembles matériel en RAID 0. On obtient donc après formatage, un seul disque d'une capacité de 60Go.



Chaque ensemble matériel en RAID 0 a des performances en lecture/écriture trois fois plus élevées environ qu'un disque seul. L'ensemble logiciel en RAID 1 propose donc des performances triples à celles d'un disque seul.

Si l'un des six disques vient à tomber en panne alors, toutes les données stockées sur l'ensemble matériel RAID 0 auquel il appartient sont perdues. On se retrouve avec seulement deux ensembles matériels en RAID 0 et donc 6 disques utilisés. Si on perd ensuite un second disque il ne reste plus qu'un ensemble RAID 0. Enfin si un dernier disque tombe en panne alors, l'ensemble subsistant et donc toutes les données seront perdus.

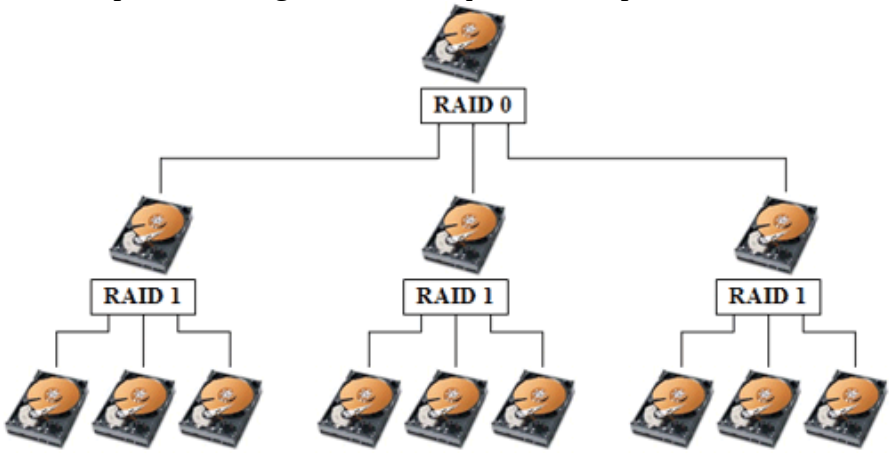
Dans notre exemple, grâce au RAID 01, on obtient des performances triplées par rapport à un disque seul et le système supporte la perte de deux disques dur.

II.C Le RAID 10

On divise les disques en trois groupes de trois. On met trois disques sur chacune des cartes RAID et on crée trois ensembles RAID 1.

Chaque ensemble RAID 1 comporte 3 disques de 20 Go chacun, ce qui donne une capacité de 20Go par ensemble.

Ensuite, on crée un ensemble RAID 0 sous Windows Server 2003 avec les trois ensembles matériel en RAID 0. On obtient donc après formatage, un seul disque d'une capacité de 60Go.



Chaque ensemble matériel en RAID 1 a des performances en lecture/écriture égales à un disque seul. L'ensemble logiciel en RAID 0 propose donc des performances triples à celles d'un disque seul.

Si l'un des six disques vient à tomber en panne alors, l'ensemble RAID 1 auquel il appartient continue à fonctionner normalement.

Si on perd un second disque dans le même ensemble RAID 1, l'ensemble RAID 0 logiciel n'est pas affecté non plus.

Si on perd le troisième disque l'ensemble matériel RAID 1 ne fonctionne plus ce qui implique que les données stockées sur l'ensemble logiciel RAID 0 sont perdues. Ainsi en RAID 10, on perdre jusqu'à 2 disques par ensemble RAID 1 (dans notre exemple) soit un maximum de 6 disques sur les 9 disques.

Tableau comparatif RAID01/RAID 10 avec 9 disques de 20 Go (50Mo/s en lecture/écriture) et trois contrôleurs RAID :

	RAID 01	**RAID 10**
capacité	60Go	60Go
débit en lecture/écriture	150Mo/s	150Mo/s
tolérance de panne	2 disques maxima	6 disques maxima

Pour conclure, les performances en RAID 01 sont les mêmes qu'en RAID 10. Cependant l'ensemble RAID 10 se distingue au niveau de la tolérance de pannes car il peut perdre (dans notre exemple) jusqu'à 6 disques contre 2 uniquement pour le RAID 01.

C'est pour cette raison essentielle, que le RAID 10 est préféré au RAID 01.

II.D Le RAID 50 et les autres combinaisons

Une autre combinaison très utilisée est le RAID 50. En effet, elle permet d'obtenir d'excellentes performances tout en gérant la tolérance de panne.

D'une manière générale toutes les combinaisons en RAID X0 sont misent en place. En effet, on implémente d'abord un mode X qui gère la tolérance de panne (1, 3, 4, 5 ou 6) puis un mode RAID 0 afin de maximiser les performances. Bien entendu, il est tout à fait possible au niveau théorique de combiner des modes qui implémente la tolérance de panne. On peut par exemple, penser au RAID 55 ou au RAID 53. Cependant, en pratique, les constructeurs n'implémente quasiment jamais de tels modes car si l'on combine deux modes qui demande un contrôle de parité alors il faut disposer d'une carte RAID extrêmement puissante afin de pouvoir réaliser les calculs de parité sur deux niveaux.

Les combinaisons les plus utilisées sont donc le RAID 10, 30 et 50. Les combinaisons RAID 40 et 60 sont rarement rencontrées en raison du succès commercial mitigé (voire nul de nos jours) du RAID 4 et du RAID 6.

II.E Le RAID 0+1 et RAID 53 : des appellations trompeuses?

Les cartes RAID d'entrées de gamme sont soit au format PCI 32bits, soit directement intégrées à la carte mère. Actuellement ces circuits disposent en standard de contrôleurs RAID gérant le RAID 0 et le RAID 1. Cependant de nombreux constructeurs annoncent que leur carte supporte un niveau combiné nommé « RAID 0+1 ».

La signification de cette appellation varie selon les constructeurs. Elle désigne soit du RAID 01 ou soit du RAID 10. Par exemple, certains contrôleurs RAID IDE de la marque Promise gèrent du RAID 10, alors que la plupart contrôleurs RAID IDE de la marque Highpoint gèrent du RAID 01.

De manière générale, les constructeurs implémentent du RAID 01 car il est beaucoup plus simple à mettre en place au niveau électronique que le RAID 10 (il faut des composants plus performants et plus nombreux pour le RAID 10). Il est dommage que les constructeurs utilisent cette notation avec des « + » qui trouble par rapport à la notation classique.

Le RAID 53 est aussi une appellation fallacieuse. En effet la plupart du temps, le terme RAID 53 ne désigne pas un niveau RAID 5 combiné avec un niveau RAID 3, mais niveau de RAID 0 combiné avec du RAID 3. Ce terme désigne donc du RAID 03. Il arrive parfois que les constructeurs appellent RAID 53 du RAID 30.

Ces appellations trompeuses sont un argument marketing utilisé par les fabriquant de carte RAID (la plupart du temps, le consommateur se fait duper car il pense que sa « carte RAID 0+1 » supporte le RAID 10 alors qu'en réalité ce n'est que très rarement le cas). Cela est vraiment dommage car cela crée un climat d'incertitude qui rend cette technologie encore plus obscure pour les novices.

II.F Conclusion

On peut donc combiner des niveaux RAID simples afin d'augmenter les performances et/ou la tolérance de panne. Le niveau combiné le plus utilisé est de loin le RAID 10 (pour sa fiabilité).

La façon la plus simple de créer un niveau combiné est de mettre en place du RAID matériel et du RAID logiciel.

Mais en pratique, cette solution n'est quasiment jamais utilisée puisque on lui préfère pour diverses raisons (cf. : ces raisons sont abordées dans la partie traitant du RAID matériel) une solution 100% matérielle.

Il est donc impératif de bien choisir le matériel à utiliser pour chaque niveau (simple ou combiné) de RAID que l'on veut mettre en place.

III Le RAID MATERIEL

III.A Introduction

Le RAID matériel est très utilisé dans le monde professionnel. En effet, beaucoup d'Entreprises ne peuvent se permettre de perdre des données ou bien de voire le réseau paralysé à cause d'une panne sur un disque dur.

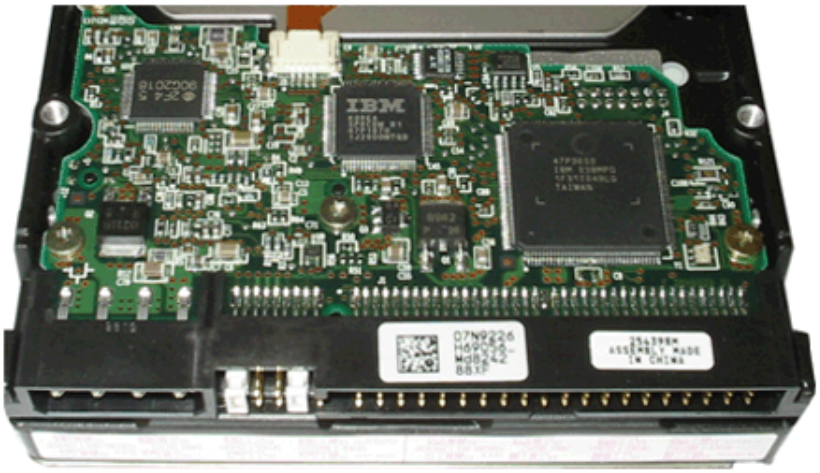
C'est pourquoi de nombreux serveurs qui reçoivent beaucoup de requêtes (serveur Web, serveur de messagerie, serveur hébergeant une base de données, …) ou qui manipulent de gros volumes de données (serveur de fichiers) utilisent le RAID notamment matériel.

III.B Quelques petits rappels

III.B.1 La notion de contrôleur

Très souvent dans cet article, les composants qui jouent le rôle d'interface entre le chipset sud de la carte mère et les disques durs sont nommés contrôleurs.

Par exemple, on nomme contrôleur RAID IDE, un jeu de composants qui permet d'interconnecter les disques durs IDE et le chipset sud (ou southbridge) de la carte mère tout en supportant la mise en place d'un RAID matériel. En réalité, cette appellation est fausse car le contrôleur est la puce qui contrôle les actions du disque dur (déplacement des têtes, état du cache,…).



Cette puce est située sur le disque dur lui-même et non sur la carte mère. Le jeu de composants qui permet de relier le disque et la carte mère est en fait un simple adaptateur et devrait être nommé comme tel.

Cependant, par abus de langage, les utilisateurs mais aussi les constructeurs désignent ce jeu de composant « contrôleur ». C'est pourquoi dans la suite de cedocument, quand vous verrez le mot « contrôleur RAID SCSI », il faudra comprendre adaptateur situé sur la carte mère ou sur une carte d'extension qui propose une interface SCSI et gère la technologie RAID.

III.B.2 RAID PATA, RAID SATA et RAID SCSI

Dans la suite de cet article, lorsque le terme RAID IDE (ou RAID PATA) est utilisé cela signifie RAID avec des disques durs compatibles avec l'interface PATA (Parallel ATA), et lorsque le terme RAID SCSI est utilisé cela signifie RAID avec des disques durs compatibles avec l'interface SCSI. De même le terme RAID SATA désigne des ensembles de disques durs utilisant la technologie RAID composés de disques durs utilisant l'interface SATA (Serial ATA).

III.B.3 Bande passante et débit

Les termes bandes passantes et débits sont très souvent utilisés dans cet article et en informatique de manière générale. C'est pourquoi il ne faut pas les confondre. La bande passante représente la quantité de donné maximale que l'on peut transférer en un temps donné. Le débit représente la quantité de données réelles que l'on peut transférer en un temps donné.

III.C le RAID IDE (ou RAID PATA)

Pour les stations de travail, le problème du matériel à utiliser pour faire du RAID ne se posait pas il y a quelques années.

En effet, il y a 6 ans la technologie RAID n'était disponible que pour les disques durs SCSI qui par leurs prix élitistes les réservaient au monde professionnel. Cependant grâce à la société Promise technology qui a commercialisée le premier contrôleur RAID IDE nommé FastTrack, la technologie RAID a pu être adaptée pour les disques durs à la norme IDE.

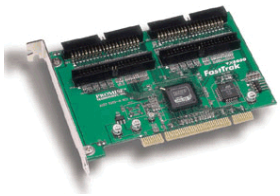
L'apparition des contrôleurs RAID IDE a fortement popularisé la technologie RAID. Cette technologie est maintenant utilisée par beaucoup d'utilisateurs avancés ou « power users » qui souhaitent obtenir de un plus grand espace de stockage, de meilleures performances ou bien encore maximiser la sécurité de leurs données.

Dorénavant, un grand nombre de cartes mères « grand public » intègrent en standard un contrôleur RAID IDE. Cependant les constructeurs de cartes mère (Asus, Abit, Gigabyte, MSI, Elite …) peuvent intégrer ce contrôleur de deux manières différentes :

- Il peut être relié directement au chipset sud (ou southbridge) de la carte mère. Ce cas de figure permet un maximum de performance car le contrôleur est généralement relié au southbridge avec des bus (ou liaisons électriques) rapides (exemple : 533Mo/s).
  Il peut être rajouté sur la carte mère grâce à une connexion sur le bus PCI. Cependant, le bus PCI possède une bande passante limitée (133Mo/s par canal). En outre, La plupart du temps les constructeurs de carte mère n'utilise qu'un seul canal pour des raisons de coût. Dans ce cas la bande passante réservée au contrôleur RAID est de 133Mo/s moins la bande passante que les périphériques branchés sur les ports PCI (cartes son, cartes réseau, …) vont consommer. Cela pourra sérieusement brider les performances en mode RAID 0 par exemple.

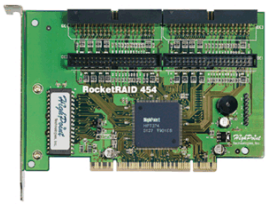
La seconde façon de mettre en place du RAID matériel est d'utiliser une carte contrôleur dédiée. Généralement ces cartes sont au format PCI 32 bits et en de rares occasions au format PCI 64bits. Il faut être vigilant lorsque l'on met en place du RAID sur une carte d'extension notamment au format PCI 32bits. En effet, selon les cartes mères l'ensemble des ports PCI 32bits se partage un ou plusieurs canaux avec une bande passante de 133Mo/s chacun. En admettant que la carte mère ne propose qu'un seul canal et donc une bande passante de 133Mo/s pour tous les ports PCI, il faudra veiller à ne pas mettre deux cartes RAID possédant chacune 4 disques en RAID 0 par exemple car dans ce cas les performances se retrouveraient sérieusement bridées par l'interface limitée elle à 133Mo/s.

De manière générale les cartes contrôleur IDE ne possèdent pas ou très peu de mémoire cache. Elles sont quasiment toutes au format PCI 32 bits (avec une bande passante de 133 Mo/s) et implémentent le RAID 0, 1 et 0+1 (soit le niveau 10, soit le niveau 01).



Ci-contre une carte FastTrak TX4000 de la marque Promise au format PCI 32 bits. Cette carte supporte 4 canaux ATA 133 et donc jusqu'à 8 disques durs en Ultra-DMA 133. Elle gère les niveaux RAID 0,1 et 10.

Cependant, il existe quelques cartes contrôleur RAID IDE « haut de gamme ». Ces cartes au format PCI 64 bits (avec une bande passante de 266 Mo/s par canal) implémentent le niveau RAID 5 et peuvent parfois gérer jusqu'à 16 périphériques et accueillir jusqu'à 256 Mo de mémoire (comme la carte Promise SuperTrack SX6000).



Ci-contre, une carte RocketRAID 454 de la marque HighPoint qui possède quatre canaux Ultra ATA 133 et permet donc l'installation de 8 disques durs en UDMA 133.Cette carte implémente les niveaux RAID 0, 1, 5, 10 et JBOD.

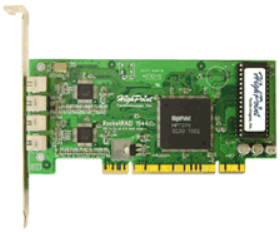
III.D Le RAID SATA

Depuis la mise en place du Serial-ATA, les cartes RAID fleurissent sur le marché.

En effet, cette nouvelle interface est bien plus propice à la mise en place du RAID que l'ancienne interface ATA (encore appelée IDE).

D'une part grâce aux débits plus élevés (jusqu'à 1,5Gb/s soit 187,5Mo/s par canal et bientôt 320Mo/s par canal avec le SATA2) et d'autre part grâce à la gestion du branchement à chaud ou hotplug.

Les constructeurs rivalisent donc d'ingéniosité pour sortir de nouvelles cartes contrôleur RAID SATA.



On peut évoquer la carte RocktetRAID 1544 de la marque HighPoint (format 64 bits).

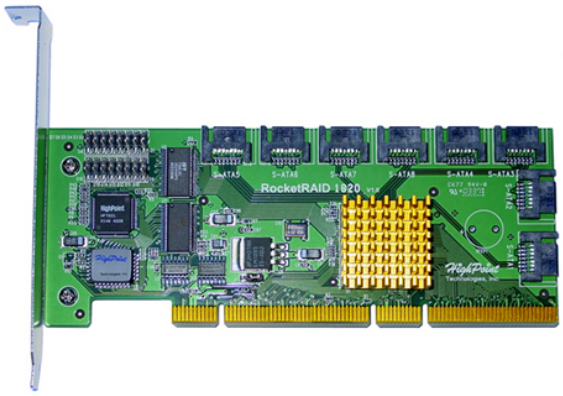
Cette carte innove car elle possède 4 canaux eSATA (external SATA) proposant chacun une bande passante de 1.5 Gb/s soit 187.5 Mo/s.

L'eSATA est une technologie propriétaire mise en place par HighPoint permettant de connecter des disques durs SATA externes (par l'intermédiaire de câbles à la norme IEEE 1394). On peut donc mettre en place un RAID matériel même si le boîtier ne contient plus d'emplacements 3,5 pouces de libre. Cette carte aussi supporte les niveaux RAID 0, 1, 5, 10 et JBOD.

Il est noter que la marque Hightpoint commercialise aussi une version plus "huppée" de la RocketRAID 1544, nommée RocketRAID 1820.

Elle propose de hautes performances et est spécialement destinée au monde de l'Entreprise. En effet, si elle ne gère pas l'eSATA, cette carte dispose de 8 canaux Serial ATA "classique" ce qui permet de connecter jusqu'à 8 périphériques.

De plus elle utilise le format PCI-X 133MHz qui permet une bande passante de 1200 Mo/s. Comme sa petite soeur, elle implémente les niveaux RAID 0, 1, 5, 10 et JBOD.



carte RocketRAID 1820 de la marque Highpoint

III.E Le RAID SCSI

La technologie RAID a été pensée pour le SCSI. Ainsi, certaines cartes mères destinées aux serveurs implémentent des contrôleurs SCSI qui gèrent le RAID directement sur la carte. Mais dans la plupart des cas on doit installer une carte d'extension au format PCI-X, PCI 64 bits voire même PCI 32 bits (peu recommandé) pour mettre en place un RAID matériel.

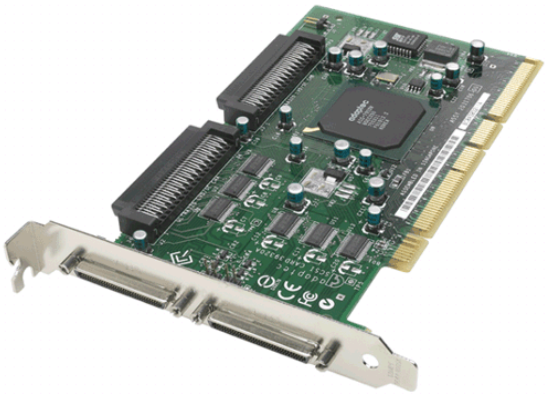
La plupart des cartes contrôleur supportant le RAID SCSI sont très onéreuses. Ce coût élevé est dû en partie la présence de microprocesseurs performants (pour synchroniser les échanges de données entre le bus PCI et les disques branchés sur la carte et surtout pour le calcul des informations de parité) ainsi qu'à la présence d'une grande quantité de mémoire cache. Généralement la mémoire utilisée est de la mémoire SD-RAM ECC. La quantité peut varier de 32 Mo à 256 Mo de mémoire.

La présence d'autant de mémoire est justifiée. En effet, la quantité de mémoire est un facteur déterminant pour les performances surtout lorsque l'on met en place un niveau de RAID avec parité (RAID 3, 4, 5, 6, 30, 50, ...). Grâce à des cartes contrôleurs et des disques très performants, le RAID SCSI permet d'atteindre des débits impressionnants. Il n'est pas rare de dépasser les 500 Mo/s en lecture avec un niveau RAID 10 ! C'est pourquoi lorsque l'on veut mettre en place un niveau de RAID, il faut sélectionner la bonne carte.

En effet, si l'on souhaite mettre en place un niveau RAID 5 avec quatre disques tournant à 10 000 tr/min, une carte au format PCI 64 bits (c'est-à-dire avec une bande passante de 266 Mo/s) suffira.

Une carte Tekram DC-390U3W au format PCI 64 bits

Par contre, si l'on souhaite mettre en place un niveau RAID 10 avec seize disques tournant à 15 000 tours/min, une carte au format PCI-X (le format PCI-X permet une bande passante de 1.06 Go/s) possédant beaucoup de mémoire est fortement recommandée.



Une carte Adaptec 39320A-R au format PCI-X

Dans le cas où l'on doit mettre en place un ensemble RAID performant, il faut essayer dans la mesure du possible d'harmoniser les composants entre eux. En effet, si l'on branche beaucoup de disque dur SCSI Ultra Wide 320 sur une carte possédant un seul canal limité à 160 Mo/s, il y a de fortes chances pour que le canal sature.

III.F RAID IDE, SATA, ou SCSI : Que choisir ?

La foudroyante propagation de la technologie RAID IDE ces dernières années est due au prix des disques durs utilisant cette interface. En effet, ils sont bien moins onéreux que leurs homologues SCSI à capacité égale.

Ils sont très intéressants pour les particuliers. Mais ils sont aussi une opportunité pour les petites entreprises ayant un budget limité et qui souhaitent sécuriser leurs données à moindre coût.

L'interface PATA qui a atteint ses limites (largeur des nappes, débits limités à cause de la trop grande sensibilité des transfert parallèles aux interférences électromagnétiques, pas de hotplug, ...) est en passe d'être remplacé par l'interface Serial ATA. Le SATA se positionne dans la même gamme de prix que le PATA. C'est pourquoi nous allons faire l'impasse sur le PATA et essayer de savoir quelle interface est la plus intéressante. Alors : SCSI ou SATA ?

Voici un petit tableau qui compare les caractéristiques de deux disques durs de la marque Seagate : le premier utilise l'interface Serial-ATA et le second utilise l'interface SCSI.

Les informations contenues dans le tableau suivant ont été relevées sur le comparateur de prix : www.prixdunet.com le 25/01/2004.

	Seagate Barracuda	Seagate U320 Cheetah
Modèle	7200.7	10K.6
Interface	Serial-ATA (SATA 150)	SCSI (Ultra Wide 320)
Capacité *(non formaté)*	160 Go	146.8 Go
Vitesse de rotation	7200 tr/min	10 000 tr/min
Taille de la mémoire cache	8 Mo	8 Mo
Bande passante théorique	150 Mo/s	320 Mo/s
Débit en lecture/écriture *(chiffre très approximatif)*	~50 Mo/s	~100 Mo/s
Prix	114,70 euros	965 euros

Il est indéniable que pour une même capacité, le rapport performance/prix du disque Serial-ATA est supérieur à celui du disque SCSI. En effet, pour le prix du disque SCSI, on peut acheter 7 disques utilisant l'interface Serial-ATA.

Le disque SCSI ne peut justifier cette différence de prix malgré ses meilleures performances (de part leurs vitesses de rotation plus élevées, les disques SCSI offrent de meilleurs débits et surtout des temps d'accès moyens réels plus faibles que les disques SATA) et son temps moyen entre pannes (ou MTBF = Mean Time Between Failure) supérieur.

Voici maintenant un petit comparatif entre deux cartes contrôleurs de la marque LSI. L'une des cartes supporte l'interface SCSI et l'autre l'interface Serial-ATA. La carte à la norme SCSI permet d'accueillir plus de disques durs et gère plus de modes RAID différents, cependant elle est plus chère et les performances seront fortement bridée par le format PCI 32 bits utilisé ici.

	LSI MegaRAID SATA 150-6	LSI Mylex AcceleRAID 170
Format de la carte	PCI 64 bits (266Mo/s)	PCI 32 bits (133 Mo/s)
Taille de la mémoire cache intégrée à la carte	64 Mo	64 Mo
Niveaux de RAID supportés	0, 1, 5, 10, 50	0, 1, 3, 5, 01, 10 ,30, 50
Nombre de disques maximum	8	15
Prix	519 euros	700 euros

On peut donc conclure que les prix des cartes contrôleurs « haut de gamme » sont sensiblement les mêmes que ça soit pour le Serial ATA ou bien encore le SCSI. L'avantage des disques durs Serial ATA reste donc leur prix peu élevé par rapport aux disques SCSI.

Pour conclure, l'interface Serial-ATA est donc une excellente alternative à l'interface SCSI sauf bien entendu sur les gros serveurs de fichiers pour lesquels les performances passent avant le prix

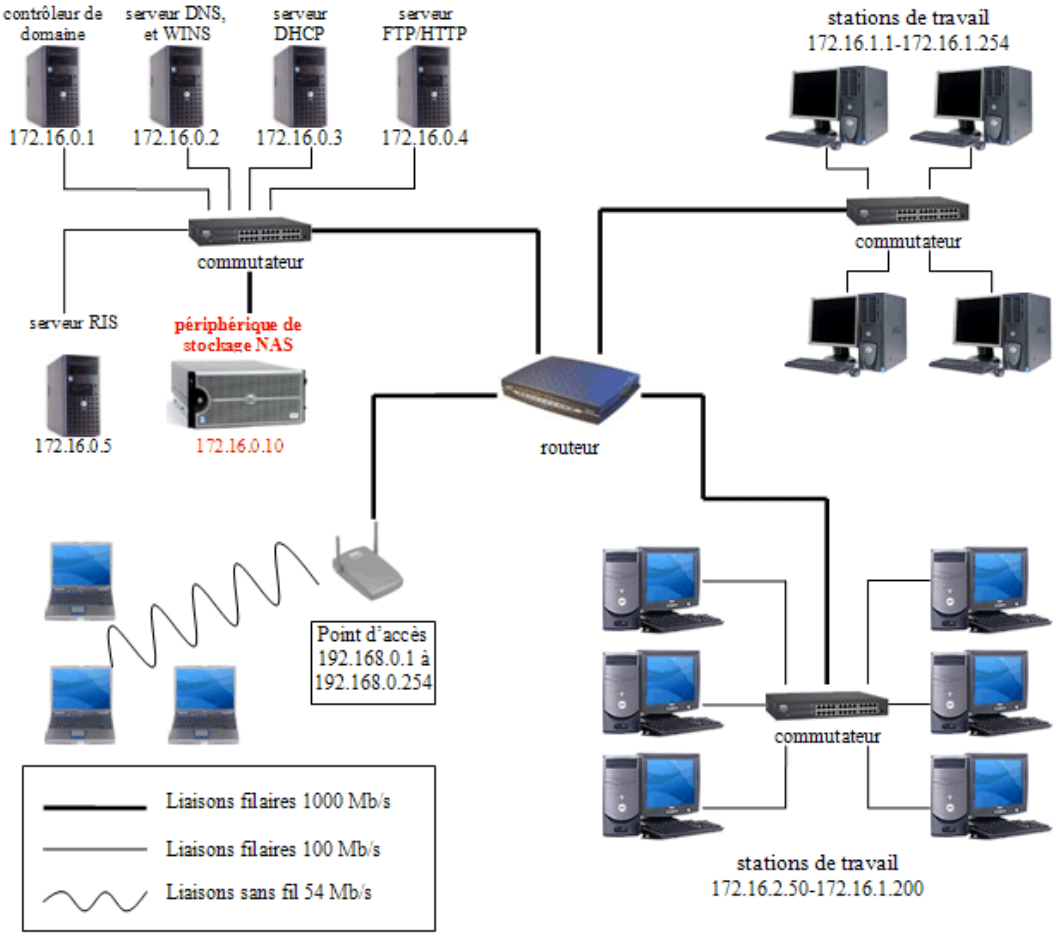
III.G Les dispositifs de stockages réseau NAS

Il existe autre solution de stockage fréquemment utilisé par les Entreprises : les périphériques de stockage en réseau. Il existe deux type de technologies de stockage en réseau : le NAS (pour Network Attached Storage) et le SAN (pour Storage Area Network).

III.G.1 Le NAS (Network Attached Storage)

Les périphériques de stockage NAS utilisent le réseau Ethernet de l'entreprise. Cela signifie qu'un périphérique de stockage NAS possèdera une ou plusieurs interfaces réseau et sera reconnu sur le réseau grâce à son adresse IP.

Dans l'exemple ci-dessous, un périphérique de stockage NAS est configuré avec l'adresse IP 172.16.0.10. Tous les ordinateurs du réseau local (LAN) peuvent le joindre et peuvent accéder aux données qu'il contient (si les autorisations d'accès le permettent).



Les NAS se présentent la plupart du temps sous la forme de boîtiers externes possédant leurs propre source d'alimentation. Ces boîtiers sont soit autonomes, soit intégrables dans des backbones (généralement dans des racks 1U, 2U ou 3U d'une taille de 19'').

Ils contiennent généralement plusieurs disques durs le plus souvent avec une interface SCSI. Les NAS permettent de mettre en place le RAID très facilement. On peut ainsi obtenir des unités de stockages très fiables et faciles à mettre en place. En effet, les serveurs de fichiers qui utilisent « une machine complète » ont beaucoup plus de chances de tomber en panne que les NAS (surchauffe du processeur, barrette de mémoire grillée ou tout simplement plantage du système d'exploitation).

Le plus souvent les NAS sont reliés au réseau via une prise RJ45 gigabit. On trouve parfois des NAS d'entrée de gamme utilisant une prise RJ45 en 10/100. La configuration d'un NAS se fait généralement grâce à une interface Web (c'est-à-dire qu'ils contiennent un serveur HTTP).
Les NAS permettent donc de partager des données sur le réseau tout en proposant une grande fiabilité (ils sont à l'abris des pannes d'ordre logiciel), des performances de premier ordre (grâce au RAID), une intégration et une maintenance facilité (le plus souvent au format 19'', monitoring en façade, système de racks, disques échangeables à chaud en cas de panne,…). En fait, le seul inconvénient des NAS est leur prix élevé qui les met hors de la portée des petites entreprises.

III.G.2 Le NDAS (Network Direct Attached Storage)

Attention !!! Il ne faut pas confondre NAS (Network Attached Storage) et NDAS (Network Direct Attached Storage). En effet, les NDAS se présentent aussi sous la forme de boîtiers contenant des disques durs.

Contrairement aux NAS, les périphériques NDAS ne sont pas directement reliés au réseau mais doivent passer par un ordinateur pour communiquer avec le réseau. Il ne sont donc pas totalement autonomes (la plupart du temps, il faut installer un pilote pour que les ordinateurs du réseau puissent communiquer avec un périphérique NDAS). De plus les NDAS contiennent généralement un seul disque et n'implémentent donc pas la technologie RAID. Cela signifie que les NDAS ne gèrent pas la tolérance de pannes.

Ci-contre le Netdisk de la marque Ximeta est un disque dur NDAS. Il contient un disque dur de 80, 120 ou 160Go (selon les versions) tournant à 7200tr/min et possédant 2Mo de cache. Il peut se raccorder à une machine ou au réseau grâce à un port USB 2.0 ou bien grâce à une prise RJ45. De plus on peut le paramétrer avec une interface web.

Pour se raccorder à la machines, les NDAS utilisent le plus souvent l'interface IEEE1394 (firewire) qui autorise une bande passante de 400Mb/s ou bien encore l'interface USB 2.0 qui autorise une bande passante de 480Mb/s (parfois les NDAS utilisent l'USB 1.1 qui offre une bande passante de 12Mb/s...). Certains NDAS, possèdent aussi des prises RJ45 et proposent une interface web.

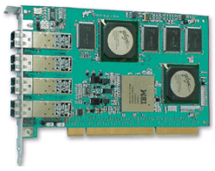
III.H Les dispositifs de stockage dans un réseau SAN

Contrairement au NAS, le SAN n'utilise pas un réseau de type Ethernet mais un réseau de type Fibre Channel (FC). Les médias utilisés pour un réseau de type Fibre Channel sont soit des câbles en cuivre, soit des fibres optiques (monomode ou multimode) qui supportent une bande passante de 2Gb/s. On préfère généralement l'utilisation de fibres optique qui permettent d'utiliser des câbles allant jusqu'à 10 Km (fibre optique monomode) contre 30 m pour les câbles en cuivre.

Le protocole utilisé dans les réseaux SAN étant différent du protocole Ethernet, il nécessite un matériel différent. Ainsi on trouve des cartes réseaux, des disques durs, des concentrateurs et des commutateurs utilisant l'interface Fibre Channel. De plus il faut installer des composants logiciels sur les ordinateurs branchés sur le réseau SAN afin d'administrer la sécurité et les autorisations d'accès aux données.

III.H.1 Les Adaptateurs de bus Fibre Channel (HBA)

Les cartes réseaux compatibles avec le protocole Fibre Channel sont généralement appelées Host Bus Adaptateur (HBA) ou adaptateur de bus Fibre Channel pour utiliser le terme francisé. Les adaptateurs de bus Fibre Channel sont généralement proposés sous la forme de cartes d'extension au format PCI-X (bande passante de 1.06 Go/s) ou bien PCI 64bits (bande passante de 266Mo/s). Mais on commence à trouver sur le marché des adaptateurs au récent format PCI Express 4X qui fournissent une bande passante de (266 Mo/s) * 4 = 1064 Mo/s.



Ci-contre, un adaptateur de bus Fibre Channel de marque Qlogic. Il utilise le format PCI-X qui propose une bande passante de 1.06 Go/s ce qui lui permet de proposer 4 ports Fibre Channel d'une bande passante de 2 Gb/s (soit 250Mo/s) chacun.

La marque Qlogic vient aussi de commercialiser un adaptateur utilisant le format PCI Express dans sa version 4X. Cette carte communique donc avec la carte mère avec une bande passante de 1064 Mo/s. Elle offre 2 ports Fibre Channel proposant chacun une bande passante de 2Go/s chacun.

III.H.1.1 Les disques durs Fibre Channel

Il existe aussi des disques durs possédants une interface Fibre Channel. On peut les brancher de deux différentes façons :

- directement sur un commutateur Fibre Channel
- dans une baie de disques durs qui, elle, sera raccordée à un commutateur

De manière générale, les constructeurs de disques durs créent des disques utilisant l'interface Fibre Channel à partir de leurs gammes SCSI.

Ainsi, on peut trouver des disques durs utilisant l'interface Fibre Channel dans la gamme Ultrastar d'Hitachi (anciennement IBNM) ou bien dans la gamme Cheetah de Seagate.

III.H.1.2 Les baies de stockage SAN

Lorsque l'on souhaite mettre en place un système de stockage performant, il faut utiliser une baie de stockage intégrée dans un réseau SAN. Les baies de stockages utilisée dans les réseaux SAN sont généralement au format 19'' ce qui permet de les « racker » dans des backbones.

On peut parfois rassembler plusieurs baies de stockage pour former un module de stockage et ainsi atteindre de plus grandes performances.

Ci-contre un rack 3U au format 19'' de la marque Adaptec. Ce modèle (le SANbloc 2Gb RAID) peut accueillir jusqu'à 14 disques SCSI de 3.5''. Il possède deux contrôleurs RAID (pour la tolérance de panne) et implémente les modes 0, 1, 3, 5, 10, 30, 50. Il est relié au réseau grâce à deux prises optiques avec une bande passante de 1 Gb/s chacune.

Voici une autre baie de stockage SAN, de marque Dell | EMC (modèle CX700). Cette baie peut contenir jusqu'à 15 disques durs avec une interface Fibre Channel ou ATA pour une capacité de stockage totale de 35To en FC et 74,16To en ATA. Elle gère les niveaux RAID 0, 1, 3, 5 et 10. Cette baie incorpore aussi 8 Go de mémoire cache.

Le CX 700 propose 8 ports FC en façade (cela signifie qu'on peut brancher jusqu'à 8 ordinateurs directement sur la baie de disque) et des ports Fibre Channel situés à l'arrière afin de relier le CX 700 à d'autres baies de stockages CX 700.
On peut ainsi créer un gros module de stockage utilisant la technologie RAID en raccordant 16 baies de stockage CX 700 entres-elles. Cela permet donc de créer un module de stockage composé d'un maximum de (15 disques/baie) * (16 baies interconnectées) = 240 disques durs accessible par un maximum de (8 ordinateurs/baie) * (16 baies interconnectées) = 128 ordinateurs.

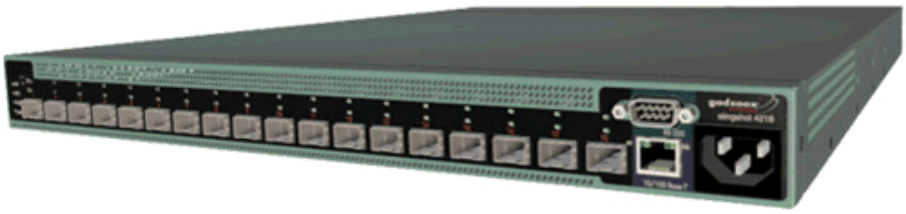
Selon Dell et EMC, les performances d'une seule baie CX 700 peuvent autorisent un débit constant supérieur à 1300 Mo/s. Les performances d'un module de stockage composé de plusieurs CX 700 peuvent donc procurer des performances de très haut niveau.

MAJ (13/02/2006) : Les caractéristiques du CX 700 ont été revues à la hausse depuis la publication originale de cet article. Ainsi les modèles actuels peuvent toujours embarquer jusqu'à 240 disques (en combinant plusieurs bais entre-elles) et la capacité maximale est passée à 120To (grâce au support des disques durs PATA de 500Go). La bande passante frontal de la nouvelle édition est de 1530 Mo/s (environ 200 000 entrées/sorties par seconde). De plus, il est possible de raccorder jusqu'à 256 serveurs à une même baie CX 700 en utilisant le commutateur approprié !

III.H.1.3 Les commutateurs SAN

Pour relier tous les ordinateurs, les disques durs et les baies de disques durs utilisant l'interface Fibre Channel, il faut utiliser un commutateur (ou switch) supportant le Fibre Channel. Il existe des commutateurs proposant des ports Fibre Channel (2Gb/s) et gigabit Ethernet (1Gb/s) afin de permettre l'interopérabilité (c'est-à-dire la communication) entre réseau SAN et réseau Ethernet.

Ci-dessous, un commutateur de marque Gadzoox qui propose 16 prises Fibre Channel (2Gb/s par port) et une prise RJ45 (1Gb/s).



III.H.1.4 Les médias utilisés

Selon la configuration matérielle utilisée, différents types de câbles sont utilisés :

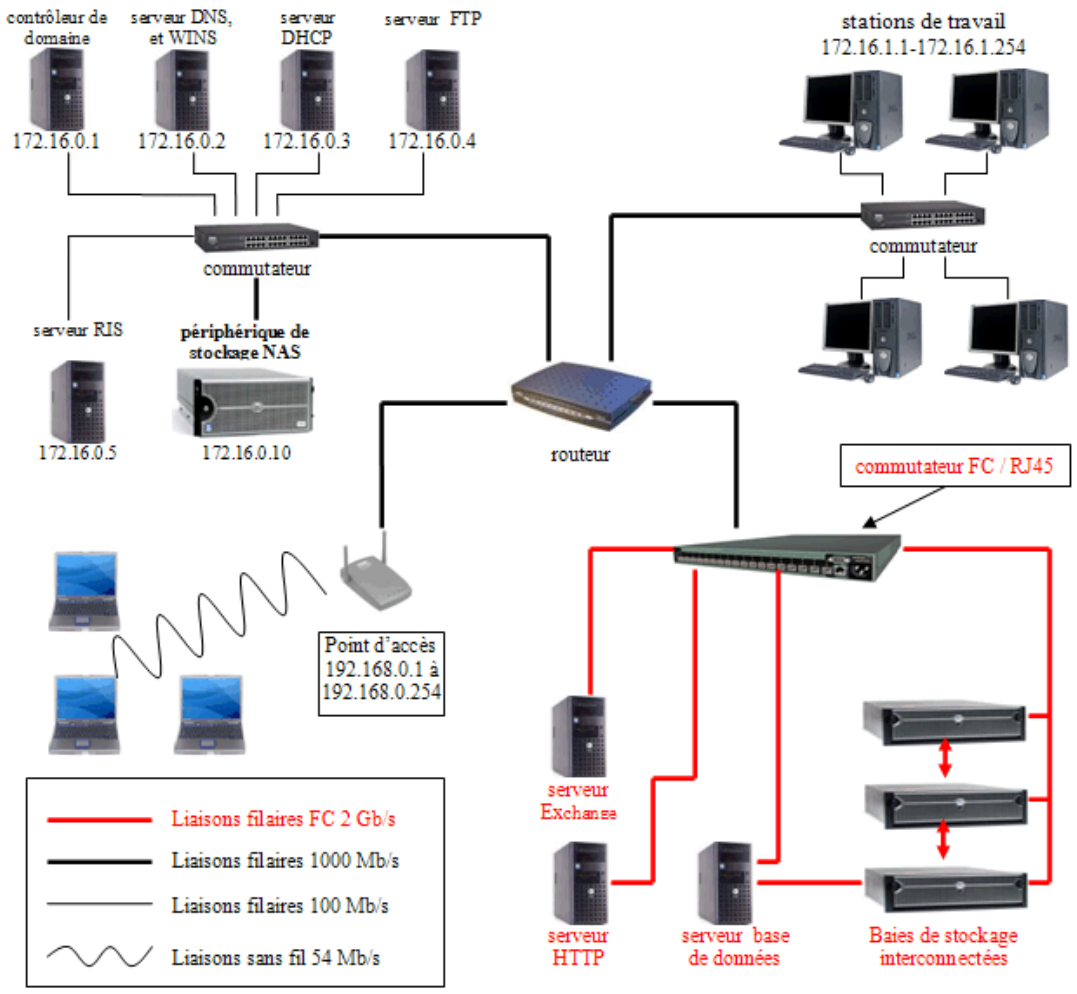
Câble SFP : (débit max : 2,5Gb/s)		Câble HSSDC : débit max : 2Gb/s)	
Câble DBm : (débit max : 2Gb/s)		Câble HSSDC2 : (débit max : 5Gb/s)	

III.H.1.5 Exemples de SAN

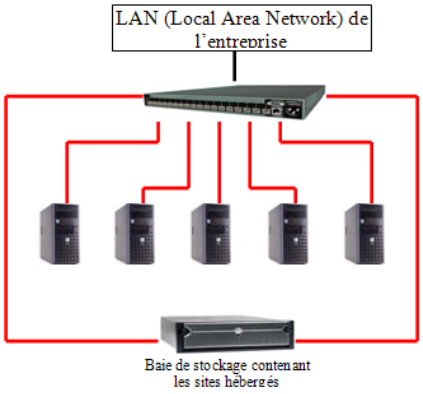
Voici comment un exemple d'intégration d'un réseau SAN au sain d'un réseau d'entreprise utilisant le protocole Ethernet. On dispose de trois serveurs. L'un de ces serveurs travaille sur une base de données stockée sur un module de stockage lui aussi dans le réseau SAN.

Un serveur HTTP et un serveur Exchange ont aussi accès à des données stockées sur le module de stockage. Le module de stockage est composé de trois baies de stockage interconnectées.

Enfin, un commutateur assure les connexions entre les différents composants du réseau SAN et permet aussi aux réseaux SAN et Ethernet de communiquer entre eux.



Une autre utilisation typique d'un réseau SAN est la mise en place d'un cluster. En effet, un cluster est un ensemble de machine vu par le réseau de l'entreprise comme étant une seule et même machine (vu du réseau un cluster possède un seul nom et une seule adresse IP).



Par exemple, on peut créer un réseau SAN contenant 5 serveurs exécutant IIS (Internet Information Service) pointant tous vers le même support de stockage (une baie de stockage par exemple).

Une fois le service de clustering MSCS installé et le Network Load Balancing (NLB) paramétré, on dispose d'un cluster qui répartit la charge réseau (les requêtes HTTP vers le cluster sont distribuées de façon équitables entre les cinq machines exécutant IIS) et qui gère la tolérance de panne (on peut perdre jusqu'à 4 machines sur 5 dans ce type de configuration).

Si vous souhaitez voir un exemple pratique de l'implémentation d'un réseau SAN, allez visiter le site du Centre Interuniversitaire de Calcul de Toulouse (http://www.cict.fr/). En effet, le CICT a mis en ligne la configuration du réseau SAN qu'il a mis en place en 2001. Pour les curieux, c'est à cette adresse http://www.cict.fr/sys/Serveurs/san.html.

III.H.2 Conclusion

Les SAN proposent de bien meilleures performances que les solutions de stockage NAS.

Cependant, les SAN nécessitent une infrastructure réseau complètement différente du réseau Ethernet classique utilisé dans le reste de l'entreprise.

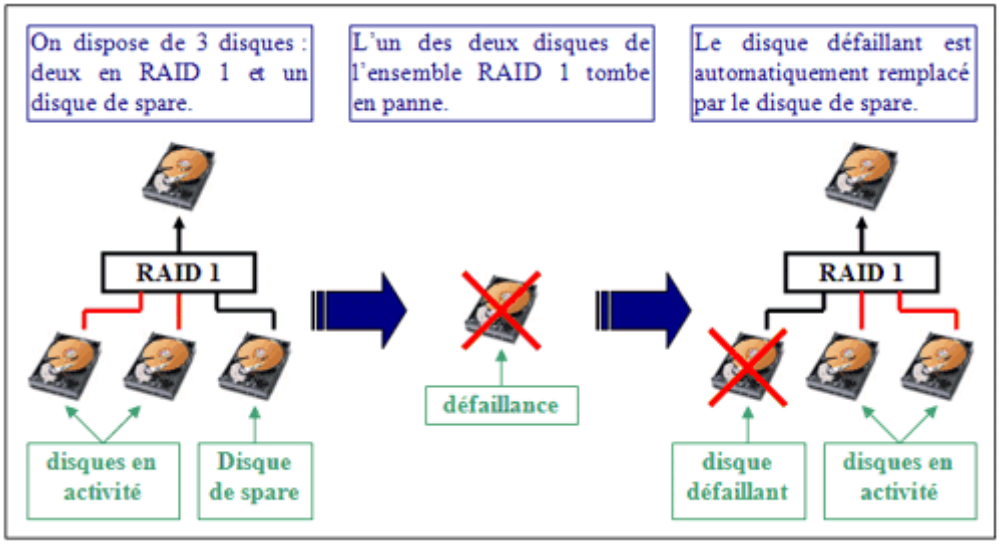
Cela apporte des inconvénients (il faut un gros investissement matériel pour mettre en place un SAN) et des avantages (meilleure fiabilité, meilleures performances, le réseaux SAN permet de décharger le réseau Ethernet).

L'utilisation d'un SAN dépend donc des besoins de l'entreprise.

III.I Le hotspare

Certaine cartes RAID et/ou périphériques NAS permettent de brancher un ou plusieurs disques de rechange (spare disks). Ces disques ne sont pas utilisés sauf en cas de panne de l'un des disques.

Voici un petit exemple :



Dans l'exemple précédent, on a mis en place un RAID 1 avec deux disques et on a aussi rajouté un disque de spare (bien entendu il faut un équipement qui supporte le hotspare).

Si l'un des disques de l'ensemble RAID un vient à tomber en panne, il sera immédiatement remplacé par le disque de spare.

Cela permet d'augmenter la tolérance de panne de l'ensemble RAID car si on utilise un disque de spare, on peut se permettre de perdre un disque supplémentaire sans perte de données.

C'est bien entendu coûteux en terme d'équipement sur des petits ensemble RAID. Par contre sur de gros ensembles (5 disques et plus) cela peut devenir très intéressant.

Par exemple, si l'on a une carte RAID adaptée on peut créer un RAID 50 avec seize disques durs (4 ensembles RAID 5 de quatre disques durs chacun montés en RAID 0) et rajouter un 17è disque de secours.

Si l'un des disques de l'ensemble RAID 5 tombe en panne alors le disque de rechange prend instantanément le relais ce qui permet de conserver la tolérance de panne. L'administrateur peut ensuite changer tranquillement le disque défectueux sachant que le système reste sécurisé.

Attention : Il ne faut pas confondre le hotspare avec le hotplug. Le hotplug (encore appelé hotswap) est une technologie qui permet de brancher à chaud diverses équipement. Par exemple, l'interface USB supporte le hotplug (on peut brancher/débrancher des périphériques USB même lorsque le système est en fonctionnement). Au niveau des périphériques de stockages, seuls deux interfaces supportent le branchement à chaud : l'interface SCSI et l'interface SATA.

Pour conclure, le hotspare permet d'augmenter la sécurité d'un ensemble RAID. Il est généralement supporté par les cartes RAID "haut de gamme" et par les NAS. Il est surtout utilisé sur de gros ensembles de disques.

III.J Conclusion

Le RAID matériel est une solution très intéressante pour améliorer la sécurité et/ou les performances des données stockées sur une machine ou sur le réseau. La solution la plus efficace mais aussi la plus chère reste l'utilisation de périphériques dédiés (NAS). Mais pour des raisons de coût, les entreprises mettent souvent en place le RAID via des cartes d'extensions.

Il faut alors faire un choix entre RAID SCSI (plus performant, plus fiable et plus cher) et RAID SATA. Il est préférable de faire une impasse sur le RAID PATA car c'est une technologie vieillissante.

On peut ausi améliorer la sécurité des gros ensemble RAID grâce à l'utilisation d'un ou plusieurs disques de spare. Enfin, on peut aussi améliorer la sécurité ou les performances en créant une combinaison entre plusieurs modes de RAID (exemple de combinaison : RAID 10) en utilisant du RAID matériel et du RAID logiciel.

L'avantage du RAID logiciel étant de permettre aux Entreprises et éventuellement aux particuliers qui ont un budget limité de mettre en place la technologie RAID.

IV Le RAID LOGICIEL AVEC WINDOWS SERVEUR 2003

IV.A Introduction

Depuis, la sortie de Windows NT 4.0, la technologie RAID est implémentée dans tous les systèmes d'exploitation de Microsoft destinés au monde professionnel. Ainsi, il est possible de créer des ensembles de disques utilisant la technologie RAID de manière 100% logicielle.

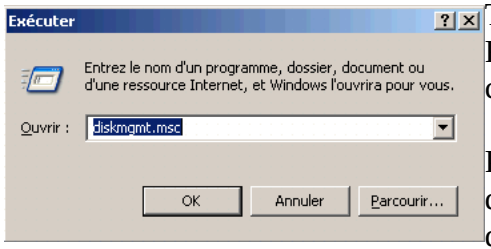
Cependant, il existe une disparité entre les versions serveurs et les versions clientes de Windows. En effet, les versions clientes (NT 4.0 workstation/2000pro/XP) ne supportent que les niveaux JBOD et RAID 0 alors que les versions serveurs (NT 4.0/2000/2003) supportent les niveaux JBOD et RAID 0 ainsi que les niveaux RAID 1 et RAID 5 qui apportent la tolérance de pannes.

Les deux utilisations les plus répandues du RAID logiciel sont :

- La création d'ensemble RAID bas prix
- La mise en place d'une combinaison entre RAID

Nous allons expliquer comment créer et dépanner un ensemble RAID sous Windows Server 2003.

IV.B Utilisation de la console de gestion des disques



Toutes les manipulations qu'il faut réaliser pour mettre en place du RAID sous Windows Server 2003, sont réalisées à l'aide de la console de gestion des disques (ou Disk Management).

Pour lancer cette console, il suffit de taper la commande « diskmgmt.msc » dans la boite de dialogue exécuter (menu démarrer/exécuter).

On peut aussi accéder à cette console par d'autres chemins :

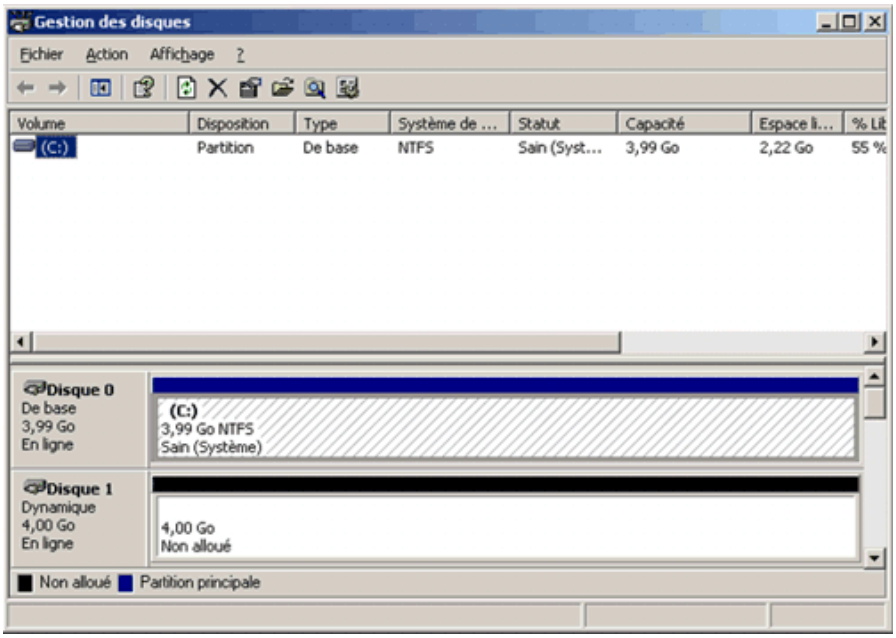
- Démarrer/paramètres/panneau de configuration/outils d'administration/gestion de l'ordinateur/stockage/gestion des disques
- Clic droit sur l'icône du poste de travail/gérer/stockage/gestion des disques

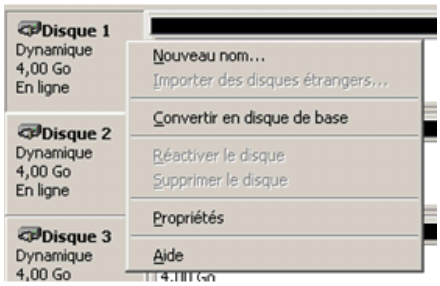
La console de gestion des disques permet de réaliser beaucoup d'actions sur les disques durs :

- activer/réactiver des disques : On doit activer un disque lorsque l'on vient de le rajouter dans l'ordinateur.
- formater une partition ou un volume avec un système de fichier donné (FAT, FAT32 ou bien NTFS)
- modifier la lettre de lecteur d'un disque (ex. : D:\ ) ou son chemin d'accès (dans le cas d'un point de montage).
- créer et supprimer des partitions
- créer et supprimer des volumes

Avec Windows Server 2003, on peut mettre en place la technologie RAID en créant des volumes. Pour être en mesure de créer des volumes, il faut commencer par convertir les disques dits « de base » en disque « dynamiques ». Lors de l'installation de Windows, le disque système est toujours un disque de base.

Voici une capture d'écran de la console de Gestion des disques où l'on observe un disque de base avec une partition de 4Go contenant le système d'exploitation. On remarque aussi un disque dynamique sur lequel on n'a pas encore défini de volumes.





Pour créer un volume, il suffit de faire un clic droit sur un disque dynamique contenant de l'espace disque non alloué.

Si l'on choisit « nouveau nom », alors l'assistant création de volume s'exécute.

On peut aussi choisir de reconvertir le disque dynamique en disque de base. Cependant, cette option apparaît uniquement si le disque ne contient aucun volume. Cela signifie que pour convertir un disque "dynamique" vers un disque "de base", il faudra sauvegarder les données de tous les volumes, puis supprimer tous les volumes du disque.

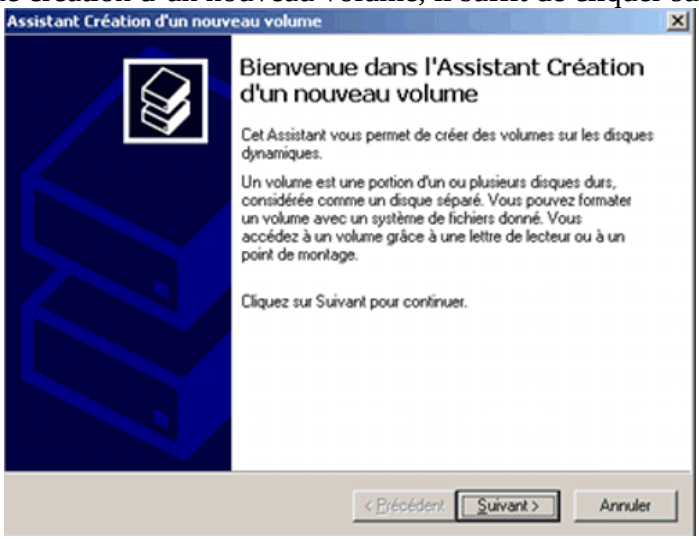
Voici les différents types de volumes proposés sous Windows Server 2003 :

- volume simple et volume fractionné : ils correspondent au RAID JBOD. En effet, un volume simple est une partition qui peut ensuite être étendue grâce à de l'espace disque contenu sur plusieurs disques et/ou partitions. Un volume fractionné est une partition qui additionne l'espace disque de plusieurs disques et/ou partitions.
- volume agrégé par bandes : ce type de volume correspond au RAID 0.
- volume en miroir : correspond au RAID 1 « mirroring ».
- volume RAID 5

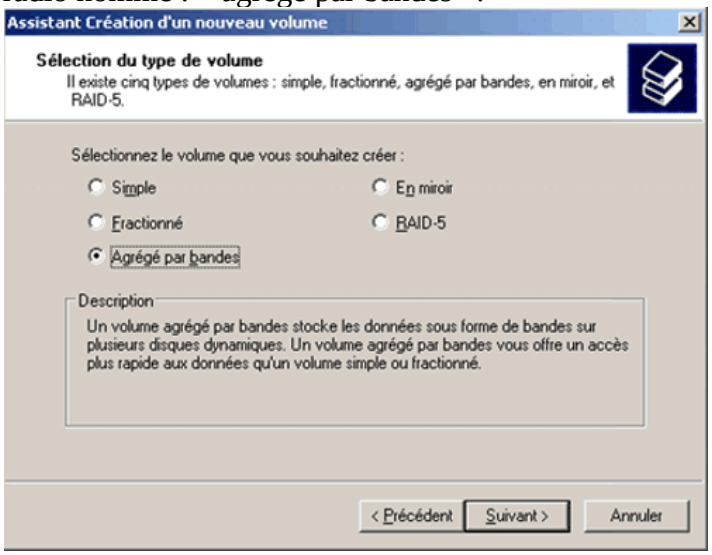
IV.C Mettre en place un volume agrégé par bandes (RAID 0)

Comme nous l'avons vu précédemment, il suffit de faire un clic droit sur un disque dynamique, puis de cliquer sur « nouveau nom » pour créer un volume.

Une fois dans l'assistant de création d'un nouveau volume, il suffit de cliquer sur le bouton suivant.

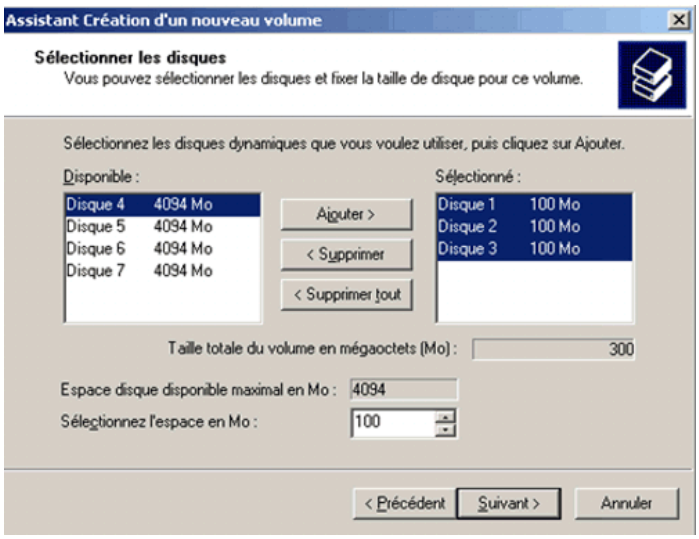


On sélectionne le bouton radio nommé : « agrégé par bandes ».

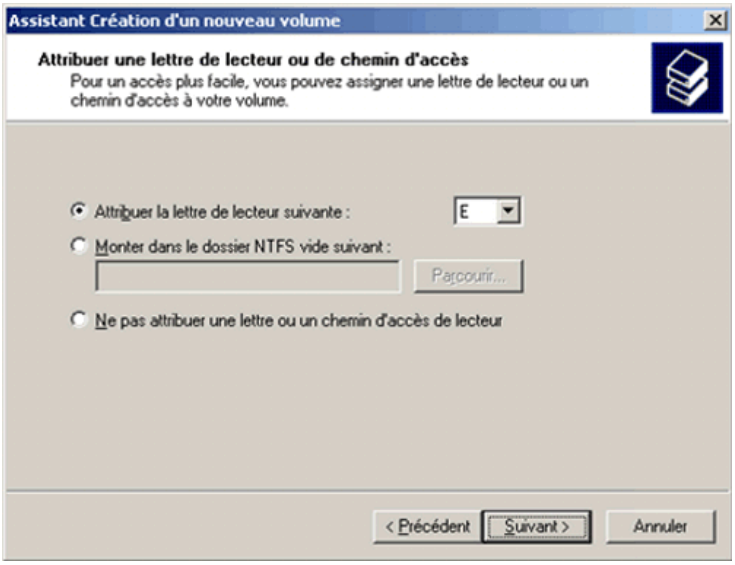


On doit ensuite sélectionner le nombre de disque qui vont composer le volume agrégé par bandes, puis la taille qui sera utilisée sur ces disques.

Dans notre exemple, on a réservé un espace de 100 Mo sur 3 disques d'une capacité totale de 4Go. On va donc créer un volume d'une taille de 300 Mo. L'espace restant sur chacun des disques peut être utilisé pour créer d'autres volumes.

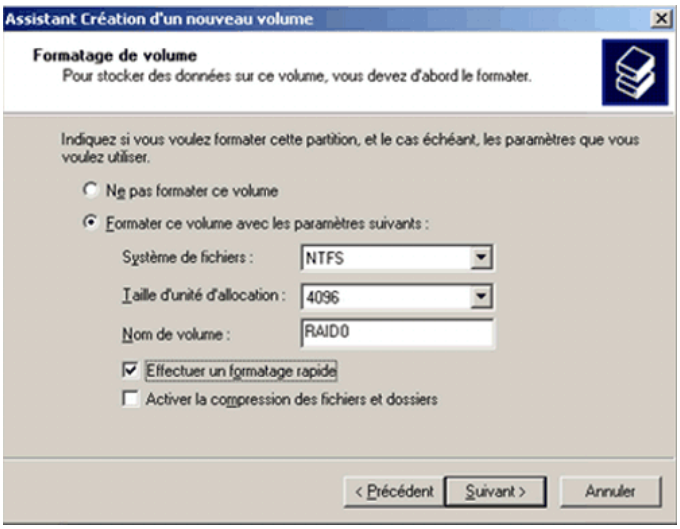


A l'étape suivante, on peut choisir la lettre de lecteur qui sera attribuée au volume ou bien choisir de créer un point de montage sur un dossier.



On doit ensuite formater le volume avec un système de fichiers particulier. Dans le cas d'un volume, Windows Server 2003 propose uniquement le système de fichiers NTFS. (par contre, on peut formater des partitions crées sur des « disques de bases » avec le système de fichiers FAT, FAT32 ou NTFS).

On doit ensuite choisir la taille des unités d'allocation (ou clusters). Cette option influera fortement sur les performances de votre serveur.

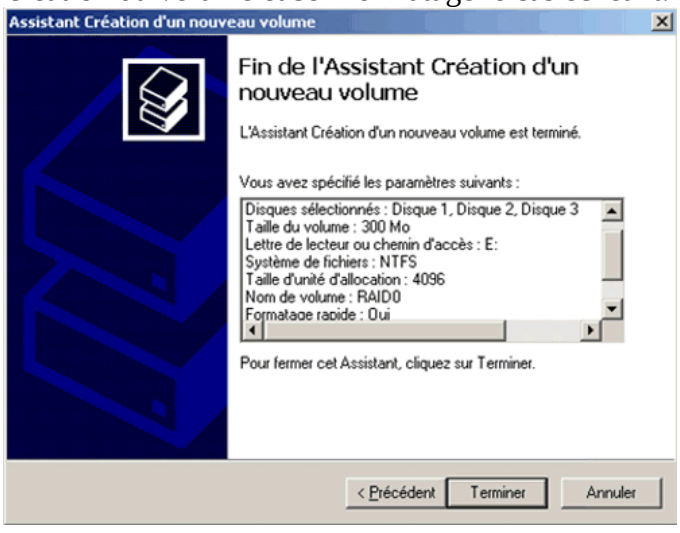


En effet, dans le cas d'un volume agrégé par bandes, la taille des unités d'allocation représente la taille occupée au minimum par un fichier, mais aussi la taille des segments qui seront répartis entre les disques de du volume. La taille des unités d'allocation est un multiple de la taille d'un secteur de disque dur (soit 512 octets). Elle est comprise entre 512 octets (1 secteur) et 65536 octets (soit 128 secteurs de 512 octets chacun). Pour le système de fichier NTFS la taille par défaut des unités d'allocation est de 4096 octets.

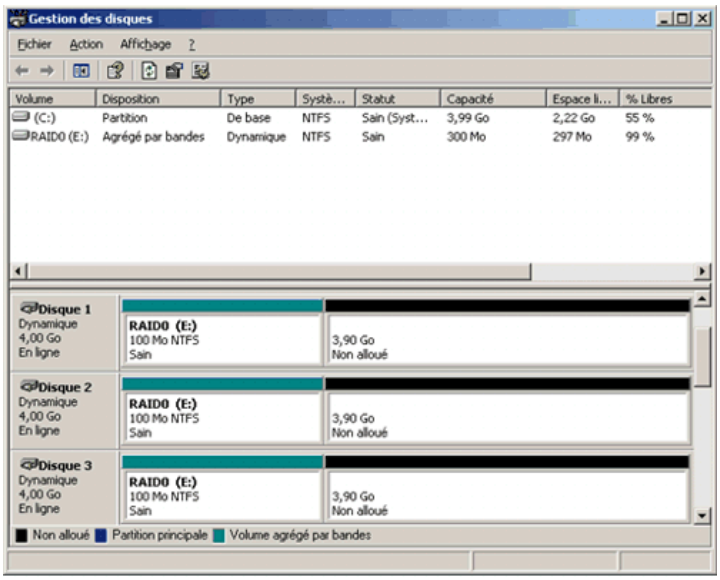
Ainsi, si le volume reçoit un grand nombre de requêtes pour des fichiers de petites tailles (si il héberge un serveur HTTP par exemple), il sera utile de choisir une petite taille pour les unités d'allocation (512 octets, 1024 octets ou 2048 octets). En revanche, dans le cas d'un serveur de fichiers qui contient par exemple des images de CDs ou des images RIS, une grande taille pour les unités d'allocation est recommandée (ex. : 16, 32 ou 64 Ko).

Les autres options que l'on peut sélectionner sur cette fenêtre sont l'activation de la compression des fichiers et des dossiers et la sélection du formatage rapide. On peut aussi donner un nom au volume. On peut noter que la compression des fichiers et des dossiers n'est possible que sur des volumes avec des unités d'allocation dont la taille est inférieure ou égale à 4096 octets.

Une dernière fenêtre récapitule les choix effectués. Si les paramètres semblent bons, il suffit de cliquer sur terminer pour lancer la création du volume et son formatage le cas échéant.



On peut ensuite visualiser le volume que l'on vient de créer dans la console de gestion des disques.



IV.D Mettre en place un volume en miroir (RAID 1)

Pour rappel, lorsque l'on met en place un volume miroir, les informations sont dupliquées entre tous les disques du volume miroir.

Le programme utilisé par Windows pour écrire simultanément les informations sur différents disques physiques est « ftdisk.sys ».

Lorsque l'on met en place un volume miroir sur une partition contenant le système d'exploitation, Microsoft recommande de suivre la procédure suivante :

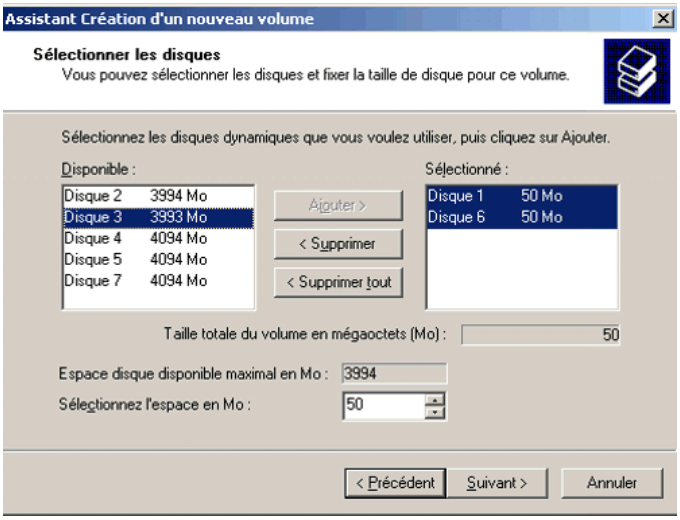
- créer le volume miroir
- modifier le fichier boot.ini pour qu'il pointe vers le deuxième disque physique
- créer une disquette d'amorçage contenant le fichier boot.ini modifié

Cette procédure permet de démarrer sur le deuxième disque via la disquette d'amorçage en cas de panne de l'un des disques.

Nous ne développerons pas inutilement la configuration du volume en miroir car elle très proche dans son approche que celle de l'agrégat par bandes.

Il faut tout d'abord créer un nouveau volume, puis passer la page de présentation de l'assistant en cliquant sur le bouton « suivant ».

Ensuite, on sélectionne volume en miroir et la fenêtre suivante s'affiche :



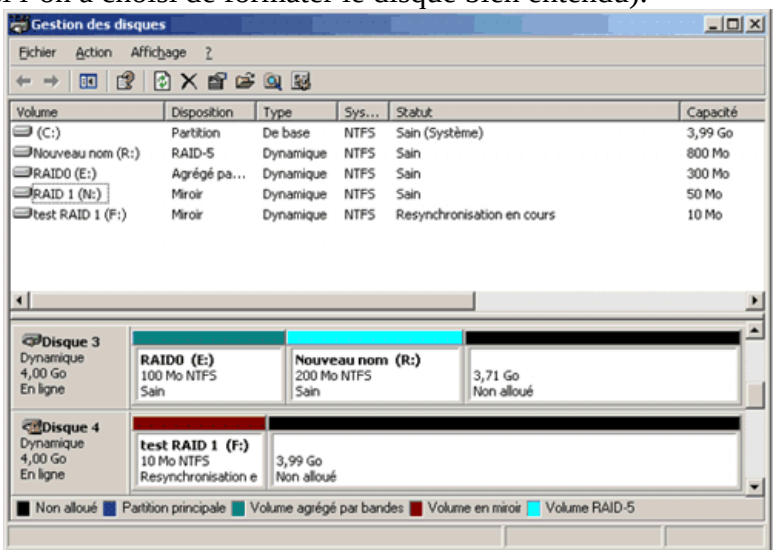
A l'instar du volume agrégé par bandes, on doit sélectionner le nombre de disques.

Cependant, pour le RAID 1 (ou volume en miroir), Windows Server 2003 limite le nombre de disques à 2. Il faut donc choisir deux disques et la taille qu'occupera le volume sur chacun des disques.

Dans cet exemple, on crée un volume de 50Mo ce qui signifie que 50 Mo seront utilisés sur chacun des disques pour créer le miroir.

Les dernières étapes de la configuration sont identiques à celle du RAID 0 (nom du volume, taille des clusters, choix de la lettre de lecteur, …).

Au moment où l'on termine l'assistant une synchronisation a lieu entre les deux disques, puis le formatage s'effectue (si l'on a choisi de formater le disque bien entendu).

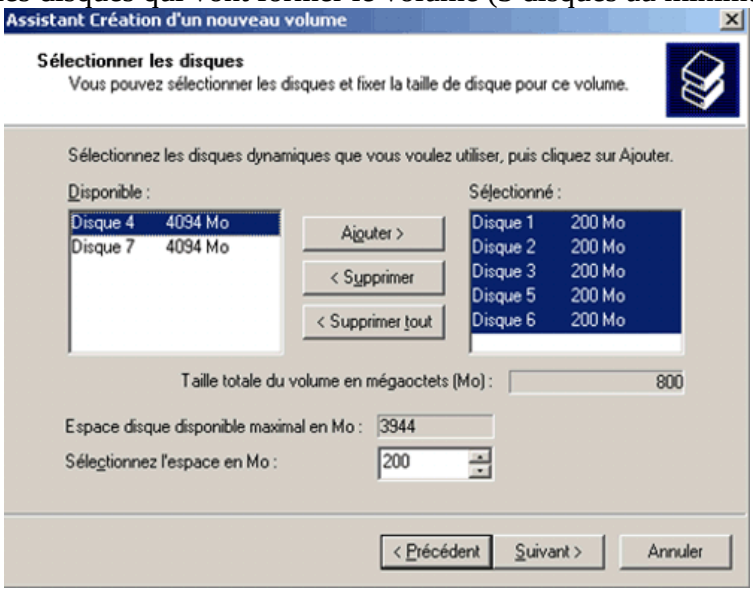


Une fois le formatage terminé, le volume en miroir est directement opérationnel. En cas de panne de l'un des disques, le système d'exploitation utilise le disque restant.

IV.E Mettre en place un volume RAID 5

Pour mettre en place un volume RAID 5, il faut lancer l'assistant « création d'un nouveau volume » et sélectionner volume « RAID-5 ».

On doit ensuite choisir les disques qui vont former le volume (3 disques au minimum) :



Dans notre exemple, on réserve un espace de 200 Mo sur chacun cinq disques afin de créer le volume RAID 5. On obtiendra donc un volume d'une taille de 4 * 200 Mo = 800 Mo.

En effet, le stockage des informations de parité utilise l'espace d'un disque entier soit 200Mo.

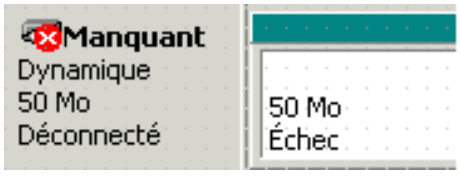
Bien entendu si l'un des disques contient de l'espace non alloué, on peut l'utiliser pour créer un autre volume.

Cela fonctionne très bien, mais ça n'est par recommandé au niveau des performances.

Le reste des options de paramétrages est strictement identiques aux autres volumes. Nous ne les détaillons donc pas.

IV.F Dépannage d'un ensemble RAID

Si un disque dur connaît une défaillance alors le statut du volume auquel il appartient est soit « manquant » soit « en erreurs ».



Ci-contre, l'un des disque dur d'un volume agrégé par bande est mal branché (suite à une manipulation à l'intérieur du boîtier par exemple), il porte l'attribut : « manquant ».

La première chose à faire lorsqu'un des volumes porte l'attribut « manquant » ou en erreur est de tenter de réactiver le disque.

Si cela ne fonctionne pas, vérifiez que les disques dur composant le volume défaillant sont bien branchés, puis re-tentez de réactivez le disque.

Enfin si l'un des disques durs se révèle physiquement endommagé, il faut suivre une procédure différente selon les cas :

- dans le cas d'un volume fractionné ou d'un volume agrégé par bande, il faut supprimer le volume existant, rajouter un nouveau disque, signer le nouveau disque, convertir le nouveau disque en disque dynamique, re-créer le volume, restaurer les données présentes sur l'ancien volume à partir d'une sauvegarde.
- Dans le cas d'un volume miroir, il faut supprimer le disque miroir marqué portant l'indication manquant, rajouter un nouveau disque, signer le nouveau disque, convertir le nouveau disque en disque dynamique et enfin ajouter le nouveau disque en tant que disque miroir pour le premier disque. Le système va automatiquement régénérer le miroir.
- Dans le cas d'un volume RAID 5, il faut rajouter un disque neuf, signer le nouveau disque, convertir le nouveau disque en disque dynamique et réparer le volume RAID 5. On peut ensuite supprimer le disque dur marqué comme défaillant.

IV.G Conclusion

La console Gestion des disques permet de créer facilement des ensembles utilisant la technologie RAID. C'est aussi un outil de monitoring intéressant (état des disques, pourcentage d'espace disque restant, …).

Un autre avantage du RAID logiciel est que sa mise en place est indépendante de la technologie utilisée. On peut par exemple mettre en place un volume RAID 5 avec un disque dur IDE branché sur le contrôleur intégré à la carte mère, un disque dur SCSI branché sur une carte d'extension et un disque dur externe branché sur un port IEEE1394 (ou firewire) ce qui est impossible avec du RAID matériel.

Le RAID logiciel est souvent utilisé pour créer des combinaisons entre différents niveaux de RAID (exemple RAID 10). Cependant le RAID logiciel est moins performant que le RAID matériel ce qui limite son utilisation aux entreprises n'ayant pas les moyens d'investir dans une solution matérielle.

V La Technologie RAID Mesures de Performances

V.A Introduction

Il est très délicat de parler de performances entre les différentes sortes d'ensembles RAID. Le sujet des performances entre le RAID logiciel et le RAID matériel, par exemple, alimente des débats sans fin.

En effet, les performances sont influencées par un trop grand nombre de paramètres (RAID logiciel ou matériel, niveau de RAID utilisé, interface SCSI ou IDE, marque et modèle des disques durs, système de fichiers utilisé, tailles des segments et/ou des clusters, …) pour que l'on puisse comparer de façon objectives et équitables deux ensembles RAID.

Nous allons tout de même essayer de mettre en valeur les performances que l'on peut obtenir selon :

- le type de RAID (matériel ou logiciel)
- le niveau de RAID mis en place
- le matériel utilisé

Pour cela, je vais m'appuyer sur les résultats de tests que j'ai effectué moi-même ou bien qui ont été publiés sur la toile.

V.B Configuration de test

Les mesures de performances suivantes ont été réalisées sur la même machine afin que les tests ne soient pas influencés par la configuration matérielle.

Voici le détail de la configuration utilisée :

Boîtier	Antec Sonata
Alimentation	Antec True Power de 350 Watts
Carte mère	Abit NF7-S V2.0 (avec un chipset Nvidia Nforce 2 ultra 400)
Processeur	AMD Athlon XP 1800+ (1533MHz) sur cadencé à 2300 MHz
Mémoire vive	2 * 256 Mo DDR PC3200 (modules de marque Samsung) utilisant le dual band
Carte RAID	carte Medley (chipset Silicon image 0680) au format PCI 32 bits et avec une interface ATA 133. Elle gère les niveau 0, 1, 10 et JBOD.
Disques durs	o disque système : IBM 80Go avec 2Mo de cache (série 120GXP) o disques testés (cad branchés sur la carte RAID) : 2 * Western Digital 80Go special edition avec 8Mo de cache
Système d'exploitation	Windows Server 2003 Entreprise Edition
Pilote du chipset	nForce Driver v3.13 (pilote IDE activé)
Pilote de la carte RAID	v1.0.1.7 (31/05/2002)

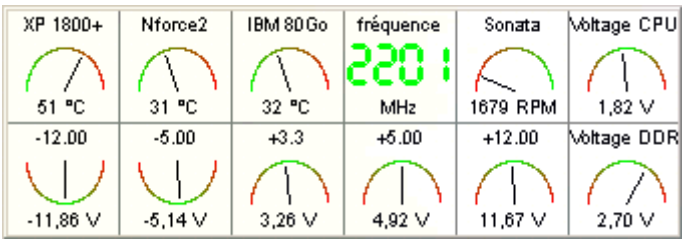
Lors de tous les tests, le disque dur système était raccordé au contrôleur intégré à la carte mère et les deux disques durs testés étaient branchés en maître sur la carte RAID chacun étant seul sur son canal.

Lors des tests en RAID logiciel le RAID était désactivé sur la carte RAID qui se comporte alors comme un simple contrôleur. Afin que l'occupation du temps processeur n'interfèrent pas sur les résultats, aucun programme n'a été exécuté pendant les tests (hormis les processus lancé au démarrage de Windows et le benchmark bien sur).

V.C Méthodologie des tests

Les tests ont été réalisés à l'aide de bancs d'essai ou benchmark (programmes qui mesures les performances des périphériques matériels) reconnus comme PCMark 2k2 et SiSoft Sandra 2004. Au niveau des disques durs, ces deux benchmarks testent différents paramètres (débit en lecture, débit en écriture, vitesse du cache, temps d'accès moyen, …) et renvoient un indice de points qui permet de comparer les performances des différents disques entres eux.

Pour vérifier les résultats obtenus avec les benchmarks précités, j'ai réalisé un petit test pratique. Il consiste simplement en la copie d'un fichier situé sur le disque système (IBM 80 Go) vers l'ensemble RAID composé des deux disques Western Digital afin de mesurer le débit en écriture. Le pourcentage d'occupation CPU pendant la copie a été mesuré précisément grâce au logiciel Motherboard Monitor (MBM).



le panneau de contrôle de MBM

Les tests ont d'abord réalisés en RAID 0, puis en RAID 1. Les performances d'un disque Western Digital seul ont été rajoutées afin de bien visualiser l'impact de l'utilisation de la technologie RAID.

V.D RAID matériel contre RAID logiciel

V.D.1 RAID 0

Voici les résultats obtenus en utilisant le niveau RAID 0 :

Comparatif des performances RAID 0 logiciel / RAID 0 matériel			
	Disque seul	RAID logiciel	RAID matériel
Tests synthétiques			
PCMark 2k2 (indice de points)	1276	2073	1957
SiSoft Sandra 2k4 (débit en Ko/s)	32 373	58 021	62 291
Tests pratiques			
temps pour la copie d'un fichier de 701 Mo	36s	17s	18s
débit pour l'écriture d'un fichier de 701 Mo	19,47 Mo/s	41,23 Mo/s	38,94Mo/s
pourcentage d'occupation processeur	11%	37%	12%

On remarque que les performances entre RAID 0 logiciel et RAID 0 matériel sont très proches avec un léger avantage pour le RAID 0 logiciel. Cela peu paraître étrange car il est bien connu que le RAID matériel est bien plus performant que le RAID logiciel. Ces résultats peuvent s'expliquer simplement. En effet en RAID logiciel, les débits et les temps d'accès profitent de la puissance du CPU (qui est dans notre cas bien plus performant que la puce installée dans le contrôleur de la carte RAID) et des performances de la mémoire vive (qui est ici plus rapide et présente en plus grande quantité que la mémoire cache intégrée à la carte RAID).

Cependant le revers de ces bonnes performances est une occupation des ressources de l'unité centrale bien plus importante qu'en RAID matériel (occupation processeur de 12% en matériel contre 37% en logiciel !!!). Si le RAID logiciel reste intéressant pour les petites configurations (avec peu de disques), il est indéniable que pour mettre en place de « gros » RAID (par exemple du RAID 5 avec 6 disques SCSI 15 000 tours/min) son utilisation est exclue. En effet, la majorité des serveurs doivent réserver au maximum leurs ressources mémoires et processeurs pour les tâches auxquelles ils sont dédiés (contrôleurs de domaine, serveurs de fichiers, serveurs HTTP, serveurs FTP, …).

Ce petit tableau permet aussi de d'observer l'exceptionnelle augmentation des performances induite par le RAID 0. La théorie (deux disques en RAID 0 doublent les performances par rapport à un disque seul) est donc bien vérifiée en pratique.

V.D.2 RAID 1

Voici les résultats obtenus en utilisant le niveau RAID 1 :

Comparatif des performances RAID 1 logiciel / RAID 1 matériel			
	Disque seul	RAID logiciel	RAID matériel
Tests synthétiques			
PCMark 2k2 (indice de points)	1276	1270	1318
SiSoft Sandra 2k4 (débit en Ko/s)	32 373	32 148	34 790

Cette fois encore les performances sont très proches. Les performances du RAID logiciel sont quasiment identiques à celle d'un disque seul alors que le RAID matériel se démarque en offrant des performances légèrement supérieures.

V.E L'influence du matériel sur les performances

Les performances dépendent fortement du matériel utilisé (carte mère, processeur, mémoire, disques durs, carte RAID,…). Cela est encore plus flagrant dans le cas d'un RAID matériel notamment au niveau du choix de la carte RAID.

En effet, selon le niveau de RAID voulu et le nombre de disques utilisés, on s'orientera vers une carte RAID au format PCI 32bits, PCI 64bits ou bien encore PCI-X. Le choix du format de la carte RAID est crucial pour les performances (il faut à tout prix éviter que le format devienne un goulot d'étranglement pour les performances).

Afin de mettre en valeur cette assertion, voici une série de tests réalisés avec 6 disques durs (identiques) très performants.

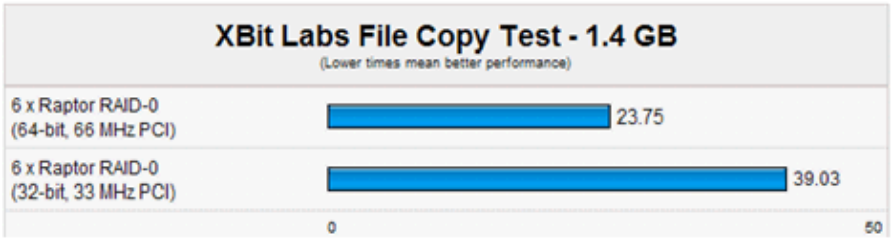
Ci-dessous, voici un petit résumé des caractéristiques de ces disques (données obtenues sur le site du constructeur).

Marque	Western Digital
Référence commerciale	Raptor
Modèle	WD740GD
Capacité (formatée)	74,356 Go
Taille du cache	8Mo
Vitesse de rotation	10 000 tr/min
Interface	Serial ATA

Les tests ont été réalisés avec la carte MegaRAID-6 de la marque LSI. Cette carte utilise le format PCI 64bits (qui offre une compatibilité descendante avec le format PCI 32bits).

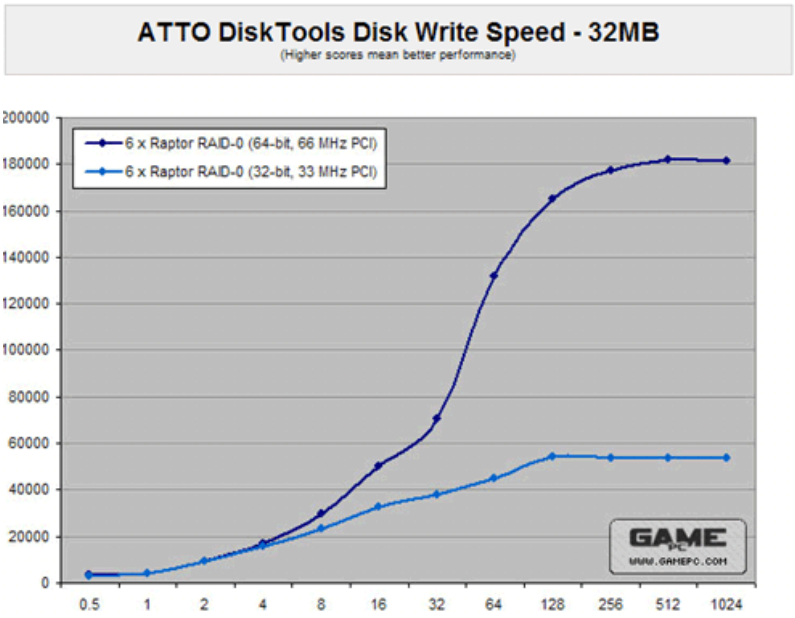
Elle gère les niveaux de RAID 0, 1, 5, 10 et 50 et possède huit canaux Serial ATA. Toutes les mesures de performances suivantes sont tirées de l'excellent article du site anglophone http://www.gamepc.com/labs/view_content.asp?id=raptor74&page=1 que je vous invite à consulter.

Le test ci-dessous compare les performances obtenues avec les six disques en mode RAID 0 pour l'écriture d'un fichier de 1400 Mo. La carte RAID est d'abord branché sur un port PCI 64bits, puis sur un port PCI 32bits.



On remarque que la carte offre de bien meilleures performances une fois branchées sur un port PCI 64bits. Cela est dû à la bande passante trop faible du port PCI 32bits (133Mo/s contre 266Mo/s).

Pour vérifier les résultats obtenus précédemment, les testeurs du site gamepc.com ont utilisés le benchmark ATTO qui donne des résultats plus précis.

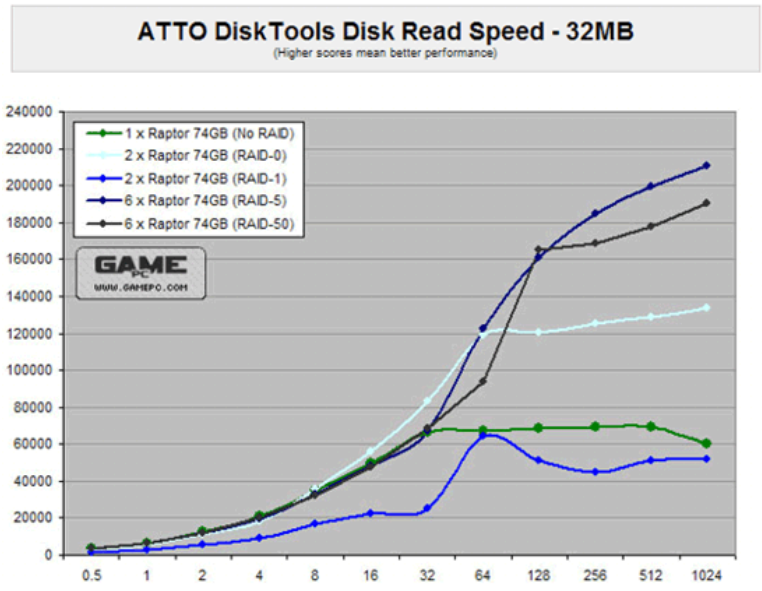


Il est flagrant que la bande passante du port PCI 32bits (133Mo/s) est un goulot d'étranglement pour les performances. Étant donné les importants débits en écriture obtenus avec cette configuration (6 disques durs Western Digital Raptor) il n'aurait pas été superflu de réaliser des tests complémentaires avec une carte RAID au format PCI-X (qui offre une bande passante de 1,06 Go/s).

Ces résultats démontrent qu'il est indispensable de choisir des composants matériels adaptés lorsque l'on souhaite monter un système performant. Il faut donc être vigilent lors de l'achat de la carte mère, de la carte RAID, des disques durs et éventuellement de la carte réseau (en effet, une carte réseau 100Mb/s voire même 1Gb/s peuvent s'avérer nettement insuffisante selon la configuration du RAID crée).

V.F L'influence du niveau de RAID sur les performances

Voici deux graphiques toujours réalisés par le site gamepc.com qui soulignent les performances en lecture et en écriture selon le niveau de RAID choisi.

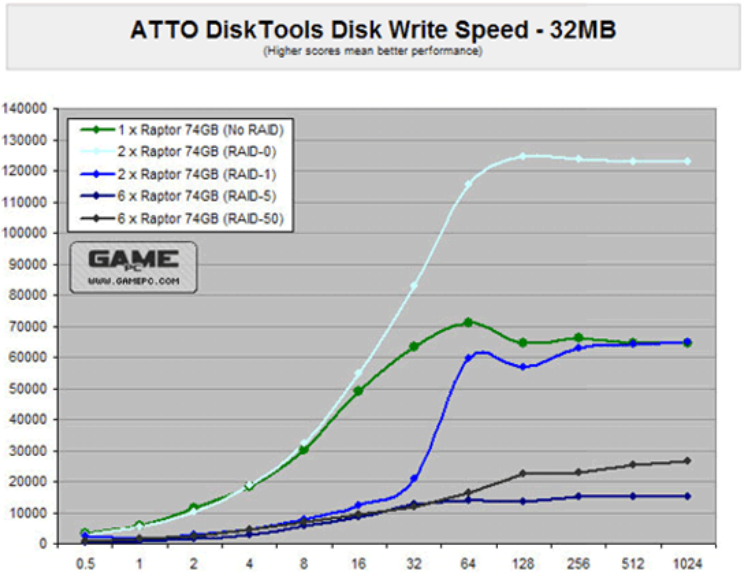


Sur ce graphique, on remarque que les performances en lecture de deux disques durs en RAID 1 sont inférieures à celles d'un disque seul alors que deux disques en RAID 0 doublent les performances en lecture d'un disque seul.

Le débit en lecture obtenu avec 6 disques durs en RAID 50 est environ trois fois plus important que celui d'un disque seul. Cela est intéressant car l'ensemble RAID 50 peut perdre jusqu'à deux disques.

Les performances en lecture de 6 disques en RAID 5 sont identiques à celles des 6 disques en RAID 50 alors qu'elles devraient être bien supérieures. Cela est probablement dû à la puissance trop faible de la carte RAID ou bien à cause du format PCI 64bits qui offre une bande passante de 266Mo/s alors que les débits mesurés atteignent jusqu'à 220 Mo/s.

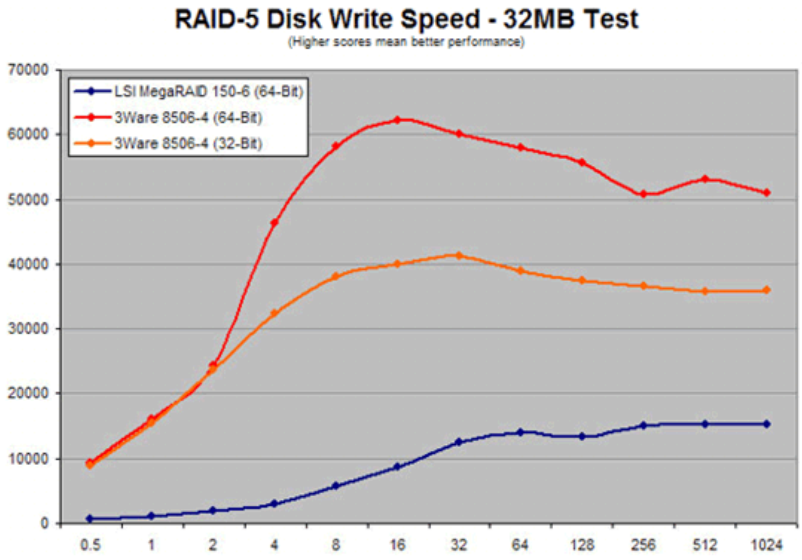
Voici maintenant un petit graphique qui montre les performances en écriture :



Le débit obtenu avec deux disques en RAID 1 est identique à celui obtenu avec un disque seul. De plus le débit en écriture obtenu avec deux disques en RAID 0 est le double de celui obtenu avec un seul disque. Ces résultats vérifient bien ceux obtenus précédemment.

Par contre les débits en écriture obtenus avec les niveaux RAID 5 et RAID 50 sont très faibles par rapport à un disque seul. Cela est principalement dû à la carte RAID qui est trop peu performante et qui manque de mémoire cache.

En effet, un autre test du site gamepc.com démontre que les performances dépendent énormément de la carte RAID utilisée :



Ce test a été réalisé dans les mêmes conditions que le test précédent c'est-à-dire avec 6 disques durs Western Digital Raptor.

Les résultats ci-dessus sont obtenus en créant un niveau RAID 5. On remarque qu'avec la carte RAID de marque 3Ware le débit en écriture est jusqu'à 3,5 fois supérieur à celui obtenu avec la carte RAID de marque LSI.

V.G Conclusion

Au niveau des performances, le meilleur choix reste le RAID 0. Cependant, ce niveau de RAID n'implémente pas la tolérance de pannes, il est donc à proscrire dans la majorité des cas. De son côté, le RAID 1 permet d'implémenter le plus haut niveau de sécurité mais sans amélioration de performances. Il faudra donc se rabattre sur le RAID 5 qui offre d'excellents débits en lecture mais pas en écriture.

Une autre alternative reste d'utiliser des niveaux de RAID combinés comme le RAID 50. On peut mettre en place très facilement le RAID 50 en utilisant le RAID logiciel intégré à Windows XP, 2000 et 2003. La meilleure solution au niveau des performances restant bien entendu le RAID 10.

Le RAID logiciel reste donc vraiment intéressant au niveau des performances dans les cas suivants :

- Mise en place d'un niveau RAID0 ou RAID 1 avec peu de disques (éviter le RAID 5 logiciel si possible car il utilise trop de ressources notamment en cas de panne de l'un des disques). Cette alternative permet d'implémenter la tolérance de panne sur un serveur lorsque l'on a peut de moyen.
- Mise en place d'un niveau combiné composé d'un RAID matériel implémentant la tolérance de pannes (1, 3 ou 5) et d'un RAID 0 logiciel. On parle donc des niveaux combinés 10, 30 et 50.

Pour conclure, la configuration d'un ensemble RAID doit être adaptée aux besoins de l'entreprise en matière de performances mais surtout de sécurité. Avant de mettre en place un ensemble de disques utilisant la technologie RAID, il faut donc commencer par une phase d'investigation, puis choisir le type de RAID (logiciel/matériel), le niveau (simple ou combiné), les disques (interface, nombre, ...) en fonction de votre budget.

Conclusion générale

La technologie RAID est de plus en plus utilisée au sain des entreprises comme chez les particuliers. Elle permet de combiner une amélioration des performances à la tolérance de pannes.

Cependant, pour mettre en place un ensemble de disques durs utilisant le RAID, il faut être vigilant à différents niveaux :

- choix du matériel : il doit être adapté afin de fournir des performances maximales.
- choix du niveau de RAID : selon les exigences du cahier des charges, il va privilégier la performances ou bien la sécurité. On peut mettre en place un niveau simple ou combiné.
- choix de la taille des segments : lorsqu'il y a lieu (RAID 0, 4, 5, 6 et toutes les combinaisons qui les utilisent), cela permet de maximiser les performances.

Si la technologie RAID protège les données si un ou plusieurs disques durs et/ou contrôleur tombe en panne, elle a aussi ses limites !!!

En effet, elle ne met pas à l'abri :

- des coupures de courant et autres surtensions. Seule l'utilisation d'un onduleur et d'une alimentation de qualité (ex. : une alimentation Active PFC) peuvent parer à ces éventualités.
- d'une erreur d'un des utilisateurs. Pour cela vous devez restreindre l'accès aux ressources stockées sur votre serveur et/ou réseau aux utilisateurs qui en ont besoin et limiter les autorisation d'accès au minium (ne pas permettre à n'importe quel utilisateur de modifier une base de donnée ou de supprimer des fichiers à tout va).
- d'une catastrophe naturelle ou bien criminelle (incendie, inondation, vol, …). Seules des sauvegardes sur support amovible régulières sont efficaces. Précaution élémentaire pour certains, un peu moins pour d'autres : il faut éviter de stocker les backup au même endroit que les données originelles.

La vague de popularité connue par la technologie RAID va grandissante. Cela est notamment dû aux constructeurs de chipsets qui intègrent la gestion du RAID en natif dans leur architecture. Étant donné les nombreux débats sur le thème de la sécurité des données on peut penser que cette technologie a encore de beaux jours devant elle.