

Les disques durs



Cours de périphérique de Mme Buseyne
Les disques durs

Sommaire

Sommaire	3
Définition :	4
Historique :	4
Géométrie	8
Performances	9
Capacité de stockage	12
Mécanique	12
Électronique	16
Types d'interfaces des disques durs.....	17
Formats.....	31
Autres caractéristiques	31
Fabricants	32
Émulation	32
Le RAID.....	33
Le formatage	39
Le partitionnement	41
Les systèmes de fichiers.....	43
Ce que l'on peut trouver actuellement	45
Tables de matières	47
Bibliographie :.....	49

Définition :

Le disque dur est un périphérique de stockage magnétique. Il a remplacé efficacement les tambours et les bandes, seulement utilisées de nos jours pour l'archivage et la sauvegarde. Inventés dans les années 1950 par IBM, leur capacité augmente très rapidement tandis que leur encombrement se réduit.

Historique :

Les ingénieurs d'IBM n'étaient pas satisfaits des systèmes de stockage sur tambours magnétiques : l'efficacité volumétrique était très faible, les tambours occupaient beaucoup d'espace pour peu de capacité.

En 1953, un ingénieur récemment embauché eut l'idée de superposer des plateaux le long d'un axe et d'y adjoindre une tête de lecture/écriture mobile, située sur un axe parallèle à celui des plateaux. Cette tête venait s'insérer entre les plateaux pour lire les informations, mais devait se retirer complètement pour passer d'un plateau à un autre.

Un prototype fut construit avec une vitesse de rotation d'environ 1000 tours/minute. À cette vitesse il était compliqué de maintenir les têtes au-dessus de la surface des plateaux.

En juin 1954 J. J. Hagopian, ingénieur IBM, a l'idée de faire « voler » les têtes de lecture/écriture au-dessus de la surface des plateaux, sur un coussin d'air. Il propose le design de la forme de ces têtes. En septembre 1954 il dessine l'équivalent des disques durs actuels : des plateaux superposés et un axe sur lequel sont fixées les têtes de lecture/écriture

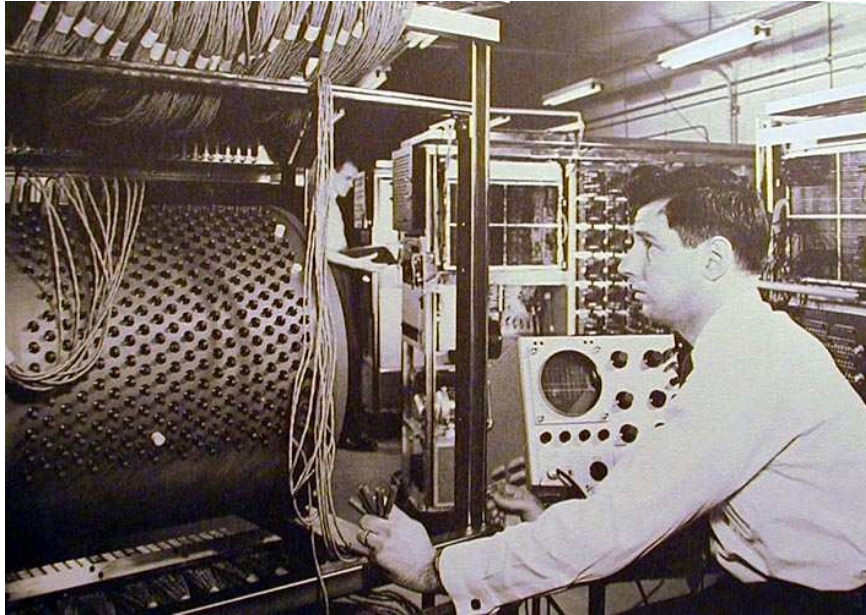
En 1956 le premier système de ce type a été dévoilé au public par IBM, il fut baptisé **RAMAC** (Random Access Method of Accounting and Control), modèle 305, et la production commerciale commença en juin 1957. Jusqu'à 1961, plus d'un millier d'unités furent vendues. Son prix : 10 000 dollars (de l'époque) par mégaoctet.

Le RAMAC 305 était constitué de 50 disques de 24 pouces de diamètre, 2 têtes de lecture/écriture qui pouvaient se déplacer d'un plateau à un autre en moins d'une seconde. La capacité totale était de 5 millions de caractères.



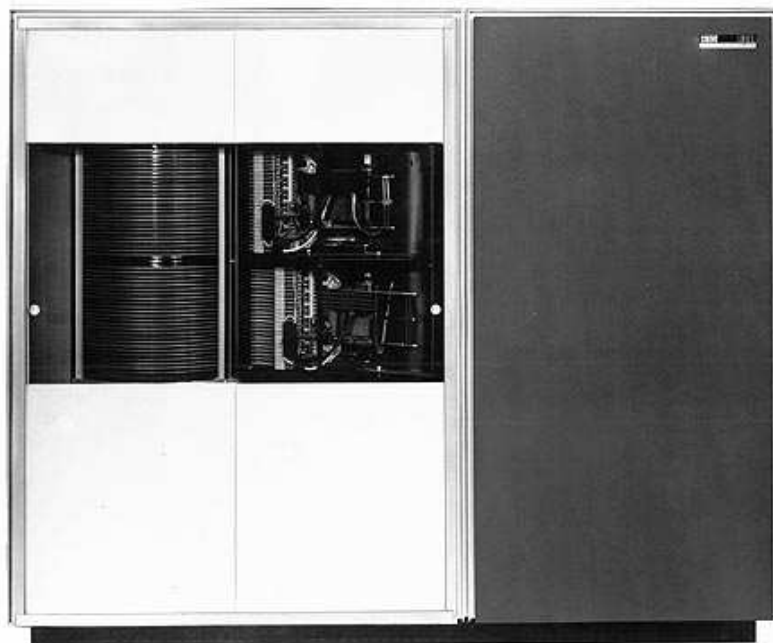
Le RAMAC 305

À noter que le RAMAC avait déjà un concurrent : le **Univac File Computer**, composé de 10 tambours magnétiques chacun d'une capacité de 180 000 caractères. Malgré une vitesse supérieure, c'est le RAMAC, qui pouvait stocker trois fois plus d'informations, qui avait le rapport coût/performance le plus intéressant pour le plus grand nombre d'applications.



L'Univac File Computer

En 1961 un nouveau produit commercial voit le jour sous la dénomination « **IBM 1301 Disk Storage** ». Il pouvait contenir 25Mo de données et jusqu'à 10 appareils pouvaient être reliés entre eux.



Fin 1969 trois ingénieurs réfléchissent à ce qui pourrait être pour eux le système disque idéal. Ils tombent d'accord sur un modèle composé de deux disques de 30 Mo chacun, l'un amovible, l'autre fixe. « 30 — 30 » donc, qui est aussi un modèle de carabine Winchester. Le nom est resté, et encore aujourd'hui un disque **Winchester désigne** un disque dur non amovible (soit pratiquement tous les disques produits aujourd'hui).



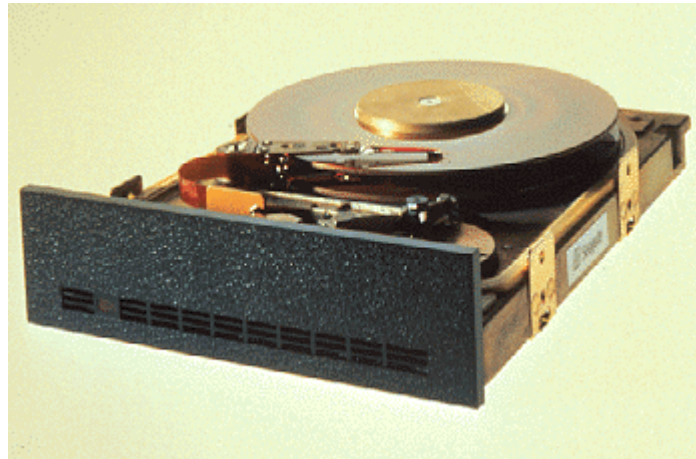
IBM 3340 disk unit (Le 1^{er} Winchester)

Présenté en 1979 par Seagate Technologies le **ST-506** fut le premier disque dur prévu pour un usage spécifique sur des micro-ordinateurs. Le ST506 était un disque 5.25 pouces qui pouvait stocker jusqu'à 5Mo après formatage, en même temps on pouvait trouver le **ST412** en tout point similaire, si ce n'est sa taille de 10Mo et son prix... Tous les deux ont été présentés en utilisant le codage MFM (déjà utilisé dans les disquettes), mais plus tard on passera au RLL pour un gain de 50 % de capacité et de débit.

Le ST506 était relié à l'ordinateur par un contrôleur S-100 qui pouvait supporter jusque 4 disques. Contrairement à nos disques actuels où les données sont envoyées au disque qui est censé savoir quoi faire, dans le ST506 le système entier était piloté par le contrôleur et le disque était « sourd-muet ». La connexion au contrôleur se faisait via 2 câbles.

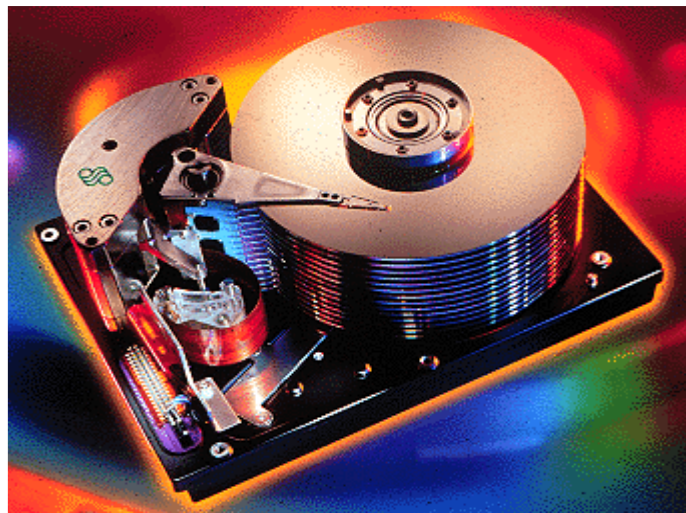
Le premier câble de 34 pin contrôlait la partie mécanique du disque, comme les mouvements des têtes, ensuite les données étaient lues ou écrites via le câble de 20 pin. Cela conduit à des performances assez médiocres dues à la bande passante limitée du câble de données, mais ce n'était pas un problème à l'époque.

De nombreuses autres compagnies ont rapidement créé leurs propres disques en utilisant les mêmes connecteurs et signaux créant ainsi le standard ST506. Vers 1980 on vit apparaître le standard IBM PC qui avait choisi d'utiliser ce type de disques. À cette époque pratiquement tous les disques étaient basés sur le ST506. Bien que la complexité du contrôleur et des câbles ont mené à de nouvelles solutions comme le SCSI et plus tard l'ATA(IDE), la plupart du temps les disques étaient en réalité des ST506 avec un adaptateur SCSI vers ST506 (ou dans certains vers ESDI). Ce fut comme ça jusqu'au début des années 1990 lorsque les contrôleurs ATA et SCSI de Western Digital, Adaptec et Emulex/QLogic devenaient courants.



Le ST506

En 1998 et après un peu moins de 10 ans de recherche, Seagate sortit son disque de 47GB, bien que cela semble peu de nos jours c'était une taille impressionnante à l'époque ! Ces années de recherches ont permis de stocker 100.000 fois plus de données sur la même surface.



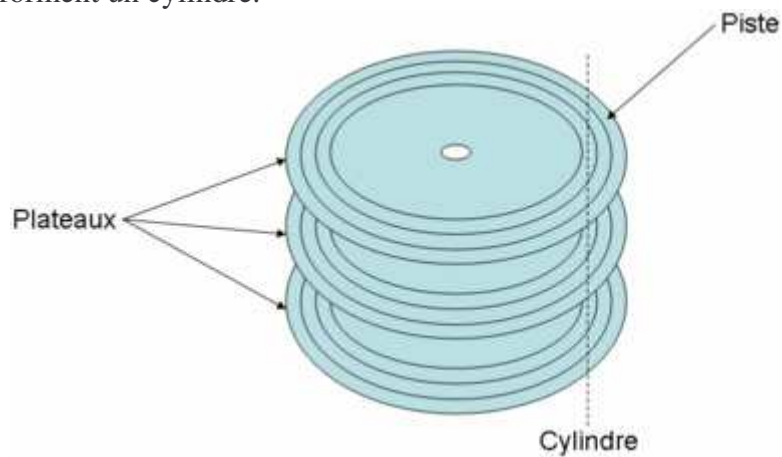
Il faut savoir qu'aujourd'hui, cette même quantité de données peut être stockée sur une seule face d'un unique plateau...

En 2004, Toshiba lance le premier disque 0.85'' de 2GB et un autre de 4GB en 2005 dont vous pouvez admirer la taille ci-dessous.

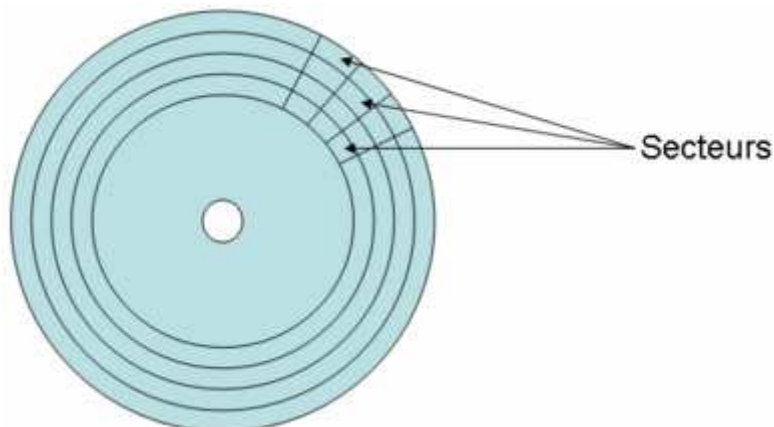


Géométrie

Chaque plateau (2 surfaces) est composé de pistes concentriques. Les pistes situées à un même diamètre forment un cylindre.



Sur une piste les données sont délimitées en secteurs, aussi appelés blocs.



Géométrie d'une surface. Les pistes sont concentriques, les secteurs contigus.

Il faut donc trois coordonnées pour accéder à un bloc :

- le numéro de la tête (choix de la surface)
- le numéro de la piste (détermine le déplacement de la tête)
- le numéro du bloc sur cette piste (détermine la position de départ).

Cette conversion est faite par le contrôleur du disque à partir de l'adresse absolue du bloc (un nombre compris entre 0 et le nombre total de blocs (moins 1) contenu sur le disque).

On notera que les secteurs extérieurs et intérieurs n'ont pas la même taille.

Sur les premiers disques, une surface était formatée en usine et contenait les informations permettant au système de se synchroniser (de savoir quelle était la position des têtes à tout moment). Cette surface était dénommée « servo ». Par la suite, ces zones de synchronisation ont été mixées entre les blocs de données, mais elles sont toujours formatées en usine.

Typiquement donc, on trouvera sur une piste une succession de :

- un petit « blanc » ou « espace » (« gap » en anglais),
- une zone servo,
- un en-tête avec contenant le numéro du bloc qui va suivre,
- les données,
- une somme de contrôle permettant de corriger des erreurs.



Format d'un secteur. Il ne contient pas seulement les données stockées, mais aussi un préambule permettant de synchroniser le système d'asservissement du disque, un en-tête avec l'identifiant du bloc et enfin une somme de contrôle (Σ) permettant de détecter d'éventuelles erreurs.

Performances

Le temps d'accès et le débit d'un disque dur permettent d'en mesurer les performances.

Les facteurs principaux à prendre en compte sont :

1. **Le temps de recherche**, ou seek time en anglais, est le temps que met la tête pour se déplacer jusqu'au cylindre choisi. Il est très variable étant donné que ni la position de la tête ni la distance jusqu'à la piste voulue ne sont fixes. On prendra donc un temps moyen calculé sur base de la distance maximale de la piste (la plus longue possible), et d'une piste à l'autre (la plus courte possible). La manière habituelle de déterminer le temps de recherche est de mesurer un grand nombre d'accès aléatoires au disque et d'y soustraire le temps de latence et ensuite de faire la moyenne. Le temps de recherche est toujours mesuré en millisecondes et est souvent considéré comme le facteur déterminant des performances d'un disque.

2. Le temps de latence : c'est le temps qui s'écoule entre le moment où la tête atteint l'emplacement désiré et le moment où le premier byte requis arrive sous la tête. On peut donc dire pour un disque standard 7200 RPM qu'il est compris entre 0 (si la donnée se trouve justement où il faut quand il faut) et la période de rotation du disque ici 8.4ms donc. La latence moyenne est toujours égale à la moitié de la période de rotation du disque donc quelque soit la marque ou le type de disque un modèle 5400rpm aura 5.56ms de latence moyenne, un 7200rpm 4.17ms, les 10 000 3.0ms et les 15 000 2.0ms, tout comme le temps de recherche, la latence est un facteur déterminant et est toujours mesurée en millisecondes

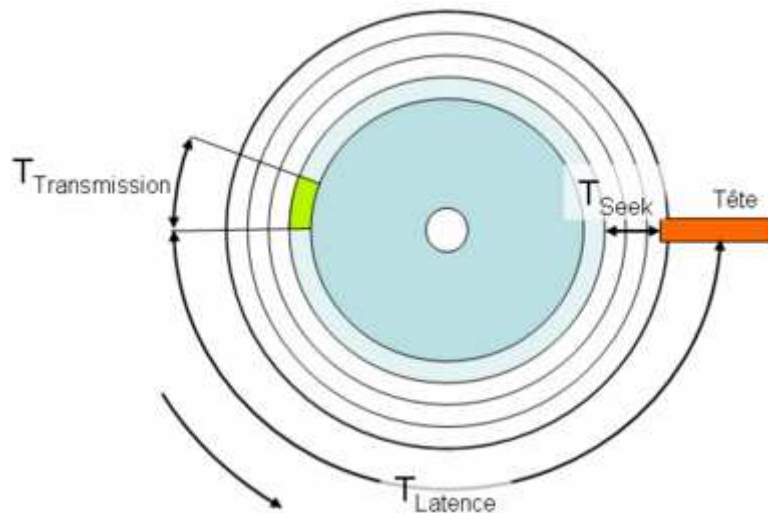
3. Le temps de transfert interne est le temps que vont mettre les données à être transférées entre le disque magnétique et le canal de données du disque. Dans le passé, ce fut un facteur très important, il l'est toujours, mais en moindre mesure, car les disques actuels ont un taux de transfert très élevé.

On peut citer d'autres facteurs subsidiaires comme :

Le temps d'accès : c'est simplement la somme du temps de recherche et de la latence. Il faut faire attention à ne pas confondre temps d'accès et temps de recherche. Le temps d'accès détermine à lui seul la vitesse d'un disque ! Certaines personnes se fient au taux de transfert d'un disque pour le choisir alors que cela fait très rarement une différence de performances... Dans de rares cas, cela pourrait être l'inverse, c'est pourquoi il vous faut connaître votre système avant d'acheter un disque !

Le débit de données externe : c'est la vitesse à laquelle les données transitent entre le PC et le buffer du disque dur. Bien qu'en théorie cela semble crucial, en pratique ce n'est pas un problème. Les interfaces de communication actuelles sont toutes au moins deux fois plus rapides que les disques qu'on leur attache. Par exemple, la différence entre l'ATA100 et l'ATA133 est purement marketing à l'heure actuelle, aucun fabricant n'est capable d'utiliser la bande passante complète de l'ATA-100.

Le temps d'instruction : c'est le temps qu'il faut à la partie électronique pour interpréter les instructions de l'ordinateur et les transformer en commande envoyées au mécanisme de lecture/écriture. Il est négligeable dans les disques modernes.



Pour lire le secteur (en vert) situé sur une piste interne à l'opposée de la tête de lecture (en rouge), il faut déplacer la tête vers l'intérieur (T_{Seek}), attendre que le bloc arrive sous la tête (T_{Latence}) puis lire la totalité du bloc ($T_{\text{Transmission}}$). Il est possible d'optimiser le temps d'accès en prenant en compte la vitesse de rotation durant le déplacement de la tête.

Pour estimer le temps de transfert total, on additionne ces trois temps. On pourra rajouter le temps de réponse du contrôleur, etc. Il faut se méfier des spécifications des constructeurs, ceux-ci auront tendance à communiquer les valeurs de pointe au lieu des valeurs soutenues (par exemple pour les débits).

Voici deux disques comparés : le premier, le DEC RP07 équipait les ordinateurs DEC des années 70-80, tandis que le second, le Maxtor est un disque de 3,5 pouces récent (2004). Ils peuvent tous deux être considérés comme des disques haut de gamme au moment de leur mise sur le marché.

	DEC RP07	Maxtor Atlas 15k
Hauteur (cm)	118	2,6
Largeur (cm)	67,3	10,1
Profondeur (cm)	83,8	14,7
Poids (Kg)	181	0,81
Capacité (Mo)	516	74 752 (73 Go)
Vitesse de rotation (t/m)	3 633	15 000
Temps de latence moyen (ms)	8,3	2
Seek time piste à piste (ms)	5	0,3/0,5
Seek time maximum (ms)	—	9
Seek time moyen	23	3,4/3,8
Taux de transfert maximum (Mo/s)	2,1	100
Taux de transfert soutenu (Mo/s)	—	75
Nombre de surfaces	16 + 1	8
Nombre de plateaux	9	4
Secteur/piste	—	50
Octets/secteur	512	512
Interface	MASSBUS	SCSI Ultra 320

L'ajout de mémoire vive sur le contrôleur du disque permet d'augmenter les performances. Cette mémoire sera remplie par les blocs qui suivent le bloc demandé, en espérant que l'accès aux données sera séquentiel. En écriture, le disque peut informer l'hôte qui a initié le transfert que celui-ci est terminé alors que les données ne sont pas encore écrites sur le média lui-même. Comme tout système de cache, cela pose un problème de cohérence des données.

Capacité de stockage

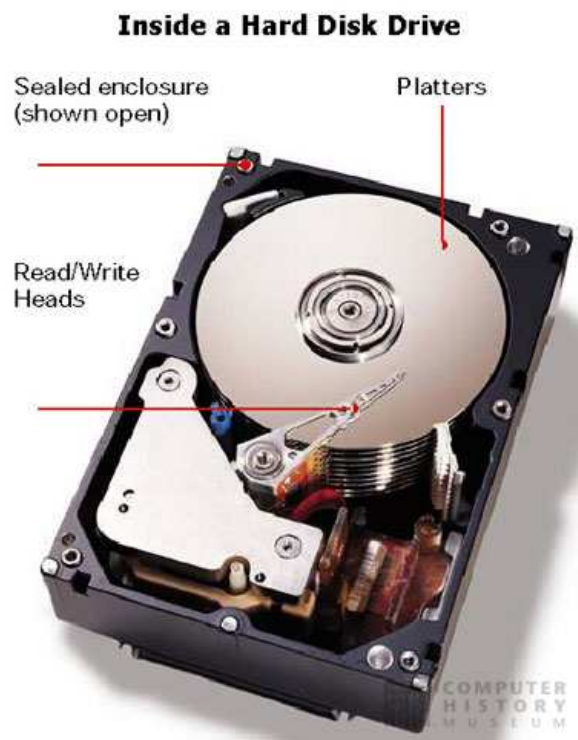
Les capacités actuelles (2005) s'échelonnent entre 20 Go et 2 To (environ 2000 Go). Ce dernier étant le LaCie Bigger Disk, vendu pour la bagatelle de 1850 €.

La capacité des disques durs a augmenté beaucoup plus vite que leur rapidité, limitée par la mécanique.

Capacité réelle VS capacité « Marketing »

Il est important de noter que les fabricants utilisent souvent la définition métrique du préfixe « Giga » alors que pratiquement tous les systèmes d'exploitation utilisent sa définition binaire. L'IEC a décidé de nommer 2^{10} kilobytes, car 1024 était assez proche du préfixe métrique kilo qui lui vaut 10^3 . Ce qui fait que lorsque vous achetez un disque de 200GB vous avez 200×10^9 bytes. Votre système comptant les gigabytes comme 2^{30} soit 1073741824 bytes, la capacité reportée sera d'environ 186.26GB ; ce qui fait une différence de plus de 10GB ! Pour pallier à ce problème, l'IEC a créé de nouveaux préfixes binaires (KiB, MiB, GiB) qui malheureusement ne sont que très rarement utilisés par les fabricants.

Mécanique



Plateaux (Platters)

Les plateaux sont solidaires d'un axe sur roulements à billes. Cet axe est maintenu en mouvement par un moteur électrique. La vitesse de rotation est actuellement (2005) comprise entre 3 600 et 15 000 tours/minute (l'échelle typique des vitesses est 3 600, 4 200, 5 400, 7 200, 10 000 et 15 000 tours/minute). La vitesse de rotation est maintenue constante.

Les plateaux sont composés d'un substrat, autrefois en aluminium, de plus en plus souvent en verre, traité par diverses couches dont une ferromagnétique recouverte d'une couche de protection. L'état de surface doit être le meilleur possible.

La chambre scellée (Sealed enclosure)

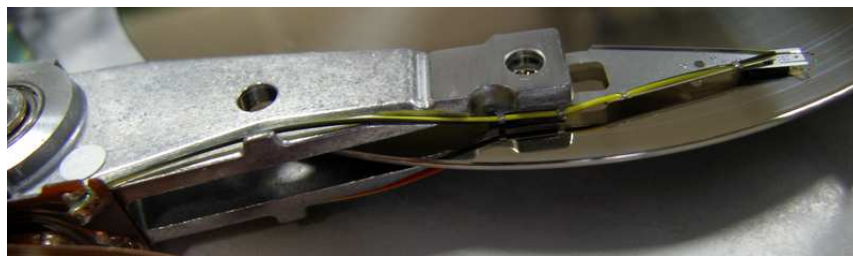
La chambre scellée (ou presque) protège le disque de la poussière, de la condensation et d'autres sources de contamination. Les têtes du disque planent sur un coussin d'air de seulement quelques nanomètres au-dessus de la surface du disque. Cette surface ainsi que l'intérieur doivent donc rester parfaitement propres, les traces de doigt, les cheveux, la poussière et même les particules de fumée ont une taille gigantesque comparée, à la hauteur microscopique à laquelle les têtes se maintiennent.

Certaines personnes pensent qu'un disque dur contient un aspirateur qui permettrait aux têtes de voler, c'est faux, le système repose uniquement sur la pression de l'air à l'intérieur du disque pour soutenir les têtes à leur hauteur de vol lorsque le disque est en mouvement. Une autre erreur courante est de penser qu'un disque est totalement scellé. Un disque nécessite une certaine pression atmosphérique pour fonctionner correctement. Si la pression est trop faible, l'air n'exercera pas assez de force sur les têtes, elles ne seront pas à la bonne hauteur et il y aura donc un gros risque de crash disque et de perte de données. On peut quand même signaler qu'il existe des disques qui eux sont réellement scellés et sont prévus pour travailler dans des conditions extrêmes, par exemple : à une altitude de plus de 10000 pieds (cela ne s'applique pas aux cabines d'avions qui elles sont pressurisées). Les disques modernes ont des sondes de température et peuvent modifier leurs réglages pour travailler dans des conditions idéales selon leur environnement.

Les disques ne sont pas hermétiques. Ils possèdent une petite ouverture munie de filtres pour permettre au disque d'équilibrer la pression interne par rapport à l'externe tout en prévenant l'introduction de poussières. Par contre, le filtre laisse passer avec l'air les particules de moisissure, ce qui posera un problème si le disque est utilisé de manière prolongée dans des environnements très humides en donnant aux têtes tendance à rester collées au disque. L'air dans le disque est constamment en mouvement, entraîné par la rotation des plateaux ; on trouvera d'ailleurs un filtre interne qui permet de nettoyer un peu l'air ambiant en retenant les particules qui seraient entrées d'une façon ou d'une autre dans le disque ainsi que les particules métalliques produites par l'atterrissage des têtes.

Tête de lecture/écriture (Read/Write heads)

Fixées au bout d'un bras, elles sont solidaires d'un second axe qui permet de les faire pivoter en arc de cercle sur la surface des plateaux. Toutes les têtes pivotent donc en même temps. Il y a une tête par surface. Leur géométrie leur permet de voler au dessus de la surface du plateau sans le toucher : elles reposent sur un coussin d'air créé par la rotation des plateaux. En 1997, les têtes volaient à 25 nanomètres de la surface des plateaux, aujourd'hui (2005) cette valeur est d'environ dix nanomètres.



Le bras supportant les deux têtes de lecture/écriture. Les rayures visibles sur la surface du plateau indiquent que le disque dur était en panne, victime d'un « atterrissage ».

Le moteur qui les entraîne doit être capable de fournir des accélérations et décélérations très importantes. Un des algorithmes de contrôle des mouvements du bras porte-tête est d'accélérer au maximum puis de freiner au maximum pour que la tête se positionne sur le bon cylindre. Il faudra ensuite attendre un court instant pour que les vibrations engendrées par le freinage s'estompent.

À l'arrêt, les têtes doivent être parquées, soit sur une zone spéciale, soit en dehors des plateaux.

Si une ou plusieurs têtes entrent en contact avec la surface des plateaux, cela s'appelle un atterrissage et provoque la destruction des informations situées à cet endroit. La mécanique des disques durs est donc assemblée en salle blanche et toutes les précautions (joints, etc.) sont prises pour qu'aucune impureté ne puisse pénétrer à l'intérieur du boîtier (appelé « HDA » pour « Head Disk Assembly » en anglais).

À cause de l'espace extrêmement mince entre les têtes et la surface des plateaux, la moindre contamination des têtes ou plateaux peut conduire à un crash de la tête au cours duquel elle a de très fortes chances d'arracher une partie du film magnétique.

De plus, la tête va surchauffer à cause du frottement et rendra donc le disque illisible tant que la température ne sera pas retombée.

Un crash peut avoir de multiples causes dont celles-ci sont quelques exemples : un problème dans la partie électronique du disque, une brusque chute de tension, un choc ou simplement un disque de mauvaise qualité.

Normalement lorsqu'un disque n'est plus alimenté il positionne ses têtes dans une zone sûre prévue à cet effet, appelée zone d'atterrissage qui ne contient pas de données.

Malgré cela, et surtout sur les anciens disques, une brusque coupure de courant peut résulter en un arrêt du disque avec les têtes positionnées sur une zone de données ce qui a de fortes chances d'altérer les données.

Les nouveaux disques sont conçus de telle façon que l'inertie de la rotation des disques permette aux têtes de se ranger dans une zone sûre en cas de coupure de courant. Ces dernières années, IBM a créé un nouveau système appelé « head unloading » qui dépose les têtes sur des « rampes d'atterrissage » plutôt que sur le disque, ce qui évite donc aux têtes de coller au disque. D'autres fabricants ont également commencé à utiliser cette technique.

Les technologies pour la conception des têtes (en 2005) :

Tête inductive :

Pour exécuter les fonctions de lecture et d'écriture, le système à tête inductive n'est constitué que d'un seul électroaimant.

En mode lecture

Lorsqu'une zone magnétique, grâce à la rotation du disque, passe à proximité de l'électroaimant, il se forme un courant électrique dans le bobinage. Ce courant est détecté et permet de connaître les informations enregistrées sur le disque.

En mode écriture

La tête d'écriture est dite « inductive », c'est-à-dire qu'elle est capable de générer un champ magnétique. C'est en polarisant négativement ou positivement la surface du disque que la tête est capable d'y inscrire des informations. Pourquoi positivement ou négativement ? Cela vient du fait que le disque dur ne stocke que des données binaires, donc ces deux états suffisent. Quand l'électroaimant est sous tension, il produit un champ magnétique qui, par induction, crée sur le disque une zone magnétique durable.

Tête MR – MagnétoRésistive

Ce système de lecture utilise en fait deux têtes de lecture : une pour l'écriture et une pour la lecture, celle de lecture étant basée sur un nouveau principe.

En mode lecture :

La tête magnétorésistive est réservée à la lecture. En présence d'un champ magnétique, sa résistance électrique se trouve modifiée. Ces changements, détectés, permettent de connaître les informations enregistrées sur le disque.

En mode écriture :
Une tête inductive est chargée de l'écriture.

Tête GMR — Giant MagnétoRésistive

Une tête GMR est comparable, dans sa structure, à une tête magnéto-résistive.

L'écriture continue à être assurée par une tête inductive.

La lecture, quant à elle, repose sur un phénomène découvert par deux physiciens (Albert Fert et Peter Gruenberg). Comme avec une tête magnéto-résistive, on mesure les changements affectant la résistance électrique d'un élément placé au voisinage d'un champ magnétique, mais le dispositif met en œuvre des matériaux en couches extrêmement minces dont la réponse est beaucoup plus forte que celle obtenue avec le procédé magnéto-résistif (d'où le terme giant). Le dispositif étant plus sensible, on peut exploiter des traces magnétiques beaucoup plus petites sur le disque.

Électronique

Elle est composée d'une partie dédiée à l'asservissement des moteurs et d'une autre à l'exploitation des informations électriques issues de l'interaction électromagnétique entre les têtes de lecture et les surfaces des plateaux. Une partie plus informatique va faire l'interface avec l'extérieur et la traduction de l'adresse absolue d'un bloc en coordonnées à 3 dimensions (tête, cylindre, bloc).

L'électronique permet aussi de corriger les erreurs.

La technologie S.M.A.R.T

Le S.M.A.R.T ou Self-Monitoring, Analysis, and Reporting Technology est un système de surveillance interne du disque dur qui lui permet en se basant sur plusieurs indicateurs de fiabilité de reporter des pannes à venir.

Certaines pannes, surtout les mécaniques apparaissent graduellement avec le temps. En surveillant le disque, on peut détecter cela. Par contre, d'autres, comme les pannes électroniques, sont imprévisibles ; on ne sait pas quand un composant grillera...

Le S.M.A.R.T peut prédire environ 60 % des types de pannes. Il préviendra alors l'utilisateur ou l'administrateur que le disque est sur le point de le lâcher et qu'il faut penser à prendre des mesures préventives, comme faire un backup des données.

Par exemple, le WDC800JB-00CRA1 (disque Western Digital de 80Go) peut surveiller les paramètres suivants :

- Raw read error rate
- Spin up time

- Start/stop count
- Reallocated sector count
- Seek error rate
- Power on hours count
- Spin retry count
- Recalibration retry count
- Power cycle count
- Reallocation event count
- Current pending sector count
- Uncorrectable sector count
- CRC error count
- Write error count

(notons que ce disque-ci n'est pas équipé d'une sonde de température)

Il n'est pas toujours facile de savoir ce que surveille exactement un disque particulier, car les fabricants ne donnent pas toujours tous ces paramètres.

Compaq fut le pionnier pour le disque avec SMART, mais à l'heure actuelle pratiquement tous les disques en sont équipés et les cartes mères récentes sont capables d'afficher un avertissement lorsqu'un disque approche de sa mort.

On peut trouver un tas de petits logiciels permettant de surveiller son disque via le smart et même y définir ses propres seuils d'alerte.

Types d'interfaces des disques durs

Les interfaces des disques durs ont largement évolué avec le temps dans un souci de simplicité et d'augmentation des performances.

Voici quelques interfaces possibles :

SMD (Storage Module Device), très utilisé dans les années 1980, elle était principalement réservée pour les disques de large capacité installés sur des serveurs.

ST506, très utilisé au début de la micro informatique dans les années 1980.

ESDI (Enhanced Small Device Interface), a succédé au ST506, qu'il améliore.

(Ces trois ci n'étant plus d'actualité, je ne les détaillerai pas.)

L'interface IDE,

la plus courante dans les machines personnelles :

Tour d'horizon

Le standard **ATA** (Advanced Technology Attachment) est une interface standard permettant la connexion de périphériques de stockage sur les ordinateurs de type PC. Le standard ATA a été mis au point le 12 mai 1994 par l'ANSI (document X3.221-1994).

Malgré l'appellation officielle "ATA", ce standard est plus connu sous le terme commercial IDE (Integrated Drive Electronics) ou **Enhanced IDE (EIDE ou E-IDE)**.

Le standard ATA était originalement prévu pour connecter des disques durs, toutefois une extension nommée **ATAPI** (ATA Packet Interface) a été développée afin de pouvoir interfacier d'autres périphériques de stockage (lecteurs de CD-ROM, lecteurs de DVD-ROM, etc.) sur une interface ATA.

Depuis l'émergence de la norme **Serial ATA** (notée **S-ATA** ou **SATA**), permettant de transférer les données en série, le terme « **Parallel ATA** » (noté **PATA** ou **P-ATA**) remplace parfois l'appellation "ATA" afin de marquer le contraste entre les deux normes.

Principe

La norme ATA permet de relier des périphériques de stockage directement à la carte mère grâce à une **nappe IDE** (en anglais, ribbon cable) généralement composée de 40 fils parallèles et de trois connecteurs (un connecteur pour la carte mère, généralement bleu, et les connecteurs restants pour deux périphériques de stockage, respectivement noir et gris).



Sur la nappe un des périphériques doit être déclaré comme **maître** (master), l'autre en **esclave** (slave). Par convention le connecteur à l'extrémité (noir) est réservé au périphérique maître et le connecteur du milieu (gris) au périphérique esclave. Un mode appelé **cable select** (noté **CS** ou **C/S**) permet de définir automatiquement le périphérique maître et l'esclave pour peu que le BIOS de l'ordinateur supporte cette fonctionnalité.

Par abus de langage, il est considéré que le disque maître contrôle le disque esclave, il y aurait en quelque sorte une hiérarchie entre les disques, ce qui est faux ; le seul intérêt du cavalier est de permettre une cohésion entre les disques étant donné qu'ils utilisent la même nappe pour transmettre les informations, autrement dit le disque maître n'a pas nécessairement la priorité sur l'esclave.

Modes PIO

La transmission des données se fait grâce à un protocole appelé **PIO** (Programmed Input/Output) permettant aux périphériques d'échanger des données avec la mémoire vive à l'aide de commandes gérées directement par le processeur. Toutefois, de gros transferts de données peuvent rapidement imposer une grosse charge de travail au processeur et ralentir l'ensemble du système. Il existe 5 modes PIO définissant le taux de transfert maximal :

Mode PIO	Débit (Mo/s)
Mode 0	3.3
Mode 1	5.2
Mode 2	8.3
Mode 3	11.1
Mode 4	16.7

Modes DMA

La technique du DMA (Direct Memory Access) permet de désengorger le processeur en permettant à chacun des périphériques d'accéder directement à la mémoire. On distingue deux types de modes DMA :

- Le DMA dit "single word" (en français mot simple) permettant de transmettre un mot simple (2 octets soient 16 bits) à chaque session de transfert,
- Le DMA dit "multi-word" (en français mots multiples) permettant de transmettre successivement plusieurs mots à chaque session de transfert.

Le tableau suivant liste les différents modes DMA et les taux de transfert associés :

Mode DMA	Débit (Mo/s)
0 (Single word)	2.1
1 (Single word)	4.2
2 (Single word)	8.3
0 (Multiword)	4.2
1 (Multiword)	13.3
2 (Multiword)	16.7

À partir de l'Ultra DMA mode 4 un nouveau type de nappe a été introduit afin de limiter les interférences ; il s'agit d'une nappe ajoutant 40 fils de masse (soit un total de 80), entrecalés avec les fils de données afin de les isoler et possédant les mêmes connecteurs que la nappe de 40 fils.



Seuls les modes Ultra DMA 2, 4, 5 et 6 sont réellement implémentés dans les disques durs.

Les normes ATA

Le standard ATA se décline en plusieurs versions, ayant vu le jour successivement :

ATA-1

Le standard **ATA-1**, connu sous le nom de **IDE**, permet la connexion de deux périphériques sur une nappe de 40 fils et propose une transmission 8 ou 16 bits avec un débit de l'ordre de 8.3 Mo/s. **ATA-1** définit et supporte les modes **PIO** (Programmed Input/Output) 0, 1 et 2 ainsi que les **modes DMA** (Direct Memory Access) 0, 1 et 2.

ATA-2

Le standard **ATA-2**, connu sous le nom de **EIDE** (parfois **Fast ATA**, **Fast ATA-2** ou **Fast IDE**), permet la connexion de deux périphériques sur une nappe de 40 fils et propose une transmission 8 ou 16 bits avec un débit de l'ordre de 16.6 Mo/s.

ATA-2 permet le support des modes **PIO** 0, 1, 2, 3 et 4 et **modes DMA** 0, 1 et 2. De plus, ATA-2 permet de repousser la limite de la taille maximale de disque de 528 Mo imposée par la norme ATA-1 à 8.4 Go grâce au **LBA** (Large Block Addressing).

ATA-3

Le standard **ATA-3** (également appelé ATA Attachment 3 Interface) représente une révision mineure de l'ATA-2 (avec une compatibilité descendante) et fut publié en 1997 sous le standard X3.298-1997. Le standard ATA-3 apporte les améliorations suivantes :

- **Fiabilité améliorée** : L'ATA-3 permet d'augmenter la fiabilité des transferts à haute vitesse.
- **Le S.M.A.R.T** (Self-Monitoring Analysis and Reporting Technology, traduisez Technologie d'auto surveillance, d'analyse et de rapport) dont on a déjà parlé.
- **Fonction de sécurité** : les périphériques peuvent être protégés à l'aide d'un mot de passe ajouté dans le BIOS. Au démarrage de l'ordinateur, celui-ci vérifie que le mot de passe codé dans le BIOS correspond à celui stocké sur le disque. Cela permet notamment d'empêcher l'utilisation du disque sur une autre machine.

ATA-3 n'introduit pas de nouveau mode, mais supporte les modes **PIO** 0, 1 et 2 ainsi que les modes **DMA** 0, 1 et 2.

ATA-4

Le standard **ATA-4**, ou **Ultra-ATA/33**, a été défini en 1997. **ATA-4** modifie également le mode LBA afin de porter la capacité maximale des disques à 128 Go. En effet, le mode LBA permet un adressage codé par un nombre binaire de 28 bits. Or chaque secteur représente 512 octets, ainsi la capacité maximale exacte d'un disque dur en mode LBA est la suivante :

$$2^{28} * 512 = 137\,438\,953\,472 \text{ octets}$$
$$137\,438\,953\,472 / (1024 * 1024 * 1024) = 128 \text{ Go}$$

ATA-5

En 1999 le standard **ATA-5** définit deux nouveaux modes de transfert : **Ultra-DMA mode 3 et 4** (le mode 4 est aussi appelé Ultra ATA/66 ou Ultra DMA/66) De plus il propose la détection automatique du type de nappes utilisées (80 ou 40 fils).

ATA-6

Depuis 2001 **ATA-6** définit le support de l'**Ultra DMA/100** (aussi appelé **Ultra DMA mode 5** ou **Ultra-ATA100**) permettant d'atteindre des débits théoriques de 100 Mo/s.

D'autre part, ATA-6 définit une nouvelle fonctionnalité, appelée Automatic Acoustic Management (**AAM**), ajustant automatiquement la vitesse d'accès aux disques supportant cette fonction afin d'en réduire le bruit de fonctionnement.

ATA-7

Depuis 2002 **ATA-7** définit le support de l'**Ultra DMA/133** (aussi appelé Ultra DMA mode 6 ou Ultra-ATA133) permettant d'atteindre des débits théoriques de 133 Mo/s.

L'ATA-8 est actuellement en développement.

SCSI (Small Computer System Interface),

plus coûteuse, mais offrant des performances supérieures.

Historique

Le bus SCSI a été initialement créé par la société Shugart, du nom de son fondateur Al Shugart, un pionnier de l'informatique, ancien ingénieur chez IBM (spécialiste des disques durs), sous le nom de SASI (Shugart Associates Systems Interface) en 1979.

Shugart se lie avec NCR Corporation en 1981 dans l'espoir d'en faire un standard ANSI. Un comité technique fut créé, qui améliorera sensiblement le bus SASI, le renomme SCSI et la norme X3.131 le décrivant verra le jour en 1986.

La première version définissait trop de paramètres optionnels pour la norme (notamment dans le jeu de commandes, mais aussi dans les possibilités matérielles comme pour les bits de parité etc.). Les incompatibilités entre fabricants étaient nombreuses.

Des améliorations furent apportées en 1994 avec la norme SCSI-2 et récemment avec la norme SCSI-3, réduisant ces problèmes d'interopérabilité.

Spécificité

Ce qui différencie ce bus des autres est qu'il déporte l'intelligence vers le périphérique lui-même. De ce fait les commandes envoyées au périphérique peuvent être complexes, c'est sous le contrôle du périphérique qu'elles seront (éventuellement) décomposées en sous-tâches plus simples, ce qui est extrêmement avantageux si l'on travaille avec des systèmes d'exploitation multitâches.

Cette interface est donc plus rapide, plus universelle et plus complexe que l'interface E-IDE dont le principal inconvénient est d'accaparer un pourcentage non négligeable du processeur, ce qui constitue un handicap quand de nombreux flux de données sont simultanément ouverts. Plus « intelligente » et moins dépendante vis-à-vis de l'unité centrale, l'interface SCSI peut gérer des périphériques internes et externes très variés, tels que disques durs, scanners, graveurs, unités de sauvegardes, etc.

Le nombre de périphériques pouvant être branché dépend de la largeur du bus SCSI. En effet, avec un bus 8 bits il est possible de connecter 8 unités physiques, contre 16 pour un bus 16 bits. Le contrôleur SCSI représentant une unité physique à part entière, le bus peut donc accepter 7 (8 - 1) ou 15 (16 - 1) périphériques.

Adressage des périphériques

L'adressage des périphériques se fait grâce à des numéros d'identification. Le premier numéro est l'ID, il s'agit d'un numéro permettant de désigner le contrôleur intégré à chaque périphérique (celui-ci est défini grâce à des cavaliers à positionner sur chaque périphérique SCSI ou bien logiciellement). En effet, le périphérique peut avoir jusqu'à 8 unités logiques (par exemple un lecteur de CD-ROM comportant plusieurs tiroirs). Les unités logiques sont repérées par un identificateur appelé LUN (Logical Unit Number). Enfin, un ordinateur peut comporter plusieurs cartes SCSI, c'est pourquoi un numéro de carte est assigné à chacune d'entre elles.

De cette façon, pour communiquer avec un périphérique, l'ordinateur doit donner une adresse de la forme « numéro de carte - ID - LUN ».

Périphériques concernés

La norme SCSI-2 précise que le bus peut relier des ordinateurs avec des périphériques tels que :

- disques durs
- lecteurs de disquettes
- lecteurs de bandes magnétiques
- imprimantes
- lecteurs de disque optique (WORM)
- lecteurs de disque optique (CD-ROM)
- scanners
- changeur de média
- périphériques de communication

On note que la norme ne restreint pas l'utilisation du bus à l'interconnexion d'un ordinateur avec des périphériques, mais qu'il peut être utilisé entre des ordinateurs, ou pour partager des périphériques entre ordinateurs.

La norme SCSI-3 est plus généraliste. On se référera à la page du comité technique pour en avoir le détail (www.t10.org).

Modes de transfert

Norme	Interface	Vitesse Bus (MOctets/s)	Fréquence de bus (Mhz)	Taille Bus (bits)	Longueur de câble SE max. (mètres)	Longueur de câble LVD max. (mètres)	Longueur de câble HVD max. (mètres)
SCSI-1	SCSI	5	5	8	6	-	25
SCSI-2	Fast SCSI	10	10	8	3	-	25
	Fast Wide SCSI	20	10	16	3	-	25
SCSI-3	Ultra SCSI	20	20	8	1,5	-	25
	Ultra Wide SCSI	40	20	16	3	-	-
	Ultra2 SCSI	40	40	8	-	12	25
	Ultra2 Wide SCSI	80	40	16	-	12	25
	Ultra3 SCSI	80	80	16	-	12	-
	Ultra-160 SCSI	160	80	16	-	12	-
	Ultra-320 SCSI	320	160	16	-	12	-
	Ultra-640 SCSI	640	160	16	-	12	-

Améliorations du SCSI-3 par rapport au SCSI-2

Le SCSI-3 apporte comme énorme changement l'apparition d'un bus série, dans une technologie jusqu'à présent exclusivement parallèle. Il apporte aussi quelques nouveautés et améliorations dans l'interface parallèle.

Interfaces parallèles

Ultra 3

La fréquence du bus est doublée. On passe donc de 40Mhz à 80 Mhz. Les périphériques SE ne sont plus du tout compatibles avec ces fréquences à cause des phénomènes de réverbération.

Ultra 160

L'interface Ultra 160 utilise le LVD, elle n'est absolument plus compatible avec SCSI-1 et 2. L'Ultra 160 apporte de lourdes modifications dans la gestion du transfert de données :

- Mise en œuvre du CRC. Le bus étant en LVD, on a de plus grandes vitesses et donc de bien plus grands risques d'erreurs dans les transmissions. C'est pourquoi on a ajouté la génération d'un CRC afin de pouvoir vérifier l'intégrité des données.
- Domaine Validation. La carte hôte négocie avec le périphérique la meilleure vitesse à utiliser (un peu à l'image de ce que font les modems), ce qui évite qu'on ne communique plus vite avec le périphérique qu'il n'en est capable (ce qui auparavant le rendait inaccessible)
- Double Transition Clocking. On transfère les données sur deux flancs (montant et descendant), ce qui permet de doubler le débit. C'est pourquoi on parle d'ultra160 (80Mhz * 2)

Ultra 320

L'interface 320 apporte des modifications supplémentaires par rapport à l'ultra160, ce qui permet d'augmenter encore la fréquence de travail à 160Mhz et d'augmenter les débits utiles.

- **Packetized SCSI.** L'ultra320 introduit un protocole de gestion de paquets de données. Ce qui a pour conséquence de réduire le nombre de commandes à transférer en même temps que les données. On peut aussi transmettre plusieurs commandes dans un seul paquet.
- **QAS (Quick Arbitration and Selection).** Nouvelle méthode de gestion du bus qui réduit le nombre de commandes nécessaires et donc le trafic "inutile "

- **Read and Write Data Streaming.** On peut envoyer plusieurs paquets d'affilée et n'attendre la commande de confirmation de réception qu'au moment de la réception de tous les paquets, ce qui réduit encore une fois le trafic "inutile "
- **Flow Control.** Prétraitement des données à transférer dans une pile FIFO, ce qui permet de regrouper des données et d'effectuer un transfert en rafale quand un paquet est prêt.
- **Precompensation.** L'augmentation de la fréquence signifie la baisse de l'amplitude du signal. Un décalage qui n'aurait eu aucune conséquence avant devient gênant. À l'initialisation du bus, on teste donc chacun des fils et on crée une précompensation afin que lors de l'envoi de données sur le bus, toutes les données arrivent en même temps à destination (même si certains fils sont plus rapides que d'autres)

Interfaces série

Améliorations par rapport à l'interface parallèle

Avec l'explosion des fréquences, les problèmes de décalage entre les signaux et leur sensibilité au bruit et aux capacités parasites deviennent problématiques et causent des restrictions dans la taille des câbles. On passe donc sur des bus série (un seul fil) ce qui évite les problèmes de courants induits.

La réflexion sur des interfaces séries fut commencée avant la mise au point de l'Ultra160 et 320. Mais les débits offerts sont tels que ces technologies restent tout à fait viables.

Le but de l'interface série est d'empaqueter les commandes et données SCSI et de les transférer via un seul fil, tout en conservant la compatibilité SCSI (afin de garder avantage notamment de la possibilité de stockage des commandes, très utile en multitâche) On retrouve de nombreux avantages dans ces technologies, on peut citer notamment :

- des architectures point à point
- double accès : on peut accéder à un point de deux manières différentes, ce qui introduit une meilleure résistance aux pannes sur le bus
- Possibilités plus grandes au niveau des systèmes RAID, on peut monter à plus de 100 disques durs grâce à SSA ou Fiber Channel contre 5 avec une interface SCSI classique

Technologies utilisées

On peut citer comme technologies série utilisant les commandes SCSI :

- FC-AL
- SSA
- FireWire
- iSCSI

Autres technologies utilisées avec les commandes SCSI-3				
	Interface parallèle	Interfaces série		
	Ultra SCSI	FC-AL	SSA	IEEE P 1394
Protocole	SCSI Interlock Protocol (SIP)	Fiber Channel Protocol (FCP)	Serial Storage Protocol (SSP)	Serial Bus Protocol (SBP)
Vitesse Max	5 à 640 Mo/s	100 à 400 Mo/s	20 à 80 Mo/s	12,5 à 200 Mo/s
Nombre de périphériques	7 à 15	126	128	63
Distance	12 à 25m	20 à 200m (cuivre) 10Km (fibre optique)	20m (cuivre) 680m (fibre optique)	72m

Améliorations du SCSI-2 par rapport au SCSI-1

La norme SCSI-2 de 1994 est une amélioration du SCSI-1. Certains points ont été améliorés ou rendus obligatoires. Théoriquement SCSI-1 et SCSI-2 ont une compatibilité descendante. On peut noter que l'utilisation des bits de parité a été rendue obligatoire avec le SCSI-2

Fast SCSI

On utilise des transferts synchrones de haut débit, ce qui permet des taux de transfert de 10Mo/s sur un câblage 8bits, et de 20 ou 40Mo si on est sur du 16 ou 32bits (avec une fréquence de 10Mhz)

Wide SCSI

On utilise toujours le même type de transfert à 10Mhz, mais on peut atteindre des débits doubles par rapport au Fast SCSI. Il existe aussi en version 8, 16 et 32 bits

File d'attente des commandes

Le SCSI-1 ne permettait d'envoyer les commandes qu'une à une ; le SCSI-2 permet d'envoyer jusqu'à 256 commandes à un périphérique. Elles seront stockées et traitées dans l'ordre optimal par le périphérique. Ceci permet d'augmenter les performances de travail du périphérique et prend un réel intérêt quand on travaille avec des systèmes d'exploitation multitâche (Linux, MS-Windows NT etc.) qui peuvent être amenés à faire plusieurs requêtes simultanément à un périphérique (plusieurs accès en lecteur sur un disque dur par exemple)

Types de câblage

Les périphériques SCSI sont reliés à leurs contrôleurs par un connecteur 50,68 ou 80 broches.

- Singled Ended (SE) : On travaille en mode asymétrique. Pour chaque signal à envoyer par le bus correspond un fil unique.

- High Voltage Differential (HVD) : On travaille en mode différentiel. Pour chaque signal, on utilise deux fils : le premier porte le signal, le second porte son opposé. Le récepteur traite la différence entre les deux
- Low Voltage Differential (LVD) : Idem HDV, mais avec des tensions plus petites, permettant des câbles plus longs (meilleure tolérance aux capacités parasites et au bruit)



Les connecteurs SCSI

Résistances de terminaison

- Résistance passive : En fin de câble, on établit deux résistances de 220 et 330 Ω , soit 132 Ω . La 220 est reliée à la masse, la 330 à l'alimentation qui est fournie par un des composants du bus (qui ne doit donc jamais être éteint si on veut que le bus fonctionne). Mais à fréquences trop élevées, ce type de résistance de terminaison provoque des réverbérations du signal.
- Résistance active (ou Alternative2) : Résistance de 110 Ω obtenue par des régulateurs de tension. Elle évite les phénomènes de réverbération.
- Résistance parfaite forcée : Variante de l'Alternative2 qui permet des câbles plus longs

Serial ATA (ou S-ATA),

est une interface série, peu coûteuse et plus rapide (normes SATA et SATA II).

Le standard **Serial ATA** (S-ATA ou SATA) est un bus standard permettant la connexion de périphériques de stockage haut débit sur les ordinateurs de type PC.

Le standard Serial ATA est apparu en février 2003 afin de pallier les limitations de la norme ATA, qui utilise un mode de transmission en parallèle.

En effet, le mode de transmission en parallèle n'est pas prévu pour supporter des fréquences élevées en raison des problèmes liés aux interférences électromagnétiques entre les différents fils.

Principe du Serial ATA

Le standard Serial ATA est basé sur une communication en série. Une voie de données est utilisée pour transmettre les données et une autre voie sert à la transmission d'accusés de réception. Sur chacune de ces voies, les données sont transmises en utilisant le mode de transmission LVDS (Low Voltage Differential Signaling) consistant à transférer un signal sur un fil et son opposé sur un second fil afin de permettre au récepteur de reconstituer le signal par différence. Les données de contrôle sont transmises sur la même voie que les données en utilisant une séquence de bits particulière pour les distinguer.

Ainsi, la communication demande deux voies de transmission, chacune effectuée via deux fils, soit un total de quatre fils pour la transmission.

Connecteurs Serial-ATA

Le câble utilisé par le Serial ATA est un câble composé de sept fils, le plus souvent de couleur rouge, et terminé par un connecteur de 8 mm :



Trois fils servent à la masse et les deux paires servent à la transmission de données.

Le connecteur d'alimentation est également différent : il est composé de 15 broches permettant d'alimenter le périphérique en 3.3V, 5V ou 12V et possède une allure similaire au connecteur de données :



Caractéristiques techniques

Le Serial ATA permet d'obtenir des débits de l'ordre de 187.5 Mo/s (1,5 Gb/s), or chaque octet est transmis avec un bit de démarrage (start bit) et un bit d'arrêt (stop bit), soit un débit utile théorique de 150 Mo/s (1,2 Gb/s). Le standard Serial ATA II permet d'avoisiner les 375 Mo/s (3 Gb/s), soit 300 Mo/s utiles théoriques, puis à terme 750 Mo/s (6 Gb/s), soit 600 Mo/s utiles théoriques.

Les câbles Serial ATA peuvent mesurer jusqu'à 1 mètre de long (contre 45 cm pour les nappes IDE). De plus, le faible nombre de fils dans une gaine ronde permet plus de souplesse et une meilleure circulation de l'air dans le boîtier qu'avec des nappes IDE (même si des nappes IDE rondes existent). Contrairement à la norme ATA, les périphériques Serial ATA sont seuls sur chaque câble et il n'est plus nécessaire de définir des "périphériques maîtres" et des "périphériques esclaves"

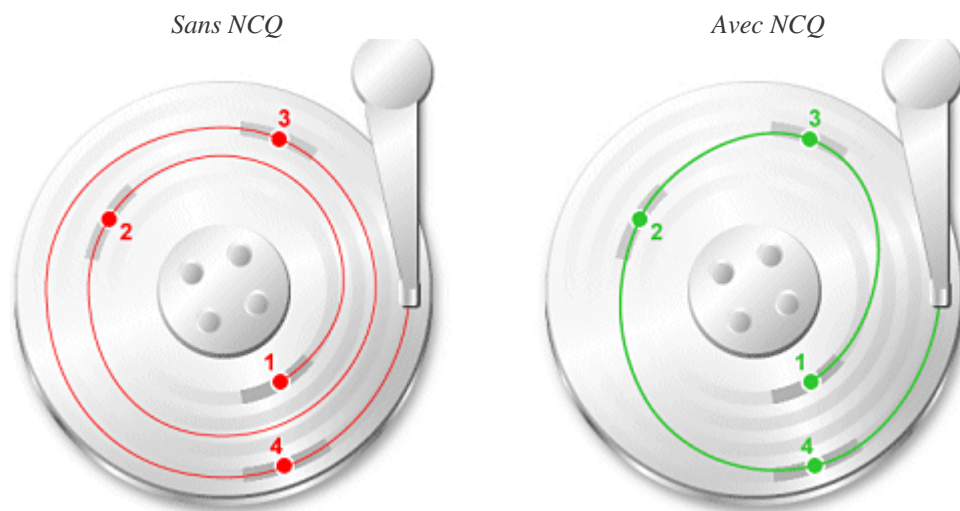
D'autre part, la norme Serial ATA permet le raccordement à chaud des périphériques (Hot Plug).

Le NCQ (Native Command Queuing)

Avec l'arrivée du SATA-II on voit apparaître la technologie NCQ qui lui est spécifiquement réservée.

Mis au point par Seagate et Intel, le NCQ permet aux applications d'envoyer, d'organiser et de traiter plus rapidement de nombreuses requêtes d'entrées/sorties envoyées en même temps. Cette technologie permet aussi d'interroger directement le processeur sur les données qui doivent être traitées par le disque dur, celles-ci sont ensuite classées et ordonnées afin d'optimiser au maximum le flux de données traité.

Voici comment Seagate représente le fonctionnement du NCQ :



On peut voir qu'en réorganisant les commandes et en lisant donc les informations dans un autre ordre on a pu limiter les déplacements et donc optimiser l'opération.

Les premiers disques équipés de cette technologie commencent à arriver sur le marché.

SAS (Serial Attached SCSI),

combine les avantages du SCSI avec ceux du Serial ATA (elle est compatible avec cette dernière).

Le SAS est en fait une version sérielle du SCSI qui a été conçu en 1986 de façon parallèle. Le SAS fut ratifié en 2003 comme la nouvelle génération de SCSI. C'est une technologie point à point utilisant un contrôleur de disque avec quatre canaux ou plus qui peuvent fonctionner simultanément. Chaque canal est full-duplex et peut atteindre un débit de 300MBps dans chaque direction, on attend un débit de 600MBps pour la prochaine génération. La compatibilité avec le Serial ATA permet des configurations mixtes SCSI/SATA.

Le protocole SSP (Serial SCSI Protocol) est utilisé pour les disques SAS alors que le protocole STP (Serial ATA Tunneling Protocol) est utilisé pour les disques SATA.

On peut connecter directement un disque aux différents canaux du contrôleur ou utiliser un expander qui permet de connecter alors jusque 128 disques par canaux. Dans ce cas on utilisera le protocole SMP (Serial Management Protocol). On peut étendre la capacité d'un système SAS jusqu'à 16256 disques.

Les disques SAS utilisent les mêmes connecteurs que le SATA mais on y a ajouté un 2^{ème} connecteur permettant de le relier à 2 contrôleurs ou expanders, pour les problèmes de tolérance d'erreur.

Fibre-Channel,

est un successeur du SCSI. La liaison est série et peut utiliser une connectique fibre optique ou cuivre (coaxial ou paires torsadées). Principalement utilisé sur les serveurs.

Le fibre channel est surtout utilisé pour les SAN (Storage Area Network) nécessitant une bande passante très élevée. Le FCP (Fibre Channel Protocol) envoie les commandes SCSI en série dans des trames Fibre Channel. Il est aussi utilisé pour transporter d'autres protocoles, comme l'IP.

Le Fibre Channel peut aussi bien être utilisé dans une configuration point à point, avec des switches, ou en boucle (FC-AL) avec ou sans hub qui peut connecter jusqu'à 127 nœuds. Il supporte un taux de transfert de 2.12Gbps dans chaque direction et bientôt 4.25Gbps. Il utilise les couches physiques du Gigabit Ethernet et la méthode d'encodage IBM 8B/10B ou chaque bit est transmis comme 10.

Les interfaces externes :

On trouve aussi des disques durs externes qu'on peut brancher en USB ou Firewire et il existe maintenant des disques SATA externes. Je ne détaillerai pas ces interfaces qui ne sont pas spécifiquement liées aux disques durs.

Formats

Les dimensions des disques durs sont normalisées :

19 pouces pour les anciens disques (à interface SMD).

8 pouces génération suivante, permettant de mettre deux disques sur une largeur de baie.

5 pouces 1/4 : format apparu dans les années 1980, exista aussi en demi-hauteur.

3 pouces 1/2 est la taille standard à ce jour (2005).

2 pouces 1/2 pour les ordinateurs portables.

1 pouce 8 pour les baladeurs mp3 et certains disques durs externes.

Les plus petits disques entrent dans la catégorie des microdrives, avec une taille de 1 pouce (0.85 pour les plus petits).

Autres caractéristiques

En plus des caractéristiques ayant été citées dans les paragraphes précédents, on peut retrouver les suivantes :

- La consommation électrique : importante surtout pour les ordinateurs portables et autres appareils fonctionnant sur batterie
- Le niveau de bruits émis : Exprimé en dBA. Les PC de salon étant de plus en plus répandus, on tente de les rendre le plus silencieux possible pour ne pas être dérangé pendant un film par exemple.
- La résistance aux accélérations : Elle est mesurée en g (lui-même valant 9.81m/s^2). C'est simplement la résistance aux chocs du disque. Elle est étonnement élevée dans les disques haut de gamme, comme les Raptors de Western Digital, capables de résister à un choc de 250g.

Fabricants

Le nombre de fabricants de disques durs est assez limité de nos jours (2005), en raison de divers rachats ou fusions d'entreprises, voire l'abandon par certains de cette activité.

- Cornice
- Hitachi GST
- ExcelStor
- Fujitsu
- GS-Magic
- Maxtor
- Seagate
- Western Digital
- Samsung
- Toshiba

Constructeurs historiques :

- Bull Périphériques
- CDC (Imprimis)
- Conner Peripherals
- Hewlett-Packard
- IBM
- Micropolis
- NEC
- Quantum
- Storage Technology
- Tandem
- Univac

Émulation

Parfois il est nécessaire d'avoir un périphérique en tout point similaire à un disque dur, mais avec des temps d'accès beaucoup plus rapide, au détriment de la capacité. Il y a deux façons d'atteindre ce but : soit par l'utilisation d'un disque SSD, soit par la création d'un disque virtuel, comme décrit ci-dessous.

La technologie SSD

SSD signifie Solid State Disque : extérieurement un disque SSD aura la même apparence qu'un disque dur classique, y compris pour son interface. Seulement, il ne contient aucun élément mécanique, les données sont stockées sur de la mémoire flash. Les temps d'accès sont très rapides, mais la capacité ne dépasse pas 16 Go à ce jour (2005). Cette technologie est utilisée principalement dans les environnements où les disques durs habituels ne peuvent fonctionner (vide, accélérations importantes, etc.). On peut aussi utiliser ce type de disques pour leur vitesse qui en général sont largement supérieure aux disques standards, ou encore, lorsque des problèmes de sécurité se posent : il est beaucoup plus facile d'effacer définitivement les données sur ce type de disque.

Les disques virtuels

Parfois aussi appelés RAM-disks. C'est un artifice qui permet d'émuler un disque dur à partir d'un espace alloué en mémoire centrale. Sa création, son effacement et son accès se font par le biais d'appels systèmes (le noyau peut contenir des pilotes adéquats). Les temps d'accès sont excessivement rapides, par contre la capacité ne peut excéder une fraction de la taille de la mémoire centrale. Les données sont perdues si la mémoire n'est plus alimentée électriquement.

Gigabyte a sorti cette année une carte PCI sur laquelle on peut installer des modules DIMM DDR-SDRAM jusqu'à une taille de 4Go qui se branche en SATA et qui possède une batterie évitant la perte des données.

Le RAID

Le terme de RAID (Redundant Array of Independent/Inexpensive Disks, c'est-à-dire un groupe de disques redondants et indépendants/bon marché) désigne une architecture matérielle (et parfois logicielle) permettant d'accélérer, de sécuriser et/ou de fiabiliser les accès aux données stockées sur disques durs. Cette architecture est basée sur la multiplication des disques durs, par opposition à la méthode sled (Single Large Expensive Disk) où toutes les données sont rassemblées sur un seul disque de prix élevé.

La première description de cette architecture apparut dans une publication de 1987 intitulée « A Case of Redundant Arrays of Inexpensive Disks (RAID) (Patterson, Gibson & Katz - Université de Californie - Berkeley) ». Cet article comparait le RAID au sled et proposait cinq niveaux différents de RAID, chacun d'eux ayant ses avantages et ses inconvénients.

Description

Dans une architecture de type sled (utilisation d'un seul disque), la bonne conservation des données est dépendante de la moindre défaillance du disque dur. Lorsqu'une panne survient, la seule façon de récupérer les données est le recours à la dernière sauvegarde, sachant que la restauration d'une sauvegarde à partir d'une bande peut prendre plusieurs heures, pendant lesquelles l'ordinateur est inutilisable.

Pour un particulier, un tel temps d'inactivité de la machine (downtime) est acceptable, mais pour une société travaillant 24 heures sur 24, une telle panne peut rapidement devenir critique pour la santé de l'entreprise. Dans ce cas, l'utilisation de certains niveaux de RAID permet de minimiser les risques de perte de données, car :

- la défaillance d'un des disques du RAID ne gêne pas le fonctionnement des autres disques et la machine continue de fonctionner normalement,
- les données du disque défaillant sont reconstruites à partir des autres disques pendant l'utilisation normale de l'ordinateur.

Ainsi, même en cas de panne de l'un des disques, l'activité de l'entreprise peut continuer de façon ininterrompue.

D'autre part, si une application nécessite un disque dur permettant des temps de lecture et d'écriture extrêmement rapides, le RAID permet d'atteindre d'excellents temps de réponse, même en utilisant des disques bon marché et de performance moyenne.

Dans de telles situations, les architectures RAID se révèlent donc idéales, tant du point de vue de leurs performances que de leur fiabilité. D'autre part, si le RAID semble compliquer l'architecture de l'ordinateur, il reste complètement invisible à l'utilisateur de l'ordinateur ; en effet, quel que soit le nombre de disques physiques utilisés pour construire le RAID, l'utilisateur de la machine ne verra jamais qu'un seul grand disque logique, auquel il accédera de façon tout à fait habituelle...

Les différents types de RAID

Les différents types d'architectures RAID sont numérotés à partir de 0 et peuvent se combiner entre eux (on parlera alors de RAID 0+1, 1+0, etc.).

RAID 0 (striping)

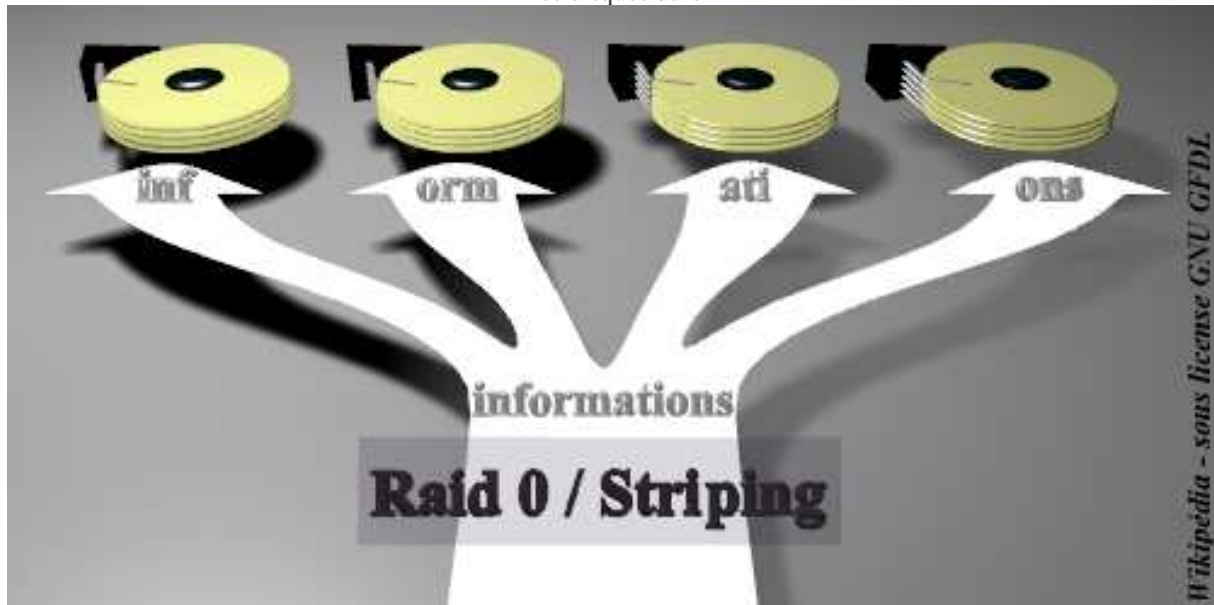
Dans cette configuration, chaque octet est divisé en autant de morceaux qu'il y a de disques. Exemple : avec un RAID 0 composé de quatre disques, si l'on veut écrire l'octet 00111001, le stockage des différents bits composant cet octet se fera de la façon suivante :

- disque 1 : 00
- disque 2 : 11
- disque 3 : 10
- disque 4 : 01

Ainsi, sur un RAID 0 de n disques, chaque disque ne doit lire et écrire que 1/n des données, ce qui a pour effet de décupler les taux de transfert des données entre le CPU et les disques, et donc d'accélérer les traitements.

Ce type de RAID est parfait pour des applications requérant un traitement rapide d'une grande quantité de données. Mais cette architecture n'assure en rien la sécurité des données ;

en effet, si l'un des disques tombe en panne, la totalité des données du RAID est perdue, ce qui fait du RAID 0 une solution moins fiable que l'utilisation d'un seul disque de stockage, puisque la probabilité de défaillance d'un des disques du RAID est largement supérieure à la probabilité de défaillance d'un disque unique.



RAID 1 (mirroring)

Le RAID 1 consiste en l'utilisation de disques redondants, c'est-à-dire n disques (en général deux) sur lesquels sont copiées exactement les mêmes données. Si cette solution n'apporte aucun gain de performance, elle permet en revanche de sécuriser les données en cas de défaillance d'un des disques. Il est à noter que dans ce type de RAID, la perte de capacité de stockage liée à l'utilisation des disques, est égale à 50% de la capacité totale des disques utilisés. Cela revient à dire que l'on perd l'équivalent de la capacité de stockage d'un disque dur entier (si l'on a deux disques durs de 120 Go soit une capacité totale de 240 Go, on perd donc 120 Go).

Chaque écriture de données s'effectue simultanément sur tous les disques participants au RAID. Tous les disques du RAID sont donc à tout moment interchangeables. Lors de la défaillance de l'un des disques, le contrôleur RAID (matériel ou logiciel) désactive instantanément le disque incriminé, sans que le fonctionnement général de l'ordinateur n'en soit affecté : les traitements continuent à s'effectuer sur les disques restants. Une fois le disque défectueux remplacé, le contrôleur RAID entame, soit automatiquement, soit suite à une intervention de l'administrateur de la machine, une tâche de synchronisation du nouveau disque avec les disques actifs.

En général, cette tâche s'effectue pendant le fonctionnement normal de la machine, sans perturber les autres traitements. Ce n'est que lorsque cette synchronisation est terminée (c'est-à-dire quand le contenu du nouveau disque est à nouveau identique à celui des autres) que ce nouveau disque est intégré dans le RAID en tant que disque actif.



RAID 3

Le RAID 3 nécessite une matrice de trois disques au minimum.

n-1 disques contiennent les données (type RAID 0) et le dernier disque stocke la parité.

En effet si un des disques de données tombe en panne, il est possible de reconstruire l'information avec le disque de parité et dans le cas où celui-ci tombe en panne, le système devient alors un RAID 0.

Il est important que le disque de parité soit de bonne qualité, car il est à tout instant sollicité à l'écriture. Ce dernier point est la limitation du RAID 3.

RAID 4

Le RAID 4 est sensiblement semblable au RAID 3 sauf qu'il travaille par blocs et non par octets, ce qui ne nécessite davantage de synchronisation entre les disques.

RAID 5

Le RAID 5 associe le striping et un système à parité répartie, il permet une bonne disponibilité (même en cas de défaillance d'un des périphériques de stockage).

Exemple pratique : sur la base de trois disques durs de taille identique A, B et C, le système va enregistrer les données sur les disques A et B comme en mode RAID 0 (striping) et, sur le disque C, le résultat de l'opération ou-exclusif entre A et B ($A \text{ xor } B$).

Ainsi, en cas de défaillance du disque A, les données qui y étaient accessibles le sont toujours avec les disques B et C, par l'opération $B \text{ xor } C$, en effet $B \text{ xor } C = A$. Il en va de même pour le disque B, et si le disque C tombe en panne, les informations sont toutes sur A et B.

Important : les disques doivent être de même taille ; vous ne pourrez stocker que sur les deux tiers de la place totale des disques (le dernier tiers étant le disque C). En fait la capacité utile réelle est $(n-1 \text{ disques}) \times \text{capacité}$. Trois disques de 100 Go en RAID 5 vous offriront 200 Go utiles, dix disques, 900 Go utiles.

RAID 6

Une évolution du RAID 5 qui accroît la sécurité en utilisant deux parités au lieu d'une. Si la sécurité est plus grande, le coût en est plus élevé (deux disques au lieu d'un) et la vitesse moindre (deux écritures au lieu d'une).

Le RAID 6 est peu utilisé à cause de son surcoût et la maigre probabilité d'une panne simultanée de deux disques.

RAID 7

L'évolution du RAID 3 permettant de faire fonctionner le tout de manière asynchrone, il a été annoncé comme plus performant de 1,5x à 6x que n'importe quel autre RAID.

Tout comme le RAID 3, un seul disque contient les parités de tous.

Chaque disque a un contrôleur SCSI et le système est régi par une carte calculant la parité, gérant le cache et contrôlant les disques.

Cette version, développée par Storage Computer Corporation, est propriétaire et s'avère très coûteuse à mettre en place.

RAID 0+1

La cascade d'un RAID 0 et d'un RAID 1. Deux groupes de n disques en RAID 0, ces deux groupes sont eux en RAID 1.

Une panne dans un des groupes et le groupe est mort. Si un disque dans les 2 groupes meure, le tout est perdu.

RAID 10 (ou RAID1+0)

Cascade d'un RAID 1 et d'un RAID 0. n groupes de deux disques en RAID 1, tous ceux-ci en RAID 0. Ici, il faut que deux disques d'un même groupe rendent l'âme pour que le tout soit perdu. Ce qui réduit la probabilité.

Le RAID 10 est plus fiable et plus rapide à se reconstruire en cas de panne que le 01. Il est donc plus courant.

RAID 50

Cascade d'un RAID 5 et d'un RAID 0. n groupes de trois disques en RAID 5, ceux-ci en RAID 0. Ici, il faut que deux disques d'un même groupe rendent l'âme pour que le tout soit perdu. Ce qui réduit la probabilité. Un des meilleurs compromis lorsque l'on cherche la rapidité, lors de gros flux vidéo par exemple, donne tout son potentiel lors de l'utilisation d'un contrôleur de RAID par groupe de disques.

Nombre d'interfaces vers les disques

Un paramètre à ne pas oublier dans la conception d'un RAID est le nombre de cartes d'interface vers les disques (cartes IDE ou SCSI). En effet, si le RAID que l'on veut mettre en place se compose d'un nombre important de disques, la carte d'interface à laquelle sont connectés les disques deviendra vite un goulet d'étranglement lors du transfert des données du CPU vers le RAID. Si la fiabilité du RAID n'en est pas affectée, les performances, elles, se révéleront médiocres, voire mauvaises.

Il ne faut donc pas hésiter à multiplier le nombre de cartes d'interface. D'ailleurs, idéalement, chaque disque participant à un RAID devrait être connecté à sa propre carte IDE ou SCSI de façon à ce qu'il puisse utiliser le taux de transfert maximal permis par l'interface à laquelle il est connecté.

Matériel ou Logiciel

Le système RAID est une redondance ou une répartition pour des raisons de performances de données sur plusieurs disques. Il faut donc gérer d'une manière ou d'une autre, la répartition et la cohérence de ces données. Deux moyens peuvent s'offrir pour gérer ceci :

Le RAID logiciel

Il est assuré par le système d'exploitation lui-même. Méthode la moins onéreuse puisque ne demandant pas de matériel complémentaire, il est aussi le moins souple et le moins performant en général, l'ensemble des calculs de répartition ou de parité devant être gérés par le système. Les migrations à la volée sont rarement faisables (passer d'une RAID 1 à un RAID 5 par exemple).

Le RAID matériel

Une carte ou un composant est dédié à la gestion des opérations. Il peut s'agir d'un composant sur les cartes mères récentes en 2004/2005, qui gère souvent les RAID 0 et 1 sur des disques IDE ou SATA. Pour les machines demandant davantage de performances, tels que les serveurs, des cartes supplémentaires, en PCI pour la plupart, intègrent un processeur dédié à la gestion du RAID, souvent de la mémoire consacrée au cache disque, et parfois des batteries pour sauvegarder la mémoire du cache d'écriture en cas de coupure de courant.

Ces cartes additionnelles offrent la possibilité de reconstruire de manière transparente les disques défectueux, de changer les propriétés ou d'étendre la capacité des unités RAID. Cette dernière fonction n'est toutefois pas toujours reconnue par le système d'exploitation. En conséquence, des disques transférés d'un système à l'autre ne pourront pas être récupérés si le contrôleur RAID n'est pas le même, au contraire d'un RAID logiciel.

Utilisation de disques Hot Swappable

Ces disques sont connectables et extractibles à chaud, c'est-à-dire qu'ils ne nécessitent pas d'arrêt de la machine hôte. Ainsi, il est possible de les retirer d'un RAID lorsqu'ils sont morts et de les remplacer pour les reconstruire avant ou pendant l'utilisation.

Le formatage

Le formatage bas niveau

Le formatage physique ou de bas niveau consiste à organiser la surface de chaque plateau en pistes et secteurs, en polarisant grâce aux têtes d'écriture des zones du disque. Les pistes sont numérotées en partant de 0, puis les têtes polarisent concentriquement la surface des plateaux. Lorsque l'on passe à la piste suivante, la tête laisse un "trou" et ainsi de suite. Chaque piste est elle-même organisée en secteurs (numérotés en commençant à partir de 1) séparés entre eux par des gaps. Chacun de ces secteurs commence par une zone réservée aux informations du système appelée préfixe et se termine par une zone appelée suffixe.

Le formatage de bas niveau a donc pour but de préparer la surface du disque à accueillir des données (il ne dépend donc pas du système d'exploitation et permet grâce à des tests effectués de "marquer les secteurs défectueux).

Somme de contrôle

Effectivement, pendant le formatage, des tests de contrôle (algorithme permettant de tester la validité des secteurs grâce à des sommes de contrôle) sont effectués et chaque fois qu'un secteur est considéré comme défectueux, la somme de contrôle (invalide) est inscrite dans le préfixe, dès lors il ne pourra plus être utilisé par la suite, on dit qu'il est "marqué défectueux".

Lorsque le disque lit des données, il envoie une valeur qui dépend du contenu du paquet envoyé, et qui est initialement stockée avec ceux-ci. Le système calcule cette valeur en fonction des données reçues, puis la compare avec celle qui était stockée avec les données. Si ces deux valeurs sont différentes, les données ne sont pas valides, il y a probablement un problème de surface du disque.

Le contrôle de redondance cyclique (CRC: en anglais cyclic redundancy check), est basé sur le même principe pour contrôler l'intégrité d'un fichier.

Les utilitaires d'analyse tels que scandisk ou chkdsk opèrent autrement: ils inscrivent des données sur les secteurs à priori marqués valides, puis les relisent et les comparent. Si ceux-ci sont similaires, l'utilitaire passe au secteur suivant, dans le cas contraire ils marquent le secteur défectueux.

Formatage de haut niveau

Le formatage logique ou de haut niveau s'effectue après le formatage de bas niveau, il crée un système de fichiers sur le disque qui va permettre à un système d'exploitation (DOS, Windows 95, Linux, OS/2, Windows NT, ...) d'utiliser l'espace disque pour stocker et utiliser des fichiers.

Les systèmes d'exploitation utilisent des systèmes de fichiers différents, le type de formatage logique dépend donc du système d'exploitation que vous installez. Si vous formatez votre disque en un seul système de fichiers, cela limite naturellement le nombre et le type de systèmes d'exploitation que vous installez (en effet, vous ne pourrez installer que des systèmes d'exploitation utilisant le même système de fichiers).

Heureusement, il y a une solution à ce problème qui consiste à créer des partitions. Chacune des partitions peut effectivement avoir son propre système de fichiers, vous pourrez par conséquent installer des systèmes d'exploitation de natures diverses.

Le partitionnement

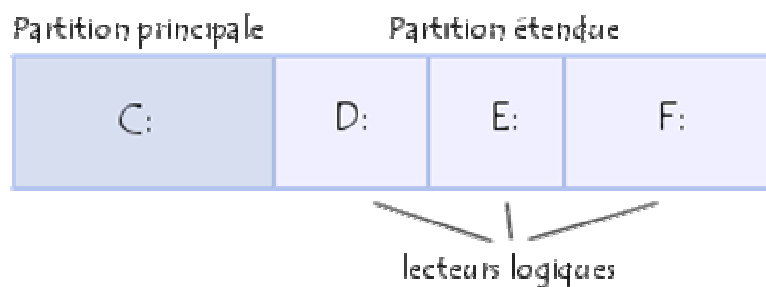
Qu'est-ce qu'une partition?

On appelle partitionnement le processus qui consiste à écrire les secteurs qui constitueront la table de partition (qui contient les informations sur la partition: taille de celle-ci en terme de nombre de secteurs, position par rapport à la partition principale, types de partitions présentes, systèmes d'exploitation installés,...).

Lorsque la partition est créée, on lui donne un nom de volume qui va permettre de l'identifier facilement.

Le partitionnement d'un disque dur se fait après le formatage physique de celui-ci et avant le formatage logique. Il consiste à créer des zones sur le disque dont les données ne seront pas mélangées. Cela sert par exemple à installer des systèmes d'exploitation différents n'utilisant pas le même système de fichiers. Il y aura donc au minimum autant de partitions que de systèmes d'exploitation utilisant des systèmes de fichiers différents. Dans le cas d'un utilisateur d'un système d'exploitation unique, une seule partition de la taille du disque peut suffire, sauf si l'utilisateur désire en créer plusieurs pour faire par exemple plusieurs lecteurs dont les données sont séparées.

Voyons un exemple, dans lequel le disque contient une partition principale et une partition étendue composée de trois lecteurs logiques (nous verrons par la suite les partitions principales multiples):



Pour les systèmes DOS (DOS, Windows 9x), seule la partition principale est bootable, c'est donc la seule sur laquelle on peut démarrer le système d'exploitation.

Master Boot Record

Le secteur de démarrage (appelé Master Boot Record ou MBR en anglais) est le premier secteur d'un disque dur (cylindre 0, tête 0 et secteur 1), il contient la table de partition principale (en anglais partition table) et le code, appelé boot loader, qui, une fois chargé en mémoire, va permettre d'amorcer (booter) le système.

Ce programme, une fois en mémoire, va déterminer sur quelle partition le système va s'amorcer, et il va démarrer le programme (appelé bootstrap) qui va amorcer le système d'exploitation présent sur cette partition.

D'autre part, c'est ce secteur du disque qui contient toutes les informations relatives au disque dur (fabricant, numéro de série, nombre d'octets par secteur, nombre de secteurs par cluster, nombre de secteurs,...). Ce secteur est donc le secteur le plus important du disque dur, il sert au setup du BIOS à reconnaître le disque dur. Ainsi, sans celui-ci, votre disque dur est inutilisable, c'est donc une cible de prédilection pour les virus.

Utiliser des partitions multiples

Il y a, comme on l'a vu précédemment, trois sortes de partitions: les partitions principales, la partition étendue et les lecteurs logiques. Un disque peut contenir jusqu'à quatre partitions principales (dont une seule peut être active), ou trois partitions principales et une partition étendue. Dans la partition étendue l'utilisateur peut créer des lecteurs logiques (c'est-à-dire faire en sorte que l'on ait l'impression qu'il y a plusieurs disques durs de taille moindre).

Partition principale

Une partition principale doit être formatée logiquement, pour ensuite contenir un système de fichier correspondant au système d'exploitation installé sur celle-ci.

Si vous avez plusieurs partitions principales sur votre disque, une seule sera active et visible à la fois, cela dépendra du système d'exploitation sur lequel vous avez démarré l'ordinateur. En choisissant le système d'exploitation que vous lancez au démarrage, vous déterminerez la partition qui sera visible. La partition active est la partition sur laquelle un des systèmes d'exploitation est démarré au lancement de l'ordinateur. Les autres partitions seront alors cachées, ce qui empêchera d'accéder à leurs données. Ainsi, les données d'une partition principale ne sont accessibles qu'à partir du système d'exploitation installé sur cette partition.

Partition étendue

La partition étendue a été mise au point pour outrepasser la limite des quatre partitions principales, en ayant la possibilité de créer autant de lecteurs logiques que vous désirez dans celle-ci. Un lecteur logique au moins est nécessaire dans une partition étendue, car vous ne pouvez pas y stocker de données directement.

Beaucoup de machines sont formatées en une grande partition utilisant l'intégralité de l'espace disponible du lecteur. Ce n'est pourtant pas la solution la plus avantageuse en terme de performances et de capacité, la solution est de créer plusieurs partitions, ce qui vous permettra :

- d'installer plusieurs systèmes d'exploitation sur votre disque
- d'économiser de l'espace disque
- d'augmenter la sécurité de vos fichiers
- d'organiser vos données plus facilement

Voyons à quoi pourrait ressembler un système comportant plusieurs systèmes d'exploitation :



Les systèmes de fichiers

Qu'est-ce qu'un système de fichiers

Les disques durs, aussi petits soient-ils, contiennent des millions de bits, il faut donc organiser les données afin de pouvoir localiser les informations, c'est le but du système de fichiers. Un disque dur est, rappelons-le, constitué de plusieurs plateaux circulaires tournant autour d'un axe. Les pistes (zones concentriques écrites de part et d'autre d'un plateau) sont divisées en quartiers appelés secteurs (d'une taille de 512 octets). Le formatage logique d'un disque permet de créer un système de fichiers sur le disque, qui va permettre à un système d'exploitation (DOS, Windows 9x, UNIX, ...) d'utiliser l'espace disque pour stocker et utiliser des fichiers. Le système de fichiers est basé sur la gestion des clusters (en français « unité d'allocation »), c'est-à-dire la plus petite unité de disque que le système d'exploitation soit capable de gérer.

Un cluster est constitué d'un ou plusieurs secteurs, ainsi plus la taille d'un cluster est importante, moins le système d'exploitation aura d'entités à gérer...

En contrepartie, étant donné qu'un système d'exploitation ne sait gérer que des unités d'allocation entière, c'est-à-dire qu'un fichier occupe un nombre entier de clusters, le gaspillage sera d'autant plus grand qu'il y a de secteurs par cluster. On comprend alors toute l'importance du choix du système de fichiers.

Systeme de fichiers et système d'exploitation

En réalité, le choix du système de fichiers se fait en premier lieu selon le système d'exploitation que vous utiliserez. D'une manière générale, plus le système d'exploitation est récent plus le nombre de systèmes de fichiers supportés sera important. Ainsi, sous DOS et sur les premières versions de Windows 95 la FAT16 est de rigueur.

À partir de Windows 95 OSR2, vous aurez le choix entre les systèmes de fichiers FAT16 et FAT32. Si jamais la taille de la partition est supérieure à 2Go, le système de fichier FAT est exclu, vous devez donc utiliser le système FAT32 (ou modifier la taille de la partition).

En dessous de cette limite, la FAT16 sera recommandée pour des partitions d'une capacité inférieure à 500Mo, dans l'autre cas, l'utilisation de FAT32 sera préférable.

Dans le cas de Windows NT (jusqu'à la version 4) vous avez le choix entre le système FAT16 et NTFS, par contre celui-ci ne supportera pas la FAT32. D'une manière générale, le système NTFS est conseillé, car il procure une plus grande sécurité ainsi que des performances accrues par rapport à la FAT. Microsoft recommande en fait d'utiliser une petite partition (comprise entre 250 et 500Mo) de type FAT pour le système d'exploitation, afin de pouvoir démarrer à partir d'une disquette DOS bootable en cas de problème, et de conserver les données sur une seconde partition pour les stocker.

Sous Windows NT5 l'éventail s'agrandit puisqu'il accepte des partitions de type FAT16, FAT32 et NTFS. Une fois de plus, le système de fichiers le plus récent (NTFS 5) est conseillé, puisqu'il offre des fonctionnalités plus nombreuses que les systèmes FAT. Pour les mêmes raisons que précédemment, vous pourrez toutefois opter pour une partition de type FAT.

Systeme d'exploitation	Types de systeme de fichiers supportés
Dos	FAT16
Windows 95	FAT16
Windows 95 OSR2	FAT16, FAT32
Windows 98	FAT16, FAT32
Windows NT4	FAT, NTFS (version 4)
Windows 2000/XP	FAT, FAT16, FAT32, NTFS (versions 4 et 5)
Linux	Ext2, Ext3, ReiserFS, Linux Swap
MacOS	HFS (Hierarchical File System), MFS (Macintosh File System)
OS/2	HPFS (High Performance File System)
SGI IRIX	XFS
FreeBSD, OpenBSD	UFS (Unix File System)
Sun Solaris	UFS (Unix File System)
IBM AIX	JFS (Journaled File System)

La cohabitation de plusieurs systèmes de fichiers

Lorsque plusieurs systèmes d'exploitation cohabitent sur une même machine, le problème du choix du système de fichiers est à son paroxysme. En effet, le système de fichiers est étroitement lié au système d'exploitation, ainsi lorsqu'il y a plusieurs systèmes d'exploitation, il faut choisir pour chacun d'entre eux le système d'exploitation en tenant en compte qu'il est possible que l'on ait à accéder à des données de l'un à partir d'un autre. Une première solution consiste à utiliser des partitions FAT pour tous les systèmes, en faisant attention à n'utiliser que des partitions d'une taille inférieure à 2Go. La solution la mieux adaptée est d'utiliser pour chacun des systèmes une partition dont le système de fichiers sera le mieux adapté, et de dédier une partition en FAT16 aux données vouées à être partagées par les différents systèmes d'exploitation.

Ce que l'on peut trouver actuellement

Tous les prix proviennent du site <http://www.ldlc.be> et datent du 05/11/2005

Les disques les plus courants :

Maxtor DiamondMax Plus 9 120 Go 8 Mo 7200RPM ATA133	70€
Maxtor DiamondMax Plus 9 120 Go 8 Mo 7200 RPM Serial ATA	70€
Western Digital WD1200JB 120 Go 7200 tpm 8 Mo Serial IDE	77€
Western Digital WD1200JD 120 Go 7200 tpm 8 Mo Serial ATA	84€
Hitachi Deskstar T7K250 - 160 Go 7200 RPM 8 Mo Serial ATA II	80€
Samsung SpinPoint P - HD160JJ - 160 Go 7200 RPM 8 Mo Serial ATA II	96€

Les grandes capacités :

Maxtor MaXLine III 300 Go 7200 RPM 16 Mo IDE	150€
Maxtor MaXLine III 300 Go 7200 RPM 16 Mo Serial ATA	146€
Maxtor DiamondMax Plus 11 - 500 Go 16 Mo IDE	383€
Hitachi Deskstar 7K500 - 500 Go 7200 RPM 8 Mo Serial ATA II	363€
Seagate Barracuda 7200.9 NCQ 500 Go 7200 RPM 16 Mo Serial ATA II	403€

Les hautes performances :

Seagate 73 Go 10000 RPM SCSI 80 broches	322€
Seagate Cheetah 10K.7 146.8 Go Ultra320 SCSI 10000 RPM 80 broches	545€
Western Digital Raptor 74 Go 10000 RPM 8 Mo Serial ATA	192€
Maxtor Atlas 15K II - 73.5 Go 15000 RPM 8 Mo Ultra320 SCSI 68 broches	474€
Seagate Cheetah 15K.4 - 73.4 Go 15 000 RPM 8 Mo Ultra320 SCSI 68 broches (bulk)	606€
Seagate Cheetah 15K.4 - 146 Go 15 000 RPM 8 Mo Ultra320 SCSI 68 broches	1208€

Les signatures :

Hitachi Microdrive 4 Go	166€
Hitachi Microdrive 6 Go	227€

Tables de matières

Sommaire	3
Définition :	4
Historique :	4
Géométrie	8
Performances	9
Capacité de stockage	12
Capacité réelle VS capacité « Marketing »	12
Mécanique	12
Plateaux (Platters)	13
La chambre scellée (Sealed enclosure)	13
Tête de lecture/écriture (Read/Write heads).....	14
Les technologies pour la conception des têtes (en 2005) :	15
Tête inductive :	15
Tête MR – MagnétoRésistive.....	15
Tête GMR — Giant MagnétoRésistive.....	16
Électronique	16
La technologie S.M.A.R.T	16
Types d'interfaces des disques durs.....	17
L'interface IDE,	17
Tour d'horizon	17
Principe.....	18
Modes PIO.....	19
Modes DMA.....	19
Les normes ATA	20
ATA-1	20
ATA-2	20
ATA-3	20
ATA-4	21
ATA-5	21
ATA-6	21
ATA-7	21
SCSI (Small Computer System Interface),	21
Historique	21
Spécificité.....	22
Adressage des périphériques	22
Périphériques concernés	23
Modes de transfert.....	23
Améliorations du SCSI-3 par rapport au SCSI-2	24
Interfaces parallèles.....	24
Ultra 3.....	24
Utra 160.....	24
Ultra 320.....	24
Interfaces série.....	25
Améliorations par rapport à l'interface parallèle	25
Technologies utilisées	25
Améliorations du SCSI-2 par rapport au SCSI-1	26
Fast SCSI.....	26

Wide SCSI.....	26
File d'attente des commandes.....	26
Types de câblage.....	26
Résistances de terminaison.....	27
Serial ATA (ou S-ATA),.....	27
Principe du Serial ATA.....	28
Connecteurs Serial-ATA.....	28
Caractéristiques techniques.....	29
Le NCQ (Native Command Queuing).....	29
SAS (Serial Attached SCSI),.....	30
Fibre-Channel,.....	30
Les interfaces externes :.....	31
Formats.....	31
Autres caractéristiques.....	31
Fabricants.....	32
Émulation.....	32
La technologie SSD.....	32
Les disques virtuels.....	33
Le RAID.....	33
Description.....	33
Les différents types de RAID.....	34
RAID 0 (striping).....	34
RAID 1 (mirroring).....	35
RAID 3.....	36
RAID 4.....	36
RAID 5.....	36
RAID 6.....	37
RAID 7.....	37
RAID 0+1.....	37
RAID 10 (ou RAID1+0).....	37
RAID 50.....	38
Nombre d'interfaces vers les disques.....	38
Matériel ou Logiciel.....	38
Le RAID logiciel.....	38
Le RAID matériel.....	39
Utilisation de disques Hot Swappable.....	39
Le formatage.....	39
Le formatage bas niveau.....	39
Somme de contrôle.....	40
Formatage de haut niveau.....	40
Le partitionnement.....	41
Qu'est-ce qu'une partition?.....	41
Master Boot Record.....	42
Utiliser des partitions multiples.....	42
Partition principale.....	42
Partition étendue.....	42
Les systèmes de fichiers.....	43
Qu'est-ce qu'un système de fichiers.....	43
Système de fichiers et système d'exploitation.....	44
La cohabitation de plusieurs systèmes de fichiers.....	45
Ce que l'on peut trouver actuellement.....	45

Les disques les plus courants :	45
Les grandes capacités :	46
Les hautes performances :	46
Les signatures :	46
Tables de matières	47
Bibliographie :	49

Bibliographie :

<http://www-03.ibm.com/ibm/history/index.html>
<http://fr.wikipedia.org/>
<http://en.wikipedia.org/>
<http://www.answers.com/>
<http://www.computerhistory.org/>
<http://www.t13.org/>
<http://www.commentcamarche.net/>
<http://www.ldlc.be/>