

# LIVRE BLANC

L'état de l'art du stockage des données

## INTRODUCTION

En 2003, L'université de Berkeley a réalisé une étude qui prouve que le stockage de données est en expansion continue.

Aussi, l'un des grands défis de ce nouveau siècle sera celui de la conservation et de l'accès à tous types de contenus (vidéo, son, photo, écrit, plans,...) dont l'essence est essentiellement virtuelle. L'évolution incessante des technologies associées aux supports de stockage entraîne l'apparition de problématiques nouvelles.

Les supports tels que les CDs, DVDs, DONs (disques optiques-numériques) , Clés USB, mémoires flash, disques durs magnétiques, cartouches de sauvegardes ont des durées de vie largement inférieures aux fossiles, à la pierre , aux papyrus et plus près de nous au papier et nécessitent des conditions de conservations précises pour rester lisibles ou en marche.

La connexion de ces disques durs ou lecteurs peut aussi se révéler difficile à terme car l'architecture des machines évolue en même temps que leurs périphériques. Il faudra donc aussi conserver des ordinateurs et leurs OS datant de l'époque des lecteurs en état de fonctionner afin de garder la possibilité de transférer les données sur de nouveaux supports plus pérennes.

La sauvegarde sur disque accompagnée de la déduplication vont permettre aux opérateurs-ingénieurs d'optimiser les opérations de sauvegarde/ restauration et d'archivage et de mettre en route des PRA (plans de reprise d'activité) beaucoup plus rapides.

L'arrivée récente des disques flash (Solid State Disks) va changer la donne à l'horizon de 2 à 3 ans en proposant enfin un support plus pérenne , consommant peu et fortement capacitif avec accès on-line aux contenus.

La gestion de ces questions nécessite donc l'apparition d'une transcription moderne de cette mémoire qui pourra remplacer à terme les instruments et supports d'écriture transmis par nos ancêtres: c'est le défi du XXIème siècle.

## 1/ La problématique du stockage

### Quelques exemples de données à stocker :

photos, vidéos, images, sons, documents, textes, schémas, plans, MP3, ...

**Pièces comptables:** factures, commandes, courriers, administration,

**Gestion courante:** Création, fabrication, distribution: stocks, historique clients, emails, CAO, données de production, logistique; intranet/extranet

**Marketing:** description produits, photos, vidéos, schémas,

**Formats :** presse, radio, TV, internet, Mobile, 3G, etc...

### 1.1/ La sauvegarde-restauration

Il faut non-seulement copier les données mais aussi les retrouver. Le support doit être suffisamment pérenne et bénéficier d'un suivi logiciel pour continuer à être lu au cours du temps. On pourra recopier des données sur les mêmes supports sachant qu'une sauvegarde hebdomadaire ou mensuelle sera conservée (temps de rétention).

### 1.2/ L'archivage

De nombreux documents doivent être conservés par l'entreprise (de 5 à 30 ans) et une nouvelle norme permet de le faire numériquement : NF Z42-013.

Pour les sociétés industrielles, l'archivage historique est une nécessité pour les données sur les pièces détachées : Secteurs automobile, aéronautique, défense, marine,...

### 1.3/ Données en ligne :du mirroring au clustering

La haute disponibilité qui permet d'éviter l'arrêt de service aux données en ligne représente un coût à l'achat :

- doublement des données : RAID1 ou RAID 0+1
- Sécurisation par RAID 4, 5 ou 6
- doublement des alimentations, ventilations
- accès via plusieurs serveurs aux mêmes données: logiciels de clustering et de virtualisation

### 1.4/ Les différents degrés de disponibilité

#### 1.4.1/ Très Critiques : *Défense, banques, assurances, industrie, pharmacie, génétique, ...*

Miroir en temps réel de toutes les opérations sur un site de secours distant d'au moins 500Kms (tremblement de terre..)

#### 1.4.2/ Critiques : *tourisme, commerce, services, presse, TV, e-commerce,, VOD, plateformes de services, réseaux sociaux...*

Miroir distant d'environ 10-15Kms de toutes les opérations sur un site de secours

**1.4.3/ Haute disponibilité: PME-PMI**

Sécurisation sur un seul site : Raid 5 + sauvegardes régulières

**1.4.4/ Accès rapide aux données: on-line** : poste de travail

Données présentes sur disque dur, Mémoires Flash, clés USB

**1.4.5/ Accès moins rapide aux données : off-line ou sur étagère**

Cdrom, DVD, DON, ZIP, REV, Bandes DDS4, LTO, DLT, VXA, AIT, etc...

**1.5/ La hiérarchie des données**

- . **Critique** : accès vital : comptabilité, brevets, savoir-faire, e-commerce, presse en ligne, VOD, ...
- . **Stratégique** : accès rapide: bases de données produits, clients, stocks, fabrication, marketing ...
- . **Courante** : emails, workflow, bureautiques,...
- . **Assurance** : sauvegardes de ces données
- . **Gisement** : archivage historique, patrimoine, presse, Audiovisuel, Sons, photos, ...

**2/ Principaux supports et techniques****2.1/ La bande**

Elle permet de gérer une volumétrie qui peut aller au Pétaoctet ou plus avec un HSM et des robots adaptés

<b>DAT DDS4</b>	<b>20/40Go</b>	<b>6Mo/s</b>
<b>VXA-172</b>	<b>80/160Go</b>	<b>6 Mo/s</b>
<b>VXA-320</b>	<b>160/320Go</b>	<b>12 Mo/s</b>
<b>DLT</b>	<b>40/80Go</b>	<b>5 Mo/s</b>
<b>SDLT</b>	<b>110/220Go</b>	<b>11 Mo/s</b>
<b>AIT5</b>	<b>400Go/1040Gos</b>	<b>24 Mo/s</b>
<b>LTO</b>	<b>100/220Go</b>	<b>15Mo/s</b>
<b>LTO 2</b>	<b>200/400Go</b>	<b>30Mo/s</b>
<b>LTO-3</b>	<b>400/800Go</b>	<b>60Mo/s</b>
<b>LTO-4</b>	<b>800/1600Go</b>	<b>120Mo/s</b>
<b>LTO-5</b>	<b>1,5To/3To</b>	<b>300Mo/s</b>

Les **librairies** de bandes sont aujourd'hui à la portée de nombreuses organisations et restent un moyen privilégié de sauvegarder, archiver, accéder et restaurer les informations. Les capacités varient de quelques centaines de Go à plusieurs Pétaoctets (soit 1000 Tos)

**2.2/ Les disques durs**

**Accès aux données:** de 0,1 ms (SSD) à quelques ms

Taux de transferts : de 5 à 200Mo/s par disque

- Flash : 128 Mo,256Mo, 512Mo, 1Go, 2Go, 4Go,8Go, 16Go
- clés USB: 512Mo,1G, 2Go, 4Go, 8Go, 16Go
- ZIP : 100, 250 et 750 Mo
- REV: 35Go, 90Go, 120Go
- Disque dur ATA ou S-ATA : de 160Go à 1To
- Disque dur SAS: 73Go, 146Go, 300Go, 450Go,600Go
- Disque dur SCSI: 36Go, 73Go, 146Go, 300Go
- Disque dur Fibre Channel : 36Go à 500Go
- Disques durs SSD: 16Go, 32Go, 64Go, 128Go,256Go, 512Go...

**2.3/ Les graveurs CD-R, CD-RW, DVD-R, DVD-ROM, DVD-RW et RAM**

Ces systèmes permettent de graver de 4,7Go à 9,4Go sur un seul support très pratique au niveau de l'accès aux données. Les DVD-RAM et DVD-RW peuvent servir de sauvegardes et les DVD-R et DVD-ROM sont réservés aux archives.

Le support reste fragile : attention aux rayures et traces de doigt, températures, ...

## 2.4/ Blu-Ray

## 3/ Architectures de stockage

### 3.1/ Projet

Le système d'information a évolué d'une architecture centrée vers les serveurs dans les années 90 vers une architecture basée sur les réseaux de stockage à partir de 2000.

Une architecture NAS/SAN est aujourd'hui la norme des systèmes d'informations des grandes entreprises.

Les outils de SRM (Storage Resource Management) sont nécessaires au bon fonctionnement de ces réseaux de stockage.

Les stratégies antisinistres font aujourd'hui partie des schémas directeurs de ces architectures modernes:

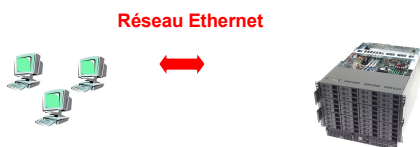
**CDP** (continuous data protection) enregistrement continu des données et de leurs modifications (par bloc ou par fichier)

**PCA** (plan de continuité d'activité) : plan regroupant toutes les actions, processus et modes d'organisations assurant la continuité des activités critiques pour la vie de l'entreprise

### 3.2/ Les technologies Raid

- . **Raid 0** vitesse : vidéo, calcul, HPC... ex: disques A et B.  
Les disques A et B sont parallélisés: capacités et vitesses sont additionnées mais l'arrêt de l'un des disques entraîne la perte de toutes les données.
- . **Raid 1** miroir : comptabilité, haute sécurité si miroir de Raid 5; ex: disques A et B  
Le disque A est la copie du disque B
- . **Raid 4** haute sécurité: ex: 5 disques A,B,C,D,E  
Les disques D=E contiennent, via un algorithme, les données des 3 disques A,B,C: chaque disque permet de retrouver toutes les données en cas de perte d'un ou de 2 disques.
- . **Raid 5** sécurité : ERP, Bases de données, CRM, Streaming...  
Les disques A, B, C contiennent, via un algorithme, les données des 3 disques sur 1/3 de leur espace: la parité unique est rotative et permet de retrouver toutes les données en cas de perte d'un disque.
- . **Raid 6** haute sécurité, archivage :ex: 5 disques A,B,C,D,E  
Les disques A, B, C,D,E contiennent, via un algorithme, les données des 5 disques sur 2/5 de leur espace: la parité est doublée ainsi que rotative et permet de retrouver toutes les données en cas de perte d'un ou de 2 disques.
- . **Raid 0+1 ou 10** :miroir et vitesse : calcul, R&D

### 3.3/ Le stockage NAS



**Vu comme un périphérique sur le Réseau ethernet:** On lui affecte une adresse IP, on l'intègre dans le domaine et le NAS est immédiatement vu comme une nouvelle ressource de stockage. Les NAS sont aussi appelés appliances de stockage. Ils sont raccordés via des switchs Ethernet. Leurs capacités varient de 250Go à 1Po. Il peut aussi fournir des blocs en mode Iscsi (SAN IP) ou se raccorder à un SAN.

#### . Usages:

- . Partage de fichiers CIFS, NFS, FTP, HTTP, Iscsi
- . Sécurité : raid + miroir via WAN
- . Travaille au niveau des fichiers :
- . Composé de disques durs en RAID 1,5,6,10,50,60 + DVD
- . Géré sous Windows, Unix, Linux.
- . Permet de gérer directement des copies miroir ou des snapshots (clichés du contenu des disques).
- Gestion espace par utilisateur, par projet. Possibilité d'allocation dynamique d'espace disque (Thin Provisioning : tous les utilisateurs disposent de 20Go mais le système ne réserve que l'espace utilisé réellement par chacun).

### 3.4/ La Fibre Channel

## . Introduction

Son développement a démarré en 1988.

La Fibre Channel est approuvée en 1994 et commence à se déployer en 1998.

La Fibre Channel est une solution fiable et mature pour les communications à haute vitesse des SANs déployés dans le monde.

La Fibre Channel est idéale pour les applications suivantes:

- Stockage haute performance
- Librairie de bandes haute performance
- Bases et entrepôts de données à forte volumétrie: Oracle, Sap
- Sauvegarde et restauration
- Clusters de serveurs, HPC (High power computing)...
- SAN
- Groupes de travail à haute performance
- Réseaux audio/vidéo : TV, radios
- Réseaux cinéma numérique.
- Web 2,0 : Juke-box, VOD et réseaux sociaux
- Gaming
- IP SANS

## Avantages de la Fibre Channel

- Jusqu'à 126 ports par boucle FC-AL sur 10Km
- Leadership rapport qualité/prix/performance
- Leadership des solutions : des centaines de Millions de ports installés.
- Bande passante 8 Gigabit aujourd'hui (1600Mo/s en full duplex)
- Nombreuses topologies : La Fibre Channel supporte les protocoles les plus utilisés SCSI, TCP/IP, des standards ouverts et de nouveaux produits pour les protocoles comme ESCON, VI et AV
- Evolutive : de lignes point à point à hautes vitesses à l'intégration de centaines de serveurs.
- Pas de congestion : Le contrôle intégré permet de vérifier si le buffer de destination est capable de recevoir le flux de données.
- Haute efficacité : la simplicité de son fonctionnement garantit le maximum de performances.
- des services complets : adressage, Lun zoning, externalisation (ISSP), sécurité (remote site).
- Vitesse
- Transfert

## .Technologie d'interconnexion

La Fibre Channel peut être vue comme un mécanisme de transport très performant qui se présente comme une boucle où les colis se déplacent dans les 2 sens (choix du sens le plus rapide) vers leur port de destination..

Elle organise par exemple les commandes Scsi en paquets, les transporte aux unités appropriés , et fournit un contrôle d'erreur pour s'assurer que le paquet est bien arrivé.

3 topologies sont possibles: point à point , partagée (FC-AL) ou switchée (chaque machine connectée au switch dispose de 4 ou 8Gbit/s par point d'entrée .

La Fibre Channel est une topologie indépendante qui est aujourd'hui un standard pour les réseaux de stockage et le transfert de données.

La Fibre Channel établit des liaisons point à point, partagées ou switchées entre les unités, en fournissant la bande passante complète (4 ou 8 Giga-bits) à chaque connexion. C'était la plus rapide des interfaces mise en service avant l'apparition de l'infiniband : 20 ou 40Gbit/s

La Fibre Channel est une technologie fonctionnant en série qui se comporte comme une liaison téléphonique.

## Drivers et utilitaires

### Entités logiques:

- les ports FC (ou nœuds) se voient attribuer une adresse unique qui doit être géré.

### Initialisation et contrôle

- Un switch va lui-même scruter les ports occupés ou non et pourra lancer une alerte s'il découvre un problème.

### Zoning:

Le SAN gérant des données sur de nombreux serveurs et OS, plusieurs types de trafic et de données sont envoyés à travers la Fabric. Pour assurer la sécurité des données, l'administrateur va créer des zones qui vont restreindre l'accès à des domaines spécifiés. Le Zoning va permettre de générer des environnements hétérogènes et permet à la Fabric d'être dynamiquement configurée pour des applications particulières.

**Routage interswitch** FSPF ( Fabric shortest Path First) est le protocole de routage qui permet la cascade de switches provenant de vendeurs multiples.

**Upper Layer Protocol (ULP) mapping**

La Fibre Channel fournit une couche de transport pour les ULPs qui contiennent les applications et l'interface utilisateur. Standards ANSI et IETF (Internet Engineering task Force).

Virtuellement, toute technologie peut être cartographiée par la Fibre Channel.

**. Fonctionnement de la Fabric ou switch**

Le switch ou Fabric fournit une solution de connectivité complète, à bas coût, similaire au système téléphonique.

Quand un appelant décroche son téléphone et entre un numéro, le système téléphonique dirige l'appel ,et fait toutes les connexions intermédiaires requises pour faire sonner le numéro composé.

Si le téléphone a répondu, le routage bien exécuté est confirmé à l'appelant.

Si le switch échoue, l'appel est re-routé via d'autres circuits.

La Fibre Channel établit comme le système téléphonique des connexions temporaires, dédiés , à bande passante complète entre les unités. Elle utilise des adresses uniques pour connecter les unités (processeurs, périphériques) à des distances pouvant aller jusqu'à 10Km et plus (Via des réseaux IP)...

Sites à visiter pour plus d'informations : [www.snia.org](http://www.snia.org)

[www.fibrechannel.com](http://www.fibrechannel.com)

**3.5/ Les réseaux de stockage SAN**

**Définition:**

Le SAN est un réseau de stockage ouvert et évolutif qui relie à des périphériques de stockage et d'archivage des serveurs/stations et postes de travail , par ailleurs reliés au réseau d'entreprise.

Le SAN constitue une plate forme de communication qui exploite les protocoles SCSI sur des technologies d'interconnexion à haut débit (Fibre Channel).

Le stockage sécurisé et centralisé permet le partage des ressources entre les serveurs/stations connectés et les données sont accessibles à haute vitesse par un grand nombre d'utilisateurs.

**Avantages:**

Garantie de débit élevé : jusqu'à 8 Giga-bits par port :possibilité de les additionner.

Distance élevée : 10Kms (SCSI 25M maximum)

Pérennité de l'investissement : ajout de capacité et de puissance machine aisé

Centralisation, consolidation et sécurisation du stockage, de la sauvegarde et de l'archivage.

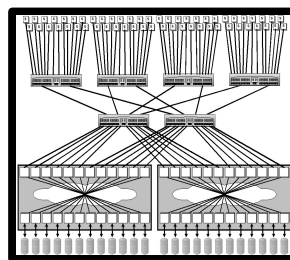
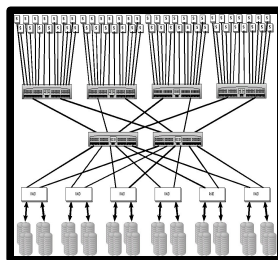
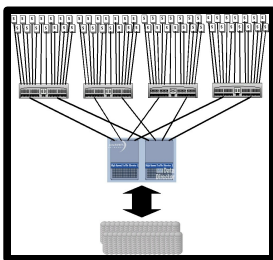
Complément indispensable au LAN pour les réseaux d'entreprise.

Interopérabilité multi-plateformes et multi-OS.

**3.5.1/ La connectique**

**. Topologie switchée / Appliance**

Appliance	Contrôleur RAID	Contrôleurs RAID switchés
Datadirect Networks	EMC2 Clariion , Engenio	EMC2 Symmetrix, Hitachi



Le réseau SAN peut être configuré pour de nombreuses applications: ERP, CRM, Sciences, Sauvegarde, virtualisation,...

Avantages : performances, souplesse, sécurité.

Inconvénient : le coût de possession élevé

#### . HBA



Un HBA (host bus adapter) est une carte s'intégrant à l'intérieur des serveurs et permettant d'avoir un lien de type Fibre Channel avec l'extérieur : switch, Disque,...

Son installation dépend de l'OS: 2000/2003/unix/linux/Novell, le type de bus pour le serveur: PCIexpress, ...

- le type de connexion physique :

.Optique mono-mode(10Km)

.Optique multi-mode (500m)

.GBIC

Fabricants : Qlogic, Adaptec, Atto, Emulex,...

#### . Bridge / router (pont / routeur)

Le bridge ou router permet à l'utilisateur de conserver les investissements effectués dans les périphériques Scsi. Il permet de convertir les connexions existantes de type Scsi en Fibre Channel et vice versa.

##### **Caractéristiques :**

Nombre de canaux Scsi (ex: 1,2,4)

Nombre de canaux Fibre Channel

**Fabricants:** Crossroads, Qlogic, Brocade

#### . Switch



Le switch est utilisé comme point de central de connexion d'une architecture Fabric.

##### **Caractéristiques:**

Nombre de ports : de 8 à 224 ,Présence de connecteurs fixes ou GBIC, vitesse de commutation

**Avantages :** chaque lien dispose de la bande passante complète, évolutivité.

**Inconvénients :** coût , problèmes de concentration.

**Marques:** Brocade, Qlogic, Cisco, Emulex.

**Tarifs :** à partir de 2000 euros les 8 ports.

#### .GBIC



Le GBIC est un adaptateur s'enfichant sur les cartes, ou switches Fibre Channel , qui donne la connexion Fibre Channel vers l'extérieur.

Le changement d'un GBIC s'effectue à chaud et permet à l'utilisateur de migrer d'une connexion cuivre ou multi-mode vers une connexion mono-mode et vice-versa.

La durée de vie d'un GBIC est de 2 ans!

**Tarifs :** GBIC Fibre Optique/ 300 euros

#### . Disques intégrés dans les JBOD ou les tours Raid 0,1,5

**Caractéristiques** des disques internes intégrés dans des boîtiers amovibles et qui peuvent être ventilés collectivement ou individuellement:

IDE, S-ATA, SCSI, SAS, SSD, Fibre Channel

Ces systèmes sont aujourd'hui standardisés dans les entrepôts de données et proviennent de quelques sources bien identifiées (Seagate, Hitachi, Western Digital, ...)

### 3.5.2/ Les baies de stockage

#### A/ Les JBOD



Ce sont des chassis bien sécurisés (Alimentations et ventilations redondantes, disques échangeables à chaud de disques destinées à être connectés à:

- des cartes Raid de type PCI
- des contrôleurs RAID 0,1,5,6,10,50,60 Fibre Channel ou S-ATA
- des serveurs qui piloteront leur « striping » (gestion des grappes de disques) via un logiciel.

La haute densité de stockage Par U permet de réaliser des économies importantes et les disques peuvent aujourd'hui être disposés dans le chassis et non en face avant (ex: 60 disques sur 4U chez Datadirect Networks)

#### B/ Les chassis de disques RAID 0,1,5,6,10,0,60

Ces chassis disposent des mêmes caractéristiques que les JBOD et intègrent directement un ou deux contrôleur(s) RAID 0,1,5,6,10,50,60.

L'avantage est de pouvoir les connecter directement sur la carte Fibre Channel HBA d'un serveur de secours sans souci de configuration.

Leurs capacités sont limitées par le nombre de disques gérés par un bus SCSI (14 éléments) , par un bus S-ATA jusqu'à 12 disques) , sur une boucle FC AL (126 disques) ou sur SAS (128 disques max.)

**Constructeurs:** Promise, Engenio, Datadirect Networks, Xyratex ...

### 3.5.3/ Les outils d'administrations Fibre Channel

#### .IBM: Tivoli SAN Manager

L'outil le plus répandu car il correspond exactement aux besoins des entreprises:  
Hautes performances, sécurité, contrôleur d'accès

#### . Quantum : Stornext

#### . Tiger Technology : Metasan

### 3.5.4/ ISCSI

Ce protocole au niveau bloc permet l'encapsulation de données SCSI sur IP et ainsi de réaliser des réseaux de stockage à petits prix sur réseaux ethernet.

L'avantage est de passer des données bureautiques et des données de type bloc sur le même tuyau.

La norme est suivie par l'IETF.

Fabricants de cartes Iscsi norme TOE : Adaptec, Qlogic...

### 3.5.5/ la virtualisation

La virtualisation entend unifier les ressources de stockage hétérogènes en réseau.

Elle comporte 2 familles de solutions:

#### 1/ Symétrique (in-band)

Point de passage obligé par le serveur de virtualisation sur qui on alloue les espaces physiques de baies du réseau. On réserve des luns correspondant aux partitions créées sur les baies du réseau puis on crée des volumes logiques à partir de ces Luns. Ils sont alors vus comme des volumes physiques constitués de disques internes par le serveur applicatif.

La virtualisation peut aussi être un goulot d'étranglement : tous les flux passent par le serveur de virtualisation.

Ce serveur doit être redondant pour assurer la disponibilité des données...

#### 2/ Asymétrique (Out-band).

Le serveur de virtualisation ne sert qu'à définir des volumes logiques.

Chaque serveur applicatif accède directement aux ressources physiques via son agent spécifique, qui possède – grâce au serveur de virtualisation – la correspondance entre vue logique et vue physique.

En cas de panne, le serveur applicatif accède toujours à son stockage, seule la redéfinition des volumes devient impossible.

### 3/ Autres architectures de virtualisation:

**S2A Silicon Storage Appliance** :Le **SAN S<sup>2</sup>A** permet un accès simplifié et beaucoup plus rapide car intelligent et parallèle : un moteur de gestion de flux analyse les requêtes et les achemine en temps réel vers 12-20 voies fibre/fibre disponibles. Le **SAN S<sup>2</sup>A** virtualise et consolide sur une ou plusieurs milliers d'unités logiques (LUNs) d'importants pools d'information agrégée sur plusieurs Tos.

**Fabricants**: Datadirect Networks

**Virtualisation NAS ou FAN (File Area Network)** : serveurs de fichiers puisant leurs ressources de stockage sur le SAN.

**Fabricants**: HP avec Polyserve, Pillar Data Systems, Equallogic, Intellique...

**Cluster de NAS** équipés de disques SAS ou Sata: le raid est géré par un file system global .

La reconstruction des données est extrêmement rapide et constitue l'avantage de ces solutions.

**Fabricants**: Isilon, 3par, Pillar data systems, Intellique, ...

#### 3.5.6/ Réplication asynchrone

La réplication asynchrone correspond à une copie exacte d'un serveur A vers un serveur B qui se déroule à heure fixe de manière décalée avec une fenêtre de réalisation précise.

**Exemple**: la commande Rsync sur linux.

#### 3.5.7/ Réplication synchrone ou temps réel

La réplication à distance est indispensable aux Plans de reprise d'activité

Les 2 serveurs sont identiques et l'un peut remplacer l'autre sans aucune interruption de service. On peut alors réparer ou échanger le serveur défaillant et reconstruire le miroir sur le nouveau serveur en ligne.

**Editeurs**: Double Take, Netvault replicator, EMC2, Netapp, Hitachi, ...

## 4/ La sauvegarde et l'archivage

### 4.1/ La sauvegarde des données

#### 4.1.1/ Stratégies

Si la volumétrie est élevée , Réaliser un miroir par réplication ou des snapshots est indispensable.

Le coût d'achat est composé des éléments suivants : serveur + logiciel + lecteur ou robot de sauvegarde + cartouches

Coût de possession est constitué des postes suivants: maintenance, Mise à jour, administration, régie,.

La connexion de nouveaux serveurs ou utilisateurs peut entraîner l'achat d'agents supplémentaires.

Obtenir un RSI rapide (Retour sur investissement)

Réduire les Fenêtres de sauvegarde

Réduire le temps de Restauration

#### 4.1.2/ Différents modes de sauvegarde

Au niveau de l'entreprise, les opérations journalières, les email, les bases de données, les fichiers ERP et CRM, Exchange, Lotus Domino et notes,... Tous ces fichiers nécessitent une sauvegarde journalière sans erreur.

Au niveau de l'utilisateur, les fichiers à sauvegarder sont ceux qu'il modifie de façon régulière et dont la perte n'est pas envisageable.

Le logiciel de sauvegarde est une assurance-vie sur les données.

#### A/ Sauvegarde générale :

Tous les fichiers du ou des disques réseau seront sauvegardés sur cartouche(s).

Si la volumétrie le permet, cette opération sera journalière.

Pour ce faire, elle nécessitera au maximum 12H.

Sinon, elle se fera tous les week-ends : jusqu'à 72H.

**B/ Sauvegarde incrémentale :**

Le logiciel de Back-up sélectionne les fichiers à sauvegarder suivant les critères définis dans les règles de gestion édités par l'administrateur :

- Fichiers présents sur les ressources disques spécifiées
- Fichiers modifiés suivant la dernière sauvegarde
- Délai minimum entre 2 sauvegardes d'un même fichier
- Elle se déroule du lundi au jeudi : prévoir suffisamment de cartouches

**C/ Sauvegarde différentielle (à la demande) :**

Le logiciel de back-up sélectionne les fichiers à sauvegarder suivant les critères définis dans les règles de gestion édités par l'administrateur :

- Fichiers présents sur les ressources disques spécifiées
- Fichiers modifiés suivant la date indiquée par l'administrateur
- Répétition

**D/ Sauvegarde Sélective :**

L'utilisateur sélectionne les fichiers à sauvegarder en spécifiant suivant ses droits, un disque, un système de fichiers, un répertoire, un sous- répertoire , un ou des fichiers contenus dans ces derniers.

Cette fonction tend à apparaître avec l'apparition d'interface Web sur ce type de logiciel et la volonté de gérer les données sur portable.

**E/ Période de rétention**

L'administrateur va définir une période durant laquelle les fichiers vont être conservés sur les cartouches de sauvegarde. A la fin de celle-ci, les cartouches seront proposés à l'effacement et à la réutilisation pour de nouvelles sauvegardes.

Il faut savoir qu'une bande LTO coûte environ 75 euros et dispose d'une durée de vie limitée (100.000 heures).

Aussi faut-il prévoir de les changer afin d'assurer la relecture des sauvegardes

**F/ La restauration**

Elle peut être simple dans le cas d'une récupération de fichiers perdus dans la journée mais aussi complexe :

- récupération de données Exchange
- récupération d'index d'une base de données de type Oracle

Aussi, la tendance est aujourd'hui d'utiliser des sauvegardes sur disque dur du type NAS ou DAS du fait de leur faible coût et d'un accès rapide aux données perdues. On appelle cela du stockage Near-line.

De plus, les NAS disposent de technologies de miroir ou de snapshot (clichés) intégrés qui permettent de récupérer instantanément des données complexes et de rétablir le service interrompu en un temps record.

En cas de virus, ils peuvent constituer une parade plus efficace que les sauvegardes sur bande.

**G/ La sauvegarde sur disque et la déduplication**

En utilisant des grappes de disques sécurisées RAID5 ou 6 comme des bibliothèques de bandes virtuelles , les organisations peuvent réduire fortement leurs fenêtres de sauvegarde et de restauration.

Elles peuvent alors écrire et surtout restaurer des données beaucoup plus rapidement qu'avec des bandes. Toutes les opérations de sauvegarde et restauration peuvent être recrées à l'identique: complète, incrémentale ou différentielle. En effet, des outils spécialisés vont créer virtuellement sur des espaces disques NAS et/ou SAN des bibliothèques complètes (lecteurs et cartouches de votre choix, nombre de slots à définir) qui vont permettre à l'utilisateur de s'affranchir des limitations imposées par les fenêtres de sauvegarde. L'idée forte est d'utiliser un volume disque suffisant pour éviter l'acquisition d'une bibliothèque à grand nombre de slots et préférer le choix d'un autoloader à format 1 ou 2U à coût réduit pour les entrées-sorties de cartouches à des fonctions de délocalisation ou d'archivage des données . Les logiciels de déduplication souvent pré-installés permettent de gagner entre 30 et 80% du volume disque utilisé en optimisant l'occupation de l'espace disque dédié à la sauvegarde.

Il ne fait aujourd'hui aucun doute que les technologies bandes utilisées uniquement pour la sauvegarde sont en fin de vie car les technologies disques de type VTL (Virtual Tape Library) plus sûres et plus simples les remplacent avantageusement, mais leur avenir reste assuré pour les délocalisations (PRA) et l'archivage pour le moment . Les SSD devraient les remplacer à terme (5 à 10 ans...)

**Editeurs** : Falconstor, CA, Bakbone, Veritas, Legato, Tivoli, Synerway, Déduplication Diligent, ...

**Constructeurs**: Overland, EMC2, Netapp, IBM, Hitachi, Datadirect Networks, Quantum,...

## 4.2/ L'archivage

### 4.2.1/ Définition de l'archivage

Une archive ne doit jamais être effacée. Elle doit être indexée, localisée et pouvoir être retrouvée.

Les supports d'archivage les plus couramment utilisés sont des cartouches non réinscriptibles.: DON, Bandes LTO, UDO, disques durs MAID (en veille si non utilisés) en RAID6, ....

Les fichiers archivés peuvent être accessibles suivant les options suivantes : ON-LINE :VTL, robots et juke-boxes

OFF-LINE : classés sur étagères

### 4.2.2/ Restauration

La restauration de fichiers sauvegardés ou archivés s'effectue à la demande de l'utilisateur ou l'administrateur quand cela est nécessaire après un désastre quelconque ou quand il faut aller consulter les archives.

Cette fonction est accessible par différentes voies et s'effectue en quelques clics :

- Restauration des sauvegardes et archives déjà réalisées
- Le logiciel peut disposer d'un moteur de recherche évolué permettant de retrouver un fichier par rapport à un archivage normé.

## 4.3/ La norme AFNOR NFZ42-013

Cette norme permet aux entreprises de donner des documents sur support numérique aux autorités administratives:

- Gain de place et Accès rapide

## 4.4/ Les logiciels de gestion de hiérarchie de stockage

Le principe des HSM est l'extension virtuelle de l'espace de stockage disque dur sur un média de stockage secondaire du type bibliothèque de bandes ou near-line (disques en veille si non utilisés).

Les fichiers sont copiés sur ce média depuis les disques , d'où ils peuvent être purgés. L'utilisateur croit qu'ils sont toujours sur les disques car l'accès y est permis.L'espace disque est un cache pour les données contenus sur le stockage secondaire.

Leur intérêt est de réduire le coût de possession des données :

- migrations automatique du disque vers le stockage secondaire
- Purges possibles vers du média Off-line
- Mécanismes de corbeille
- Talon et méta-données attachés aux données
- Volumétrie très importante
- Transparence pour l'utilisateur au niveau des accès mais attention aux vitesses d'accès aux données...

**Editeurs** : IBM Tivoli Storage Manager, EMC diskxtender, SGI DMF,...

## 4.3/ La Gestion du cycle de vie des données (ILM Information lifecycle Management)

La gestion du cycle de vie des données existe depuis bien longtemps mais le coût du stockage informatique ayant baissé drastiquement ces 5 dernières années, le problème se pose aujourd'hui de manière différente quand on sait que la croissance de la volumétrie des données est exponentielle.

Il pourrait sembler évident de créer un code génétique unique et identificateur pour toutes les informations générées mais personne ne sait par où commencer pour faire ce que la Nature génère depuis la nuit des temps.

Chaque organisation dispose d'un circuit d'information unique et souvent opaque afin en limiter l'accès pour des raisons de sécurité ou de confidentialité.

Le partage de données semble donc utopique pour nombre d'administrations et d'entreprises qui préfèrent toutes se réserver un accès privilégié aux informations de leurs îlots digitaux.

Comme en mathématiques, le raisonnement par l'absurde peut permettre de trouver une réponse:

Tout d'abord il faudrait une volonté de créer un format universel et libre de document (écrit, image, audio, vidéo,...) qui serait accompagné d'un code génétique évolutif (ADN) composé d'un nombre aléatoire et unique (identification, certification et signature) et d'éléments de description simples dont certains pourraient varier au cours du temps: *Raison sociale, date de création, utilisateur, nom du fichier, projet, version, durée de vie légale s'il y a lieu, criticité (confidentiel, aspect légal, workflow, assurance ou copie, gisement ou archivage),...*

L'intérêt de ce type de marquage est de pouvoir tracer la vie et les localisations de tout document et de pouvoir l'archiver ou le détruire dès qu'on en a plus l'usage.

Formats normalisés: XML, ODF pour OASIS Open Document Format for Office Applications, PDF (adobe), MXF (video), MP3 (son), ...

**Editeurs:** Documentum, Alfresco, OpenCube, Princetonsofttech, ...

## 5/ Le Marché

Le marché du stockage pèse aujourd'hui 25 Milliards d'euros dont 20% est dédié aux logiciels de gestion (croissance 25% /an).

Le standard S-ATA a permis de baisser drastiquement les coûts d'achat de systèmes de stockage et entraîner l'apparition de nouveaux besoins :

- Librairies virtuelles de bandes sur disques durs (VTL Virtual Tape Library)
- Vidéo à la demande ou hébergement de vidéos, jeux, series TV, Films, ...
- Vidéosurveillance numérique continue
- Réseaux sociaux

**Les leaders mondiaux :** EMC2, HP, Netapp, Hitachi, IBM, Sun, Dell, Datadirect Networks, ...

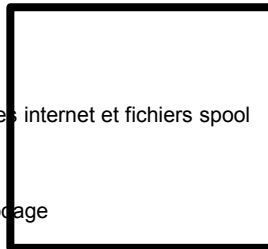
**Les acteurs émergents :** Isilon, 3Par, Exanet, Rackable, Pillar Data Systems, Blue Arc, ..

**Les acteurs grand public:** Iomega, Seagate, Western Digital,...

## 6/ Les applications

### 6.1/ La lecture automatique de documents

- . papier, fax, fichiers électroniques, formulaires internet et fichiers spool
- . identification et tri p/r base de modèles
- . Gestion automatique des lots
- . Gestion des courriers
- . Traitement des images OCR-ICR, vidéo-codage
- . GED et SGBD documents XML
- . Banques, assurances, administrations, commerce



### 6.2/ La Gestion des Connaissances (knowledge-management)

- . **Portail web (Wikipedia,...)**  
Intranet, extranet
- . **moteurs de recherche**  
structuré, flou, plein texte, images, vidéo
- . **applications**  
accueil clientèle  
publication dynamique  
publicité à la demande  
communication interne  
E commerce

### 6.3/ ERP: Back-Office

- . Structuration de l'entreprise  
Gestion de la production  
Gestion des stocks  
Gestion des tâches

- . Elimination du middle-office
  - Entrée des données immédiate
  - Traitement journalier
  - Extraction
  - Tableaux de bord

#### **6.4/ CRM: Front-Office**

- . Structuration des contacts clients
  - Accueil téléphonique, Web, courrier
  - Réponse personnalisée
  - Envoi mailings
- . Elimination du middle-office
  - Automatisation des tâches de relance
  - . Tél, Web, Centre d'appels, intranet

#### **6.5/ Places de marché**

- . Structuration des achats
  - Mise en place par métier: pétrole, chimie.
  - Mise en relation des vendeurs et des acheteurs
  - Possibilité d'enchères
- . Elimination du middle-office
  - Moins de papiers
- . Groupements: alimentaire, automobile

#### **6.6/ e-commerce B to C**

- . Catalogue en ligne
- . Possibilité de rappel téléphonique
- . Logistique : suivi des livraisons
- . facturation
- . Suivi des comportements d'achat : Amazon vous propose des CD ou livres correspondant à vos goûts.

#### **6.7/ Décisionnel**

- . Datawarehouse
- . Bases de données marketing
- . Logiciel de type Pertinence : règles établies à partir de données aléatoires
- . Réduire le temps des entreprises

#### **6.8/ PLM Product Life Management**

De la création à la fabrication  
Industriels :Renault, Peugeot, EADS, Thales, ...  
Bureaux d'étude  
Sous-traitants  
Partage des informations multi-entreprises  
Suivi des projets par groupe de travail  
Workflow

#### **6.9/ Santé : Opérations à distance ,Imagerie numérique**

- Scanners des hôpitaux
- Cabinets de radiologie
- Echographies
- Numérisation des films archivés
- Accès aux informations rapide et partage aisé.

---

**6.10/ WEB 2,0**

- réseaux sociaux: facebook, myspace, linkedin, viadeo,...
- Hébergement vidéo : Youtube, dailymotion,...
- Hébergement photos: Flickr, kodak, ...
- Hébergement jeux vidéo on-line: metaboli, ...
- Juke box MP3:
- VOD : Free, Canal+, TF1, Itunes,...

**6.11/ La vidéosurveillance**

- lieux publics : Rues, Métro, Bus, ...
- Propriétés privées
- Entreprises
- tribunaux
- commissariat
- sécurité

**CONCLUSION**

Même si le volume de données créées chaque année semble gigantesque, il ne rassemble qu'une infime partie de la mémoire des hommes présents sur la terre: Les neurones présents dans chacun de nos cerveaux pouvant stocker environ 11 péta-octets , l'ensemble de l'humanité disposerait d'un potentiel de 1,75 millions fois supérieur à l'ensemble des données numériques créées en 2003 !

De plus, Le volume d'un cerveau humain n'a rien à voir avec les 40 mètres cube nécessaires pour stocker aujourd'hui 11 Péta-octets sur 11000 disques durs de 1To (17 racks 44U de 660To (60Tos/4U chez Datadirect Networks).

La mémoire neuronale est donc largement supérieure aux meilleures technologies de stockage sur le marché mais bénéficie de millions d'années d'évolutions. Or, ce bon vieux disque dur n'a que 46 ans et a démarré avec 5Mo disposé sur 50 plateaux de 61cm de 100Ko chacun...

Il est donc normal que nous ayons encore un long chemin à parcourir.

Le vieil adage « en Afrique, chaque vieillard qui meurt est une bibliothèque qui brûle » est vrai aujourd'hui pour tout être humain.

Mais il est possible ,qu'un jour , notre mémoire ne meure jamais et que nos connaissances soient enfin préservées de l'oubli ,de la destruction ou de l'érosion du temps.