

# LIVRE BLANC

L'essentiel du stockage et de l'archivage  
des données

## INTRODUCTION

En 2003, L'université de Berkeley a réalisé une étude qui prouve que le stockage de données est en expansion continue, soit la même règle que l'univers.

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, Blu-Ray, Clés USB, mémoires flash, disques durs magnétiques, disques SSD, bandes LTO, ... ont des durées de vie largement inférieures aux fossiles, à la pierre, à l'argile, aux papyrus et plus près de nous au papier et nécessitent des conditions de conservation 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 va permettre aux ingénieurs d'optimiser les opérations de sauvegarde/restauration et d'archivage actif et de mettre en route des PRA (plans de reprise d'activité) beaucoup plus rapides.

L'arrivée récente des disques flash (Solid State Disks) change la donne en proposant enfin un support plus pérenne, consommant peu et fortement capacitif avec accès on-line aux contenus.

La gestion de ces questions a nécessité 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 et sa réponse est « le cloud ».

## 1/ La problématique du stockage

La période 2009-2012 a vu l'éclosion de nouveaux usages de la donnée dont les volumes non structurés ont augmenté exponentiellement ; ce qui oblige toutes les organisations à revoir leur infrastructure informatique et leur gestion.

### 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, plans,...

**Formats :** presse, radio, TV, internet, 3G, 4G, 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, ... L'archivage peut être passif ou actif.

Il existe aujourd'hui des outils permettant de retrouver des documents et de les remettre au goût du jour.

Ainsi, de nombreux films ont pu être restaurés et nous les visionnons aujourd'hui mieux que nos grand-parents...

### 1.3/ Données en ligne : du mirroring au clustering, puis au stockage objet

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 6 (le Raid 5 est aujourd'hui obsolète dès qu'il faut protéger plus de 10To nets.
- doublement des alimentations, ventilations
- accès via plusieurs serveurs aux mêmes données: logiciels de clustering et de virtualisation

L'augmentation très rapide du volume des données a permis l'apparition d'un nouveau type de stockage : le stockage objet (illimité) à mettre en opposition au stockage de fichiers (relativement limité), qui s'installe sur des matériels de toute marque.  
Le triplement automatique des données dans 3 silos de stockage distants va permettre de pouvoir récupérer immédiatement des données perdues sans aucune reconstruction, par simple recopie de ces données via un algorithme..  
La rentabilité de ce type de stockage démarre aujourd'hui vers les 100To nets mais augmente rapidement en cas de fortes capacités à stocker.

## 1.4/ Les différents degrés de disponibilité

### 1.4.1/ Très Critiques : Défense, banques, assurances, industrie, pharmacie, génétique, e-commerce, ...

Miroir ou triplement 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 ou triplement distant d'environ 10-15Kms de toutes les opérations sur un ou 2 site(s) de secours

### 1.4.3/ Haute disponibilité: PME-PMI

Sécurisation sur un seul site : Raid 6 + sauvegardes régulières sur bandes ou disques

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

Données présentes sur disque dur ou SSD , sauvegardes sur disques durs, mémoires Flash ou clés USB rapides

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

DVD, DON, Blu-ray, Bandes DDS4, LTO, ...

## 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 jusqu'à des centaines de Po avec un HSM et des robots adaptés

|        |      |             |          |
|--------|------|-------------|----------|
| LTO    | 2000 | 100/220Go   | 20Mo/s   |
| LTO 2  | 2002 | 200/400Go   | 40Mo/s   |
| LTO-3  | 2004 | 400/800Go   | 80Mo/s   |
| LTO-4  | 2006 | 800/1600Go  | 120Mo/s  |
| LTO-5  | 2010 | 1,5To/3To   | 140Mo/s  |
| LTO-6  | 2012 | 3,2To/8To   | 160Mo/s  |
| LTO-7  | 2015 | 6,4To/15To  | 300Mo/s  |
| LTO-8  |      | 12To/30To   | 472Mo/s  |
| LTO-9  |      | 25To/62,5To | 708Mo/s  |
| LTO-10 |      | 50To/120To  | 1100Mo/s |

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 12To à des centaines de Pétaoctets (soit 1000 Tos)

**Constructeurs** : Fujifilm, IBM, HPE, Quantum, Overland, Maxell, ...

## 2.2/ Les disques durs

**Accès aux données:** de 0,1 ms (SSD) à quelques ms  
Taux de transferts : de 5 à 400Mo/s par disque

- Flash : 16Go, 32Go, 64Go, 128Go, 256Go
- clés USB: 8Go, 16Go, 32Go, 64Go, 128Go, 256Go
- Disque dur S-ATA : de 1To à 10To , apparition des disques Hélium chez HGST plus fiables, consomment 2 fois moins.
- Disque dur SAS 15KT : 146Go, 300Go, 450Go, 600Go, 900Go
- Disque dur SAS 7200Tr/min : jusqu'à 10To

- Disques durs SSD: 128Go, 256Go, 512Go, 1To, 2To, 3To, 4To, 16To... accès très rapide, plus milliers d'E/S par seconde, consommation très réduite , pas de bruit.

## 2.3/ Les graveurs 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/ les graveurs-lecteurs Blu-Ray

Les Blu-ray sont plus sûrs , codage MPEG2, capacité 25Go simple couche et 50Go double couche (128Go 4 couches) et sont appuyés par les studios et les grands constructeurs.

## 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 a longtemps été la norme des systèmes d'informations des grandes entreprises mais de nouveaux acteurs full-flash sont apparus et ont proposé de nouvelles architectures adaptées aux VM (virtual machine) et aux clouds avec des architectures évolutives basées sur une brique serveur qu'il suffit d'acheter en nombre suffisant pour ses besoins.

De nouveaux outils comme les SDS (Software defined storage) et l'hyperconvergence (Virtual San) sont apparus et peuvent s'installer sur tout matériel et changent la donne des constructeurs en leur donnant une souplesse supplémentaire .

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

**PRA** (Plan de reprise d'activité) : plan regroupant toutes les actions, processus et modes d'organisations assurant la reprise 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 5 sécurité : ERP, Bases de données, CRM, Streaming...**

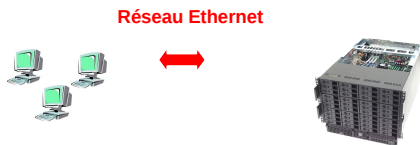
Les disques A, B, C contiennent , via un algorithme, les données des 3 disques sur 1/3 de leur espace: la parité unique est rotative et permet de retrouver toutes les données en cas de perte d'un disque.

. **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
  - . Géré sous Windows, 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 100Go 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 (Iscsi)
- Réalité virtuelle
- Réalité augmentée

#### 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 (16 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 du switch

Le switch 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 ou ethernet (SAN iscsi)).

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 |
|-------------|------------------------------|---------------------------|
| DDN Storage | Netapp, EMC2 , Infotrend,... | 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: Microsoft ou linux , 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,...

##### . 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** : concurrencés par la technologie Ethernet et Infiniband plus performantes aujourd'hui.

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

\* **.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!

■ **Disques intégrés dans les JBOD ou les tours Raid 0,1,6,10**

**Caractéristiques** des disques internes intégrés dans des boîtiers amovibles et qui peuvent être ventilés collectivement ou individuellement:  
S-ATA, SAS, SSD

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 HGST, 84 disques sur 4U chez DDN)

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

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

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 présents sur chaque baie.

**Constructeurs:** Netapp, EMC2, Infortrend, Seagate, DDN Storage, Promise...

### 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

Efficace, nécessite l'aide d'experts pour sa mise en oeuvre.

. Tiger Technology : Metasan

économique; répandu dans le monde de la post-production et des Tvs sous forme d'appliance

. Scale Logiciel : Hyper FS

Simple et complet ; le plus souple du marché en termes de fonctionnalités



### 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. Ses performances en 10Gbit se révèlent excellentes. Une bonne alternative à la fibre Channel.

L'avantage est de passer des données bureautiques et des données de type bloc (images, vidéos,...) sur le même tuyau.

La norme est suivie par l'IETF.

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

**Fabricants de cartes Ethernet utilisables en SAN IP** : Myricom, emulex, Brocade, ...

### 3.5.5/ la virtualisation

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

#### 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. Cette architecture est leader dans le HPC (High Power computing)

**Fabricant**: DDN Storage

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

**Fabricants**: EMC2, Netapp, Oracle, Dell,...

**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** : EMC2-Isilon, HPE 3Par, DDN Storage Exascaler, ...

**Virtualisation de SAN ou Hyperconvergence** : les serveurs équipés de simples disques SAS ou SSD (Briques) sont transformés en SAN virtuel via des Sds intégrés dans des solutions de virtualisation de serveurs (San virtuel en mode bloc, réplication synchrone, snapshots...)

**Editeurs** : Vmware – virtual SAN, Datacore SANSymphony, Simplivity Omnicube Nutanix, ...

### 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, Datacore, EMC2, Netapp, Hitachi, DRBD, ...

## 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, ... 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 complète :**

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 dispose d'une durée de vie limitée (100.000 heures d'usage).

Aussi faut-il prévoir de les changer périodiquement 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. C'est le stockage Near-line qui peut être aussi utilisé pour l'archivage actif.

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 RAID 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 recréées 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 pourraient les remplacer à terme (3 à 5 ans...)

**Editeurs :** EMC Data domain, IBM Diligent, Veritas, Quantum (Rocksoft), Exagrid...

**Constructeurs:** Overland, EMC2, Netapp, IBM, Hitachi, DDN Storage, Quantum,...

**Logiciel Libre :** Sun ZFS, Nexentastor, Openendedup SDFS,...

## 4.2/ L'archivage actif

### 4.2.1/ Définition de l'archivage actif

L'archivage actif consiste à déplacer les données utilisées vers un support de stockage moins coûteux pour une rétention long-terme. En général, on va mirroring cette donnée sur un deuxième support pour plus de sécurité.

Une archive active ne doit jamais être effacée. Cependant, on peut estimer sa durée de vie moyenne entre 20 et 50 ans aujourd'hui. Une procédure d'effacement peut être programmée en cas d'obsolescence avérée (Big Data dépassée)

Elle doit être indexée, localisée et pouvoir être retrouvée.

Une archive active a 3 âges: courant, intermédiaire et définitif.

Elle peut être stockée sur 3 types de support :

**Disques durs, SSD, ...**

**bandes LTO (TAR format ouvert ou LTFS format ouvert directement lisible comme un disque dur) )**

**Cloud (support disques ou bandes LTO (glacier Amazon))**

Les supports d'archivage les plus couramment utilisés sont des cartouches non réinscriptibles.: DON, Bandes LTO, UDO, disques durs miroirs ou en RAID6, ....

Les fichiers archivés peuvent être accessibles suivant les options suivantes :

**ON-LINE** : Cloud, direct, VTL, robots et juke- boxes

**OFF-LINE** : supports classés sur étagères : disques magnétiques, Bandes LTO, Sony UDO, Bluray,...

**Principaux éditeurs et clouds** : EMC2, Oracle, Xendata, Archiware, Intelligence (StorIQ One), Dropbox,BOX, ...

#### 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 (HSM)

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

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, EVS, Princetonsofttech, Intelligence (StorIQ One)...

#### 4.4/ Le stockage Objet

Le volume exponentiel des données non structurées (photos, vidéos,...) dans les clouds privés et publics (Web2.0) a nécessité l'arrivée d'un nouveau type de stockage en opposition aux stockages NAS et SAN traditionnels qu'on a dénommé : Le stockage objet qui est en format SDS (software defined storage)

Avantages

- évolutivité illimitée
- accès via des protocoles internet (API REST sur HTTP)
- métadonnées personnalisées
- compatibilité avec des matériels de toutes marques
- doublement ou triplement automatique des données (suivant leur valeur dans le temps)
- coût contenu

**Editeurs-outils :** IBM cleversafe, HDS HCP, DDM WOS, Scalify RING, EMC2 Elastic cloud storage, HGST active archive system, Netapp storage grid, RedHat Ceph, storage, ...

#### 4.5/ le stockage dans le cloud

La tendance profonde du marché est d'utiliser le stockage cloud pour l'accès direct en tout lieu de ses données stratégiques ou non.

Un grand compte disposera de datacenters internes mais pourra utiliser des datacenters externalisés en vue de sécuriser ses process en cas d'attaque extérieure ou de panne interne...

De même, les particuliers comme les entreprises peuvent utiliser des services externes de stockage comme dropbox, box, ..., pour sécuriser leur données et y accéder facilement via des Pcs, tablettes, mobiles...

Par exemple, facebook va conserver indéfiniment toutes nos photos et vidéos sans nous en faire supporter le coût...

L'achat de Cds musicaux, de DVDs et de blu-rays ne peut rivaliser avec l'offre de géants comme Netflix ou Canal+ qui donnent accès à des nouveautés sans s'encombrer de supports physiques et qui rendent quasi inutile la copie privée...

## 5/ Le Marché

Le marché du stockage hors cloud pèse aujourd'hui 30 Milliards d'euros dont 30% est dédié aux logiciels de gestion (croissance 25%/an).

Le standard S-ATA qui a remplacé SCSI a permis de baisser les coûts d'achat des disques et de systèmes de stockage et a entraîné l'apparition de nouveaux outils de même que les disques SSD ont permis des centaines de milliers d'accès simultanés sur les mêmes données (pratique pour des événements comme les keynotes, les concerts ou les écoutes de nouvelles musiques...)

Librairies virtuelles de bandes sur disques durs (VTL Virtual Tape Library)  
VOD films, séries, jeux en ligne,...  
Vidéoprotection  
Réseaux sociaux  
Objets connectés  
Réalité virtuelle  
Réalité augmentée  
vidéos et films 4K UHD-HDR – 8K  
Médical : télé-opération, radios haute définition, ...

**Les leaders mondiaux** : EMC2, HPE, Netapp, Hitachi, Intel , IBM, Dell, Oracle, DDN Storage, Brocade, ...

**Les acteurs émergents** : Pure Storage, Scality, Nutanix, Ctera Networks, Qumulo....

**Acteurs grand public**: Seagate, Western Digital, Toshiba, Samsung, ...

## 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 4 mètres cube nécessaires pour stocker aujourd'hui 11 Péta-octets sur 1140 disques durs de 10To (2 racks 44U de 6Po (600ToS/4U chez HGST).

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 60 ans et a démarré grâce à IBM avec le RAMAC de capacité 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 puisse être sauvegardée sur un support immortel (céramique, verre, ADN, ...). Les connaissances de l'humanité seraient alors préservées de l'oubli, de la destruction ou de l'érosion du temps, un peu comme l'ADN, les fossiles, la pierre et le carbone 14...