
Gérer le patrimoine informatique

La sauvegarde

BTS SIO - B1/U4 Support et mise à disposition des services informatiques

1. Objectifs et contraintes de la sauvegarde	3
1.1. Qu'est-ce que la sauvegarde informatique ?	3
1.2. Objectifs	3
1.3. Analyse des risques	3
1.4. Critères de choix	4
2. Stratégies de sauvegarde	5
2.1. Sauvegarde sur serveur	5
2.2. Sauvegarde sur système client	5
2.3. Sauvegarde sur Internet	6
2.3.1. Sauvegarde en ligne	6
3. Méthodes de sauvegarde	7
3.1. Mécanisme	7
3.2. Sauvegarde Complète	7
3.2.1. Détail technique	8
3.3. Sauvegarde Différentielle	8
3.3.1. Détail technique	8
3.4. Incrémentale ou incrémentielle	9
3.4.1. Détail technique	9
3.5. Décrémentale ou Reversed incremental	9
3.6. Formule de calcul de l'espace de sauvegarde nécessaire	10
3.7. Les snapshots des systèmes de fichiers	10
3.8. Les architectures à base de disques durs	11
3.9. Les sauvegardes de machines virtuelles à chaud	11
3.10. Les techniques de sauvegardes par bloc	11
4. Sauvegarde des fichiers ou du système	12
5. Les modèles des data repositories et les supports utilisés	13

1. Objectifs et contraintes de la sauvegarde

La protection des données est un point important dans un système d'information. Il est donc nécessaire de mettre en place les services adéquats pour pouvoir restaurer des données en cas de perte, vol, corruption...

1.1. Qu'est-ce que la sauvegarde informatique ?

En informatique, la sauvegarde (backup en anglais) est l'**opération qui consiste à mettre en sécurité les données contenues dans un système informatique**.

Par extension et par emprunt à l'anglais, ce mot s'applique à tout enregistrement d'information sur un support permanent (par ex. l'écriture sur disque dur d'informations en mémoire vive). Dans ce sens, il est synonyme de «stockage d'information». Mais le sens premier se rapproche de celui d'archivage, de conservation.

1.2. Objectifs

Les sauvegardes ou *copies de sûreté* sont utiles principalement à deux choses :

- La première et la plus évidente est de **permettre de restaurer un système informatique dans un état de fonctionnement à la suite d'un incident** (perte d'un support de stockage tel que disque dur, bande magnétique, etc., ou de tout ou partie des données qu'il contient).
- La seconde, incluse dans la première mais certainement la plus courante, **est de faciliter la restauration d'une partie d'un système informatique** (un fichier, un groupe de fichiers, un système d'exploitation, une donnée dans un fichier, etc.) à la suite d'une suppression accidentelle ou d'une modification non désirée.

La technique la plus fréquente est la copie des données sur un support indépendant du système initial (ordinateur local, serveur, etc.).

L'**opération inverse qui consiste à réutiliser des données sauvegardées s'appelle une restauration**. On parle de «stockage» pour les données sauvegardées en attente d'une éventuelle restauration.

En résumé, elle consiste à :

- Maximiser la disponibilité des données et minimiser la perte de données
- Adaptées à l'environnement et en fonction des ressources disponibles

1.3. Analyse des risques

Il peut être intéressant, avant de choisir une solution de sauvegarde, de procéder à une **analyse des risques**. Pour un particulier, cette analyse peut être très simple. Pour une entreprise, cela peut être plus complexe compte tenu de l'éco-système du système d'information. Les principaux risques auxquels la plupart des entreprises sont confrontés sont les suivants :

- perte de données par effacement (erreur humaine ou logiciel malveillant)
- perte de données par défaillance matérielle (disque dur défectueux, destruction accidentelle de support)
- sinistre (incendie, dégât des eaux)
- vol de support (cambriolage)

Une réflexion sur les différents risques énoncés ci-dessus vous indiquera la stratégie de sauvegarde à suivre.

Pour les points 1 et 2 (les plus fréquents) il suffit de faire des sauvegardes sur disque dur externe, sur disque réseau ou sur support amovible. Pour les points 3 et 4 (événements rares mais très destructeurs) une sauvegarde hors site devient essentielle.

1.4. Critères de choix

- le type et la fréquence
- la nature et la vitesse du matériel employé
- les tests de sauvegarde
- l'emplacement de stockage des supports de sauvegarde
- la durée de sauvegarde (rétention des sauvegardes)

2. Stratégies de sauvegarde

On distingue la sauvegarde d'un poste individuel et la sauvegarde sur serveur. L'une et l'autre s'adressent à la **même nature d'information** (la donnée informatique) et ont **le même objectif** (protéger l'information et permettre de la retrouver si elle était perdue), mais les méthodes de sauvegarde sont différentes pour plusieurs raisons :

- les données sur poste client sont réputées moins importantes que les données gérées sur des systèmes centraux ;
- les utilisateurs sont moins sensibilisés au risque de perte de données que les professionnels de l'informatique ;
- ils ont également moins de formation sur les techniques de sauvegarde ;
- les moyens techniques sont moins développés sur poste individuel que sur serveur, même si des progrès importants ont été réalisés ces dernières années (chute du rapport coût/volume des supports de sauvegarde, simplification des interfaces de sauvegarde, sauvegarde sans intervention de l'utilisateur, etc.)

De ce fait, la sauvegarde des données des postes individuels reste marginale (à tort en 2021) dans la stratégie d'utilisation des ordinateurs.

Cependant les entreprises, en généralisant l'usage des micro-ordinateurs et du partage des ressources en réseau, ont ressenti un besoin de sécurité qui a favorisé le développement d'outils de sauvegarde sur micro-ordinateurs, lesquels gagnent de plus en plus le monde de la micro-informatique personnelle.

2.1. Sauvegarde sur serveur

La sauvegarde s'inscrit dans une démarche plus globale qui consiste à assurer la continuité d'activité d'un système informatique ou, en cas de défaillance, son redémarrage le plus vite possible. Cette démarche est souvent formalisée dans un document qui peut porter des noms divers, par exemple le **Plan de reprise d'activité** (PRA) ou le plan de secours, et qui fait appel soit à des automatismes (ex. donner l'alerte en cas de coupure de courant ou de perte d'accès à une unité de stockage) soit à des gestes manuels (ex. remplacer des bandes magnétiques défectueuses). **La tendance est à l'automatisation, réputée plus sûre dans les situations d'urgence que les opérations manuelles.**

En termes de support, les serveurs ont depuis toujours requis des supports à grande capacité de stockage. La bande magnétique a longtemps été le principal vecteur, du fait de sa grande capacité, de son faible coût (par rapport aux autres supports), de sa capacité de réutilisation et de sa relative stabilité au temps et à l'usure. Puis sont venues les cartouches numériques (bandes magnétiques intégrées dans un boîtier plastique type DAT, DLT, SDLT, LTO), les disques durs et plus récemment les médias optiques, réinscriptibles ou non, tels que les CD-R, DVD-R ou formats similaires.

2.2. Sauvegarde sur système client

Au cours des années 1975-95, la plupart des utilisateurs d'ordinateurs personnels (PC) associaient principalement le terme "backup" au fait de faire des copies sur disquettes. Avec le développement de micro-ordinateurs mieux équipés, les utilisateurs personnels ont adopté des supports plus performants : disques optiques (CD-ROM ou DVD), clés USB. De même, les ordinateurs intègrent des fonctions de sauvegarde de plus en plus évoluées, par exemple :

- des outils intégrés au système d'exploitation tels que les "points de restauration" que l'on peut exécuter avant d'installer un nouveau logiciel et qui remettront le système en l'état d'avant l'installation si l'utilisateur le demande ;
- des logiciels capables de faire une image parfaite du système à un moment donné (image appelée un "ghost", en référence au logiciel du même nom, mot qui signifie "fantôme" en anglais) ; cette image sera stockée sur l'ordinateur lui-même ou sur un support externe.

2.3. Sauvegarde sur Internet

Avec la banalisation des connexions Internet à large bande et à haut débit, de plus en plus d'utilisateurs recourent à ce type de service de sauvegarde. On peut différencier deux méthodes :

2.3.1. Sauvegarde en ligne

Aujourd'hui, les copies de sûreté dites « en ligne » deviennent populaires. Elles consistent à se connecter à un site Internet, appelé « hébergeur », et à y transférer ses données. Les avantages sont multiples :

- minimiser le risque de perte puisque le site est géré par un professionnel qui fait lui-même des sauvegardes
- accéder à ses données à partir de n'importe quel ordinateur connecté à Internet
- souvent le coût de cette prestation est modique, parfois même gratuit pour les petites sauvegardes.

Sur le marché de la sauvegarde de données, les entreprises et administrations expriment des attentes spécifiques vis-à-vis des acteurs du marché, si bien qu'elles se tournent en priorité vers des intégrateurs, des éditeurs de solutions de sauvegarde, des fournisseurs d'infrastructure matérielle et des fournisseurs de services d'externalisation de sauvegarde, voire vers des prestataires couvrant un spectre complet d'expertises dans le domaine de la sauvegarde ainsi que de services à distance.

L'inconvénient majeur du recours à ces solutions est de laisser ses données à disposition d'un tiers qui peut à loisir les consulter, les modifier, les dupliquer, les publier ou en faire commerce, voire les rendre indisponibles (cas des faillites, rachats de sites par des concurrents, ou différend commercial avec l'hébergeur). Évidemment, des dispositions contractuelles viennent réguler ces risques mais elles ne peuvent empêcher l'hébergeur d'agir techniquement de façon malveillante. Une des parades à la consultation abusive consiste à chiffrer / crypter les données.

Un autre inconvénient vient des limites imposées sur le stockage ou la récupération des données : pour maîtriser l'usage de ses disques et de sa bande passante, un hébergeur peut limiter contractuellement son client à un volume de stockage ou de données consultées au-delà duquel il bloque l'accès aux données.

Les critères importants à retenir dans l'utilisation de ce processus de sauvegarde en ligne sont les suivants :

- Les données doivent être cryptées/chiffrées avant de remonter via Internet chez l'hébergeur. Cela empêche le prestataire d'exploiter les données du client par définition.
- L'hébergeur se doit d'avoir deux copies de vos données, pour se prémunir aussi d'une panne de son côté (si possible sur des réseaux informatiques distincts, sur des réseaux électriques indépendants, voire sur deux infrastructures géographiques distantes). Ne pas oublier que, si le client a une panne en même temps que le prestataire, la solution devient caduque pour le client dans ce cas de figure.
- L'hébergeur doit avoir vis-à-vis des professionnels une assurance « Responsabilité d'exploitation » adéquate avec le service proposé, afin que le client dans un cas ultime « puisse être couvert ».
- Le tarif des solutions n'est pas le plus important. Il faut d'abord et surtout compter sur la mise à disposition des données et la couverture du prestataire.

3. Méthodes de sauvegarde

Pour préserver de l'espace - qui dans le passé était souvent très cher - en plus de la compression des données, différents types de sauvegardes sont proposés par la plupart des logiciels.

Elle est longue et très coûteuse en termes d'espace disque, ce qui empêche de l'utiliser en pratique pour toutes les sauvegardes à effectuer.

Afin de gagner en rapidité et en temps de sauvegarde, il existe des méthodes qui procèdent à la sauvegarde des seules données modifiées et/ou ajoutées entre deux sauvegardes totales. Afin de gagner en rapidité et en temps de sauvegarde, il existe des méthodes qui procèdent à la sauvegarde des seules données modifiées et/ou ajoutées entre deux sauvegardes totales. On en recense deux :

- **La sauvegarde différentielle**
- **La sauvegarde incrémentielle**

La restauration d'un disque avec l'une de ces méthodes s'avère plus longue et plus fastidieuse puisqu'en plus de la restauration de la sauvegarde différentielle ou des sauvegardes incrémentielles, on doit également restaurer la dernière sauvegarde complète. Les fichiers supprimés entre-temps seront restaurés ou non (en fonction des fonctionnalités du logiciel de sauvegarde utilisé).

Afin de comprendre la différence entre les deux méthodes, nous prendrons l'exemple d'un plan de sauvegarde selon le cycle suivant :

- Une sauvegarde complète au jour J (dimanche soir par exemple)
- Une sauvegarde des fichiers modifiés ou nouveaux du jour J+1 au jour J+6 (du lundi soir au samedi soir inclus)
- Une sauvegarde complète au jour J+7 (dimanche soir suivant)

3.1. Mécanisme

Pour pouvoir différencier ces différentes méthodes de sauvegarde/archivage (complète, incrémentielle, différentielle), **le mécanisme mis en place est l'utilisation d'un marqueur d'archivage**. Chaque fichier possède ce marqueur d'archivage, qui **est positionné à "vrai"** lorsque l'on crée ou modifie un fichier. On peut comprendre cette position comme "Je viens d'être modifié ou créé : je suis prêt à être archivé donc je positionne mon marqueur à vrai". **Ce marqueur est appelé aussi attribut d'archivage** (ou bit d'archivage). Sous Windows, cet attribut est modifiable et peut être visualisé par la commande ATTRIB (attribut A pour archive). Le système de sauvegarde peut aussi constituer une base de données contenant les définitions des fichiers et utiliser un marquage interne

3.2. Sauvegarde Complète

Toutes les données sont sauvegardées. C'est la sauvegarde qui prend le plus d'espace, mais aussi la plus sûre et facile à récupérer car elle contient toutes les données.

Lors d'une sauvegarde complète, on va remettre à "0" l'attribut du fichier pour mémoriser le fait que le fichier a été enregistré. Lorsque l'on travaille avec la date, on mémorise la date de la dernière sauvegarde de façon à pouvoir différencier les fichiers qui ont été sauvegardés des autres (date de dernière modification).

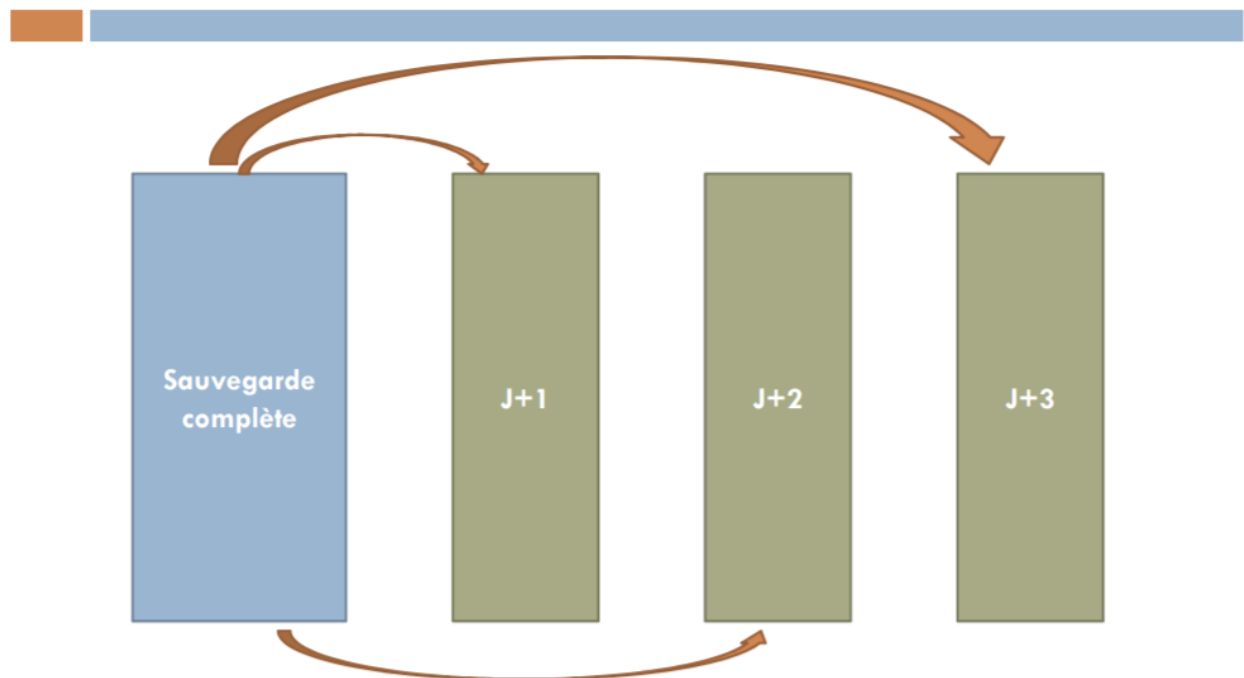
3.2.1. Détail technique

Lors d'une sauvegarde complète, tous les fichiers sont sauvegardés, indépendamment de la position du marqueur (vrai ou faux). Une fois le fichier archivé, celui-ci se voit attribuer la position de son marqueur (le bit d'archive) à "faux" (ou à "0").

3.3. Sauvegarde Différentielle

La sauvegarde différentielle englobe tous les fichiers ayant changé depuis la dernière sauvegarde pleine uniquement. Elle utilisera plus d'espace que les sauvegardes incrémentales mais la récupération des données complète ne nécessitera que la dernière sauvegarde pleine ainsi que la dernière sauvegarde différentielle.

La sauvegarde différentielle



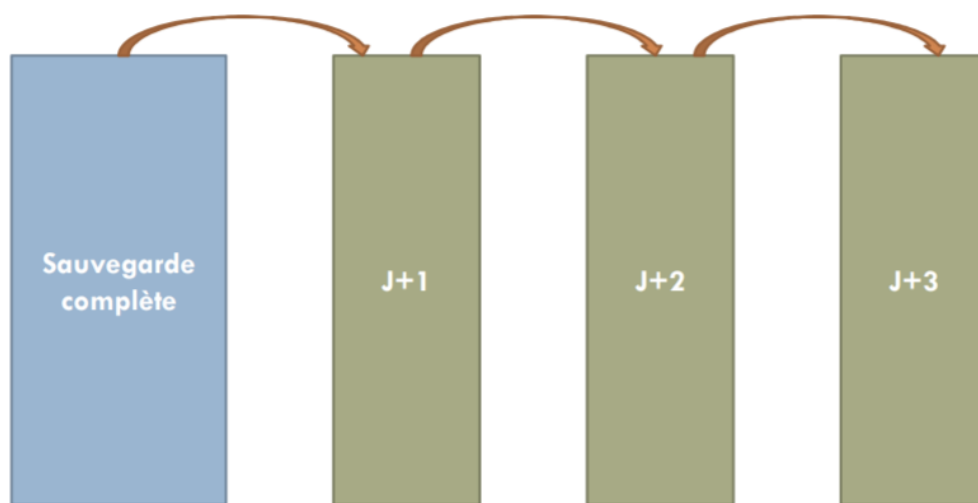
3.3.1. Détail technique

Lors d'une sauvegarde différentielle, tous les fichiers dont le marqueur est à "vrai" sont sauvegardés. Une fois le fichier archivé, celui-ci garde la position de son marqueur tel qu'il l'avait avant la sauvegarde.

3.4. Incrémentale ou incrémentielle

Une sauvegarde incrémentale suit toujours, soit une sauvegarde pleine, soit une autre sauvegarde incrémentale et ne contient que les fichiers qui ont changé depuis la dernière sauvegarde, quel que soit son type. Pour récupérer tous les fichiers, il faudra donc disposer de toute la série de sauvegardes depuis la dernière sauvegarde pleine. Ce type de sauvegarde est celui qui utilise le moins d'espace.

La sauvegarde incrémentale ou incrémentielle



3.4.1. Détail technique

Lors d'une sauvegarde incrémentielle, tous les fichiers dont le marqueur est à "vrai" sont sauvegardés. Une fois le fichier archivé, celui-ci se voit attribuer la position de son marqueur à "faux".

3.5. Décrémentale ou Reversed incremental

Contrairement à la sauvegarde incrémentale où la sauvegarde la plus ancienne est complète et les suivantes différentielles, le principe de la sauvegarde décrémentale consiste à obtenir une sauvegarde complète comme sauvegarde la plus récente et des sauvegardes différentielles pour les plus anciennes.

L'avantage tient au fait que la restauration complète du système dans son état le plus récent est simple et rapide, on n'utilise que la dernière sauvegarde, (contrairement à la méthode incrémentale qui implique la restauration de la plus ancienne (complète) puis de toutes les suivantes, incrémentales). Si maintenant on souhaite récupérer le système dans l'état de l'avant dernière sauvegarde, il faut restaurer la dernière sauvegarde (complète) puis la précédente (dite "décrémentale" parce qu'elle donne la différence à appliquer au système de fichier pour atteindre

l'état N-1 à partir de l'état N). Autre avantage, le recyclage de l'espace de stockage des sauvegardes est simple car il consiste à supprimer les sauvegardes les plus anciennes, alors que dans le cas des sauvegardes incrémentales le recyclage implique usuellement plusieurs jeux de sauvegarde (complète + incrémentales).

Le désavantage de cette approche est qu'elle nécessite plus de manipulation de données à chaque sauvegarde, car il faut construire une sauvegarde complète à chaque nouvelle sauvegarde et transformer l'ancienne sauvegarde la plus ancienne (qui était donc une sauvegarde complète) en une sauvegarde décrémentele.

3.6. Formule de calcul de l'espace de sauvegarde nécessaire

Cette formule permet de dimensionner une librairie de sauvegarde (bande ou disque VTL).

Dans le cas d'une sauvegarde classique, c'est-à-dire sauvegarde totale le week-end (vendredi soir) et sauvegardes incrémentielles les autres jours ouvrés de la semaine, du lundi au jeudi (pas le vendredi) soit quatre jours :

- soit D l'espace de donnée utile à sauvegarder
- soit R la durée de conservation des travaux souhaitée, exprimée en semaine
- soit T le taux de modification par jour des fichiers de l'espace à sauvegarder

La formule suivante est obtenue : **$D \times R + (D \times T\%) \times 4$** = capacité de sauvegarde.

Exemple chiffré : 100 Go au total à sauvegarder avec une rétention de 3 semaines et un taux de modification de 20% par jour donne $100 \times 3 + (100 \times 20\%) \times 4 = 380$ Go. 380 Go seront nécessaires pour sauvegarder nos 100 Go de données avec une rétention de 3 semaines et une modification de 20% par jour.

Des innovations technologiques telles que les snapshots ou la déduplication permettent de réduire cette valeur d'une façon très intéressante.

3.7. Les snapshots des systèmes de fichiers

Les systèmes de stockages centralisés sont les premiers éléments sur lesquels il est intéressant de mettre en place des mécanismes de sauvegarde. En effet, pourquoi multiplier les mécanismes à mettre en œuvre si le système hébergeant les données peut lui-même proposer des protections contre la perte ou l'altération involontaire des données. De plus ce premier niveau de sauvegarde coûte souvent peu cher à mettre en place, il faudra juste prévoir la place nécessaire au stockage des blocs qui sont modifiés, en général de 15 à 20% d'espace disque supplémentaire.

Plusieurs systèmes utilisant le principe de clichés instantanés (snapshots) sont maintenant proposés. Les solutions de stockage réseaux des constructeurs embarquent ces fonctions depuis déjà bien longtemps. À un instant précis, le système de fichiers est figé et seules les modifications de blocs suivantes seront écrites dans un espace différent. L'utilisateur peut alors accéder à l'instantané pour en récupérer les données lui-même sans avoir besoin de contacter son administrateur.

Les systèmes d'exploitation de Microsoft proposent le service Volume Shadow Copy (VSS) pour réaliser la prise de ces instantanés sur les partages proposés par des serveurs Windows 2003 ou plus. Le client VSS permet d'accéder aux différents instantanés via un nouvel onglet dans les propriétés des dossiers et fichiers pour leur restauration.

Les autres systèmes d'exploitation ne sont pas en reste puisque le système de fichier ZFS permet également de mettre en place une prise d'instantanés de façon très simple. Pour les systèmes GNU/Linux, les outils de gestion de volumes logiques LVM permettent de mettre en place des instantanés lors de la création d'un volume.

3.8. Les architectures à base de disques durs

La sauvegarde sur bande ne correspond plus aux problématiques actuelles. Les architectures à base de disques durs sont plus adaptées. On pourra citer comme avantages :

- les disques sont beaucoup plus souples d'utilisation : c'est un grand espace de stockage à plat modulable,
- ils sont plus performants en termes de temps d'accès et d'accès simultanés,
- ils peuvent être rendus robustes grâce à des mécanismes de redondance.

3.9. Les sauvegardes de machines virtuelles à chaud

Dans le cas de virtualisation à l'aide des produits VMWare, beaucoup proposent de se reposer sur le produit interne de backup de cet éditeur. Ce module permet facilement de faire une sauvegarde à chaud des machines virtuelles sans impact sur leur fonctionnement.

Une sauvegarde de machine virtuelle se décompose de la façon suivante :

- le serveur VCB déclenche la prise d'un instantané de la machine virtuelle,
- copie des fichiers composant l'instantané sur le serveur VCB,
- l'image peut alors être sauvegardée depuis le serveur VCB grâce à la solution logicielle en place.

La restauration se déroule selon le même principe puisque les images sont d'abord restaurées sur le serveur VCB avant de pouvoir réellement être restaurées.

3.10. Les techniques de sauvegardes par bloc

De nouvelles solutions de sauvegardes reprennent le principe des clichés instantanés qui sont mis en place sur le stockage primaire. Ces mécanismes ayant accès au plus bas niveau des données, les sauvegardes sont réalisées en se plaçant au niveau des blocs. C'est à dire qu'après une première sauvegarde complète du système, seuls les blocs ayant subi des modifications seront copiés sur le média de sauvegarde.

Ceci réduit fortement la consommation d'espace disque et surtout a pour avantage de s'affranchir des sauvegardes totales, on ne fait plus que de la sauvegarde incrémentielle.

Bien qu'ayant pour objectif la sauvegarde de leurs équipements propres, certains constructeurs proposent des agents à déployer sur des machines afin de réaliser la sauvegarde en utilisant leurs mécanismes, remplaçant alors les solutions génériques.

4. Sauvegarde des fichiers ou du système

Vos fichiers sont importants et leur sauvegarde primordiale en cas de perte de données. Mais avez vous également pensé à la disponibilité de votre système ? Que se passerait-il si votre disque dur venait à avoir une panne ? Vous auriez bien tous vos fichiers importants dans votre sauvegarde, mais il faudrait acheter un nouveau disque, réinstaller votre système d'exploitation et finalement réinstaller aussi tous vos logiciels avant de pouvoir récupérer vos documents.

Pour éviter ce type de désagrément certains logiciels permettent une sauvegarde non seulement des fichiers mais aussi de tout le disque. Ces logiciels font ce qu'on appelle une « image » du disque et permettent en cas de panne de prendre un nouveau disque et de recopier toute l'image sur celui-ci, rendant la panne moins douloureuse en temps et en efforts.

5. Les modèles des data repositories et les supports utilisés

La question du choix du support de sauvegarde est au moins aussi importante que celle de la stratégie. Les plus communs sont les suivants :

- supports amovibles (CD, DVD) ; très bon marché, mais de capacité limitée et rapidement détruits avec le temps ;
- les disques durs magnétiques ; bon marché, offrant les capacités de stockage les plus élevées, résistent assez bien à la dégradation au fil du temps ;
- les mémoires flash, SSD ; plus chères que les disques durs conventionnels, résistent moins bien à la dégradation au fil du temps ;
- les bandes magnétiques ; offrant de bonnes capacités de stockage, mais chères et peu pratiques, elles sont réservées à un usage professionnel ;
- services en ligne ; prix, disponibilité et pérennité dépendent du fournisseur.

Dans tous les cas, l'espace de stockage sera souvent le facteur déterminant le prix, mais aussi la faisabilité de la sauvegarde. Ceci nous amène au concept d'organisation du dépôt de données.