Exercice 01 – L'aspect légal des opérations de sauvegarde et de restauration :

Les **champs d'application de la loi liés à la sauvegarde des données** sont :

- Leur protection
- L'archivage des documents ayant une valeur fiscale
- L'archivage de la documentation et la communication interne
- L'archivage des fichiers log chez les FAI (selon l'employeur)

Pour **assurer la sécurité d'un fichier contenant des données personnelles**, on peut :

- Y exiger une authentification
- Y assigner des droits d'accès différenciés (NTFS par exemple)
- Le crypter
- Limiter le local dans lequel il se trouve par un badge ou une clé
- Sauvegarder l'inventaire des supports du fichier, la protection des supports de sauvegarde contre les sinistres, et la destruction des supports en fin de vie

En tant qu'informaticien, et **dans le domaine de l'archivage des données**, on doit :

- Documenter les processus d'archivage
- Journaliser les accès aux supports d'archivage
- Effectuer des inventaires de ces supports
- Mettre à jour la documentation

Les problèmes techniques qui peuvent ressortir de cet archivage sont la durée de vie des supports ainsi que le format des fichiers enregistré, qui doit être garanti 10 ans quant à sa lisibilité.

Liste des ordonnances, textes et différents et lois à connaître :

Nom	Abréviation	But
Préposé fédéral à la protection des données et à la transparence	**PFDDT**	Protection des données
Constitution (art. 13)	**-**	Protection des données
Loi fédérale sur la protection des données	**LPD**	Protection des données
Ordonnance relative à la loi fédérale sur la protection des données	**OLDP**	Protection des données
Code civil	**CC**	Protection des données
Ordonnance concernant la tenue et la conservation des livres de comptes (art. 9)	**Olico**	Régence de la conservation des documents*
Ordonnance sur la surveillance de la correspondance par poste et télécommunication (art. 26)	**OSCPT**	Régence des obligations des fournisseurs d'accès à internet

NOTE : Ces documents sont édités par le **Conseil fédéral et divers départements.**

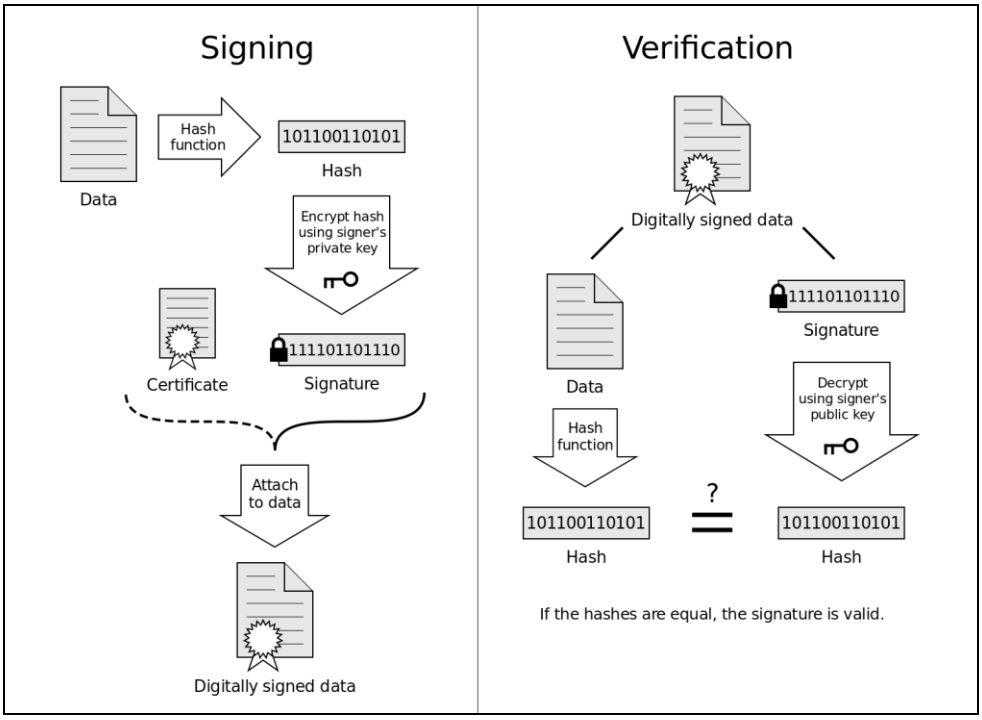
*****C'est cette ordonnance qui doit être consultée (art. 10) si une migration d'un support obsolète vers un support récent doit être faite. Pour conserver des documents, les supports autorisés sont :

- **Les supports d'infos non modifiables** (papier par exemple)
- **Les supports d'infos modifiable** si l'intégrité des données y est assurée (par exemple grâce à une signature éléc.), si le moment où les infos y ont étés enregistrée peut être prouvé (grâce à un système d'horodatage/timestamping par exemple), si les prescriptions relatives à leur utilisation sont respectées et que leurs procédures et modes d'utilisation sont conservés (logs, journaux de bords, protocoles, etc.).

Les **pièces comptables** doivent être conservées **au moins 10 ans** après la fin de l'exercice les concernant. Dans le cas **des logs** (identifications des usagers et données relatives à la facturation et au trafic) **d'un FAI**, cette durée est de **6 mois**. Dans un but de protection, ces logs ne peuvent être fournit qu'à un juge d'instruction ou un procureur ayant fait la demande à l'autorité compétente.

WORM (pour ***Write Once, Read Many***) est une propriété d'un support sur lequel l'information ne peut être modifiée une fois écrite.

Ci-dessous un tableau expliquant les notions de hachage et de signature électronique avec les mêmes étapes que dans le corrigé, donc techniquement suffisant et clair. C'est de l'anglais, mais à priori pas du Shakespeare donc ça devrait aller…



En cas de vol ou autre litige durant l'opération, l'envoyeur et possesseur de la clé privée est responsable car il ne peut pas se replier sur la perte de la clé.

Organismes reconnus par la loi pour **émettre des certificats électroniques** : Swisscom Solution, QuoVadis, SwissSign, BIT. Organismes en émettant : Symantec (Verisign), Thawte, Geotrust, etc…

Exercice 02 – Les supports de sauvegarde :

Dans le doute, quelques caractéristiques et ordres de grandeur de certains supports importants…

- **CD-ROM** : capacité native de 0.64 Go, prix moyen de 0.5.-, durée de vie moyenne de moins de 10 ans
- **DVD** : capacité native de 9.4 Go, prix moyen de 0.75.-
- **LTO-7** (la bande magnétique la plus actuelle) : capacité native comprise entre 6000 et 15000 Go, débit de 315 Mo/s, prix moyen de 140 francs, durée de vie moyenne de 30 ans.

Exercice 03 – Les types de sauvegarde :

Type de sauvegarde	Explication	Avantage(s)	Inconvénient(s)
Complète	Copie de la totalité des données vers un support de stockage (bandes magnétiques, cartouches, disques).	Lors de la restauration, les fichiers sont faciles à trouver car présents sur le support de stockage.	Demande beaucoup de temps et de place. S'il y a peu de modifications entre les saves, elles sont quasi identiques.
Différentielle	Traite uniquement les fichiers qui ont étés ajoutés ou modifiés depuis la dernière save **complète**.	Demande uniquement le support des dernières saves complètes et différentielles. Plus rapide	La restauration complète d'un système peut prendre beaucoup de temps. S'il y a eu de grosses modifications, elle peut être plus longue qu'une incrémentale.
Incrémentale	Concerne les fichiers qui ont étés ajoutés ou modifiés depuis la **dernière** sauvegarde.	Demande moins de place et permet des saves rapides.	La restauration complète d'un système peut prendre beaucoup plus de temps qu'avec une save complète ou différentielle.

À connaitre également, les sauvegardes (certaines proviennent d'autres exercices) **:**

- **Décrémentielle** : on garde la dernière version des données en entier. Les saves précédentes sont alors gardées comme différences de la save suivante.
- **En mode Delta Bloc** : incrémentielle améliorée. Seuls les blocs modifiés des fichiers concernés sont sauvegardés**.**
- **Par déduplication** : vise les sauvegardes de systèmes entiers. L'espace à sauvegarder est analysé et les fichiers doublons ne sont sauvegardés qu'une seule fois afin de limiter la taille de la save. L'avantage est la réduction de l'espace occupé par les saves (il semblerait que cette méthode permette de diviser par 20 voire 30 les besoins en espace de stockage). En revanche, on s'expose à des risques de pertes car les données ne sont plus en double. Le support doit donc être fiable.
- **Par point de restauration** : vise la sauvegarde de (fichiers) système en enregistrant les paramètres d'un OS dans le but de pouvoir le recréer au besoin.
- **Image** : destinée à la sauvegarde d'une partition entière non utilisée sous forme d'image.
- **Distribuée** : un serveur de backup va sauvegarder tout ce qu'il y a à sauvegarder dans l'entreprise.
- **Le snapshot** : capture d'une « photographie », à un instant spécifique, d'un état des données d'un volume à des fins de sauvegarde et/ou de protection, à ne pas confondre avec la save image : si on perd la base, le snapshot est inutile. Il est plus léger que la sauvegarde image (il se limite généralement à 10-20% du volume d'origine de base).
- **Miroir** : assure l'identité du contenu de deux copies de save en les comparant et en y ajoutant des nouveaux fichiers. Permet de sauvegarder les données dans deux directions, contrairement aux saves complètes et incrémentielles.
- **Synthétique** : permet de synthétiser une save complète antérieure et les saves incrémentielles suivantes vers une nouvelle sauvegarde complète.

Sur Windows, le mécanisme mis en place qu'utilisent les logiciels de sauvegarde est le **bit d'archivage**. Ce bit passe à vrai/1 pour signifier qu'une sauvegarde doit être faite lorsqu'il est créé ou modifié. Pour lister tous les fichiers prêts à être sauvegardés, on peut utiliser la commande **Attrib –A**, Attrib permettant de voir tous les attributs d'un fichier. On peut voir ces attributs dans l'explorateur de fichiers également, les différents attributs possibles sont **A** (Archive : fichier normal), **H** (Hidden : fichier caché), **R** (Read Only : lecture seule, non modifiable), **S** (System : fichier système selon Windows, plus complexe à supprimer) et **N** (Not Content Indexed : pas indexé par le service d'indexation.

Exercice 04 – Les types de sauvegarde et la planification :

Exercice purement pratique, pas vraiment de théorie. On peut néanmoins mentionner le **cycle de permutation Grand-père (mois)/Père (semaines)/Fils (jour)**. Les sauvegarde mensuelles (Grand-père), sont en général des complètes qui sont entreposées hors site pour des questions de sécurité. Ensuite, on effectue des sauvegardes complètes de manière hebdomadaire (Père), que l'on conserve sur site pendant la semaine de leur utilisation avant de les en sortir pendant deux ou trois cycles hebdomadaires. Enfin viennent les sauvegardes journalières incrémentales (Fils), dont le traitement est le même que pour les sauvegardes hebdomadaires. À noter que l'on conserve, en général, les bandes hebdomadaires pendant 3 semaines après leur départ du site avant de les réutiliser, contre 12 mois pour les bandes mensuelles.

Exercice 05 – Insérer un titre : Disparu ad vitam æternam dans de profonds et obscurs méandres…

Exercice 06 – Les logiciels de sauvegarde :

Un **agent de sauvegarde** est le terme générique désignant l'hôte qui gère la sauvegarde au nom du client de backup. Il permet de sauvegarder de manière simple et efficace des fichiers ouverts sur le réseau, comme des bases de données par exemple. Dans le cas de la sauvegarde de ces bases ainsi que de celui des programmes de courriels, un agent de sauvegarde est nécessaire car ce dernier est capable d'extraire les informations de manière très fine grâce à une connaissance importante de leur structure. Une solution standard ne serait pas capable de le faire. Ces agents de sauvegardes optimisent le trafic et le sécurisent. Pour sauvegarder une base de données, une option « **OpenFile** », qui permet de résoudre des problèmes pouvant survenir durant certaines opérations comme la protection de fichiers ouverts par exemple est préconisée, bien que très chère.

L'acronyme **VTL** (*Virtual Tape Library*) désigne une forme d'émulation d'une librairie de stockage par bandes magnétiques à l'aide de disques, dans le but d'améliorer les performances en temps d'accès et en écriture, et de compenser les inconvénients du système de stockage sur bandes grâce à une couche intérmédiaire sous forme de disques. Une VTL fonctionne sous réseau de stockage Fiber Channel ou TCP/IP et se place à côté du serveur d'applications, avant les librairies de stockage par bandes.

Le principe **D2D2T** (*Disk To Disk To Tape*) consiste à sauvegarder d'abord sur un disque (rapide) puis sur une cassette (lente), dans le but d'optimiser les performances.

Exercice 07 – Sauvegarde du contrôleur de domaine avec Sauvegarde Windows Server :

Les éléments à sauvegarder d'un serveur qui est DC et serveur de fichiers sont la partition (C :), l'état du système, les données d'annuaire (S :) ainsi que les données d'utilisateurs (D :). Normalement, les fichiers ouverts ne peuvent être sauvegardés, et les fichiers DB, du DNS ou de l'AD sont sensés l'être en permanence, d'où l'intérêt des options « OpenFile » citées précédemment.

Pour restaurer l'AD, il y a deux méthodes :

- **La restauration non-autoritative** : principalement utilisée dans le cas où un DC tombe en panne. l'OS restaure le contenu du DC à partir d'une save. Après ça, le DC reçoit tous les changements effectués depuis la sauvegarde des autres DC du réseau, grâce à la réplication.
- **La restauration autoritative** : plus couramment utilisée lorsqu'une modification opérée dans l'AD doit être inversée, comme des suppressions d'OU accidentelles par exemples. Le processus rétablit le courant continu de la save puis la réplique, et remplace tous les DC du réseau en fonction du DC restauré.

Exercice 08 – Découverte d'un logiciel de sauvegarde du marché :

Exercice purement pratique. Le peu de théorie s'y trouvant se résume à la définition d'une sauvegarde synthétique (expliquée à la page 3) ainsi que de relever pourquoi une save par déduplication pèse moins lourd et se fait plus rapidement qu'une normale…

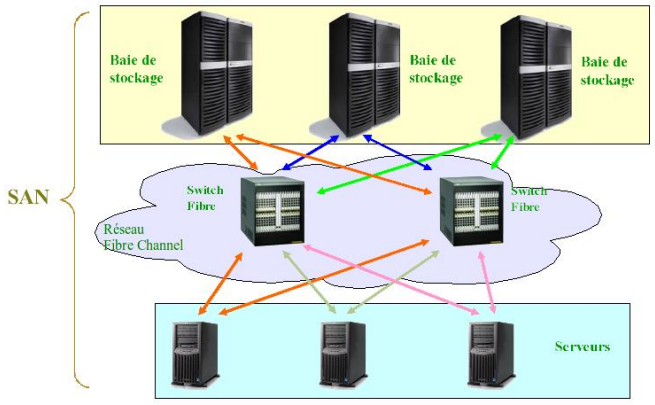
Exercice 09 – Les système de Backup (Snapshots, NAS, SAN, iSCSI) :

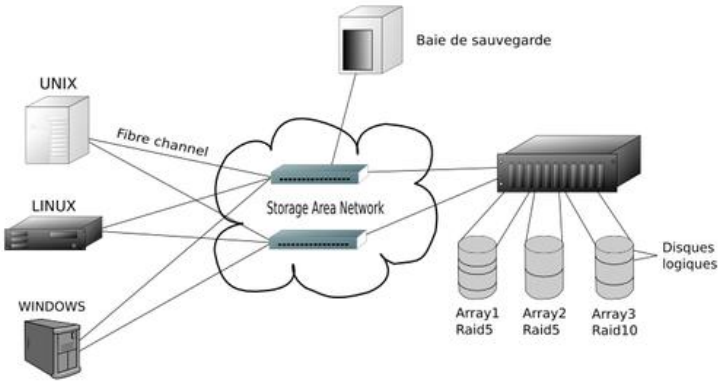
Le snapshot a, lui aussi, déjà été traité à la page 3. En complément, il convient de savoir que, hors de l'utilisation d'une VM (donc plutôt dans le cas d'un NAS/SAN), ces snapshots sont transférés en temps réel vers un serveur de sauvegarde. Durant ce transfert, il est possible de sauvegarder certaines instances/itérations de ces données sur une baie de disques dédiée, afin d'en effectuer une **restauration instantanée** si nécessaire. On notera également que les snapshots ont été développés pour protéger (d'autant plus en cas d'erreur) et accéder rapidement à des informations via la restauration, ainsi qu'à la confidentialité desdites informations. Le **snapclone**, quant à lui, est un clone, une copie complète d'une machine virtuelle par exemple, lorsque le **snapshot** est vraiment le delta de cette machine à un instant voulu.

- **DAS** (*Direct Attached Storage*) : méthode de connexion où la baie de disque est directement connectée sur un serveur.
- **NAS** (*Network Attached Storage*) : la connexion d'une baie de stockage se fait directement sur le réseau. Elle est alors accessible par tous les périphériques connectés au même réseau sans devoir passer par un hôte intermédiaire.
- **SAN** (*Storage Area Network*) : on reste sur les baies de stockage connectées au réseau, mais ce réseau leur est dédié.
- **iSCSI** (*Internet Small Computer System Interface*) : protocole de stockage en réseau basé sur le protocole IP destiné à relier les installations de stockage de données.

Les avantages du **NAS** par rapport à un serveur de fichiers sont la place qu'il offre pour de nombreux disques durs, le fait qu'il soit préinstallé avec un OS et des outils associés, qu'il fournit de l'espace pour différents types de clients (Unix, Windows, Mac) et qu'il présente moins de risques de panne. Les protocoles associés sont NFS (Unix), CIFS (Windows), AFP (Apple) et SMB (Windows).

Quant au **SAN**, ses avantages résident dans le fait que l'espace disque y est évolutif en fonction des ajouts de disques/baies, qu'il y est plus facile de répliquer des données, que les performances sont meilleures tout en allégeant le réseau local, que la redondance de tous les éléments le rendent hautement disponible, et qu'il peut fonctionner dans un environnement complètement hétérogène (serveurs Windows + Netware + Unix, etc..). Voilà sa représentation graphique :

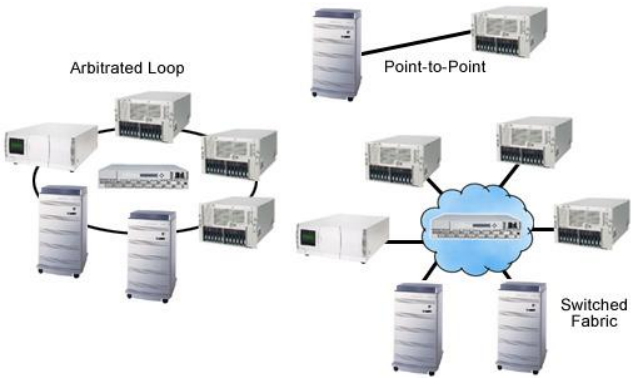




Comme on peut le voir, c'est une infrastructure complexe, qui est donc onéreuse, lourde à mettre en place, et qui nécessite du personnel très compétent. Ce sont les principaux défauts du SAN. On voit le protocole Fibre Channel sur les schémas de la page précédente, mais un SAN peut également fonctionner avec le protocole **iSCSI** (encapsulé dans les trames TCP/IP, ne nécessite que très peu de matériel spécifique) ou **FCoE** (*Fibre Channel Over Ethernet*, pas encapsulé dans les trames TCP/IP, mais requiert des cartes réseau et des switches très spécifiques et coûteux).

Le SAN se décline en trois topologies visibles ci-contre :

- **Point à point** : topologie la plus simple, les périphériques sont directement reliés entre eux. Offre la meilleure bande passante.
- **FC-ArbitratedLoop** : les périphériques sont reliés pour former une boucle, ce qui permet de relier davantage d'objets entre eux ou de pallier la défaillance d'un périphérique avec un ou plusieurs hubs.
- **Fabric** : En mailles. Apporte le meilleur taux de disponibilité car on peut doubler chaque lien. De plus, chaque communication possède sa propre bande passant contrairement à la topo. en boucle. C'est la topologie qui demande le plus grand investissement.



Dans une infrastructure iSCSI, les deux éléments sont :

- Un **initiator** : composant comportant un pilote pour gérer et transporter des blocs de commandes sur le réseau IP.
- Une **target** : périphérique qui reçoit et traite les commandes, typiquement un périphérique de stockage mais peut également être un pont réseau entre IP et Fibre Channel.

Enfin, il faut connaître la notion d'***iSCSI Software Initiator***, qui permet à Windows de se connecter à un réseau de stockage iSCSI.