

# Sauvegarde versus Archivage: sauvegarde

- sauvegarde: c'est dupliquer et externaliser les données pour se protéger :
  - des erreurs humaines
  - des crash matériel
  - des erreurs logicielles
  - des sinistres (incendie, inondation, ...)
  - de la malveillance (vol, virus, ...)
- la durée de vie d'un jeu de sauvegarde est limitée dans le temps (quelques mois au plus)
- Sauvegarde des données / du système

# Sauvegarde versus Archivage: sauvegarde

- on sauvegarde certaines données pour se protéger de certains risques
  - faire la liste des risques dont il faut se protéger
  - adapter le cahier des charges des sauvegardes et des autres mesures en fonction des ces risques
  - les sauvegardes ne sont qu'un des éléments permettant de garantir la continuité du service à côté d'autres : disques raid, redondance des serveurs, ...

# Sauvegarde versus Archivage: archivage

- l'archivage, c'est le stockage hors ligne de données peu utilisées
  - le temps d'accès a souvent peu d'importance
  - on peut avoir une hiérarchie de mode de stockage (du plus rapide au moins rapide (bande à aller chercher) suivant la fréquence de l'utilisation des données concernées)
- la durée de vie d'une archive se compte au minimum en années, voire en décennies
- il faut donc penser à l'obsolescence des matériels, des supports, des logiciels, des standards utilisés

# Solutions qui ont prouvé leur inefficacité

- Faire faire les sauvegardes par les utilisateurs : pour dégager sa responsabilité mais aucune garantie qu'elles seront faites
- Sauvegardes sur des supports peu fiables (disquettes, DAT, ...)
- Sauvegarde en écriture seule : il faut valider sauvegarde et procédures de restauration
- Sauvegarde d'un système en cours d'exécution
- Un seul support de sauvegarde
- Sauver sur une partition du même disque
- Pas de sauvegarde hors site (incendie, ...)

# Sauvegardes : procédure/planification

- Planifier les sauvegardes, tester leur réalisation
  - il faut avoir l'assurance que les sauvegarde prévues ont eu lieu
  - les procédures de sauvegardes doit être écrites
  - procédures testées
  - procédures et planification validées par les utilisateurs/propriétaires des données

# Sauvegardes : planification

- choix des données à sauvegarder :
  - données des utilisateurs (y compris : boîtes aux lettres, profil, ...), fichiers de configuration, ..;
    - retrouver un système en état suppose de réinstaller le système puis de restaurer les données
    - volumétrie plus faible
  - Système entier avec procédure de redémarrage:
    - restauration directe et rapide d'un système opérationnel
    - volumétrie plus importante
- périodicité, choix des données ont un impact fort sur la volumétrie et donc sur le coût  
=>compromis

# Problèmes liés à la volumétrie:

- Coût
- Charge réseau
- Durée des sauvegardes
- Indisponibilité des serveurs/applications dans le cas où le logiciel de sauvegarde impose l'arrêt des logiciels.

# Restauration

- procédures de restauration
  - écrites
  - au résultat validé par les propriétaires des données (pour ne pas en oublier)
  - **testées régulièrement en grandeur réelle** de façon à garantir :
    - que toutes les données pertinentes ont été sauvegardées
    - d'être capable de tout restaurer correctement

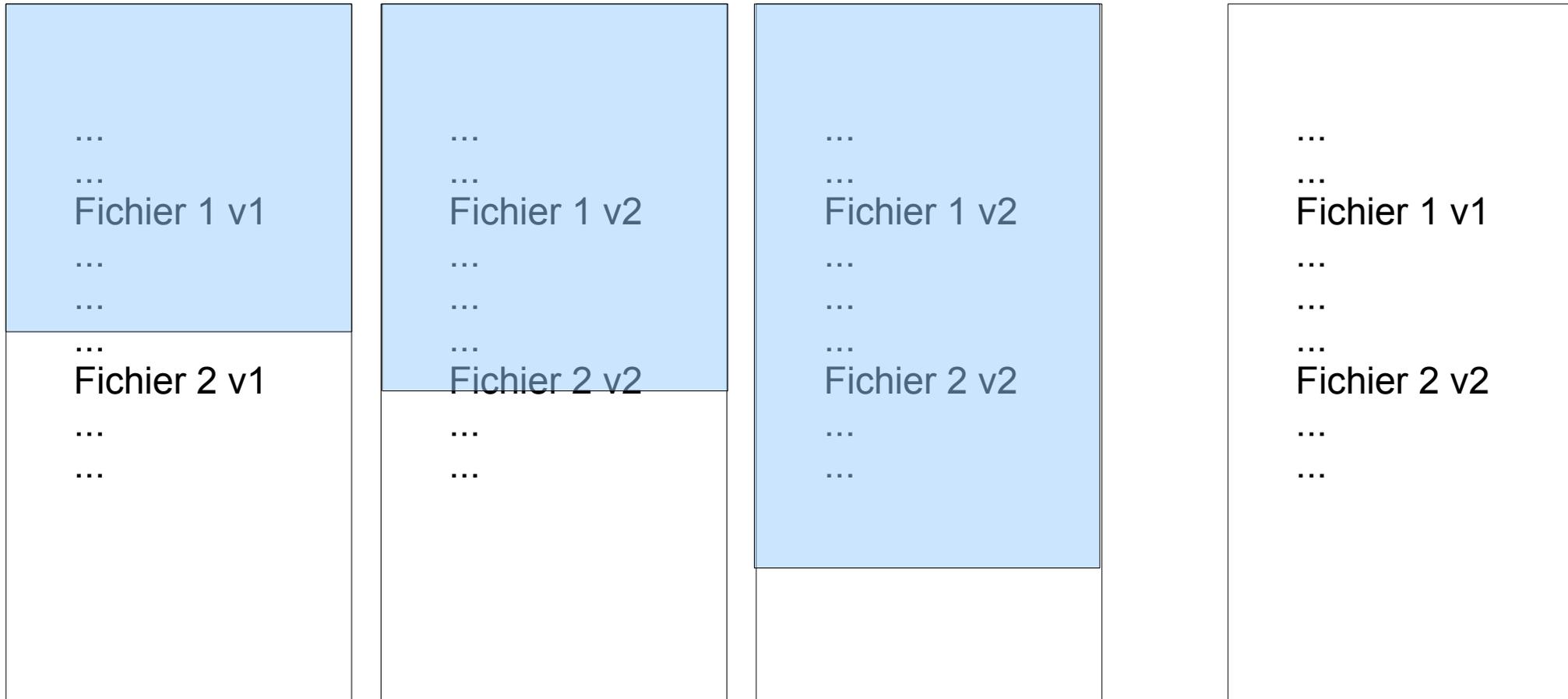
# Fiabilisation

- utiliser des média fiables
  - éviter disquettes, CD RW, DAT, ..et leur préférer CD R (bof), DLT, AIT, ...
  - varier les marques de media pour palier un défaut de fabrication dans une série, ...)
- gérer le vieillissement des média de sauvegarde/archivage (reprise sur des média récents, ...)

# Fiabilisation (2)

- fiabiliser l'environnement des media de sauvegarde:
  - inondation, incendie, vol (coffre ignifugé, ...)
  - sauvegarde hors site (penser à la confidentialité des données, au risque de vol, à l'interception des données, ...)
- indexer les données pour s'y retrouver dans le volume total des jeux de sauvegardes
  - indexer les données
  - étiqueter les media

# Sauvegarde live: problématique



Si l'application modifie plusieurs fichiers durant la sauvegarde, les versions sauvées peuvent ne pas être cohérentes.

Sauvegarde

# Sauvegarde live: solutions

- Arrêt de l'application pendant la sauvegarde:
  - garantit de plus qu'aucun verrou n'empêchera l'accès à un fichier
- Instantané (Snapshot): image des blocs du sgf, en cas d'effacement de fichier, les blocs ne sont pas réalloués tant que la sauvegarde n'est pas finie
  - Suppose un support dans le sgf
  - Proposé par les NAS, par afs, LVM, UFS2 (FreeBSD, ...), ...
  - Application supplémentaire : proposer aux utilisateurs une image des états antérieurs de leur compte à moindre coût.

# Sauvegarde live: snapshot

- Là, on mettra une illustration du fonctionnement des snapshot
- Fait en live au tableau.

# Sauvegardes incrémentales/différentielles

- Sauvegarde incrémentales : fichiers créés ou modifiés depuis la sauvegarde précédente
  - Diminue le volume à sauver
  - Restauration nécessite toutes les sauvegardes, restaure toutes les versions d'un même fichier
  - Peu adapté si la totalité des fichiers changent
- Sauvegarde différentielle : fichiers créés/modifiés depuis la dernière sauvegarde de référence
  - Diminue le volume à sauver
  - Plus de volume qu'en incrémental
  - Restauration nécessite la sauvegarde de référence et la dernière sauvegarde différentielle

# Exemple:

- Comparer la taille des sauvegardes et les procédures de restauration dans les cas suivants : (ST: sauvegarde totale, I: sauvegarde incrémentale, D: sauvegarde Différentielle)
- Cas 1) ST, I1, I2, I3, I4, I5, I5
- Cas 2) ST, D1, D2, D3, D4, D5
- Cas 3) ST, I1, I2, I3, D1, I4, I5

# Dump: un outil de sauvegarde sous unix

- Dump est un outil unix
- Il est efficace (travail directement au niveau du sgf)
- Sauvegarde non portable d'un unix à un autre (lié au sgf)
- Niveau (0 à 9) permettant un gestion très souples des sauvegarde incrémentales/differentielles:
  - Une sauvegarde niveau  $n$  sauvegarde tous les fichiers modifiés depuis la dernière sauvegarde de niveau  $n-1$