



Préambule

- Remerciements
 - Catherine Dejancourt (UREC)
 - L'UREC
 - IAP
 - IN2P3 (retransmission vidéo)
 - Catherine, Gilles et Jean-Yves
 - Les intervenants



Définir une politique de sauvegarde

pour garantir l'intégrité des données, la fiabilité et la continuité de l'activité du laboratoire en cas d'incident.

Françoise BERTHOUD

Laboratoire de Physique et Modélisation des Milieux Condensés
Unité des REseaux du CNRS

Quelques définitions

- Sauvegardes

Activité de copie d'un ensemble ou sous-ensemble (partitions, répertoires, fichiers uniques, ...) de données sur un média en vue de leur préservation

- Archivage

Conservation des données (non modifiables) pour les besoins internes ou pour des raisons juridiques, fiscales, administratives.

- Stockage

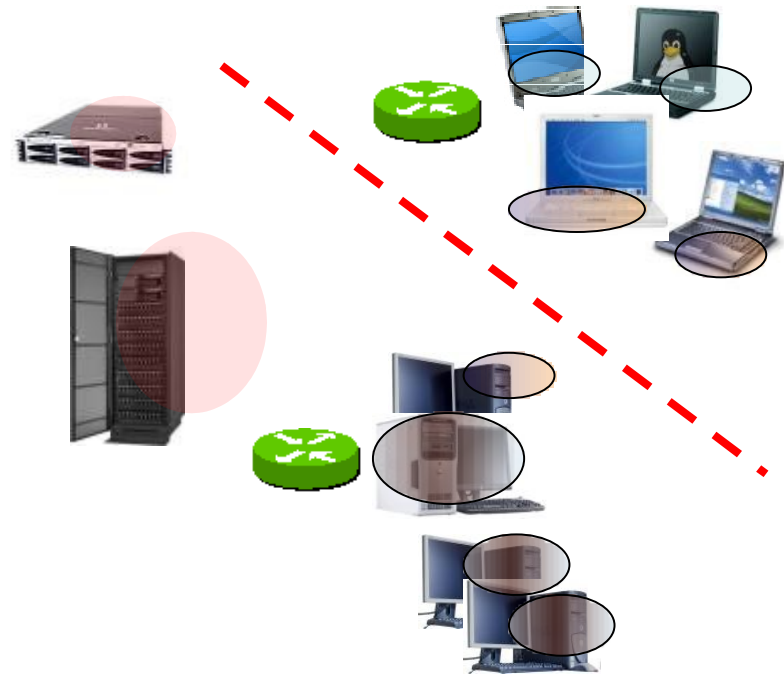
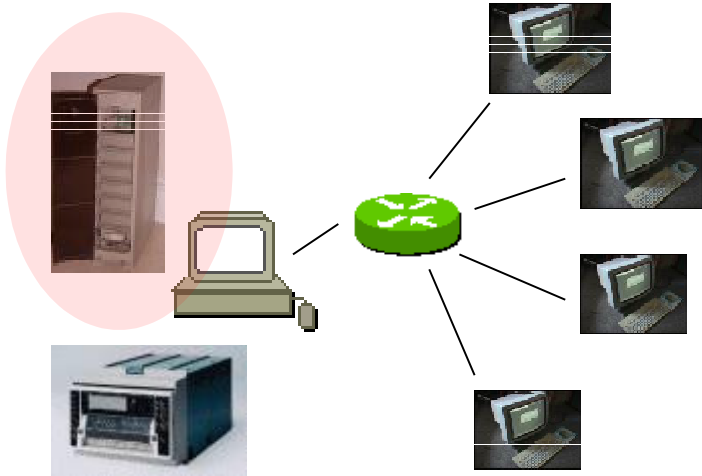
désigne l'ensemble des périphériques matériels permettant de conserver une information



une certaine complexification ...

Avant

Maintenant



.. dans la distribution des données, le type de données, les systèmes impactés, la connectivité réseau etc.



Contexte : laboratoire de recherche

Population : administratifs, chercheurs / techniciens (*définition du périmètre*)

qui manipulent *différents type de Données* (cf CAPSEC-PSSI <http://www.jres.org/slides/112.pdf>)

- documents auto-produits, échangés (techniques / scientifiques / administratifs)
- Résultats d'expériences
- Fichiers de travail temporaires
- Emails
- Certificats / signatures / documents administratifs / brevets (archivage ??)
- Fichiers système
- Bases de données
- Et d'autres ... (logiciels récupérés sur le net, photos des enfants, MP3, DIVX etc.)

Toutes ces données ne sont pas critiques ...

Le type de données à sauvegarder évolue avec le temps.

Les données peuvent être chiffrées sur le poste client .. Attention !



Dans le temps ...

- Sauvegardes totales
- Sauvegardes incrémentales : on sauvegarde les éléments qui ont changé depuis la dernière sauvegarde quelque soit son type.
- Sauvegardes différentielles : on sauvegarde les éléments qui ont changé depuis la dernière sauvegarde totale.
- Synchronisation

Choix fonction de :

- Capacité de stockage, volume à sauvegarder et débit
- Besoins en rapidité de restauration

À quelle fréquence ?

Combien de temps garde-t-on quelles sauvegardes ?

➔ VOLUME et SUPPORT

Contre quoi veut on se prémunir ?

- Erreurs de manipulation (utilisateur / administrateur)
 - Problèmes matériels (pannes disques – fiabilité du matériel)
 - Malversations (internes ou externes)
 - Vols / incendie
- Localisation du (des) supports de sauvegarde et comment (redondance ?)
- Protection des données -sensibles- sauvegardées, chiffrement ?, externalisation -interne-)
- Dans la grande majorité des laboratoires, la sauvegarde se trouve à moins de 1m des données à protéger -



Moyens

- Quels moyens humains ?
 - Disponibilité
 - Expertise

- Quels moyens financiers ?
 - Obtenus
 - Possibles



Sauvegardes et architecture informatique


- Architecture du parc informatique (types d'environnement)
 - Postes clients autonomes et hétérogènes - architecture centralisée ?
- ➔ Un bon équilibre dans l'architecture du schéma informatique du laboratoire peut grandement faciliter la mise en œuvre de la politique de sauvegarde !



En résumé : des questions

qui n'ont pas une unique réponse
(ni dans l'espace, ni dans le temps)

- Quel périmètre (services, matériels, sites, utilisateurs)
- Quel type de données (classification, données chiffrées)
- Quel volume ?
- Quelle fréquence de sauvegarde, quelle durée de rétention des sauvegardes
- Quels moyens humains/techniques, quels moyens de stockage
- Ou se trouvent les sauvegardes ? (interne et/ou externalisé en interne et/ou externalisé - supports mensuels et annuels en dehors du site)
- Politique de restauration (qui restaure ?)



Répondre à ces questions, c'est s'engager dans un processus d'analyse qui implique aussi les personnels non informaticiens → important travail de communication au sein du laboratoire

Les réponses à ces questions permettent de préciser la politique de sauvegarde

- Tester la bonne récupération des données
- Prévoir le contrôle régulier d'un journal de sauvegardes



Mise en oeuvre

- Ressources internes
 - ➔ Analyse + Maîtrise technique et fonctionnelle, dans la durée
- Sous-traitance interne (déjà déployée dans certains campus / centres de ressources pour certains labo)
- Sous-traitance externe (attention à la protection des données !!)



Pour une mise en œuvre en interne ...

- Connaître les technologies de stockage
- Et les logiciels de gestion de sauvegardes évolués (présentés aujourd'hui au travers de retour d'expérience)
 - Libres : backuppc, bacula, amanda
 - Commerciaux : Arkeia, Time Navigator , Netbackup, éléments de windows, NetVault
- Présentations / Documentations prochainement accessibles à partir de la page web de JoSy:sauvegarde



Briques de base

Les outils historiques (qui ont fait leurs preuves ...)

- `dd` : copie de données
Entrée/Sortie standard ou système de fichier,
Gestion de la taille des blocs,
Manipulations sur les blocs.
- `cpio` : copie de fichiers dans une archive
Entrée/Sortie standard ou système de fichiers,
Formats de sortie possible : binaire, hexa, tar, ...
Prise en compte des expressions régulières,
Tailles de blocs paramétrable,
Compression possible.

```
find /home/ | cpio -o > /mnt/backup/home-backup.cpio
```



Briques de base

- **Tar : Tape Archiver (Outil GNU)**
 - Créer une archive,
 - Rajouter/Extraire/Remplacer des fichiers,
 - Editer la table des index,
 - Réaliser des différences,
 - Prise en compte de la taille des périphériques de stockage,
 - Reprise sur erreur,
 - Préservation de toutes les propriétés des fichiers,
 - Compression avec gzip.

```
tar czf /mnt/backup/home-backup.tar.gz ./home/
```



Briques de base

- Rmt : Outil de gestion des bandes magnétiques
 - Opérations bas-niveau sur les périphériques de stockage,
 - Ouverture/fermeture d'un périphérique,
 - Statut des périphériques,
 - Lecteur/écriture de données.
- Dump : Backup de systèmes de fichiers
 - Prise en compte du niveau de backup :
 - 0 : backup complet,
 - N (1 à 9) : backup des fichiers ajoutés ou modifiés depuis le dernier backup de nombre inférieur ou égal.
 - Utilise Rmt pour le stockage sur bande,

MAIS : Dump est déconseillé par Linux depuis le noyau 2.4.x du fait de la gestion de son cache.



Briques de base

- Avantages :
 - Outils en ligne de commande,
 - Aisément « scriptables »,
 - Utilisation très précise.
- Inconvénients :
 - Convivialité,
 - Interactivité obligatoire (si le backup dépasse la taille d'une bande, génère des erreurs, etc.),
 - Nécessité de maîtriser le flux de données de façon très précise,
 - Outils limités au monde Unix.
 - Ne sont pas adaptés aux parcs hétérogènes

Un peu plus évolué : rsync

- `rsync --force --ignore-errors --delete --delete-excluded --exclude-from=exclude_file --backup --backup-dir='date +%Y-%m-%d' -av -e ssh /home/moi/ moi@autre.serveur.net:archives/`
- Les options sont:
 - `--backup`: crée des sauvegardes des fichiers avant de les écraser en les écrivant sur eux-mêmes
 - `--backup-dir='date +%Y-%m-%d'`: crée un répertoire de backup pour ces sauvegardes, qui ressemble à ça pour la sauvegarde du 15 août 2003: 2003-08-15
 - `-av`: mode archive et mode verbeux.
 - `--delete`: détruit les répertoires et fichiers qui n'existent plus sur le source –force les détruit même si ils ne sont pas vides)
 - `--delete-excluded`: détruit même les fichiers correspondants aux fichiers exclus
 - `-e ssh`: on utilise le protocole ssh (et non rcp) entre les deux machines (manipulation clefs privées/publiques)

Syntaxe du fichier exclude :

- *.o : exclure tous les fichiers se terminant par ".o" (utile pour les programmeurs, ce sont des fichiers intermédiaires pour la compilation d'un programme)
- /foo : exclure un fichier dans le répertoire de base, nommé "foo".
- foo/ : exclure n'importe quel répertoire nommé "foo".
- /foo/*/bar : exclure n'importe quel fichier "bar" situé dans le sous-répertoire de "foo".
- /foo/**/bar : exclure n'importe quel fichier "bar" situé dans un (ou plusieurs) sous-répertoires de "foo".
- + /foo/allo/bar : explicitement inclure le fichier bar pour un cas précis, même si une autre règle l'exclut. Cette règle doit être placée *avant* la règle générale.