Éléments de réflexion préalable à la mise en place de sauvegardes individuelles

Arnaud TISSERAND

Avril 2025

Organisation de la séance

1. Premiers éléments (questions après cette partie)
2. Autres éléments
3. Démos
4. Mots de la fin

Avertissement

Ceci est un résumé *personnel* d'éléments de réflexion préalable à la mise en œuvre et utilisation de sauvegardes dans un cadre individuel.

La mise en place de sauvegardes est une chose critique qu'il convient de faire seulement si on s'estime en capacité d'en maîtriser les différents aspects ! Sinon, mieux vaut demander de l'aide...

Chaque personne doit s'assurer que les informations données ici sont compatibles avec les règlements et pratiques de son cadre d'usage, employeur, laboratoire ou établissement d'hébergement.

Dans un cadre professionnel, il convient de prendre appui sur les solutions de sauvegarde proposées par le service informatique.

Je ne conseillerai pas d'outils, techniques, ou matériels. Enfin, ce qui est présenté dans la suite est indicatif, partiel, potentiellement erroné ou incompatible avec certains usages, et n'offre aucune garantie d'une quelconque sorte.

Sauvegardes individuelles

Ici, cela signifie que la même personne endosse tous les rôles suivants :

- utilise la machine à sauvegarder
- bénéficie de droits d'administration suffisants
- effectue les sauvegardes
- restaurera les données en cas de problème

D'autres cadres de travail, comme celui d'utilisateurs et utilisatrices multiples, ne seront pas considérés.

Vocabulaire

- données / méta-données
- sauvegarder / restaurer ou récupérer

- redondance
- diversité
- total / incrémental / différentiel
- tester / vérifier
- analyser / planifier
- documenter
- automatique / manuel
- isoler / cloisonner / sécuriser

Pourquoi faire des sauvegardes ?

Pour restaurer des données en cas de problème :

- erreur humaine
- problème logiciel
- panne matérielle
- attaque informatique
- vol
- incendie / destruction
- contrainte réglementaire
- ???

Impact du type de problème sur la restauration

Après le problème, le couple machine et OS :

- reste fonctionnel, mais certaines données sont inaccessibles / corrompues
- est physiquement intègre mais n'est plus fonctionnel (requiert réinstallation)
- n'est plus physiquement intègre (requiert réparation / changement de pièce)
- doit être intégralement remplacé

Besoins et étapes différents pour la restauration.

Important : anticiper le pire scénario...

Quoi sauvegarder ?

- Tout ce qui est nécessaire pour restaurer : outils, scripts, configurations, paramètres, documentations, clés de chiffrement
- fichiers sources
- configurations, scripts d'administration, logs importants
- documents importants
- bases de données
- données de mesures coûteuses / difficilement reproductibles

- documents réglementaires
- méta-données (dates, propriétaires, droits, groupes, ACL)
- ???

Exclure : caches, fichiers temporaires / intermédiaires / objets, fichiers redondants, clones dépôts externes, outils installés, etc..

Où sauvegarder (dispositifs physiques) ?

- disque interne dédié (actif uniquement lors des sauvegardes)
- disque externe (mis en sécurité en dehors des sauvegardes)
- clé USB ou carte mémoire
- dispositifs spécifiques : NAS, baies RAID, bandes
- dispositif / service distant : autre pièce / bâtiment / site

Éléments clés :

- redondance : multiples copies des données
- diversité technologique
- diversité spatiale

Quand sauvegarder ?

Élément clé : diversité temporelle

Combinaisons de fréquences :

- journalier (plusieurs fois par jour ?)
- hebdomadaire
- mensuel
- trimestriel / annuel

Important : combiner les « moments » avec différents dispositifs et processus automatiques / manuels.

Qui sauvegarde ?

- utilisateur (*user* unix, commande `id`)
- utilisateur dédié (avec privilèges)
- administrateur
- (par accès réseau depuis une autre machine)

Les fonctionnements ne sont pas exclusifs, p. ex. un script lancé manuellement par l'utilisateur peut nécessiter des privilèges administrateur pour certaines tâches (via `sudo`).

Remarque : des sauvegardes multi-utilisateurs sont bien plus complexes (préserver propriétaires et droits tout au long du processus sauvegarde–restauration, gestion des nouveaux/anciens utilisateurs, droits de restauration, . . .).

Comment sauvegarder ?

Principe : dupliquer les données pour les mettre en « sécurité »

- outils « classiques » de copie de données du système (directement ou via un script)
- outil de sauvegarde (je suis incapable de conseiller)
- service distant (je suis incapable de conseiller)

Éléments clés :

- adéquation aux besoins, simplicité et robustesse
- diversité de fonctionnement : les modes manuel et automatique sont complémentaires
- tester
- documenter

Tester, re-tester, . . . ; puis utiliser

Impératif avant utilisation courante :

- restauration partielle de quelques fichiers (erreur humaine)
- restauration totale en local (ex. partition temporaire dédiée)
- restauration totale sur une *autre* machine (vol, incendie. . .)
- documenter les tests

Régulièrement, pratiquer des tests :

- restauration partielle
- restauration totale sur une autre machine (sans tricher, partir seulement des sauvegardes)
- maintenir les compétences et vérifier que tout se passe bien
- analyser l'évolution de la volumétrie

Premières questions ?

. . . avant de regarder plus de détails . . .

Première étape : préparer un dispositif de sauvegarde

- dispositif dédié aux sauvegardes (pas d'autre utilisation)
- privilégier matériel robuste (éviter : grosse volumétrie, haut débit, gros cache)
- formater (système de fichiers) pour préserver les méta-données (propriétaires, droits, groupes, ACL, etc.)
- identifier physiquement chaque dispositif et son contenu
- tester chaque dispositif (écriture gros volume)
- refroidir si test intensif ou remplissage aléatoire

Deuxième étape : sauvegarder TOUT pour restaurer !

Moins évident que cela peut paraître sans préparation...

- outils / scripts
- configurations / paramètres
- documentations
- clés de chiffrement

Éléments clés :

- ne rien oublier ! (analyser plusieurs fois dans le temps)
- redondance
- diversité technologique et spatiale
- doit être simple à utiliser, tester pour vérifier
- (1+) copie isolée sur dispositif différent des sauvegardes

Troisième étape : sauvegarde totale de base

- quoi : tout le système de fichiers / partition
- où : disque externe dédié (mis en sécurité après sauvegarde)
- qui : administrateur ou utilisateur
- quand : à la demande (avant ré-organisation, test, changement important configuration ou machine, ...)
- comment : copie totale
- documenter : date, système de fichiers, machine

Important de toujours savoir faire rapidement !

Étapes suivantes

Une fois le mécanisme de restauration et une copie totale sauvegardés, il y a des choses à améliorer :

- augmenter la redondance
- diversifier : technologique, spatial, temporel
- automatiser
- copie totale longue $\Longrightarrow$ techniques pour accélérer

Étape de fin

Important :

- vider les caches du système de fichiers du dispositif de sauvegarde
- effectuer un « démontage » (`umount`) propre du système de fichiers
- arrêter proprement le dispositif de sauvegarde
- identifier et documenter
- ranger le dispositif de sauvegarde en sécurité

Techniques

Au minimum avoir des sauvegardes « 3–2–1 » :

- 3 copies des données (redondance)
- 2 dispositifs physiques différents (diversité technologique)
- 1 copie hors site (diversité spatiale)

Il existe des techniques « ?–?–?–? » ou « ?–?–?–?–? », où les chiffres/lettres peuvent signifier des propriétés comme :

- copie hors ligne
- copie non modifiable
- copie dont l'intégrité est vérifiée (hachage plus ou moins fort)
- 0 erreur durant la sauvegarde (checksum, ECC)

Attention « ?–?–?–?–? » signifie des choses (très) variables selon les outils et services (bien lire la documentation) !

Techniques pour accélérer les sauvegardes

Sauvegarde totale (ou intégrale) :

- assez simple à mettre en œuvre
- restauration simple (juste besoin de la copie)
- mais lente et volume important

Autres techniques : incrémentale, différentielle
Attention à la définition selon les outils / services !

- copie des nouveaux fichiers
- copie des fichiers modifiés :
  - complets
  - différence avec la dernière copie complète
  - différence avec la dernière différence

Important : évaluer l'impact sur la restauration

Exemple de « calendrier » de sauvegardes

dispositif	fréquence	méthode	étendue
disque interne	jour	automatique	partition
service distant	jour	automatique	sélection
disque externe 1	semaine	manuel	sélection
clé USB 1	semaine	manuel	sélection
disque externe 2	mois	manuel + hors site	sélection
clé USB 2	mois	manuel + hors site	sélection

Adapter la fréquence : passage de hebdomadaire à 2 ou 3 fois par semaine

Organiser les données

- ne rien oublier, identifier :
  - (parties de) systèmes de fichiers à sauvegarder
  - types de fichiers à sauvegarder

- ne pas sauvegarder des « choses » inutiles (caches), fichiers générés ou facilement reconstructibles
- pour réduire le temps de sauvegarde (interruption accès)
- pour reduire le coût des dispositifs
- évaluer la volumétrie nécessaire et son évolution dans le temps
- avoir des partitions différentes aide : home | système + libs | tmp | outils installés ...

Mais sauvegarder avant de réorganiser votre machine !!!
(copie totale partition sur disque externe mis à l'abri quelques mois)

Versions précédentes de données modifiées

Ne pas confondre avec un outil de gestion de versions (p. ex. `git`) !

Possibilité dans de nombreux outils de conserver des versions précédentes de données modifiées.

Éléments clés :

- identifier clairement les versions : p. ex. suffixe et date (AAAA-MM-JJ)
- faire attention à l'augmentation du nombre de versions (nb fichiers et volume)
- faire des bilans (au moins au début)
- adopter une stratégie claire (p. ex. 1/J sur 1S, 1/S sur 1M, 1/M sur 1A)

Durée de vie des sauvegardes

- évolution de la volumétrie *vs* taille disponible
- usure des dispositifs
- limiter le nombre de copies précédentes sur un dispositif
- possibles contraintes réglementaires (min. et/ou max.)

Attention aux besoins contradictoires :

- majoritairement : suppression de données par erreur mais sauvegardes pour restaurer
- ponctuellement : suppression de données *et* suppression des sauvegardes associées

Gestion des dispositifs de sauvegarde dans le temps

Facteurs à prendre en compte :

- usure des dispositifs (complexe à évaluer et gérer)
- évolution de la volumétrie
- évolution des besoins et contraintes

Éléments :

- renouveler les dispositifs régulièrement (ex. 3–5 ans)
- diversifier
- adapter
- tester intégrité
- décommissionner ($\Longrightarrow$ gérer redondance avant)

Problématiques différentes $\Longrightarrow$ solutions différentes

Ne pas confondre :

- sauvegarde automatique (pas d'intervention humaine)
- sauvegarde manuelle

- sauvegarde totale
- sauvegarde partielle / différentielle / incrémentale

- sauvegarde
- archivage

- sauvegarde avec conservation de versions précédentes
- système de gestion de versions (`git`)

- diversité technologique
- uniformité dispositifs pour RAID, miroir, ...

Impact du système de fichiers

Problème potentiel si modification (involontaire) d'un fichier pendant sa sauvegarde dans certains systèmes de fichiers (anciens)

Moins de risque de problème avec des :

- systèmes de fichiers journalisés (`ext4`)
- systèmes de fichiers avec *copy on write* (COW) (`Btrfs`, `ZFS`)
  notion d'instantannés (*snapshots*)

Attention à bien vider les caches (du système de fichiers) avant et après chaque sauvegarde (sous linux : `sync -f DIR`)

Remarques

NAS (*Network Attached Storage*) : peut servir à certains types de sauvegardes mais attention aux caractéristiques « parasites » : partage, continuellement actif, coût, vol/incendie, exposé sur le réseau...

Attention aux outils avec des formats internes propres qui empêchent de lire les données sauvegardées avec d'autres outils.

Sécuriser physiquement le stockage les dispositifs de sauvegarde externes (boîte / coffre, EMI, température, humidité) et identifier chaque dispositif (date achat / mise en service, fréquence utilisation, contenu, machine source).

Sauvegarde chiffrée

Possibilités :

- outil de sauvegarde qui chiffre (je suis incapable de conseiller)
- outil(s) classique(s) dans une partition chiffrée (ex. LUKS)
- attention à la sauvegarde et au stockage des clés de chiffrement
- utiliser des clés différentes pour les systèmes de fichiers et les dispositifs de sauvegarde
- attention à l'évolution de la volumétrie

Identification précise des dispositifs

Attention au nom « automatique » attribué lors du montage d'un dispositif, externe et interne, comme `/dev/sdx` (avec `x` comme `a`, `b`, `c`, etc.) !

Il peut être prudent d'utiliser des identifiants stables dans le temps et en cas de restauration sur une autre machine (ex. UUID).

`/dev/disk/by-uuid/...` contient des liens vers les chemins physiques des dispositifs.

```
lsblk -f
udisksctl info -b DEVICE
udevadm info DEVICE
```

Démo : `rsync`

- copie locale / distante (p. ex. via `ssh`)
- rapide car détection modifications et envoi des différences
- capable de préserver les méta-données
- configurable (lire la documentation)
- largement disponible et éprouvé

Précautions :

- bien lire la documentation des options
- attention aux possibles changements de noms
  (`rsync ... src/ dest` *vs* `rsync ... src/ dest/`)

Démo : `unison`

- synchronise le contenu de deux systèmes de fichiers (internes, externes, distants sur le réseau), appelés réplicats
- à la demande et automatisable
- gestion d'éventuels conflits
- rapide car propage seulement les différences (`librsync`)
- robuste aux pannes

https://github.com/bcpierce00/unison (aussi dépôts Linux)

Documentations

- CNIL https://www.cnil.fr/fr/securite-sauvegarder
- Gouvernement https://www.cybermalveillance.gouv.fr/tous-nos-contenus/bonnes-pratiques/sauvegardes
- ANSSI (serveurs) https://cyber.gouv.fr/publications/fondamentaux-sauvegarde-systemes-dinformation

Mots de la fin : « anticiper » et « rigueur »

- se préparer en avance (sans pression)

- faire un bilan de ses besoins et contraintes
- évaluer des solutions techniques (simples et adaptées)
- (re-)tester pour différents types de problèmes
- documenter
- pratiquer régulièrement des restaurations (partielles et totales)
- suivre le plan mis en place
- faire un bilan de temps en temps (volumétrie, versions, outils)
- partager expériences avec d'autres personnes