

Amanda; solution de sauvegarde centralisée.

Table des matières

I	Sauvegarde physique des données.....	2
1	Introduction.....	2
1.1	Pourquoi sauvegarder ?.....	2
1.2	Quoi sauvegarder ?.....	2
1.3	Quand et comment sauvegarder ?.....	3
1.4	Support, support, mais qu'est ce donc qu'un support ?.....	5
2	Sauvegarde avec les outils POSIX.....	6
2.1	Qu'est ce que POSIX ?.....	6
2.2	Mt, mtx,.....	6
2.3	Dd, tar, cpio, dump/restore, afio, pax.....	7
2.4	Rsync.....	7
2.5	Outils spécialisés.....	7
2.6	Conclusion.....	8
3	Sauvegarde centralisée.....	8
4	Le travail réalisé.....	9
5	Consistance des données.....	13
6	Conclusion sur les sauvegardes.....	15

ISauvegarde physique des données.

1Introduction

Le petit Robert nous dit que sauvegarder provient de la réunion de sauver et de garder. Il nous indique conserver, défendre, préserver, protéger comme synonymes.

1.1 Pourquoi sauvegarder ?

Un système informatique est très souvent un constituant stratégique d'une entreprise, même s'il est fréquemment considéré comme la cinquième roue du carrosse. En effet, imaginez l'incidence du non fonctionnement du système dans les entreprises autour de vous, ne serait-ce que votre agence bancaire.

Les sauvegardes pérennisent le fonctionnement de l'entreprise, elles constituent une sorte d'assurance sur sa capacité à continuer de fonctionner.

En effet, l'arrêt du système informatique coûte à une grande entreprise entre 10 000 et 4 millions d'euros par heure (source : Contingency Planning Research), et la seule reconstitution de 20 Mo de données perdues peut coûter jusqu'à 100 000 euros (source : IDC).

Plus que jamais, il est clair que le véritable patrimoine de l'entreprise réside bien plus dans l'information que dans le matériel informatique. S'engager dans une démarche de sécurité sans s'assurer de la réelle efficacité du dispositif de sauvegarde choisi revient à partir pour un tour du monde à la voile, en solitaire, sans canot de sauvetage. Bien évidemment, on espère toujours ne pas en avoir besoin, mais prendre le risque de le laisser à quai relève plus de l'inconscience que de l'optimisme.

Le système peut être victime de plusieurs types d'atteintes à son intégrité.

Les risques - incendie, inondation, catastrophes naturelles, erreur humaine, vol, malveillance ou attaque par un virus - sont réels et en augmentation constante chaque année. Afin de prévenir ces risques, la plupart des entreprises assurent convenablement leur matériel, dont la valeur de remplacement semble toujours importante. Néanmoins, celle-ci reste marginale comparée au coût de la perte de données et de l'immobilisation du système informatique.

1.2 Quoi sauvegarder ?

Dans le cadre qui nous intéresse, c'est à dire celui des systèmes informatiques, l'acception courante est la copie d'information stratégiques sur un autre support, à fin d'éventuelles récupération. Mais sous quelles formes trouve-t-on ces données stratégiques ?

Elles sont réparties en deux familles:

1. Données physiques --> sauvegarde physique

Les données inhérentes au système lui même, typiquement les programmes qui servent à le faire fonctionner. C'est ce qui servira à l'administrateur système à redémarrer, en cas de défaillance du matériel par exemple. On trouvera aussi des termes comme « clonage »

2. Données logiques --> Sauvegarde logique

Les données liées à la fonction du système, comme les fichiers des utilisateurs, des bases de données, une comptabilité, ...

D'aucun utilise le terme de « miroir », mais on ne peut déduire de quelle famille la sauvegarde sera la lignée. Car on peut aussi bien faire un miroir physique, dupliquer intégralement un disque dur, qu'un miroir logique, qui est constitué de l'ensemble des objets formant un système de fichiers. L'ambiguïté de ce terme recommande de l'éviter.

Quelle que soit le type de données, il convient d'en évaluer la valeur pour l'entreprise. En cas de perte, les données peuvent elles être reconstituées ? Si non; que nous en coûte t-il ? Si oui; à quel prix ? De ceci va découler la ou les méthodes à choisir, et les investissements à apporter pour la réalisation.

La stricte définition de ce qu'il faut sauvegarder, sa localisation dans le système, son mode de fonctionnement, les plages horaires de servitude, etc, tout cela va imposer un cadre aux opérations de sauvegarde, et aussi à la restauration, qui est l'opération inverse. Baptisons ceci « plan de sauvegarde », et intéressons nous à son contenu.

Il sera réalisé à partir de quatre éléments essentiels.

1. Une analyse fonctionnelle:

définissant les attentes de l'entreprise en termes de besoin et de pérennité des données, sur des critères de validation, d'intégrité et de confidentialité

2. Une analyse organique:

définissant fréquence durée de rétention, cycle de rotation, besoin de réplication, mode d'identification, règles de validation, sécurisation des lieux de stockage, localisation.

3. Un plan d'exécution:

détaillant toutes les actions devant être effectuées par le service d'exploitation pour la réalisation des sauvegardes, l'annotation des supports, le contrôle d'intégrité des données, la manutention.

4. Un plan de tests de restauration:

Permettant de vérifier la qualité et la pertinence des sauvegardes, et s'assurer de l'adéquation des procédures.

1.3 Quand et comment sauvegarder ?

Une opération de sauvegarde n'est pas neutre dans un système. Elle engendre une charge des machines et du réseau, elle implique la fermeture, et donc la non utilisation des fichiers. Cela induit

que les bases de données doivent être arrêtées.

La méthode (ou type) la (le) plus simple, consiste à réaliser une sauvegarde totale du système. C'est dans cette méthode que l'on trouve les sauvegardes physiques. Cette méthode est envisageable dans le cas où le volume de données sauvegardé est faible, c'est à dire que le temps nécessaire pour réaliser la sauvegarde est inférieur à l'intervalle de temps entre deux sauvegardes, et que ce volume est compatible avec le matériel utilisé. Suivant les cas de figure, on sauvera ou la totalité des fichiers du système, ou une sélection, ceci afin d'optimiser les temps et les volumes.

Ensuite, vient la méthode différentielle, assez peu utilisée à ma connaissance. Pour la faire, on commence par réaliser une sauvegarde logique totale. Puis, on n'effectue que la sauvegarde de ce qui a été modifié de puis cette dernière sauvegarde totale. On en déduit que, si le volume de modification et/ou l'intervalle de temps entre deux sauvegardes totale est/sont important(s), le bénéfice est amoindri, voir nul.

Et donc, on utilise la méthode dite incrémentale, qui est la proche cousine de la méthode différentielle, à la différence que l'on n'effectue la sauvegarde que de ce qui a été modifié ou créé depuis la précédente sauvegarde, de quelque méthode fût-elle. Ainsi, on reste, tout au long de l'intervalle de temps entre deux sauvegarde totale, avec un volume de données restreint.

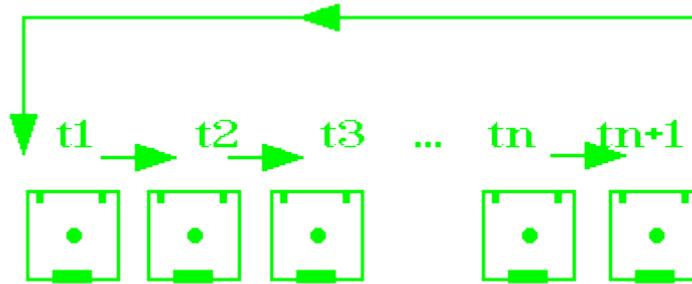
Une application courante est de réaliser une sauvegarde physique d'une machine à la fin de son installation, afin de pouvoir la reconstituer rapidement. Ensuite, lors de période de faible activité de l'entreprise, on réalise une sauvegarde logique totale, par exemple pendant la fermeture de fin de semaine, et, le soir, une sauvegarde incrémentale.

On a vu que sauvegarder consiste à dupliquer des informations, à les mettre sur un autre support. Certes, mais la gestion des support revêt une importance quasiment aussi grande que la copie sur ces supports. En effet, quelle est l'utilité d'une sauvegarde, si, le support présent à coté de la machine, un incendie ou un dégât des eaux ravage les lieux ? Ou qu'un malfrat s'introduisse dans les lieux, et dérobe ET la machine sauvegardée, ET la sauvegarde ? Ou, beaucoup plus probable, que le support ayant servi ne soit pas lisible ?

Contre ce genre de désagréments, un éventail de précautions élémentaires existe. Toutefois, il convient de les prendre AVANT une hypothétique catastrophe !

Concernant les catastrophes, si l'on a pas les moyens d'avoir les données à sauvegarder dans un lieu, le système de sauvegarde dans un autre, et de la communication entre les deux, le plus simple est d'éloigner les supports, de les isoler ailleurs, hors de portée.

Concernant la lisibilité du support, il suffit qu'il y en ait plusieurs, aurait dit monsieur De LaPallice. Et de les recycler quant elles sont toutes écrites. On introduit une rotation des supports, un cycle d'utilisation.



Cette solution induit un avantage. Il permet d'avoir les différentes versions d'un objet, au fil des jours. Vous avez 6 supports, un par jour ouvré, plus un pour le week-end, vous pouvez honorer une demande de restauration pour chaque jour de la semaine. Poussons un peu la logique du raisonnement, en alternant les supports pour le week-end, vous pouvez aussi assurer pour une semaine avant, et rien ne vous empêche de faire de même pour le mois, le trimestre, ...

Il n'est pas ridicule d'avoir 30 jeux de supports pour les sauvegardes différentielles journalières, et ~15 jeux de supports pour les sauvegardes totales hebdomadaires, ainsi que quelques jeux pour des sauvegardes physiques, ou autres.

1.4 Support, support, mais qu'est ce donc qu'un support ?

Bien sûr, on peut sauvegarder sur des disquettes, si l'on a un goût avancée pour le préhistorique, quoi que pour certaines situations, ce soit l'unique voie, je pense. particulièrement en cas de « crash recovery¹ ». Les CD, re-inscriptibles notamment, sont une solution. Plus rare, mais répondant à des besoins particuliers, comme des débits élevés, il y a l'utilisation de baies de disques durs. Pour des petits volumes, on dispose de dispositifs externes, tel les lecteurs Zip, que l'on peut connecter par un port parallèle ou USB, mais de capacité limitée, et dont le taux de transfert n'est pas très brillant.

Le support le plus courant reste la bande magnétique. Sous forme de cartouches ou cassettes, de nombreux formats se concurrencent.



DDS 4 mm



D8 8mm

Depuis l'antique bobine ½ pouce, jusqu'aux récent LTO, AIT, en passant par les très répandus QIC ¼ pouce, 4mm DAT DDS, 8mm D8 ou DLT, la bande magnétique est le support de choix pour la sauvegarde². Le DLT est très répandu, avec une base installée qui approche les deux millions de dérouleurs. Quantum, son principal promoteur, joue la carte de la compatibilité avec les cartouches de générations précédentes. Mais il est en perte de vitesse en face du LTO qui lui, joue la rupture

¹ Recours en cas de destruction

² Quoiqu'au vu de la baisse du prix du mégaoctet sur disque dur, on peut être amené à s'interroger.

technologique. Ils offrent tout deux de grandes capacités et de forts taux de transfert. L'AIT de Sony jouit d'une grande robustesse, bien supérieure à son prédécesseur DDS. Les autres formats (SLR, VXA, Travan, ...) conviennent pour des sauvegardes mono postes ou à petite échelle. Les formats cités précédemment conviennent, eux, fort bien pour des robots de sauvegarde, aussi appelés « silo », sorte de très gros meubles dans lesquels on trouve énormément de cartouches, un ou plusieurs dispositifs pour les manipuler, et quelques dérouleurs pour lire/écrire les cartouches.



Robot (source Storagetek)

Avec de tel dispositifs, certes, fort onéreux, on change d'ordre de grandeur quant aux unités. Il est courant de s'exprimer en PETAoctets pour le volume, et en TERAoctets par heure, pour les taux de transfert.

2Sauvegarde avec les outils POSIX

Ce petit chapitre est mon *credo*, car POSIX apporte la normalisation des commandes, non seulement de leur invocation, mais aussi, entre autres, des formats. Cela rend les travaux réalisés conformément aux normes POSIX déployables entre systèmes hétérogènes.

2.1 Qu'est ce que POSIX ?

POSIX signifie Portable Operating System Interface. C'est un comité normatif issu de l'IEEE³. Il a pour but de normaliser les langages, les interfaces, ... afin de faciliter les évolutions, l'interopérabilité, ...

³ IEEE = Institute of Electrical and Electronic Engineers

2.2 *Mt, mtx,*

Les utilitaires *mt* et *mtx* servent à la manipulation des supports. *Mt* permet de retendre la cartouche, de la positionner à un endroit précis, de la formater, de positionner des options telle la compression matérielle du dérouleur, si celui ci en est équipé. *Mtx* est une extension de *mt*, car il sert à manipuler les supports au sein des robots. Son action se limite à sortir une cartouche du dérouleur, ou à y déposer la cartouche numéro N, ou à appeler la suivante.

2.3 *Dd, tar, cpio, dump/restore, afio, pax*

Ce sont les utilitaires destinés à lire ou écrire les données sur les supports. L'ordre suivant lequel ils sont présentés va du plus élémentaire au plus sophistiqué. Sans aucun caractère péjoratif, chacun ayant ses avantages et ses inconvénients.

- Par exemple, *dd* peut absolument tout lire ou écrire, puisqu'il est capable de le faire dans la plus petite unité (l'octet) que peut manipuler la machine, et agir sur des éléments physiques. Par contre, ceci est au prix de beaucoup de lourdeur.
- *Tar* agrège des fichiers et leurs attributs en une seule archive, que ce soit dans un autre fichier, ou sur un support. C'est cet utilitaire qui offre le plus de portabilités d'une architecture à l'autre. *Cpio* est une alternative à *tar*, et choisir l'un ou l'autre peut facilement relever de la querelle de chapelles. Ce sont tant les contraintes particulières à une situation que la culture des personnels qui feront choisir une solution ou une autre.
- *Dump* et *restore* apporte un degré de sophistication, dans le sens où ils sont moins destinés à sauvegarder des systèmes de fichiers complets, qu'à apporter une solution à la gestion des versions des objets sauvegardés. Un historique de la sauvegarde est écrit sur le disque, avec les attributs du moment et des précédentes sauvegardes. Ceci, et l'interpréteur de commande interactif intégré à *restore*, offre la possibilité de rechercher une version précise d'un objet dans l'historique.
- *Afio* ou *pax* sont des outils généralistes, dont le but et l'interopérabilité des supports créés avec *tar* et avec *cpio*. D'initiative récente, ils n'ont pas encore rencontrés le succès auquel ils sont destinés. Toutefois, *pax* est porté par un désir de normalisation, désir en provenance d'un comité POSIX.

2.4 *Rsync*

Rsync n'est pas à proprement parler un outil de sauvegarde. Mais il dispose de caractéristiques tellement utiles, qu'il leur est souvent adossé. C'est une évolution d'une ancienne commande peu sécurisée⁴, qui effectue des copies à distance. Son atout réside sur son utilisation du protocole SSH qui sécurise les transmissions, dans un algorithme de comparaison, qui évite la copie de fichiers s'ils sont identiques sur la source et la destination, la compression par fichier, et non pour l'ensemble.

4 rcp

2.5 Outils spécialisés

Unison, sitecopy sont, entre autres, des programmes qui permettent de faire des copies entre deux machines, même si les systèmes d'exploitation sont différents. Un lot d'options permettent un contrôle fin de ce que l'on copie ou supprime sur telle ou telle machine. Par contre ces outils sont monolithiques et ne sont pas interopérables. Leur but est principalement de faire des « miroirs » entre deux machines, et peuvent précéder une sauvegarde.

2.6 Conclusion

Il n'est guère possible d'indiquer *une* méthode de sauvegarde/restauration par outils POSIX, car il en existe beaucoup plus que de machines à sauvegarder !

Chaque administrateur va concevoir une ou plusieurs solutions adaptées au cas de figure qui se présente devant lui. Cette démarche fut utilisée au Conseil Régional à plusieurs reprises, notamment pour sauvegarder dans l'urgence une machine distante.

3 Sauvegarde centralisée

Il existe quelques programmes de sauvegarde centralisée. Nous allons voir les principaux.

- BRU
L'application BRU (Backup and Restore Utility) est une application de sauvegarde et restauration très populaire. Il apporte quelques facilités d'utilisation, comme la sauvegarde par utilisateur. BRU dispose aussi d'une possibilité de remise en État suite à problème, qui est conçu pour remettre en service aussi rapidement que possible un système ayant subi une perte de données catastrophique. Les archives BRU sont portables sur d'autres plate-formes, comme Windows NT. Hélas, BRU n'est pas disponible suivant les termes de la GPL⁵.
- Arkeia
Arkeia est une solution très célèbre. Elle est imaginé dès le départ pour gérer la sauvegarde sur de très grands réseaux. Elle contient des agents très solides, reconnaît plusieurs protocoles, et

⁵ General Public Licence, base du logiciel libre, voir <http://www.gnu.org>

s'adapte sur de nombreuses architectures. Un monitoring du réseau est incorporé pour préserver les ressources. Bien que n'étant pas un logiciel libre, il est possible de l'utiliser sans problème de licence sur une configuration incluant une machine Arkeia, et deux postes de travail faisant fonctionner des agents.

- Amanda
AMANDA (Advanced Maryland Automatic Network Disk Archiver) est un outil de sauvegarde en provenance de l'université du Maryland. Jaime Da Silva, comme tout administrateur système qui se respecte, avait une collection de scripts sur de multiples machines. Devant l'accroissement de la tâche à accomplir, il fédéra cet ensemble dans Amanda.

Dans cet ensemble, c'est Amanda qui correspond le mieux aux désirs de l'équipe du Conseil Régional. Et Amanda y est déjà utilisé.

- Amanda est un logiciel sous licence GPL,
- Il est basé sur des outils standards (dump, restore, tar, ...)
- Son architecture modulaire répartit, tel un chef d'orchestre, les tâches sur des machines périphériques. Par exemple, le serveur peut ne pas accueillir de dérouleur.
- Il implémente une gestion des bandes simple et efficace: on ne peut effacer une bande par erreur, et il existe des outils de recherche de fichier dans toute la collection de support.
- Interface générique des changeurs de bandes, donc, non seulement on peut utiliser des dérouleurs différents, mais aussi des robots différents.
- Si besoin, peut encrypter ou compresser tant les communications avec les agents de sauvegarde, que les supports eux-mêmes
- Ajuste l'ordonnancement par des règles d'optimisation
- Intègre des outils de rapport, hautement configurables et modifiables,
- Autorise l'inclusion de tout module que l'on désire

4Le travail réalisé.

Ainsi, Amanda effectuait son travail depuis plus de deux ans, sans heurts. Les rapports générés indiquait depuis peu une nette augmentation du temps de sauvegarde, et le travail a consisté dans une première phase à trouver le pourquoi, dans une seconde phase à y remédier, et, enfin, d'essayer d'optimiser.

Dans un premier temps, une mise à jour du logiciel a été faite, ainsi que l'installation d'un agent sur une machine de test. Ceci a permis de faire des petits tests, avec de petits volumes de données. La configuration matérielle étant figée, les sources de données à sauvegarder également, nous ne pouvions qu'agir sur la méthodologie opératoire d'Amanda. A cette fin, nous n'avons besoin que des seuls fichiers de configuration.

Le principal, `amanda.conf`, ne fait que décrire les options générales, ainsi que les spécificités du serveur de sauvegarde, comme la gestion de l'historique, la façon dont sont nommés les cartouches, la description des règles de sauvegarde (dumptype), etc.

Le fichier `tapelist` décrit les caractéristiques techniques des dérouleurs de bande. Ici, un seul

dérouleur de 12 giga-octets est utilisé, ce qui, aujourd'hui, est suffisant, mais n'autorise guère de marge de manoeuvre. Donc, aucune optimisation possible par cette voie.

Le fichier `filelist` quant à lui décrit ce qu'il faut sauvegarder, et la méthode pour ce faire. La méthode doit correspondre au nom d'une règle présente dans le fichier de configuration général, `amanda.conf`. Par contre, il peut être surchargé par de nouvelles options, voir modifier des options définies dans les règles.

Afin de rentabiliser l'utilisation des matériels, le serveur Amanda fonctionne sur une machine ayant peu de charge la nuit, mais beaucoup en journée. Les essais en grandeur réelle, c'est à dire incluant tous les systèmes de fichiers, n'ont donc pu être faits qu'une fois par nuit. Ainsi la phase d'optimisation a duré quelques jours. Voici des extraits de trois rapports.

On peut y voir successivement:

Le temps de fonctionnement

La taille des données compressées sur la cartouche,

La taille réelle de ces données,

Le taux de compression moyen obtenu

Le nombre de systèmes de fichiers sauvés

Le débit obtenu pour les données compressées.

Run Time (hrs:min)	13:02		
Output Size (meg)	5752.7	5752.7	0.0
Original Size (meg)	15164.5	15164.5	0.0
Avg Compressed Size (%)	37.9	37.9	--
Filesystems Dumped	19	19	0
Avg Dump Rate (k/s)	126.6	126.6	--

Run Time (hrs:min)	7:03		
Output Size (meg)	5963.2	5963.2	0.0
Original Size (meg)	15280.3	15280.3	0.0
Avg Compressed Size (%)	39.0	39.0	--
Filesystems Dumped	19	19	0
Avg Dump Rate (k/s)	246.2	246.2	--

Run Time (hrs:min)	3:28		
Output Size (meg)	9095.5	9095.5	0.0
Original Size (meg)	15553.3	15553.3	0.0
Avg Compressed Size (%)	36.7	36.7	--
Filesystems Dumped	19	19	0
Avg Dump Rate (k/s)	824.0	824.0	--

On constate que, malgré une très légère augmentation du volume de données, le temps de sauvegarde a été réduit de façon drastique. Pour ce faire, il a fallu décider si la compression des données devait être réalisée côté agent, ou côté serveur. Comme plusieurs machines font partie du jeu de sauvegarde, le cas est différent pour chacune d'elle. De plus, les données en elles même influencent le temps et le taux de compression. Si, par exemple, ce sont déjà des fichiers compressés, la machine va consacrer beaucoup de temps pour un résultat nul. Et inversement si les données sont un ensemble de fichiers de type texte.

Observons la partie détaillée d'un rapport.

(Les noms des machines ont été remplacés par xxx)

DUMP SUMMARY:

HOSTNAME	DISK	DUMPER STATS				TAPER STATS			
		L	ORIG-KB	OUT-KB	COMP%	MMM:SS	KB/s	MMM:SS	KB/s
xxx /aquarel	0	1058399	641056	60.6	114:22	93.4	114:22	93.4	
xxx /home1	0	1276319	241760	18.9	102:04	39.5	102:11	39.4	
xxx /home2	0	203647	158464	77.8	13:25	196.9	13:25	196.8	
xxx /home3	0	18412791	363360	74.0	122:48	185.0	122:48	185.0	
xxx /home5	0	2572127	487328	18.9	212:56	38.1	212:56	38.1	
xxx /home6	0	1076223	128224	11.9	54:03	39.5	54:03	39.5	
xxx /home7	0	256575	73376	28.6	15:18	79.9	15:26	79.2	
xxx /var	0	545151	192896	35.4	57:26	56.0	57:36	55.8	
xxx /etc	0	7940	1344	16.9	0:07	186.0	0:07	185.9	
xxx /home	0	1636640	444864	27.2	21:54	338.5	21:54	338.5	
xxx /home2	0	1250260	980896	78.5	19:50	824.1	19:50	824.0	
xxx /home3	0	632980	144544	22.8	4:36	524.1	4:37	522.7	
xxx /root	0	387030	46144	11.9	3:06	248.5	3:06	248.3	
xxx -ar/backups	0	412350	193728	47.0	5:40	569.9	5:40	569.8	
xxx /var/log	0	566480	110848	19.6	4:03	457.1	4:03	456.9	
xxx /var/mail	0	1511860	594752	39.3	19:53	498.4	19:53	498.4	
xxx /var/spool	0	242730	75264	31.0	3:17	382.0	3:17	381.8	
xxx /var/www	0	49740	12096	24.3	0:43	277.6	0:44	276.9	
xxx -ar/www-ssl	0	700	160	22.9	0:00	322.2	0:01	275.3	

On y observe que /home5 est sauvegardé en deux heures et demie, mais avec un taux de compression faible.

Nous créons de nouveaux dumptypes, et affectons, les systèmes de fichiers qui prennent le plus de temps à ces nouveaux dumptypes.

extrait du fichier disklist:

```
# File format is:
#
# hostname diskdev dumptype
#
# where the dumptypes are defined by you in amanda.conf.
### BMR, le nn juillet 2003
## xxx: passe de comp-medium a cra_tar-client-fast
xxx /root cra_tar-client-fast
xxx /etc cra_tar-client-fast
xxx /var/www cra_tar-client-fast
xxx /var/www-ssl cra_tar-client-fast
xxx /var/spool cra_tar-client-fast
xxx /var/backups cra_tar-client-fast
xxx /var/log cra_tar-client-fast
xxx /var/mail cra_tar-client-fast
xxx /home cra_tar-client-fast
xxx /home2 cra_tar-client-fast
xxx /home3 cra_tar-client-fast
### BMR, le nn juillet 2003,
## xxx: /home5, /home3, /aquarel, /home1, /var, /home6
## passent de comp-dump a cra_dump-client-fast
### BMR, le nn juillet 2003
## xxx: Les autres aussi
### BMR, le nn juillet 2003
## xxx: /home3 = 1800Mo vers 1400Mo en 70 min
## xxx: /aquarel = 1000Mo vers 660Mo en 37 min
## xxx: /home5 = 2600Mo vers 560Mo en 52 min
# xxx /aquarel cra_dump-client-fast
# xxx /home3 cra_dump-client-fast
# xxx /home5 cra_dump-client-fast
xxx /aquarel cra_dump-no-compress
xxx /home3 cra_dump-no-compress
xxx /home5 cra_dump-no-compress
xxx /var cra_dump-client-fast
xxx /home1 cra_dump-client-fast
xxx /home2 cra_dump-client-fast
xxx /home6 cra_dump-client-fast
xxx /home7 cra_dump-client-fast
```

Extrait de amanda.conf, où l'on définit, entre autres, les types de sauvegarde, les dumptypes précédemment cités.

```
#
# amanda.conf - sample Amanda configuration file.
#

.../...

define tapetype DDS-3 {
    comment "Digital Data Storage Format"
    length 11631 mbytes
    filemark 434 kbytes
    speed 1139 kbytes
}

.../...

define dumptype global {
    comment "Global definitions"
    # This is quite useful for setting global parameters, so you don't have
    .../...

    ### BMR, le nn juillet 2003
    ### La machine ne connait le programme dump,
    ### qui est le programme par défaut.
    ### Donc, utilisation de tar
}

# sauvegarde avec GNUTAR, compression maximum et index, sans sauvegarde temporaire sur
disque

define dumptype comp-medium {
    program "GNUTAR"
    comment "tar with compression"
    options compress-best, index, no-hold
    priority medium
}

define dumptype comp-dump {
    comment "dump with compression"
    options compress-best, index, no-hold
    priority medium
}

define dumptype nocomp-dump {
    comment "dump without compression"
    options no-compress, index, no-hold
}

define dumptype nocomp-medium {
    program "GNUTAR"
    comment "tar without compression"
    options no-compress, index, no-hold
    priority medium
}

define dumptype cra_tar-client-fast {
    global
    comment " Tar, avec compression fast cote client"
    program "GNUTAR"
    compress client fast
    options index
    priority medium
}

define dumptype cra_tar-client-best {
    global
    comment " Tar, avec compression best cote client"
    program "GNUTAR"
    compress client best
    options index
}
```

```

    priority medium
}

define dumptype cra_dump-client-fast {
    global
    comment " Dump, avec compression fast cote client"
    compress client fast
    options index
    priority medium
}

define dumptype cra_dump-client-best {
    global
    comment " Dump, avec compression best cote client"
    compress client best
    options index
    priority medium
}

define dumptype cra_dump-no-compress {
    global
    comment " Dump, sans compression"
    compress none
    options index
    priority medium
}

```

Et nous obtenons successivement les rapports suivants:

Run Time (hrs:min)	7:03		
Output Size (meg)	5963.2	5963.2	0.0
Original Size (meg)	15280.3	15280.3	0.0
Avg Compressed Size (%)	39.0	39.0	--
Filesystems Dumped	19	19	0
Avg Dump Rate (k/s)	246.2	246.2	--

Run Time (hrs:min)	3:28		
Output Size (meg)	9095.5	9095.5	0.0
Original Size (meg)	15553.3	15553.3	0.0
Avg Compressed Size (%)	36.7	36.7	--
Filesystems Dumped	19	19	0
Avg Dump Rate (k/s)	824.0	824.0	--

En examinant le dernier rapport détaillé, on voit bien ce sur quoi ont porté les efforts, ce sont particulièrement les gros volumes, et les taux de compression faible. Lorsqu'il n'y a utilisation de la compression, le rapport détaillé dispose – en lieu et place du taux de compression.

xxx /aquarel	0	10587201058720	--	13:431286.9	13:431286.6
xxx /home1	0	1276319 286400	22.4	24:58 191.1	24:59 191.1
xxx /home2	0	203647 159680	78.4	8:01 332.0	2:50 938.6
xxx /home3	0	18537921853792	--	27:111136.3	27:121136.0
xxx /home5	0	25721282572128	--	18:002382.5	18:002381.6
xxx /home6	0	1076223 158336	14.7	17:34 150.2	2:48 944.1
xxx /home7	0	256575 78944	30.8	5:44 229.4	1:34 843.4
xxx /var	0	554687 207936	37.5	14:21 241.6	3:38 952.2
xxx /etc	0	7990 1664	20.8	0:07 243.5	0:021013.6
xxx /home	0	1698890 522208	30.7	12:23 702.5	12:24 702.3

```

xxx /home2      0 1251030 996032 79.6 17:18 960.0 17:18 959.3
xxx /home3      0 675080 170368 25.2 3:38 780.5 2:59 950.9
xxx /root       0 386160 57408 14.9 2:45 347.2 1:03 909.3
xxx -ar/backups 0 412350 205568 49.9 2:261408.6 3:37 949.5
xxx /var/log    0 608490 134464 22.1 3:43 602.6 2:23 942.3
xxx /var/mail   0 1713010 743232 43.4 12:45 970.9 12:46 969.6
xxx /var/spool  0 270270 92896 34.4 3:04 503.4 1:42 915.1
xxx /var/www    0 50500 14048 27.8 0:41 341.3 0:141004.9
xxx -ar/www-ssl 0 700 192 27.4 0:00 662.2 0:003408.8

```

Il est possible d'optimiser encore, pour réduire le temps de sauvegarde, mais, vu l'augmentation croissante du volume de données à sauvegarder, la prudence impose de laisser une marge de sécurité.

5 Consistance des données.

Il convient de sauvegarder des données cohérentes. C'est à dire, n'ayant pas subi de modifications entre le début et la fin de l'opération. Pour atteindre ce but, on procède à une sauvegarde des bases de données, en bloquant toute écriture sur les bases. Un annuaire LDAP est également sauvegardé de cet manière, car, sans être véritablement une base de données, il s'en approche fortement.

Donc, avant le lancement de la sauvegarde, un utilisateur muni des droits suffisants, procède à quelques opérations pour fabriquer un jeu de données dans un État satisfasse ces besoins. Un script réalise cette opération. Il est simplement lancé par l'ordonnanceur cron.

Les valeurs sensibles ont été remplacées par xxx

```

#! /bin/bash

# Sauvegarde des bases MySql et des annuaires LDAP

# les variables :
# u      comme user
# ts     comme timestamp
# tsd    comme timestamp date
# bd     comme base directory
# wd     comme working directory
# td     comme temporary directory
# t1da  comme temporary 1 days ago
# t2da  comme temporary 2 days ago
# t7da  comme temporary 7 days ago

#u=xxx
ts=`date +%Y%m%d_%H%M%S`
tsd=`date +%Y%m%d`

```

```

t1da=`date --date '1 days ago' +%Y%m%d`
t2da=`date --date '2 days ago' +%Y%m%d`
t7da=`date --date '7 days ago' +%Y%m%d`
bd=/xxx/xxx/
wd=$bd$tsd/$ts/

# Purge des anciennes sauvegardes
# mail pour verifier
#echo ${bd}${t7da} | mail xxx -s "xxx fait sauve.sh" xxx@xxx.tld
rm -rf ${bd}${t7da}

# Creation d'un repertoire de travail horodate
mkdir -p $wd
cd $wd

# Copie de quelques fichiers essentiels
td=${wd}etc/
mkdir -p $td
# les fichiers relatifs a xinetd
cp /etc/xinetd.conf $td

# les fichiers relatifs a pam
cp /etc/pam.conf $td
cp /etc/pam_ldap.conf $td
cp -r /etc/pam.d $td

# les fichiers relatifs a LDAP
cp -r /etc/ldap $td
cp -r /etc/ldap.save $td
cp /etc/libnss-ldap.conf $td
cp -r /etc/openldap $td
/usr/sbin/slappcat -l ${wd}cra.ldif

# les fichiers relatifs au courrier
cp -r /etc/postfix $td
cp /etc/qpopper.conf $td

# les fichiers relatifs a MySQL
cp -r /etc/mysql $td
/usr/bin/mysqldump -u xxx -pxxx --all-databases > mysqldump.txt

```

Rien de bien compliqué en somme.

6 Conclusion sur les sauvegardes

De l'évaluation des besoins de l'entreprise, à la gestion quotidienne des supports, nous voyons bien que la mise en oeuvre d'une solution de sauvegarde est loin d'être triviale.

On ne connaît la validité de la solution de sauvegarde, uniquement lorsqu'on y a recours. Afin d'éviter d'amères déceptions, il convient de vérifier régulièrement la validité des informations préservées, ainsi que la pertinence des procédures. Ceci peut avoir lieu lors de simulations poussées de reprise après incident, de continuité de service. Bien entendu, une planification rigoureuse, un plan de secours est nécessaire pour y aboutir !

