

## Rsync: le meilleur des systèmes de sauvegarde



par Brian Hone  
<bhone(at)eink.com>



### *L'auteur:*

Brian Hone est administrateur système et développeur dans la société E Ink corporation. Il fait du surf pendant son temps libre.

*Traduit en Français par:*  
Jean Peyratout  
<jean.peyratout-at-abul.org>

### *Résumé:*

La sauvegarde ou "backup" est l'une des plus difficiles et l'une des plus négligées des tâches d'administration système. C'est aussi l'une des plus importantes. C'est le dernier rempart de défense contre les pannes matérielles, des brèches de sécurité et le danger ultime: les utilisateurs. Bien sûr il existe de nombreux systèmes de backup coûtant plusieurs milliers d'euros qui archivent les données sur de coûteux lecteurs de bande et utilisent du logiciel propriétaire cher et farci de bugs; mais on peut faire mieux, en utilisant rsync et une pile de disques durs bon marché.

---

## Le problème

Je peux vous énumérer longuement les raisons pour lesquelles la sauvegarde est le cauchemar des administrateurs-système. Si vous êtes vous-même administrateur, c'est probablement inutile. Voici quelques-une de ces raisons : le matériel, cher, qui est plus souvent en panne qu'opérationnel, les logiciels coûteux dont la gestion tourne au cauchemar, les longues heures passées à restaurer de multiples versions de fichiers. Pour rendre les choses encore plus difficile, les sauvegardes ont rarement une haute priorité en entreprise – jusqu'au jour où, inévitablement, on en a vraiment besoin. Si vous avez déjà eu à faire des sauvegardes/restaurations, il y a des chances que vous ayez eu une conversation de ce genre :

**Utilisateur:** "J'ai perdu un fichier. J'ai besoin de le récupérer. C'est urgent."

**SysAdmin:** "D'accord. Quel nom, le fichier ?"

**Utilisateur:** "Je ne sais pas, je pense qu'il y a un 'e' dans le nom."

**SysAdmin:** "Bon. Dans quel répertoire était-il ?"

**Utilisateur:** "Je ne sais pas, peut-être dans l'un de ces trois-là..."

**SysAdmin:** "\*Soupir\* Quelle date, la dernière utilisation ?"

**Utilisateur:** "Et bien.... Je crois que c'était un mardi ; en février ... ou en avril. Mais c'est quoi, le problème ? Je pensais que quelqu'un comme vous avait mis en place un système pour faire face à ce genre de choses."

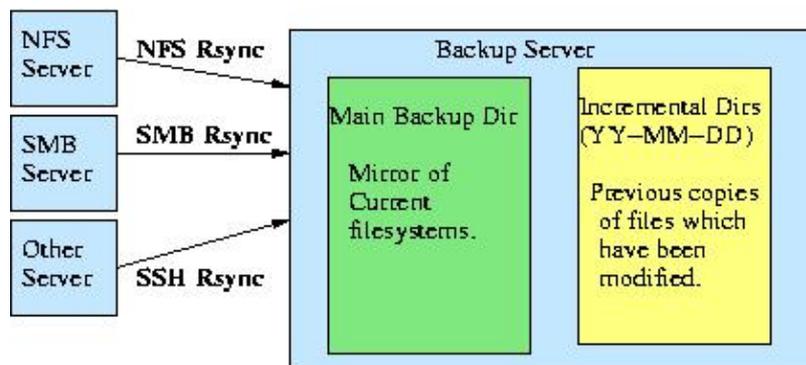
## L'alternative rsync

Rsync est une puissante implémentation d'un beau petit algorithme. Cette puissance vient d'abord de sa capacité à faire un "miroir" efficace d'un système de fichiers. En utilisant rsync, il est facile de paramétrer un système qui conservera une copie à jour d'un système de fichiers en utilisant un ensemble souple de protocoles réseau comme nfs, smb ou ssh. La seconde fonctionnalité de rsync exploitée par ce système de sauvegarde est sa capacité à archiver d'anciennes copies de fichiers modifiés ou effacés. Il existe bien trop de fonctions pour que cet article sur rsync soit exhaustif, je vous conseille donc de poursuivre votre lecture sur <http://rsync.samba.org>.

## Le système

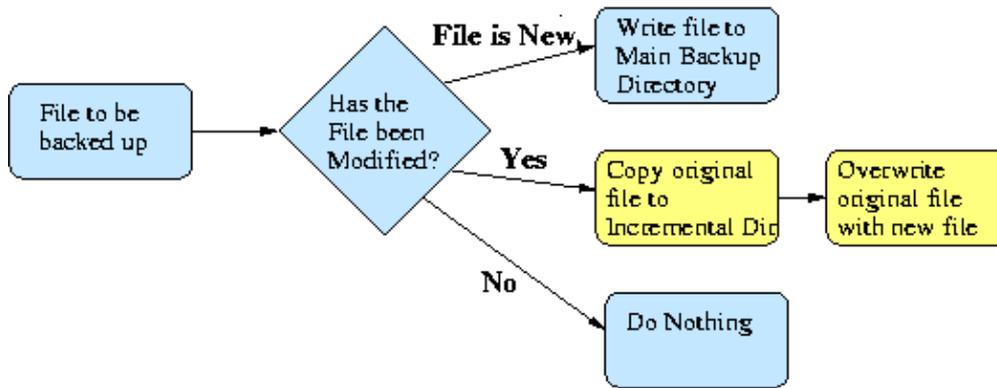
En bref, ce système utilise un ordinateur bon marché sous Linux avec plusieurs disques bon marché et un petit script shell qui appelle rsync. **[Fig 1]** Pour faire une sauvegarde, on demande à rsync de créer un répertoire appelé 'AA-JJ-MM' qui servira à stocker les modifications incrémentales. Ensuite rsync examine les changements intervenus sur les serveurs à sauvegarder. Si un fichier a changé, il copie la version ancienne dans le répertoire incrémental puis copie la version actualisée du fichier dans le répertoire principal de sauvegarde. **[Fig 2]**

**Figure 1: Structure of an Rsync Backup Server**



En général, les modifications quotidiennes ne représentent qu'un petit pourcentage du système de fichiers total. J'estime que le pourcentage moyen ne dépasse pas 0,5% à 1%. Par conséquent vous pouvez, avec un jeu de disques de sauvegarde d'une capacité double de celle des disques à sauvegarder, conserver 50 à 100 jours de sauvegardes incrémentales sur disque dur. Lorsque le disque commence à se remplir, basculez simplement sur un nouveau jeu de disques et conservez l'ancien à l'abri, à l'extérieur du site. En pratique, il est possible de conserver plus de six mois de sauvegarde incrémentale sur disque. En fait, si vous trouvez de l'espace libre quelque part, vous pouvez copier vos sauvegardes incrémentales sur un autre serveur avant d'effectuer la rotation de disques. De cette façon il est possible de conserver sur disque un aussi grand nombre de sauvegardes que vous voulez.

**Figure 2: Backup of a file using Rsync**



## Avantages: Récupération en cas de crash et restauration de fichiers deviennent faciles

Revenons à notre conversation imaginaire de tout à l'heure. Maintenant, au lieu d'un encombrant système à base de bandes, imaginez six mois de sauvegardes incrémentales qui vous attendent tranquillement sur votre machine Linux. En utilisant la combinaison favorite locate/find/grep, il est possible de trouver toutes les occurrences des fichiers de notre utilisateur imaginaire contenant un 'e' et dont la date correspond à un mardi en février ou avril, puis de les copier dans son répertoire personnel. Quant au problème de retrouver la bonne version, c'est devenu le genre de problème que je préfère : celui de quelqu'un d'autre que moi.

Ensuite, imaginons notre scénario favori – la panne complète. Supposons qu'un gros serveur nfs/samba vous lâche. Et bien, si vous avez sauvegardé vos configs samba, vous pouvez remonter votre serveur de sauvegarde en lecture seule en quelques minutes seulement. Essayez toujours avec des bandes.

## Pourquoi une sauvegarde rsync/disque dur est un meilleur choix que des bandes

	Sauvegarde sur bande	Rsync
Coût	Très élevé	Faible
Sauvegarde complète	Rapide	Rapide
Sauvegarde incrémentale	Rapide	Rapide
Restauration complète	Très lente, probablement plusieurs bandes	Rapide – tout est sur disque dur
Restauration de fichier	Lente, souvent sur de multiples bandes, fréquente difficulté à trouver la version correcte.	Très rapide – tout est sur disque et vous disposez de la puissance des outils de recherche des UN*X tels que find, grep et locate
Panne complète	La seule option est la restauration complète	Possibilité de redémarrer avec la sauvegarde comme

		serveur de fichiers en un claquement de doigts.
--	--	---

## Les outils

Il y a de nombreuses façons de configurer cela. Tous les outils présentés sont open-source, inclus dans les distributions standard, et très souples d'emploi. Nous décrivons ici une configuration possible, mais c'est loin d'être la seule.

- **Le serveur:** J'utilise RedHat Linux. N'importe quelle distribution conviendra, comme le ferait un UN\*X quelconque (je l'ai même fait avec Mac OS X). Un avertissement: la présence de beaucoup de RAM est une aide précieuse.
- **Disque:** La façon la plus facile que nous ayons trouvée pour bâtir une pile de disques bon marché est une carte PCI firewire connectée à un groupe de disque IDE dans des boîtiers firewire externes. Configurer Linux pour les utiliser comme une seule grosse partition RAID se fait sans trop de peine.
- **Le logiciel:** Rsync est un outil génial. C'est une sorte de couteau suisse du miroir de système de fichiers. Pour en savoir plus, allez voir [rsync.samba.org](http://rsync.samba.org).
- **Connexion aux serveurs de fichiers:** Rsync est très souple d'emploi. Nous utilisons nfs et smbfs. Vous pouvez aussi utiliser le propre protocole réseau de rsync en lançant un démon rsync sur le serveur de fichiers. Vous pouvez également demander à rsync d'utiliser ssh pour sauvegarder de façon sûre des sites à distance. Voyez les ressources ci-dessous pour des informations sur la configuration de ces connexions.

## élaborer le script

La base de ce script vient du site Web de rsync. Il n'y a en réalité qu'une seule commande:

```
rsync --force --ignore-errors --delete --delete-excluded --exclude-from=exclude_file --backup
--backup-dir=`date +%Y-%m-%d` -av
```

Les options clés 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.

Le script qui suit peut être lancé chaque nuit en utilisant le système cron intégré à Linux. Pour lancer le script à 23 h chaque soir, utilisez la commande "crontab -e", puis tapez ce qui suit:

```
0 23 * * * /chemin/de/votre/script
```

## Le script complet

Voici donc mon script shell pour tout lier ensemble. Une fois de plus, il y a d'autres manières de faire ceci. C'est juste une façon de s'y prendre.

```
#!/bin/sh

#####
# Script to do incremental rsync backups
# Adapted from script found on the rsync.samba.org
# Brian Hone 3/24/2002
# This script is freely distributed under the GPL
#####

#####
# Configure These Options
#####

#####
# mail address for status updates
# - This is used to email you a status report
#####
MAILADDR=your_mail_address_here

#####
# HOSTNAME
# - This is also used for reporting
#####
HOSTNAME=your_hostname_here

#####
# directory to backup
# - This is the path to the directory you want to archive
#####
BACKUPDIR=directory_you_want_to_backup

#####
# excludes file - contains one wildcard pattern per line of files to exclude
# - This is a rsync exclude file. See the rsync man page and/or the
#   example_exclude_file
#####
EXCLUDES=example_exclude_file

#####
# root directory to for backup stuff
#####
ARCHIVEROOT=directory_to_backup_to

#####
# From here on out, you probably don't #
# want to change anything unless you #
# know what you're doing. #
#####

# directory which holds our current datastore
CURRENT=main

# directory which we save incremental changes to
INCREMENTDIR=`date +%Y-%m-%d`

# options to pass to rsync
OPTIONS="--force --ignore-errors --delete --delete-excluded \
--exclude-from=$EXCLUDES --backup --backup-dir=$ARCHIVEROOT/$INCREMENTDIR -av"
```

```

export PATH=$PATH:/bin:/usr/bin:/usr/local/bin

# make sure our backup tree exists
install -d $ARCHIVEROOT/$CURRENT

# our actual rsyncing function
do_rsync()
{
    rsync $OPTIONS $BACKUPDIR $ARCHIVEROOT/$CURRENT
}

# our post rsync accounting function
do_accounting()
{
    echo "Backup Accounting for Day $INCREMENTDIR on $HOSTNAME:">/tmp/rsync_script_tmpfile
    echo >> /tmp/rsync_script_tmpfile
    echo "#####">>/tmp/rsync_script_tmpfile
    du -s $ARCHIVEROOT/* >> /tmp/rsync_script_tmpfile
    echo "Mail $MAILADDR -s $HOSTNAME Backup Report < /tmp/rsync_script_tmpfile"
    Mail $MAILADDR -s $HOSTNAME Backup Report < /tmp/rsync_script_tmpfile
    echo "rm /tmp/rsync_script_tmpfile"
    rm /tmp/rsync_script_tmpfile
}

# some error handling and/or run our backup and accounting
if [ -f $EXCLUDES ]; then
    if [ -d $BACKUPDIR ]; then
        # now the actual transfer
        do_rsync && do_accounting
    else
        echo "cant find $BACKUPDIR"; exit
    fi
else
    echo "cant find $EXCLUDES"; exit
fi

```

## Ressources

- Rsync: <http://rsync.samba.org>
- NFS: <http://nfs.sourceforge.net/nfs-howto>
- SMBFS: <http://samba.org>
- Linux RAID: <http://linas.org/linux/raid.html>

Site Web maintenu par l'équipe d'édition LinuxFocus  
 © Brian Hone  
 "some rights reserved" see [linuxfocus.org/license/](http://linuxfocus.org/license/)  
<http://www.LinuxFocus.org>

Translation information:  
 en --> -- : Brian Hone <bhone(at)eink.com>  
 en --> fr: Jean Peyratout <jean.peyratout-at-abul.org>