

Rsync based backups

Rsync is essentially a directory syncing program.

This presentation is about using it as the basis of a disk based backup system that can store multiple incremental backups using a minimal overhead of disk space.

whoami

I am Kevin Korb

I am the Systems Administrator at FutureQuest.net. We are a shared server web hosting company that specializes in Customer Service, Unique features, and not over-selling our capabilities.

I am also very active in the Greater Orlando Linux User's Group (golug.org). GOLUG meets every month for a technical presentation by volunteer members followed by socializing over dinner.

When/Why would you use rsync for backups?

- Disk based – no tapes
- Uses Secure Shell for all network communications
- Only backs up changes – Only do a full backup once EVER
- Very simple restore procedure
- There are pre-made backup systems out there that use rsync OR you can make your own using basic scripting.

What is so special about rsync?

It all comes down to hard links. Rsync uses UNIX style hard links to put the same file into two different directories at the same time without consuming any additional disk space.

This means you can backup the same stuff over and over again each time to a different directory on the backup system. In each backup directory there will be a complete tree of what was there at the time of the backup and files that are the same on multiple backups will be the same files.

When/Why wouldn't you use rsync for backups?

- Databases and VM images – Rsync can't handle databases on its own (we will discuss this in more detail)
- Windows – Rsync can backup data from Windows systems but it can't backup the Windows OS
- Compression/Encryption – Rsync stores backups in the file system as is. It does not handle encrypted or compressed storage on its own.

Why not just use RAID, DRBD, or some other mirroring system?

A mirror is not a backup. Especially if the mirror is updated in real time.

The purpose of RAID and mirroring is to protect you from hardware failures. Backups also protect you from file deletions and corruptions whether by accident or on purpose.

My philosophy is that if I don't have a backup copy of something then it must be a temp file.

How to organize backups

- Forget everything you know about yearly/monthly/weekly/daily/hourly backups. These are obsolete concepts that were forced on us by tape rotations.
- I prefer to organize my backup names in directories with date+time stamped names such as: `/backup/hostname/filesystem.$(date +%Y-%m-%d.%H-%M-%S')`
- Also, to make scripting easier I create a `/backup/hostname/filesystem.current` symbolic link pointing to the latest backup.

How to manage old backups

- This one is really simple. When you start running low on disk space it is time to either add more capacity or delete some old backups.
- To delete old backups just `rm -rf` the oldest backups until you are no longer low on disk space.
- Deleting backups out of the middle rarely helps much.

How to restore

- This is pretty simple. It is generally just an rsync in the opposite direction.
- If you have to do a “bare metal” restore my suggestion is to boot from a live environment of some kind (I prefer a network booted SystemRescueCD), format the storage however you want it, rsync restore, then install the boot loader (grub-install).
- TEST and document your restore procedure!

Databases and VM images

Databases and VM images are constantly modified at the block level and if this happens while using any file-based tool like rsync you will get a corrupt backup. The best solution is to freeze the data so it can be backed up safely (LVM2 snapshot, locked tables, shut down the application, etc) and then run the rsync. If you can't do that then remaining solution is to use the backup utility included in such programs (such as mysqldump for MySQL) to make a backup file that can be copied with rsync.

Freely available backup systems that use rsync to do their work

- Rsnapshot
- Dirvish
- Backuppc
- Rspaghetti Backup (mine)
- Make your own ;)

More info

- I have written a much more technical document about how this works which I presented at the Greater Orlando Linux User's Group. It is available here:
http://sanitarium.net/golug/rsync_backups/
- I am generally available as BasketCase in the #rsync channel on irc.freenode.net