Backup and Recovery

Intro video link…https://youtu.be/tZwl7zSfxCI

Backup – why bother

- Hard drives fail at approximately 1% per year
- People ‘rm *’
- Often a legal requirement

Where to backup

- Disk
- Optical
- Tape

Tape's "Not quite dead yet"

- LTO-6
  - 2.5TB / cartridge
  - 160 MB/sec
- Power efficient – only draws power when running
- Large automated silos hold up to 100,000 tapes

Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.

—Tanenbaum, Andrew S. (1989). *Computer Networks*

https://en.wikipedia.org/wiki/Andrew_S._Tanenbaum

Common backup retention

- Dailies for a week
- Weeklies for a month
- Monthlies for a year
- Yearly's for seven years

That's a lot of backups. How will we keep track of all of them?

Software for backup - Commercial

Commercial shops will use packaged software like NetBackup, Spectrum Protect, Data Protector – to name but three.

- Network and ‘local’ backup
- Tape management
- Integration with Database servers

Software for backup – lab/ad hock

These tools are not just for the lab, they are often used by administrators to create ad hock backups and to move data around

Two main classes: tools that copy file system tree to file system tree (cp, scp, rsync, cpio) and tools that create a single archive file from a file system tree (tar, cpio)

tar

- Short for tape archive
- Can write to tape or disk
- The most popular format for Unix/Linux systems
- Supports incremental backups natively (recent versions)

cpio

- Once the ‘darling’ of sys admins
- Three modes: copy-in, copy-out, copy-pass
- Copy-in and copy-out read and write to archive files or tape like tar
- Copy-pass works like cp –A (*used to be way faster than cp –A*)
- Has largely been replaced by tar and rsync

Full vs. Incremental Backups

Full backups copy every file

Incremental backups copy every file that has changed since some previous backup

Parting thoughts

- Enterprise backup is big business. Big companies have many people who do nothing but manage backups.

- Get good at moving data around, a significant portion of many admin's time is spent managing storage

- Learn rsync

Examples

Let's walk through some examples…