IT Challenges When Dealing With Long-Term / Infinite Data Retention

Sep 2012

Chris Millet, Sr Mgr, IT

Qualcomm Overview

- Server backups:
  - Software: Symantec Netbackup 7.x
  - Hardware: 57 master servers in 27 countries
  - Tape libraries: Oracle & Spectra Logic
  - Tape drives: Hundreds of LTO drives
  - Other technologies used: Data Domain, VTL, NBU Appliances
  - 10.5PB total backup volume per month across ~10,000 servers
  - Deployed since 2000
- Desktop / Laptop backups:
  - Software: Connected Backup PC
  - Hardware: 30 servers in 6 countries regionally located
  - # of active clients: 29,000 (98,000 clients all-time)
  - Disk based backup – 1,300TB of NAS is storing client data online
  - Deployed since 2005
- ~93% success rate across both solutions

Data Retention

- Server backups
  - Monthly Full backups – Infinite retention
  - Weekly Cumulative Incremental backups – 90 days retention
  - Daily Differential incremental backups – 30 days retention
  - Application backups: varies
  - For most backups, final storage medium is tape. Tapes go offsite after 14-30 days to offsite tape storage for DR.
  - No recycle policy in effect means any tapes that go off-site are not re-used, even if the backup is expired. Ouch!

- Desktop / Laptop backups
  - Keep 10 versions of actively changing files
  - After 1 year on inactivity on a file, start pruning versions down to 1 which is kept forever
  - At any given time, at least one version of every file backed up from your system(s) is retrievable

Why do we keep data so long?

- Primary reason: eDiscovery
  - About 60 restores requested per month
  - Primarily desktop/laptop backups wanted (PSTs!)
  - Exchange mailbox (5GB quotas)
  - Email archive
  - Rarely requesting data from a file server

- Other Restore Facts:
  - 50+ server restores performed monthly.
    - Mostly near-term data (<90 days old)
  - Countless desktop / laptop restores (self service through agent that sits in the systray)

Challenges

- Explosive data growth (+50% YOY)
  - Plus our no recycle policy means we consume lots of tape
- Can be difficult to implement, especially retroactively
- New tape costs increasing
- Offsite tape storage costs increasing
  - You pay $.xx per month per tape
  - No recycle means we are continually adding to storage inventory
  - 10 years of storage roughly equal to the cost of the media itself. This is a factor when planning tape technology upgrade timing to get the most bang for buck.
  - Hundreds of thousands of tapes stored offsite. It adds up.
- Tape shelf life of older tapes
  - We have 9 track tapes and 4mm & 8mm DAT from the early 90s
  - Believe it or not, we routinely read from these tapes even today
  - Evaluating media refresh which is a complex project

More Challenges…

- Legacy hardware
  - Legacy tape drives require legacy HVD SCSI adapters, which only work in older OS / drivers, which only work on older hardware.
    - We keep a couple old Ultra 60 workstation with old drives attached.
- Documentation / Tribal Knowledge
  - Even if you have the gear, who knows how to find and restore this 10-20 year old data from legacy in-house backup applications?
- Lost tapes
- Excessive tape drive wear & tear
  - Primarily new tapes going through tape drives
- Many audits need to be put in place:
  - Setting configurations consistently on all your backup servers
  - Monitor for expiring images that should not be expiring due to misconfigurations or bugs

..and More Challenges...

- Large Connected SQL databases
  - 1TB SQL databases common on most servers
  - Billion row tables
  - More than 100 million metadata objects for a client in the database breaks Windows 32 bit
- Large Netbackup catalogs
  - 750GB recommended max per Symantec.
  - We have numerous 3-4TB catalogs on master servers.
  - Makes daily backup of the catalog challenging. Not to mention add time when you do a recovery.
  - Built-in catalog archiving feature – not fully baked
  - This means once the catalog gets too large, we retire the server (but keep it around) and deploy a new one just to start with a fresh catalog

..and Even MORE Challenges..

- Legacy backup servers
  - Need to be kept around indefinitely (at least the catalogs)
  - Have to keep maintaining them (hardware refresh, patching, etc)
- Adds red tape
  - Approvals needed to make any changes (IT security, legal support, exec mgmt). Once you go down this path, it is hard to revert back.
  - Time consuming, thorough processes to “destroy” any data on disk or tape
- Limits use of disk technologies
  - VTLs end up just being a short-term storage area for things like daily backups. We are phasing them out because the management is worth more than the benefit.
  - Data Domain – again only good for short-term backups, of which we have few.
  - Netbackup appliances – still have to get data to a tape somehow.

Summary

- Consider long-term retention carefully, especially if you have a large environment as it has tangible and intangible ripple effects in all aspects of your backup infrastructure.
  - Increases tape costs
  - Increases offsite storage costs
  - Increases complexity to manage
  - Difficult to undo later
  - Realistically, only possible with tape.

Q & A