



THE GLOBAL LEADER IN INDEPENDENT STORAGE SERVICES



Seriously: Tape Only Backup
Systems are Dead, Dead, **Dead!**

Agenda

- Overview
- Tape backup rule #1
- So what's the problem?
- Intelligent disk targets
- Disk-based backup software

Overview

- We're still talking disk to disk to tape, or D2D2T
- Not yet getting rid of tape, only placing disk between it and the backup system
- However, some environments have completely gotten rid of tape

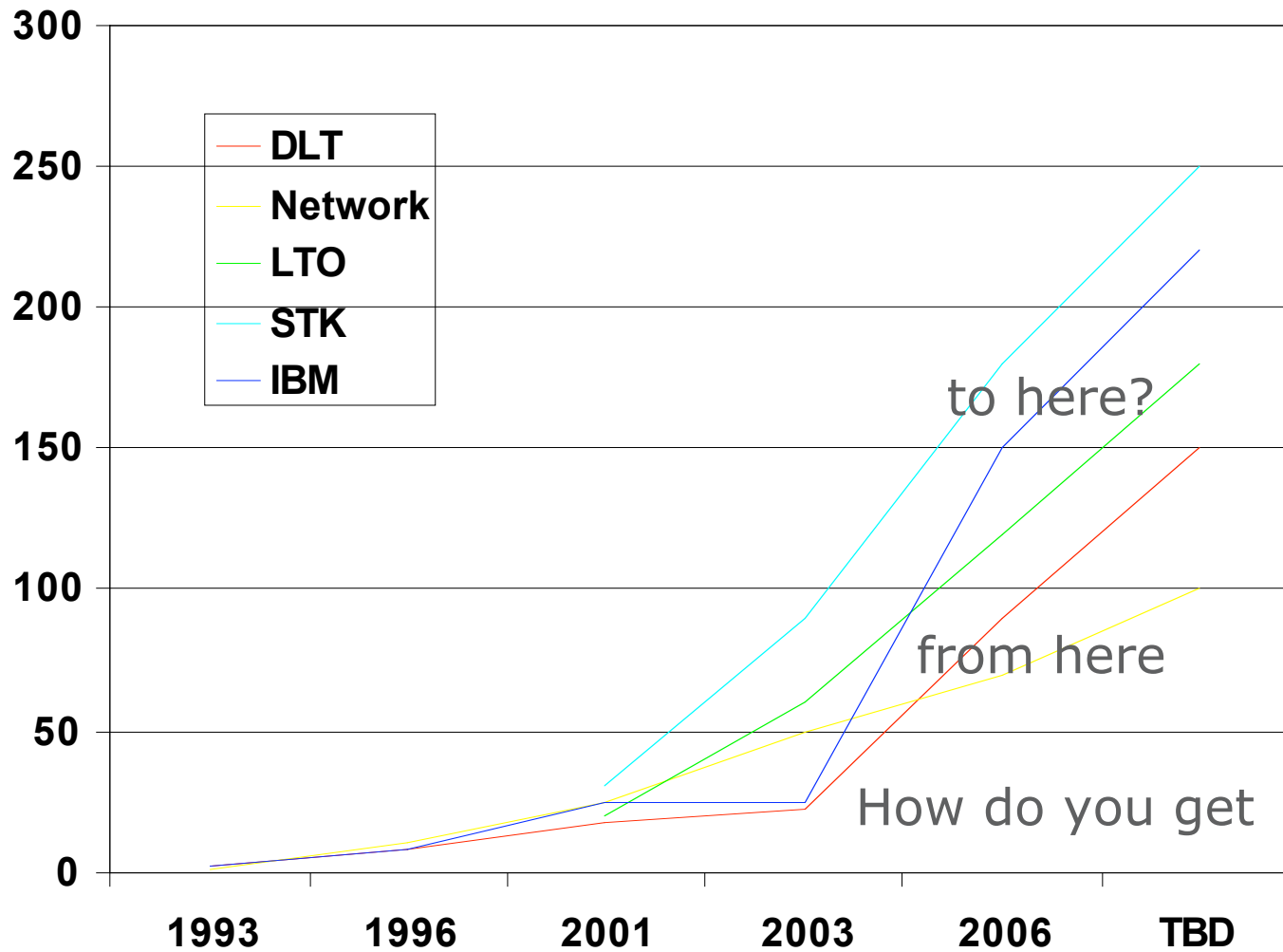
Tape backup's rule #1

- Rule #1: Match the speed of the network to the speed of the tape
- Every modern tape drive has a minimum speed at which it can reliably write data
- A streaming tape drive **cannot** write slower than its minimum speed! If it looks like it is, trust me, it's not. It's faking it.

Variable speed tape drives

- Variable speed tape drives can step down to keep up with slower speeds, but they still have a minimum speed
 - LTO-3: 27 MB/s to 80 MB/s
 - TS1120: 40 MB/s to 105 MB/s
 - T10000: 50 MB/s to 120 MB/s
 - Despite vendor claims, these drives have not eliminated shoe-shining
- Times your compression ratio!*

So what's the problem?



Disk: Infinitely Variable

- It'll take 100s of slow or fast backups all at once without multiplexing
- You can then
 - Dump them to tape at its max speed
 - Replicate them to another location

What are the options?

- Products
 - Disk-as-disk
 - SAN disk arrays
 - NAS filers
 - Disk-as-tape
 - Virtual tape drives
 - Virtual tape libraries
 - Virtual tape cartridges

Filesystem or virtual tape?

Disk-as-disk

- Pay for disk, possibly volume mgr & filesystem
- Possibly pay per GB to backup s/w vendor
- Provisioning/sharing issues
- Speed of filesystem
- Fragmentation
- Some backup s/w not fully filesystem aware
- If not already copying tapes, will need to figure out how

Disk-as-tape

- Pay for disk & value of VTL software
- Pay per slot, drive, or GB to backup s/w vendor
- No provisioning/sharing issues
- Speed of mult. raw disks
- Usually no fragmentation
- All backup s/w tape aware
- Should be faster
- Can continue not copying tapes if you use an integrated VTL

Standalone or Integrated VTL?

Standalone

- Sits *next to* your physical tape library (PTL)
- Pretends to be another tape library
- Backup to virtual tape
- Use dupe/clone software to copy virtual tape to physical tape
- Backup s/w will always know status of all “tapes”
- May take longer to make physical tapes

Integrated

- Sits *in front of* your physical tape library (PTL)
- Pretends to be your PTL
- Creates virtual tapes to match physical tapes in PTL
- Ejecting virtual tape creates physical tape
- Certain conditions may require manual intervention
- Don't break relationship btw backup s/w and real tapes
 - HSM style VTLs
 - Tape stacking



Single node or clustered?

- VTLs tend to be single nodes that range from 100-800 MB/s
- If you need more throughput, you buy another node
- “Clustered” VTLs allow you to expand capacity or throughput by adding additional data movers, but manage as a single VTL

De-duplication

- Examines and eliminates redundant blocks of data from any source
- Best thing since sliced disks
- Reduces effective cost of disk by 10:1 or more
- All major VTL vendors releasing de-dupe products right now
- All products are not the same

Fingerprinting

- Lining up incoming data with existing data to get the best match. Think real fingerprints
- The more content aware a de-dupe app is, the better this phase will go

Redundancy Identification

- SHA-1: 160 bit hash
- MD5: 126 bit hash
- Custom
 - 16 bit hash
 - Other hashes
- Bit-level compare
- Most products can use two methods

Forward vs reverse referencing

- Reverse
 - If you see a new block that you've already seen before, don't write the new block
 - Write a pointer to the old block
- Forward
 - Write all new blocks, then look at old blocks
 - If an old block matches a new block, make the old block a pointer to the new block
- The jury hasn't even left yet

In-band or out-of band

- In-band de-dupes in RAM, never writes redundant data to disk
- Requires quick processing to minimize impact on incoming backup performance
- Easier on the disk
- Requires less I/O
 - Possible hash lookup
 - Possible read of matched data
 - Write new data
- Cannot divide and conquer an individual backup
- May limit speed of some backups
- Out-of-band writes original data to disk, reads it later and de-dupes it
- Requires more I/O
 - Initial write
 - Read to calc hash
 - Read to lookup hash
 - Possible read of matched data
 - Write pointer or not
- Can divide and conquer a single large backup btw CPUs
- *Supposed to* never slow down any backups. Some are concerned that all the I/O may do this anyway

Clustering

- VTL clustering and de-dupe go hand in hand
- Two VTLs with different de-dupe databases are not the same as one cluster
- Difference between hundreds of MB/s and thousands of MB/s

Do it right from the start

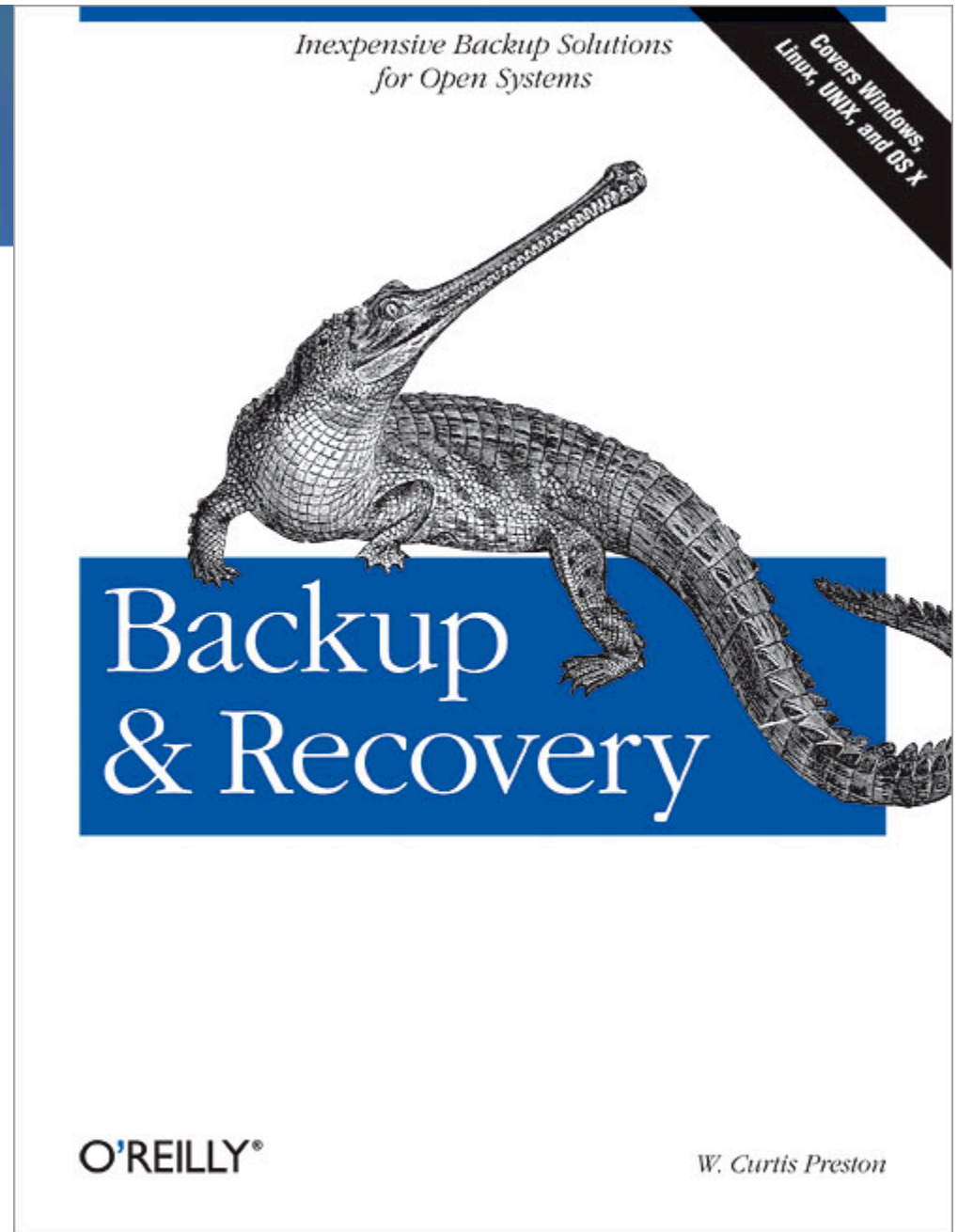
- Backup systems designed to go to disk
 - De-duplication backup software
 - Snapshots & replication
 - Continuous data protection
- Open source tools
 - BackupPC
 - Rdiff-backup
 - Rsnapshot

Vendors

Vendor	GA	Target or Source	Forward or Reverse	In/out of Band	Primary Check	Secondary Check
ADIC	Now	Target	Reverse	In	Custom	Custom
Data Domain	Now	Target	Reverse	In	SHA-1	Custom
Diligent	Now	Target	Reverse	In	Custom	Bit-level
EMC/AVAMAR	Now	Source	Reverse	In	SHA-1	None
Falconstor	Q1 07	Target	Reverse	Out	SHA-1	Opt. MD5
NetApp	Now	Target	Reverse	Out	Custom	Bit-level
SEPATON	Now	Target	Forward	Out	Custom	Bit-level
Symantec Puredisk	Now	Source	Reverse	In	SHA-1	None

Seriously

- Buy my book!
- 750 pages dedicated to free & open source backup
- AMANDA, BackupPC, Bacula, rdiff-backup, rsnapshot
- Oracle, DB2, Sybase, SQL Server, Exchange, mySQL, PostgreSQL
- AIX, HP-UX, Solaris, Linux, Windows, MacOS bare metal recovery
- Available Dec 20!



 **GLASSHOUSE**