Managing Data Migration

September 2012

LC Overview

Preserving data is really about curating data, making timely and informed decisions to ensure the usability of digital information in the face of constantly changing technologies. So we shouldn't be wooed into thinking that any outside organization is going to solve all of our data transformation needs over time.

LC Overview

Over 1000 physical and over 400 VMs
Linux, Unix, Windows
7+PB of spinning disk 25% yearly increase
9 tape libraries, 5 enterprise and 4 LTO
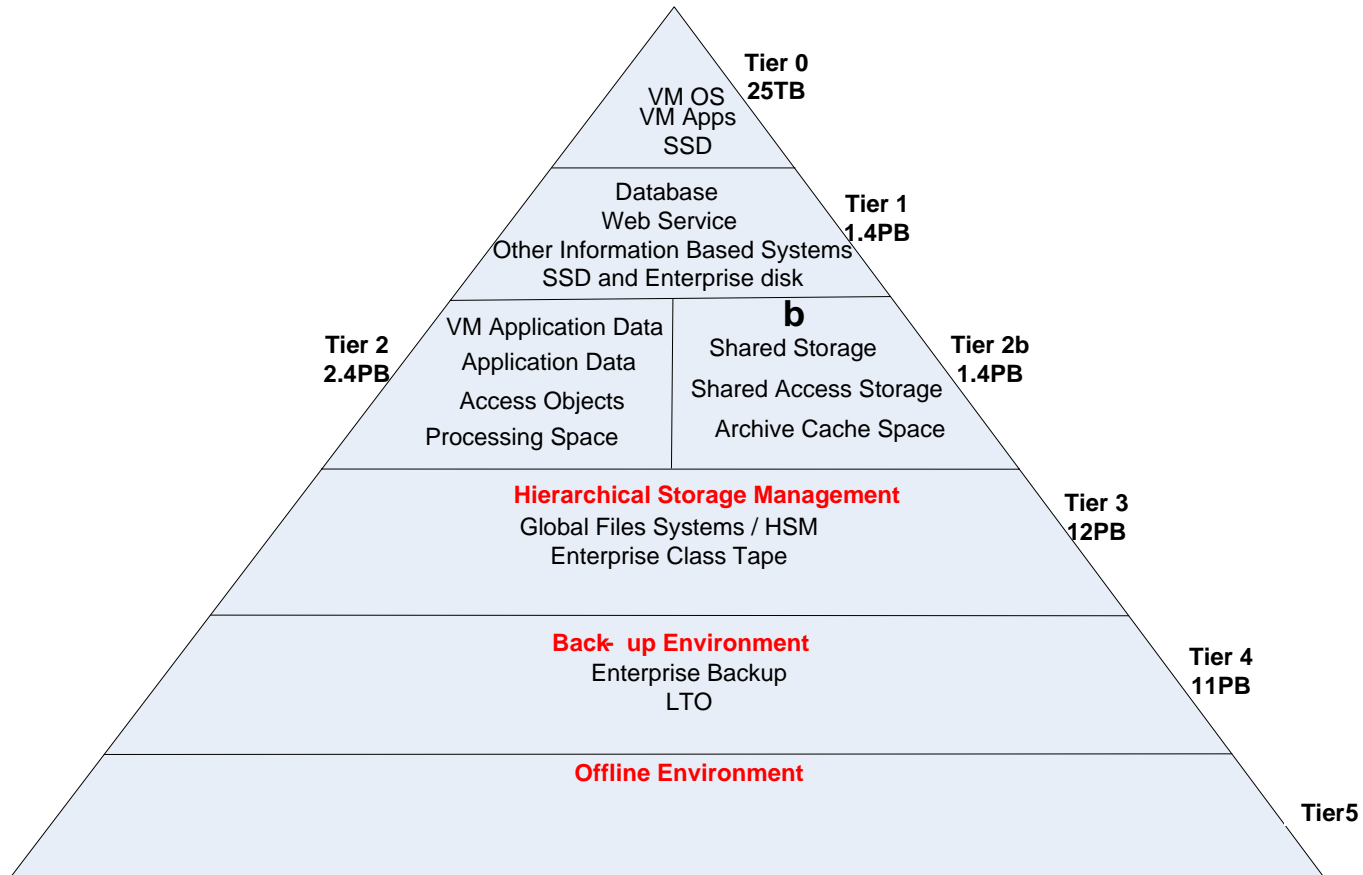Ingest 40-50TB week 50+% increase per year
HPSS and SAM

LC Overview

Application to Storage Pattern   Updated June 2012

Tier 0
25TB
VM OS
VM Apps
SSD

Tier 1
1.4PB
Database
Web Service
Other Information Based Systems
SSD and Enterprise disk

b

Tier 2
2.4PB
VM Application Data
Application Data
Access Objects
Processing Space

Tier 2b
1.4PB
Shared Storage
Shared Access Storage
Archive Cache Space

Tier 3
12PB
Hierarchical Storage Management
Global Files Systems / HSM
Enterprise Class Tape

Tier 4
11PB
Back up Environment
Enterprise Backup
LTO

Tier5
Offline Environment

Migration experience

- Disk migration is constant, and is both host based and storage based.

- Case Study - Tape migration – LC has done one complete technology upgrade.

- Case Study - HSM migration – LC has been migrating data for over 1.5 years from a defunct system.

Disk Migration

- Something everyone has experience in.

- Would like to eliminate all host based migration

- Would prefer a heterogeneous solution

Case Study -Tape Migration – Tech Upgrade

SAM running on Oracle and using an Oracle SL8500

T10Ka to T10Kb upgrade

Data a mix of packaged data (with checksums) and unpackaged data (without checksums)

Less than 1000 tapes

Case Study -Tape Migration – Tech Upgrade

Data Integrity – Packaged data checked and rewritten

Data Integrity – Unpackaged data packaged and checksums done

Customized scripts

Manual processes to correct errors

Migration done in the same cache space as ingest.

Eight (8) months

Not enough tape drives for both ingest and migration.

- EMC DiskXtender to SAM

- 580TB on LTO3 tape and Centera – millions of files, originally set up to optimize ingest.

- 1.5 years and still unfinished

- Less that 10TB left to migrate.

- What happens when a product is no longer supported.

Tape Migration HSM to HSM

- Data Validation
- No checksums were taken when originally ingested into DiskXtender.
- Crude data validation – byte counts and file counts
- New manifests and checksums taken when written to SAM
- Content Validation not done.

Tape Migration Issues

- Data Integrity (Fixity)
- Data path integrity – T10 PI/DIF
- Separate ingest from migration
- Tape contention
- Too many manual steps