Tiered Storage and
LC Storage Environment Update

Library of Congress
ITS/OPSESE Storage Group
Introduction

Storage Tiers Defined
- Tier 0
- Tier 1
- Tier 2
- Tier 3
- Tier 4
- Tier 5

How these tiers relate to the current Library of Congress Environment

Current Environment Improvements for 2012
Tier 0 – High Speed Data

➡ Characteristics
- Solid State Disk SLC/MLC, DRAM or a mixture
- Low Density - 100's of Gigabytes to 10's of Terabytes
- 100+K IOPS
- High acquisitions cost $10K+ per TB

➡ Use Case
- Database logs and indexes
- Operating System
- Virtualization: Servers and Virtual Desktop Infrastructure
- Data Caching

➡ Connectivity
- SAN (FC, FCoE), PCIe, iSCSI, SAS
Tier 1 – Transactional Data

⇒ Characteristics
  ➢ Integrated large scale disk array
  ➢ Centralized controller and cache system
  ➢ Ability to replicate between one or more devices
  ➢ 10+K IOPS
  ➢ Primarily structured data
  ➢ Moderate acquisition cost of $2K to $5K per TB

⇒ Use Case
  ➢ Database
  ➢ Transaction Processing
  ➢ Mission critical application

⇒ Connectivity
  ➢ SAN
Tier 2 – Active Data

⇒ Characteristics
  ➢ Higher capacity (100's Terabytes)
  ➢ High speed drives (15K to 10K RPM drives)
  ➢ Sequential Performance
  ➢ Scale-out design
    ▪ decreased disk-to-controller or increased sub-system to gain performance
  ➢ Both structured and unstructured data
  ➢ Moderate acquisition cost of $1K to $2.5K per TB

⇒ Use Case
  ➢ Application data
  ➢ Transformation and transitional data
  ➢ Tier 4 thru 5 cache storage

⇒ Connectivity – SAN, iSCSI, NAS
Tier 2b – Data for Access

 ➤ Characteristics
   ➢ Highest capacity drives (1 TB or greater)
   ➢ Lower speed drives (less than 10K RPM)
   ➢ Higher disk to controller ratio
   ➢ Scale-up
     ▪ hundreds of drives per controller
   ➢ Primarily unstructured data
   ➢ Lower acquisition cost of $500 to $650 per TB

 ➤ Use Case
   ➢ Access storage
   ➢ DR Data
   ➢ Tier 3 thru 5 cache storage

 ➤ Connectivity
   ➢ SAN, iSCSI, NAS, SAS
Tier 3 – Data at Rest

⇒ Characteristics
  ➢ Mixture of disk, tape and software
  ➢ Hierarchical Storage Manager / Automated Tiering
  ➢ Transparent to application or end-user of data locality
  ➢ Predictable latency between data request to data received
  ➢ Primarily unstructured data
  ➢ Lower acquisition cost of $25 to $55 per TB

⇒ Use Case
  ➢ Master file storage
  ➢ Collection storage

⇒ Connectivity
  ➢ SAN, NAS
Tier 4 – Backup Data

➔ Characteristics
  ➢ Mixture of disk, tape and software
  ➢ Back-up storage product
  ➢ Administrator assistance required for data recall

➔ Connectivity
  ➢ SAN, Server Agents
Tier 5 – Long-term Storage - Offline

- Characteristics
  - Mixture of disk, tape and software
  - Offline storage
  - Off-site storage
  - Hours to weeks to recall data

- Use Case
  - DR of critical data

- Connectivity
  - Truck
Example Products to Tiers

Tier 0
- Avere Nimble
- EMC VNX
- 25TB

Tier 1
- EMC DMX
- IBM XIV
- 1.4PB

Tier 2
- Avere
- EMC VNX, CX4
- IBM DS3000 Series
- Pivot3 Cloud Stor
- RAID Inc. Xanadu 260
- Sun StorEdge 6xxx
- 2.4PB

Tier 2b
- DDN SFA10K
- IBM DS5000 Series
- RAID Inc. Xanadu 260
- Scale Computing NS4000
- Whitebox NAS Servers
- 1.2PB

Hierarchical Storage Management
- IBM GPFS-HPSS-TS1140
- Cache: IBM DS5000 Series and RAID Inc. Xanadu 260
- 800TB

Backup Environment
- Oracle QFS-SAMFS-T10K (B/C)
- Cache: Sun StorEdge 6xxx
- DDN SFA10K and RAID Inc. Xanadu 260

Tier 3
- 12PB

Tier 4
- 11PB

Symantec NetBackup
- Tape: IBM LTO4
- Disk: EMC Clariion
System Improvement Efforts FY’13

• Technology Refresh
  – Server and storage refresh
  – Tape technology change for HPSS

• Ongoing Efforts
  – Migration from older tape system 80% complete
  – Static storage update
  – Content Transfer Environment
  – Storage Resource Management

• Big Data environment
Any Questions

Contact: Carl Watts
Enterprise Systems Engineering
email: cwat at loc dot gov