Library of Congress

Thomas Youkel

Storage projections

- Presentation space
  - FY09 178 TB DCF
  - FY10 210 TB DCF
  - FY11 200 TB DCF
- Preservation storage
  - FY09 289 TB DCF 2.0 PB Culpeper
  - FY10 338 TB DCF 6.0 PB Culpeper
  - FY11 327 TB DCF 12.0 PB Culpeper

DCF – Madison Computer Room
Culpeper – Packard Campus (LoC) (NAVCC)

Overall Challenges

- Data Integrity – both on disk and tape
- Workflow – automating from ingest to preservation
- Content management
- Migration – points back to data integrity

LoC goals

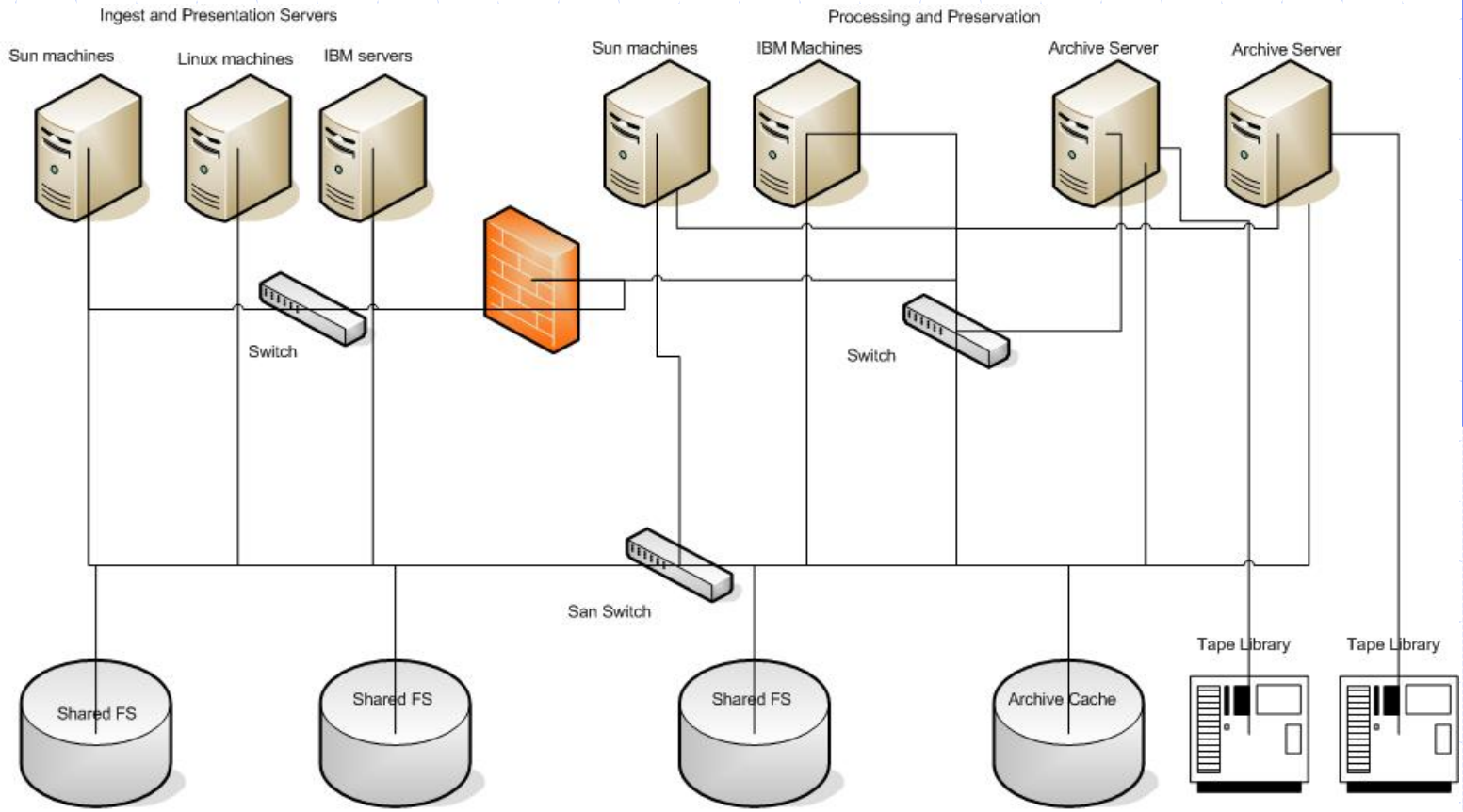
hsn2

- a high-performance data network to accomplish basic bit preservation as the foundation for distributed preservation
- scalable, reliable and secure local services and tools for data transfer, ingest and management
- reusable and modular components generalized and made available for inter-operability with the broader network, and
- protocols and metrics for preservation processes, including exchange of data

Slide 4

hsn2 Is this order of importance??

What about standard ILM framework??
Henry S Newman, 9/17/2009



Shared file systems are attached to the ingest as well as the archive machines allowing rapid movement of data from the ingest machines to the archive. They are also attached to processing servers as well as presentation servers wherever possible to facilitate transformation and presentation of data

Projected Preservation and Preservation

- 10GbE backbone
- SAN upgrades to 4gb, 8gb, and FCoE
- 2 long term storage solutions, Sun SAMFS and IBM HPSS
- Shared Storage – SQFS, GPFS, etc
- Data moved across the SAN and fewer times across the network
- Multiple machines sharing data