September 27, 2010

UNT DIGITAL LIBRARY
STORAGE ARCHITECTURES FOR DIGITAL COLLECTIONS

History

- Similar to most academic libraries
- Library manages library infrastructure
- Average growth in technology needs
- Investment in high-end SAN solution
- Moving happily along…

- Until…

LOC - Storage Architectures for Digital Collections

UNT Digital Library Program

- Sudden growth in digital content acquisition
  - Web archiving
  - Digitization projects
  - Born digital content
- Unknown growth patterns

- Unplanned system architectures

Interim Solutions

- Scale out existing infrastructure
  - Not scalable at growth rate because of $$$

- Build R&D infrastructure using new technologies
  - PetaBox infrastructure
  - Low-end SATA arrays
  - Great for R&D but weren't adequate for production services

Regrouped Efforts

- Define needs based on real patterns
- Collaborative partnership with UNT Library Core IT services and Digital Library
- Give and take on both sides
  - Moved to write once digital object model
  - Data spread across logical volumes and nodes
  - Just in time lookup of digital object locations
  - Abstract storage in application logic instead of overcoming with complex storage architectures

Current Solutions

- Large investment in Dell SATA based iSCSI Infrastructure (MD3000i – 500 TB)
- Arrays have different configurations based on need (6 different arrays)
- Different RAID configuration based on system and data integrity needs (RAID-5, RAID-6 primarily)
- Add new storage volume to DL environment, don't grow existing volumes
- Rebalance digital objects as storage arrays are replaced
- Replicate across systems with technology neutral replication framework

Next Step

- Scale it up!!!

- Invest in geographically separate replication location:
  - Tape infrastructure
  - iRods
  - Not managed by UNT (hopefully)