Archive Storage Infrastructure

At the Library of Congress

September 2014

The Packard Campus of the NAVCC

Mission

- The National Audiovisual Conservation Center develops, preserves and provides broad access to a comprehensive and valued collection of the world's audiovisual heritage for the benefit of Congress and the nation's citizens.

Goals

- **Collect, Preserve, Provide Access to Knowledge**

- The National Audiovisual Conservation Center (NAVCC) of the Library of Congress is the first centralized facility in America especially planned and designed for the acquisition, cataloging, storage and preservation of the nation's collection of moving images and recorded sounds. This collaborative initiative is the result of a unique partnership between the Packard Humanities Institute, the United States Congress, the Library of Congress and the Architect of the Capitol.

- The NAVCC will consolidate collections now stored in four states and the District of Columbia. The facility houses more than 1.25 million film and video items and 3.5 million sound recordings, providing endless opportunities to peruse the sights and sounds of American creativity.

The Packard Campus – Many Formats

The Packard Campus – Past, Present and Future

- Growth since production
  - February 2009: 10 TB / month
  - February 2010: 45 TB / month
  - February 2011: 91 TB / month
  - February 2012: 118 TB / month
  - February 2013: 71 TB / month
  - Peak in August 2012: 141 TB / month
  - February 2014: 40 TB / month
  - August 2014: 121 TB month
- Current: 5.1 PB and 1.1 Million files replicated in 2 locations. Newspapers, internet archive, prints and photographs are over 2 PB and 160 Million files.
- 53 Points of Digitization (PODs):
  - 34 Solo (16 in robotic cabinets)
  - 9 Pyramix
  - 10 Linux(OpenCube,etc)
  - Daily each POD generates: 2GB-150GB for audio and 50GB-1,200GB for video
  - Additional PODs coming in the future include 2K and 4K scan for film, digital submission for Copyright and other (Live capture-264 DVRs, PBS, NBC Universal, Vanderbilt TV News, SCOLA, etc)

The Packard Campus – Past, Present and Future

Interesting changes

- Television Pilot: Bring audio-video copyright materials in via networks with minimal intervention and automated processes

- HistoryMakers Interviews: 184 TB of LTO digital files processed through "orderless ingest" where the content can create relevant meta data from embedded and included data.

- SNL: 40 years of video, practice, sound (Up to 15 TB per season)

- Automated processes to generate required proxies (Radiant Grid) and verify content (Baton)

- Less than 20 ingest streams per day last year to almost 30 ingest streams today
    - How does this change our architecture?

The Challenge

- Projected: 300 TB / week or 1.3 PB / month – at least 4 years off

- Counting on doubling of tape density and computing power to keep us in our current 3000 Sq feet of computer room.

Performance

We've changed from less than 20 streams to almost 30 streams

- Tested our filesystems using different I/O block size and found the best performance at 512 MB blocks. Tested to 6000 MB/sec

- With 14 LUNs and 30 streams we are seeing 700 MB/sec max

Why?

- Still investigating, but 30 streams with a write and a read (SHA1) is 60 tasks on 14 LUNs or over 4 tasks per LUN. The heads in the LUNS are not operating efficiently.

- Smbget file transfer protocol is small block size (64 KB). Win XP won't allow adjustments, Win 7 will. Still not better.

- Digest (for SHA1) appears to use small block size (64 KB). Investigating options that will leverage the SSL accelerator cards in the hardware.

- Operations per second are limiting us at network and disk. Likely also within the operating system but we haven't proven that yet.

Performance

Architectural Solutions?

- Reduce # of streams coming in to reduce contention
    - Limit # of ingests at any one time
    - Deploy another server. This will require application software changes requiring months of coordination and planning
    - Deploy more LUNs. More hardware, more complexity. Likely still have issues with processing # of I/O blocks to storage and packets to network.

- Replace small block I/O
    - Use signiant and other file transfer protocols
    - Use different digest

- Change technology
    - SSD won't have the head latency but does have issues with overall bandwidth. We already planned on evaluating SSD so accelerate the timetable.

LIBRARY OF CONGRESS

Doveryai, No Proveryai
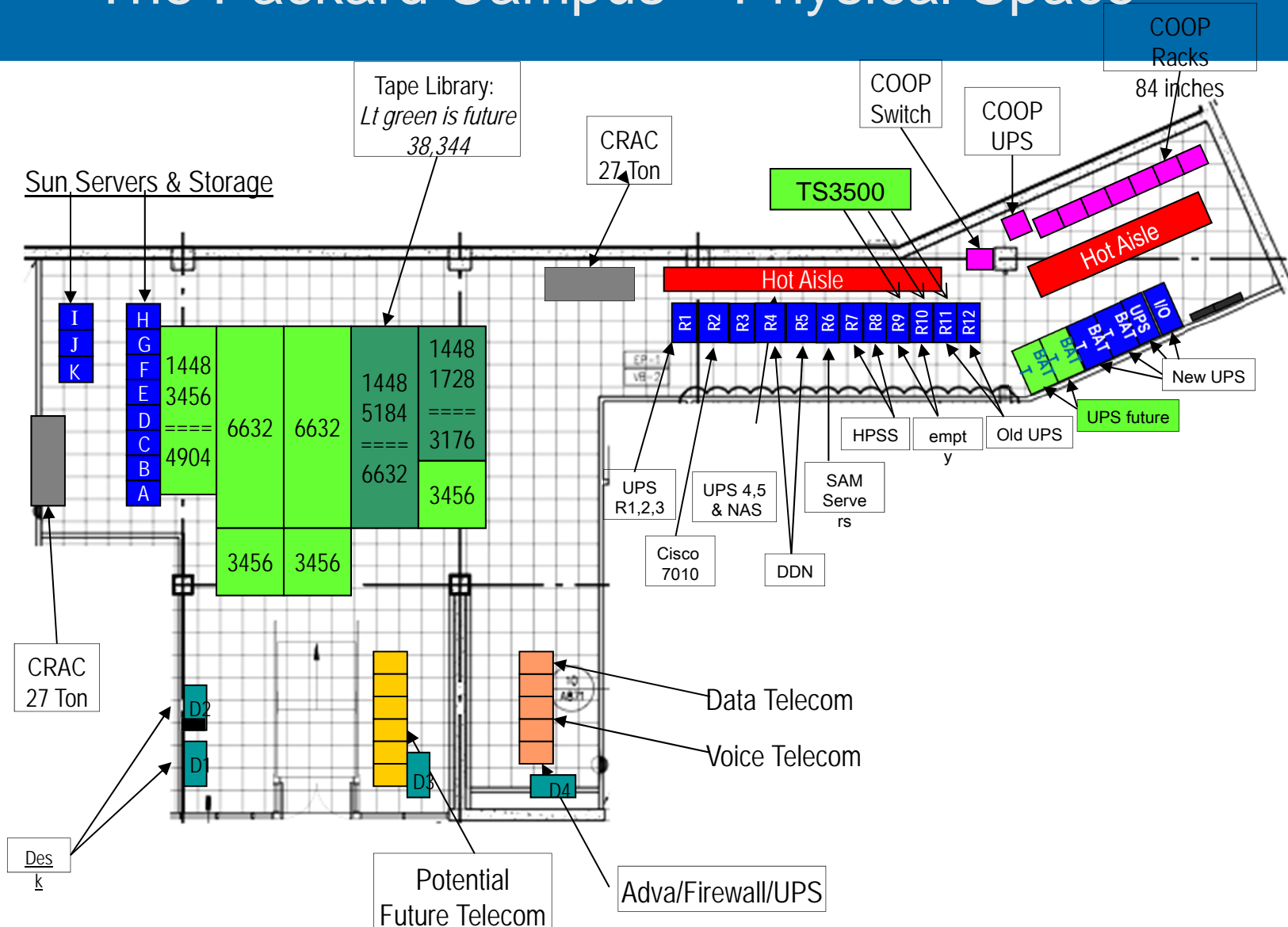## Trust but Verify

Content versus data

- We want to reduce the likelihood of losing content while still recognizing that data loss is inevitable.

- Catch and correct all marginal errors as soon as possible

- Some of the regular verification processes that we run:

    - Samfsbackup (meta data backup) 5X/day
    - Verify samfsbackup size and frequency. Send an email if missing.
    - Fix damaged files. Occasionally a file will be marked damaged because it cannot be retrieved from tape. Usually because a tape was stuck in a drive/robot/pass thru port. Find these everyday and attempt to stage. If we can't, then send an email. Send an email when we find damaged files so we know issues are occurring and being corrected
    - Stats: Watch the # and size of files waiting to archive. Warn when the # of files or size of files exceeds thresholds. Usually an indication of some marginal error condition. Fix before file system fills up or we fail to deliver a file for customers.
    - Samfsck: Run this daily with filesystem mounted. Warns when there are marginal conditions with file system before they are catastrophic.
    - # of tapes/TB available: Know when we are running low so we can correct before a failure
    - Tpverify: Running in test, verifying all blocks of data on tape with CRC.
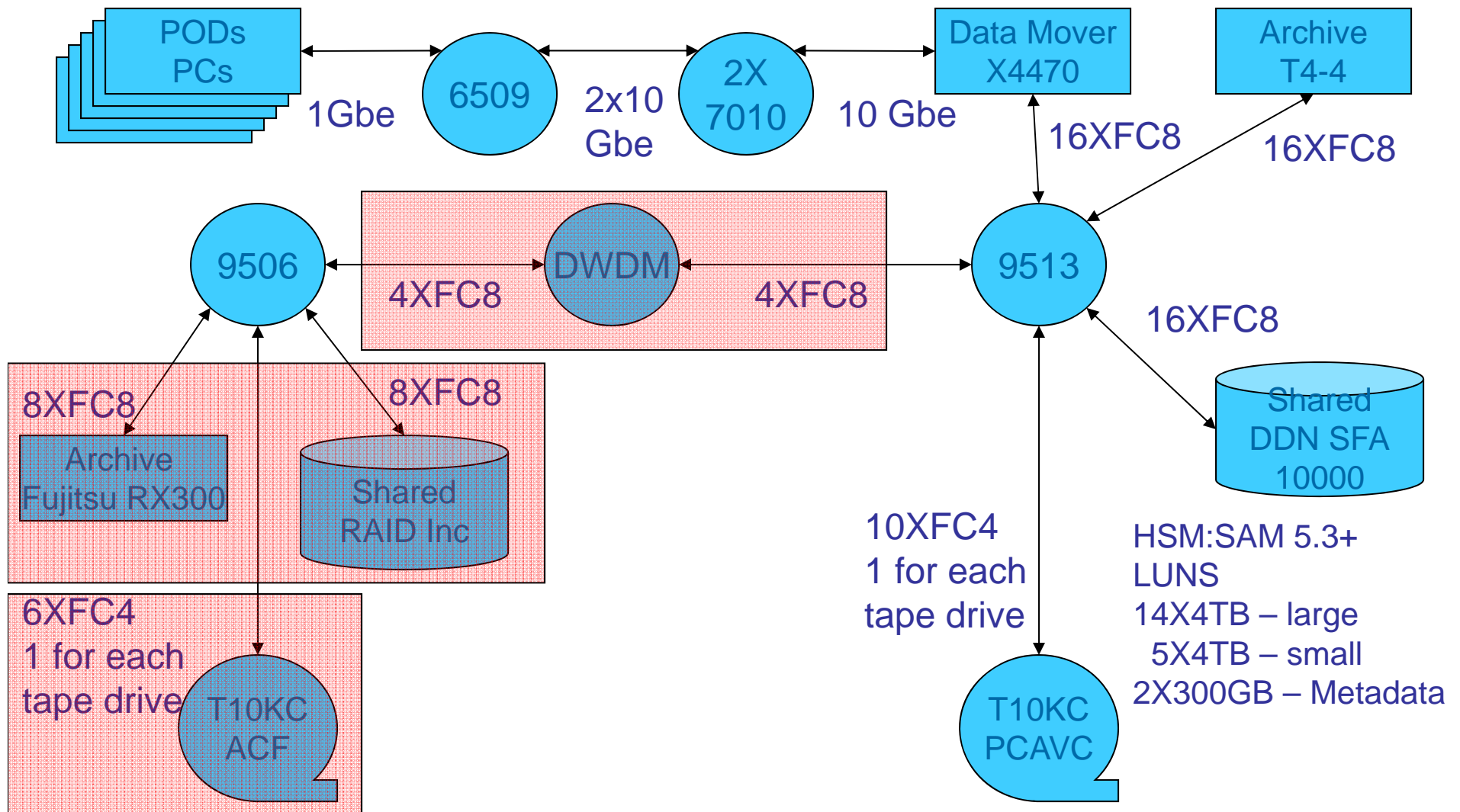
Addendum / Attachments

- Current Physical Space
- Current Hardware Architecture
- Future Hardware Architecture

LIBRARY OF CONGRESS

The Packard Campus – Physical Space

Physical Implementation V2.1: 6.5 GB/s throughput Infrastructure Upgrades

PODs PCs

1Gbe

6509

2x10 Gbe

2X 7010

10 Gbe

Data Mover X4470

Archive T4-4

16XFC8

16XFC8

9506

DWDM

4XFC8

4XFC8

9513

16XFC8

8XFC8

8XFC8

Archive Fujitsu RX300

Shared RAID Inc

Shared DDN SFA 10000

6XFC4 1 for each tape drive

T10KC ACF

10XFC4 1 for each tape drive

T10KC PCAVC

HSM:SAM 5.3+ LUNS
14X4TB – large
5X4TB – small
2X300GB – Metadata

Physical Implementation V2.2: 6.5 GB/s throughput Infrastructure Upgrades