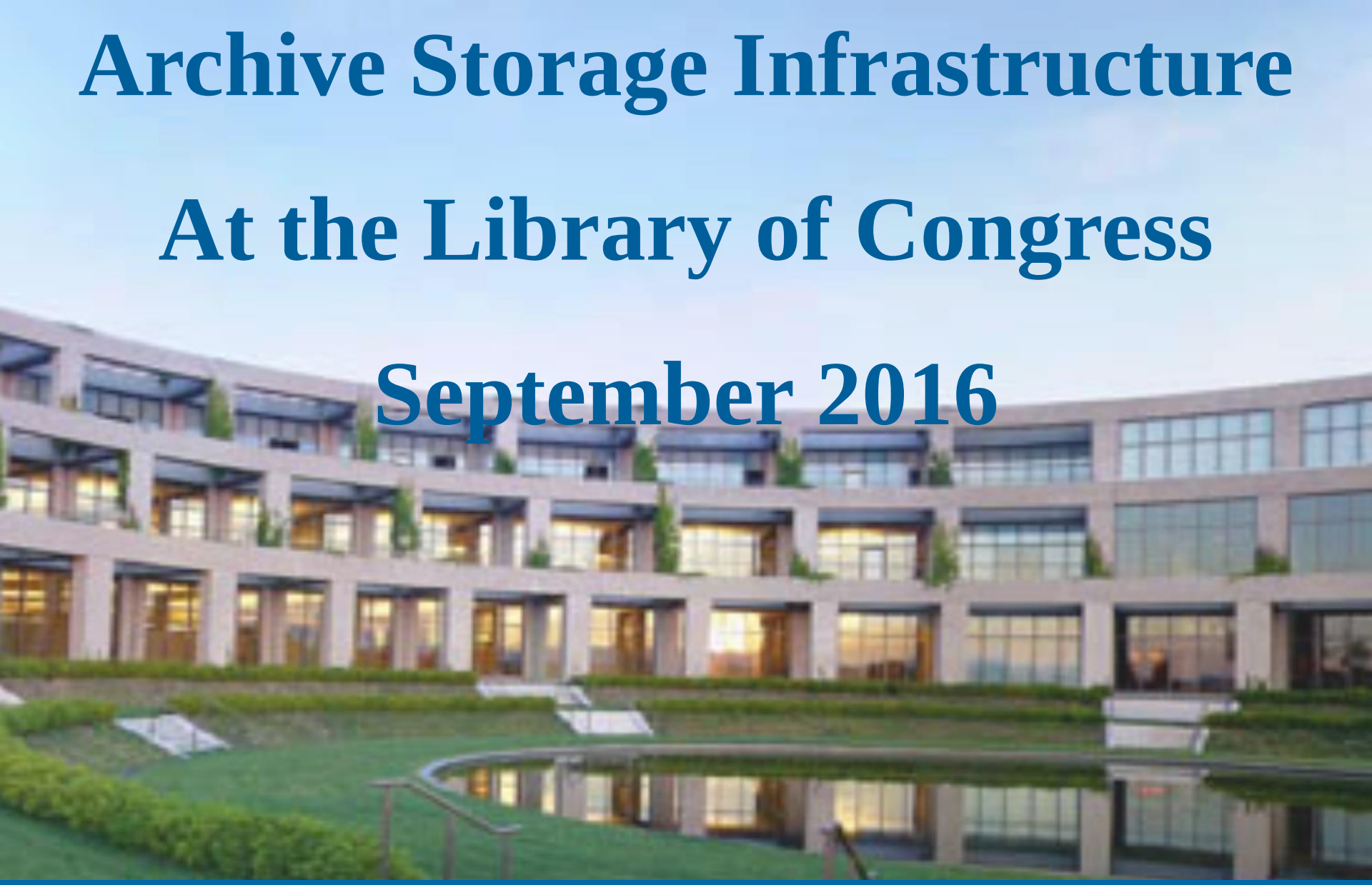
Archive Storage Infrastructure

At the Library of Congress

September 2016



LIBRARY OF CONGRESS

Packard Campus for Audio Visual Conservation
http://www.loc.gov/avconservation/packard/

The Packard Campus

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 will be 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 consolidated collections stored in four states and the District of Columbia. The facility boasts more than 1.5 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

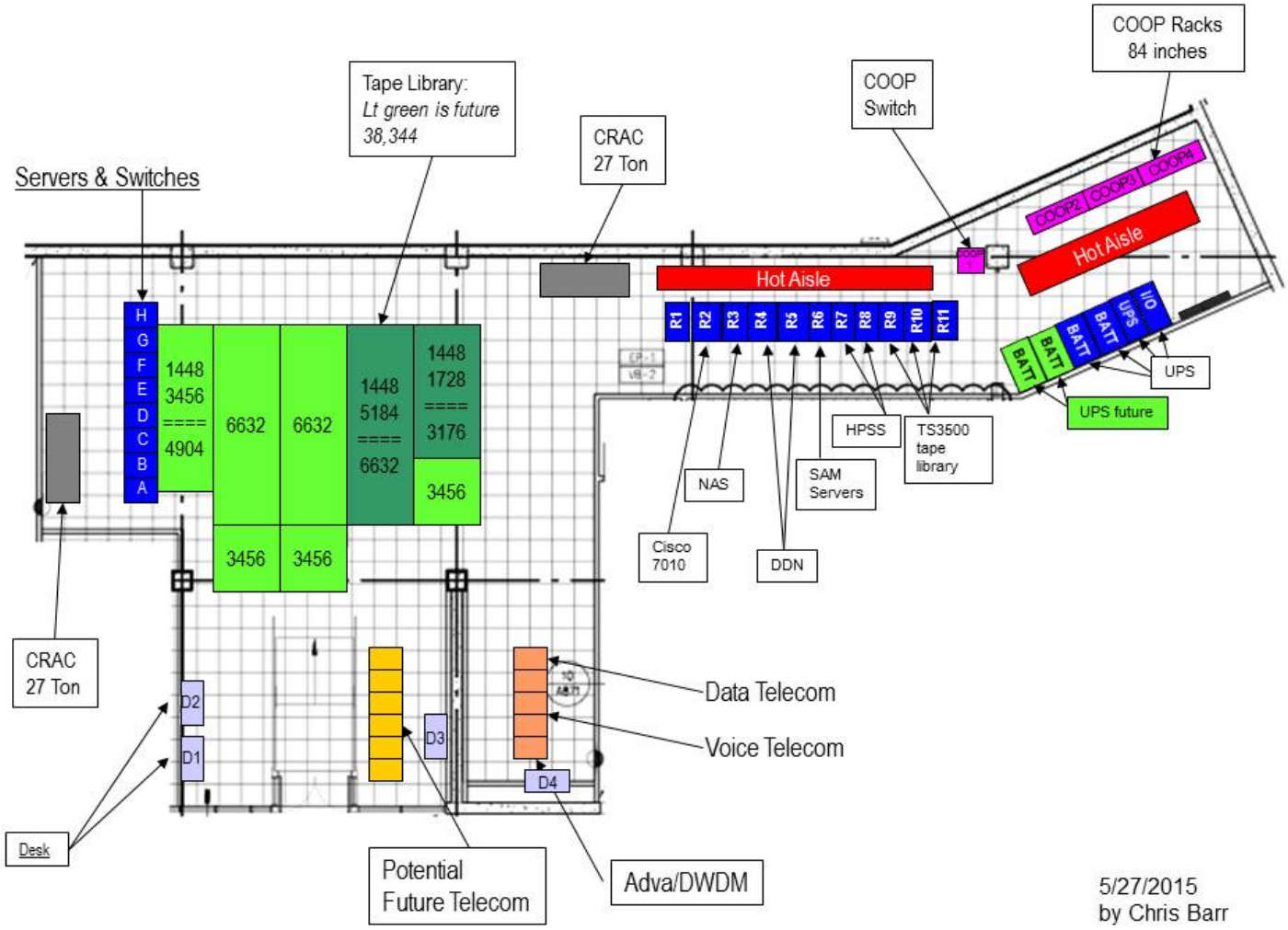
The Packard Campus – Past, Present and Future

- Growth since production
  - February 2009: 10 TB / month
  - February 2010: 45 TB / month
  - February 2011: 91 TB / month
  - February 2012: 118 TB / month
  - *Peak in September 2015: 235 TB / month*
  - February 2013: 71 TB / month
  - February 2014: 40 TB / month
  - February 2015: 45 TB / month
  - February 2016: 64 TB / month
- Current: 6.8 PB and 1.8 Million files replicated in 2 locations. 3.9 PB and 200 Million files for Newspapers, internet archive, prints and photographs
- 53 Points of Digitization (PODs): 2K is new this year, NBC Universal, AAPB
  - 34 Solo (16 in robotic cabinets), 9 Pyramix, 10 Linux(OpenCube,etc) 1 Quadriga, 2 DVD Rippers, 1 CD Ripper, Oxberry, Arrilaser, Spirit, Vario, Clipster
  - Daily each POD can generate: 2GB-150GB for audio and 50GB-1,200GB for video
  - Additional PODs coming in the future include 4K scan for film, digital submission for Copyright, Live capture-264 DVRs, PBS and others.

The Challenge

- **Projected: 300 TB / week or 1.3 PB / month – at least 5 years off**
- **Counting on doubling of tape density and computing power to keep us in our current 3000 square feet computer room with two 20 ton CRACs and 300 KVA of power**
  - **Using 51 KVA now**

The Packard Campus – Physical Space



Doveryai, No Proveryai

Trust but Verify

Content is different than data

- Reduce the likelihood of content loss while recognizing that data loss is inevitable
- Catch and correct all marginal errors and failures as soon as possible
- Verify all the content at a regular interval
- 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: Verify all tapes with data every 6 months. Verifying all blocks of data on tape with CRC.

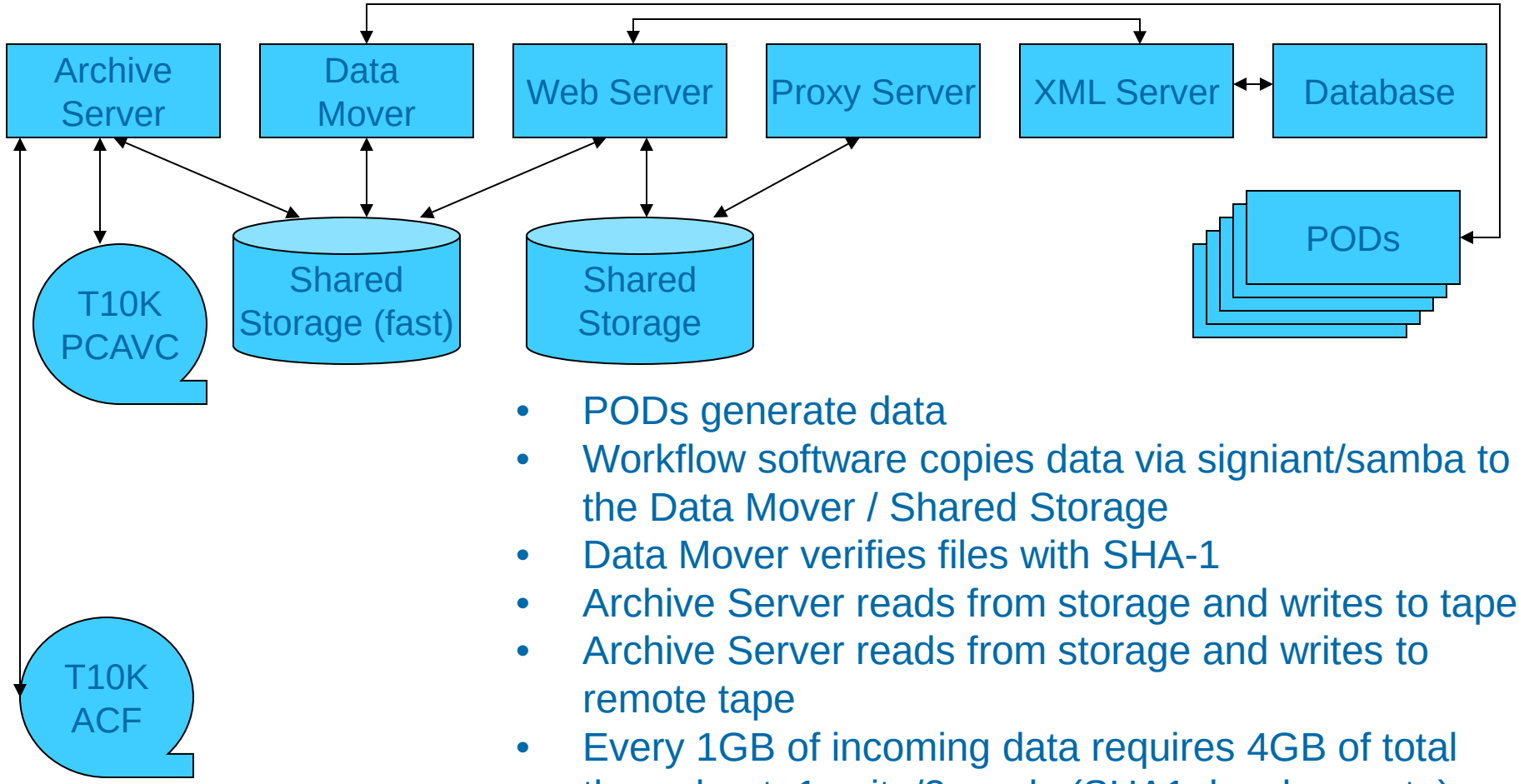
The Packard Campus – Status

Current initiatives (Partially due to questions brought up in 2015 Storage Conference)

- Reduced complexity and cost by writing directly to tape over DWDM. Revisted due to improvements in DWDM bandwidth/latency and fiber channel buffer credits
- Upgrade 2010 disk cache. 6 years means the same performance is an order of magnitude less in cost. Less than 12 month ROI over maintenance cost
- On schedule for second migration of 6 PB of content from T10KC to T10KD over a 9 month time frame. No errors so far.
  - Didn't need to migrate so soon, but the ability to increase capacity by 70% (7 PB) by investing in newer drives proved to be more cost effective.
  - First migration exposed 27 errors, most due to a corruption on disk. We added a verification step after tape write and have seen no more occurrences, validating our analysis about the root cause.
- Oracle has a roadmap that includes tape to tape migration and storing our SHA1 values in extended file attributes. This will move our verification process away from migration. With tpverify/tape CRC verification every 6 months, we look to verify the whole collection every 2 years
- Piloting a partnership with a local University to provide greater access over Internet II
- Enhancing requirements for a Content Abstraction Layer to simplify customer submission / access and technology maintenance / refresh

Functional Architecture – Data Movement

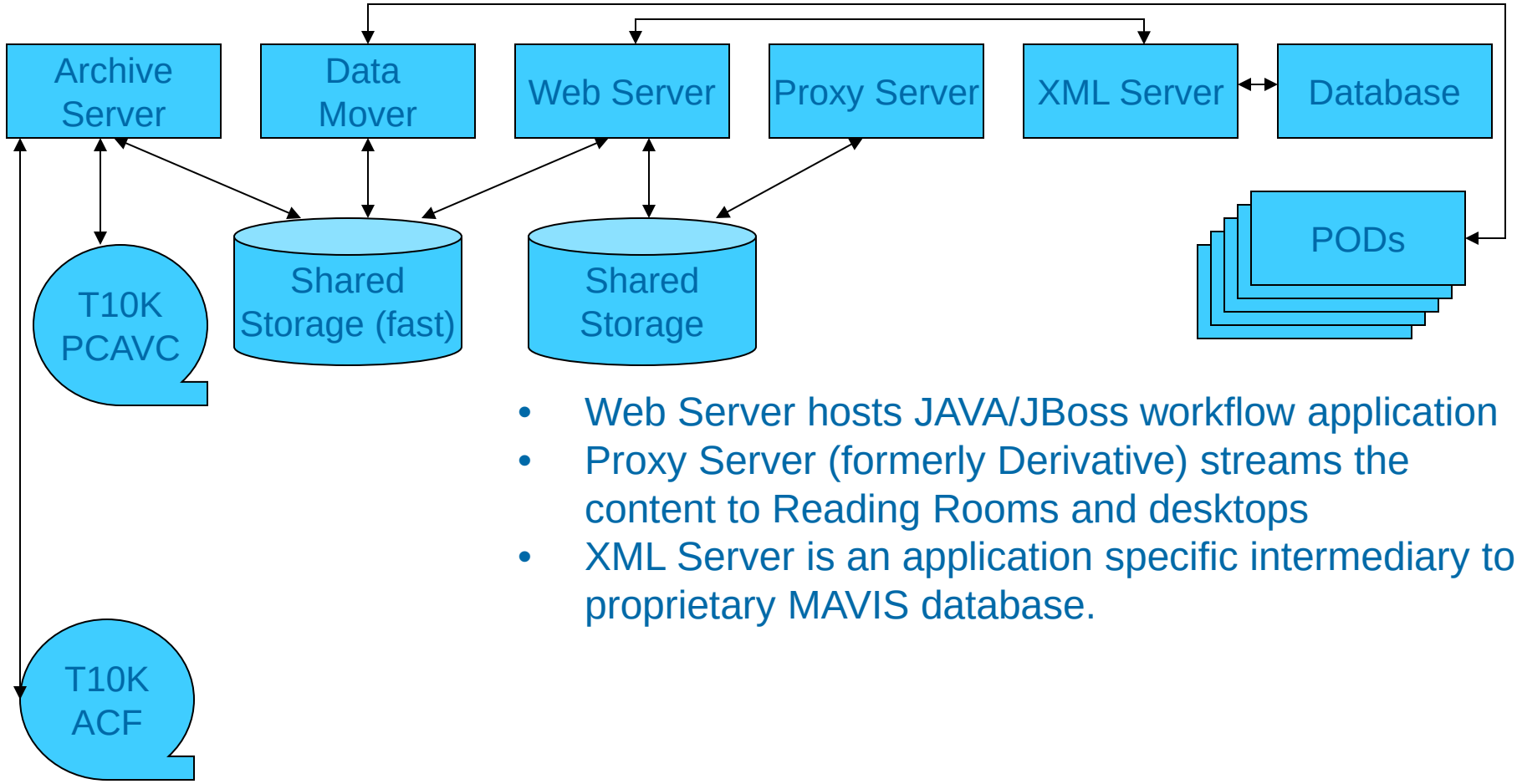
Archive Storage Infrastructure



- PODs generate data
- Workflow software copies data via signiant/samba to the Data Mover / Shared Storage
- Data Mover verifies files with SHA-1
- Archive Server reads from storage and writes to tape
- Archive Server reads from storage and writes to remote tape
- Every 1GB of incoming data requires 4GB of total throughput: 1 write/3 reads (SHA1, local, remote)

Functional Architecture – User Interface

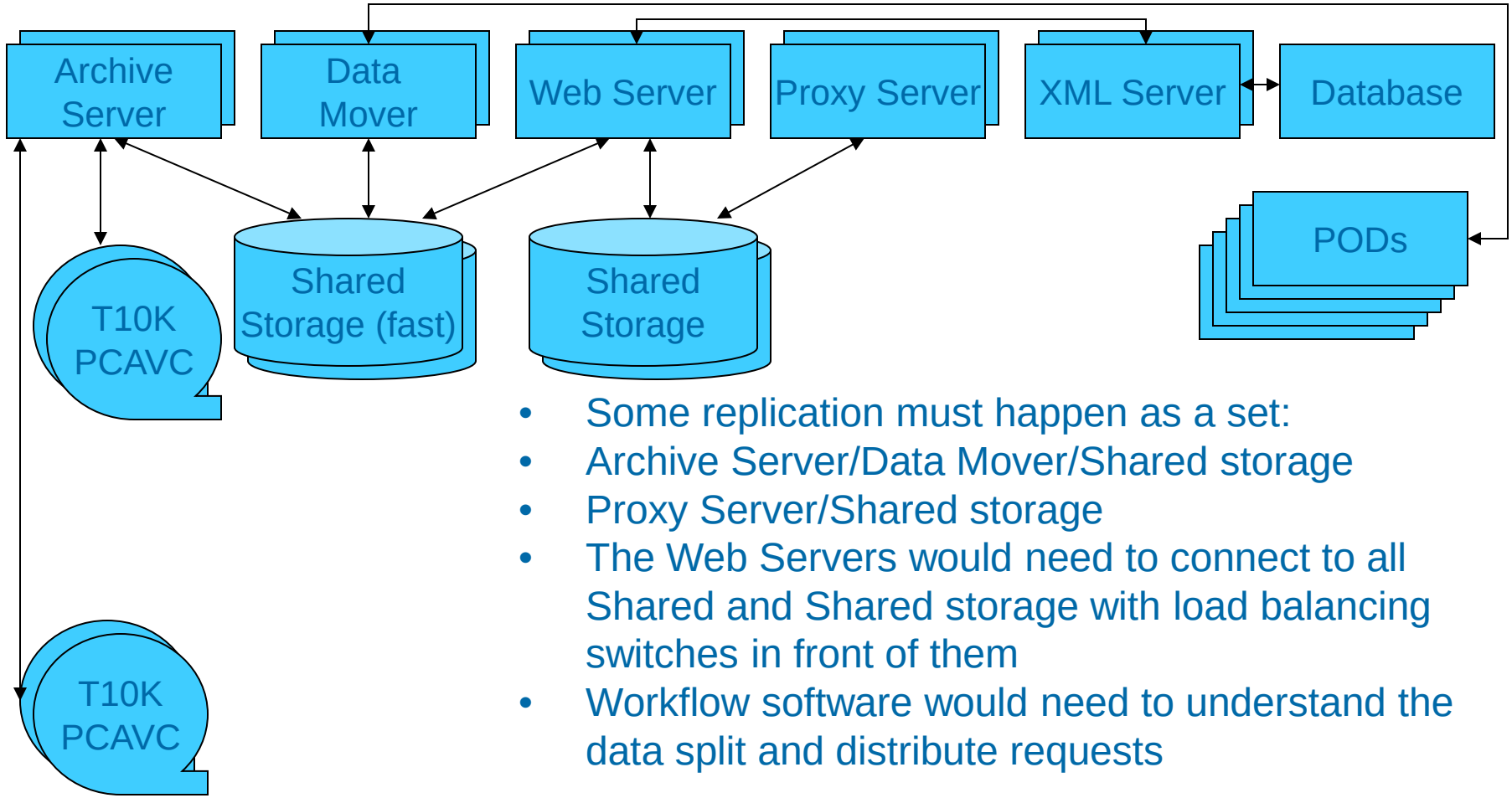
Archive Storage Infrastructure



- Web Server hosts JAVA/JBoss workflow application
- Proxy Server (formerly Derivative) streams the content to Reading Rooms and desktops
- XML Server is an application specific intermediary to proprietary MAVIS database.

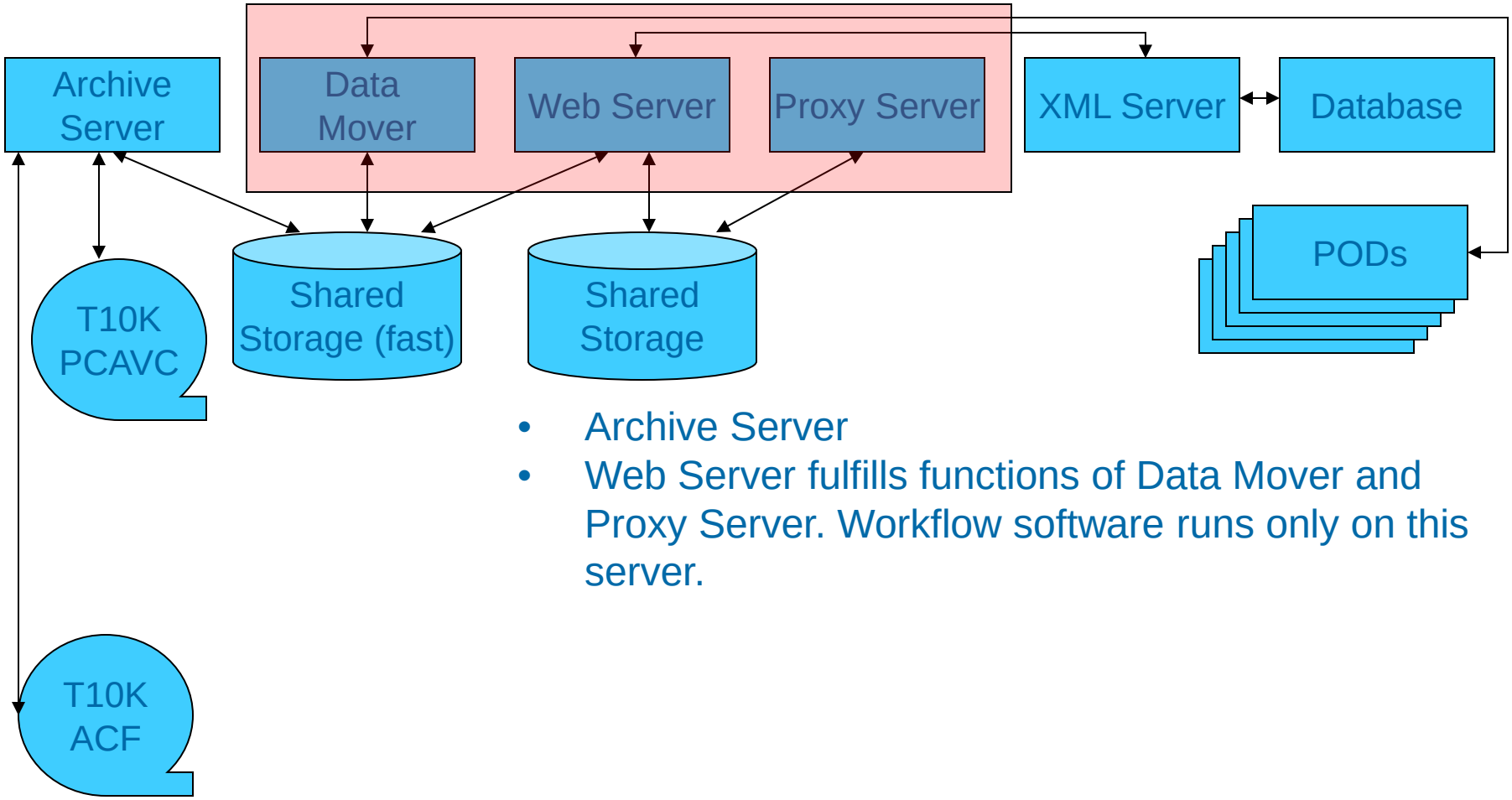
Functional Architecture - Scaling

Archive Storage Infrastructure



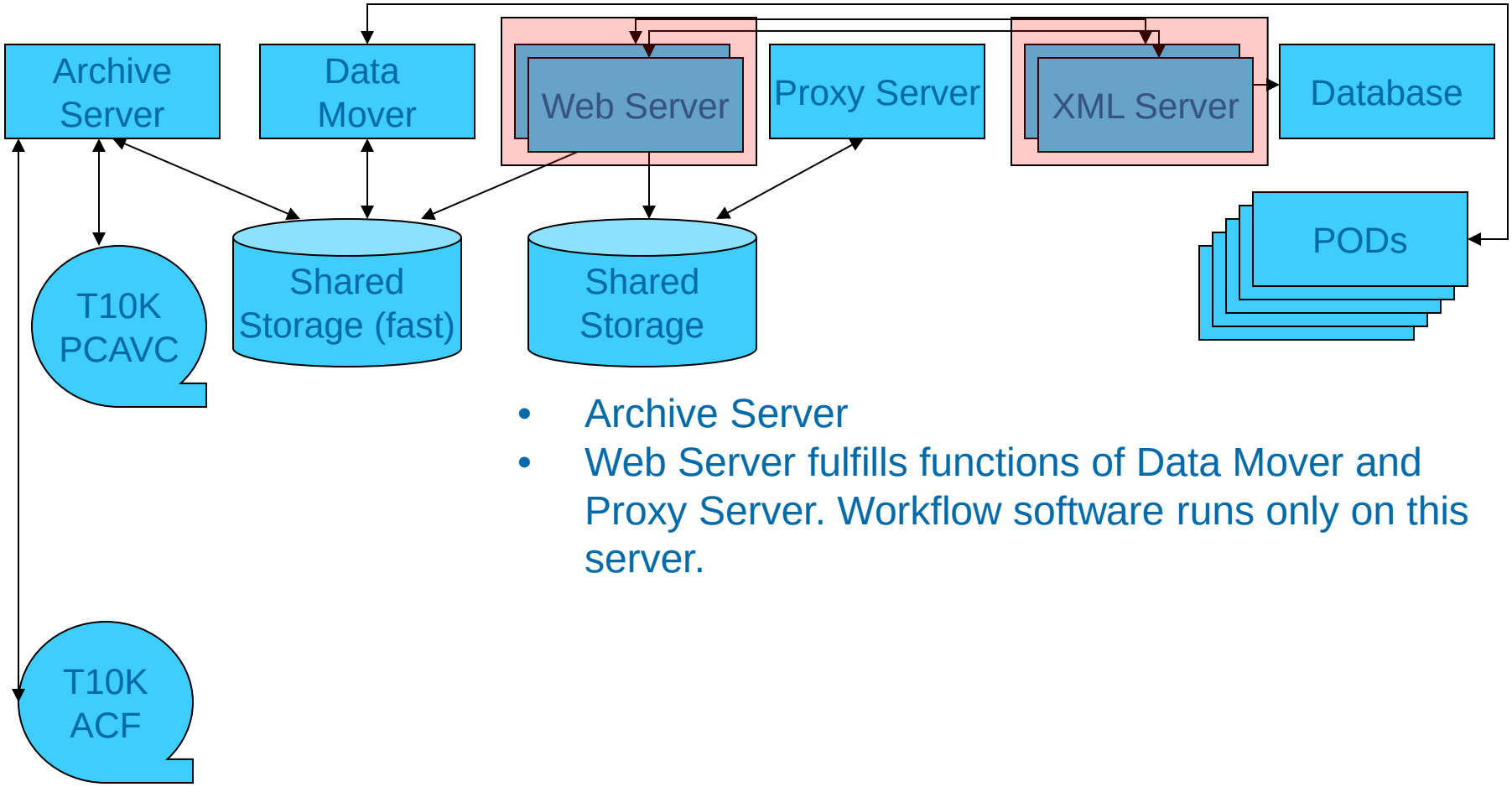
- Some replication must happen as a set:
- Archive Server/Data Mover/Shared storage
- Proxy Server/Shared storage
- The Web Servers would need to connect to all Shared and Shared storage with load balancing switches in front of them
- Workflow software would need to understand the data split and distribute requests

Functional Architecture – Current



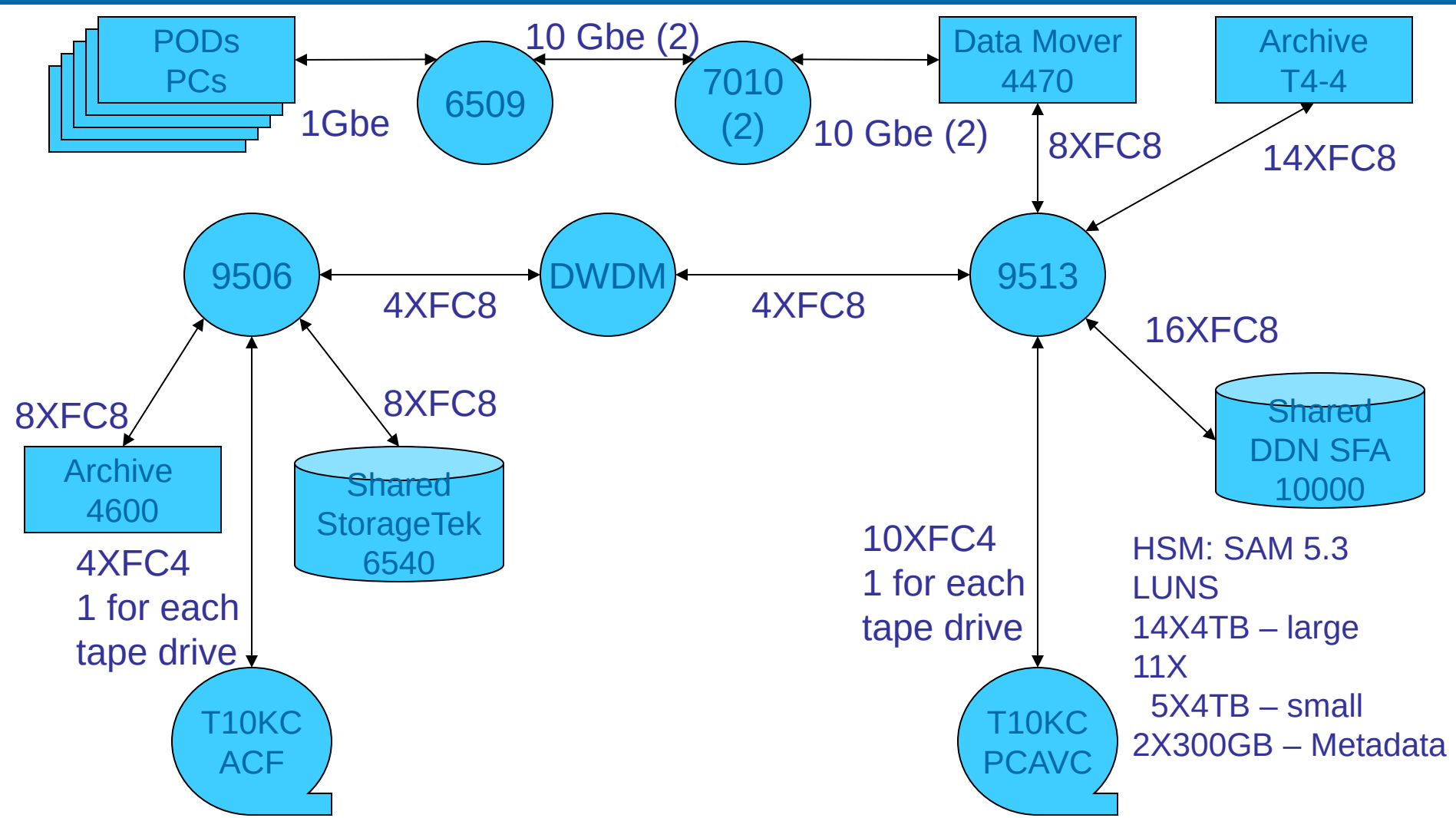
- Archive Server
- Web Server fulfills functions of Data Mover and Proxy Server. Workflow software runs only on this server.

Functional Architecture – In Development

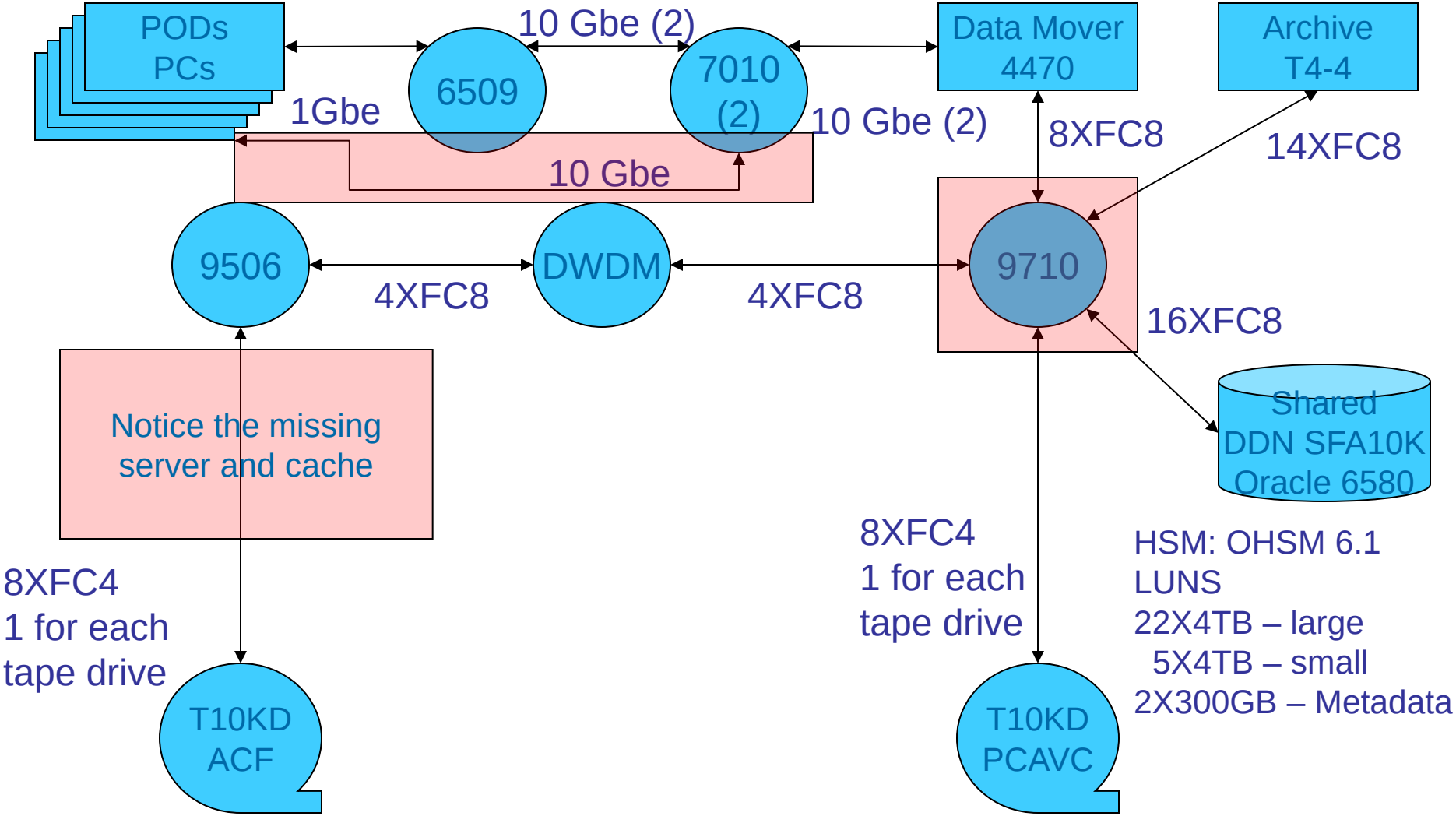


- Archive Server
- Web Server fulfills functions of Data Mover and Proxy Server. Workflow software runs only on this server.

Physical Implementation V2.2: 6 GB/s throughput 2015



Physical Implementation V3: 6 GB/s throughput Current



Physical Implementation V4: 6 GB/s Future