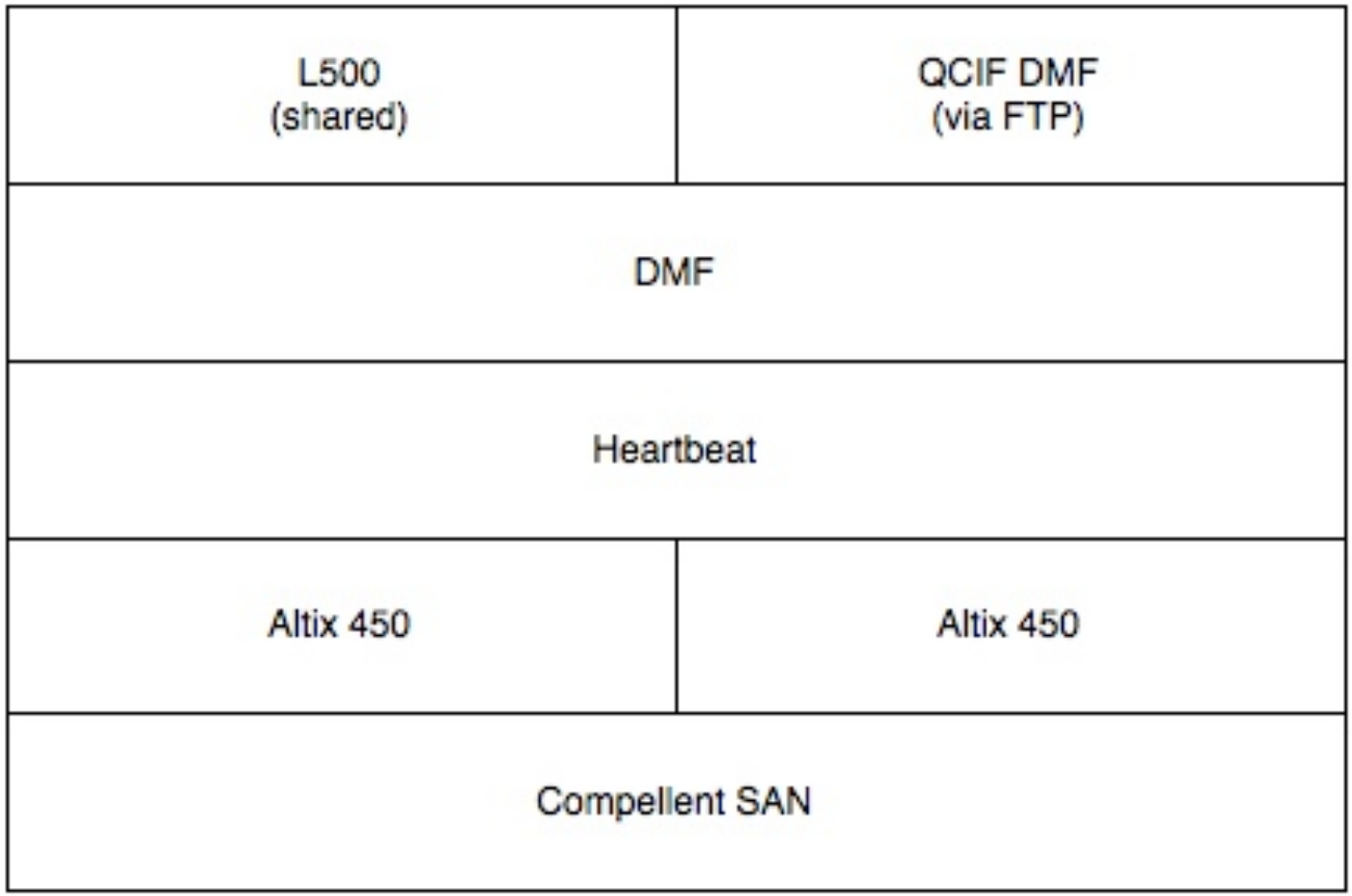
DMF Configuration for JCU HPC

Dr. Wayne Mallett
Systems Manager
James Cook University

DMF Environment



- Home=8GB, Journal=8GB, Spool=80GB, Move=80GB, Others=2x8GB
- Sizes chosen for LTO2, now using LTO4
- Two copies held on tapes locally
- FTP sends copies to QCIF DMF server
- 5 day old files are migration candidates
- File size isn't included in policies
- 1 robot in tape library, 4 tape drives for DMF
- Heartbeat installed/configured by SGI

- Altix servers provision storage to clients
- Clients connect to HPC storage via NFS (cluster) and Samba (Windows/OSX)
- NFS clients trigger mostly one-off recalls
- Samba clients often perform mass recalls
  - A number of users have tried to copy their entire directory tree onto USB disks
  - We only have 3.2TB of filesystem space
  - On three occasions, recalls of over 10TB have been triggered

Cascaded DMF

- Easy to configure
- Policies should be compatible
- Manual intervention may be required. Prolonged outages require removing FTP, followed by a syncing of files post-outage.
- Obtainable bandwidth needs to be considered when sizing filesystems
- Wouldn’t recommend as a primary destination but has worked well in a “DR” sense
- Time required for FTP transfer significantly increases time windows for migrations

General Issues at JCU

- 3 year funding cycle (no upgrade money)
- Shared infrastructure (SAN, tape library)
- Jobs running for 6 months or more
  - No schedule maintenance windows
- Network problems/outages
  - Internal: HA failover (stale NFS, prior to HA)
  - External: disrupting FTP transfers etc.

DMF Issues at JCU

- Dealing with growth (size and file count) on filesystems on “static” storage
- Keeping recalled files on disk long enough for client download (mass recalls)
- Client requests don’t get optimized (for tape)
- Performance (e.g., dirty SFPs, FTP)
  - How to prevent FS consumption during migrations
- Hardware/Software compatibility

Wishlist (not just DMF)

- Native windows experience from Samba?
  - enabling ITR to use DMF behind fileshares
- Server capturing recall requests (ActiveMQ like) and optimizing them on the fly?
- Method for preventing timeouts in client file transfer applications?
- A means to reduce migration (triggered by backups) windows – allow DMF to resume normal operations earlier
- More funding ($)

JCU HPC Hardware Overview

- 40 x (AMD, 2-cores, 8GB RAM, 72GB SCSI)
- 35 x (AMD, 8-cores, 16GB RAM, 300GB SATA)
- 2 x (AMD, 8-cores, 16GB RAM, 146GB SAS)
- 3 x (AMD, 16-cores, 32GB RAM, 146GB SAS)
- 2 x Altix 450 (NAS/HSM servers)
- Compellent Storage (7.5TB – FC and SATA)
- Brocade 5300 FC switches (shared)
- Nortel 1Gb/s IP switches (ATS protection)
- SL500 tape library (shared)
- 4 LTO4 tape drives, 160 LTO4 tapes

JCU HPC Software Overview

- xCAT used for cluster management
- XEN used for delivery of VMs
- CentOS 5 (recently signed onto RHEL CAUDIT)
- NFS delivery of filesystems to cluster
- Samba delivery of files to user desktops
- Torque/Maui

JCU HPC Usage Summary

- ~240K jobs submitted in 2009
- Maximum number of jobs in queue 8726
- 4 users with jobs un-runnable – required more than 32GB of memory
- Longest running job (PAUP, 152 days and counting)
- 84% of jobs require only 1 core, but mostly more than 1-cores share of RAM
- Most MPI jobs consumed 8-cores (1 system).
- Several users using more capacity than we have disk space
- New file creations have exceeded 1TB/day on several occasions