GridPP
UK Computing for Particle Physics

RAL Site Report

HEPiX - Rome 3-5 April 2006

Martin Bly

CCLRC

Overview

- Intro
- Storage
- Batch
- Oracle
- Tape
- Network
- T1-T2 challenges

RAL Tier1/A

- Rutherford Appleton Lab hosts the UK LCG Tier-1
  - Funded via GridPP project from PPARC
  - Supports LCG and UK Particle Physics users and collaborators
    - VOs:
      - LCG: Atlas, CMS, LHCb, Alice, (dteam)
      - Babar
      - CDF, D0, H1, Zeus
      - bio, esr, geant4, ilc, magic, minos, pheno, t2k
    - Expts:
      - Mice, SNO, UKQCD
    - Theory users
    - ...

Storage I

- New: 21 servers
  - 5U, 24 disk chassis
  - 24 port Areca PCI-X SATA II RAID controller
  - 22 x 400GB Western Digital RE series SATA II HDD
  - 8TB/server after RAID overheads (RAID 6)
  - 168TB (10^12) total useable
  - Opteron server
    - Supermicro motherboard, 2 x Opteron 275 (dual-core) @ 2.2GHz, 4GB RAM, 2 x 250GB RAID 1 System disks, 2 x 1Gb/s NIC, redundant PSUs
  - Delivered March 10
    - Now in commissioning
    - Issues with some RE drives
    - Expected in service late May
  - Running SL4.2, possibly SL4.3, with ext3
    - Issues with XFS tests

Storage II

- Existing: SCSI/PATA and SCSI/SATA
  - 35TB of 1st year storage now 4 years old
    - Spares difficult to obtain
    - Array PSUs now considered safety hazard
    - To be decommissioned when new capacity ready
      - Unless the power fails first!
  - ~40TB of 2nd year storage out of maintenance
    - Obtaining spares for continued operation
  - ~160TB of 3rd year storage
    - Stable operation, ~20 months old
  - Migration of 2nd and 3rd year servers to SL4 in May/June

Batch capacity

- New: 200+kSI2K delivered March 10
  - Tyan 1U chassis/motherboard (S2882)
  - Twin dual-core Opteron 270s
  - 1GB RAM/core (4GB/chassis)
  - 250GB SATA HDD
  - Dual 1Gb NIC
  - In commissioning - 4 week load test
    - Expected to enter service late April early May
- Existing: 800kSI2K
  - Some now 4 years old and still doing well
    - Occasional disk and RAM failures
    - 2nd year units more prone to failures
- All running SL3.0.3/i386 with security patches

Oracle Services

- 2004/5:
  - Test3d project database machine: re-tasked batch worker
    - Lone Oracle instance within Tier 1 (RHEL 3)
  - FTS backend database added 2005
- 2005:
  - Two further machines for SRB testing (RHEL 3)
- 2006:
  - Load on Test3D system with FTS very high during transfer throughput tests to T2s
    - Impact both FTS and Test3D
    - Migrate FTS database to dedicated machine (RHEL 4 U3)
  - New specialist hardware for 3D production systems:
    - 4 Servers (2 x RAC) + FC/SATA array (RAID 10)
    - RHEL 4 U3, ASM
    - Commissioning

Services

- LCG UI and traditional front end services migrated to ‘new’ racked hosts
  - Faster, 4GB RAM, 1Gb/s NIC
  - DNS quintet with short TTL
  - Additional system specifically for CMS
    - Needs service certificate for Phedex
- Migration of NIS, mail etc from older tower systems
- Nagios monitoring project
  - Replace SURE
  - Rollout April/May

Tape Store

- New:
  - 6K slot Storagetek SL8500 robot
  - Delivered December 2005
  - Entered production service 28th March 2006
  - 10 x 9940B drives in service + 5 on loan
  - 10 x T10000 drives + 2 EPE drives on evaluation
  - 1000 x T10K media @ 500GB each (native capacity)
  - Expand to 10K slots summer 2006
  - ADS caches: 4 servers, 20TB total
  - Castor2 caches: 4 servers, 20TB total
  - New ADS file catalogue server
    - more room for expansion and load sharing if Castor2 implementation is delayed
- Existing:
  - 6K slot Storagetek Powderhorn silo to be decommissioned after expansion of new robot to 10K slots.
  - All tapes now in new robot


StorageTek
StorageTek
Emergency Exit

Castor2

- 4 disk servers
- 10 tape servers
- 6 services machines
- Test units

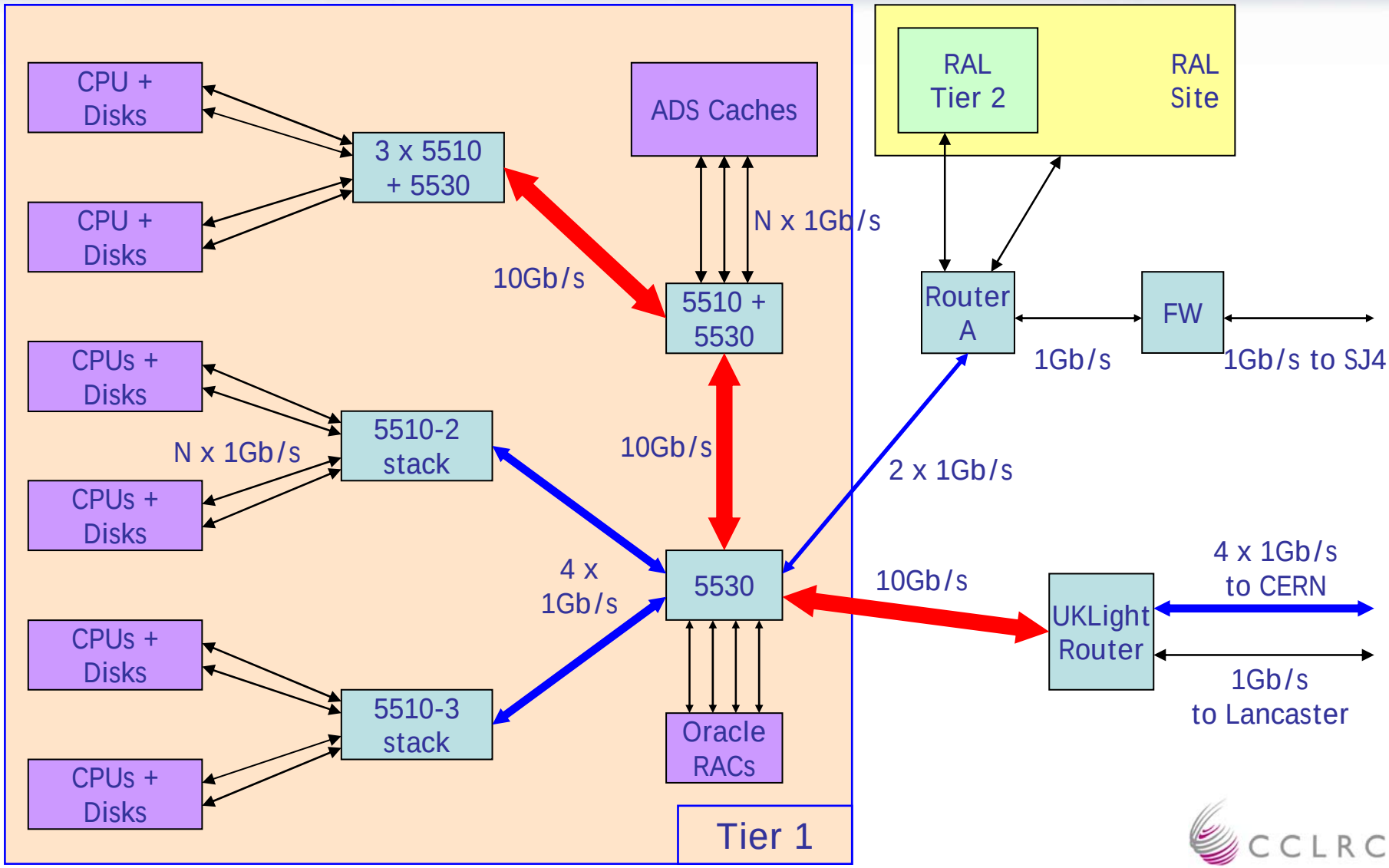
Schedule for CASTOR 2 Deployment (evolving)

- Mar-Apr 06
  - Testing: functionality, interoperability and database stressing
- May-Sep
  - Spec and deploy hardware for production database
    - May: Internal throughput testing with Tier 1 disk servers
    - Jun: CERN Service Challenge throughput testing
    - Jul-Sep: Create full production infrastructure; full deployment on Tier1
- Sep-Nov
  - Spec and deploy second phase production hardware to provide full required capacity
- Apr 07
  - Startup of LHC using CASTOR at RAL

Network developments

- 10GE backbone for Tier1 LAN
  - Nortel 5530s with SR XFPs
    - Partially complete, some 4 x 1Gb/s remaining
    - Full backbone in May
  - Looking at potential central 10GE switch solutions
- 10GE link to UKLight
  - CERN link @ 4x1Gb/s, Lancaster Tier 2 @ 1Gb/s
  - Expect CERN link @ 10Gb/s summer 2006
- Tier 1 link to RAL site backbone now @ 2 x 1Gb/s
  - Expect 10GE site backbone late spring 2006
    - Tier 1 will get 10GE link thereafter
- RAL link to SJ4 (WAN) @ 1Gb/s
  - Expect new link to SJ5 @ 10Gb/s during autumn 2006
    - High priority in SJ5 rollout program
    - Firewalling @ 10Gb/s will be a problem!

RAL Tier1 Network Connectivity March 2006



GridPP service challenges

- Program of tests of the UK GridPP infrastructure
  - Aim: stress test the components by exercising both the T1 and T2 hardware and software with extended data throughput runs
  - 3 x 48-hour tests
  - Opportunity to demonstrate UK T2s and T1 working together
- First test: trial at high data-rate (100MB/s) to multiple T2s using production SJ4 link (1Gb/s)
  - Severe stress to RAL site network link to SJ4 – T1 using 95% of bandwidth
    - Reports of site services dropping out: VC, link to Daresbury Lab (corporate Exchange systems etc)
    - Firewall unable to sustain the load – multiple dropouts at unpredictable intervals
    - Test abandoned
- Site combined traffic throttled to 800Mb/sec at firewall. Firewall vendors working on the issue but retest not possible before Tier 1 starts SC4 work.
- Second test: sustained 180MB/s from T1 out to T2s
  - 100MB/s on UKLight to Lancaster
  - 70-80MB/s combined on SJ4 to other sites
- Third test: sustained 180MB/s combined from multiple T2s in to T1
  - Problems with FTS and Oracle database backend limit the rates achieved
- Overall: success
  - T1 and several T2 worked in coordination to ship data about the UK
  - Uncovered several weak spots in hardware and software

GridPP challenges - throughput plots