



GridPP

UK Computing for Particle Physics

RAL Site Report

HEPiX Spring 2012, Prague

23-27 April

Martin Bly, STFC-RAL



Science & Technology Facilities Council

e-Science

- News
- Developments
- Hardware
- Networking



- **New Departmental Structure at STFC**
 - Director of e-Science Department Neil Geddes has moved to be Director of Technology Department
 - Merge of e-Science and Computational Science and Engineering Departments
 - Director of Scientific Computing to be appointed, post advertised
- **John Gordon has stepped down as a Division Head**
 - replaced by David Corney
- **eduroam for on site for visitors after a very long wait**
 - Also available for trial by STFC staff

- Switch gear in power feeds to RAL site need replacement
 - Short notice, this summer
 - Because they are ancient
 - Switch off one feed path and rely on other one
- Initial assessment - gloom
 - Very much increased hazard of power failures
 - Predicted long-time to repair most scenarios except glitches - site might be without power from 7 to 28 days!
- Latest news - better
 - Estimate that can cut over to the other cable upstream of the switchgear in less than 48 hours
 - **But RAL at-risk period is May to October**

- Govt provided unexpected capital-only funding via grants for infrastructure
 - DRI = Digital Research Infrastructure
 - GridPP awarded £3M
 - Funding for RAL machine room infrastructure
- GridPP - networking infrastructure
 - Including cross-campus high-speed links
 - Tier1 awarded £262k for 40Gb/s mesh network infrastructure and 10Gb/s switches for the disk capacity in FY12/13
 - Dell/Force10 2 x Z9000 + 6 x S4810P + 4 x S60 to compliment the existing 4 x S4810P units.
 - ~200 optical transceivers for interlinks
 - Must use Force10 DA copper with S4810P - most other DA cables won't work!



- Filled up remaining space in main machine room with APC racks
 - 3 enclosed cold aisles - 160 racks
- Roof and end-row (sliding) doors from Nubis Solutions match to racks
 - New working practices
- One enclosed cold aisle has 12 in-row cooling units
 - Liebert x 32kW units
 - Use chilled water from main building supply
- Nominal N+1 room cooling capacity increased from 1500kW to 1850kW
- Server Lifter





- E-Infrastructure South
 - Consortium of UK universities
 - Oxford, Bristol, Southampton, UCL
 - RAL has 10% share
- Hosted at RAL
 - Provides access to large production GPU facility
 - 372 Nvidia Tesla M2090 GPUs
 - 10Gb and Infiniband networking
 - Panasas Fast Parallel IO
 - 2nd biggest in Europe
- User management, registration, authentication, authorisation, accounting by NGS Core Services



- The JASMIN super-data-cluster
 - UK and European climate and earth system modelling community.
 - Climate and Environmental Monitoring from Space (CEMS)
 - Facilitating further comparison and evaluation of models with data.
- 4.6 PB Storage Panasas at STFC
 - Fast Parallel IO to Compute servers (360 Cores)
 - Biggest Panasas installation in Europe
- Exploring NGS authentication models
- Cloud interface to the compute facilities



- Summary of previous report:
 - Various Dell R510s for small data servers for Facilities Data Service, provides interfaces into Castor for RAL site facilities and others.
 - 68 x 40TB 4U servers ordered for capacity storage - two suppliers
 - 10GbE, 2TB HDD, single CPU, 24GB RAM, 2.66PB total
 - Note that disks may be hard to get ☹
 - 15,000 HEP-SPEC tender completed evaluation, result just announced
- New since October
 - 68 x 40TB SATA disk servers with 10GbE
 - 26 x 4-node double-twin WNs, X5650/E5645, 4GB/core, twin disk
 - 40GbE/10GbE and 10Gbe/1GbE switches, more tape servers
- Gone: 86 x 6TB servers - 2006 generation
- To come: 2012 procurements

- Problem with one batch of 2011 storage servers
 - SM 24-bay, 2TB SATA drives, Areca 9260-8i RAD card, Mellanox 10BbE hybrid IB/NIC
- Vendor completed their acceptance testing requirement using on-board 1GbE NIC attached to the management network which should be reserved for the IPMI cards
- Problem 1: PXE stack not included in IB/NIC card firmware
 - Fixed: load correct firmware, but:
- Problem 2: Custom vendor BIOS unable to cope
 - White screen of death
 - Fixed: standard SM bios, but:
- Problem 3: Card not initialised quickly enough for Anaconda to see a NIC rather than IB
 - Bug in Anaconda
 - Fixed by manipulations in Quattor

- **Site**
 - Sporadic packet loss in site core networking (few %)
 - Fixed
 - ‘Early Summer’ - boundary routers to be replaced by two pairs of Extreme x670V units to provide resilient 40Gb/s connectivity for site and multiple 10GbE and 40GbE connections within site
 - Tier1 will get 2 x 40GbE resilient connection direct from own top-level routers
- **Asymmetric Data Transfer rates in/out of Tier1**
 - Investigations ongoing
- **LAN**
 - New Force10 routers and core switches to be deployed with mesh-type arrangement linked at multiple 40Gb/s with storage connectivity at 10Gb/s
 - PerfSonar monitoring for LHCOPN
 - Rollout across GridPP infrastructure in progress

