Royal Holloway Site Report

Tom Crane, Szymon Gadomski, Simon George, Barry Green, Govind Songara, June 2016, HEP Sysman meeting @ RAL

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Royal Holloway Particle Physics

ROYAL HOLLOWAY UNIVERSITY OF LONDON

ATLAS

- benefit from strong collaboration software support
- large Tier3 batch compute and storage resources for data analysis
- DAQ test systems

Dark Matter

- detector development (lab DAQ systems)
- growing need for compute and storage resources to analyse data
- need help with things like sw installation, data movemement

Accelerator Physics

- small DAQ systems
- many small activities around the world which generate unique, valuable data sets
- simulation: both embarrassingly parallel and multi-process (MPI) computing
- software development infrastructure (e.g. cdash server)

Theory

- occasional use of Tier3 cluster
- interest in MPI

RHUL PP Computing

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Newton Tier-2:

- 3600 job slots
- 1439 TB of storage (DPM Grid SE)
- some critical services are virtualized
- located in the modern computer center on the Huntersdale site

PP Admin Team:

- Simon George
- Barry Green
- Tom Crane
- Govind Songara (Tier-2)
- Szymon Gadomski (Tier-3)

Faraday Tier-3:

- recycled worker nodes
- 610 job slots
- 138 TB on NFS
- 186 TB in Hadoop
  - ~60 batch nodes have 3 TB disks, 3 copies of the data
- 1 node 64-core system for MPI jobs
- 1 GPU server
- critical services are virtualized
- located in the machine room of the Physics building
- Linux desktops in the department mount the same /home and other NFS directories

Newton Tier-2

ROYAL HOLLOWAY UNIVERSITY OF LONDON

8 racks of the T2 @Huntersdale

Tier 2 news - the latest upgrade (Spring 2016)

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Limited by rack space and cooling

- Forced to decommission 300 TB of storage from 2008 to make way for new equipment.
  - (Discovered DPM bug in draining which caused most file to be lost.)

Spending summary:

- £125k from GridPP
- £75k from from RHUL infrastructure funds
- Spent £187k on compute nodes from XMA
- Remaining £13k from RHUL spent on network equipment (incl. upgrade to 10Gb link), spares and service node.

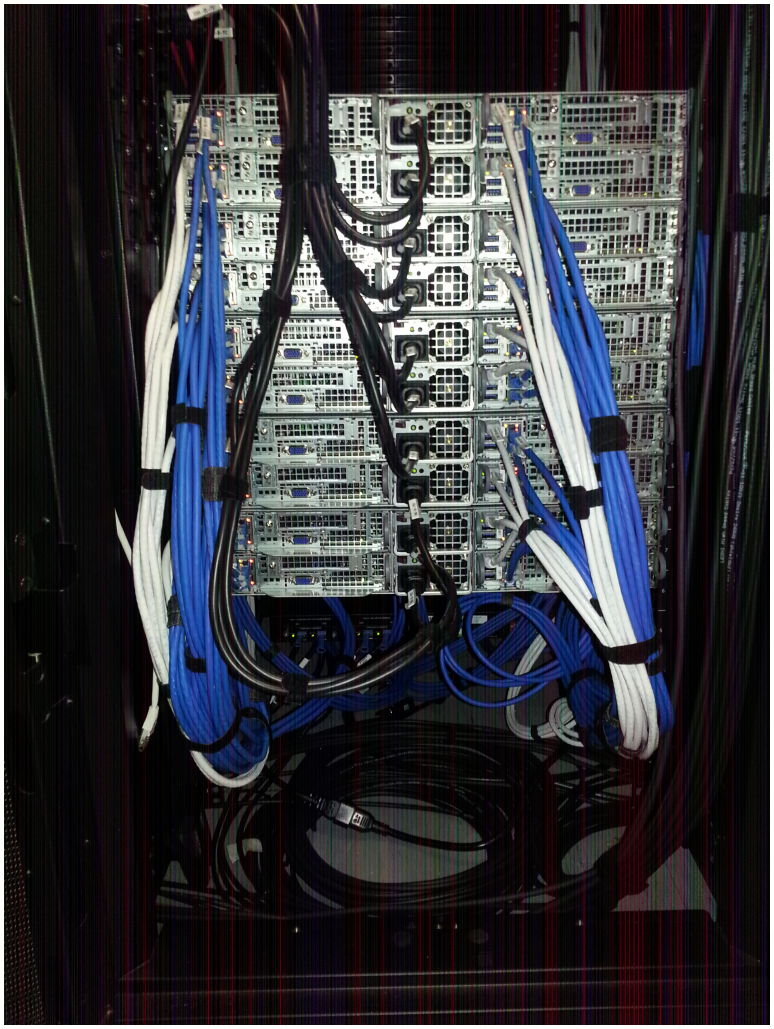
New compute nodes

13 x XMA HX625T2i 2U quad node chassis; each node has:

- 2x Xeon E5-2640v3 CPUs (16 virtual cores with HT)
- 128GB DDR4-2133 ECC RAM (8 GB/core)
- 2x2TB SATA disks, RAID0 stripe for job working area performance
- 2x1Gb ethernet (which we bond in alb mode)
- IPMI v2.0 with KVM over LAN.
- 3 year NBD support

Total 16.4kW 832 cores 19kHS06 (nominal)

New servers in the Newton Tier-2 farm



Tier 2 news - upgrade experience

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Purchased as mini-competition through NSSA, XMA won.
- Rack depth was a key issue in tender (800mm usable depth from front due to vertical PDUs at back)
- Install and commission went smoothly and just about on schedule
- No problem for them to integrate with SL6-based Alces stack (no longer supported)
- Unrelated internal network issues delayed commissioning; now fully integrated behind CREAM CEs.
- Struggling to reproduce claimed HEPSPEC06 figure
- Mixed experiences with XMA support so far but they are clearly eager

Faraday Tier-3 farm

ROYAL HOLLOWAY UNIVERSITY OF LONDON

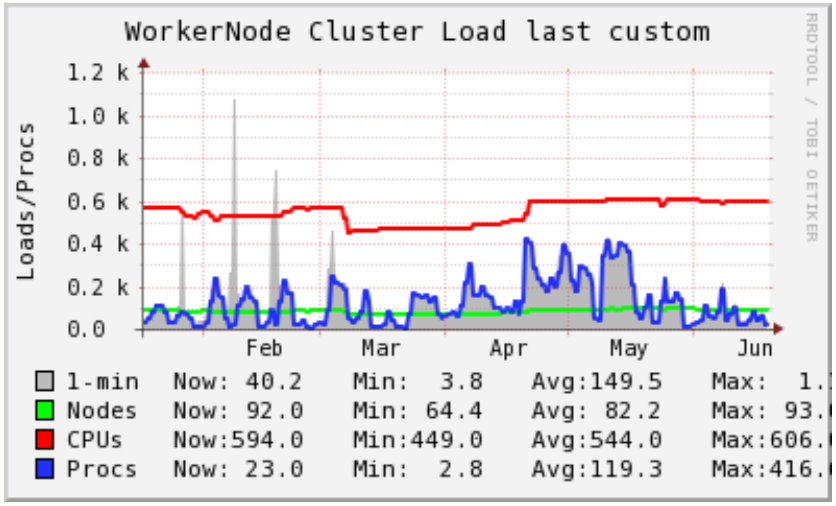
In total 9 racks, 6 shown here are the compute/hadoop nodes

Faraday farm report

Running smoothly and getting work done

Recent developments

- Non user facing services moved from SLC6 to CC7
- Completed transition from SLC5 to SLC6 for batch workers and UI servers
- A new GPU server (1 GeForce GTX 970 card, 1664CUDA Cores)
- Adding recycled storage servers
- Replacing H/W router NAT with a Linux VM
- Virtualised Win2003 server when H/W failed
- A backup firewall on a VM for emergencies



- Nagios event handler to prevent Black Hole Nodes
- Retired problematic Dell switches and replacing them with HP ones
- Improved script to clean up WNs after jobs have finished
- physics dept twiki running on RHEL6 VM hosted by IT Service

Recycled storage servers in the Faraday farm

Supermicro, purchased in 2008
IPMI cards from the USA via ebay!
A few like that will be used.
No shortage of spare parts.

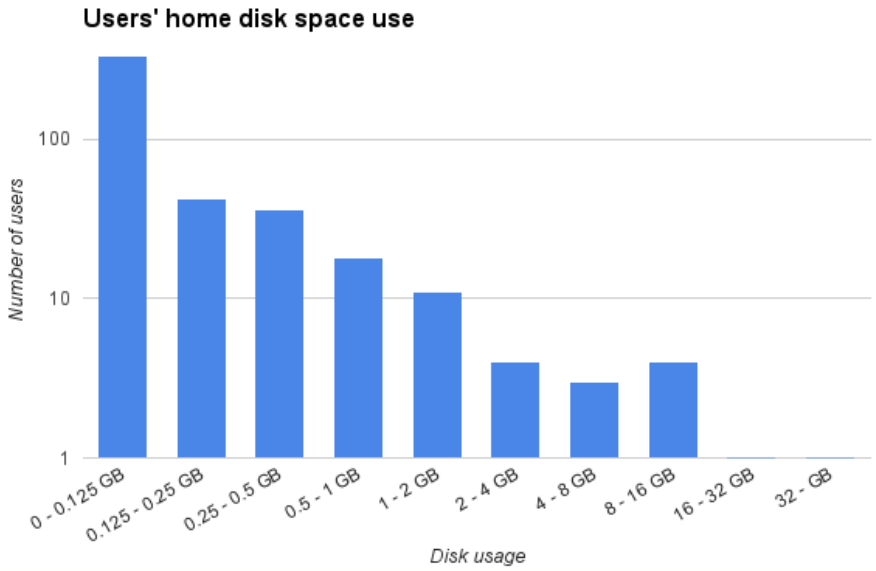
The "home" file service

Concept: a safe and reliable place to keep your files

Current system

- Two x86_64 servers with 10-15 TB RAID6 ZFS Solaris 10 (with Oracle sw support for updates)
- One is the /home system
- Second is a backup system: every night home is rsynced to backup, then a ZFS snapshot is done on backup
- 30 days of snapshots, users can access them to self-recover lost files
- Off-site backup of the backup (daily via HP Protector, a College IT service)

Success! Extremely stable and reliable for ~8 years.



Current usage of the /home

- default quota has evolved from 50 MB to 2 GB
- even now only a few power users have more

Re-imagining the "home" file service

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Requirements and usage patterns are changing

- Shift from desktop to laptop O(100GB), with (ad hoc) backup strategies, e.g. cloud, external drive, nothing
- NFS is not laptop-friendly, no dropbox-like sync service, and we don't export beyond our firewall
- One group actually set up its own NAS server with sync client to keep all user's laptop data on a more systematic basis
- Home still used for code dev, e.g. to run on cluster

Considering the following model

- 1TB per user (>current laptop disk size)
- NAS appliance(s)
- NFS-mounted on workstations and cluster as before
- Sync client for Linux/Windows/Mac to keep laptops backed up and allow access from multiple devices (Android, IOS)
- File versioning - better than backups for fixed number of days
- Possible hybrid local/cloud leveraging existing, under-used cloud storage (OneDrive for business)

Watching college data preservation strategy for other options

Summary

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Both farms running well and reliable
- An upgrade of the T2 was done with success
- Hardware retired from the T2 is useful in the T3
- A few ideas how to develop the T3 further
  - update of PP group /home service
  - migrate authentication to college AD (ldap)
  - 2nd name node for hadoop
  - maybe set up a small ~400 core MPI cluster