Royal Holloway site report

Simon George

HEPSYSMAN @ RAL 10-11 Jun 2010

Contents

- Staff & basics
- Tier2 relocation & upgrade
- Tier3 cluster upgrade
- Local storage upgrade
- CC services
- Forthcoming procurements

Computing staff & basic portfolio

- Diego Bellini (main sysadmin)
  - Leaving 2 July
- Govind Songara (Grid site admin)
- Simon George (local/Tier2 manager)
- Barry Green (hardware & general support)

- Maintaining usual local systems & services
  - Desktops (SLC5), laptops (Win/Mac/Linux), printers
  - Windows TS
  - Network and firewall, Wifi
  - Nagios, DHCP/PXE/kickstart, DNS, email
  - TWiki, web, SVN, elog, etc.
  - Videoconference suite

Tier2 cluster

- Govind Songara, Duncan Rand (LG tech coord)
- Feb10 – successful relocation from IC to RHUL machine room
  - Installed in new APC racks
- Upgraded to SL5 and BCM
- Completely reinstalled gLite3.2 WN, SE, DPM
  - Kept old SL4 CE
- Now back up and testing/running for main VOs.
- It took much longer than anticipated to build the site back up from scratch, but it was a good learning experience.

Total CPU: 7.9 HS06 x 400 cores = 3160 HS06

Total storage: ~21 TB x 14 servers = ~294 TB

Ratio HS06/TB ~11

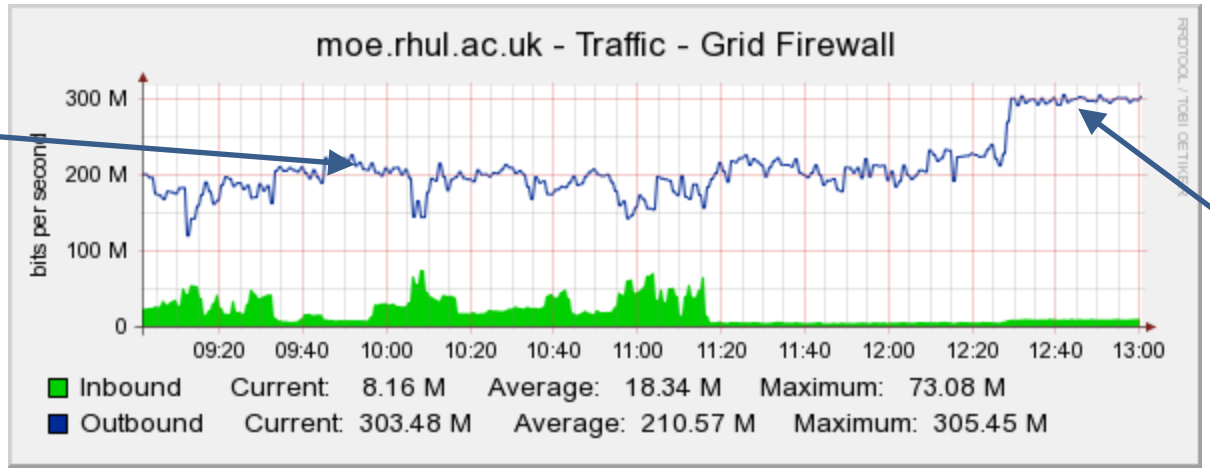
WN network: 2Gb/s bonded per 8 cores

Storage network: 4Gb/s bonded per server

Tier2 network

- Bad news: sharing 1Gb/s network with college
- Dedicated 1Gb/s line still not provided
- Long chain of request (includes LMN, BT)
- Delayed by road digging problems
- Hoping to get temporary additional connection through existing fibre.

Atlas transfers running at ~200 Mb/s



Capped at 300 Mb/s

Tier3

- Govind Songara, Diego Bellini, Barry Green
- Old Tier2 cluster circa 2004:
  - 75 nodes x 2 P4 Xeon 3GHz, 2 GB RAM (4.17 HS06)
- Upgraded 40 nodes:
  - AMD Phenom II 955 3.2GHz, (4 core) + 8GB RAM, to fit inside current power & heat constraints.
- WNs built by academics!
  - from our detailed instructions
  - ~4 man days
- For local use:
  - Atlas analysis
  - Accelerator simulations (MPI)
- Rapid deployment using our standard workstation PXE/kickstart install
- Torque/maui
- Next: MPI, cobbler/cfengine.

Total CPU:

13.25 HS06 x 160 cores = 2120 HS06

(was 625)

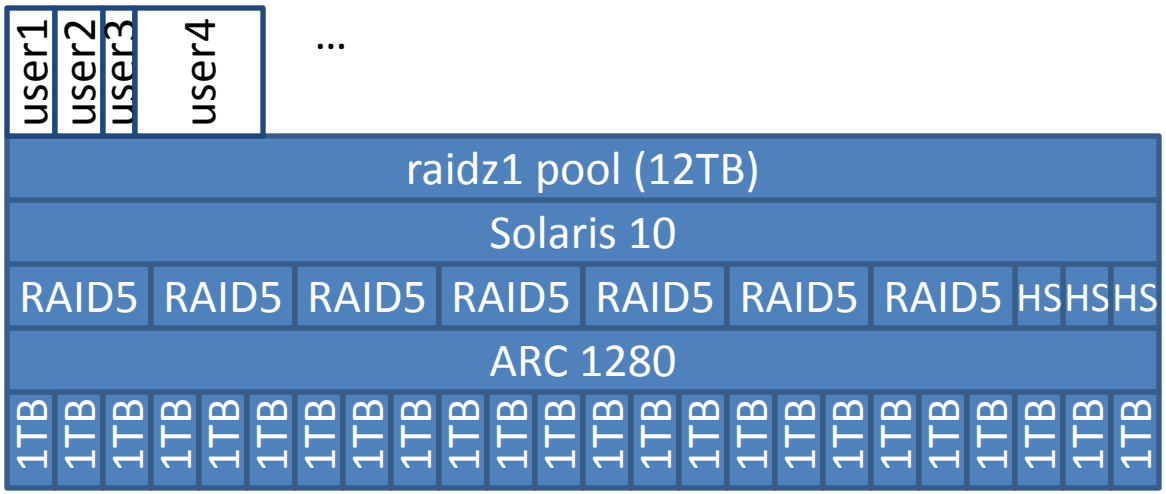
Total storage: 60 TB scratch via NFS

Local storage

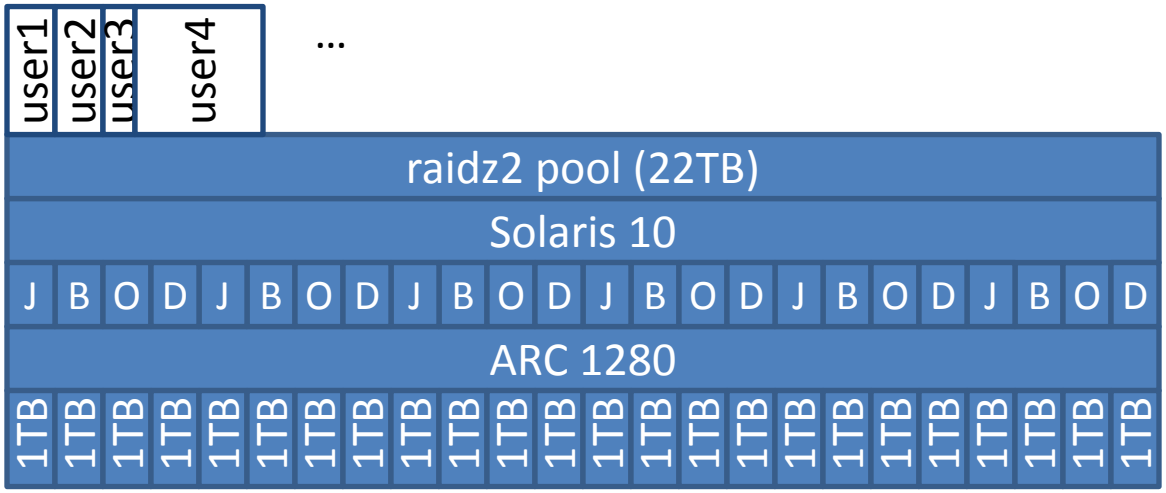
- To meet demands of local ATLAS analysis and expanding accelerator group
- **Scratch server**
  - PC hardware, 48-slot server, 2TB disks
  - 2x ARC1280
  - 4xRAID6+8HS (lower rebuild times) => 64TB
  - Heavily run in (repeatedly fill and file system) to tease out ~10 dodgy disks
  - Linux SLC5, XFS
  - NFSv3, bonded 2Gb/s ethernet
  - ~60 clients including 40 Tier3 WNs
- **'Home' file space with backups**
  - To replace old sun E450 ~100 GB and LTO2 20-tape library with ~12 TB + off-site backups
  - PC hardware (like Tier2 storage)
  - Solaris 10, ZFS, software support contract
  - NFSv3, bonded 2Gb/s ethernet
  - Backup to CC system
  - Problems encountered:
    - HP protector: won't visit nested file systems – see later
    - Solaris/ZFS ARC cache exhausting system memory - limited it and plan mem upgrade
    - Disk (Samsung HD103UJ) timeouts too slow for controller – drop out although fine. Tuned some settings in firmware.

‘Home’ server configurations

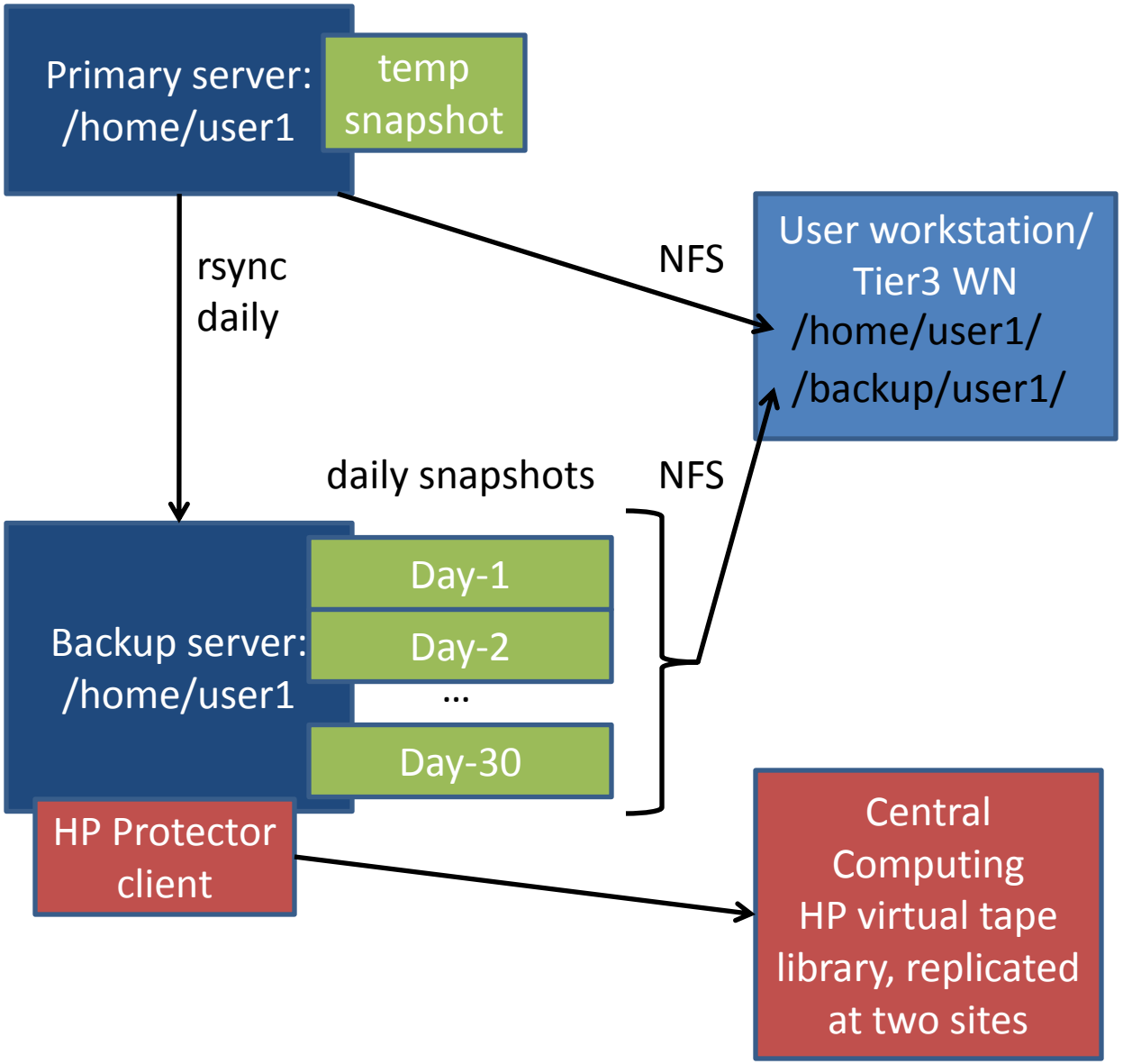
Primary server: optimised for reliability (double raid5) and recovery time



Backup server: optimised for space but still raid6 for reliability



Backup procedure



- Provide users with self-service backups
- Snapshots => no duplication, space efficient
- 4 weeks is easily accommodated

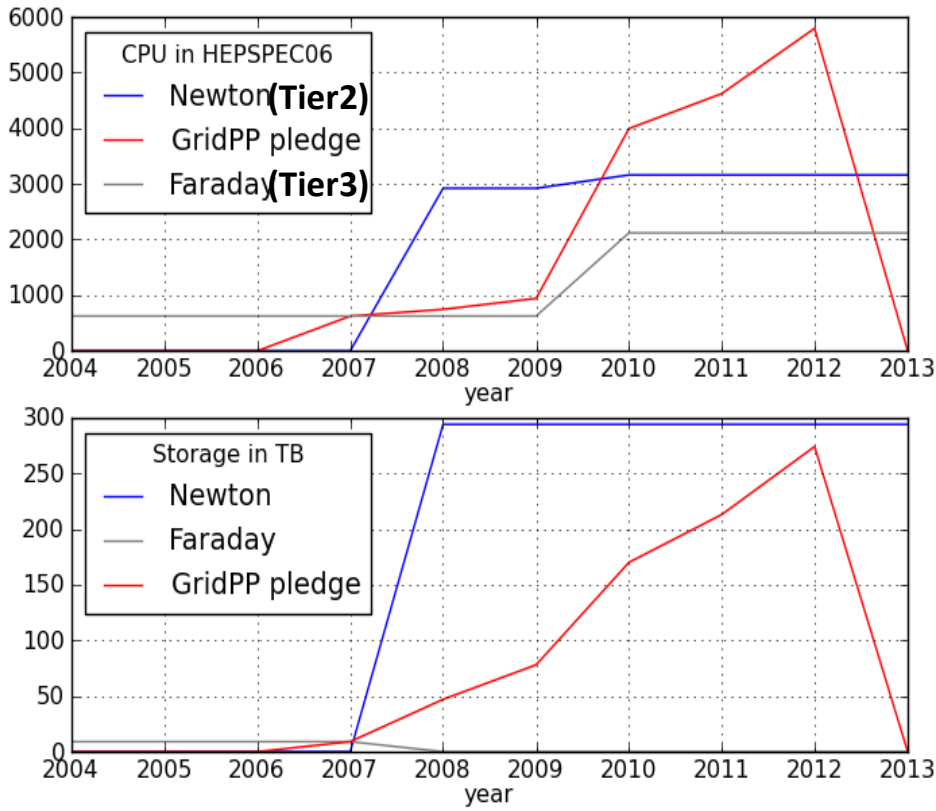
- Replace old LTO2 tape system
- Using HP protector for off-site backup to CC virtual tape library
- Auto replication across 2 sites
- Problem so far: cannot visit nested fs – maintenance of /home/* list by hand in GUI.
- Workaround: copy snapshot to temporary single fs before backup.

Central Computing services

- In the light of:
  - Good experience of CC hosting our Tier2 cluster
  - CC having upped their game significantly in recent years
  - Reduction in funded HEP sysman effort
- Review what can be done by CC
  - Benefit from their economies of scale
  - Reduce HEP sysman effort
  - Refocus on what we do that is unique
  - Maybe increase equipment/consumables costs
- Examples: RHEL5 VMs, NAS, virtual desktops, backups
- So far
  - just migrating services that are replicated centrally
  - E.g. email & DNS
  - Hope to start tests with VMs and NAS soon

Forthcoming procurements

- Tier2: need to double HS06 by 2012 to meet GridPP pledge
- Storage within pledge but will also increase to maintain balance with CPU
  - Target ratio ~10 HS06/TB
- First round later this year – focus on CPU
- Anyone using the ATLAS/Dell deal?
  - How good are the prices?
- Push for 2 Gb/s network to LMN once we have 1 Gb/s



Conclusions

- Trends:
  - Grid expansion continues, both Tier2 & 3
  - Network link to LMN is still bottleneck for Tier2
  - Lost support effort in grant => do less
  - Moving to model where computer centre provide services and hosting.
  - Emphasis switching from local IT support to running Tier2/3 and managing SLAs.