

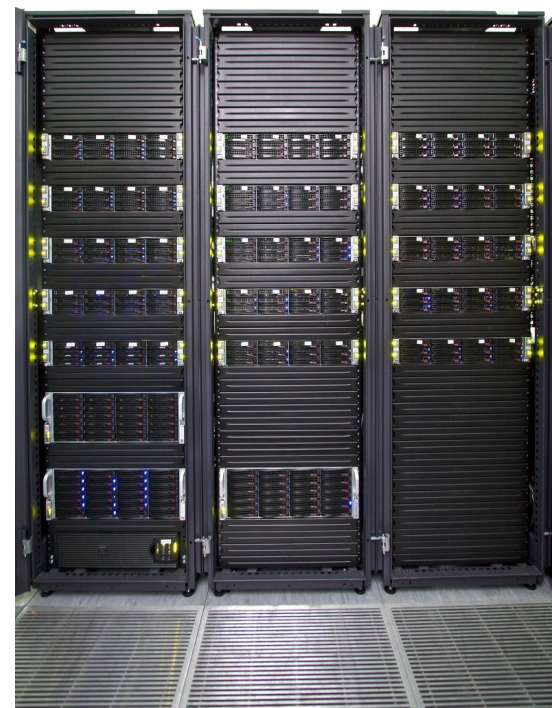


UNIVERSITY OF
LIVERPOOL

Liverpool HEP - Site Report

June 2011

John Bland, Robert Fay



Staff Status



No changes to technical staff since last year:

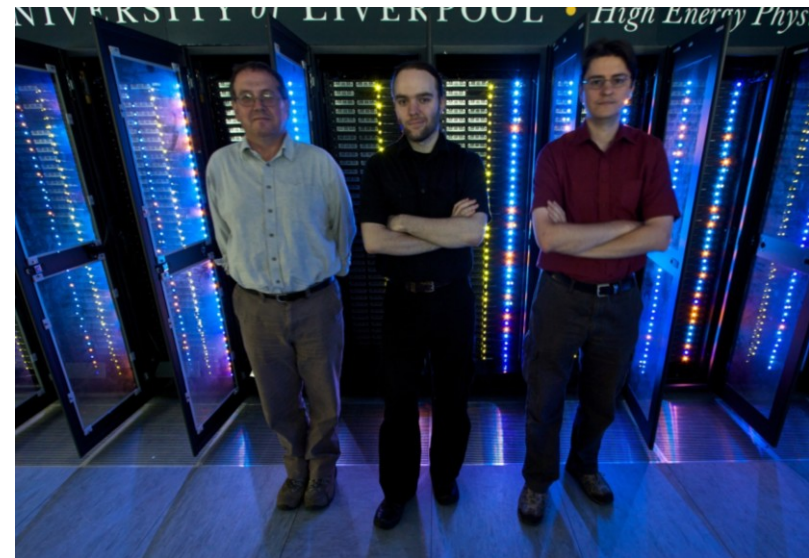
Two full time HEP system administrators

- John Bland, Robert Fay
- David Hutchcroft (Academic)

One full time Grid administrator

- Steve Jones

- Sir Not-appearing-in-this-report (TBD)





Current Hardware - Users

Desktops

- ~100 Desktops: Scientific Linux 5.5, Windows 7+XP, Legacy systems
- Minimum spec of 2.66GHz Q8400, 4GB RAM + TFT Monitor
- Recently upgraded, clean installs, single platform
- Opportunistic batch usage (~60cores)

Laptops

- ~60 Laptops: Mixed architecture
- Windows+VM, MacOS+VM, Netbooks

Printers

- Samsung and Brother desktop printers
- Various HP model heavy duty group printers



Current Hardware – ‘Tier 3’ Batch



‘Tier3’ Batch Farm

- Software repository (0.5TB), storage (3TB scratch, 13TB bulk)
- Old 32bit kit and queues recycled
- ‘medium64’, ‘short64’ queues consist of 9 64bit SL5 nodes (2xL5420, 2GB/core)
- 2 of the 9 SL5 nodes can also be used interactively
- 5 older interactive nodes (dual 32bit Xeon 2.4GHz, 2GB/core)
- Using Torque/PBS/Maui+Fairshares
- Used for general, short analysis jobs
- Grid jobs now also run opportunistically on this cluster (not much recently due to steady local usage)

Current Hardware – Servers

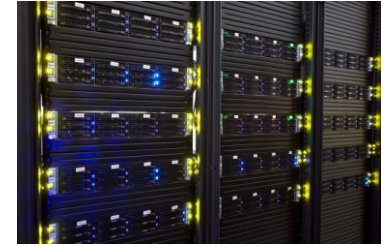
- ~40 core servers (HEP+Tier2)
- Some rack Gigabit switches
- 1 High density Force10 switch (400 ports)
- Console access via KVMoIP (when it works) + IPMI

LCG Servers

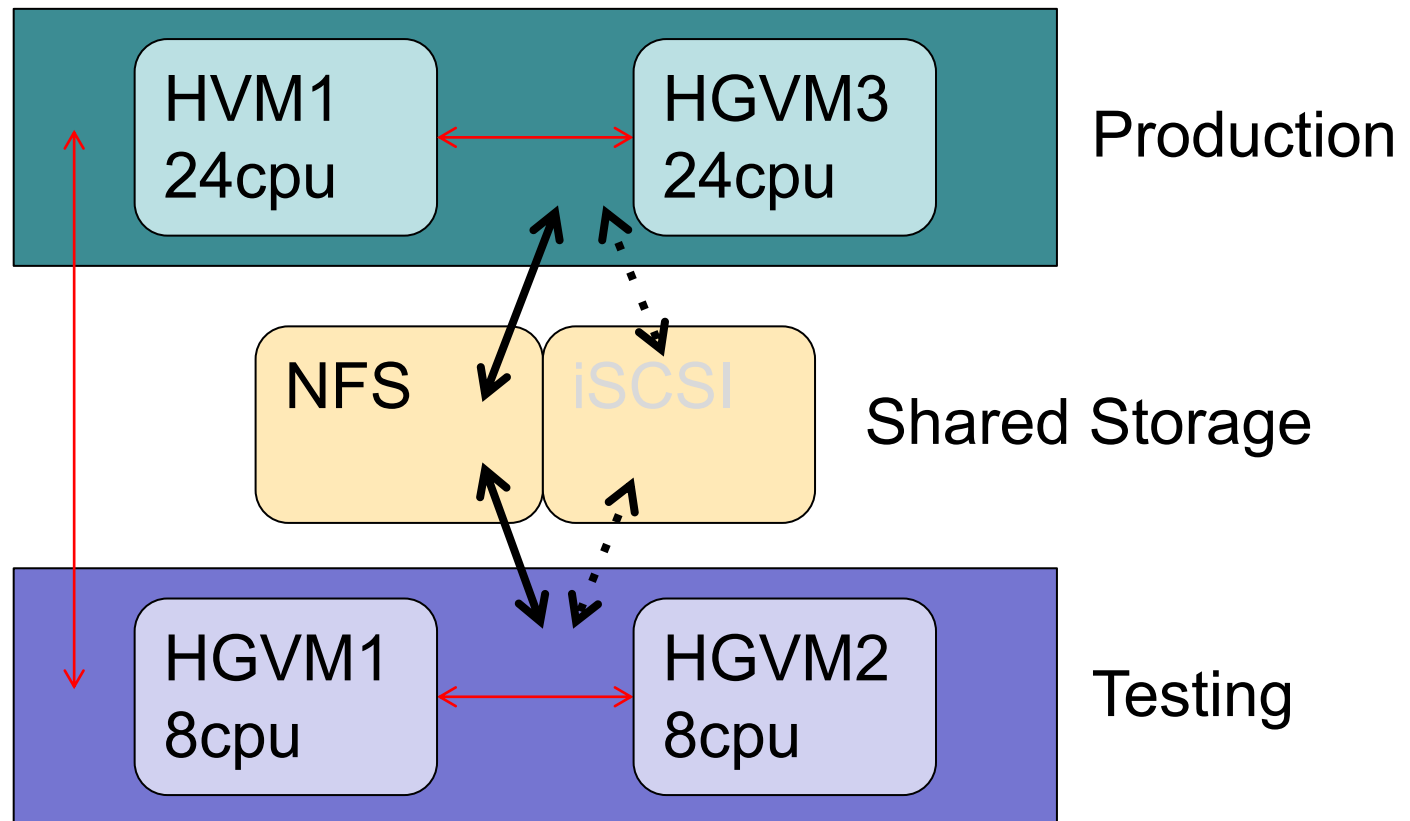
- Many upgrades since last year
 - Most services on Virtual machines (lcg-CE, CREAM*2, sBDII, APEL, Torque, ARGUS)
 - Still more to come (UI, mini virtual test cluster)
 - Old SL4 service nodes upgraded to SL5+kickstart/puppet
- VMware Server retired, KVM much more reliable
- Setting up redundant KVM servers for load sharing and testing



Current Hardware – Virtual



- Heavy duty KVM servers for grid and HEP services



Current Hardware – Nodes

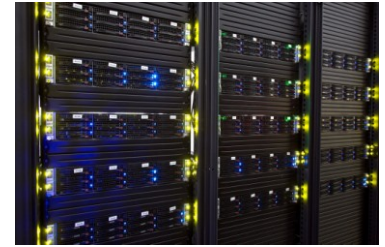


MAP2 dead and buried:
Good riddance!

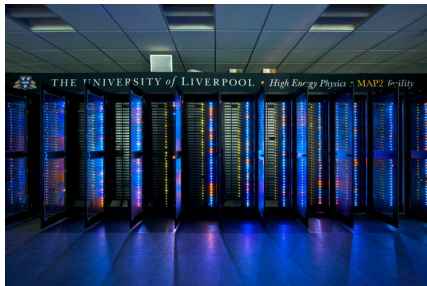


They even paid us to take it away! (Suckers)

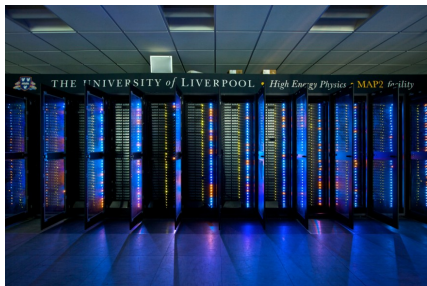
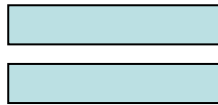
Current Hardware – Nodes



Replaced with shiny new boxes:



12 Racks x 40 Dell 650



12 Racks x 40 Dell 650



The Smug brothers with 10x4x5620s



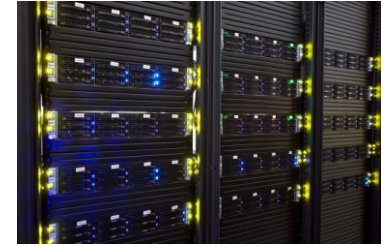
Current Hardware – Nodes

New cluster dubbed HAMMER

- Similar load per rack (~6kW)
- Lots of room to work in
- Rack space for new storage
 - Shame we haven't got enough cables
- IPMI+KVM makes life so much easier
 - Also hot swap drives and mboards
- IPMI monitoring not entirely trustworthy
 - Sensor reading spikes
 - Occasionally needs rebooting



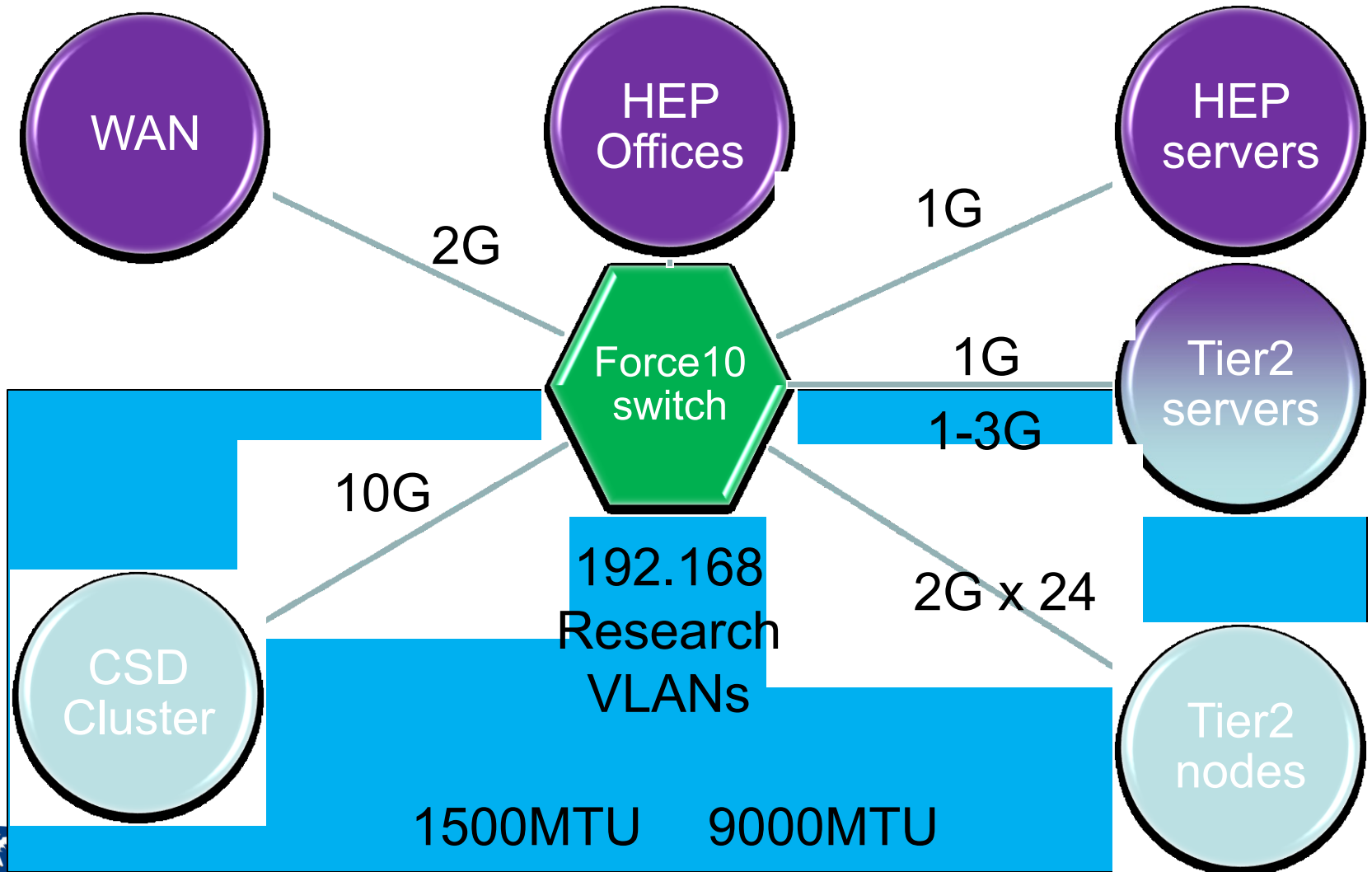
Current Hardware – Network



- Getting low on cables
- Network capacity our biggest upcoming problem
- Really need to move to 10G
 - No money locally
 - Research computing strategy may help
- Bonding starting to get ridiculous as density increases
 - But works very well



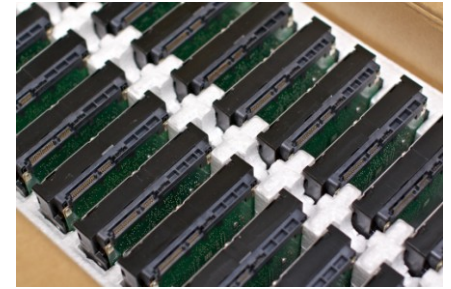
HEP Network topology



HEP Network (they told me to say this bit)

- Grid cluster is on a sort-of separate subnet (138.253.178/24)
 - Shares some of this with local HEP systems
 - Most of these addresses may be freed up with local LAN reassignments
- Monitored by Cacti/weathermap, Ganglia, Sflow/ntop (when it works), snort (sort of)
- Grid site behind local bridge/firewall, 2G to CSD, 1G to Janet
 - Shared with other University traffic
 - Possible upgrades to 10G for WAN soon
- Grid LAN under our control, everything outside our firewall CSD controlled
 - CSD aren't the most helpful or cooperative people you might wish for.

Storage



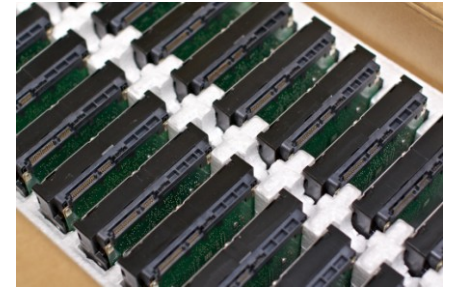
- Majority of file stores using hardware RAID6.
- Mix of 3ware, Areca SATA controllers and Adaptec SAS/SATA
- Arrays monitored with 3ware/Areca software and nagios plugins
- Software RAID1 system disks on all servers.
 - A few RAID10s for bigger arrays and RAID0 on WNs
- Now have ~550TB RAID storage in total. Getting to be a lot of spinning disks (~700 enterprise drives in WNs, RAID and servers).
 - Keep many local spares
- Upgraded some older servers 1TB->2TB
 - Trickle down of 1TB->0.75TB->0.5TB->0.25TB upgrades to come



UNIVERSITY

LIVERPOOL

• Also beefed up some server local system/data disks



Storage - Troubles

- Adaptec/SuperMicro firmware problems
 - Weeks of investigating delayed other upgrades
 - 'Fixed' with a firmware downgrade
 - Seems ok now but vague sense of unease persists
 - SuperMicro UK support is rubbish
- Areca firmware problems
 - Latest firmware fixes this
- Far too many drive failures
 - One or two instances of double disk failure - knuckle-biting rebuilds!

Joining Clusters



- Spent last few years trying to hook up UKI-NORTHGRID-LIV-HEP to NWGRID over at CSD
- Never really worked
 - Too many problems with OS versions
 - SGE bugs
 - Awkward admins
- Still sorta works for tests but LHCb+ATLAS software still not working fully
- Available resources lot less significant now
 - We have our own dedicated 64bit cluster now
- 10G fibre link may be useful to tap into fast WAN link so not a complete loss

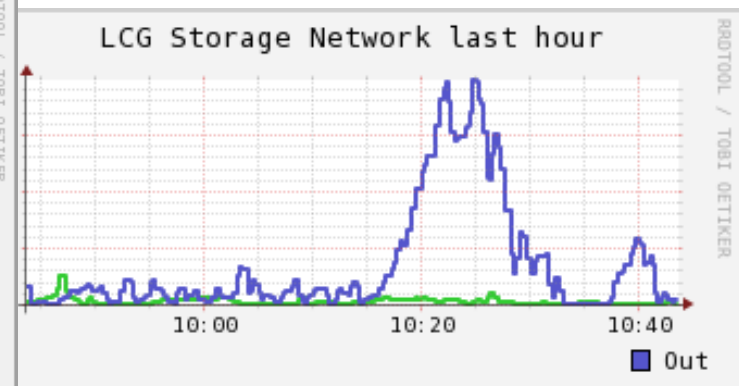
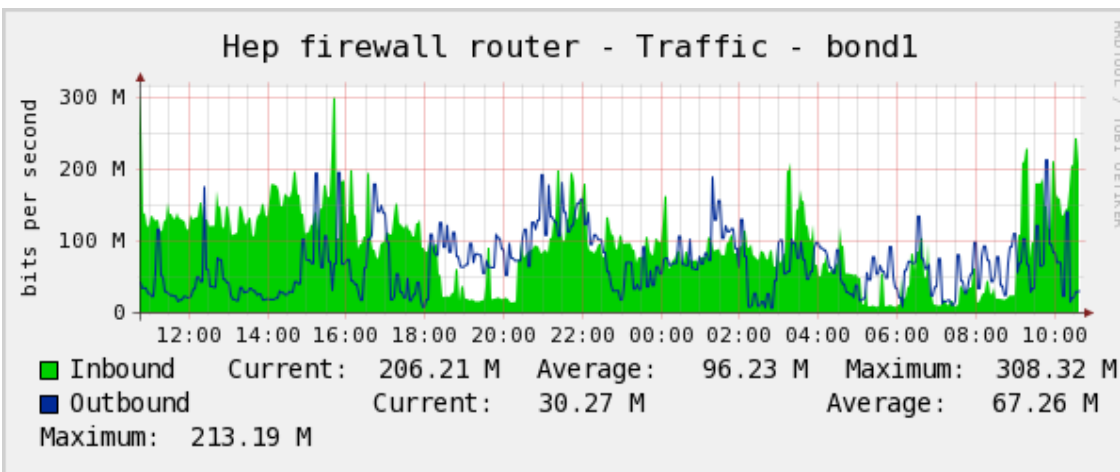
Configuration and deployment

- Kickstart used for OS installation and basic post install
 - Used with PXE boot for some servers and all desktops
- Puppet used for post-kickstart node installation (glite-WN, YAIM etc)
 - Also used for keeping systems up to date and rolling out packages
 - And used on desktops for software and mount points
- Custom local testnode script to periodically check node health and software status
 - Nodes put offline/online automatically
- Keep local YUM repo mirrors, updated when required, no surprise updates

Network Monitoring

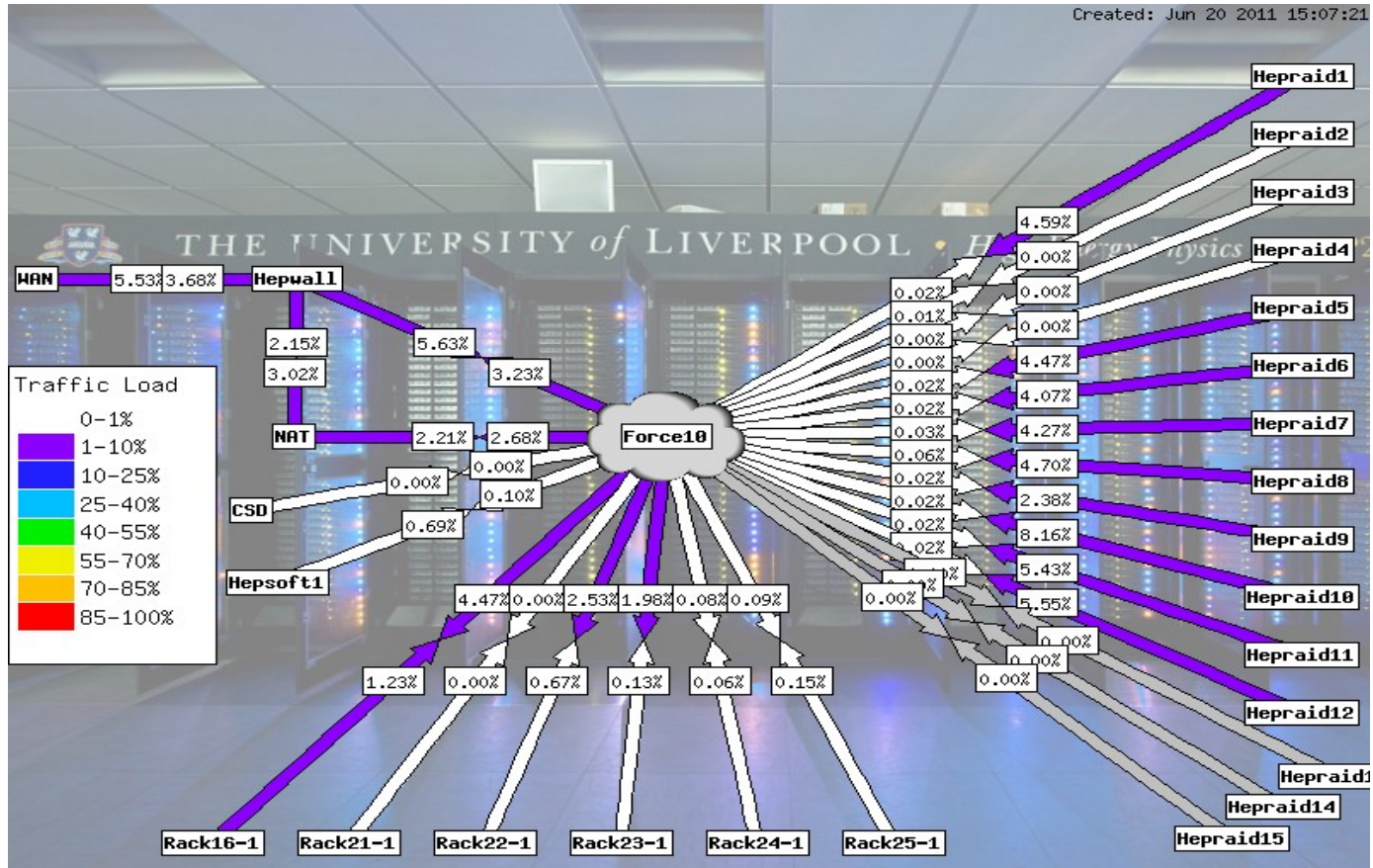


- Ganglia on all worker nodes and servers
- Cacti used to monitor building switches and core Force10 switch
 - Throughput and error readings + weathermap
- Ntop monitors core Force10 switch, but still unreliable
 - sFlowTrend tracks total throughput and biggest users, stable
- LanTopolog tracks MAC addresses and building network topology
- arpwatch monitors ARP traffic (changing IP/MAC address pairings).



Monitoring - Cacti

- Cacti Weathermap





Security

- Network security
 - University firewall filters off-campus traffic
 - Local HEP firewalls to filter on-campus traffic
 - Monitoring of LAN devices (and blocking of MAC addresses on switch)
 - Single SSH gateway, Denyhosts
 - Snort and BASE (need to refine rules to be useful, too many alerts)
- Physical security
 - Secure cluster room with swipe card access
 - Laptop cable locks (occasionally some laptops stolen from building)
 - Promoting use of encryption for sensitive data
 - Parts of HEP building publically accessible
- Logging
 - Server system logs backed up daily, stored for 1 year
 - Auditing logged MAC addresses to find rogue devices

Plans and Issues

- Liverpool Uni doesn't have a research computing strategy
 - Effort from the admin staff to create one
 - Physics should be a major player in this
 - Might even get some kit/money out of it!
- Local interest in GPGPU increasing
 - Getting a test Tesla box in to play with
- A few local users want SL6 (already!)
 - Far too early for experiments
 - Many useful software upgrades that users would like
 - Beta testing now, mostly everything's fine apart from LCG software
- Cluster room cooling still very old, regular failures
 - Bit more slack after the MAP2 removal but failures increasing
 - University data centre strategy meeting soon
 - With more groups interested in our room we might have more clout

They stole it from us!



- Tricksy Computing Services Department wants our network
 - In return we get a network... and no control
- They also wants to take our printers
 - In return we get some printers... and no control
- We mustn't let them have the precious(es)
- General University policy of centralising anything and everything
 - One size fits all computing doesn't work when you're Physics

Conclusion

- * New kit in, older kit recycled
- New kit+IPMI+KVM etc making life a lot easier
 - Might even have time to develop interesting things
- Big territorial fight(s) with CSD and university bureaucracy ongoing

