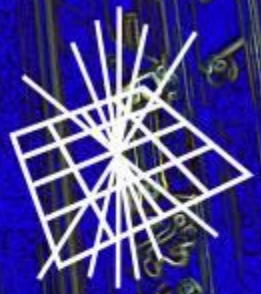


GridPP

UK Computing for Particle Physics

UKI-SouthGrid Update Hepix

Pete Gronbech
SouthGrid Technical Coordinator

April 2012

UK Particle Physics Computing Structure

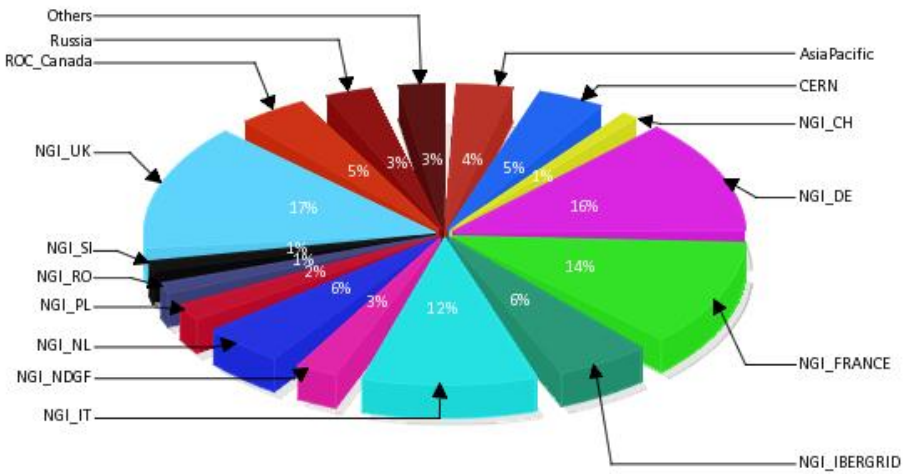
- One Tier 1 center - RAL
- 20 University Particle Physics Departments
- Most are part of a distributed Grid Tier 2 center

- All have some local (AKA Tier 3) computing

- SouthGrid is comprised of all the non London based sites in the South of the UK.
- Birmingham, Bristol, Cambridge, JET, Oxford, RAL PPD, Sussex and Warwick.

UKI Tier-1 & Tier-2 contributions

PRODUCTION Normalised CPU time (kSI2K) per REGION



© CESGA EGI View: PRODUCTION / normcpu / 2011-4-20123 / REGION-VO / lhc (x) / ACCBAR-LIN / i

R2

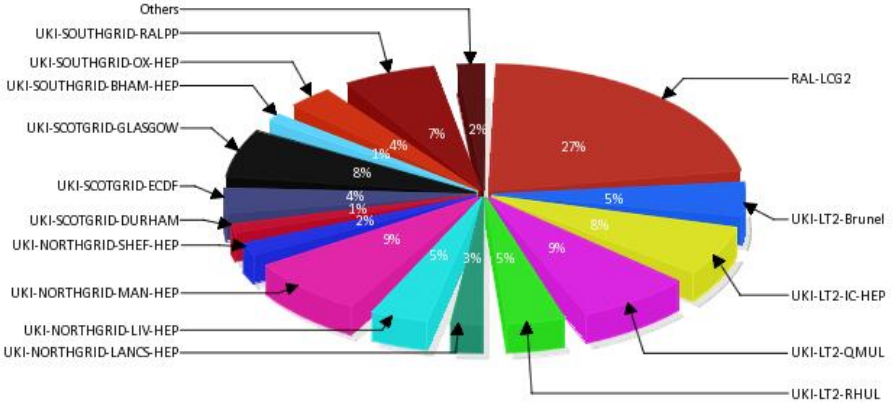
2012-08-29 20:44

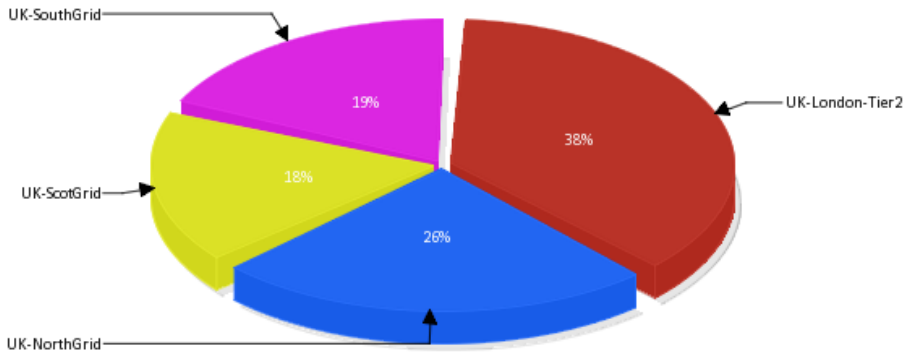
The UK is a large contributor to the EGI.

Tier-1 accounts for ~27%

Tier-2s share as below

ROC Normalised CPU time (kSI2K) per SITE



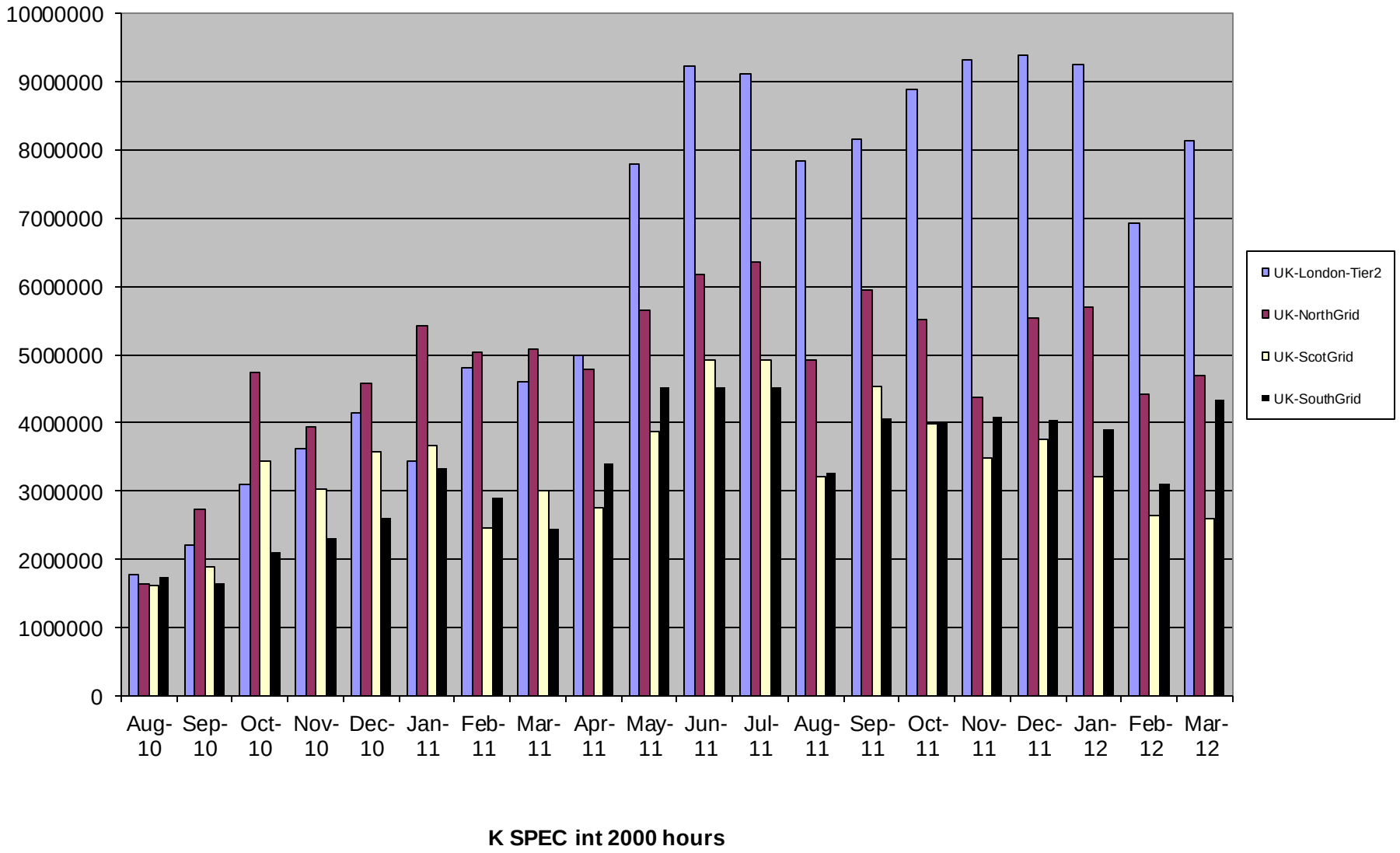


© CESGA EGI View: United Kingdom / normcpu / 20113-20124 / SITE-VO / all (x) / ACCBAR-LIN / i

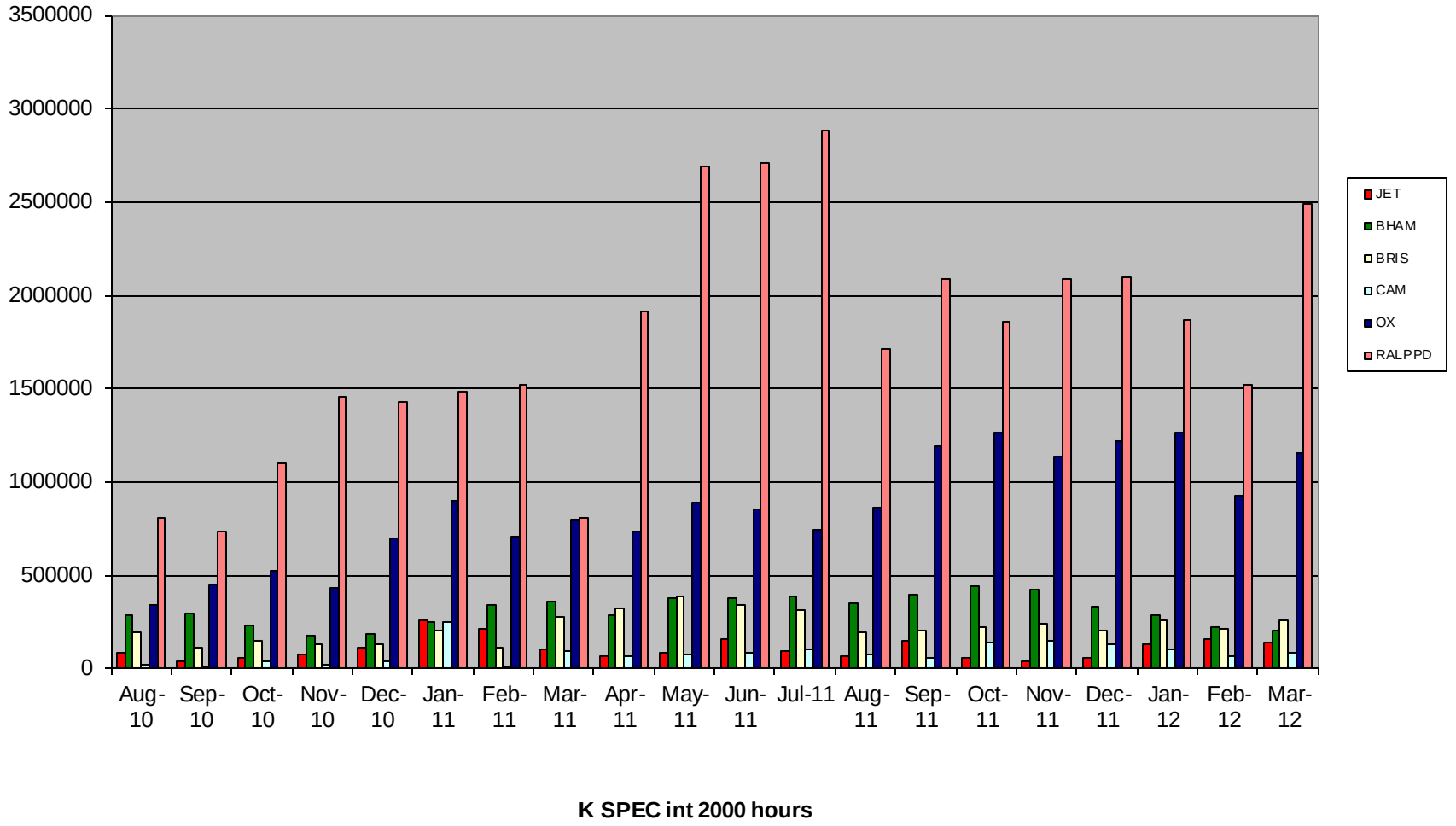
2012-04-06A EGI View: UK (Tier2) / normcpu / 20113-20124 / TIER2-VO / all (x) / ACCBAR-LIN / i

2012-04-09 20:1

UK Tier 2 reported CPU – Historical View to present



SouthGrid Sites Accounting as reported by APEL



Birmingham

Current capacity
3345HS06 195TB

GridPP4 and DRI Funds

Capacity Upgrade

4 Dell C6145 chassis complete with 96 AMD6234 cores per chassis, and 60 TB of RAID storage.

The C6145 servers will therefore contribute 384 cores at around 8 HS06 per core. This will double our number of job streams.

Networking Infrastructure upgrades

Connectivity to JANET has been improved by a dedicated 10GbE port on the West Midlands router, dedicated to the GridPP site.

New Dell S4810 switches with 48 10GbE ports per switch to enhance the internal interconnection of the GridPP clusters in both Physics and in the University centre.

Single GbE links on existing workers in Physics to be enhanced to dual GbE.

3 S4810 switches being deployed in Physics for current & future grid storage nodes, servers, and existing GridPP workers at 2x1GbE and at 10GbE.

2 S4810 switches to be deployed in Uni central gridPP clusters.

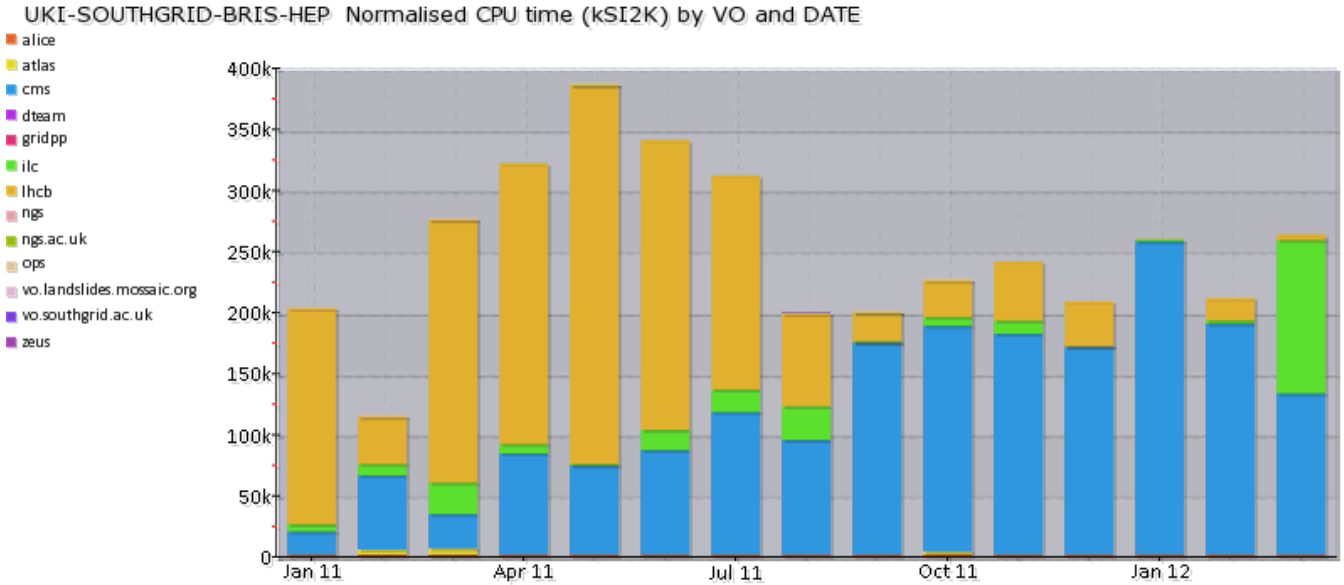
The switches were purchased with sufficient fibre inserts and 1GbE RJ45 inserts as well as many 10GbE direct cables, for current and future needs over the next 4 or more years.

Bristol

Current capacity
2247HS06 110TB

- Status
  - StoRM SE with GPFS, 102TB "almost completely" full of CMS data
  - Currently running StoRM 1.3 on SL4, plan to upgrade to 1.8 soon.
  - Bristol has two clusters, both controlled by Physics. The university HPC clusters are currently being used.
  - New Dell VM hosting node bought to run service VMs on, with help from Oxford.

- Recent changes
  - New Cream ce's front each cluster, one glite 3.2 and one using the new UMD release. (Installed by Kashif)
  - Glexec and Argus have not yet been installed.
  - Improved 10G connectivity from the cluster and across campus, plus improved 1G switching for WNs.



SouthGrid

Cambridge

Current capacity
2700HS06 295TB

- Status (Following recent upgrades)
  - CPU : 268 job slots
  - Most services glite 3.2, except the LCG-ce for the condor cluster.
- DPM v1.8.0 on of the DPM disk servers, SL5
- Batch System – Condor 7.4.4, Torque 2.3.13
- Supported VOs: Mainly Atlas, LHCb and Camont

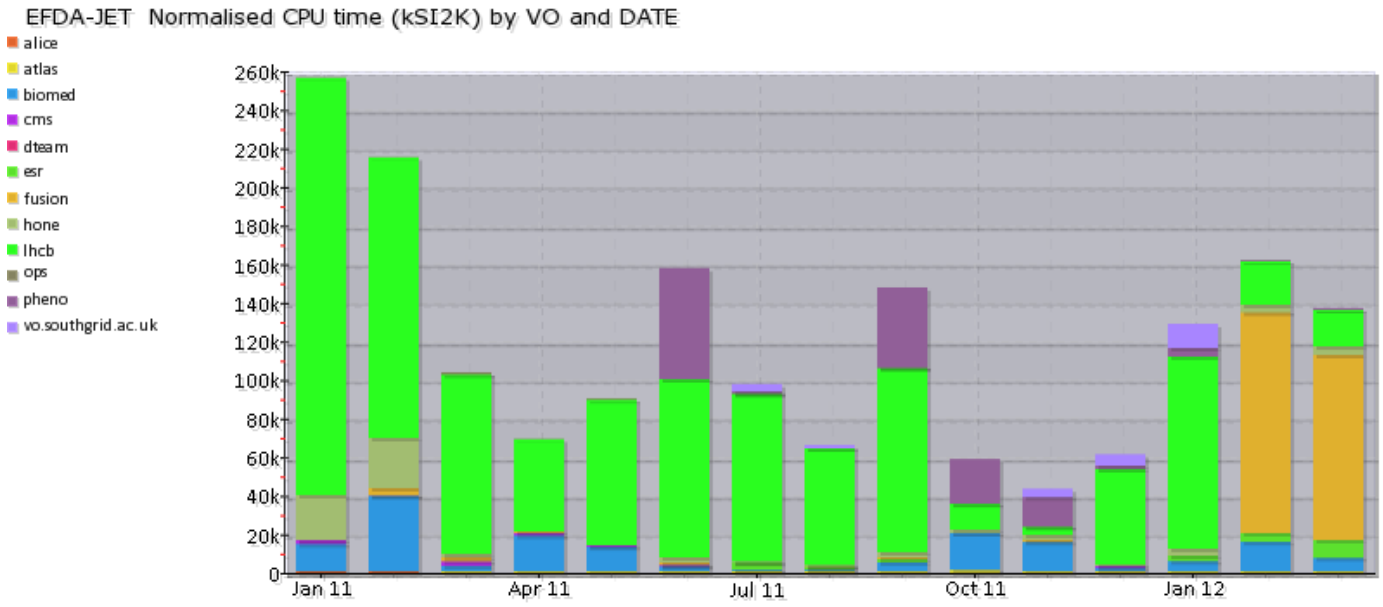
- With GridPP4 funds 40TB of disk space and 2 off dual six-core CPU servers (ie. 24 cores in total) have been purchased.

- The DRI grant allowed us to upgrade our campus connection to 10Gbps.
- We were also able to install dedicated 10Gbps fibre to our GRID Room, 10GBE switches to support our disk servers and head nodes and 10GBE interconnects to new 1GBE switches for the worker nodes.
- We also enhanced our UPS capability to increase the protection for our GRID network and central servers.

JET

Current capacity
1772HS06 10.5TB

- Essentially a pure CPU site
- All service nodes have been upgraded to glite3.2, with CREAM ce's. SE is now 10.5TB
- Aim is to enable the site for Atlas production work, but the Atlas s/w will be easier to manage if we setup CVMFS.
- New server purchased to be the virtual machine server for various nodes, (including the squid required for CVMFS)
- Oxford will help JET do this.



RALPP

Current capacity
26409HS06 980TB

- 2056 CPU cores, 19655 HS06
- 1180TB disk
- We now run purely CreamCEs: 1 * glite 3.2 on a VM (soon to be retired), 2 * UMD (though at time of writing, one doesn't seem to be publishing properly).
- Lately a lot of problems with CE stability, as per discussions on the various mailing lists.
- Batch system is still Torque from glite 3.1, but we will soon bring up an EMI/UMD torque to replace it (currently installed for test).
- SE is dCache 1.9.5 – planning to upgrade to 1.9.12 in the near future.
- GridPP4 purchases were:
- 9 * Viglen/Supermicro Twin^2 boxes (i.e. 36 nodes) each with 2 * Xeon E5645 CPUs, which will be configured to use some of the available hyperthreads (18 job slots per node) to provide a total of approximately 6207 HS06.
- 5 * Viglen/Supermicro storage nodes, each providing 40TB pool storage, for a total of 200TB.

- DRI funds enabled the purchase of 6 * Force10 s4810 switches (plus optics, etc) plus 10Gb network cards which will allow us to bring our older storage nodes to 10Gb networking. 2 of the new switches will form a routing layer above our core network.
- After testing hyper-threading at various levels we are now increasing the number of job slots available on hyper-thread capable worker nodes to use 50% of the available hyper-threads.
- See http://southgrid.blogspot.com/2012/03/following-recent-discussion-on.html

Sussex

- Sussex has a significant local ATLAS group, their system (Tier 3) is designed for the high IO bandwidth patterns that ATLAS analysis can generate.
- Recent spending at Sussex on the HPC infrastructure from GridPP, DRI and internal EPP budgets has been as follows:

4 R510 OSS's to expand lustre by 63TB
18 infiniband cards for the OSS's and for upcoming CPU spend
4 36-port infiniband switches

Integrated all sub-clusters at Sussex into one unified whole. Set up a fat-tree infiniband topology using the extra switches. Better IB routing throughout the cluster and more ports for expansion. Users are reporting that the cluster is running faster.

The total capacity of the cluster (used for the entire university) is now ~500 cores (equivalent to intel x5650) with about 12 hepsepc06 per core. (**6000HS06**) We have **144TB** of lustre filespace in total, again shared by the entire university.

Tier 2 Progress

The Sussex site will become a Tier2 shortly. The bdii, cream-ce, storm, CVMFS and apel services are now working correctly. We should be able to get online by early next week.

Initially, we are going to restrict grid jobs to using only a small part of the cluster - 24 job slots, with 50TB of disk space allocated to the grid.
Any future disk spend will all be allocated to the grid as we now have enough available for internal needs for the foreseeable future.

Also, most of the future EPP CPU spend will be allocated to the grid, and due to the unified cluster, the grid will be able to backfill to any spare capacity on CPU added by other departments

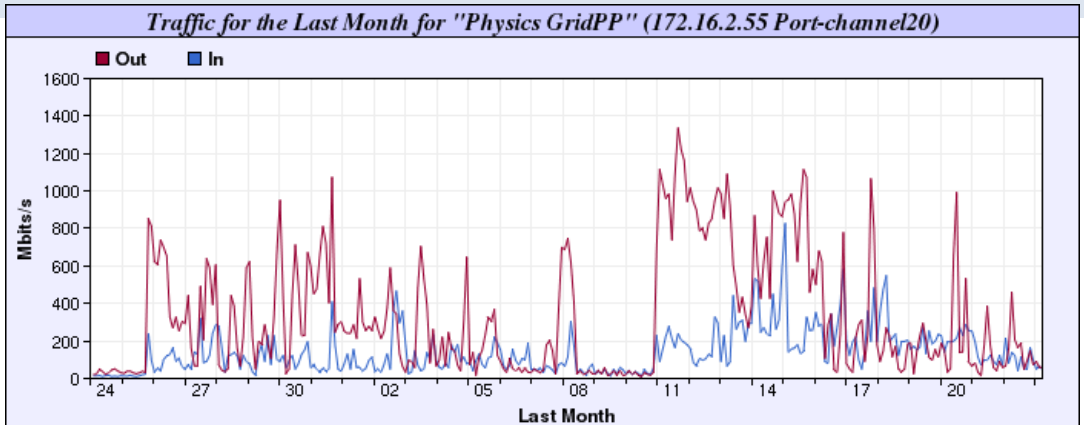
Oxford

Current capacity
8961HS06 620TB

- Significant upgrades over the last year to enhance the performance and capacity of the storage.
- A move to smaller faster Dell 510 servers (12*2TB raw capacity). 14 installed during Autumn 2011and a further 5 this Spring.
- A mixture of Intel (Dell 6100) and AMD CPU worker nodes have been installed.
- Four AMD 6276 Interlargos 16 core CPU's on each of the two motherboards in the new Dell C6145's. 4 servers provides 512 job slots.
- New total capacity will be ~**700TB** and **1360cores** with a total of **11500HS06**.

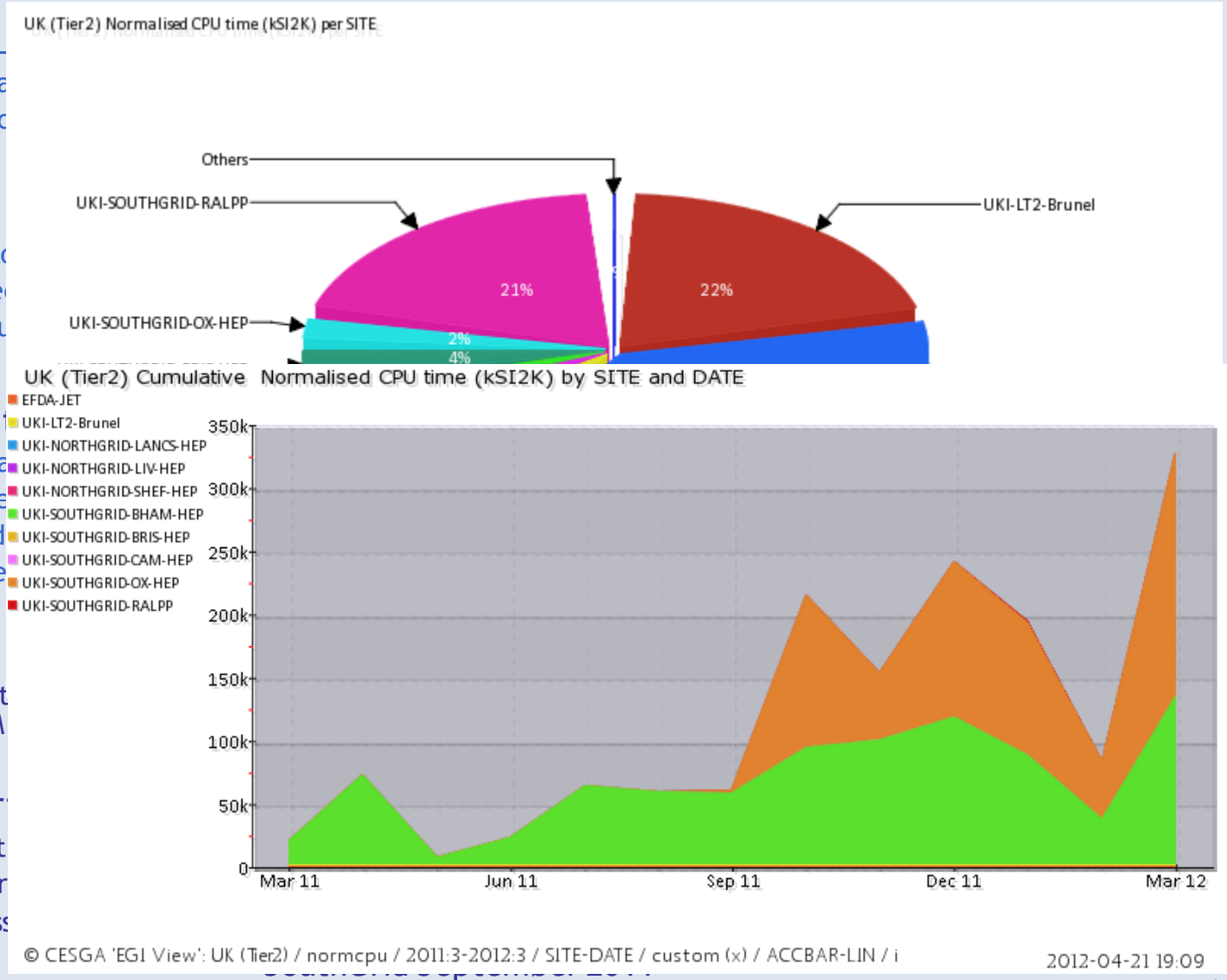
Network Upgrades

- Current University backbone is 10Gbit, so additional links are being installed to provide a 10Gbit path from the Grid and Tier 3 clusters to the JANET router.
- In due course this will allow full 10Gbit throughput without saturation of the links used by the rest of the University.
- Storage servers and high core count WNs will be connected to Force 10 s4810s, with older WNs connected at gigabit.
- The initial 1Gbit link from the computer room was upgraded to 2 Gbit as an interim measure in Autumn 2011. Currently peak usage is 1.6Mbit/s.



Other Oxford Work

- CMS Tier 3
  - Supported by RAL
  - Useful for CMS, a
  - However can bloc
- ALICE Support
  - There is a need t
  - Made sense to ke
  - Site being configu
- UK Regional Moni
  - Kashif runs the na
  - These include the
  - The WMS is an ad
  - There are very re
- Early Adopters
  - Take part in the t version of CREAM
- SouthGrid Suppor
  - Providing support
  - Landslides suppor
  - Helping bring Suss



Conclusions

- SouthGrid sites utilisation generally improving, but some sites small compared with others.
- Recent hardware purchases will provide both capacity and performance (Infrastructure) improvements.
- Enabling of CVMFS at JET should allow Atlas Production jobs to run there to soak up the spare CPU cycles
- Sussex in the process of being certified as a Grid site.