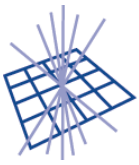


GridPP Project Management Board

GridPP5 – Tier-2

Document identifier :	**GridPP-PMB-162-GridPP5-Tier-2**
Date:	**20/02/2013**
Version:	**1.0**
Document status:	**Final**
Author	**Steve Lloyd/PMB**

Introduction

The UK has established four successful distributed Tier-2s that have been used as a model elsewhere in WLCG, such as France and Germany. The strength of this model is that it combines the best attributes of local facilities (such as on-site user support and local engagement) with the advantages of larger managed facilities (such as quality of service and access to expertise).

Functionality

The LHC computing Grid was originally conceived as a hierarchical structure of Tier Centres based around a single Tier-0 at CERN, a dozen or so large national Tier-1 Centres in different countries, and many 10s of regional Tier-2 facilities. This architecture was driven by the assumption that network bandwidths when the LHC started would prove to be the bottleneck. In practice, network bandwidths grew much faster than assumed and are now typically 10Gbit/sec. Thus, the strict hierarchy, which reflects the various steps of data-reduction, is now unnecessary. Currently, as discussed below, the computing models of the LHC experiments are being modified to make use of networks as the third resource, alongside Computational and Storage resources. Allowing duplication or migration of data across the network enables more efficient use of computational resources as popular datasets can be accessed locally t more sites.

ATLAS

The new ATLAS computing model begins to dissolve the boundaries between the Tier-1s and Tier-2s, and between the High Level Trigger, the Tier-0 and the Tier-1s. ATLAS will be using the High Level Trigger (HLT) farm, as well as the Tier-0 in reprocessing campaigns and Monte Carlo simulation, when other demands permit. The increased use of virtualization makes this use of the HLT Farm possible. Similarly, contingency plans are in place to offload first pass processing to selected Tier-1s if required. CERN still retains the role of preserving the primary copy of the raw data, with the Tier-1s storing the second copy, requiring good uptime for the long-term storage system.

Given the bandwidth available and the ability to dynamically place data, Tier-2s are being used in the reprocessing previously restricted to Tier-1s. This has been successfully demonstrated, to the extent that large fractions of the US reprocessing have been transferred to their Tier-2 sites. The Tier-1s retain their role in the long term archival of data, and serve any raw data to be reprocessed to the Tier-2s. However, another change that eases this new role for the Tier-2s is the increased role of ‘patching’ data, where reprocessing can be driven in many cases from the formats already resident on that Tier-2.

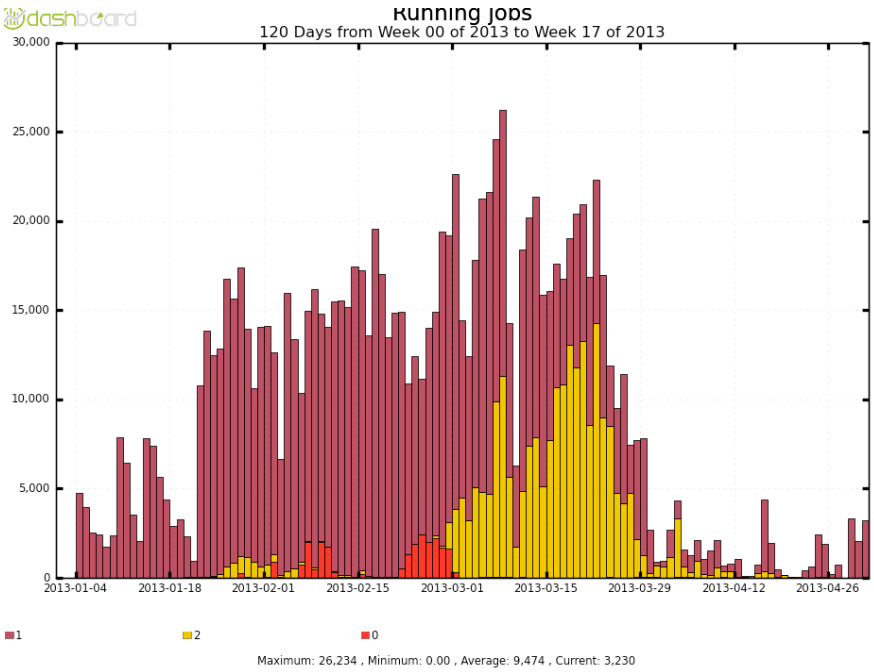
Group selection and analysis becomes a more integrated part of the processing and reprocessing step, with selections and the production of a reduced number of analysis formats without overlaps. It is to be shared between Tier-1s and Tier-2s.

In the ATLAS computing model Tier-2s remain the key to the vast bulk of the end user analysis and the large majority of the simulation. Selected Tier-2 sites also host the group data sets, and have accordingly a larger fraction of the associated analysis load. They also provide capacity for national communities and for national priorities. An important change, which has been underway for some time, is to open the major Tier-2s (T2Ds) to receive and send data directly to Tier-1s and Tier-2s elsewhere in the world, breaking the close dependence on a ‘local’ Tier-1, leading to higher resilience. Similarly, the catalogues have been largely moved from the Tier-1s to CERN.

CMS

In a manner similar to ATLAS, the evolution of the CMS computing models has blurred the distinction between the services offered by different Tiers for production activities (such prompt reconstruction and re-reconstruction). Two of the key aspects that have enabled this blurring is the division of the archival storage and the processing aspects of the Tier-1 centres, and the increased performance and reliability of the Wide Area Network (WAN).. The processing parts of the Tier-1s now essentially provide the same services as are available at the Tier-2 centres but with a greater reliability. The use of virtualisation and cloud technologies means enabled the use of the, large HLT resources as an additional processing centre between periods of

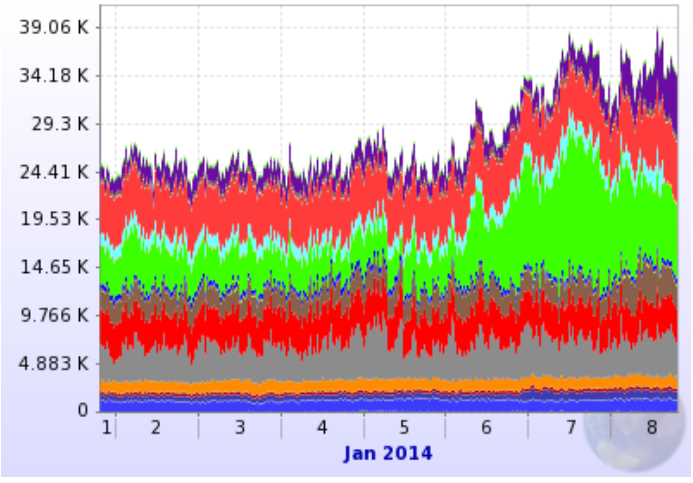
data taking. This means that it is possible to share production processing workflows across the Tier-0, the Tier-1s, the Tier-2s and the HLT. For example, it is planned to use the processing functionality of the Tier-1 centres (and perhaps the HLT and larger Tier-2 centres) to perform up to 50% of the prompt reconstruction of new data for periods when the Tier-0 resources are insufficient to keep up. Another example that has already demonstrated this blurring between Tiers was the 2013 reprocessing of the data taken in 2011. This tasks is traditionally confined solely to the Tier-1 centres, however as can be seen in the figure below, it was successfully shared between the Tier-1s, Tier-2s and the Tier-0.



2013 reprocessing of the 2011 datasets at the three different Tiers (Tier-0 = red, Tier-1 = burgundy, Tier-2 = yellow)

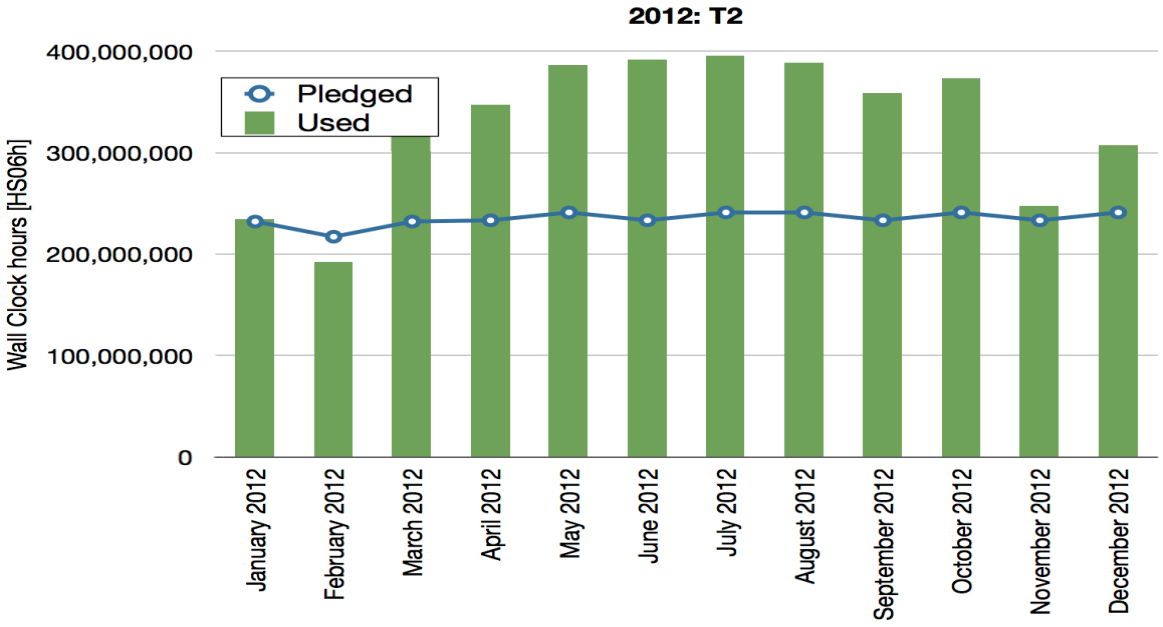
While there has been a blurring of the boundaries between the processing aspects of the different Tiers, only the Tier-0 and Tier-1s are capable of providing the custodial storage of the primary data.

Currently, the primary role of the Tier-2s within CMS is end user analysis and this is expected to remain the case into the future. However, the increased capacity of the WAN has enabled analyses to run at one Tier-2 analysing data hosted by a different Tier-2 (or Tier-1). The connections and data access between sites is performed using the AAA services (based on xrootd). This is already a production service and the figure below shows the number of AAA connections (i.e. jobs running at one site accessing data at another) running for a week in January 2014.



Number of CMS AAA connections in the first week of 2014.

The efficient use of the Tier-2 resources is particularly important to CMS because at times of intense physics activity, the use of these resources has exceeded the pledged capacities. This has only been possible because of the additional resources leveraged at individual Tier-2 institutes; however, the lack of these discoveries would have delayed many of the important physics results of 2012 and 2013. The figure below shows the Tier-2 usage by CMS in the Higgs discovery year of 2012.



• Tier-2 CPU by month for the Higgs Discovery year 2012.

LHCb

The LHCb computing model has evolved during the Run-I period of the LHC. At the outset, for a planned trigger rate of 2 kHz, it depended entirely upon Tier-1 centres for data storage, reconstruction and analysis, with Tier-2 sites providing the large amount of CPU needed for Monte Carlo simulation. Whilst this performed adequately for the initial processing as data was recorded, it had a number of shortcomings due to the peaks of resources needed for re-processing. The situation also became difficult because LHCb expanded its physics scope to be both a "b" meson and "c" meson experiment, roughly doubling the data volumes with the trigger rate rising from 2kHz to 5 kHz.

To mitigate this the limitation of running reprocessing on Tier-1s was relaxed in 2011. Reprocessing jobs were executed on a selection of Tier-2s by downloading the input RAW file from Tier-1 storage, running the job, and uploading the reconstruction output to the same storage. This was generalised in 2012 when 45% of reconstruction CPU time was provided from outside CERN and Tier-1s. In fact in 2012, only 30% of the RAW data was processed by the first pass prompt reconstruction and was used mainly for monitoring and calibration. The reprocessing was then run continuously on the full dataset once calibration constants were available, within 2-4 weeks of data taking. This new scheme evens out the load and removes the need for a big end of year reprocessing.

This concept will be generalised for Run 2 where the trigger rate will rise to 12.5 kHz (more than doubling the data rates). Monitoring and calibration will be performed automatically on the HLT farm and picked up by the offline reconstruction, which will run continuously shortly after the data taking. No reprocessing is planned before the end of Run 2. Should resources be insufficient for the continuous reconstruction of the full dataset, it is foreseen to ‘park’ a fraction (up to 50%) of the data to be processed during Long Shutdown 2.

In 2012 LHCb also faced severe disk storage problems and the computing model has accordingly been modified in two ways: The most important is the introduction of so called Tier-2 sites with Data (T2D sites). These T2D sites must qualify in terms of quality and volume of disk, after which they act like extra Tier-1

resources. This allows many more sites (and countries without Tier-1s) to pledge disk resources. LHCb has also targeted much more use of microDST format instead of DST format, which is an order of magnitude more compact. LHCb does not move data around in large quantities and so is relatively insensitive to the network (i.e. that which is good enough for the GPDs is more than good enough for LHCb.

In summary, in Run2 LHCb will ideally rely very heavily on the increased resources available at the UK Tier-1 Centre, and will augment this with more use of T2D sites in globally.

ALICE

With respect to the original computing TDR, the new ALICE Computing Model dispenses with the tiered centre structure and becomes more cloud-like, in that the main features of the computing centres are the service levels agreements, commitment to provide long term custodial storage for some of the centres and a stronger reliance on inter-centre network connectivity for data management. During LS1 ALICE is upgrading the readout electronics of several detectors (TPC, TRD) and expanding the HLT and DAQ systems to allow for a factor of 2 increase of the readout rate. The TRD azimuthal coverage will be extended from 60 to 100% and a second electromagnetic calorimeter (DCAL) diagonal to the existing ECAL will be installed. The additional detectors will increase the raw data output size by about 10%. The computing resources estimates for Run 2 are based on the same CPU power needed for reconstruction and 25% larger raw event size due to the additional detectors and higher track multiplicity with increased beam energy and event pileup.

Current Service Provision

While preparing for GridPP4, GridPP came to the conclusion that to perform the most demanding of the Tier-2 tasks, hosting detector performance and physics analysis data and running group and individual user analysis jobs, a site requires more than a single system-administrator to keep the site running week-after-week. The simpler Monte Carlo production task requires CPUs, some cache disk and a reasonable outbound network connection to the Tier-1, which is manageable with less effort. However, group and user analysis requires large amounts of disk with excellent tuned bandwidth to the local CPUs and good inward and outbound network connections to other sites plus many additional services; in addition, it must be highly responsive to the needs of the supported experiments and their computing teams. If a CPU fails then it can be changed some time later, but if a disk fails it has to be replaced almost immediately before a second disk in the same RAID fails leading to unrecoverable data loss, and data recovery procedures must be enacted immediately.

For GridPP4, difficult decisions were made in optimising the Tier-2 manpower within the limited funding available whilst trying to maintain and foster the excellence needed to meet the experiment requirements. The solution was to designate certain sites, based on previous performance, as Group Analysis Sites with the others engaged through more generic effort with the opportunity to develop their analysis roles in the future. For ATLAS the Group Analysis Sites were Glasgow, Lancaster, Liverpool, QMUL, and Oxford while for CMS they were Brunel, Imperial and RAL PPD. These sites were allocated 2 FTE each, the medium sized sites 0.5 to 1 FTE and most of the smaller sites had to rely on their own unfunded manpower. In April 2013, following a review, 1 FTE of manpower was transferred to Manchester, which became an ATLAS Group Analysis site, and there were other minor adjustments.

The table summarises the current provision at the UK Tier-2s. The colour coding is that large sites are in green, medium size sites in yellow and small sites in blue. FTE-1 and FTE-2 refer to the amount of effort funded by GridPP4 prior to April 2013 and after an adjustment was made to reflect changing resources. The changes are highlighted in red.

Site	CPU [HS06]	Disk [TB]	GridPP FTE-1	GridPP FTE-2
Brunel	11,600	484	1.4	1.4
Imperial	26,900	2,026	3.0	3.0
QMUL	29,800	1,680	2.0	2.0
RHUL	14,500	640	0.5	1.0

UCL	2,600	189	-	-
Lancaster	25,298	1,032	2.0	2.0
Liverpool	11,129	550	2.0	1.0
Manchester	22,809	883	1.0	2.0
Sheffield	5,476	360	0.5	0.5
Durham	8,140	50	0.25	0.25
Edinburgh	17,200	355	0.5	0.5
Glasgow	31,600	1,300	3.0	3.0
Birmingham	8,121	320	0.5	0.5
Bristol	2,247	105	-	-
Cambridge	1,600	275	0.5	0.5
EFDA Jet	1,772	11	-	-
Oxford	11,770	650	2.0	1.5
RAL PPD	25,975	1,260	1.35	1.35
Sussex	2,365	50	-	-
Total	**261,000**	**12,200**	**20.5**	**20.5**

A brief description of each Tier-2 is given below. The LHC experiments listed in brackets are those that the institute actively participates in. In practice almost all sites support most experiments in some way.

LondonGrid

Brunel (CMS)
The Brunel GridPP Tier-2 is a medium sized site that is located in a purpose built data centre shared with the University Computer Centre who also provide significant network support. There is plenty of space, power and cooling capacity for future expansion. Brunel is a designated CMS analysis site with 1.4 FTE of GridPP effort to support CMS Group Analysis and devotes the majority of its capacity to CMS, however it also supports a large number of other VOs. It has four dedicated 1Gb/s links directly to JANET. Brunel has taken a leading role in middleware deployment and was the first site in WLCG to debug and completely migrate to xroot for both local and fallback access for CMS and ATLAS. Brunel has led effort in WLCG in debugging both CMS and ATLAS on SL6, operation with glexec, Argos authentication and IPV6 storage passing all Nagios tests. The site has published articles on computational statistics, multi-core computing, and computer security in a number of ACAT and CHEP conferences and hosted the 2010 CERN School of Computing and the 2011 ACAT Conference.

Imperial (CMS, LHCb)

The Imperial College site is a leading and significantly sized designated CMS analysis Tier-2 with 2 FTE of GridPP effort to support CMS Group Analysis plus 1 FTE of Grid operations expertise for the whole of LondonGrid. Imperial also hosts a 0.5 FTE LHCb GANGA post and a 0.225 FTE User Technical Support post. Imperial also hosts the GridPP Technical Director funded at 0.25 FTE. Imperial delivers almost a half of the UK's CMS computing contribution and coordinates CMS' computing activities in the UK. There is close involvement with the central CMS computing team and the site led the project to use the CMS HLT as

a cloud computing resource. It also contributes to LHCb and ATLAS MC production and supports a variety of other VOs including T2K, D0, ILC, Biomed etc. The site maintains a number of supplementary Grid services including a load-balanced Top-BDII, twin WMSs, a backup VOMS server, GridPP's Dirac work-load management system, the Real Time Monitor and a perfSONAR dashboard. Staff coordinate the staged-rollout of EMI software within the UK and provide dedicated support for small VOs such as NA62 and MICE. The site hosted the Big Data and Cloud technical workshops plus a WLCG Collaboration workshop. There is notable expertise in security and a member of staff is in the EGI software vulnerability group risk-assessment team. The GridPP Cloud is hosted and maintained at Imperial College. The site contributes to the HEPiX IPv6 working group by running an IPv6 testbed comprising a small compute cluster, test storage elements, an FTS3 server and SAM Nagios monitoring. The site was, to our knowledge, the first major WLCG site to support the IPv6 protocol across the majority of its production services. With a 20 Gb/s link to Janet (soon to become to a shared 40 Gb/s link) the site is well-connected and is actively engaged in network testing in preparation for Run 2 of the LHC.

QMUL (ATLAS)

QMUL is a designated ATLAS group analysis site with 2 FTE of GridPP effort to support ATLAS Group Analysis. QMUL also hosts the GridPP Operations Coordinator (and CB Chair) funded at 0.25 FTE. It is one of the largest Tier-2 sites in the UK. The majority of the resources support the ATLAS experiment, providing 21% of ATLAS UK computing use in 2013. It was first site to use the high performance Lustre parallel file system more commonly used in supercomputers. Lustre high performance enables QMUL to excel in the analysis of data. As a result, from 2011-2013, QMUL has processed over 68 PB of data in 80 million data files, the 3rd largest for Tier-2 sites worldwide. For ATLAS computing, it has been the only UK site constantly in the top 10 over the last 3 years. QMUL also provides extensive support for non LHC VOs both in terms of man power and computing resources, for example it is the largest T2K site in the world doing 35% of all Grid work in 2013 (about 17% of all T2K processing) and the 2nd largest supporter of non LHC VOs in the UK in 2013. QMUL uses its expertise in data analysis to support a number of initiatives outside of particle physics requiring "Big data", leading to high impact publications that would not have been possible without the QMUL's Grid resources.

RHUL (ATLAS)

The Royal Holloway GridPP Tier-2 is a medium-sized site dedicated to providing a reliable service for the ATLAS experiment. With 1 FTE of GridPP-funded effort to support unscheduled analysis and simulation for all VOs, together with a contribution from LondonGrid central effort, it represents good value for a site of this size. The site has maintained its position within the top five in the ATLAS UK region. The cluster, with a high disk to CPU ratio and good internal network bandwidth, has been specifically designed for ATLAS analysis. It is housed in a purpose-built data centre run by the University computing service with a dedicated 4 x 1 Gb/s link to Janet. Possibilities for an upgrade of the wide area network link to 10 Gb/s are being explored. Previous and anticipated GridPP grants have allowed the site to leverage three significant investments from the University.

UCL (ATLAS)

UCL currently provides access to a batch farm and a DPM Storage Element dedicated to ATLAS. The HEP group intends to provide Grid access to the large shared university cluster. This is becoming more of a prospect as the central Research Computing department at UCL is moving towards a more flexible approach to service provision. UCL has a dual 10 Gb/s link to JANET, and there is a 10 Gb/s link to the HEP grid cluster. UCL receives no manpower support from GridPP.

NorthGrid

Lancaster (ATLAS)

Lancaster is a designated ATLAS group analysis site with 2 FTE of GridPP effort to support ATLAS Group Analysis. Lancaster is the leading example of a Tier-2 site delivering resources to GridPP through the use of a shared university cluster facility (the High End Computing cluster, managed jointly by Lancaster ISS and Physics staff). This has lead to high leverage, as they obtain 50% of the University capital contributions for CPU, and also significant additions to the storage they can offer, which are solely for GridPP. The close working with ISS staff provides additional cross-coverage and resilience. Further, they have the direct use of two new, state-of-the-art machine rooms designed with space for considerable expansion. They benefit from water-cooled racking and other central infrastructure. The well-integrated operation also allowed them to make excellent use of the DRI network funding opportunity to take the site to dual-10Gbps connection (one northbound, one southbound) and to future-proof the connections between the cluster and the WAN.

The site has significant experience in data storage and produced with Edinburgh colleagues a white paper on the configuration of Dell storage solutions for grid usage. Lancaster staff are also the the EMI WN and UI Tarball Product team, which has been used by both ATLAS and CMS within CVMFS as well as at sites in the UK and abroad. Staff at the site are well integrated into the ATLAS experiment activities and have leading roles in ATLAS distributed computing operations (Global Grid data Processing Co-ordinator, pilot factories, ADC expert on call, opportunistic Cloud exploitation, alternate Cloud middleware); the GridPP PI is also head of the ATLAS Upgrade Computing and the ATLAS UK Software Computing Coordinator. The site is very closely involved with the ATLAS computing and software upgrades and is has co-organized future-computing workshops.

Liverpool (ATLAS, LHCb)

At the start of GridPP4, Liverpool was a designated ATLAS group analysis site with 2 FTE of system management effort but now has 1 FTE of GridPP effort. The Liverpool Tier-2 site is situated in a large, dedicated cluster room within the Oliver Lodge Laboratory, with ample capacity for future expansion. Liverpool has an active support and research team, specialising in site optimisation and providing a broad range of support across disciplines that include High Energy Physics and Life Sciences. Physics staff at Liverpool develop software for the ATLAS and LHCb experiments at the LHC, and various other experiments including T2k and SNO+. Liverpool also supports 27 smaller VOs. To support their large number of small VOs, Liverpool has developed tools for VO life cycle management (VomsSnooper) and a simple set-up process (instantUI) to allow new VOs to access the grid more easily. The site has maintained some of the highest levels of reliability, efficiency and cluster utilisation across the LCG, adopting innovative technologies such as 'black-hole' detection software, automatic build and disaster recovery systems and Grid node virtualisation. The team also conducts staged-rollout of middleware developments, research into optimal configuration and adoption of emerging technology. The results and outputs of this research and development have been presented at WLCG workshops and CHEP. Liverpool has also been active in trialling XROOTD data federation and other aspects of ATLAS data access optimisation.

Manchester (ATLAS, LHCb)

Manchester became a designated ATLAS group analysis site with 2 FTE of GridPP effort to support ATLAS Group Analysis after the transfer of effort from Liverpool. The Manchester GridPP Tier-2 is a large site located together with the University central facilities. The Tier-2 cluster shares one wing of the computer centre with the University shared cluster. Manchester has 2 FTE of GridPP system management effort. It devotes the majority of its capacity to ATLAS and is also an LHCb primary site. It was one of the first ATLAS T2Ds (T2 hosting data for other T2s) and has recently become one of only 2 LHCb T2Ds in the UK. It makes a major contribution to the UK's commitment of CPU to both experiments and also devotes a significant fraction of its resource to the benefit of smaller experiments inside and outside particle physics, theorists and accelerator physicists. It has a dedicated 10Gb/s link directly from the cluster to JANET. Manchester provides the primary VOMS servers (replicated at Imperial and Oxford). The site is well integrated with both ATLAS (ADC monitoring coordinator and central operations) and LHCb (computing model review and Tier-2 coordination) computing, is also heavily involved in WLCG operation coordination (WLCG Ops Coord co-chair, SL6 TF leader, multicore deployment TF co-leader) and EGI ops (ROD shifts). The group has made a leading contribution to the investigation and development of cloud solutions for use by the LHC experiments.

Sheffield (ATLAS)

The Sheffield GridPP Tier-2 is a medium size site that mainly supports the ATLAS, T2K and LHCb VOs. It recently acquired space in a University managed computing room, which means there is plenty of room, power and cooling capacity for a further expansion in the next few years. The Sheffield Tier-2 has 0.5 FTE of GridPP effort to support unscheduled analysis and simulation for all VOs and has an excellent performance record with an availability of 95% or better over the past 5 years. The staff in Sheffield are mostly associated to ATLAS providing cloud support, distributed computing support, software support and T2K data distribution support.

ScotGrid

Durham

The Durham Tier-2 site is one of the smaller sites in the GridPP collaboration. It is currently housed in a small machine room in Physics however they are currently preparing to move the cluster to the Universities HPC machine room. This will provide better cooling and more stability for its users. Durham University also expects to upgrade the JANET connection to 10Gb around summer 2015. As the site is located within the

IPPP it is a key link with particle physics phenomenology research in the UK, represented by the Pheno VO. The IPPP also plan to spend a further £150k on increasing the size of the cluster. The site also has links with the DiRAC community based in Durham. Durham was one of the first sites in the collaboration to use Puppet for configuration and management, ahead of its growing adoption in the community. The site intends to leverage this advantage in developing UK expertise in puppet configuration at sites. Durham receives 0.25 FTE of GridPP effort to support unscheduled analysis and simulation for all VOs.

Edinburgh (ATLAS, LHCb)

Edinburgh is a leading example of a Tier-2 site delivering resources to GridPP through the use of a shared university cluster facility (ECDF) and has 0.5 FTE of GridPP effort to support unscheduled analysis and simulation for all VOs. The site has strong expertise in data storage and hosts a 1 FTE Data Storage post responsible for the development of DPM software and who is currently chair of WLCG data working group. As well as providing significant CPU and storage for LHC experiments, the site has specialised in the exploitation of many-core devices such as GPUs and MICs for use on the Grid. In addition, the site has recently explored the possible use of High Performance Computing systems (such as the HECToR supercomputer hosted by the Advanced Computing Facility in Edinburgh) for Grid-based workloads. Staff at the site are closely aligned to the ATLAS experiment activities and have taken lead roles in ATLAS distributed computing operations. The site has an excellent publication record as part of EGI and CHEP conferences as well as representation on the CHEP programme committee. Edinburgh has also organised and hosted a number of data management and future-computing workshops.

Glasgow (ATLAS, LHCb)

Glasgow is a designated ATLAS group analysis site with 2 FTE of GridPP effort to support ATLAS Group Analysis plus 1 FTE of Grid operations expertise for the whole of NorthGrid. Glasgow also hosts 1 FTE Data Storage and 0.5 FTE Data Processing posts and the GridPP Project Leader funded at 0.75 FTE. The Glasgow Tier-2 site is, at least by CPU-core count, the largest UK installation. It is currently housed in two dedicated machine rooms that were refurbished by the University several years ago. Currently, a further £700k is being spent on upgrading the electrical and cooling infrastructure. In the longer term, Glasgow is developing plans to custom build a University Data Centre as part of the campus development plan. The Glasgow Tier-2 would migrate to this new facility as one of the key academic clients. At present, the site has 160 Gb internal interconnects between the two rooms and two redundant 10 Gb links to JANET. The team at Glasgow has particular expertise in networking (currently provide an IPV6 testbed); monitoring (currently leading the development of Graphite-based monitoring); and data management (particularly data transfers and contributions to the DPM collaboration). In addition, the team is investigating containerization as a more efficient alternative to virtualisation on cluster resources.

SouthGrid

Birmingham (ALICE, ATLAS, LHCb)

Birmingham particularly supports ALICE (60%) and ATLAS (30%). Most (75%) of the cluster is located in a small dedicated server room in our group with the remaining ~25% of the resources provided by central IT in their main server room. They have a dedicated 10Gb/s network link to JANET at present. The site also runs a DIRAC test server primarily to aid Ganga development for small VO support but is also available as a development/test server for others in the UK. Birmingham hosts a 0.5 FTE ATLAS GANGA post for this purpose and has 0.5 FTE of GridPP effort to support unscheduled analysis and simulation for all VOs.

Bristol (CMS, LHCb)

Bristol is primarily a CMS site but also supports LHCb and non-LHC organisations such as ILC and landslides. The site has a mixture of local departmental resources and a share of the HPC facility. The storage capacity is 117 TB on the GPFS file server but expects to increase that to 200 TB shortly and to add 277 TB of HDFS storage served by DM-Lite. The CPU capacity is expected to rise to 10408 HS06 following new purchases funded by GridPP and local funds. The site has a 10 Gb/s link out to JANET and could increase it to 20 Gb/s if the need arises. Bristol receives no manpower support from GridPP.

Cambridge (ATLAS, LHCb)

The Cambridge site supports mainly ATLAS and LHCb. The cluster is housed in a room with a capacity of 12 racks, refurbished with money from a SRIF bid. The site benefits from the excellent networking infrastructure in the University, with a shared 10Gbps connection onto JANET. Cambridge has 0.5 FTE of GridPP effort to support unscheduled analysis and simulation for all VOs.

EFDA Jet

The EFDA-Jet site is not officially part of GridPP but has been part of the WLCG Grid and contributed a small amount of CPU to LHCb, plus other VOs such as Biomed, ESR and Fusion. They receive no manpower support from GridPP.

Oxford (ATLAS, LHCb)

The Oxford GridPP Tier-2 is a medium sized site that is based in a purpose built computer centre located five miles out of town shared with the University Super Computer facilities. There is plenty of space, power and cooling capacity for future expansion. Oxford is a designated ATLAS analysis site with 1.5 FTE of GridPP effort to support ATLAS Group Analysis and devotes the majority of its capacity to ATLAS, however it also supports LHCb and ALICE at around a 16% share each. It has a dedicated 10Gb/s link directly from the cluster to JANET. Oxford runs the UK WLCG Nagios Availability monitoring service including managing the remotely located server at Lancaster. They provide one of two backup VOMS servers (the other is at Imperial) that provide backup to the primary service at Manchester. The site is part of the security team, storage and IPv6 groups and was one of the first to enable xrootd access, and provides an IPv6 enabled Grid User Interface and other services. Oxford also hosts the GridPP Project Manager funded at 1 FTE.

RAL PPD (ATLAS, CMS, LHCb)

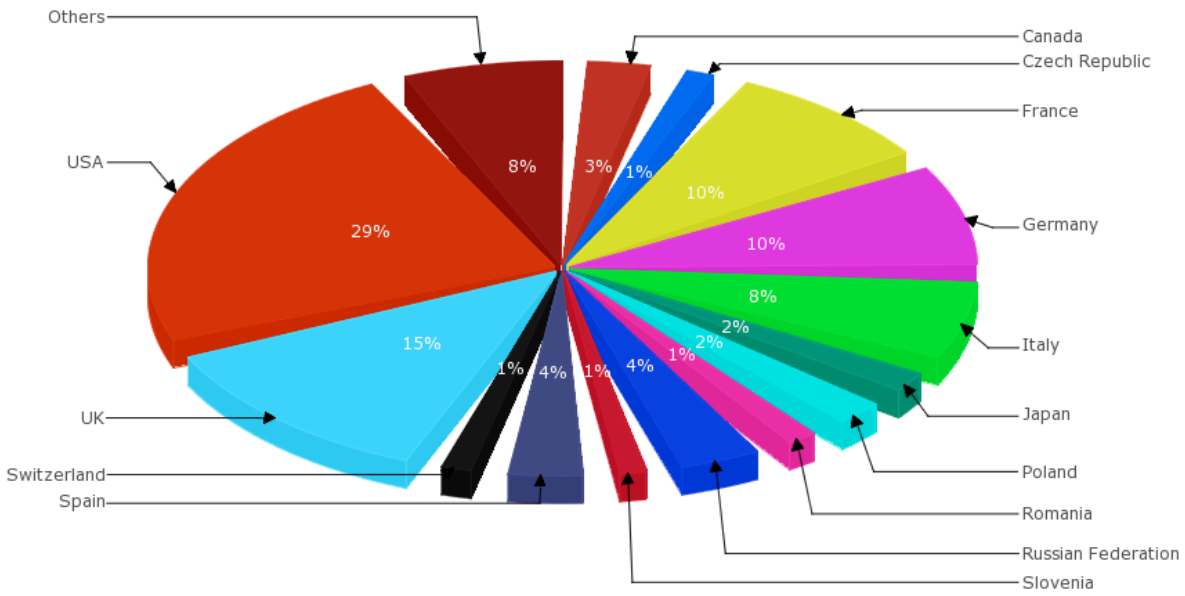
RAL PPD is a designated CMS group analysis site with 1.35 FTE of GridPP effort to support CMS Group Analysis. RALPP is the largest site in SouthGrid, and is located in a new computer room on the RAL site. It benefits from very good network connectivity to the Tier-1 and JANET. It is primarily a CMS (40%) site but also supports ATLAS (30%) and LHCb (20%) plus many of the smaller VOs. They play a key role in interfacing between CMS and the sites. RAL PPD also have 1.5 FTE GridPP effort for operational security and international security cooperation. RAL PPD also hosts the GridPP User Board Chair funded at 0.25 FTE.

Sussex (ATLAS)

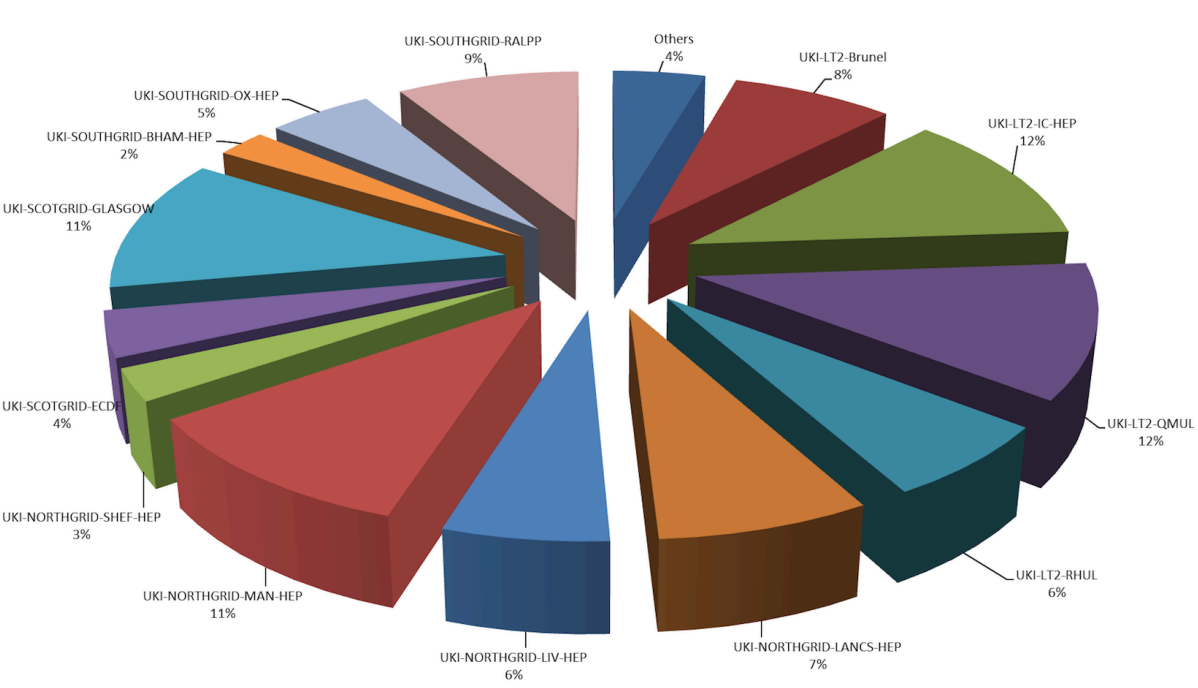
Sussex is the newest addition to GridPP and joined in 2012. This site primarily supports ATLAS. The cluster is located in a state-of-the-art data centre within IT Services; all equipment is powered via UPS with a 1.1MW standby generator to provide resilient power. The Grid cluster is part of a larger shared resource within the University that comprises over 2,300 cores with 4GB RAM per core, all with QDR infiniband interconnect, and a 500TB Lustre parallel file system. The University JANET link will shortly be upgraded to 10Gb/s and the cluster is expected to get an 80% share of this. They receive no manpower support from GridPP.

Performance

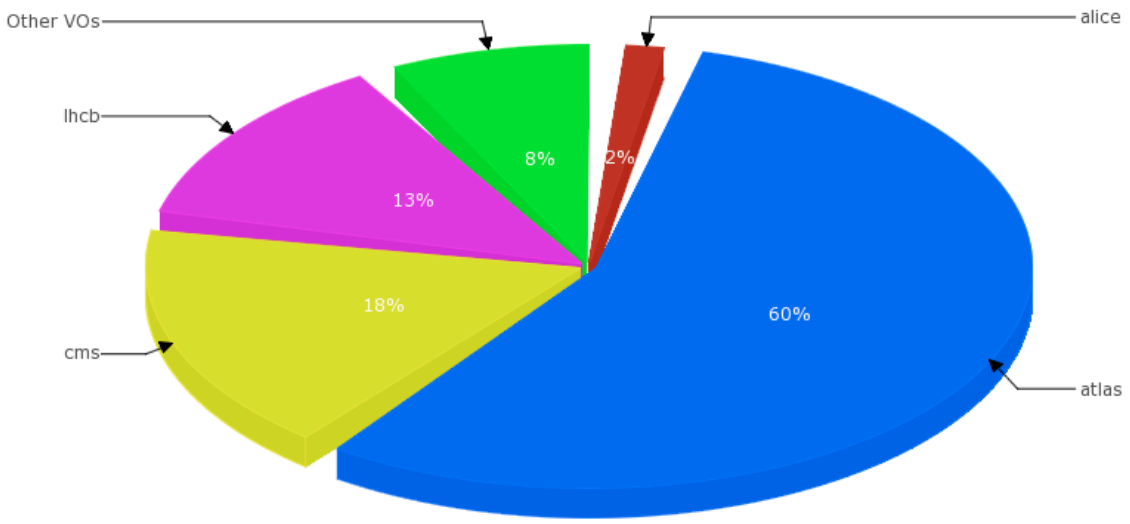
The plots below show the UK Tier-2 contributions to the WLCG and the breakdown of these to the individual sites.



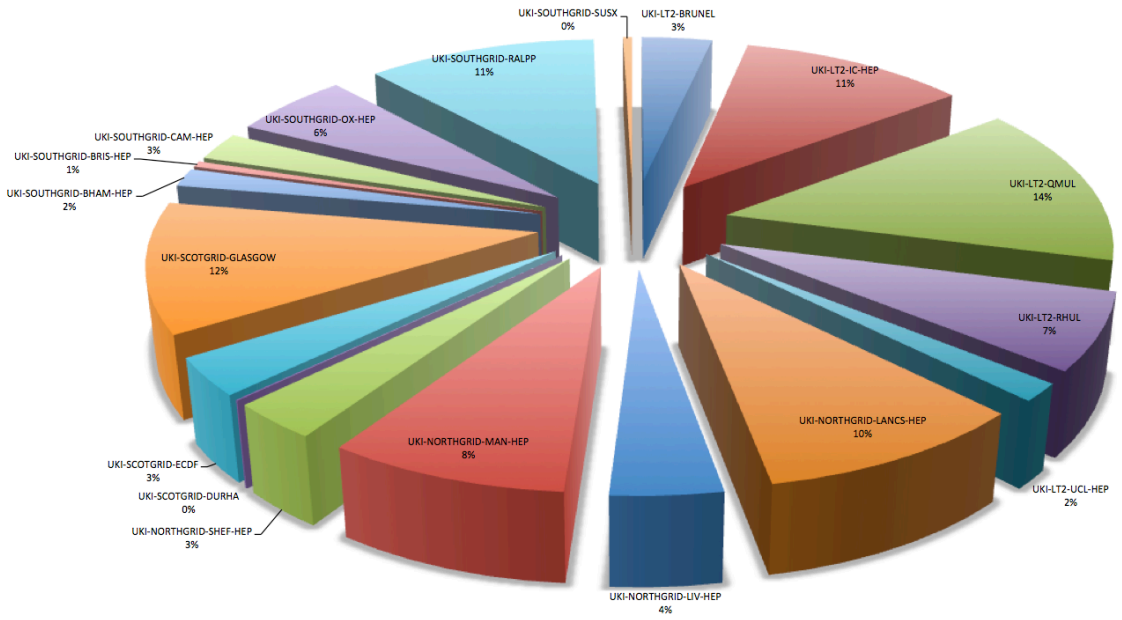
CPU Delivered to LHC VOs by all Tier-2s (Jan 2011 to Dec 2013)



CPU Delivered to LHC VOs by all UK Tier-2 sites (Jan 2011 to Dec 2013)



Delivery of UK Tier-2 CPU to the LHC experiments and other VOs (Jan 2011 to Dec 2013)



Delivery of UK Tier-2 Disk to ATLAS and CMS

Leverage

One of the great advantages of having the UK Tier-2s distributed across almost all the institutes participating in experimental particle physics, is that it enables GridPP to leverage resources, operating costs, knowledge and manpower from a variety of difference sources, such as university computer clusters, research infrastructure funding, local system and network expertise and so on. Likewise it gives local physicists ready access to Grid expertise and provides an easy conduit to disseminate information and engage with diverse communities outside particle physics.

Although this leverage is difficult to quantify exactly, especially electricity costs, a survey of all sites was undertaken in 2009 and repeated in 2013 and summarised below. Note that 2005-2009 is five years while 2010 to 2013 is four years.

	2005-2009	2010-2013
Capital Costs	£4.4M	£2.7M
Hardware	£3.3M	£2.7M
Electricity	£2.4M	£2.6M
Manpower	£1.9M	£1.8M
Total	£12.0M	£9.7M
Total/Year	**£2.4M**	**£2.4M**

Table-xx:

Capital costs includes machine rooms, network infrastructure etc. For 2005-2009 the actual costs were previously reported as £10.7M giving a total of £18.3M. However for 2010-13 the number includes the cost of machines rooms bought earlier amortised over 10 years and so for consistency we have adjusted the 2005-2009 figure downwards. Hardware refers to additional CPU and disk made available to GridPP above the amount pledged to and funded by GridPP, costed using average prices at the time. For 2005-2009, electricity costs were those estimated by the institutes. For 2010-2013 the electricity costs were estimated based on running similar equipment at the RAL Tier-1. The numbers include a Power Usage Effectiveness factor for Air Conditioning etc (PUE=1.6 in 2010-13). Manpower is the additional amount of effort provided by the institutes that complements GridPP funded effort, if any, costed at average GridPP Tier-2 manpower costs. This includes additional system management, networking support and some academic management time.

It is clear that the Tier-2 institutes have been able to leverage very significant resources above those that STFC, through GridPP, has provided over a sustained period. In the absence of capital available from the SRIF and CIF schemes in the earlier period, the capital has fallen slightly but electricity costs have increased leaving the total constant at around £2.4m/year for the last 9 years. This leverage will scale partly with the level of resource located at a site but also with the number of sites that contribute. To date, it has been spread across all 17 institutes, which has allowed much of it (such as electricity costs) to still be at a level that institutes are happy to provide. If the number of Tier-2 sites goes down then the leverage will almost certainly go down significantly, both due to the loss of sites but also because electricity costs will become more visible if resources are concentrated at fewer places and are more likely to be charged.

In summary, loss of sites is likely to significantly reduce the leverage; previous investment in server rooms would be wasted; physicists at those sites will lose local support and may become less efficient; and there will be lost opportunities for outreach and impact activities.

Conclusion

The UK has successfully constructed and run a network of Tier-2s for over 10 years that have made major contributions to the WLCG and the experiments. The UK Tier-2s are among the largest in the world. Due to their distributed nature and the fact that all but one are located at major UK universities, GridPP has been able to leverage large additional resources making for a very cost effective solution to the LHC computing challenge. The UK Tier-2s are in an excellent position to fully exploit the previous major investments made by STFC and the institutes for the upcoming LHC Run 2.