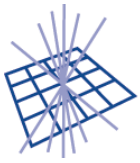

GridPP
UK Computing for Particle Physics

GridPP Project Management Board

Project Status

Document identifier:	**GridPP-PMB-168-ProjectStatus.doc**
Date:	**1/09/2014**
Version:	**1.0**
Document status:	**Draft**
Author	**PMB**

Introduction

The GridPP project was reviewed by the Oversight Committee in November 2013 and by a Strategic Review committee in March 2014. This report provides a further update and summary of the project status.

The LHC ran successfully for 18 months until Long Shutdown-1 started in February 2013. During that time, 25 fb-1 of data was processed and analysed on the Worldwide LHC Computing Grid (WLCG) infrastructure, and the Higgs Boson was observed. The Initial announcement on July 4th 2012 was confirmed later in the year as the data set grew, and on 8th October 2013 Peter Higgs and Francois Englert were awarded the Nobel Prize for Physics. The LHC is scheduled to restart in early summer 2015 and good progress is being made with upgrades to the LHC and the experiments. In the meantime, the computing Grid infrastructure has continued to (re-)process data; perform analyses; and produce simulated data. The Grid infrastructure has also continued to grow and evolve in preparation for Run-2 in step with the experiment computing models. ATLAS and CMS are currently performing large-scale tests to validate preparations for Run-2.

GridPP4 is in the final year of a 4-year project. The 32nd and 33rd collaboration meetings were held with themes on “Mini Review: A look at where we are and where we're heading”, and “Preparing for Run 2”. A proposal for a four-year GridPP5 project was submitted and reviewed. However, funding constraints led, instead, to a one-year extension of the current project (to be called GridPP4+) and the GridPP5 proposal will be resubmitted later this year.

International Context

WLCG now provides about 485,000 cores, 270PB of storage and processes in excess of 2 million jobs per day. GridPP currently provides about 11% of the total resources. Figure-1 below shows that the Tier-1 at Rutherford Lab continues to deliver 9% of the global LHC Tier-1 resources, as it has done since the start of data taking. Figure-2 shows 14% of the LHC Tier-2 resources were delivered by the UK over the same period, down slightly from 15% reported for 2011-2013. The difference between the Tier-1 and Tier-2 global fractions reflects the significant leverage of Tier-2 resources from the institutes that GridPP funding has enabled.

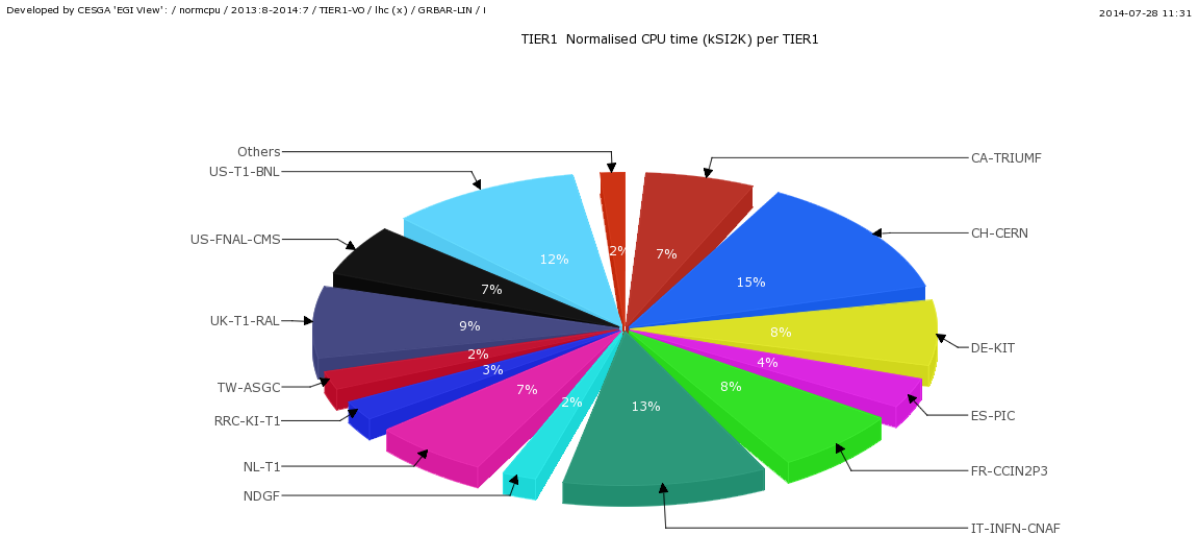

Figure-1: CPU delivered to LHC VOs (Aug 2013 to July 2014) by Tier-1.

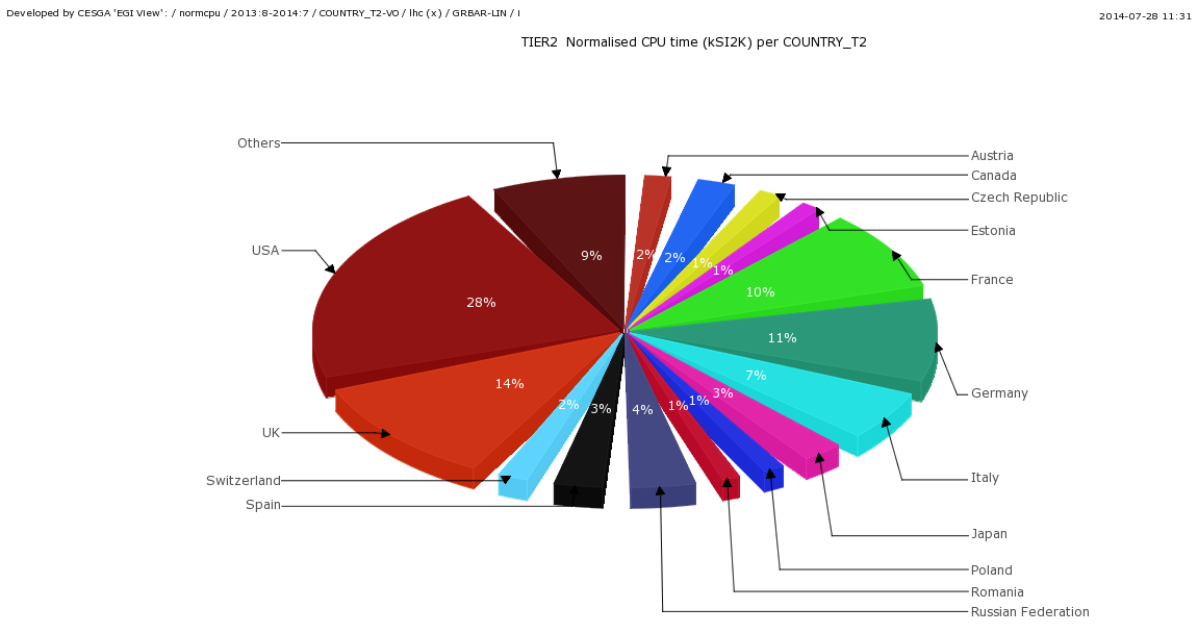

Figure-2: CPU delivered to LHC VOs (Aug 2013 to July 2014) by Tier-2s.

EGI-InSPIRE successfully completed its final annual review by the European Commission (EC) in July 2014 and has entered an extension period that is designed to bridge operations until H2020 funding becomes available in the new year. The UK continues to benefit from EGI, receiving support to provide services that are critical to WLCG (APEL accounting; Grid Operations Centre Database (GOCDB); and leadership of international security policy development). It is anticipated that these services will continue to be supported by H2020 funds, at some level, through the future EGI project. The UK is contributing to on-going discussions about the Governance model and fee-structure of the next phase of EGI.

GridPP is also engaged with, and contributing to, other H2020 initiatives, including the EU-T0 (an association of funding agencies who own substantial computational resources), the HEP Software

Foundation (a collaboration of HEP software projects designed to promote standards, reduce overlaps and provide common utilities); and VLDATA (a project that is developing and abstracting the DIRAC software/middleware for use by a wide range of other user groups).

GridPP4 Status

Since the last OC meeting the GridPP4 project has been progressing well. The production systems at both the Tier-1 and 2 sites have mainly been running smoothly. During this period the middleware has been migrated from EMI2 to EMI3 and the migration to the SL6 Operating System has been completed. Preparation for Run 2 at the LHC has included purchasing and commissioning new hardware to be ready to meet the April 2015 MoU requirements. The progress has continued to be monitored by Quarterly reports from the various Work Packages. The evolution of milestones and metrics over the period is shown in Figure-3 and the current Project Map shown in Figure-4. With three quarters left before the end of the project we are on track to complete all active milestones by the end of Q115.

	Q211	Q311	Q411	Q112	Q212	Q312	Q412	Q113	Q213	Q313	Q413	Q114	Q214
Metric OK	144	156	156	148	157	161	158	169	171	169	159	168	171
Metric close to target	18	16	19	17	14	17	9	7	6	5	17	9	7
Metric not OK	15	16	14	18	12	5	15	5	5	8	6	5	4
Not able to be measured	0	0	0	0	0	0	0	0	0	0	0	0	0
Suspended	2	12	15	21	22	18	19	19	18	18	19	19	19
Awaiting input	13	2	0	0	0	0	0	0	0	0	0	0	0
Metrics Total	192	202	204	204	205	201	201	200	200	200	201	201	201
Milestone achieved	7	12	21	23	28	33	34	40	46	54	60	62	62
Milestone close to completion	0	0	0	1	1	0	0	5	3	0	1	2	4
Milestone overdue	1	3	3	4	2	2	1	1	0	0	0	0	1
Milestone not due/ metric n/a	76	65	54	50	46	46	46	36	33	28	20	17	14
Milestones total	84	80	78	78	77	81	81	82	82	82	81	81	81
Total	276	282	282	282	282	282	282	282	282	282	282	282	282

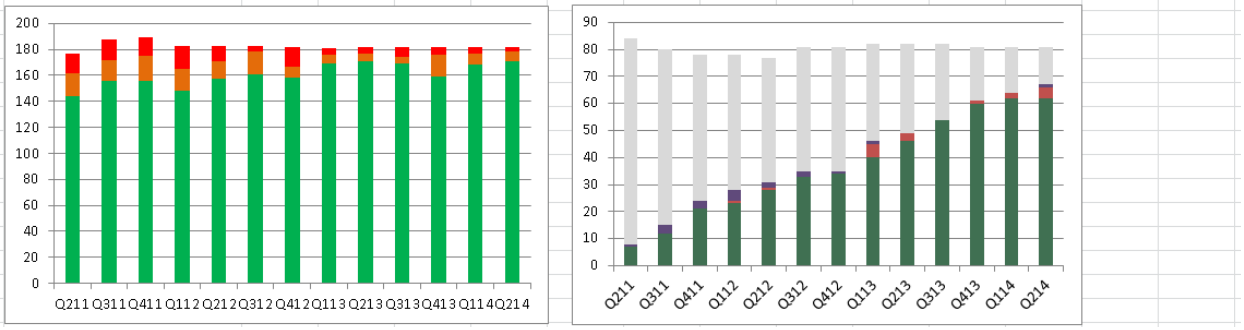

Figure-3: Evolution of Milestones and Metrics for GridPP4 to Q2 2014. The lower left plot shows the number of metrics that are currently met (green) and not met (red). The lower right plot shows the evolution of the number of milestones that have been met (red flags late ones). The data is presented numerically in the table at the top.

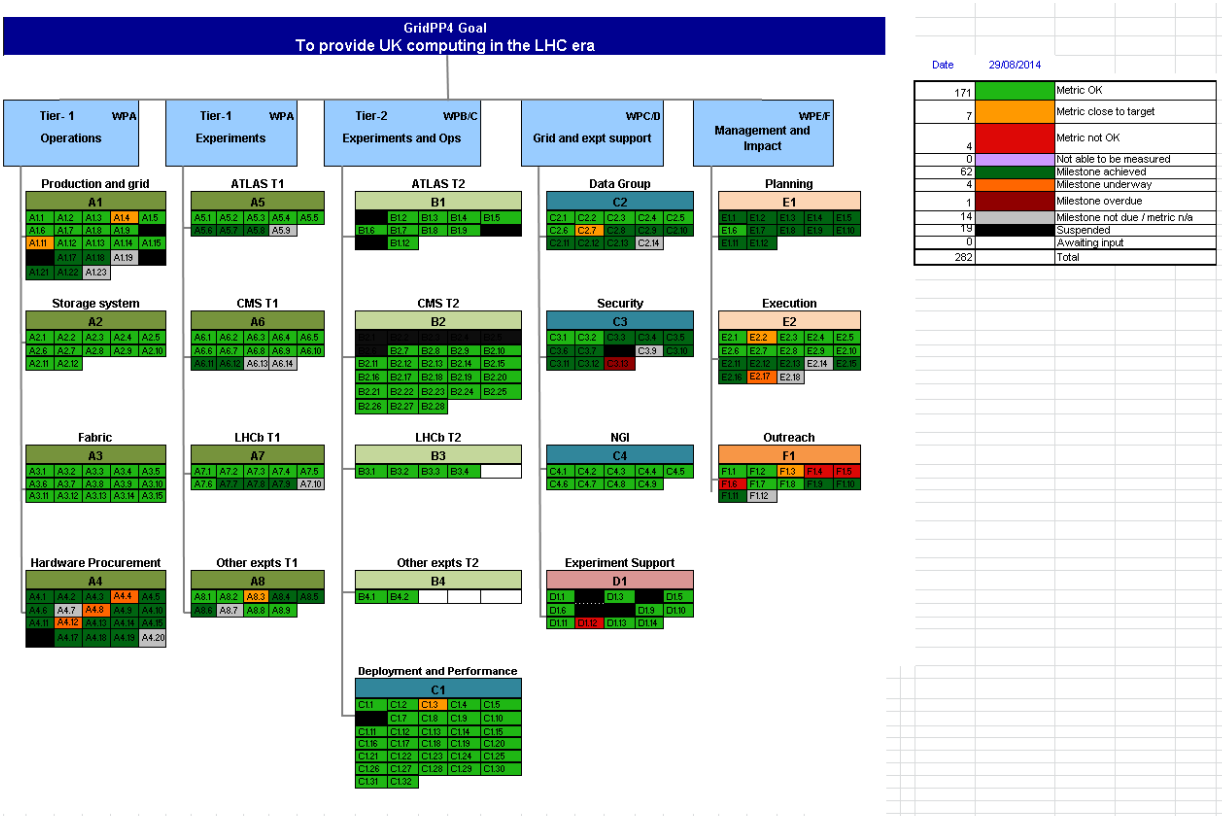

Figure-4: The Project map at the end of Q2 2014

Overall the majority of metrics are being met or are close to target, with just four metrics unsatisfied this quarter.
These are:

- D1.12 ATLAS Ganga, Dedicate one week per month to being a support shifter. Change of staff at Birmingham has meant there has been a gap this quarter as roles are reassigned.
- F1.4, F1.5 & F1.6 Number of news items on the GridPP website, GridPP press releases were low as were the number of press articles about GridPP. This is partly down to a period of steady state running where the focus was on LHC running and the consolidation of the Higgs result. However, it also reflects an attempt to refocus our effort towards the Impact Agenda, rather than simply dissemination. The GridPP Dissemination Officer appointed in October 2013 has been active in promoting the use of the grid to small VOs such as 'CERN@School'.

One milestone that is due by the end of Q2, 2014, is overdue and three are delayed.

- C3.13 Security recommendations for the future. With the extension to GridPP4 by one year and EGI-InSPIRE by 8 months it makes sense to delay planning for the future until Spring 2015.
- A4.4, 4.8, 4.12 ('Tier-1 strategy review', 'procurement plan', and 'procurement started') are all ongoing but not yet complete. We expect to submit paperwork in late August for delivery in February.

The quarterly reporting has continued to be useful to track issues and focus the different groups on their priorities. The overall performance of the different work packages in GridPP has been very successful and this has been reflected in the project map and reports. All detailed quarterly reports are available on the web at: http://www.gridpp.ac.uk/pmb/ProjectManagement/QuarterlyReports/reports.html

Risk register

The GridPP4 risk register summary is presented below in Figure-5 and the full version is available at: https://www.gridpp.ac.uk/pmb/ProjectManagement/GridPP4_Risk_Register_v1-Q214.xls

Ref.	Risk Description	Owner	Inherent Risk			Residual Risk		
			Likelihood	Impact	Total	Likelihood	Impact	Total
1	Castor Storage System Problems	AS	0.4	60	24	0.5	60	30
2	Expt. s/w runs poorly on the grid	PG	0.4	60	24	0.2	50	10
3	Middleware at T2s performs badly for user analysis	SL	0.4	60	24	0.2	60	12
4	Outage of UK T1	AS	0.4	60	24	0.1	50	5
5	Loss of experienced personnel at T2s	SL	0.6	50	30	0.5	40	20
6	Failure of T1 to meet SLA or MoU	AS	0.6	50	30	0.1	50	5
7	Security problem affecting reputation	DK	0.7	80	56	0.5	60	30
8	Los of service due to security	DK	0.3	50	15	0.1	50	5
9	Insufficient manpower to operate core T2 sites	SL	0.5	40	20	0.2	40	8
10	Recruitment retention problems at RAL	DK	0.9	60	54	0.6	50	30
11	Hardware costs aren't accurately predicted by model	DB	0.25	50	12.5	0.2	20	4
12	Core service funding insuffcient	CD	0.5	40	20	0.2	10	2
13	Insufficient travel funds	DK	0.4	25	10	0.15	25	3.75
14	Expt. Computing models change	PG	0.4	40	16	0.3	40	12
15	Breakdown of NGI/EGI infrastructure	JC	0.3	30	9	0.1	30	3
16	Critical middleware no longer supported	DC	0.4	40	16	0.1	40	4
17	Significant loss of custodial data at the T1	AS	0.1	75	7.5	0.1	50	5
18	Unplanned infrastructure costs	PG	0.25	30	7.5	0.1	10	1
19	Loss or damage to hardware at T1	AS	0.05	100	5	0.05	100	5
20	Disaster at T1 leads to prolonged outage	AS	0.05	90	4.5	0.05	80	4
21	Insufficient funding at T2s for h/w	SL	0.3	40	12	0.2	30	6
22	Loss of EGI.eu	CD	0.5	20	10	0.5	15	7.5
24	Insufficient Network Bandwidth	PC	0.4	50	20	0.2	50	10
25	Technology Mismatch	DC	0.2	50	10	0.05	50	2.5
26	Insufficient support effort	PG	0.4	40	16	0.2	20	4
27	Financial Uncertainty	DB	0.2	50	10	0.1	50	5
28	Failure to deploy or operate hardware	JC	0.5	50	25	0.1	50	5
29	Conflicting opinions amongst GridPP stakeholders	DB	0.3	50	15	0.1	50	5
30	Over contention for resources	DB	0.3	40	12	0.1	40	4
31	Capital vs Resource	DB	0.8	50	40	0.3	60	18

Figure-5: The GridPP4 Risk Register Summary

The risk register is reviewed regularly by the PMB and the key current risks for the next 6 months are:

1: Castor storage system problems.
Although reliability has been good there are some indications of performance related problems at high transaction rates. An extensive testing infrastructure, close liaison with CERN and proactive monitoring and tight change control over version changes reduce the potential risks. However an alternative solution is being developed and will be tested at scale before migrating of large fraction of the production data.

7 Security problem affecting reputation.
There is a continuous need to be vigilant on all security issues. Security problems may bring down the whole of GridPP or even WLCG, given that all Sites run very similar software. A major outage is likely to attract significant outside interest and GridPP is likely to suffer damage to its reputation. The Tier-1 in particular could be an attractive target for 'Bit Coin Miners'. Attacks of this nature have become common in the last 6 months. We will continue to act on many fronts to reduce the threat by

ensuring software is up to date and to contain and handle incidents quickly when they occur. The UK (GridPP) leads the Joint Security Policy Group and the security vulnerability group. The security team operates across GridPP sites to promote best practice and many staff have recently had security training as part of a HEPSYSMAN event. It is a high priority to fill the security officer position, to coordinate all security issues and lead the distributed team.

10 Failure to retain or recruit key technical staff at RAL
Losing key staff could impact the projects ability to meet deliverables and key milestones. This is a potential problem that has to be dealt with rapidly by STFC management by speeding up recruitment in the event. It has been noted that recruiting new staff is increasingly difficult due to non-competitive salaries and conditions.

Risks that are high but under control are:

5 Loss of experience personnel at Tier-2s
This risk is mitigated by sharing knowledge across sites, but it is important to keep the staff informed of future funding plans to reduce staff loss towards the end of the project. Uncertainty about GridPP5 and possible delays in contract extensions increase the risk at sites, although the risk the overall project is mitigated by the distributed nature of the Tier-2s.

6 Failure of T1 to meet SLA or MoU commitments
Failure to meet our WLCG obligations would damage our reputation. Many systems have been put in place to cover call out, change management, disaster recovery and hardware procurement. These continually focus operations staff on the MoU commitments to ensure excellent performance.

31 Difficulty with budgets due to Capital vs Resource limitations
Classification of different types of equipment as capital or resource can sometimes change causing problems with budgeting. It is important to have good frequent communications with STFC to track such changes and limitations.

Tier-1 Status

Overview

This section of the report covers the period from October 2013 to July 2014.

During this period, the priority areas of work have been:

- On-going routine stable operation of the production service and preparation for LHC Run 2 in 2015
- Completion of procurements and deployment of hardware to meet the 2014 MoU commitments
- Continuing development of new services (such as cloud and object store) in order to meet potential needs of new user communities

Fabric and Infrastructure

The 2013 disk and CPU procurements proceeded routinely and were deployed in time to meet 2014 MoU commitments. A tape media purchase of 1020 tapes was carried out, providing 5-8PB of capacity depending on density written.

The two new generation T10KD tape drives successfully completed testing and a further 8 drives were purchased in FY14. These drives extend the existing 5TB T10KC media to 8TB. The CASTOR tape back end was upgrade to enable production operation. In September repacking to T10KD of the T10KA and T10KB generations of tapes will commence. Approximately 8000 tapes containing 5PB of data will need to be repacked, this is expected to take at least 6 months to complete after which the tape service will be operating on only the T10KC and T10KD generations. Average maintenance costs of the tape drives will increase significantly over the following year as the low maintenance cost A and B generation (inherited when Oracle purchased Sun Microsystems) are phased out.

In May 2014 a 3-year-old generation of disk servers were phased out of operation prematurely due to unacceptably low reliability leading to data loss. The shortfall in resources was replaced from the procurement reserve, which is held to address problems of this kind. The hardware supplier (Streamline Computing) went into receivership in 2011 after which the maintenance contract of Streamline equipment ceased to be honoured. Considerable effort was being expended in resolving hardware problems and, given the small volume of hardware concerned, this was not cost effective use of staff time.

The R89 machine room has generally provided a stable operating environment during this period.

Considerable restructuring of the site network has taken place and further work is underway. As a consequence, site network reliability is much improved recently. The Tier-1 40Gb mesh network was deployed into operation and has seen considerable increase in load in recent months as experiment data processing increased substantially. Rates at the batch farm peaked at over 10GB/s in July (Figure-6) – moving approximately 1PB of data per day between batch farm and CASTOR.

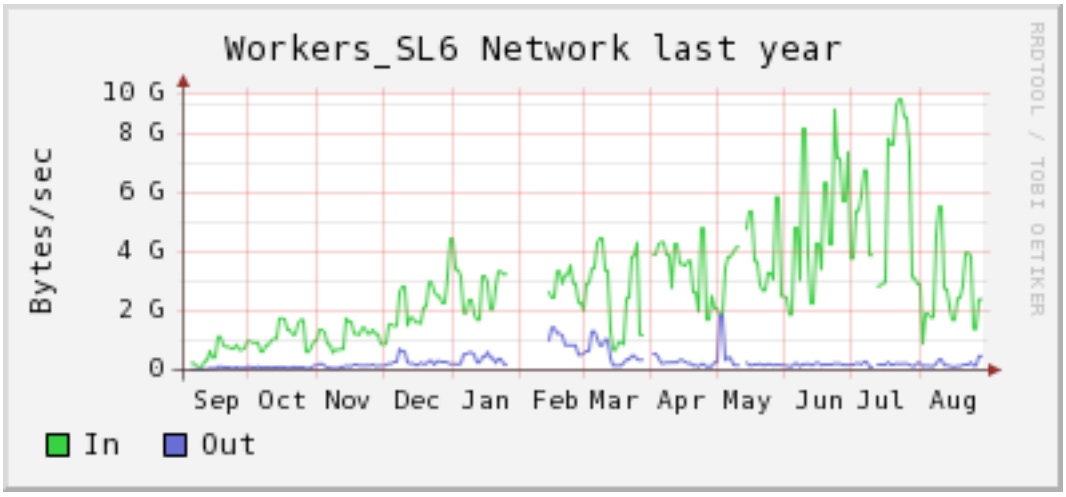

Figure-6: Batch Farm Network rates

Production, Operations and Service

Tier-1 operations have remained reliable over the period of the report despite ever increasing data rates presenting new challenges at times. Average RAL (ops test) service availability since October 2013 was 99% compared to a WLCG target of 97%. Average RAL availability since October 2013 for the LHC VOs was 98% breaking down as follows:

	4Q13	1Q14	2Q14
ALICE	97%	100%	99%
ATLAS	96%	100%	99%
CMS	99%	97%	99%
LHCb	97%	99%	99%

Figure-7: Monthly VO availability at Tier-1s

Middleware, Grid and Cloud Computing

The Tier-1 continued to be heavily used over the reporting period. Farm occupancy averaged 84% and job CPU utilisation efficiency averaged 88%. Since October 2013, RAL delivered 9% of global LHC Tier-1 CPU computing – similar to the previous report. (Figure-8)

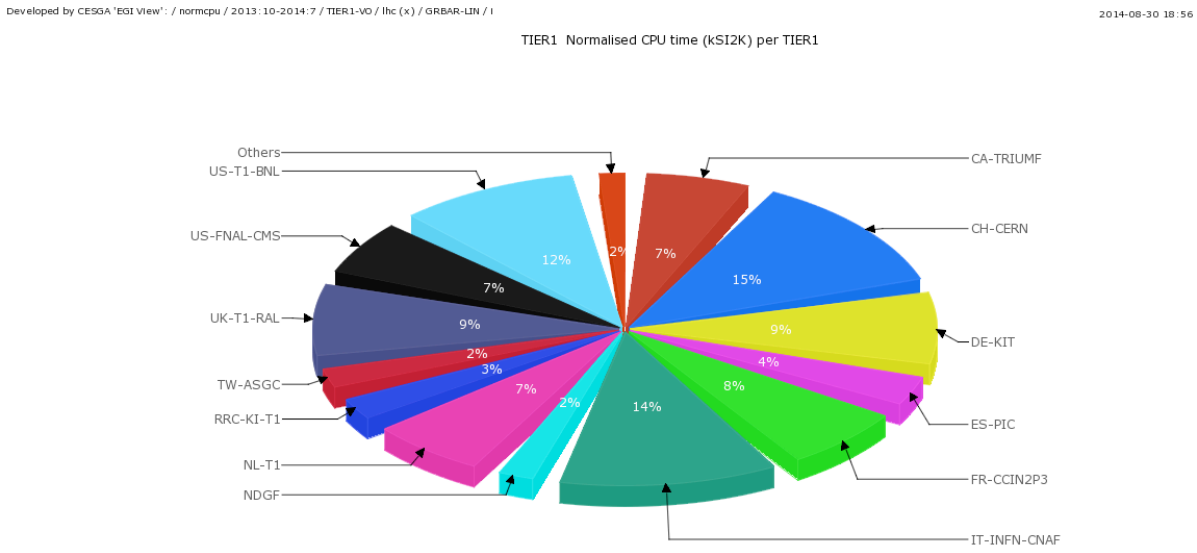

Figure-8: Global Tier-1 usage October 2013 to August 2014

The project to replace the PBS batch scheduler with HTCondor completed successfully. This successfully resolved the reliability issues described in the previous report.

Work commenced again on developing a replacement for CASTOR as a disk pool manager. This was motivated by performance issues beginning to appear in CASTOR as transaction rates have increased through the period. At the same time a number of recent developments in CEPH addressed issues identified during the previous evaluation. Work is underway with the experiments to evaluate a test CEPH instance after which a considerably larger deployment will be required to test CEPH at scale.

Little effort has been available to progress the test cloud service mentioned in the previous report. However a recent internal recruitment has provided effort required to make progress once more with the intention of an early trial service being ready by Q2 2015.

Management, Business Processes and Communications

The Tier-1 reached full planned staffing levels in 14Q1, however subsequently the situation began to deteriorate and is likely to deteriorate further. The uncertain funding situation for the Tier-1 led to a number of staff departures and in the absence of a firm funding plan it was not possible to recommence recruitment until recently and approvals from within STFC are still pending. The expected 1 year extension to GridPP4 has helped somewhat but advertising fixed term posts for only 18 months are unlikely to be particularly attractive. Furthermore STFC salaries for IT staff have become increasingly uncompetitive and applications to recent recruitments in other groups have been very disappointing. It is therefore likely that staffing levels will fall below planned levels throughout the remainder of GridPP4 with a consequent impact on both on-going development and operations.

Deployment Status

The section summarises deployment and operations activities for the period October 2013 to August 2014.

Resources deployed have seen significant changes, particularly in Q2 2014 with most Tier-2 sites deploying hardware procured with their GridPP4 2nd tranche hardware awards. There has been a 20% increase in HS06 available, which now stands at about 430,000 HS06, and a 28% increase in disk leading to 31,000 TB nominally available across GridPP sites. There are more deployments to come as acceptance tests are on-going in some institutes, and several sites had their awards deferred to this financial year due to constraints on funding from STFC. However, these increases may be partly offset, as there is a non-negligible amount of hardware now out of warranty and becoming too unreliable to be sustained. Even with the disruption of new hardware deployments, GridPP site availability and reliability (which since January is assessed separately for each of the LHC VOs) has on the whole remained well above the WLCG targets of 95%. There has not been a repeat this year of some sites having to put worker nodes offline due to high machine room temperatures – over the last year several machine rooms have been renovated while others have had some hardware migrated to central university machine rooms.

The newest hardware is generally well understood, but challenges are being faced as the character of experiment jobs change, and internally machine resources become increasingly stretched with higher core numbers. GridPP has been leading in several areas of deployment testing, particularly in relation to multi-

core approaches on different batch systems (A GridPP member is chairing the WLCG task force in this area and also led the migration to SL6 in 2013). Deployment of these (multi-core and whole node) queues is now a priority as the experiments (particularly ATLAS) make more use of multi-threaded jobs. In conjunction with this multi-core work many sites are seeking to move away from Torque/Maui based batch systems and during the current reporting period about 30% of sites have either moved to HT Condor, SGE or SLURM or set up test systems ahead of a move in the near future; most sites have also increased their use of Cobbler and Puppet configuration management (and are actively supporting one another through the sharing of modules). It is expected that the majority of sites will have moved batch system by the end of the year.

This period has witnessed a number of changes to the infrastructure with many large-scale changes driven by top-down upgrade campaigns. The first of these was a move of operating system (particularly on Worker Nodes (WNs)) to Scientific Linux 6 with a target of October 2013 that our sites met. The move led to performance improvements that required sites to rerun and publish their HS06 performance[1]. Another successful campaign in recent months required sites to move from unsupported EMI-2 to EMI-3 middleware.

The middleware products upon which WLCG/GridPP sites depend are generally supported now for the coming year(s), but there are exceptions. A notable one is ARGUS, which handles authentication. Following a drive to enable pulling down of banning lists from a WLCG and regional ARGUS server, it has become apparent that the product still has some instabilities and fixes are not being propagated quickly. At the time of writing no organization or group has agreed to take over maintenance of the product from SWITCH, an organisation who agreed to support it until 1 year after the end of EMI (March 2013).

Security operations in GridPP continue to be well managed by a distributed team who partake in a duty rota. The security officer position is still vacant despite a recruitment round. The last 10 months have witnessed a steady flow of security advisories and updates (e.g. the last IGTF release was in June) to which GridPP sites have responded well. This was particularly the case with the openSSL “heartbleed” vulnerability in March of this year. We have seen a couple of site security incidents since October 2013, but none that have caused ongoing concern. Earlier this year we fully commissioned our resilient network of VOMS servers and, where requested, took over VOMS support for a number of NGS VOs. More widely we have contributed to joint discussions on future security policy and approaches including with EGI, EUDAT and PRACE teams, and to setting timelines for the community to move to SHA-2 – the UK CA switched to issuing SHA-2 certificates in April this year. In June a security training workshop took place at RAL for our sysadmins and this was very well received; it may now become an annual event.

This period has seen steady progress in GridPP investigations into, and deployment of, cloud and VM technologies. The GridPP cloud at Imperial moved to using OpenStack and the IceHouse release (together with Gluster FS for image storage) and is well advanced in becoming part of the EGI Federated Cloud project. Oxford and the RAL Tier-1 have also deployed Openstack. Lancaster has investigated VMWare cloud, QMUL is deploying a Cloudstack prototype and Manchester is pioneering a new approach based on its own VAC and Vcycle methods. In recent months Manchester expanded VM use onto WNs under its conventional batch system. The GridPP work has been well received by the LHC experiments (who are now using cloud resources for some production work – and of course to get more benefit from HLT resources at CERN) and WLCG management. There have been an increasing number of meetings and workshops aimed at pooling the experiences of the WLCG community into a future strategy in this area. EGI is now making cloud approaches part of its core offering. One area of increasing priority for further investigation is in the area of cloud/VM security – misconfiguration of VNC on some nodes led to a security incident in August.

[1] HS06 as a benchmark is increasingly challenged as an accurate indicator for WLCG jobs and work will start on a replacement towards then end of 2014.

There have been a number of developments and improvements in the area of operations support. Within the WLCG context GridPP is making significant contributions in areas such as the monitoring consolidation task force (whose mandate includes reducing the effort needed within operations and earlier this year led changes to the SAM infrastructure) and IPv6 studies (Imperial was one of the first sites in WLCG to move to running its services IPv4/IPv6 dual-stack). The UK was one of the first regions to fully adopt perfSONAR at all its sites for network monitoring and also deployed a prototype dashboard. There is still work to do in enabling the full network mesh and resolving some site firewall issues. perfSONAR has proved a very useful tool in resolving end-to-end network issues and will be part of the considerations of a new network and transfer metrics group – clear metrics are needed to advance network status aware tools. GridPP is also working with a RIPE initiative using plug-in network probes (confusingly called ATLAS probes) to improve network monitoring coverage.

The UK is active in the WLCG middleware readiness working group – whose aim is to introduce more stringent testing of products within site infrastructures and experiment workflows before they become part of our recommended baseline. To coordinate in this area CERN has recently appointed a middleware officer.

Our sites have been responsive to middleware change requests; for example all moved within weeks to CVMFS release 2.1.19 during July as this was needed for CERN to upgrade its Stratum-0 release. UK sites have also been making good progress in resolving issues affecting the move to Glue2 – a validation exercise started in November.

Storage and data management has remained a very active area. With a move to Rucio distributed data management, ATLAS required all files at our sites to be renamed – this went smoothly and finished on target in Q4 2013. The main issue was with Rucio creating empty directories when first used in December. Some problems were seen this year in June when ATLAS localgroupdisk became full at certain sites causing those sites to fail Nagios tests, which impacted their availability. There were also a few niggling issues encountered with federated xroot in December. Although the majority of tickets/issues continue to be resolved quickly, there are a few that have eluded resolution for many months and top among these is one concerning LHCb pilot jobs losing connectivity at Bristol and RAL.

We have continued our support of the DPM Collaboration and agreed a revised collaboration agreement and work plan (in areas such as interface testing, documentation and support) earlier this year. We organized a DPM workshop in Edinburgh during December 2013. FAX and WebDAV services have been enabled at many sites and Wahid Bhimji of Edinburgh led a WLCG wide meeting on data access (and dynamic federations) in May 2014. There have been some major DPM upgrades during the period and these have gone smoothly, largely as the result of a lot of testing beforehand. Another noteworthy change is that LHCb started using storage at Tier-2s (at Manchester and RALPP) and this has gone well.

There have been two WLCG workshops during the period covered in this report: The first took place in November 2013 in Copenhagen and focused on areas such as Run2 preparations and the evolution of the experiment computing models (including storage), and the outlook for e-Infrastructure. The second was held in Barcelona in July 2014 and prompted thinking beyond Run2. Our team (particularly on storage aspects) presented at CHEP in October and at HEPiX in November.

Core operations work has continued smoothly during the last 10 months. Weekly ticket reviews have been sustained and tools like VOMsnooper updated to aid site alignment with VO portal changes. There has been progress on monitoring visualization based on graphene (an iOS app was developed as proof of concept to show sites how the data might be used), and the deployment of VO Nagios for five of our smaller VOs. The GridPP Nagios service has received regular and on-time updates. Blog posting by team members and site admins, which act as a useful way of disseminating technical findings and solutions, is improving. The ROD activities have strengthened with the Tier-1 now also contributing a team member to the weekly tasks; it should be noted that the work has recently become more difficult due to an untested

and somewhat less functional operations dashboard being released into production by EGI. Problems have included non-working email functionality, a loss of the security dashboard and bugs preventing Nagios alarms from being removed. To avoid such poor releases in future an advisory and testing group is being established between NGIs.

Our sites have seen a number of staff changes, for example at Glasgow, Durham, Birmingham and Sussex. Cross-site support has helped sites remain in production during such transitions but this comes at a cost.

Users' Reports

ATLAS

During GridPP4, ATLAS has made extensive use of Tier-2 resources for simulation, group analysis and individual user analysis, and the Tier 1 resources primarily for reprocessing, simulation and data curation. All UK Tier 2s were in the top 20 for delivered HS06-hours in the year from August 2013, and two in the top 5. Inevitably, this pattern has been slightly different in 2013 and 2014 because we are in shut-down, so the emphasis has been on simulation, analysis and scaling exercises for the new computing mode; nonetheless, reconstruction (of real data and of simulated data) still represents a significant activity. The overall UK CPU usage by ATLAS in the UK is broken down by activity in Figure-9; this shows a healthy balance between analysis, Monte Carlo production and reconstruction. There is an increasing use of major Tier-2s in dynamic data placement and reprocessing activities; the latter will already be a more crucial role in GridPP4+, with a major revision of the processing and analysis model now rolling out, and will continue to grow in GridPP5. At present, the dynamic data placement in the UK is dominated by analysis formats (ntuples and the AOD), as illustrated in Figure-10. The UK has met its disk pledge to ATLAS within the RRB year, as illustrated in Figure-11.

The UK continues to perform well, delivering roughly 10% of the global Tier-2 capacity (reflecting the UK authorship share) and 13% of the Tier-1 capacity (our authorship share of Tier-1 providers). ATLAS has also made good use of beyond-pledge resources made available by the institutes from their own funds. It has allowed UK groups to exert direct influence on the development of analyses by the provision of over-pledge capacity for work using the institutes' leveraged internal resources.

ATLAS has revised its projected future resources assuming the tightly constraining envelope proposed via the Computing RRB/CRSG. This makes evident the major challenges. The processing time per event has already been considerably reduced (by a factor of two), but this speed-up must continue. More challenging is the growth in data volume; the scope for reduction in disk-resident copies is now exhausted and serious compromises in the physics achievable have to be considered.

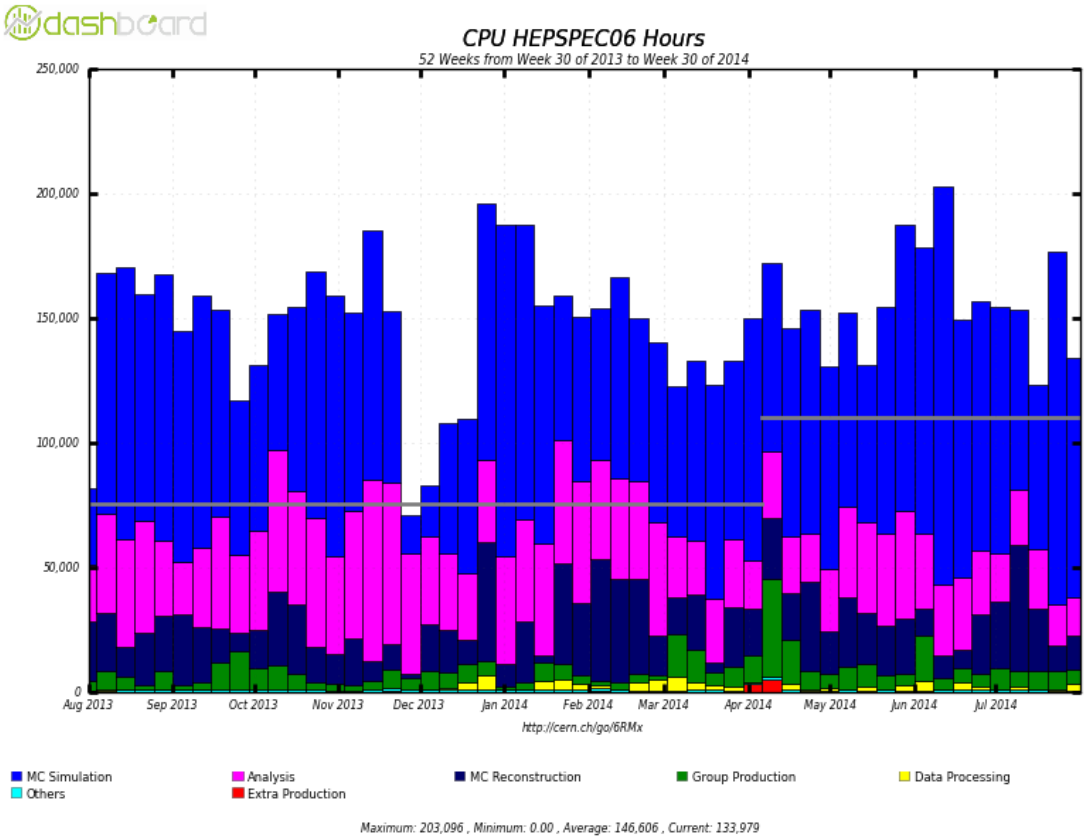

Figure-9 The ATLAS UK CPU usage by activity; the grey line represents the pledged value.

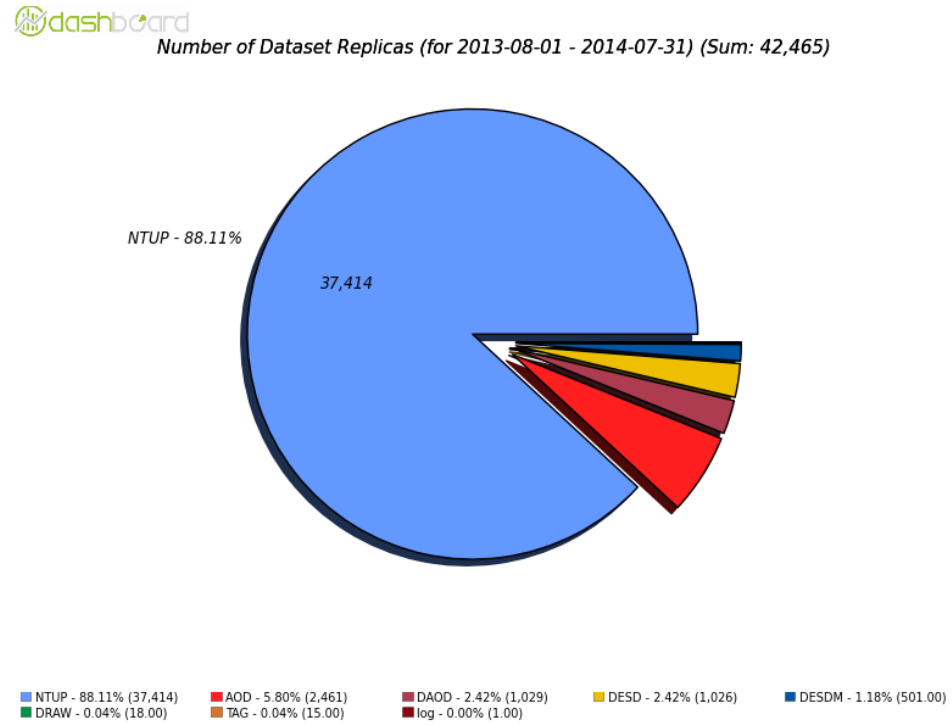

Figure-10 A snapshot of the typical data held in the UK through dynamic data placement by data type; it is currently dominated by analysis formats.

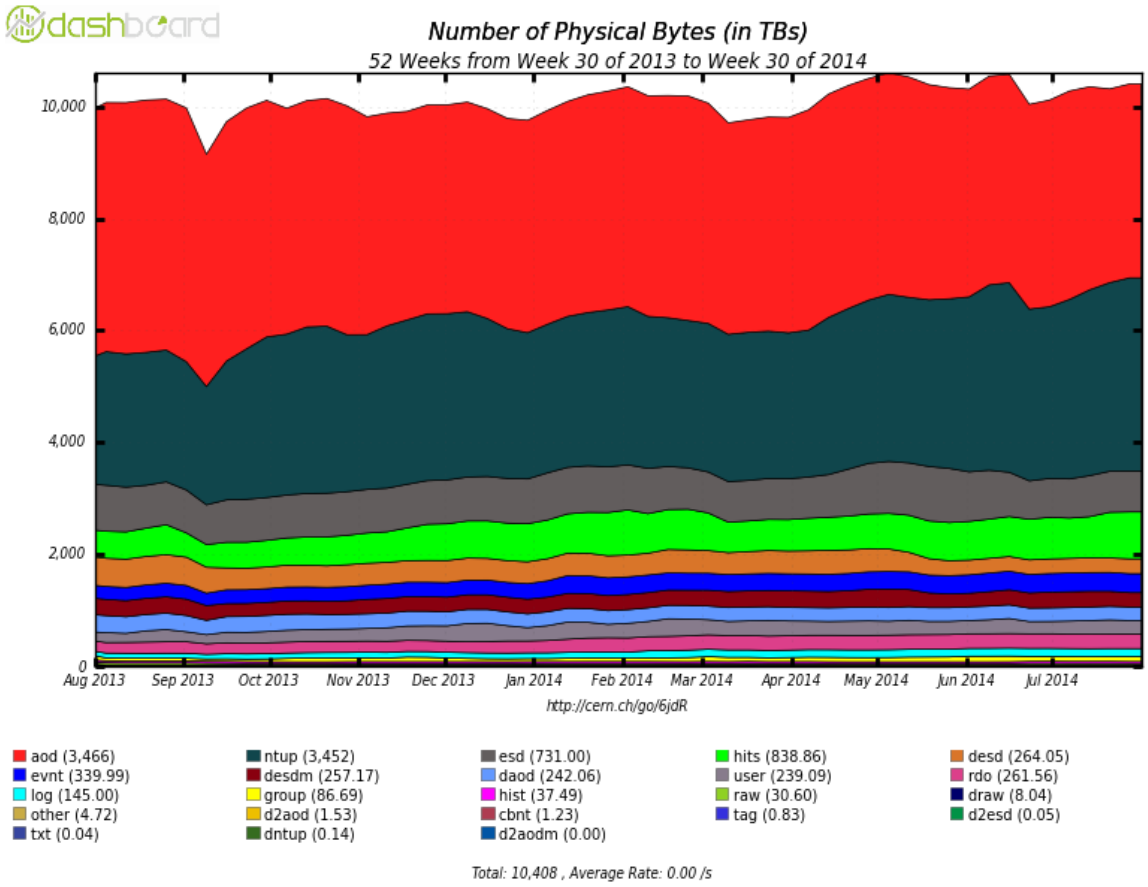

Figure-11 The UK disk capacity delivered to ATLAS through the year from August 2013, broken down by type; the grey line represents the pledge.

The role of the Tier-2s continued to grow in ATLAS in 2013 and 2014, with the major sites beginning to serve data to smaller sites with little storage using data federation. Analysis requires large amounts of disk with excellent tuned bandwidth to the local CPUs and good inbound and outbound network connections to other sites plus many additional services. These sites 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 and possible data-loss addressed. The transfer to the Tier-2s of workflows that were formerly done at the Tier-1, such as reprocessing, reinforces the conclusion that the current distributed Tier-2 structure remains a balanced and resilient solution for the ATLAS Tier-2 requirement. ATLAS continues to estimate that a region the size of the UK requires a minimum of 10.5 FTE to run an adequate Tier-2 infrastructure and, in fact, the UK has one of the smallest levels of deployment and operation staff, despite being one of the largest ATLAS regions. In GridPP4, two FTE were allocated to each of five major ATLAS sites in order to provide continuity of operations and support. These have proven invaluable in the last year, with the sites performing with good reliability and availability.

ATLAS globally has investigated distributed computing implementations other than the current Grid model, including the limited use of commercial cloud resources and the opportunistic use of HPC resources. The UK has been particularly active in this work, making contributions to the pilot job system for the clouds, running a pilot factory pointing to UK cloud resources. While most workflows have been tested and made to work at some level, cloud resources are only feasible and efficient for about 10% of the total work, and HPC resources for even less; however, work continues to make best use of all available resources when appropriate. ATLAS-UK has also been adapting and making use of the VAC system for ATLAS.

CMS

The period since the last oversight committee meeting has been one of consolidation within CMS. Activity has been a mixture of producing detailed and refined analyses of the data taken in Run 1 and preparation for Run 2. There has also been evolution of the computing model used by CMS, which aims to make more efficient use of the resources available to CMS.

The UK plays an important role in CMS computing operations by providing a Tier-1 (at RAL) and three Tier-2 sites (at Brunel, Imperial and RAL PPD). The UK only provides a small fraction of CMS' projected hardware needs. However, our very efficient use of these resources means that we are able to deliver a proportionately higher fraction of the useful work done. We are able to do this because of the high quality of the individuals, employed both through GridPP and the Consolidated Grant, who work in CMS computing in the UK. This group of people is smaller than the number of people CMS estimate would be required to deliver the UK contribution and we are regularly awarded more service credit than we have people. The UK's role is very much appreciated within CMS computing. It is pleasing to note that UK sites are at the forefront of the adoption of the changes required to evolve the computing model.

The RAL Tier-1 is pledged to deliver 8% of CMS total Tier-1 hardware requirement and funded by GridPP accordingly, but, as can be seen in Figure-12, the UK has run nearly 11% of the successfully completed jobs. This reflects the fact that CMS has been able to make more efficient use of the resources at RAL than those at some other Tier-1s and is a tangible testament to the efficient operation of the RAL Tier-1 and the CMS specific staff embedded within it. These staff are seen within CMS as being highly responsive. One significant example of this is that RAL was the first CMS Tier-1 to separate the operation disk and tape resources as required by the evolved computing model.

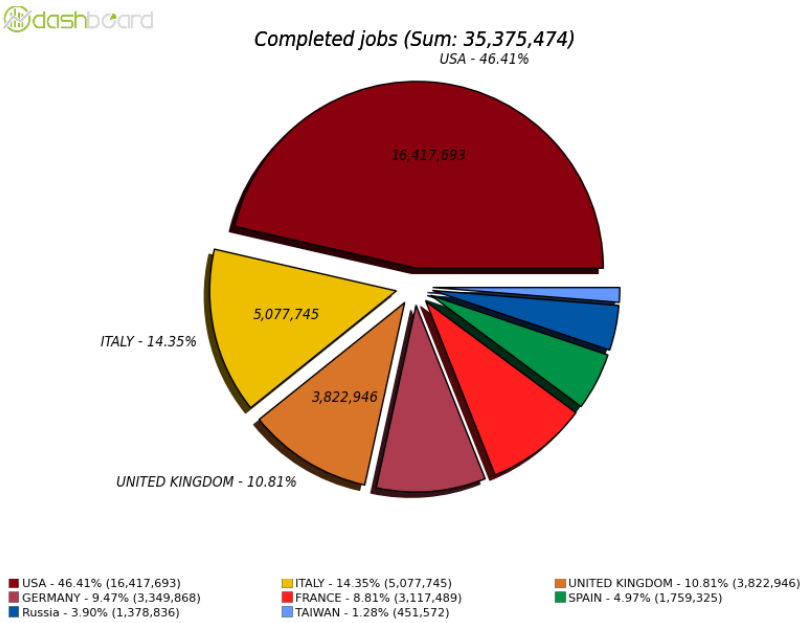

Figure-12 Number of successfully completed jobs at CMS Tier 1s since the last OC meeting (as measured by the number of successfully completed jobs).

GridPP pledges to provide 5% of the Tier-2 computing hardware that CMS projects that it will need. Figures-13 and 14 show the UK Tier-2 actual contributions to CMS (measured by successfully completed jobs). Again it is noteworthy that through the talented people employed at CMS Tier-2 sites, who have enabled particularly efficient use of the resources available and through additional resources provided by

the institutes themselves (i.e. not through GridPP) the UK has been able to deliver a far greater fraction of the used resources than the pledged figure.

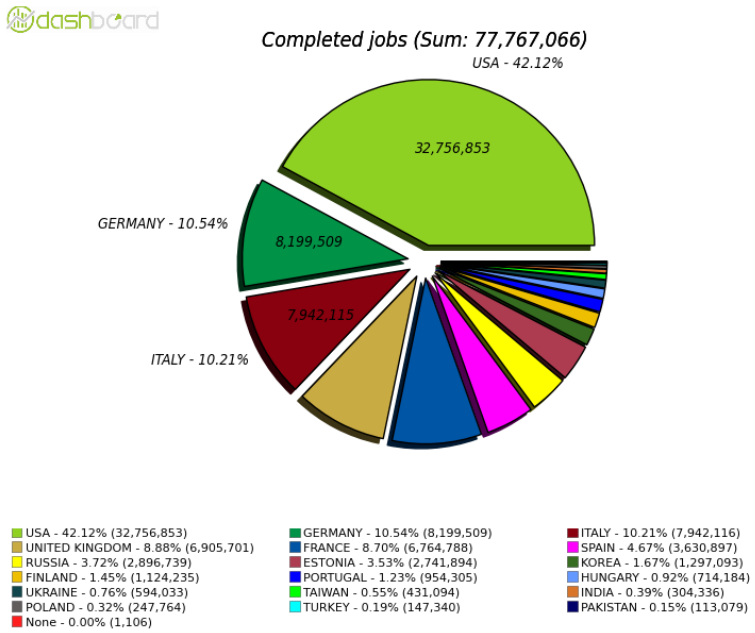

Figure-13 Contribution grouped by country since the last oversight meeting (as measured through the number of successfully completed jobs).

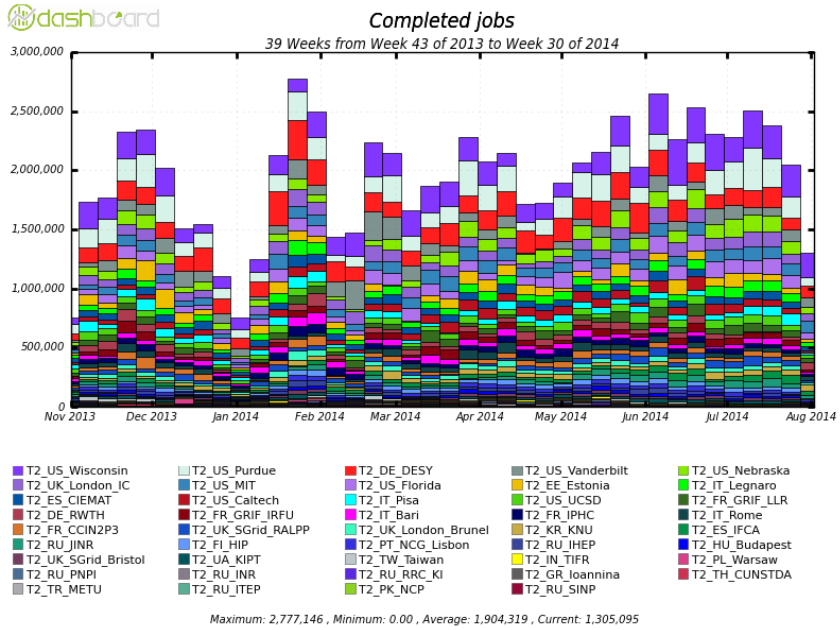

Figure-14 The contribution of individual Tier 2 sites to CMS (as measured through successfully completed jobs).

The same UK individuals who run the UK operations so efficiently for CMS also play significant play significant roles in CMS' central operations and development. One example of this is in the area of cloud computing which we expect will be of increasing importance to CMS computing. Centrally UK efforts have been focused on both using the CMS HLT farm as a cloud resource and on how CMS will interact with the CERN Agile Infrastructure. These efforts are part of a much wider team which is currently led by a UK person.

The UK's contribution to CMS computing, both in terms of operation of resources and the greater role played by GridPP people is highly appreciated by CMS management.

LHCb

The UK continues to be jointly the largest national collaborator in LHCb along with Italy. We are 21% of the collaboration for Tier-2 purposes and 31% of the collaboration within Tier-1 countries. During the last period the UK has continued to be a very reliable provider, both at Tier-1 and Tier-2 sites. This reflects both capacity but also as importantly the reliability and availability of resources. Figure-15a below, taken from the most recent report to the Computing Resource Scrutiny Group (CRSG), show the comparative UK CPU usage at Tier-1 sites (left) and Tier-2 sites (right) during 2013. From the LHCb point of view the UK pulls its weight and responds quickly and positively to any issues that arise. There have been no significant problems due to UK GridPP resources.

During 2013/14 LHCb has mainly carried out Monte Carlo simulations, a restriping of 2014 data, as well as heavy ion and low energy 2013 data processing, a full reprocessing of 2011 data as well as continuous user analysis. The balance of workload is shown in Figure-15b below.

The introduction of T2D centres (Tier-2 centres with disk) has proceeded smoothly and successfully. We now have Tier-2 sites in several non-Tier-1 countries contributing and expect 1.2 PB of disk to be provided by the end of 2014. As usual, the UK T2D sites at Manchester and RAL have proven to be extremely effective.

LHCb has evolved its computing model for the approaching Run-II to include three categories of data (1) The normal prompt data stream to be processed immediately, (2) a parked stream to be kept for later processing and (3) a turbo stream, which is processed within the trigger farm. The prompt stream will be reconstructed after a delay of a few hours, to allow calibration and alignment to be completed. This means that no further reconstruction will be needed as has been the case in the past, when a full reconstruction took place immediately after the end of data taking in any year. The parked stream will not be needed on 2015 and 2016 but may be needed in later years. The turbo stream is mainly for charm physics, and assumed that all necessary information can be created in the trigger farm itself, and therefore no further offline processing is needed. These changes will help to reduce the overall CPU requirement, and LHCb continues to use non pledged resources in Russia, and to make full use of its trigger farm.

LHCb is also developing the means to parallelise its framework, Gaudi, and is very much assuming many core-devices will be the norm in future. LHCb is also developing virtualisation mechanisms associated with the desire to use heterogeneous resources in the future.

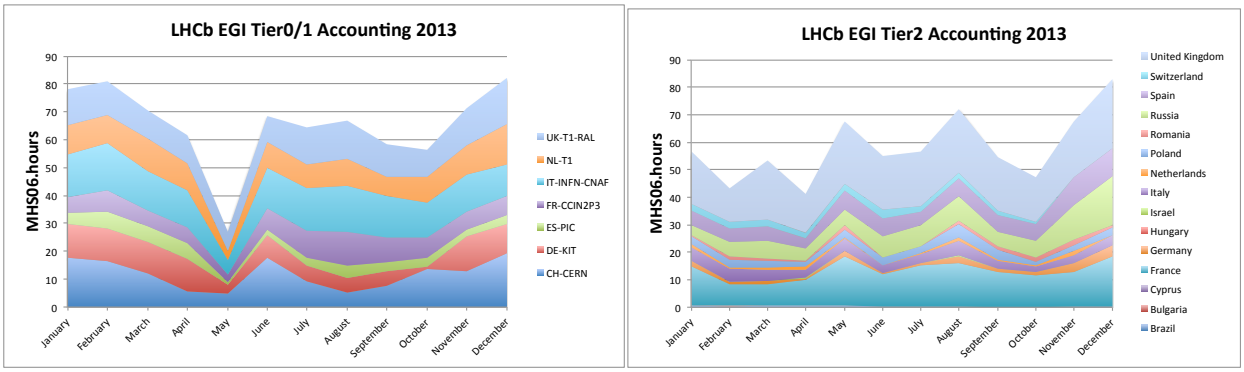

Figure-15a: LHCb Tier-1 (left) and Tier-2 (right) CPU Usage in 2013.

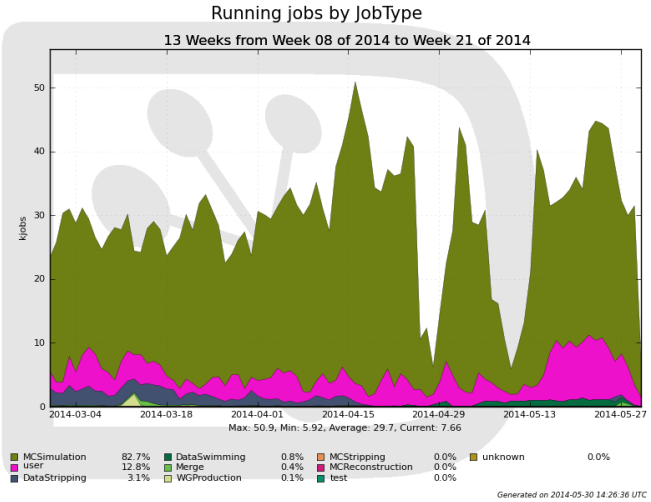

Figure-15b: Workload balance early in 2014

Other VOs

We are making steady progress with our other VO communities (i.e. those not one of ATLAS, CMS or LHCb). The fraction of GridPP Tier-1 resources used by "others" increased from 1-2% in the first quarters of 2013 to 3.5-4% in Q1 and Q2 2014. At Tier-2s the figure remains around 11% with biomed overtaking ALICE as the largest user in this category. In Q3 2014 ALICE formed 26%, T2K 17%, ILC 14%, biomed 13% and Pheno 10% of the 'other' CPU usage. In 2014 this has become biomed 40.5%, ALICE 21% and Pheno 18%.

Our SL6 migration went without major incident for these VOs, and there was no reoccurrence of the VOMS renewal issues that affected many VOs in early 2013 when user VO memberships expired with little warning. Problems that did occur during the last 10 months include SNO+ writing 1TB files to several sites (e.g. Sheffield) causing disk pools to fill and Nagios tests to fail. Biomed has been making strong demands on resources such that their job storms at time have saturated a couple of sites (the sites have since implemented maximum job quotas on the VO or in some cases banned individual users – particularly where the users also submitted multi-threaded jobs to single core resources).

The data management strategy for our smaller VOs is still a concern, especially as the LHC VOs move away from tools such as the Local File Catalogue (LFC) and WMS. To mitigate some risk in this area we have looked more closely at DIRAC. DIRAC as a service was in test early last year and has received positive feedback from several of our early adopter VOs but particularly Landslides, T2K and CERN@School. CERN@School actually moved from being 'new' to in 'production' in October 2013. Other GridPP hosted VOs have been enabled within DIRAC to provide more sysadmin experience with the system.

GridPP has continued to encourage the smaller VOs to move their software deployment mechanisms to use CVMFS. In recent months ILC moved to using CVMFS and the Tier-1 made a final decision to switch off its NFS software server in September 2014. Whilst good progress has been seen with software distribution, more work is needed to address increasing issues associated with use of shared storage resources whereby problems for other users of the same space (or other members of the VO) occur. They have little incentive to use spacetokens so better education may be out best option.

In recent months GridPP operations have sent out warnings to our smaller VO communities about a major change that may affect them. In October 2014 there is going to be a move away from lcg-utils and gfal towards newer gfal2 and gfal2-utils tools. These will require the VOs to actively adapt their frameworks.

There has not been a large change in the number of active or supported VOs over the reporting period. Oxford provide support to the European Centre for Medium-Range Weather Forecasts with some network performance tests, but they did not run work on the Grid. QMUL has been working to support HyperK.org and with storage now run on iRods at QMUL there is an open request for more sites to enable the VO. The LUX dark matter collaboration has recently been in contact with our members at Imperial.

MICE reprocessed their data in October 2013 and moved their RAW data to CASTOR. Work on their configuration database has been ongoing for many months. NA62 has made use of resources but their focus has been on software development for their Monte-Carlo data distribution (including use of DIRAC as a WMS) and work on asynchronous FTS transfers.

Impact and Dissemination

The GridPP Dissemination Officer, Neasan O'Neill, left the project at the end of July 2013 after several years to become Events Officer at EGI. A new Dissemination Officer, Tom Whyntie, was appointed at the end of October 2013. Tom has continued in his role as STFC Researcher in Residence at the Simon Langton Grammar School for Boys in Canterbury for the CERN@school project where he had previously worked actively with GridPP to conduct research based and educational activities. Tom is using the Grid for what it was designed for while promoting its existence and use to a new generation of scientists and engineers, and current researchers in the field. He has worked with various funding and industry bodies in order to gain support and foster strategic partnerships that have enabled the project to flourish. Tom's appointment is allowing us to leverage this work and fully integrate it with GridPP and will give us access to other school projects such as data from classroom based Timepix silicon pixel detectors and the space-based Langton Ultimate Cosmic ray Intensity Detector (LUCID), recently launched.

During this transition in Dissemination Officers, we took the opportunity to update the job description so that although dissemination is still important, the emphasis has changed more towards impact and two-way engagement with industry etc. Tom is very well suited to this changed role. Using newly developed tools developed for the Grid, CERN@school has successfully used GridPP resources to generate Monte Carlo simulations of the LUCID spacecraft detectors and we believe that this will make it much easier to attract SMEs and others to do likewise. GridPP is also developing Cloud interfaces to its resources and these coupled with increased virtualisation of services should also make it easier to attract new communities to our resources.

GridPP resources continue to be used by a number of diverse projects. The top users over the last year were Life Sciences, European Earth Sciences, NMR and Structural Biology, Nuclear Fusion and Veterinary Surveillance. Two examples of work done on GridPP resources and subsequently published in Nature are given below.

One example is the Tomato Genome Consortium (TGC) who had been working on unpicking the genome of the tomato since 2004. After 8 years of hard work, they were able to reveal in Nature that they have sequenced the full tomato genome. The work leads to the possibility of developing new breeds, with particular traits like drought resistance or improved taste, easily and quickly. As with most modern science a lot of the work to make this breakthrough was done on computers, usually on the researchers' desktop computers or on clusters of computers provided by their universities. But with a tomato having 35,000

genes, one of the TGC's partners the international Tomato Annotation Group (iTAG), needed a lot more computing power. Using GridPP resources, contributed by LondonGrid, they were able to use over 300 days of CPU time, giving them some of the final pieces needed to finish the jigsaw.

Another example is two biologists at QMUL who have been using GridPP computing resources to analyse the largest genetic sequence dataset ever to look for convergent evolution in 2,326 genes shared by 22 mammals, including six bats and the bottlenose dolphin. As reported in Nature on 4 September 2013, they concluded that bats and dolphins might have developed echolocation via similar mutations. It took their team roughly a month to perform its analyses using the computing cluster, rather than the years it would have taken on a desktop computer. The biologists ran a total of about 60,000 jobs using 180,556 CPU hours in the study. In fact this corresponds to only 1% of the CPU resources used by the ATLAS experiment on the QMUL cluster. GridPP has written a letter of support to enable them to obtain funding for further studies using our resources in future.

GridPP has been working with Alex Efimov, Director of Sensors and Instrumentation at the Electronics, Sensors, Photonics KTN. Alex obtained funding to help GridPP's impact agenda and wrote a 'GridPP Capability Guide' for potential industrial contacts. Alex initially introduced GridPP to a small SME interested in providing an analysis platform for a London based Bank. After some discussions this didn't progress as they were more interested in Hadoop based solution rather than a Grid solution. However, we are now involved another project involving scaling tests of a novel method of splitting files across disks and sites. A pilot study is taking place at Imperial, which if successful will be extended to QMUL and then to a larger number of GridPP sites. A number of other SME's have been approached by Tom Whyntie and discussions with some of them are ongoing.