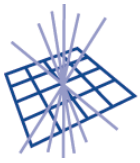


GridPP
UK Computing for Particle Physics

GridPP Project Management Board

Project Status

Document identifier:	**GridPP-PMB-176-ProjectStatus.doc**
Date:	**7/11/2016**
Version:	**1.0**
Document status:	**Final**
Author	**PMB**

Introduction

The Oversight Committee last reviewed the GridPP project in May 2016 at the end of the GridPP4+ phase of the project. The GridPP5 phase started just before that, in April 2016, and has now been underway for six months.

In 2016, the second year of Run-2 at the LHC has seen characterised by the stunning performance of the LHC machine, which has run almost at design energy and well above design luminosity. Even more significantly, the LHC has run with remarkable stability and efficiency and by the end of August 2016, more data had been accumulated than had been planned for the whole of 2016. Of course, the LHC experiments have also had to respond to the challenge and it is remarkable that they have been able to record so much data. At the time of writing, ATLAS and CMS have recorded close to 35 fb^{-1} of data in 2016 against the original goal of 25 fb^{-1}.

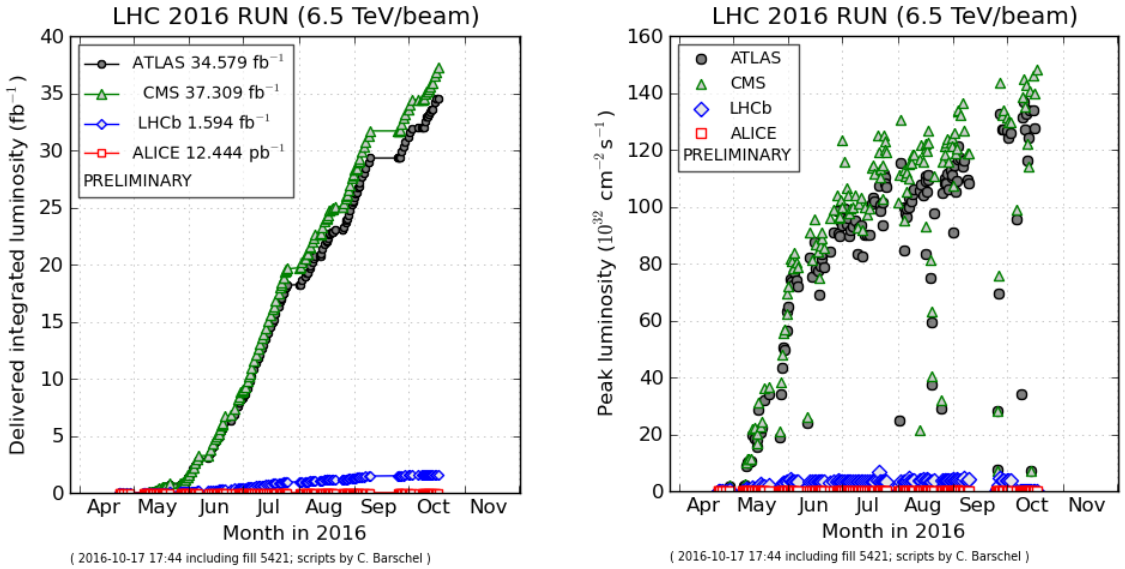


Figure-1: Left: Luminosity delivered to LHC experiments in 2016 (goal as 25 fb^{-1}); Right: Peak luminosity (design luminosity was 100 x 10^{32} cm^{-2} s^{-1}.)

The UK Grid infrastructure has continued to provide an excellent service as part of the Worldwide LHC computing Grid (WLCG) and the resources have been heavily used, though with slightly more headroom when the LHC was not running.

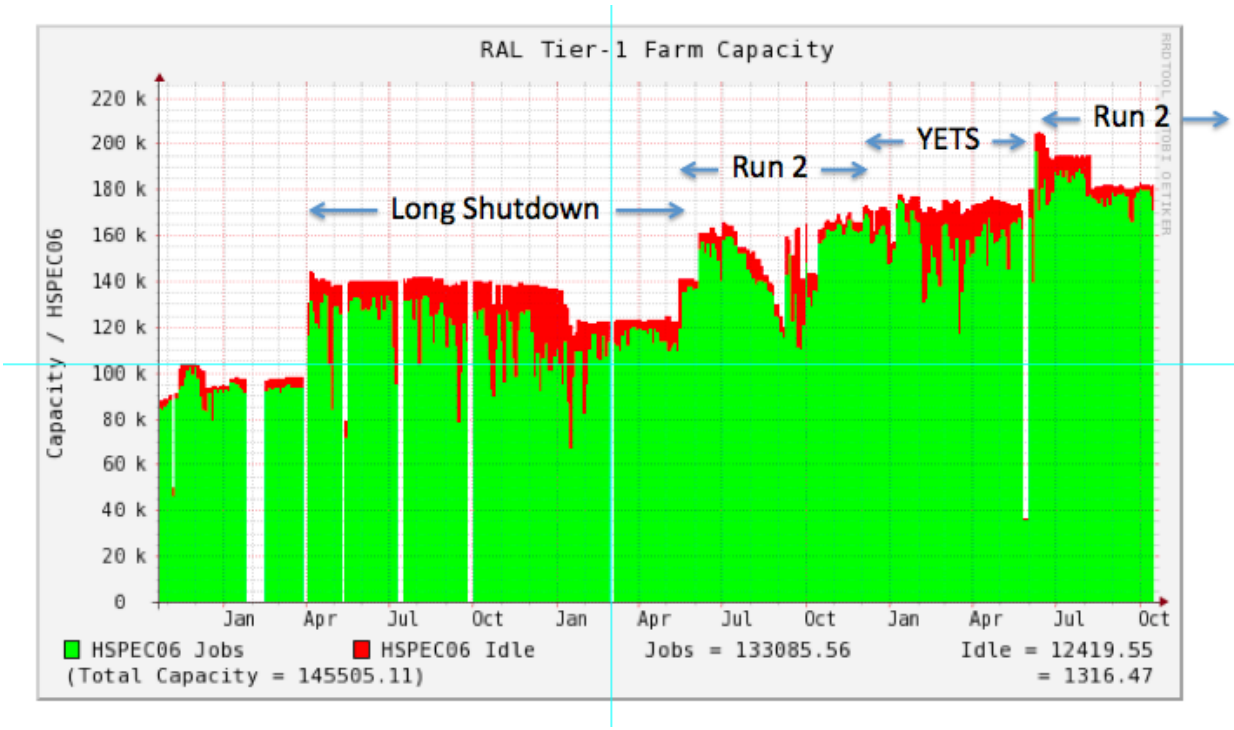


Figure-2: CPU usage at RAL from Nov 2013-2016. Green (Red) areas show used (idle) capacity.

The 36th and 37th GridPP collaboration meetings were held with themes of "New Directions in GridPP5" at Pitlochry, and "(R)evolution", in Ambleside.

Wider Context

WLCG resources have continued to be delivered according to the annual cycle wherein resources pledged at the end of August are deployed by the following April. However, due to the success of the LHC in delivering more beam and at high than the design intensity, and the success of the experiments in operating with great efficiency, the resource request for 2017 has been increased by around 20%. This is a significant perturbation on the levels on which the funding for GridPP5 was awarded and is compounded by a reduction in buying power due to the post-BREXIT drop in the value of the pound (currently 20%). Discussions have taken place with STFC about the appropriate UK response, given the flat cash funding profile. The UK position and plan on how to best respond to this request will be provided to WLCG at the October 24th Resource Review Board meeting by STFC and will be presented verbally at the GridPP Oversight Committee meeting. It is likely that GridPP will be able to re-profile funds to partly, but not fully, meet the new increased resource requests for 2017. It is clear that there will be significant problems in 2018 unless the exchange-rate recovers.

Overall, GridPP currently provides about 11% of the total global resources. Figure-3 below shows that the Tier-1 at RAL delivered 10% of the global LHC Tier-1 resources, as in previous periods. Figure-4 shows 13% of the LHC Tier-2 resources were delivered by the UK over the same period. The difference between the UK Tier-1 and Tier-2 global fractions reflects the significant leverage of Tier-2 resources from the institutes that GridPP funding has enabled.

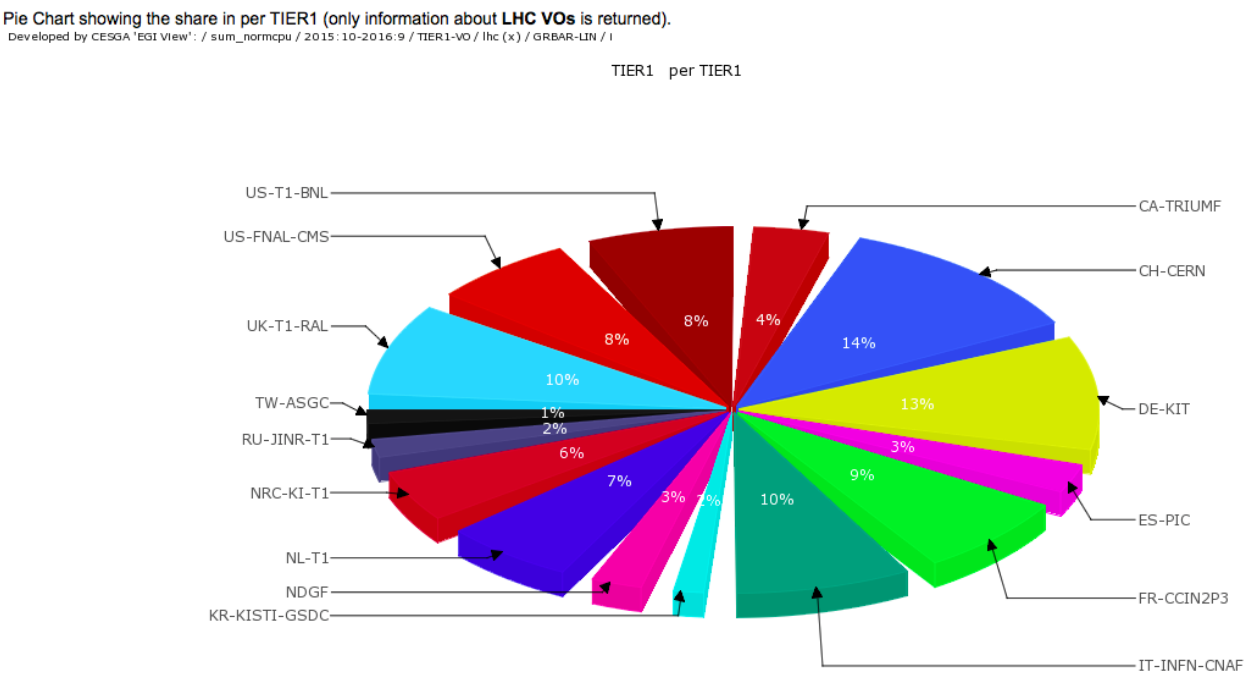


Figure-3: CPU delivered to LHC VOs (October 2015 to September 2016) by Tier-1.

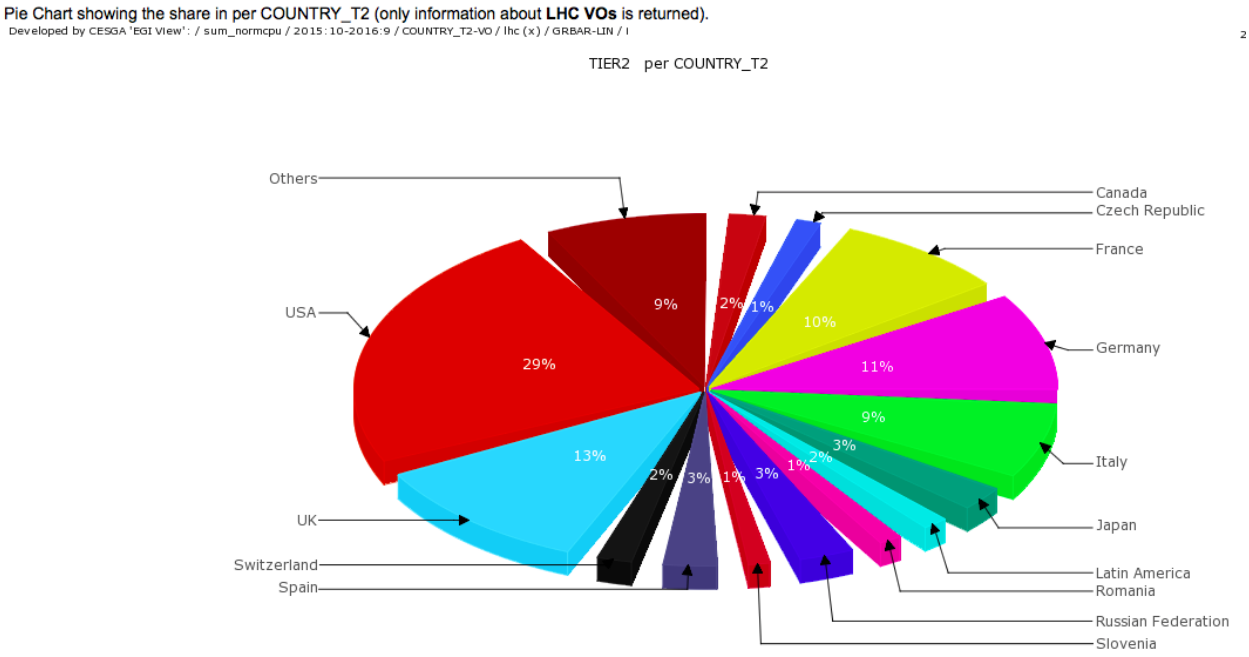


Figure-4: CPU delivered to LHC VOs (October 2015 to September 2016) by Tier-2s.

GridPP has continued to be active on several fronts in the efforts to consolidate e-Infrastructures and promote shared use across communities. Most recently, the University of Manchester hosted a productive two-day joint meeting between GridPP and SKA that should form the start of a long-term collaboration. Of particular note, the LOFAR collaboration, which involves some fraction of the UK SKA community, reported success in running “hello-world” jobs on the GridPP infrastructure. Over the next month or so, we hope to see this progress to real LOFAR work-loads.

The H2020 European Open Science Cloud (EOSC) project has been approved. The project is led by STFC-SCD and in addition to project management includes some modest funding for SCD to support a testbed.

The H2020 AENEAS project for SKA has been approved. This is to design a European Science Data Centre, and GridPP members are involved to share experience. STFC-SCD is involved with funding for 2 PM/year working on technical design and data movement tools.

Preparations, led by EGI, are underway to submit a response to the H2020 EINFRA-12 call, where again SCD in involved and is indicating interest to be involved in GOCDB, APEL, CVMFS software distribution, leading the security response team, and having responsibility for policy coordination and risk assessment. Overall there is a funding request of 46 PM/year to support STFC infrastructure delivering to LHC computing.

Within the STFC domain, members of GridPP and STFC SCD have been centrally involved in promoting the coordination of computing across PPAN science areas and the national facilities under the UKT0 banner. GridPP was able to underwrite the LSST project to commit resources to the international dark energy science consortium in a submission to the DoE. This was achieved using leveraged resources at the GridPP sites participating in LSST. The work to provide the case to BEIS for investment in e-Infrastructure has progressed well. A revised set of STFC budget lines was submitted along with a set of cases for computing across STFC, including HTC, HPC, Ada-Lovelace for Facilities, a data centre and networking. This was done at the request of BEIS and there is some chance that support may be announced in the autumn statement. If successful then UKT0 will be able to proceed and provide a set of shared resources.

The UK continues to receive support from EGI.eu and EGI Engage to provide services that are critical to WLCG (APEL accounting; Grid Operations Centre Database (GOCDB); and leadership of international security operations and policy development). The UK currently chairs the EGI Council (Dovey, JISC) and Executive Board.

GridPP is also engaged with and contributing to other H2020 initiatives. These include EUDAT, the Indigo Datacloud project and the AARC Authentication and Authorisation for Research project in collaboration with Geant and others. GridPP is also working with the HEP Software Foundation, (a collaboration of HEP software projects designed to promote standards, reduce overlaps and provide common utilities).

GridPP5 Status

This is the first report to the Oversight Committee on the progress of the GridPP5 project, which was funded from April 2016 to the end of March 2020. During the first two quarters of the project the production systems at both the Tier-1 and 2 sites have continued to function well and have contributed to the successful processing of Run-2 data. While the primary focus of the project is to continue to deliver to the LHC experiments, further work has taken place to support non-LHC experiments from other communities.

The Work Package structure was modified slightly for the start of GridPP5 into four sections. The quarterly reporting structure has been updated over the last quarter to reflect these changes. The overall tracking is recorded in the ProjectMap, which shows the on-going metrics and annual milestones.

The evolution of milestones and metrics since the start of GridPP5 is shown in Figure-5 and the current Project Map shown in Figure-6.

	Q116	Q216	Q316	Q416	Q117	Q217	Q317	Q417	Q118
Metric OK	168	140							
Metric close to target	9	8							
Metric not OK	0	0							
Not able to be measured	0	0							
Suspended	19	0							
Awaiting input	0	0							
Metrics Total	196	148	0	0	0	0	0	0	0
Milestone achieved	28	9							
Milestone close to completion	0	0							
Milestone overdue	0	0							
Milestone not due/ metric n/a	5	58							
Milestones total	33	67	0	0	0	0	0	0	0
Total	229	215	0	0	0	0	0	0	0

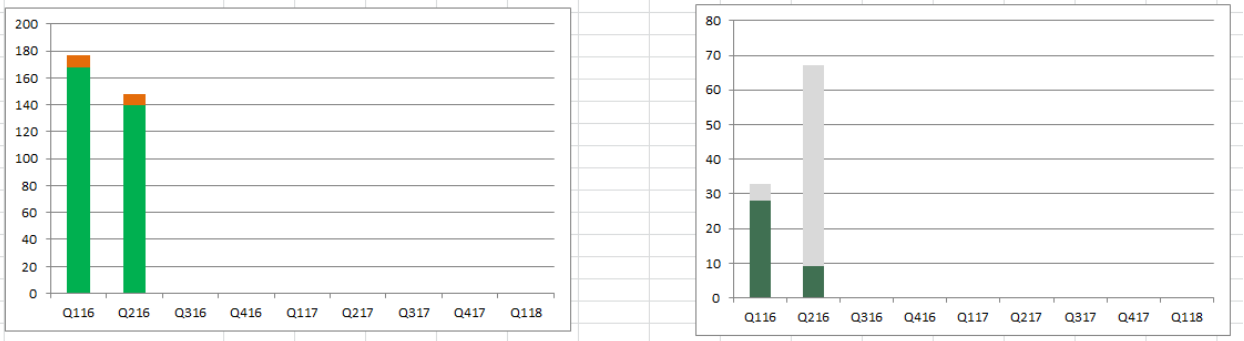


Figure-5: Evolution of Milestones and Metrics for GridPP5. The last quarter from GridPP4+ is included to show how the number of metrics and milestones has changed going into GridPP5. The lower left plot shows the number of metrics that are currently met (green) and not met (red). The lower right plot shows the evolution of the number of milestones that have been met. The large grey bar in Q216 shows how the total number of project milestones has increased at the start of GridPP5 (the green area represents those already completed). The data is presented numerically in the table at the top.

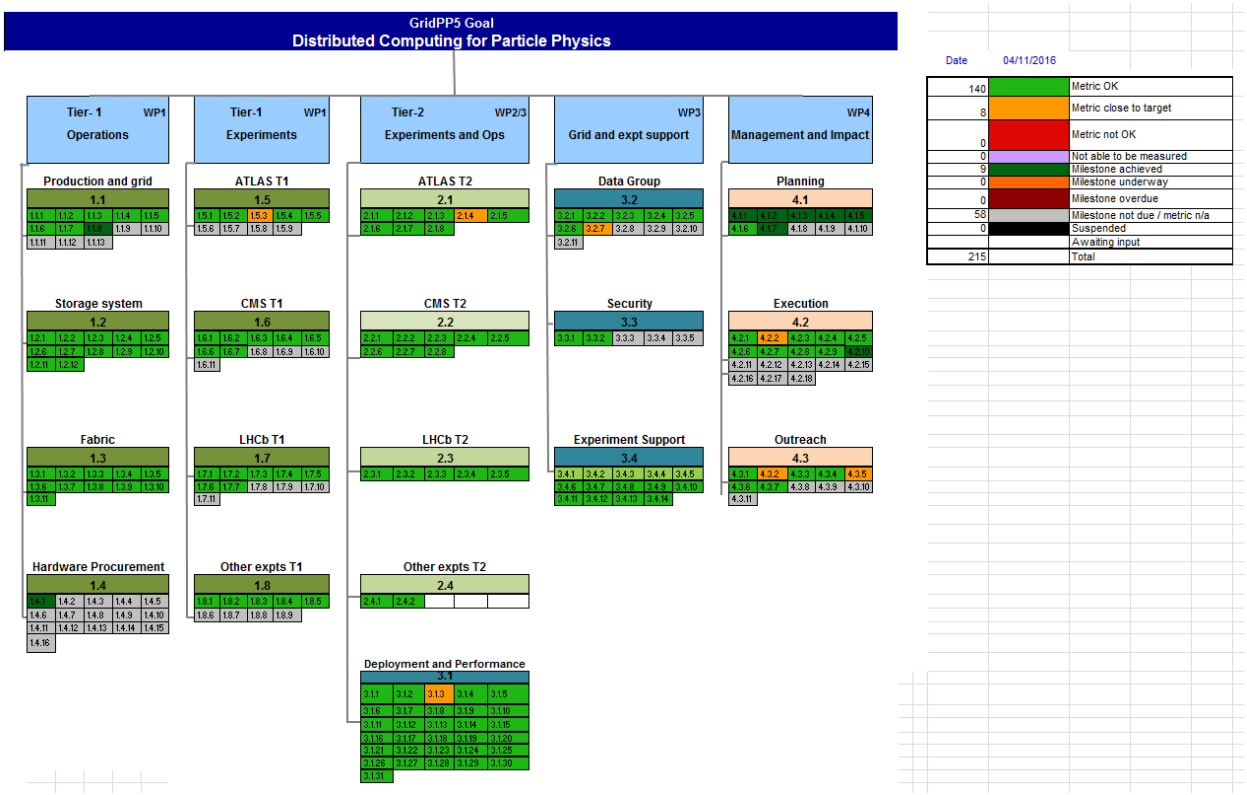


Figure-6: The GridPP Project Map at the end of Q2 2016

Overall the majority of metrics are being met or are close to target, with just eight metrics below target and marked amber this quarter.
They are:

- 1.5.3 'T1 Atlas Data availability' as measured by the ATLAS dashboard was 96% which is 2% below the target level.
- 2.1.4 'Tier-2 Data availability' as measured by the ATLAS dashboard was also down 2% below target at 93%.
- 3.1.3 ' Fraction of HEPSPEC06 used' This metric has been low for some time and is mainly as a result of multi-threaded jobs causing poorer utilisation at sites. This quarter we are higher than last quarter at 68%, but still low compared to our target of 80%.
- 3.2.7 'Number of Grid-Storage blog posts' was below target. There were 3 blog posts but the Purdah prior to the 'BREXIT' referendum prevented Tier-1 staff from publishing for some time.
- 4.2.2 '100% of quarterly reports submitted within 2 months': A couple of reports were late.
- 4.3.2 'Number of GridPP publications' was only one this quarter but is well above target at 12 for the last year.
- 4.3.5 'Number of press articles' was also only one this quarter but again over the last year we are above target.

At this early stage in the GridPP5 project only a few milestones have been completed. These are mainly to do with strategic planning and procurement planning.

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. The quarterly report format and structure has been reviewed and updated for the start of GridPP5.

All detailed quarterly reports are available on the web at:
https://www.gridpp.ac.uk/gridpp5-y1-quarterly-reports/

Risk register

The GridPP risk register was recently reviewed at the PMB Face to Face meeting following the start of GridPP5 and the current version summary is presented below in Figure-7 and the full version including the newer format suggested by STFC is available at:
https://www.gridpp.ac.uk/gridpp5-risk-register/

Ref.	Risk Description	Owner	Inherent Risk			Residual Risk		
			Likelihood	Impact	Total	Likelihood	Impact	Total
1	Castor Storage System Problems	AS	6	6	36	5	6	30
2	Tier-1 CEPH Project fails	AS	5	7	35	4	7	28
3	Outage of UK T1	AS	4	6	24	1	5	5
4	Failure of T1 to meet SLA or MoU	AS	6	5	30	2	5	10
5	Significant loss of custodial data at the T1	AS	1	7.5	7.5	1	5	5
6	Loss or damage to hardware at T1 >£2M	AS	0.5	10	5	0.5	10	5
7	Disaster at T1 leads to prolonged outage	AS	0.5	8	4	0.5	7	3.5
8	Recruitment retention problems at RAL	DK	9	6	54	7	5	35
9	Failure to deploy or operate hardware	AS,JC	5	6.5	32.5	5	6.5	32.5
10	Insufficient Network Bandwidth	PC	6	5	30	6	5	30
11	Over contention for resources	DB	6	5	30	6	5	30
12	Capital vs Resource at the Tier-1	DB	10	6	60	10	6	60
13	Technology Mismatch	DC	2	5	10	0.5	5	2.5
14	Loss of experienced personnel at T2s	SL	7	5	35	6	4.5	27
15	Insufficient funding at T2s for h/w	SL	2	4	8	2	3	6
16	Tier-2s are not fit for purpose	SL	4	6	24	1	6	6
17	Expt. s/w runs poorly on the grid	PG	5	6	30	2	5	10
18	Security problem affecting reputation	DK	7	8	56	5	5	25
19	Loss of GridPP service due to security	DK	3	5	15	2	5	10
20	Insufficient VO/user support effort	PG	7	5	35	6	5	30
21	Mismatch between budget and hardware costs	DB	9	7	63	8	7	56
22	Core service funding insuffcient	DB	5	4	20	4	2	8
23	Breakdown of NGI/EGI infrastructure	JC	3	3	9	3	3	9
24	Insufficient travel funds	DK	4	2.5	10	1.5	2.5	3.75
25	GridPP resources prove insufficient for actual requirements.	PG	8	4	32	8	4	32
26	Critical middleware no longer supported	DC	4	4	16	1	4	4
27	Unplanned infrastructure costs	PG	2.5	3	7.5	2	2	4
28	Loss of EGI.eu	DB	2	2	4	2	1.5	3
29	Financial Uncertainty	DB	5	6	30	5	6	30
30	Conflicting opinions amongst GridPP stakeholders	DB	3	5	15	3	5	15
31	Failure of achieving further intergration within PPAN community	JC	5	6	30	4.5	6	27

Figure-7: The GridPP5 Risk Register Summary

There are currently two risks marked red:

Risk-12: Difficulty with budgets due to Capital vs Resource limitations
This risk is marked as high as dealing the shortage of resource available compared to our expected requirements is causing difficulty.

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. FY16 levels have only recently been confirmed.

Risk-21: Mismatch between budget and hardware costs
BREXIT and the subsequent exchange rate shock has had a major effect on the pricing of hardware and hence significantly increased the levels of this risk.

The current risks that are somewhat elevated (orange in Figure-6 above) are discussed below:

Risk-1: Castor Storage system problems.
As the data rates continue to increase the likelihood that Castor fails to perform adequately is getting higher. There have been limits placed on some CMS jobs to alleviate the situation. The plan is to migrate as to the ECHO (CEPH based) storage system, however there is a residual risk of difficulties before the migration can be completed.

Risk-2: Tier-1 CEPH project fails
If the CEPH project (to replace CASTOR for disk storage) fails during scale testing, we would have to fall back to CASTOR, which we would not expect to cope well with the increased load. Once CEPH is put into production, new problems may arise leading to an unacceptable degradation of performance or data loss. Progress is going well, and initial testing is under way and close project management by the Tier-1 team is in place to try to prevent the risks escalating.

Risk-8: 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 of a critical loss. It has been noted that recruiting new staff is increasingly difficult due to non-competitive salaries and conditions. Retention has been affected by the earlier uncertainties in GridPP5 funding which caused some staff to seek alternative employment.

Risk-9: Failure to deploy or operate hardware
There is an increased risk that purchasing could be delayed due to changes in procurement planning caused by the increased LHC requests. Similarly the timing of the issue of Tier-2 grants may be non-optimal for gaining best value for money.

Risk-10: Insufficient Network Bandwidth
Increased usage of Tier-2's and recent saturation of Tier-1 links, show this is now a real problem. The Tier-1 has upgraded some links but that may just 'open the flood gates' showing up further issues. We are seeking costings for upgrading the OPN. Fortunately, the cost of the OPN link has fallen so the risk is manageable.

Risk-11: Over contention of resources
The likelihood has increased as the LHC experiments have announced increased resource requests and the number of new groups supported increases. Success in the UK-T0 context will mean an increased pressure on our resources.

Risk-14: Loss of experienced staff at Tier-2s.
The prolonged uncertainty of funding for GridPP5 caused some staff to seek alternative employment. Further losses could adversely affect the operation of the Tier-2s. New working models are being developed to help mitigate this by providing a lower maintenance mode of

working at sites with little manpower. However this places more demands on the remaining larger sites.

Risk-20: Insufficient VO/user support effort
We have dedicated support posts for the major LHC VOs, but the increased number of new users and groups is stretching the support teams. A dedicated GridPP-support email list has been setup to help with this. The new user groups have each been assigned a 'Champion' who takes responsibility to ensure they are progressing well. The status is reported weekly at the GridPP Operations meeting.

Risk-25: GridPP resources prove insufficient for actual requirements.
LHC requirements have increased significantly above expectations.

Risk-29: Loss of EGI.eu
There is an elevated risk here post BREXIT due to the uncertainty of the future despite some assurance from the UK Government of underwriting of European project funding.

Risk-31: Failure of achieving further integration within PPAN community.
If GridPP fails to work with and engage others within the PPAN community GridPP would not benefit from potential shared infrastructure investment and cost savings. There could be reputational damage, and lack of endorsement that the GridPP infrastructure service is relevant to a wider community. Effort is being made to bring on new users but staff resources are being stretched.

Tier-1 Status

Overview

This section of the report covers the GridPP5 period from April 2016 to September 2016.

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

- Ongoing routine stable operation of the production service supporting the second year of LHC Run 2.
- Deployment of hardware to meet the 2016 MoU commitments
- Addressing capacity and capability limitations in the Tier-1 network
- Final stages of building a production ready release of ECHO the new Object based storage system for the Tier-1

Fabric and Infrastructure

FY15 procurements were deployed into operation. The disk procurement deployed directly into the new ECHO object storage system (based on the CEPH object store). One tranche of CPU deployed routinely into production enabling the Tier-1 to meet its MoU commitments. The second CPU

deployment was more problematic requiring a number of issues to be addressed by the supplier; this equipment became operational in September.

Plans for the FY16 capacity procurements have been finalised. The original GridPP5 plan was modified in order to deliver additional capacity in 2017 in order to partly meet the increased capacity requirements of the LHC experiments expected in 2017. These late changes to the plan have increased the risk of late delivery of hardware as both the procurement mechanism and approval process will need to change in response to the increased volumes. These risks were discussed with the STFC program manager but on balance were considered to be outweighed by the benefit of delivering increased capacity for the LHC.

Reliability problems with one older generation of disk server hardware which were reported in the previous Oversight Committee report were finally resolved (by firmware changes and drive replacement of a subset of disks). All hardware generations are currently expected to reach their standard (4 year) planned phase-out date. This was nevertheless a useful reminder that appropriate reserve capacity remains essential in order to mitigate hardware reliability risks.

The tape service is currently operating on a mix of T10KC and T10KD generations of tape drive, the C drive writing media at 5TB capacity and the D using the same media at the higher 8.5TB Migration of 6PB of ATLAS data from T10KC to T10KD (in order to gain media capacity) was a continuous activity over most of the period, completing successfully in August. An additional 5 T10KD tape drives have been procured this financial year and these are now entering production. With the additional drive capacity in place, migration of 5PB of LHCb data to T10KD can now commence.

In the longer term we expect to procure T10KE drives in 2017 and recently received updates to the technology roadmap from Oracle indicates that we should be able to achieve the GridPP5 capacity plan within our projected budget and with considerably reduced media costs (which we will trade off against capital investment in tape drives. Total tape media costs over the period of GridPP5 are now estimated to be only £285K (all needed in FY17). According to current STFC finance practice tape media is classed as resource. We intend to make a case to STFC's finance team to classify tape media purchases as capital. There are, however, risks associated with this plan, in particularly media purchases already made in previous years could also be adjusted (approximately £1M of media purchases since 2010).

DiRAC media requirements (5PB) were not part of the GridPP5 proposal's capacity plan. Up until summer 2017 it will be possible to meet DiRAC's needs from unused tape released during the current round of data compression. However in 2017 new media will need to be procured to accommodate DiRAC at an estimated cost of £75K this is not yet included in any budget plan and without an additional budget increase DiRAC cannot be accommodated in the GridPP tape roadmap from 2018.

Following the start of the year's running in May, network load on the Tier-1's external links grew substantially beyond that experienced in the previous year(s). In particular the 10Gb/s OPN link from CERN to RAL began to saturate regularly and individual file transfer rates began to suffer substantially. Following discussion with CERN the unused backup link was brought into play as a second primary link and traffic load balanced over both links, much improving the situation. However, more recently the 20Gb/s pair of links has begun to saturate for short periods of time. JISC have provided prices to upgrade one or both of the CERN links to 20GB/s and options are currently being evaluated to increase the capacity to 30Gb/s. The two CERN links are physically diversely routed and in the event of a link failure the service will continue – albeit at reduced capacity.

STFC recently re-engineered the Tier-1's network bypass route to JANET. This was necessary in order to allow IPV6 packets to reach the Tier-1's storage nodes. This change also upgraded the Tier-1's link from JANET from 10Gb/s to 40Gb/s (the full capacity of site peering with JANET). While the Tier-1's 10Gb/s bypass link had rarely been fully loaded, the upgrade had unexpected and dramatic impact on the traffic throughput to JANET. Traffic flows which rarely reached 10Gb/s increased substantially both on average and peaked recently at 30Gb/s. The cause of this increase is not well understood but indicates some invisible constraint at the bypass router which the replacement has addressed. STFC site networking team are now monitoring Tier-1 traffic carefully in order to ensure it does not impact negatively on other Lab activities. This matter will no doubt be discussed at the bi-annual site network oversight meeting (TDA) scheduled for November.

wLCG has set a deadline of April 2017 by which time all Tier-1 sites must offer some limited storage services by IPV6. Only limited resources have been available within STFC's Site Network Team to support the rollout of IPV6 and this in turn has delayed Tier-1 deployment. Some progress has recently been made in re-engineering the network path to the storage systems. The site networking team has scheduled work for Q1 2017 to further progress IPV6 at RAL. While the Tier-1 expects to be able to meet the April deadline, it will only be able to meet the absolute minimum specification for a prototype IPV6 service. Recently the member of staff within the Tier-1 who was working on IPV6 left STFC and RAL is likely to remain lagging behind most other Tier-1s.

There was a machine room incident in May when the main R89 chiller system shutdown following a problem with the building management system (BMS). Rapid response by the machine room operators and site estates team managed to restart the chiller system within the 30 minutes necessary before an emergency shutdown was required.

Production, Operations and Service

Tier-1 operations were generally very reliable over the reporting period. Average RAL (ops test) service availability remained high at 99.5% (similar to the previous reporting period) compared to a WLCG target of 97%. Average RAL availability for 2015 for the LHC VOs was 99.6%, improving on the 98.8% from to the previous 12 months covered in the previous reports. Reliability broken down by VO was follows:

	2Q16	3Q16
ALICE	100%	100%
ATLAS	100%	100%
CMS	99%	99%
LHCb	99%	100%

Figure-7: Quarterly VO availability at Tier-1

There was one significant operational incident during the period, which was hardware and software problems affecting the tape robot. Tape robot availability during 16Q2 was, exceptionally, only 92% against an average of 99.9% over the whole period of GridPP4 and GridPP4+. While the tape robot was unavailable data acceptance from CERN could generally continue to disk buffer but data reading from tape was either degraded or impossible. The incident was multi-faceted and involved coupled

hardware and software problems. The problem was escalated to Severity-1 inside Oracle's management system. A number of hardware replacements were carried out and a software problem identified and firmware fixes eventually applied.

Middleware, Grid and Cloud Computing

CPU utilisation was high through the period. Farm occupancy averaged 91% (compared to 89% in 2015) and job CPU utilisation efficiency improved to 88% (increased from 84% in 2015. Over the period RAL delivered 9.92% of LHC Tier-1 delivered CPU cycles, exceeded only by CERN and KIT.

Load related issues have become increasingly common on the CASTOR storage system and have been an increasing draw on CASTOR team effort. This has been an identified risk for several years but is now becoming a reality in practice. Performance degradation can be caused by several factors including transaction rates at the Oracle catalogue and low disk spindle count in smaller storage pools. New database hardware for the primary CASTOR Oracle database was purchased early in the reporting year and when deployed may reduce issues. Responding to incidents has been an increasing drain on the limited effort available; team headcount for CASTOR has been reduced, partly in response to GridPP5 staff reductions and partly in order to release effort for CEPH. The planned upgrade to CASTOR 2.1.15 has been later than desired as effort was diverted to address performance problems. The upgrade is now ready for deployment early in the New Year.

Final disk hardware deployment has been completed into the new ECHO storage service (which utilises CEPH as the object store) planned to replace the Tier-1's CASTOR disk system. Additional software development was necessary to achieve full performance of the GridFTP and xrootd gateways, which has now been completed. Functional tests were carried out successfully with both ATLAS and CMS and ECHO is now included in the experiment's routine functional tests. Large scale load testing is underway and results from initial external wide area network load tests provide assurance that we can meet external data rate commitments. A number of technical/operational issues still need to be addressed before load tests are ramped up further in order to demonstrate CEPH can meet internal batch farm I/O requirements. Both ATLAS and CMS now have read-only copies of some of their production data on ECHO which is already used as the priority copy ahead of CASTOR.

Effectively CEPH is now classified by the experiments as a "production Tier-3 endpoint". Gradually, provided all goes well, further datasets will be moved to CEPH and its status will be moved to a full production Tier-1 service. While the project continues to be a high risk for the Tier-1 there is gradually increasing assurance that CEPH will meet the Tier-1's operational requirements. CEPH must be fully operational and reliable in time to meet a part of the Tier-1's 2017 MoU commitment from April 2017. Once CEPH is fully operational The Tier-1 will consolidate the four CASTOR instances (ATLAS, CMS, LHCb and GEN) to provide a simplified CASTOR infrastructure for a tape only service. This is expected to reduce operational costs (hardware and staff) in GridPP5 year 3 onward when staff effort will be reduced yet further

CERN have now begun planning and early development of a new tape storage system (CTA) to replace CASTOR at CERN in the early part of the next decade. At RAL we expect CASTOR to be phased out for disk based storage once ECHO enters production. At present CASTOR remains the Tier-1's tape service, and will continue to be so throughout GridPP5. A meeting with CERN is planned for May 2017 to discuss CERN's long term plans for tape storage. In the event that CERN formally announce a plan to phase out support for CASTOR, then Scientific Computing Department will begin the process of evaluating a potential replacement for CASTOR that meets the needs of all the major tape user communities at RAL. Given the planned reduction in staff effort at the Tier-1 in

the second half of GridPP5, only very limited effort will be available to support testing and deployment of a new tape system for the Tier-1.

Management, Business Processes and Communications

Tier-1 staff levels were reduced to 17.5 FTE for the first two years of GridPP5 (down from 19.5 in GridPP4). Team funding lines have been adjusted to match the GridPP5 plan. During the reporting period bookings averaged 17.1 FTE – 98% of plan (slightly reduced from plan owing to unplanned staff departures). The Tier-1 effort will be reduced to 14.5 FTE in FY18 and at that planned level it is unlikely that the service will be able to continue to evolve in response to the changing external environment in the way that it has in the past.

Scientific Computing Department (SCD) has structurally re-organised its Systems Division. One driver for this re-organisation was to consolidate the Tier-1 team with the internal STFC services to ISIS and Diamond. This change will deliver some efficiency gains over the next couple of years as Tier-1 infrastructure is consolidated with that of the STFC facilities. While this may modestly reduce operational costs, it will however lead to necessary compromises, both in terms of technical choices and also in terms of staff priorities as the tightly focussed Tier-1 team increasingly shares effort with other projects. One early change planned is a study to investigate the feasibility of consolidating the Tier-1's dedicated Hyper-V virtualisation platform with SCD's VMware service. This change will tension the benefits of platform convergence (and consequently a gain in agility and a reduction in staff effort required) against increased licence costs.

Following on from staff departures in the last reporting period, further staff left during this period. In particular the database team remains under strength and relying on contractor effort. More recently, the production team 2.8 FTE (which handles service exceptions and overseas service delivery) has suffered particularly badly, and team effort has been affected by a staff departure, sickness and the production manager (Gareth Smith) moving to reduced working hours.

Recruitment and retention continues to be is identified as a high corporate STFC risk owing to deteriorating pay and conditions for staff. STFC is attempting to mitigate this situation through a number of initiatives. For the Tier-1 there has been increasing emphasis on making use of apprentices and other early career recruitment mechanisms. However the service depends increasingly on a number of long standing core staff and there is a risk that expertise established at the start of the LHC runs at the start of the decade will become increasingly dilute as GridPP5 continues.

Recruitment of a new Tier-1 manager will commence after Christmas. Currently the role is filled partly through the former manager working 30% and partly by re-tasking some production manager effort to support service management.

STFC agreed to join and contribute a total 75K euro of GridPP funding of the HNSciCloud project (this was agreed with STFC programs directorate at the request of John Womersly). HNSciCloud is a project to carry out a pre-commercial procurement (PCP) of hybrid cloud infrastructure. Involvement in the project will enable GridPP to evaluate commercial cloud solutions and pricing in order to assist planning for the next decade. The first phase of this procurement has now completed and GridPP will contribute 30K euro from the Tier-1 budget in FY16 with the 45K euro balance falling due in FY17. As these contributions are classed as resource this is an additional pressure on the Tier-1 resource budget when the financial situation is getting increasingly challenging.

The SKA AENEAS proposal submitted to the Horizon 2020 program has been approved. AENEAS is a project to develop a design for European Science Data Centre for SKA. If funded, AENEAS will provide 6PM of effort to STFC to work on Computing Requirements for SKA. Given the close alignment between SKA and LHC Run-3 requirements the intention is to part-fund the Tier-1 manager role in order to work on technology foresight and system architectures. This is a very positive outcome and will bring benefits to both GridPP and SKA in the future.

Deployment Status

This operations update covers site progress, middleware and service evolution, developments towards easier running of sites, together with general news in this area for the period April to October 2016

Overall site performance has remained good. Aggregated resources available from GridPP sites have remained steady at around 59,000 HS06 and 41,000 TB of disk. Overall CPU resource utilization improved during Q2 and was over 70% in July - though some sites were challenged by demanding I/O needs of certain ATLAS workloads. GridPP received notification in June that EFDA-JET, a small site that has been part of SouthGrid since about 2005 was going to start the decommissioning process and close due to ageing hardware and reduced work being undertaken under the fusion VO.

Two relatively minor site availability incidents took place during the period. Glasgow suffered from a machine room power cut at the beginning of August, and Oxford was down for a few days due to an air-conditioning failure (on 12th).

In relation to wider services there are a couple of notable events. In early October the APEL accounting service experienced problems caused by sites globally republishing large numbers of jobs. This situation can arise when sites make a conscious decision to (re-)publish data due to a gap, or fix a problem that existed for some time (e.g. an out of date certificate, or parser cron not running). This event led to a request across WLCG/EGI for sites to schedule major republishing events rather than undertake them in an adhoc fashion (to reduce overlaps of high activity).

During this period the RAL T1 encountered high usage of its OPN connection, with the inbound link saturated most of the time. In August the T1 also experienced intermittent periods of high packet loss within the Tier1 network that was later traced to a problematic switch. There have not been any other significant networking issues, but the core operations team continues to chase issues picked up by our perfSONAR instances.

There was a problem with the Greek-run dteam VOMS server in August. A large number of site administrators appeared to be suspended from this core operations VO following an upgrade of the service during which a “malfunction” occurred; this malfunction propagated false information about membership and led to a lot of confusion over 2 days.

Finally on the operations front, the UK eScience CA encountered a certificate issuance problem in September following a partial but significant database corruption on the signing system for the CA. Data was restored from (offline) backups, but the rebuild was not correctly configured leading to an extension of the service downtime.

In terms of middleware there are two key items to mention – one operational and the other developmental. The first relates to an EGI target of 30th April 2016 for SL5 decommissioning. Within GridPP we took this as a soft target. To smooth the transition we allowed several sites to run some background services beyond this deadline such that in mid-May we had an SL5 WMS/LB in the process of being decommissioned at Glasgow and several at RAL, a BDII being upgraded at Lancaster, several sites running backend disk services on SL5 (with the nodes being retired together with old hardware) and a SAM/VO Nagios service undergoing retirement at Oxford and Lancaster. The GridPP Nagios at Oxford and Lancaster were both decommissioned in July. As of October, the only remaining SL5 nodes are under the RAL T1 Castor SRM systems. Upgrading these nodes has been waiting for a validated CASTOR upgrade under SL6 to become available.

As an aside it should be mentioned that the VO Nagios service mentioned above presents some concern going forward as there is no SL6 VO Nagios service, and no such service under the new (centralised) monitoring service run by EGI (called ARGO that went live on 1st July) . This means that the automated checking that jobs can run at sites under specific VO credentials may not be possible going forward.

There has been good progress in this period towards validating components of WLCG middleware with the next operating system that will be deployed across the infrastructure in the coming years, namely CentOS7. The LHC experiments presented their validation plans at an operations coordination meeting in early September, and site plans are currently being collected via a questionnaire.

GridPP sites have been leading this work in several areas, for example Brunel undertook to confirm that the site-BDII was fully compatible with CentOS7 and ECDF has recently become one of the first WLCG sites to run exclusively on SL7. This leading position is much appreciated by the community, but it has come with various issues, such as ECDF being labelled as failing SAM probe availability tests for LHCb (e.g. due to problems with the "org.lhcb.WN-sft-lcg-rm-gfal" probe) when in fact the site was functioning perfectly.

A final area related to middleware relevant to this report is that there has been a renewed look at HTCondor Machine/Job Features testing. This fits well with the VAC work being led by Manchester and their team now also lead the WLCG working group in this area.

Within WLCG operations, the last 6 months has seen decisions being made and validated in relation to the grid information system. In particular plans to move away from use of the BDII in favour of static information being placed in the GOCDB. Our sysadmin community are broadly behind these attempts to simplify the information system, but they do present some risks for our wider VO support if WLCG and EGI diverge.

Central to the ambitions of GridPP5 was the exploration and adoption of improved and less manpower intensive approaches for running Tier-2 sites. Our work to-date and insights have been very well received within WLCG which since May has also begun exploring this theme of lightweight sites, first with a GDB overview and operations coordination team focus meeting, and more recently in October with a pre-CHEP WLCG workshop session in San Francisco. Other areas where GridPP has made significant recent contributions include in a September discussion of Data Management approaches (in particular resource reporting and ways to tackle the ‘protocol zoo’) and a June workshop on IPv6 deployment. Manchester has also led part of the work of a Management Board approved WLCG Accounting Task Force to validate accounting data to aid the deployment of a new accounting portal. GridPP accounting switched to use the 'new' EGI accounting portal in mid-June.

The two quarters have seen good progress in the area of operations related to Tier-2 evolution. Since Q3 2015, this activity has formed a standing item for discussion at the weekly GridPP operations meeting and continues to report at a biweekly 'GridPP technical meeting'. A GridPP operations project was setup in the CERN JIRA service to track progress, particularly in relation to Vac (a simplified approach to VM management developed within GridPP at Manchester).

The UCL and Oxford Tier-2 sites have continued to serve as testbeds for the Vac-in-a-Box installation procedure, dramatically reducing the local effort required to maintain a WLCG site, running ATLAS, LHCb, and GridPP DIRAC workloads. Liverpool volunteered to convert a significant fraction of their resources to Vac during GridPP4+ and has demonstrated operation of such a site during Q2/Q3, with invaluable feedback to the development project at Manchester.

GridPP's complementary system for managing VMs on Cloud systems, Vcycle, has been maintained in parallel with Vac, and is in production at CERN managing the LHCb OpenStack resources at CERN and CC-IN2P3. Together with the UK Vac resources, these systems now manage as much virtualised CPU power for LHCb as one of the LHCb conventional Tier-1s (PIC).

The GridPP Regional Operations Duty team have continued to track and address UK operational issues, albeit with the team size reducing by one in July. (The RAL T1 is looking at options to cover this role again in the near future). The team's work was temporarily hampered in July when a new version of the operations portal (called VAPOR – this latest release including integration of gstat features) experienced teething problems and the ROD dashboard become slow as a result.

GridPP has long been a leader in WLCG in the area of security and this has continued in Q2 and Q3 2016. A security training course was arranged for sysadmins in association with the June HEPSYSMAN meeting, and several GridPP staff have continued in their leading roles in WLCG Security Operations Coordination Centre (SOC) work (for which there was a pre-GDB workshop in July) as well as being major contributors to a newly formed working group on Traceability and Isolation. In the current reporting period there have been several vulnerability advisories in all categories (low, medium and high/critical). For example one recent critical vulnerability related to the use of iperf3 in perfSONAR - a core component of WLCG network monitoring. GridPP sites have responded well to all advisories and we have only seen one incident in this period (wherein a site acted as an amplification reflector for a DDOS attack abusing open port 111 (Portmapper) responding on UDP).

Users' Reports

ATLAS

Since the last OSC, ATLAS has collected and processed record volumes of data reflecting the unexpectedly high (80%) live time for the accelerator and the new trigger rated. It is continuing its work on moving to greater use of data federation and its event service, allowing more work to be moved to lower tiers. It has made extensive use of Tier-2 resources for simulation, group analysis and individual user analysis, and increasingly processing at the larger Tier 2 sites. The main Tier 2 processing activity has been of group xAOD selection and production, a process that has reduced the demand on storage to compensate for the increased number of events. The Tier-1 resources are

primarily used for reprocessing, simulation and data curation. ATLAS is well advanced with its event service model, allowing sites to access data across the network on an event rather than file level. This aids the reduction in endpoints for data transfer, and is accelerating the shift to large Tier-2s with significant data storage and smaller Tier-2s providing CPU and disk cache. In effect sites with less than 1PB are transitioning to effectively be providing a rolling cache.

The UK continues to perform well, in the period April-September delivering roughly 11.9% of the global joint Tier-1 and Tier-2 capacity by CPU consumption (reflecting the UK authorship share); the share of the Tier 2 provided by the UK is 14.26%, reflecting additional resources provided by the universities. This 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.

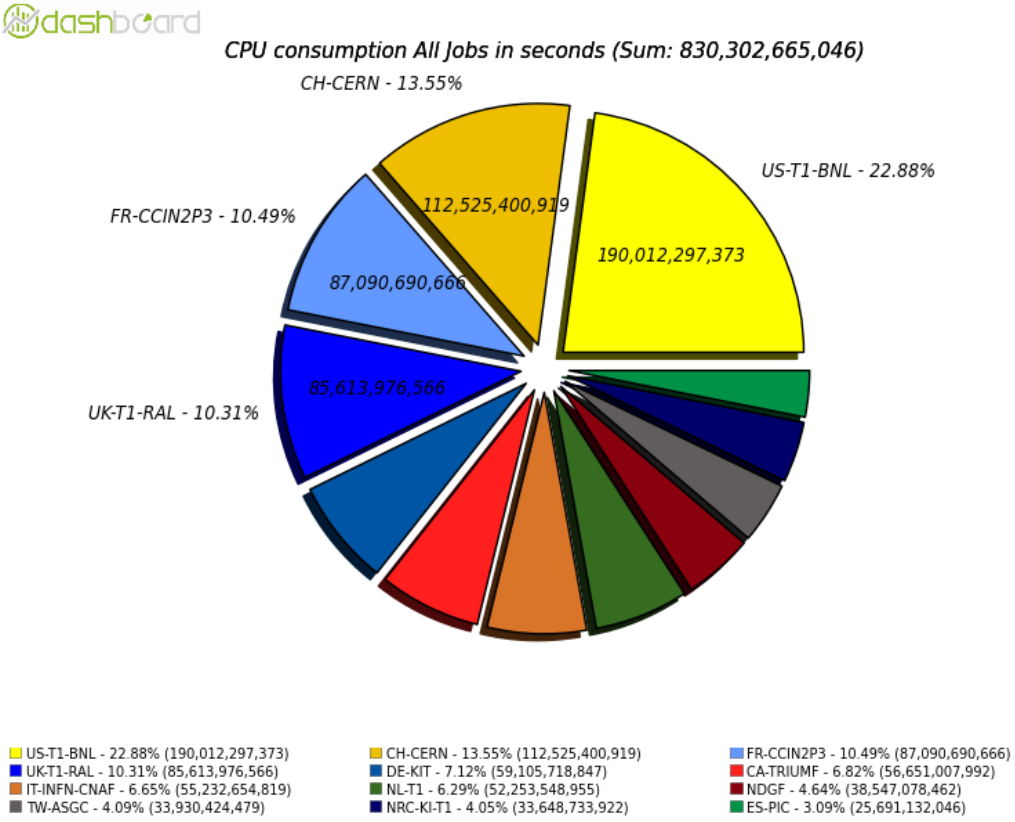


Figure-9 Sharing of ATLAS Tier-1 CPU consumption 1 April 2016- 30 Sept 2016.

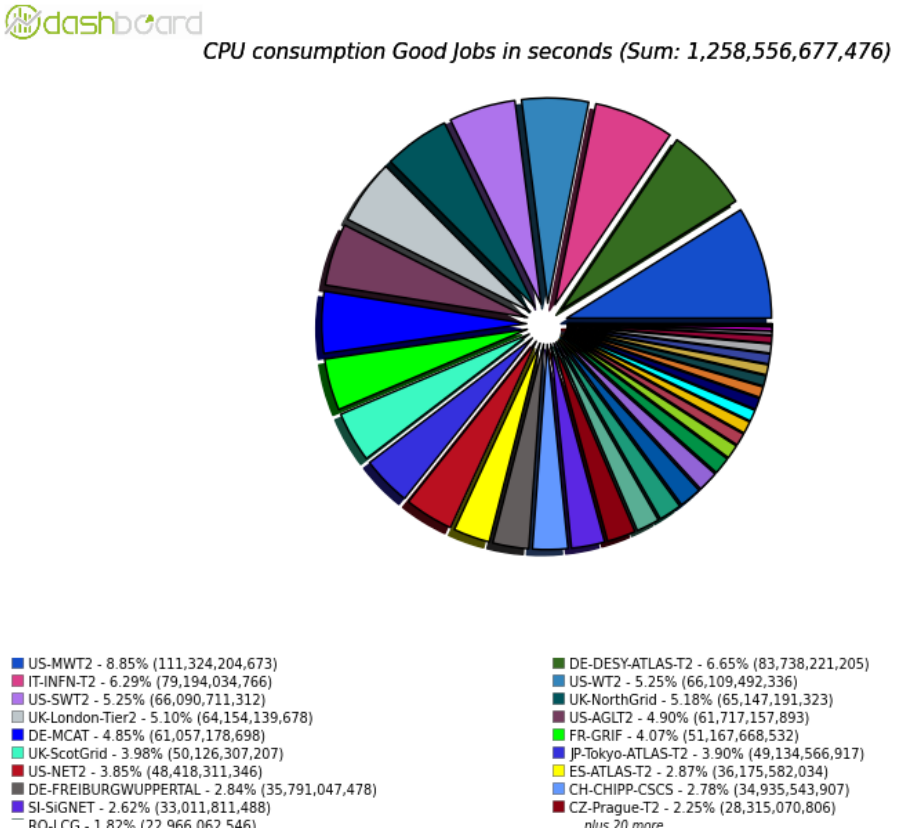


Figure-10 Sharing of ATLAS Tier-2 CPU consumption 1 April 2016- 30 Sept 2016.

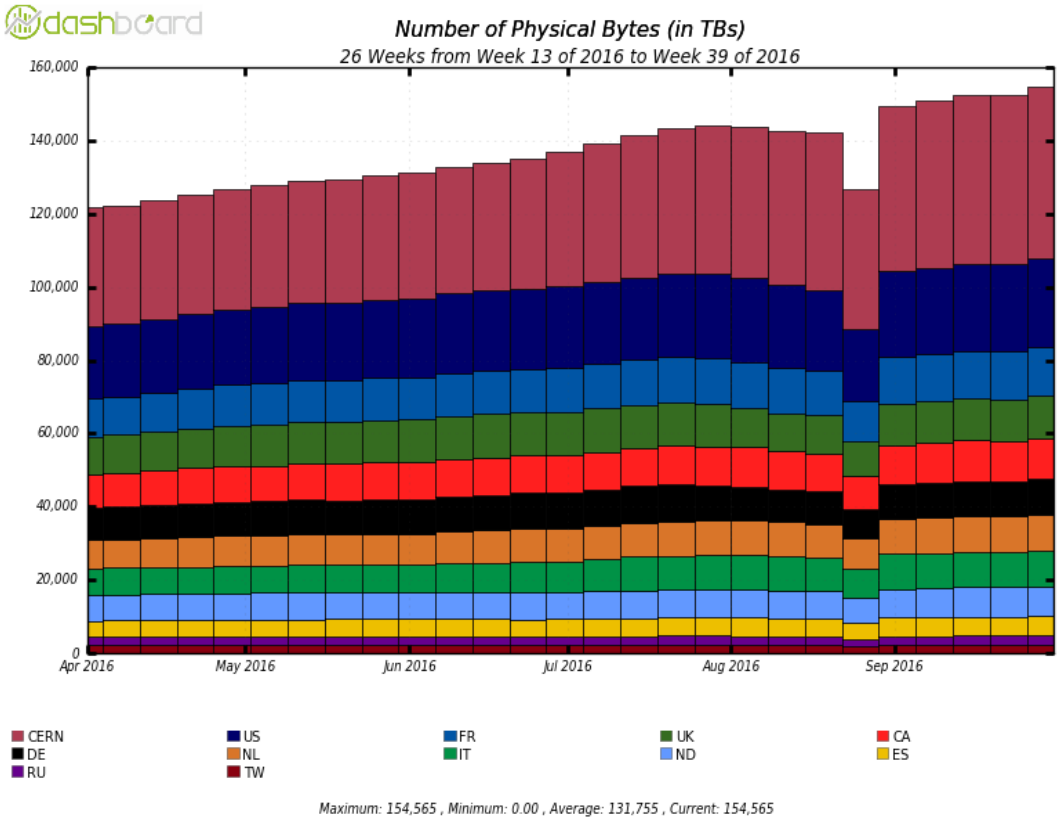


Figure-11 UK share of ATLAS data 1 April 2016- 30 Sept 2016.

The resource at major Tier-2 sites continues to be invaluable to the ATLAS operations. With the declining human resource at smaller sites, ATLAS is working to develop modes and practices that make most efficient use of those sites with the reduced effort.

ATLAS continues to press for higher efficiency in its use of resources. It runs almost all production on 8-core multi-process mode, with analysis using single cores. Much work has gone into the long tail of jobs that do not complete and into avoiding the pilot jobs absorbing resource. ATLAS continues on-going work to explore opportunistic resources and new styles of computing, with the UK in the vanguard. The Grid continues to dominate (80% of CPU); but HPC resources make a useful contribution to some workflows, particularly Monte Carlo generation. The High Level Trigger farm is used as a processing resource outside of running periods, and we have various demonstrator cloud resources.

The UK work with VAC is attracting particular attention in ATLAS as other regions try to manage the transition to many sites being diskless/cache only and very little local effort.

CMS

Overview

The period since the last OC meeting has been one of exceptional data taking at CMS (the integrated luminosity is shown in Figure-12). With over $35fb^{-1}$ during the 2016 running period. These data are currently being analysed and will provide great opportunities for both discoveries of new physics and precision measurements of parameters of the standard model.

Throughout 2016 the changes to the computing model described in the previous status report have worked well and have allowed the efficient use of the resources available to CMS. These include Dynamic Data Placement, use of a single global pool and Any Data, Any Time, Anywhere. However some inefficiencies were introduced at Tier 1 centres early in the year because of CMS running a mixture of single and multi-core jobs in multi-core pilots. These have been gradually reduced as CMS moves increasingly to multi-core jobs.

The quantity of data taken has placed pressure on the resources and CMS is making growing use of opportunistic resources. These include supercomputing centres and resources such as the high level trigger farm. While none of these resources are in the UK GridPP members are central to these activities.

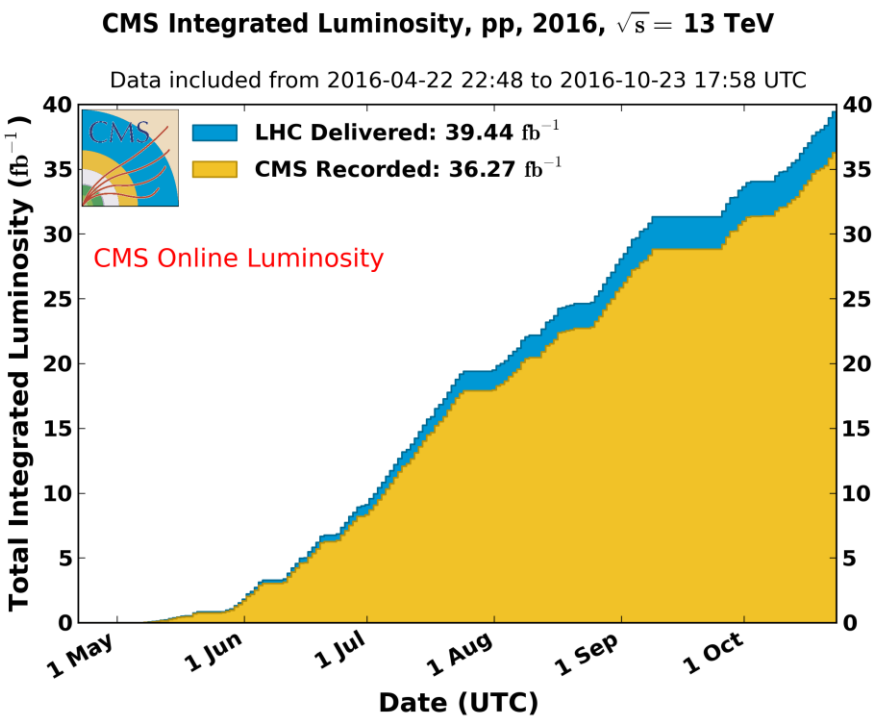


Figure12: Integrated luminosity collected by CMS in 2016

CMS Usage of UK Tier 1 and Tier 2 Resources

Tier 1 usage can be seen in Figure-13 and Tier 2 usage in Figure-14

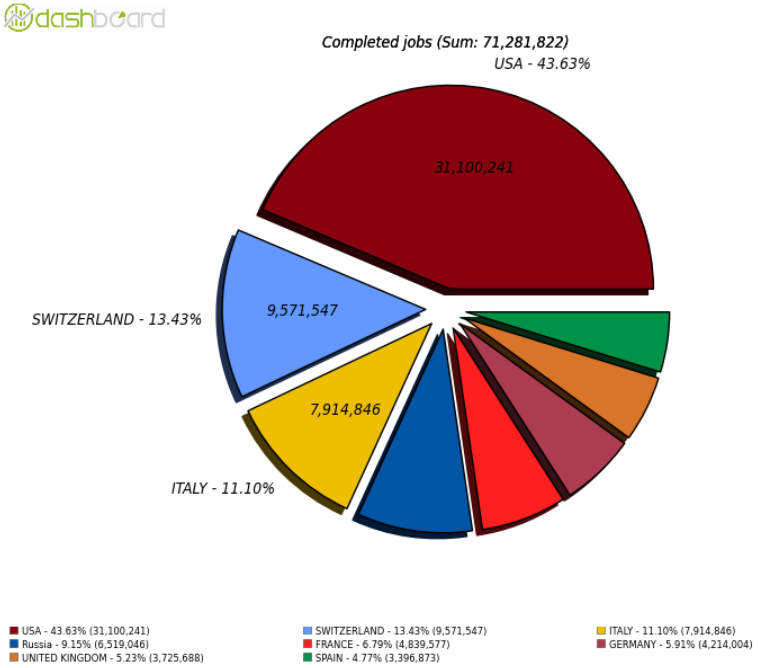


Figure-13: CMS use of Tier-1 resources in 2016

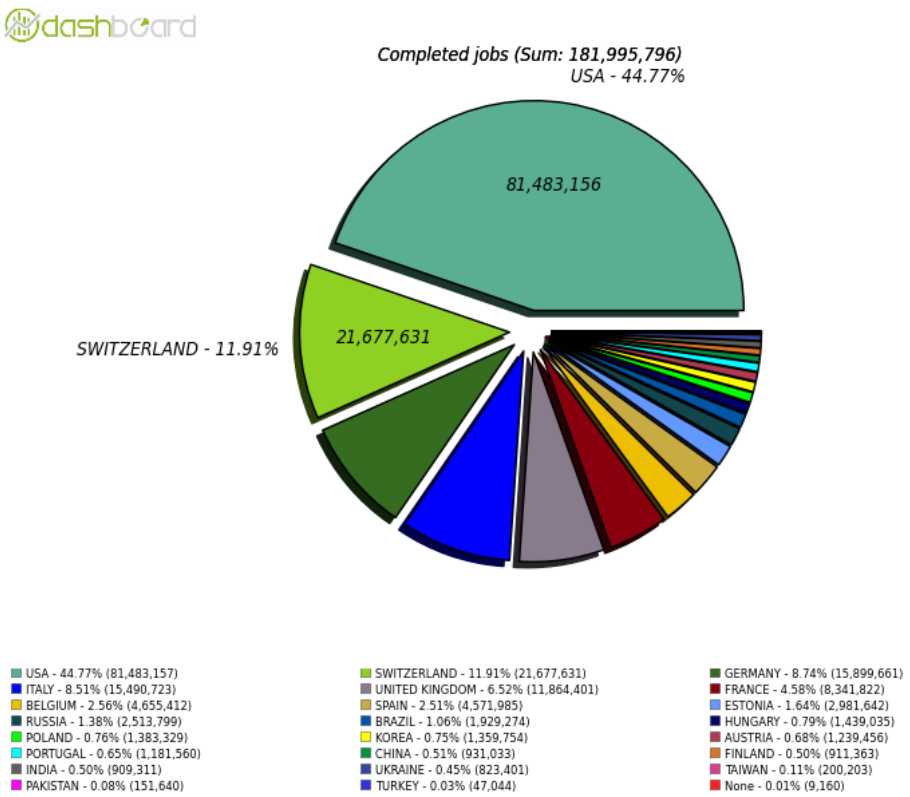


Figure-14: CMS use of Tier 2 resources in 2016

The UK has provided 5.3% of the Tier 1 resources used by CMS in 2016 and 6.5% of the Tier 2 resources used by CMS in 2016. The extra 1.5% of Tier 2 resources have been provided by from university resources rather and not through GridPP funding.

LHCb

Since the start of GridPP5, the UK has continued its leading role in LHCb and has been the largest national contributor of CPU power, reflecting the size of the UK collaboration. During this period the UK has supplied 30.3% of the Tier-2 CPU used by LHCb, and 23.1% of the Tier-1 CPU and has continued to be a very reliable provider, both at Tier-1 and Tier-2 sites. LHCb has appreciated both the capacity and, as importantly, the reliability and availability of the resources. Figure-15a below shows CPU usage at Tier-1 sites (left) and Tier-2 sites (right) during Q2/3 of 2016 across all countries, with larger UK sites in leading positions. Similarly Figure-15b below shows the disk usage at Tier-1 sites (left) and Tier-2 sites (right) during 2015. 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.

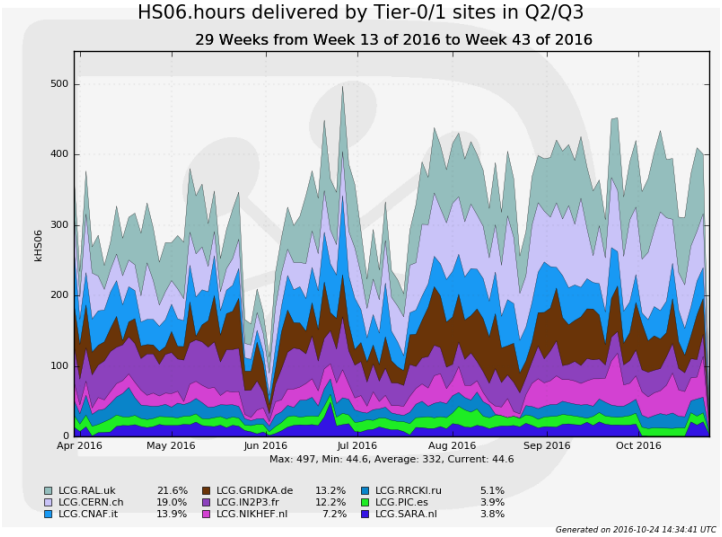
LHCb changed its computing model during GridPP4/4+ to parallel that of the GPDs, with the expectation that more Tier-2 sites will participate in data processing as well as simulation, and that nominated Tier-2 sites with storage will additionally hold real and simulated DST data for user analysis jobs. The UK provides five (Manchester, RAL PPD, Imperial, QMUL, Liverpool) of the 12 Tier-2 sites with storage. At the end of Q3 the storage at Glasgow was ready for certification by LHCb as a sixth UK site.

LHCb has evolved its computing model for the approaching Run-2 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 may be needed from 2017 onwards. The turbo stream assumes that all necessary information can be created in the trigger farm itself, and therefore no further offline processing is needed. These changes have helped to reduce the overall CPU requirement, and LHCb continues to use substantial non-pledged resources in Russia, and to make full use of its trigger farm when available.

Along with the introduction of the turbo stream, the data to be stored for each event has been reduced in several ways and a less conservative data replication strategy has been used. In 2016 the LHC live time increased by 40% compared to the figure planned for[1], which proportionately affects LHCb as the experiment operates at a fixed triggering rate throughout each fill. However this has only led to an increase by 10% in CPU and tape request to the WLCG CRSG, and a 25% increase in requested disk.

In preparation for the LHCb upgrade during LS2, the experiment is conducting a systemic programme of planning and updating its software and computing systems. This includes a continuation of the work to parallelise its framework, Gaudi, and is very much assuming many core-devices will be the norm in future. LHCb is also adapting its workload management system to support virtual machine architectures associated with the desire to use heterogeneous resources. This activity is being led by the UK in conjunction with the virtualized infrastructure provided by GridPP sites and using the virtual machine lifecycle managers, Vac and Vcycle, developed by GridPP. These developments are already giving LHCb access to resources equivalent to another larger Tier-2.

[1] See Table 3-1 of https://cds.cern.ch/record/2212512/files/LHCb-PUB-2016-022.pdf



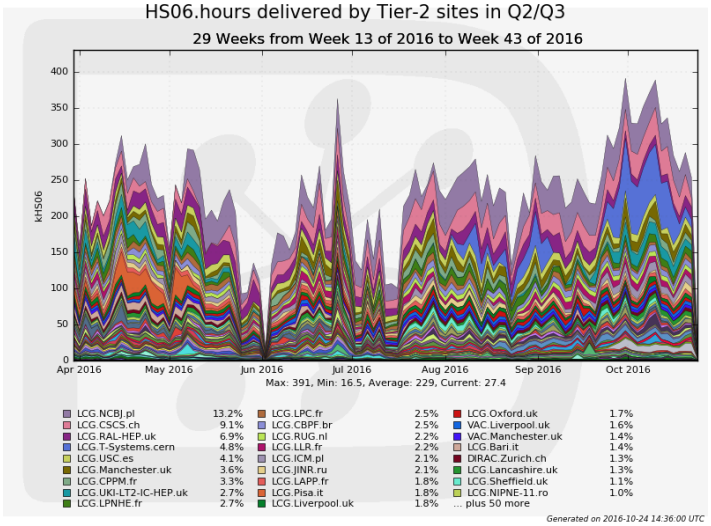
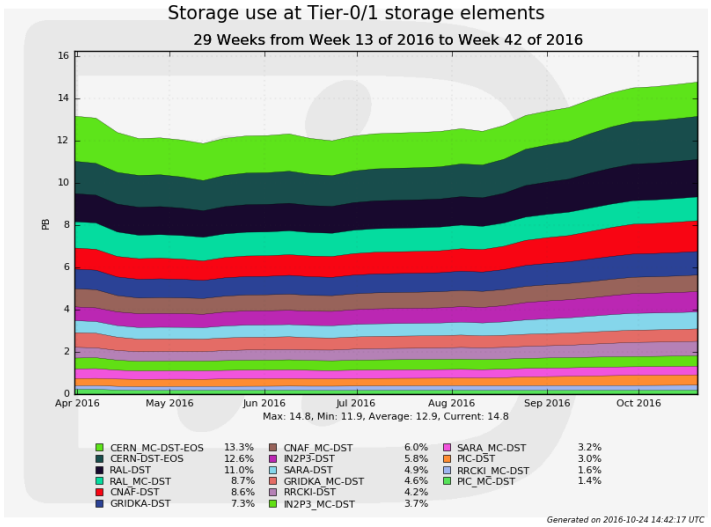


Figure-15a: LHCb CPU usage at Tier-0/1 (left) and Tier-2 (right) in Q2/3 2016



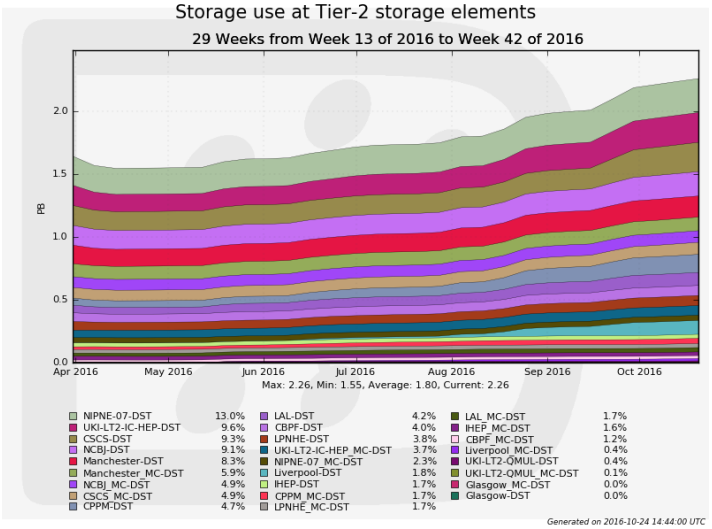


Figure-15b: LHCb disk usage at Tier-0/1 (left) and Tier-2 (right) in Q2/3 2016

Other VOs

We are making good progress with our other VO communities (i.e. other than ATLAS, CMS or LHCb). Excluding ALICE, the fraction of GridPP Tier-1 resources used by "others" was 4.9% in Q2 2016. At Tier-2s the figure reached 11.5% in Q2 2016 with Biomed, gridpp, ilc, lsst Phenogrid and snogrid all running significant numbers of jobs.

In recent months there have been several developments that demonstrate and have furthered our support for wider communities. A new joint EUCLID/GridPP researcher started at Edinburgh; our approved VOs list was expanded to include IceCube and Fermilab in late May, and we supported LZ registration in the EGI portal during June, and the sharing of VOMS information for the DUNE VO. Alongside these enablement activities we have enhanced the GridPP processes supporting such communities. Now when VO records change our approved VOs document version will have its version incremented and the RPMs of changed VOs will be released carrying the same version stamp as the document.

Recent specific VO/area updates include:

- *ALICE*: The ALICE VO has made extensive use of the CPU at the Tier-1 over the recent period. Taking advantage of a lull in use by some of the main LHC VO's to use >10% of the resources.

- *DEAP3600*: The DEAP project will generate around 10TB of (calibrated) data per year for 5 years, starting this year. The original (much larger) raw data are backed up on tape in Canada, but the calibrated data are not. For reasons of backup and access, the supporting group are discussing whether it may be possible to get their 50TB of calibrated data stored at the Tier-0 at RAL.

- The Tier-1 continues to support the HPC DiRAC service by providing tape backup on the CASTOR service. So far two sites have been configured to use this service and have transferred PBs of data for backup purposes. A third DiRAC site is currently being incorporated into the work.

- *The GalDyn Project*: A student (visiting from China) will pick up on previous grid deployment work for an upcoming paper. They will have a UCLan computing account. It is expected they will make use of the CVMFS Stratum-1 server already setup at RAL to host the VO software.

- *Infrastructure for smaller VOs*: Workload and data management strategies for our smaller VOs are still a concern. Our main strategy has been to encourage use of the GridPP DIRAC service to submit jobs. Some groups have also made use of Ganga to control their jobs and this approach has recently been included into the GridPP Users's Guide. It should be noted that GridPP funded support of Ganga has now ceased which will mean that support for this from GridPP staff will be increasingly limited.

- SNO+: There has been some initial setup work with SNO+. A Condor submission server has been setup but more work is required to move this into a stable production level solution.

- A very productive workshop organised jointly with UK based SKA contributors looking ahead at the design study for an SKA Regional Science Centre in Europe took place in early November. Some good work has already been undertaken with collaborators at Manchester under a new GridPP created and hosted skatelescope.eu VO and working with LOFAR data.

Further information related to the engagement with new user communities can be found in the next section on Impact and Dissemination.

Impact and Dissemination

Overview and status

The focus of impact and dissemination activities over the course of GridPP4+ has continued with GridPP5, and the New User Engagement Programme (described below) goes from strength to strength. The GridPP Dissemination Officer works partly for GridPP and partly as an STFC Public Engagement Fellow working on "*CERN@school and GridPP: harnessing the power of the Worldwide LHC Computing Grid for research in schools and beyond*". The CERN@School project that has now become part of the newly-formed Institute for Research in Schools, a charitable trust supporting school students and teachers to develop authentic research in schools. This 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, the space-based Langton Ultimate Cosmic ray Intensity Detector (LUCID), the Monopole and Exotics Detector at the LHC (MoEDAL), and the TimPix project that uses Timepix data taken during Tim Peake's Principia mission on the International Space Station. As GridPP5 has progressed, the emphasis has continued to emphasise towards impact and two-way engagement with new users from beyond ATLAS, CMS and LHCb, notably with the addition of other STFC-supported projects such as the Large Synoptic Survey Telescope (LSST), EUCLID, LOFAR and MoEDAL.

The New User Engagement Programme

The investment in the suite of tools, resources, and documentation that forms the New User Engagement Programme continues to pay dividends in terms of engaging new user communities with GridPP resources, as detailed below. Developments in various include:

- *Infrastructure*: The GridPP DIRAC instance, hosted at Imperial, has continued to mature and offer a production-ready service for many small VOs. Considerable work has gone into adapting DIRAC to work with multiple VOs, with much success; many smaller VOs now use DIRAC for production-level activities.
- *Tools*: CVMFS, the GridPP CernVM, and Ganga continue to be used as the de facto toolkit for new user interaction with GridPP resources. Where it has proven difficult to use the CernVM (for example, in universities with restrictive network configurations), it has often been simpler to provide new users with computing accounts on local clusters. This is where Ganga really comes in to its own; small tweaks to a user's configuration allow jobs to be run locally, sent to a local batch system, or the GridPP DIRAC system. This allows for the easy testing of new workflows with a single, Python-based User Interface, greatly reducing development time.

- *The GridPP website and documentation for new users*: as more user communities have engaged with GridPP, more case studies have been added to the website, presented as an "Impact Matrix" listing the tools and resources used by each community. The GridPP UserGuide continues to develop as user feedback is received and new tools (e.g. Ganga) are added.

User community engagement

We have continued to engage users from HPC DiRAC, GalDyn (Galaxy Dynamics, University of Central Lancashire), the Large Synoptic Survey Telescope (LSST), the Low Frequency Array (LOFAR), Euclid, the Lux-Zeplin (LZ) experiment, as well as pheno (phenomenology at Durham), T2K, Sno+ and SuperNEMO. The most recent development has come from the MoEDAL experiment; while this is technically an LHC VO, it represents a small user community with limited resources and no previous grid access. Thanks to the tools and documentation in place as part of the New User Engagement Programme, it was possible to prepare the MoEDAL VO (vo.moedal.org) for grid running within a few weeks. The MoEDAL simulation software was deployed via CVMFS and new users were brought online with CernVMs or lxplus using Ganga as a User Interface. Simulations for an upcoming paper, that would have taken a month to run with resources previously available resources (i.e. local clusters), took only a matter of days to complete. As mentioned above, a number of case studies have been written up for the website; more will follow in due course.