Les Robertson

5 September 2006

LHC Computing Grid Project
Overview Board – 11 September 2006
Status Report

1 MoU Signatures

The following countries have now signed the MoU:

China	*France*	*Germany*	*Italy*	*India*
Japan	*Netherlands*	*Pakistan*	*Portugal*	*Romania*
Taiwan	*UK*	*USA*		

Signatures are pending from the following collaboration members:

Australia	*Belgium*	*Canada*	*Czech Republic*	*Nordic*
Poland	*Russia*	*Spain*	*Switzerland*	*Ukraine*

Canada has now secured funding and expects to sign the MoU in the coming weeks. There is a growing concern that two of the countries that are planning to provide Tier-1 services (Spain, Nordic) are not yet in a position to make a formal commitment. There is now little more than a year before the first collisions.

2 Second Quarter 2006 Progress Report

The LCG quarterly status and progress report for the second quarter of 2006 is available from the planning page. The full report includes the status of each of the Tier-1 centres, as well as of the experiments and the activity areas of the project. There is a separate executive summary.

3 Level-1 Milestones

The status of Level 1 Milestones due since the last Overview Board meeting and those due in the coming 3 months is summarised in the table on the next page. Full milestones tables are available via the LCG Planning Page.

The only level-1 milestone due was the internal review of Castor 2, which took place at the beginning of June. The executive summary of the reviewers' findings and recommendations is contained in the final report.

Two milestones fall due at the end of September – SC4-5 and IS-1 – defining the end of Service Challenge 4 and the beginning of the initial LHC service. It is clear that these will not be achieved by the end of the month: the site reliability is far from the target level and has not been improving during the period of SC4. This issue is discussed in more detail later in this report. As far as the CERN→ Tier-1 data distribution test is concerned, the target data rate of an aggregate 1.6 GB/sec was demonstrated to disk in the Tier-1s already in April, but many sites will not have sufficient tape handling capacity installed to extend this to tape as required by the milestone. The milestone will therefore have to be re-scheduled.

DBS-1, operation of the distributed database services at all Tier-1 sites will also not be achieved on time. The current status is covered in a later section of this report.

Level 1 Milestones due to end August 2006			
Colour coding: completed - successful completed - partially successful postponed			
ID	**Date**	**Milestone**	**Status**
CAS-1	15Mar06	**Castor2 Readiness Review**	The review took place on 6-8 June.
Level 1 Milestones due in coming 3 months			
ID	**Date**	**Milestone**	**Status**
SC4-5	30.09.06	**Service Challenge 4: Successful completion of service phase**	1) **8 Tier-1s and 20 Tier-2s** must have demonstrated availability better than 90% of the levels specified in Annex 3 of the WLCG MoU [adjusted for sites that do not provide a 24 hour service] 2) Success rate of standard application test jobs greater than 90% (excluding failures due to the applications environment and non-availability of sites) 3) Performance and throughput tests complete: Performance goal for each Tier-1 is the nominal data rate that the centre must sustain during LHC operation: CERN-disk > network > Tier-1-tape. Throughput test goal is to maintain for one week an average throughput of 1.6 GB/s from disk at CERN to tape at the Tier-1 sites. All Tier-1 sites must participate.
DRC-4	30.09.06	**1.6 GB/s data recording demonstration at CERN:** Data generator→ disk → tape sustaining 1.6 GB/s for one week using the CASTOR mass storage system.	
DBS-1	30.09.06	**Full LCG database service in place**	Milestone for all Tier 1 sites
IS-1	30.09.06	**Initial LHC Service in operation**	Capable of handling the full nominal data rate between CERN and Tier-1s. The service will be used for extended testing of the computing systems of the four experiments, for simulation and for processing of cosmic-ray data. During the following six months each site will build up to the full throughput needed for LHC operation, which is twice the nominal data rate.

4 The SC4 Service

SC4 began as scheduled at the beginning of June, as an evolution of the production services in place at that time using LCG resources connected to the EGEE and OSG grids. In the case of the EGEE sites, this coincided with a new release of the basic middleware – gLite 3.0. There had been concerns over the state of readiness of this release, reported at the June Overview Board meeting, but the deployment went smoothly and the middleware has proven to have a good level of reliability.

The plan is that SC4 will evolve into the permanent service for LHC when a certain number of performance and reliability metrics have been achieved (milestone SC4-5). This was scheduled for the end of September, but it is clear that they will not be achieved by that time.

We are still not able to demonstrate the full nominal Tier0-Tier1 transfer rates (1.6GB/s) over extended periods. However, experiment-driven data transfers (mainly ATLAS and CMS) have shown that the service can sustain 50% of the target rate for a sustained period (see Figure 1) under much more realistic conditions than in the previous (April) tests, and with about half of this data flow going to tape. In addition, both ATLAS and CMS have managed to export over 1PB of data (1 PB of data per month for CMS over a 90-day period, 1.25 PB of data for ATLAS in the two-month period starting 19th

June). Much work has been done and more remains to be done to improve these data rates, and to improve reliability of the service on a site by site basis.

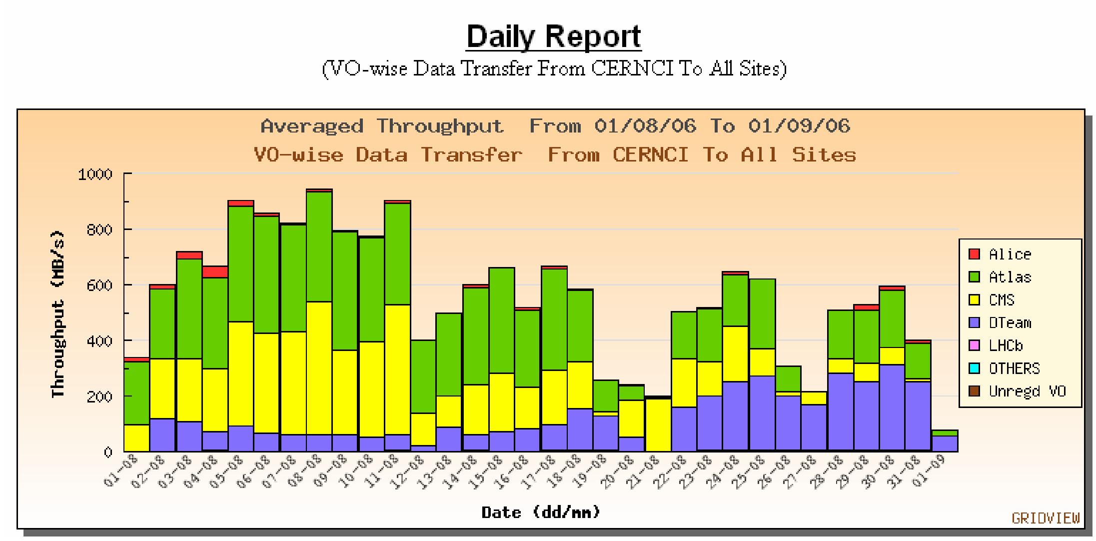


Figure 1 – Data distribution from CERN by VO – August 2006

The experiment activities have tested individual site services, such as LFCs, VO boxes, and overall production readiness significantly more than in previous test periods. Many problems have been identified and resolved, but it is clear that some sites / regions still have to make significant progress to achieve the required service level. In some cases sites appear to be able to focus their full attention on a specific experiment or challenge for only a few days only, possibly indicating workload problems at the sites. It looks as if the experiments can only expect a few days at high priority attention per month from each site. A particularly effective model, as demonstrated by Lyon for ATLAS, is to have a contact person for the experiment both at the Tier0 and the Tier1;

Upgrades to CASTOR2 at a number of sites have been the cause of instabilities. Once all such migrations have been completed, a further test needs to be made to ensure that these sites can now meet both throughput and stability targets.

Several sites have experienced significant power and / or cooling problems during the summer, resulting in prolonged service downtime. Many sites also appear to suffer from significant manpower shortages, shown very clearly during the summer holiday period, which impacts both the service level that they are able to provide and the response time to requests and problems.

The purpose of and attendance at the regular coordination meetings has been reviewed, in order to improve communications between the experiments, the sites and the grid and LCG service coordinators, while minimizing the number of meetings that must be attended by the individuals responsible for running the services. There are now four general coordination meetings:

- The LCG Resource Scheduling Meeting takes place weekly, attended by the experiments, the Service Coordination Team and the EIS team. The purpose is to review the current to medium term resource requirements of the experiments at CERN and the Tier-1s, and to follow the status of major and critical service problems, initiating escalation when appropriate.
- The weekly Joint Operations Meeting provides the operational coordination across the sites taking part in the LCG services, as well as providing the basic operational coordination of the EGEE grid infrastructure. The meeting is attended by the Tier-1s and the other major sites that are part of the EGEE infrastructure. The EIS team and the LCG Service Coordination Team provide the link with the experiments, and ensure that the planning and operational issues of the LCG services are fed into the meeting. The operational issues at Tier-2 sites that do not attend the meeting are the handled through the EGEE Regional Operations Centres and the US Tier-1s.
- The Operations Meeting at CERN, which takes place each morning, serves also as the daily LCG services meeting, through the presence of the LCG Service Coordination Team.

- The weekly LCG Service Coordination Meeting, attended by the members of the Service Coordination Team, the EIS team and the people responsible for the major LCG services at CERN provides the overall coordination.

Reporting to and attendance at the weekly Joint Operations Meetings has improved since the beginning of SC4 but still leaves considerable room for further improvement, perhaps indicating that the services for the LHC experiments are not yet the main priority at the sites.

A summary of the experiment planning for using the SC4 service is maintained on the LCG planning pages - https://twiki.cern.ch/twiki/bin/view/LCG/SC4ExperimentPlans.

5 Job Reliability Monitoring

The work on job reliability (Grid reliability) is continuing and the main points are:

- A system to process and analyse job logs has been put in place for some of the major activities in ATLAS and CMS, identifying and classifying errors. The results have been communicated to the developer teams and discussed in the EGEE TCG. We expect this activity to become effective on a fairly short timescale, particularly when the new workload management middleware enters production.
- In collaboration with grid operations the job failure analysis can be used to identify the most important site (configuration) problems impacting the experiment's activity. The data is published daily and this allows the operation team to drill down from site, to computing elements to (in certain cases if permitted by the site security policy) affected worker nodes. This was in use by the end of August, and there are already some examples of problems being identified and fixed. It is intended to use this data to show the comparative long term efficiency of sites in terms of job failure rates for each experiment.

6 Accounting

The full accounting report for CERN and the Tier-1 sites, covering January through July 2006, is available at

http://www.cern.ch/LCG/mb/Accounting/accounting_summaries.pdf

A summary of the data for the four-month period from April through July by site and by VO is contained in the following table. The resources used during the period are compared with the capacity installed for LHC experiments at the site, and with the commitment in the MoU for 2006. In these comparisons the installed or pledged capacity is reduced to take account of standard utilisation efficiency factors, and so it should be possible to reach 100% provided there is a consistent load. Where the consumption exceeds 100% of the installed capacity this indicates that LHC experiments have been able to use resources provided for other applications. The usage made by each VO of the resources at each site is summarised in Appendix 2.

Overall about 60% of the installed resources (cpu, disk and tape) have been used during the period, equivalent to only 40% of the capacity planned in the MoU. Three sites have delivered over 80% of their installed capacity, and four sites have delivered less than 40%. During this period there has been a significant load from Monte Carlo production, where high utilisation factors should be expected, but there are also other tests being carried out that have greater demands on the computing environment such as mass storage and network performance, and are more sensitive to overall reliability. It should be noted that these tests do not provide a constant load on the system. It is essential to continue with a test programme, stressing the full set of services, until we are well into the first full year of data taking. During this time we should therefore expect to see relatively low utilisation levels at the Tier-1s.

Site Summary	*cpu*			*disk occupancy*			*tape occupancy*		
	KSI2K-days	*% of installed*	*% of pledge*	*TBytes at end of period*	*% of installed*	*% of pledge*	*TBytes at end of period*	*% of installed*	*% of pledge*
CERN Tier-0+CAF	168,286	71%	48%	412	59%	76%	1,472	59%	98%
ASGC	15,895	35%	16%	21	63%	8%	13	4%	3%
BNL	68,571	94%	59%	42	33%	12%	298	71%	99%
CC-IN2P3	31,734	46%	26%	44	75%	12%	313	78%	59%
CNAF	44,816	27%	24%	94	30%	16%	198	40%	23%
FNAL	55,511	59%	74%	108	154%	154%	300	100%	120%
FZK-GridKA	21,181	38%	20%	26	33%	13%	151	38%	38%
NDGF	16,679	39%	28%	45	62%	38%	0	0%	0%
NL LHC/Tier-1	13,574	69%	43%	0	0%	0%	0	0%	0%
PIC	10,953	70%	42%	15	102%	15%	99	89%	63%
RAL	59,668	257%	59%	33	74%	10%	85	121%	13%
TRIUMF	1,486	84%	8%	4	95%	21%	0	0%	0%
Total	**508,354**	**61%**	**39%**	**844**	**56%**	**27%**	**2,929**	**56%**	**52%**

VO Summary	*cpu*		*disk occupancy*		*tape occupancy*	
	KSI2K-days	*% of total*	*TBytes at end of period*	*% of total*	*TBytes at end of period*	*% of total*
ALICE	30,387	6%	110	13%	283	10%
ATLAS	253,233	50%	307	36%	1,183	40%
CMS	110,351	22%	345	41%	964	33%
LHCb	114,383	23%	82	10%	499	17%
Total	**508,354**	**100%**	**844**	**100%**	**2,929**	**100%**

7 Availability & Reliability

The algorithm for computing site availability was agreed by the MB at the beginning of April and measurement started for CERN and the Tier-1 sites from the beginning of May. In July it was decided to measure also *reliability*, defined as the time that the site passes a set of standard tests as a percentage of the time that it is scheduled to be available. *Availability* is defined as the percentage of the time that the site passes the tests, taking no account of scheduled down time. The reliability measurement was introduced from the beginning of July. The results for availability from May through August are summarised in the table below, and fuller data for each site is given in Appendix 1.BNL and NDGF do not yet take part.

Site availability

	CERN-PROD	FZK-LCG2	IN2P3-CC	INFN-T1	RAL-LCG2	SARA-MATRIX	TRIUMF-LCG2	Taiwan-LCG2	USCMS-FNAL-WC1	PIC	average - all sites	8 best availability	8 best reliability
May	89%	85%	83%	89%	68%	58%	77%	87%	68%	61%	77%	81%	-
June	92%	15%	89%	62%	76%	49%	88%	75%	64%	88%	70%	79%	-
July	90%	54%	87%	31%	73%	84%	80%	98%	20%	87%	70%	82%	83%
August	91%	65%	89%	64%	55%	79%	82%	92%	0%	84%	70%	81%	82%
average	90%	56%	87%	62%	67%	68%	82%	89%	37%	80%	72%	78%	

target (90% of MoU)	88%	# sites meeting target	2	8 best site avge as % of target	89%
90% of target	79%	# sites > than 90% of target	5		

The target availability for SC4 is 88%, to be achieved by at least 8 sites. Only two sites achieved the target on average during the four month period, with a further three coming within 10% of the target. The average for all 10 sites is 72%. Taking the 8 best sites in each month (different sites in each month) the average hovers around 80%, and in some individual cases the situation has deteriorated rather than improved. There are additional reasons to be worried about this situation. The tests used for the measurements at present cover only a few functions, and are far from providing a comprehensive validation of the site's availability. On the other hand, in some cases a test fails at a particular site but the site is still able to run jobs for one or more VOs. However, observing the current SC4 activity there is a clear correlation between the reliability of some sites and their ability to take part in the some of the VO challenges. It is also clear from SC4 that site monitoring of local services still needs considerable further improvement – many issues that could be spotted locally are still first found by the central Service Coordination Team or – worse still – by the users. The summer holiday season has also shown that some sites do not have sufficient overlap of expertise to cover for scheduled or unscheduled absences.

8 Applications Area

The Internal Review of the Applications Area will take place on 18-20 September, just before the LHCC Comprehensive Review of LCG on 25-26 September.

The projects in the Applications area continuing to support the experiments in their preparation of the software releases that are going to be used in the various data challenges and productions this fall. Several iterations of the software packages (ROOT, CORAL, POOL, COOL, etc,) have been made available in various configurations to allow experiments to integrate the new functionality, provide feedback and be ready with a production quality releases by this summer. Special emphasis has been put in the optimization of this iterative process. For that, a new set of procedures for testing and building the software are being put in place to optimize the time that it takes to integrate by the experiments the changes and bug fixes in libraries provided by the Application area.

In the ROOT project strong development is taking place in the integration of the C++ interpreter (CINT) with the C++ reflection system (Reflex). It is planned to release the new version of system this fall. The mathematical libraries have been consolidated and additions have been added concerning Fast Fourier Transforms and Multivariate Analysis. Many developments are currently ongoing in the PROOF system as the result of the serious testing being done by ALICE in the context of their analysis facility (CAF). Important performance improvements are being introduced when accessing remote files. The first batch of these improvements is available in version 5.12.

The POOL/CORAL project has been consolidating the generic RDBMS interface for Oracle, MySQL, SQLight and FroNTier. New functionality has been developed for improving the overall reliability of user applications with database back ends. This new functionality consists of database lookup by logical name; fail over to other databases; connection pooling; authentication and monitoring facilities. In addition, the COOL project (conditions database) has been improving the versioning capabilities by the use of tags and hierarchical tags.

The Simulation project is putting considerable effort on the study of hadronic shower shapes, to understand the discrepancies observed between simulation and test-beam data. Also comparisons between Geant4 and FLUKA simulation packages are being made with the help of the set of tools that has been developed to facilitate this task.

A new version of the Geant4 has been released. It includes a new precise elastic process for protons and neutrons. It is particularly relevant to improving the accuracy of energy deposition in scintillators. In addition, the new G4 version includes a new efficient method to detect overlaps in a user's geometry and updated particle definitions to match with PDG-2005 among other improvements.

9 Distributed Database Deployment – 3D

The full database services are scheduled to be in production at all Tier-1 sites at the end of September (Level-1 milestone DBS-1). The Oracle database systems at the phase 1 sites (ASGC, BNL, CNAF, GridKA, IN2P3, RAL) are now synchronized with experiment clusters at CERN for direct experiment tests with conditions data. All phase 2 sites have now nominated database contacts and in all cases there are plans for commissioning the appropriate equipment. Additional database expertise has been acquired at two of the phase 1 sites, but appropriate staffing remains a problem – in several cases it is not clear how the services will be supported on a 24 X 7 basis. A workshop will take place on 13-14 September at which the current status of the phase 2 sites will be reviewed. It seems, however, likely that at least two of the sites will not be in production by the deadline at the end of this months.

The Frontier/SQUID systems have been set up and tested at all of the CMS Tier-1s and all but three of the CMS Tier-2s. Multi-client stress tests have been done by CERN to validate the Frontier back-end installation at the T0.

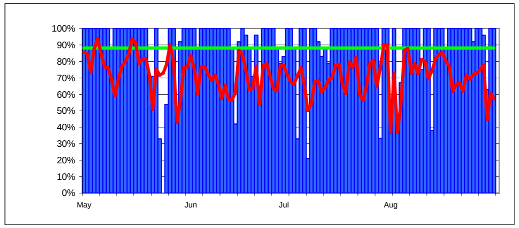
APPENDIX 1 – Site Availability – May-August 2006

Availability of WLCG Tier-1 Sites + CERN 2006

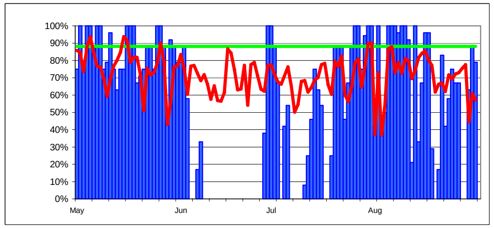
Data from SAM monitoring. Site availability and reliability as agreed in WLCG MB on 11 July 2006 (scheduled interruptions are excluded when calculating reliability)

average (all sites) 72% **target (90% of MoU)** 88% **site average colour coding: < 90% of target ≥ 90% of target ≥ target**

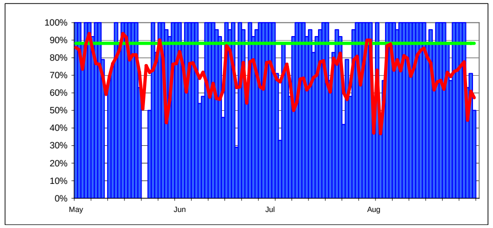
 # sites meeting target 2 **Average of 8 best sites** 78% (89% of target)



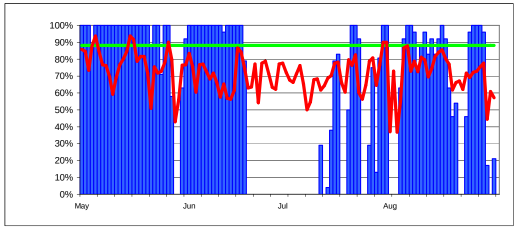
CERN-PROD **average** **90%**



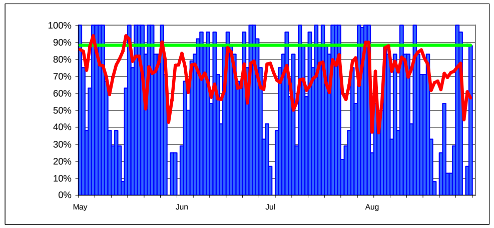
FZK-LCG2 **average** **56%**



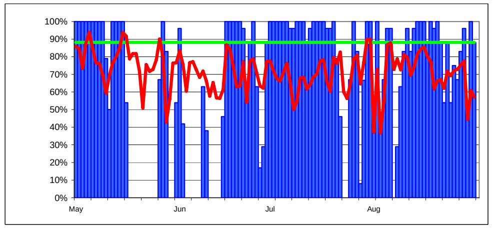
IN2P3-CC **average** **87%**



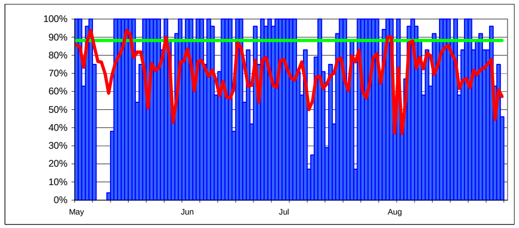
INFN-T1 **average** **62%**



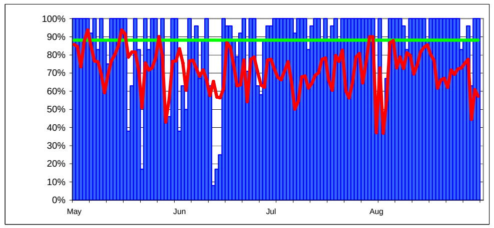
RAL-LCG2 **average** **67%**



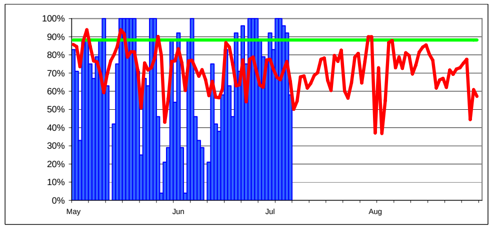
SARA-MATRIX **average** **68%**



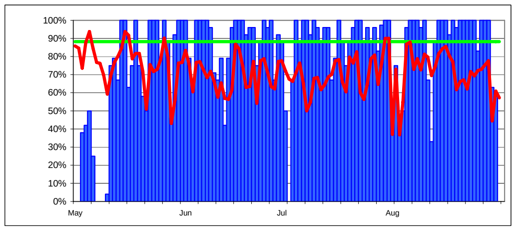
TRIUMF-LCG2 **average** **82%**



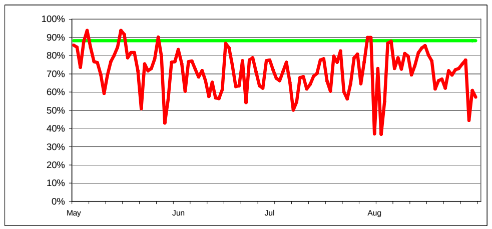
Taiwan-LCG2 **average** **89%**



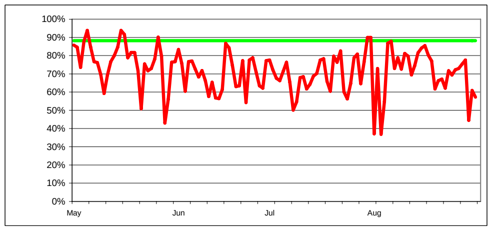
USCMS-FNAL-WC1 **average** **37%**



PIC **average** **80%**



BNL-LCG2 **average** **n/a**



NDGF **average** **n/a**

APPENDIX 2 – Summary of Accounting Data – April-July 2006

ALICE	cpu		disk occupancy		tape occupancy	
	KSI2K-days	*% of total*	*TBytes at end of period*	*% of total*	*TBytes at end of period*	*% of total*
CERN Tier-0+CAF	14,889	49%	81	74%	234	83%
ASGC						
BNL						
CC-IN2P3	4,240	14%	15	14%	31	11%
CNAF	6,947	23%	8	7%	15	5%
FNAL						
FZK-GridKA	1,847	6%	6	5%	3	1%
NDGF						
NL LHC/Tier-1	1,182	4%				
PIC						
RAL	1,282	4%				
TRIUMF						
Total	**30,387**	**100%**	**110**	**100%**	**283**	**100%**

CMS	cpu		disk occupancy		tape occupancy	
	KSI2K-days	*% of total*	*TBytes at end of period*	*% of total*	*TBytes at end of period*	*% of total*
CERN Tier-0+CAF	33,161	30%	122	35%	419	43%
ASGC	5,599	5%	6	2%	2	0%
BNL						
CC-IN2P3	2,127	2%	15	4%	69	7%
CNAF	2,117	2%	65	19%	36	4%
FNAL	50,892	46%	108	31%	300	31%
FZK-GridKA	1,847	2%	7	2%	75	8%
NDGF						
NL LHC/Tier-1	243	0%				
PIC	3,419	3%	8	2%	27	3%
RAL	10,946	10%	14	4%	36	4%
TRIUMF						
Total	**110,351**	**100%**	**345**	**100%**	**964**	**100%**

ATLAS	cpu		disk occupancy		tape occupancy	
	KSI2K-days	*% of total*	*TBytes at end of period*	*% of total*	*TBytes at end of period*	*% of total*
CERN Tier-0+CAF	72,709	29%	145	47%	461	39%
ASGC	10,296	4%	15	5%	11	1%
BNL	68,571	27%	42	14%	298	25%
CC-IN2P3	14,439	6%	11	4%	159	13%
CNAF	17,823	7%	16	5%	94	8%
FNAL	4,619	2%				
FZK-GridKA	11,429	5%	11	4%	68	6%
NDGF	16,679	7%	45	15%		
NL LHC/Tier-1	8,420	3%				
PIC	4,291	2%	7	2%	54	5%
RAL	22,471	9%	11	4%	38	3%
TRIUMF	1,486	1%	4	1%		
Total	**253,233**	**100%**	**307**	**100%**	**1,183**	**100%**

LHCb	cpu		disk occupancy		tape occupancy	
	KSI2K-days	*% of total*	*TBytes at end of period*	*% of total*	*TBytes at end of period*	*% of total*
CERN Tier-0+CAF	47,527	42%	64	78%	358	72%
ASGC						
BNL						
CC-IN2P3	10,928	10%	3	4%	54	11%
CNAF	17,929	16%	5	6%	53	11%
FNAL						
FZK-GridKA	6,058	5%	2	2%	5	1%
NDGF						
NL LHC/Tier-1	3,729	3%				
PIC	3,243	3%			18	4%
RAL	24,969	22%	8	10%	11	2%
TRIUMF						
Total	**114,383**	**100%**	**82**	**100%**	**499**	**100%**