Alberto Aimar
13 August 2006

LHC Computing Grid Project

Quarterly Status and Progress Report

2006 Q2

Executive Summary

1   Introduction

This document highlights major achievements and issues of the quarter, also providing an overview of the individual reports that were submitted by sites, experiments and projects.

For further details please refer to the complete document: "Quarterly Status and Progress Report - 2006 Q2" that contains the individual reports. All reports were reviewed by a Review Team (A.Aimar, M.Ernst, L.Robertson and A.Sansum). The reviewers then asked for additional information and, in some cases, the reports were modified and re-submitted by the original authors.

The status of the milestones in the report is described as it was at the time of the reports (end June 2006).

2   Level-1 Milestones

All the WLCG Level 1 milestones due were completed during the quarter:

CAS-1:Castor2 Readiness Review.

The review took place on 6-8 June. The review was initially scheduled for April but rescheduled for June, because a new chair had to be appointed (J.Harvey /CERN-PH). The report of the Castor 2 Readiness Review will be available during next quarter (2006Q3).

SC3-4: All services on all Tier-1 sites monitored

The Service Availability Monitor (SAM) system is in operation since mid April 2006 for all Tier-1 sites, except BNL and NDGF. The monitoring of the services has been implemented; a general tool checking site service availability was deployed with a first set of tests that hourly verify the functioning of all LCG services at each site.

SC4-3: Service Challenge 4 Set-up: Set-up complete and basic service demonstrated, capable of running experiment-supplied packaged test jobs, data distribution tested

The basic services were in operation by the end of April, with data distribution from CERN to Tier-1s fully tested. The pre-production test phase was not implemented, due to the late availability of the gLite 3.0 software distribution.

DRC-3: 1.0 GB/s data recording demonstration at CERN: Data generator to disk to tape sustaining 1.0 GB/s for one week using the CASTOR 2 mass storage system and the new tape equipment.

The 1 GB/s data recording rate was reached and sustained for one week, with a rate of 1.6 GB/s sustained over a 24 hour period.

SC4-4: Service Challenge 4: Start of stable service phase, including all Tier-1 and 30 Tier-2 sites

The Service Challenge 4 started on the 4 June 2006. During the challenge, the service must be able to support the full computing model (use cases and services to provide) of each experiment, including simulation and end-user batch analysis at Tier-2 sites. FTS services at Tier-1s for data distribution to Tier-2s and between Tier-1s were only partially in operation.

3   Service Challenges Progress

SC4 Started - SC4 began as scheduled at the beginning of June, evolving from the production service as sites upgraded their software and configured the necessary mass storage services and data transfer channels. ATLAS began Tier-0 and data distribution tests to all Tier-1s and some Tier-2s on 19 June. LHCb production for their data challenge DC06 began during the quarter, but the data processing phase has been delayed until July. The pre-production of ALICE's Physics Data Challenge PDC06 has been running since April to test the AliRoot software. A summary of the SC4 planning for experiments and sites is maintained on the web at https://twiki.cern.ch/twiki/bin/view/LCG/SC4ExperimentPlans.

Throughput Tests - The SC4 disk-disk and disk-tape throughput tests showed improvement since the SC3 re-run of January /February, but only managed to meet the target disk-disk rate of 1.6 GB/sec for a 24 hour period. Further work on service stability is required, as is testing under realistic data taking conditions if we are to meet the target of sustained 1.6 GB/sec to tape by the end of SC4. A plan to achieve this by combining the data export tests of all 4 LHC VOs in July and August has been prepared and should result in 800MB/s summed over these VOs with a further 800MB/s driven by the deployment team (dteam).

Service Challenge Workshop and Technical Day - A well attended (>160 people) 3 day workshop was held at CERN in June, focusing primarily on those Tier2s who have not yet been fully involved in these activities. This was followed by a 2-day tutorial and a 1-day Service Challenge Technical day. The results of the Technical Day included a clear list of experiment requirements and schedule, presented to the MB and carried over in a combined action list, reviewed weekly. A questionnaire on the tutorials and workshops generated very positive responses. Further such events with increased participation from the sites were encouraged, with very good ratings assigned to both presentations and tutorials. A follow-up event is foreseen for January 2007, with a further workshop immediately prior to CHEP in September.

Review of Services - The Internal Review of Services that took place at the beginning of June made several recommendations to improve communications flow within the project. A number of actions will be taken during the third quarter to improve the situation, including the introduction of a bulletin, a review of the goals and functioning of the weekly service coordination and operations meetings, and the implementation of a system to ensure that sites and experiments can contact each other rapidly when urgent operational issues arise.

4   Summary of Major Progress and Issues

4.1   Grid Sites Preparation

Tier-1 Sites - While some of the Tier-1s have completed extensive testing of Tier-1/Tier-2 data transfer services (e.g. IN2P3, RAL) in other cases there have been delays in configuring these services. Several sites are in the process of upgrading their operational infrastructure and processes to improve reliability. From the beginning of May site availability is monitored by the Management Board each month, but the figures for the first two months indicate that much work is still required if we are to achieve the SC4 availability goals. There have been delays in preparations for SC4 at several sites, for a number of reasons including hardware acquisition problems, difficulties encountered in migrating to Castor 2, and staffing problems.

Reporting of accounting data began from the beginning of April – the current summary data is maintained on the Web: http://www.cern.ch/LCG/MB/accounting/accounting_summaries.pdf.

The site status reports in this document have a common layout but the milestones are adapted to the detailed plans and status of each site and it is difficult to gauge the progress relative to the overall goals or make comparisons with other sites. A number of systems are now in operation to measure service and site performance and reliability (FTS service throughput, resource usage, site availability, ..) and these should be integrated into a "dashboard" showing the status of each site compared with the targets, adding additional factors such as installed capacity v. current VO requirements, status of new services, etc.

Tier-1 - Tier-2 Associations - An activity has been started to integrate data from all four experiments on expected relationships between Tier-1 and Tier-2 centres (data rates, storage requirements) in order that

sites can configure and test the necessary FTS services, verify that there is adequate networking performance, and initiate operational relationships with their partner sites.

Tier-0 Installations - The new tape installations from IBM and STK are successfully in production. Quite a few data challenges were done during Q2 achieving the high throughput data rates. Issues with CASTOR2 were identified and will be tackled in the next months (like the load balancing problems). The full migration to CASTOR2 has been finished. The full cooling and ventilation upgrade was completed in May after some delays. Several unexpected power interruptions caused service interruptions. The positive side of this was that, profiting from this experience, the time to recover all services was reduced considerably. Resources (CPU and disk) were made available as planned. Currently the 4 experiments have already more disk space than planned to cope with several scheduled data challenges.

DAQ-T0-T1 Data Flow - The progress in establishing a plan and tests for the full data chain (DAQ-T0-T1) was slower than expected and a review of the milestones will be presented for the next quarterly report. More in general, after some iteration with the 4 experiments the architectures and the plans have been collected in the following web page: https://cern.ch/twiki/bin/view/Main/DaqTierZeroTierOnePlanning but the fine grain details are still to be sorted out. The internal milestones of the experiments are not really compatible in all cases with the corresponding LCG milestones, which have therefore been suspended. A review of these milestones will have to take place during the next 3 month period.

VO Boxes Installations - The VO Boxes needed by the LHC experiments are installed at all required Tier-1 sites and are functioning adequately. Further work is needed between sites and experiments in order to agree how VO Box configurations and data are stored and backed up (by the sites or by the VOs). In addition, as agreed at the VO Boxes workshops, the services provided by the VO Boxes should be implemented by future services (or improvements) of the general middleware software, making unnecessary the deployment of any specific VO Box. Work in this direction is in progress under the steering of the EGEE Technical Coordination Group (EGEE TCG).

Milestones for LCG Initial Service - The quantitative measures in the level-1 milestone for successful completion of SC4 (SC4-5) cover only reliability and Tier-0/Tier-1 data distribution rates. Additional quantitative targets should now be set for the Initial Service, such as job submission rate, number of simultaneous jobs, catalogue access rates, etc. Milestones should also be defined within the Service Challenge area to follow experiment preparations for the beginning of their data challenges and to measure their success in using the services.

4.2   Procurement

During this quarter efforts were devoted to the purchasing of hardware (compute nodes and disk servers). Delays in the delivery of the purchased hardware are announced in several of the quarterly reports; but will be without major expected consequences on the initial schedule for putting in production these resources. In addition, following the reschedule of the LHC programme for 2007 and 2008, several sites will re-plan their procurement schedule in the next few months. In addition, some sites are pausing their disk expansion purchases, and wait for the availability of lower prices or for the introduction of new technologies.

4.3   LCG Middleware Releases and gLite 3.0

The preparation of the gLite 3.0 release was integrated, certified and deployed to both the pre-production and production services and on 4 June had been installed at all but three Tier-1s, with CERN and FZK upgrading before the end of the quarter (by which time more than 50% of all EGEE sites had upgraded). The release had been made available for distribution on May 4th, only 3 days behind schedule. Considering the number of problems encountered during the release preparation this was a considerable achievement. The deployment of gLite-3.0 has proceeded with minimal disruption to the production service. However, the time taken to resolve the software and integration problems meant that the release was not usable as an experiment test environment on the pre-production service as had been scheduled.

The release mainly consists of upgrades to existing components, with only a few major new components: the gLite Workload Management System – WMS (resource broker and new gLite CE) and a new File Transfer Service – FTS. The FTS had been tested in production as part of the data distribution test during April. An earlier version of the WMS had been tested over a long period through the ATLAS Task Force.

The new WMS co-exists with the previous Resource Broker and CE and it can therefore be introduced progressively into production over the coming months.

4.4   SRM Interface and Implementations

The SRM V2 coordination activity initiated at the Mumbai workshop in February has obtained an agreement between the experiments, developers and major sites on the functions to be provided for the initial LCG service and on the way in which these will be implemented within the SRM standard. An implementation plan has also been agreed for what is now defined as SRM v2.2. This is expected to lead to support by the three mass storage systems used in LCG (dCache, DPM, Castor) at the end of September. This should in turn lead to production services in the first quarter of 2007, about three months later than in the previous plan. However, it is clear that the requirements and interoperability issues are now much better understood than they were at the time of the Mumbai workshop.

A reviewed schedule has to be developed for the deployment of SRM v2.2 in production at sites and for the ramp-up of capacity. This should lead to a series of new Level-1 milestones (IS-x) and these should be translated into individual site targets and milestones.

All production and general user work at CERN was migrated to Castor 2 by June 9th. The CERN Castor 1 service for LHC will be removed on July 17th. Durable and permanent storage classes will be implemented at CERN via the SRM 1.1 interface following changes to the Castor TURL format (which will also be used long term.).

CASTOR and dCache (DPM already provides it) will also implement VOMS roles and Access Control Lists before next release in November.

4.5   Applications Area

The Applications Area projects have continued to support the experiments in their preparation of the software releases that are going to be used in the various data challenges and productions. Several iterations of the software have been made available in various configurations to allow experiments to integrate the new functionality and be ready with a production quality release by this summer.

New 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 AA.

ROOT - In the ROOT project many developments have happened in the integration of the C++ interpreter (CINT) and the C++ reflection system (Reflex). It is planned to release the new 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 being done in the PROOF system as the result of the serious testing done by ALICE. Important performance improvements are being introduced when accessing remote files. The first batch of these improvements is available in version 5.12.

POOL - 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.

Simulation - 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, which approximates the cross-section and models with higher precision the final states for projectile for Hydrogen (energies T=100 MeV to 2.0 TeV), and for d and 4He targets (T=30 MeV to 900 GeV). 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.

4.6   Distributed Database Deployment – 3D

The last quarter was mainly focused on the first throughput tests with the production database infrastructure at Tier 0 and Tier 1. On the FroNTier side the installation and installation tests (driven by CMS) went without larger problems. Several extensions on the FroNTier software side went in LCG AA releases of CORAL to implement a more efficient protocol and allow for retry and failover. The setup is largely ready for the CMS tests even though some of the CMS Tier 2 sites still have to complete their FroNTier/Squid setup (which has recently been requested by CMS also for Tier 2 sites).

On the database side several sites introduced delays, as their database setup was not fully usable when the throughput test started. As of today ASGC, BNL, CNAF, IN2P3, RAL have a working database setup. GridKa still needs to complete the database configuration. Main reasons are the learning curve for database cluster setup and the limited database administrator manpower coverage especially during the vacation period. Now all phase 1 sites have setup their database systems and apart from GridKa (some minor work to be completed by the site) all have been successfully used in throughput tests. Also two of the phase two sites (TRIUMF and SARA) have now started to actively participate in the 3D planning meetings and expect the requested database setups by October. For PIC and NDG we are in contact with the site management and hope to get a technical contact nominated soon to avoid a delay of the full production milestone.

During this quarter the Oracle license and client distribution issues have been resolved. The project has gathered license requirement from all software providers (experiments, grid middleware and Castor) and sites and centrally negotiated the required licenses for the sites. Now all WLCG tier 1 sites should be covered with licenses for Oracle software and direct access to Oracle support.

A 3D database administrator workshop has been organised and hosted by GridKa and a service review workshop is scheduled for 13-14 September.

4.7   ARDA

Task Forces Activities - During this quarter the activities connected to the task forces has progressed at a steady pace. We give here just a few highlights.

The analysis system of ALICE was presented to the users (physicists), also with a series of tutorials. The feedback of the users is very important (15-20 active users) and the system is the result of the integration of the original ALICE system with contribution of ARDA and EGEE middleware. In parallel, production activities are continuing on LCG resources.

Another very interesting example is CMS. In this case the analysis effort (negligible only 1 year ago) has doubled since end 2005 and reached 6k jobs/day. This activity was part of the finalization of the CMS physics technical design report.

ATLAS is integrating their production chain with the new data management, implementing the cloud model (production tasks assigned to a given Tier1 and its associated Tier2s) and with the gLite 3.0 WMS.

LHCb is continuing and the main activity is the support of the preparation of DC06.

CMS and ATLAS are very active in the context of the EGEE TCG (Technical Coordination Group) on job priorities: a common set of requirements has been defined. A first proposed implementation is based on mapping different VOMS groups and roles to different local UNIX accounts and groups.

Ganga - (http://cern.ch/ganga) development continued at a steady pace. A series of stable releases (4.1.x series) have been made public, with so far 4 bug fix/enhancement releases. This release series started the migration to a new internal structure of adapter classes. One of the most important changes on the user visible side is the support for splitting jobs in Ganga.

The main development effort is now on version 4.2, with already 2 public beta releases. The public beta releases feature the possibility to merge job result and a plotter object to create pie- and bar charts inside of Ganga. Enhancements for the ATLAS DQ2 system have been implemented as well as the possibility for LHCb users to query the LHCb bookkeeping database for input data sets of their jobs. Ganga was used as

the basis of a demo at the EGEE Phase 1 final review. Tutorials were held for both ATLAS and LHCb communities, reaching several tens of physicists.

Experiments' dashboard - The new version of the CMS dashboard for job monitoring (in development during the previous quarter) was released in the beginning of April (emphasis on reliability and performance). In addition, a new data source – Imperial College Resource Broker Real Time DB – was added in order to complement Grid job status information, which was originally received through the R-GMA system.

As requested by the CMS group running load transfer tests, a load transfer test monitoring framework was developed. The framework reads the information directly from the Phedex DB, which is a CMS-specific tool to manage data transfer. The load transfer information is presented in a user-friendly format on the CMS Dashboard. Both job monitoring and load transfer test monitoring are currently in production, and they are widely used by the CMS community.

The experiment dashboard is available (as preview) for ATLAS. The system reuses most of the components originally developed for CMS plus ATLAS-specific components. The system has been presented recently and it was well received.

Job Reliability - First results shown to the LHCC reviewers and discussed with GD group (operation) and with the middleware development teams. Plan agreed (concentrate on big sources of instabilities, improve the logging of the various systems, publish performance statistics vs. time). These data have been presented by P.Saiz in the last EGEE TCG (Technical Coordination Group).

5 Experiments

5.1 ALICE

Progress in AliRoot: The very first reconstruction of real data (cosmic trigger) from the TPC has been successfully achieved. The raw data as well as the reconstructed events are successfully visualized with the newly developed event display. The implementation of the raw data format in the simulation and reconstruction algorithms is being finalized. Consistency checks are being performed between the real raw data format and the software format. The first day geometry of the ALICE setup has been implemented. The implementation of the alignment procedures is progressing including the implementation of an ideal geometry, the misalignment deduced from survey data, and the software corrections for residual alignment. The calibration strategy has been defined as well online as offline. The user requirements are still being refined. The online-offline interface (the SHUTTLE program) has been implemented and released to users for the implementation of the detectors specific algorithms.

Fast analysis: a prototype (40 worker nodes) for the CAF has been configured for PROOF. Tests and debugging is ongoing by selected experts. The service has been opened for more test users. An analysis task force has been set up to validate the distributed analysis model and to train users.

Progress in distributed computing: All Tier1s and the main Tier2 have been installed and are operational. The installation of smaller Tier2s is in progress. So far, only 80% of the resources pledged for 2006 are made available. Resources are mainly missing in the Tier2s (less resources available than pledged, not properly installed sites, sites not responding, poor network bandwidth). Most of the sites have installed gLite 3.0. For the production, the WMS of LCG 2.7 is still in used. Tests of the gLite 3.0 WMS have started in parallel.

Physics Data Challenge: The pre-production of PDC06 is running since April for debugging the AliRoot software, the Grid middleware, the AliEn interface and services and the sites installation. The production has started on July, 1 with a weekly average value of 1100 jobs running in parallel and a peak at 1889 jobs. A smooth operation is not yet reached due to failures of LCG and AliEn services.

5.2 ATLAS

SC4 Tier-0 and data distribution tests to all Tier-1s and some Tier-2s started on 19 June and are continuing in July.

Global data transfer rates can be achieved but a few problems of system stability (to be addressed before next extensive testing period in September-October) have been uncovered. Data transfer subscriptions and communication with the local LFC catalogues cause overloading of the VO boxes at Tier-1s.

Software release 12.0.0 was built on 26 May; the first bug-fix release 12.0.1 was built on 1st July. Release 12 contains the functionality to generate misaligned and miscalibrated events for CSC tests, in addition to many other improvements to detector geometry, event data model, simulation and reconstruction. It is also the first release built out of separate projects; this is a stepping stone towards the simplification of the build and testing procedures.

Milestones that were marked as incomplete in the 2006 Q1 report (ATL-CSC-05-01 and ATL-REL-06-01) have been completed during this reporting period. Validation for release 12.0.X is ongoing, and it is expected that early production experience will result in further bug fixes. However, this possibility is incorporated into the planning for CSC Physics Sample production.

5.3 CMS

To demonstrate that services and facilities were ready to complete SC4, the CMS Computing Integration program implemented a series of site validation activities that are being tracked within CMS.

The computing systems began formal validation at a large scale at the beginning of May. CMS has validated roughly 75% of the sites that volunteered to participate in SC4, the list of validated sites includes all CMS Tier1's and 20 Tier2's or perspective such. The number of sites available is sufficient to meet the goals of CSA06.

Also a number of technical items have been validated. An initial version of the MC production tool , of the data management catalogues (DBS/DLS) and of the user job submission tool (CRAB) running on new framework data. The PhEDEx tool is in place and operational, transfers across all the sites are being exercised as part of a continuous activity. A tool (JobRobot) and an activity have been put in place that continuously run analysis jobs on the data replicated at all sites. A tool (CMS Dashboard) has been deployed and interfaced with the user job submission tool to monitor and track all job activities from a central location allowing relevant metrics to be measured.

There is an active team of people to carry on SC4 operations, which comprises a central team and contact persons at each site. For all the sites involved in SC4 we have well established communication channels, with name of local people, and continuous communication via mailing lists, phone calls and weekly integration meetings.

5.4 LHCb

The final alignment strategy proposal was completed. Other tracking sub-detectors have shown through studies that the strategy adopted by the VELO can be used by them. A tracking workshop to discuss progress was arranged, and LHCb will also participate in the cross-experiment alignment workshop.

Production activities for DC06 started during this quarter. In general the production runs smoothly. There was a problem with the use of GridKa which was solved in close collaboration with the site. LHCb procedures for production submission need to be speeded up. Major outstand problem is due to "flickering" of the LCG information system; this causes LHCb jobs to flood a site. Preparation for DC06 data processing continued but there is delay until July for the commencement, both on our side and the readiness of LCG.

Tests on the gLite WMS have been carried out. The results were disappointing with preliminary tests indicating an efficiency of only about 50%. Investigations will continue.