



LHC Computing Grid Project

Quarterly Status & Progress Report

2005 Q4 – Executive Summary

Alberto Aimar
Les Robertson
7 March 2006

1 Introduction

This is the first of the LCG quarterly reports using the new format, including the Tier-1 centres, and emphasising the progress in comparison with the detailed medium term planning. The format of the report was designed to make it easy to be completed by the sites, experiments and projects, but in practice many of the contributions were late and the analysis phase was also delayed. While the Christmas break and the Mumbai CHEP no doubt contributed to the delay, we must do better next time, with a target of completing the process within 4 weeks of the end of the quarter.

The milestones in the report are as they were at the end of 2005. Changes requested in the report will be implemented before the end of March, along with significant changes to the future milestones that result from the completion of the planning for Service Challenge 4 (SC4), which has been completed during the past two months.

2 Level-1 Milestones

The criteria for milestone SC3-2 - *successful completion of Service Challenge 3 (SC3)*, were agreed at the GDB in October 2005 as follows:

5 Tier-1s and 5 Tier-2s must have achieved the following targets:

- (a) appropriate baseline services operational*
- (b) availability better than 80%*
- (c) success rate of standard application test jobs greater than 80% (excluding failures due to the applications environment and non-availability of sites)*

By the end of SC3 all Tier-1 and around 20 Tier-2 sites were active.

The measurement framework for the last two criteria, the Site Functional Test (SFT) framework, was still being developed during SC3, and was only in operation for the EGEE grid. Using the SFT tests, by November the average availability was better than 80% for *all* sites in the EGEE grid (not just the LCG SC3 sites) on a 24 hour basis. Some of the experiments do not yet have standard test jobs integrated into SFT, and there is at present no consistent measure of failure rate for production jobs. However, taking all jobs submitted, ALICE, ATLAS and LHCb report grid/system failure rates of less than 20%, and analysis of the EGEE Resource Broker logs (by D.Colling /Imperial College) shows for these experiments significantly less than 20% grid-visible failure rates.

It is therefore considered that the project met the criteria for this high level milestone.

The other Level 1 milestones in the quarter were also completed:

- OPN-1 – 10 Gbps network in operation between CERN and 3 Tier-1s (FNAL, SARA, IN2P3)
- DRC-2 – 750 MB/sec data recording demonstration at CERN (950 MB/sec was achieved).

3 Summary of Major Issues

3.1 Service Challenge 3

This was the major activity of the quarter. As noted above SC3 demonstrated a marked improvement in operational reliability and availability over the situation at the end of 2004, and the criteria for the completion milestone were achieved. On the other hand, all experiments experienced start-up difficulties,

LCG – Quarterly Report – 2005 Q4 – Executive Summary

particularly as they began using the new version of CASTOR at CERN. There was also a general feeling by the experiments that many of the tools available were not at the level of functionality, reliability and performance that they expected. This has led to a much more formal process for planning SC4, including a clearer definition of the software to be delivered, insistence on adequate time for testing, including a 6 week period for experiment *beta* tests, and delaying the introduction of new components until after SC4. This results in a one month delay of the SC4 service phase, now scheduled to begin on 1 June 06.

The milestone for SC3 throughput (SC3-3: Tier-0→Tier-1 data distribution) was postponed until February 2006 as it could not be scheduled during SC3.

3.2 Progress in Tier-1 Centres

The full report provides details of the progress, achievements and outstanding issues for each of the Tier-1 centres. Almost all of the milestones planned for the quarter were completed. There were delays in equipment deliveries and network provisioning at a number of sites, but these are not expected to impact preparations for SC4.

There has been progress in interoperability between OSG and EGEE, each grid now making its LHC resources visible to the other grid through the information service (BDII). This enables job submission in both directions. Closer collaboration and compatibility on operational procedures is under discussion. There has also been progress in interoperability with NDGF, which now also publishes its resources via BDII.

3.3 Castor 2

All experiments are now using CASTOR 2 for service challenge activities. In most cases comprehensive testing was carried out only as the experiment started the service challenge, but the problems encountered were almost all resolved within two weeks. CASTOR 2 has also been used for the general production of some experiments, but the schedule for full migration from CASTOR 1, which had been agreed in the summer, had to be re-scheduled for the first quarter of 2006 as the experiments did not have time to do this during SC3.

The data recording tests at CERN at the end of the year demonstrated that CASTOR 2 is a major improvement over the previous version in terms of load balancing and error recovery, as well as showing that it is capable of achieving 50% of the nominal LHC recording requirement.

The Tier-0 buffer performance test at 500 MB/sec (milestones DR-1, DR-3) has been postponed from December 05 to February 06, when it will be done in collaboration with ATLAS Tier-0 tests.

Deployment of version 2 at all the sites using CASTOR is planned for Q1 2006.

3.4 Distributed Database Deployment – 3D

An agreement on the architecture and technology for the distributed database services was agreed in October, and the testing and deployment plan was agreed before the end of the year. The timescale is tight, with the conditions database as the initial application. Several sites have their database clusters already installed (BNL, CNAF, IN2P3) and other sites (RAL, GridKA) are awaiting the delivery of database hardware.

The *FroNtier* and *squid* evaluation is being coordinated by CMS at 10 tier-1 and Tier-2 sites. The Oracle *streams* based replication tests involve ATLAS, CMS and LHCb.

3.5 VO Boxes

As agreed, VO Boxes were deployed at many sites during SC3 for evaluation, and were successfully used by ATLAS and ALICE, and by LHCb at CERN. CMS used similar facilities at some sites, but these were agreed directly with the sites. Many sites have major reservations about the long-term support of VO Boxes, particularly from a security standpoint, and a workshop was organised for January 2006 to examine the issues in the light of SC3 experience and make proposals of how to move forward.

LCG – Quarterly Report – 2005 Q4 – Executive Summary

3.6 Applications Area

During this quarter the main work of consolidation of the AA software has continued, as well as the support for production deployments in the experiments.

The migration to the new Reflex library for the persistency framework (POOL) and Python scripting has been completed. Several iterations of POOL have been made available to the experiments to validate the changes. The final version of Reflex and Cintex has been released as part of ROOT 5.08 in December and will be used instead of the SEAL ones in the coming weeks. At the same time the appropriate components will be removed from SEAL releases and this will be an important milestone towards the completion of the SEAL and ROOT merge.

The first public release of the new re-engineered version of the relational database API package (CORAL) was made available. The adaptation of POOL and COOL (conditions database) to this new package is ongoing and finishing soon. This will be on time for the experiments to be integrated in their production software to be used in this year's major data challenges.

End of the year production versions have been released of ROOT and Geant4 which include a long list of new functionality required by the experiments.

3.7 The Baseline Services Working Group

A second round of the Baseline Services Working Group took place during the last quarter of 2005, covering extensions to the services agreed in the first round (and reported in the LCG TDR) and taking account of the experience gained using the baseline services in SC3. This round did not lead to a consolidated report (milestone GD-4) but instead will provide detailed input to a series of specific actions, including: the general planning for SC4; the VO Box decision process referred to above; prioritisation of additional EGEE middleware functionality through the EGEE Technical Coordination Group.

3.8 ARDA

The milestones due in the quarter (1.6.11.1/4 - usage of a new version of gLite middleware by each of the experiments) were not completed because the gLite version 1.5 is only becoming available during January 2006. On the other hand, the demonstrations of the different prototypes at the EGEE conference, the LHCC review, and at Supercomputing show the evolution of the prototype and some level of integration with the production systems (e.g. GANGA/DIRAC, CMS dashboard aggregating data from user jobs, production system and SC3 jobs).

ALICE - The different components of the ALICE prototype using PROOF are being integrated together. A preview was presented to the LCG Comprehensive review, showing examples of user analysis, and a demonstration and paper were presented at SuperComputing 2005.

ATLAS - A preview of the work in ATLAS was shown during the LCG Comprehensive review, with examples of user analysis via the production system and Ganga. Within the ATLAS task force, ARDA has contributed middleware studies (WMS bulk submission) and some preliminary work on the new Lexor version (together with the INFN Milano group).

CMS - The activity with CMS is continuing as foreseen, and the combination of the CRAB and ASAP analysis systems were submitting about 150K grid jobs/month by the end of the year. The next version of CRAB will integrate the main ASAP features. The CMS dashboard activity is continuing, notably focusing on improving the performance for complex querying and migrating the database back to Oracle.

LHCb - Ganga has been demonstrated both at the EGEE conference and in the LCG review. In addition developments and support of LHCb users has continued.

Metadata Activities - The AMGA system is now part of the gLite 1.5 release. The activity during the quarter concentrated on streamlining the testing process and providing additional functionality, in particular on security.

LCG – Quarterly Report – 2005 Q4 – Executive Summary

4 Experiments

4.1 ALICE

Testing of the ALICE Computing Model has continued during the Physics Data Challenge 05 executed together with the LCG Service Challenge 3. The central and distributed Grid services agreed by the Baseline Services Group have been adopted. LFC and FTS were partially tested, extensive testing of CASTOR2 and RB has been performed.

A complete set of requirements and a simple installation method of the ALICE VO software on the VO-box has been established and the scalability and robustness of the VO-box model demonstrated. MC generation and reconstruction have been continuously run during 4 months, involving 4 T1s and 10 T2s. A maximum of 2,450 concurrent jobs has been reached, evenly distributed over the participating T1s and with a good T2 involvement. The job failure rate was low: below 2% for the application software, below 15% non-persistent failures of Grid services. The majority of produced data has been stored in CASTOR2@CERN, which showed a good stability and continuous operational improvement over the 4 month running period. *xrootd* has been extensively used as a file transport protocol. The application software (ROOT, AliRoot) was debugged in detail and optimized for CPU and memory consumption.

The frequent updates of the middleware and services prevented some computing centres from taking part in PDC'05, and also introduced unevenness of resource utilization. In terms of operation, priority for updates was given to the centres in the SC3 plan. Not all services, in particular LFC and FTS, could be tested in depth. Not all sites could be switched to the VO-box model of operation, mainly because of lack of experts, who could do the installation of the software (in the absence of full documentation). Several new functions were identified as being required (e.g. additional query functions in CASTOR 2; high-level service for FTS transfer scheduling). Not all centres can cope with the high demands of the application in terms of memory utilization: ALICE requires worker nodes with at least 2GB memory.

4.2 ATLAS

Most attention in Q4 2005 was dedicated to LCG Service Challenge 3. We ran in the SC3 context two major operations: data transfer from Tier-0 to all ATLAS Tier-1s, and internal Tier-0 data flow as specified by the ATLAS computing model. Both operations resulted in the identification (and resolution) of several problems and considerable progress was achieved. Further tests will take place in the context of SC4 in 2006.

Most open issues are related to the organisation of distributed systems (Data Management, Production and Analysis). The single architectural point that is still open, but to be sorted out in early 2006, is the issue of the deployment of experiment-specific services at Grid sites (VO Boxes).

4.3 CMS

The main milestone for this quarter was to be ready for the Service Challenge 3, and to run SC3 in phases. CMS had specified as its main goal, to establish the "baseline use case" (described in detail in the C-TDR) of moving data from the T0 tape system to T1 centers, onto tape for "custodial storage". Later, "selection jobs" would run on this data at each T1 center, and "selected data" would be moved to T2 centers for fake "data analysis" jobs. At the same time, "simulated data" from T2 centers would move to the T1 centers for storage and further distribution. Also exchange of datasets between T1 centers was planned for "service phase 2", but this was de-scoped due to the problems encountered during SC3. An initial assessment of the CMS results from SC3 is available.

The new Data Management system with its components Dataset Bookkeeping System (DBS) and Dataset Location Services (DLS) was scheduled to be initially deployed in November, but due to manpower shortage (in particular loss of the key developers) this was delayed to early 2006. This eliminates all contingency on delivering the MC production system and data management system in preparation for SC4.

The new MC production system that allows to run MC with the re-engineered framework was designed and the prototype development was started, with the initial release now scheduled for early 2006.

LCG – Quarterly Report – 2005 Q4 – Executive Summary

The re-engineered EDM and framework is being delivered: The re-factoring of the simulation, reconstruction, event filter and visualization Software (CMSSW) in the context of the new framework and event data model is well underway. The release of version 0.3.0 will be the base for the Cosmic Challenge data-taking. This release provides a fully functional system, from the readout of the detector up to local reconstruction of nearly all the subdetectors, and Geant4 simulation.

This quarter saw a lot of discussion of the role of application specific services at regional centers and the use of VO boxes to operate these services. CMS has written a "White Paper" on the issue.

4.4 LHCb

The three milestones due in the 3rd quarter of 2005 were completed. This included running analysis jobs successfully at all Tier-1s, with some limitations at sites without a storage element with an SRM interface.

A major review of DIRAC was completed. The implementation of the revised event model was begun, using libCore and the new Linear Algebra. The Conditions Framework was released.

The grid middleware lacked some of the expected functionality. LHCb software releases were delayed due to late delivery of the LCG math library.