Alberto Aimar
16 May 2006

LHC Computing Grid Project
Quarterly Status & Progress Report
2006 Q1 – 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. The milestones in the report are as they were at the end March 2006.

The format of the reports is designed to make them uniform and easy to be completed. Compared to the previous quarter, when the new format was used for the first time, the format of the reports has been slightly modified following the lessons learned and the comments received. Important sections have been added in order to allow all sites to report about their exact "CPU, disk and tape capacities" installed and the "status of VO Boxes" installations.

All reports were reviewed by a Review Team (A.Aimar, H.Marten, L.Perini, L.Robertson and D.Salomoni). The reviewers then asked for additional information and in some cases the reports were modified and re-submitted by the original authors.

This process took about six weeks, with an improvement of about three weeks in the preparation of the reports. But still many contributions were late or incomplete. The analysis phase by the Review Team was also delayed because of the Easter holidays. The target of completing the process within 4 or 5 weeks of the end of the quarter seems achievable in the next quarters.

2 Level-1 Milestones

These are the WLCG Level 1 milestones that were completed during the quarter:

- *SC4-1: All required software for baseline services deployed.*

All software for baseline services was operational at all Tier-1s and at least 20 Tier-2 sites by the end of SC-3. All sites have been upgraded with the LCG 2.7 distribution at the end of January and in next quarter they will be updated to gLite 3.0.

- *OPN-2: Tier-0/1 high-performance network operational at CERN and 6 Tier-1s, at least 3 via GEANT.*

The high-performance network was operational at CERN and 6 Tier-1 sites. FNAL, SARA, IN2P3, TRIUMF, BNL and CNAF were connected but, as anticipated in the last quarterly report, only two sites (CNAF and FZK) were connected via GEANT.

- *SC3-5: Proposal on availability levels specified in Annex 3 of the WLCG MoU.*

The algorithms to calculate availability and problem response times, as in Annex 3 of the WLCG MoU, have been agreed taking account of sites that do not provide a 24 hours service yet.

- *SC4-2: Use cases and service level support defined for SC4.*

The use cases and service level was defined by the outcome of the Mumbai Service Challenge workshop. The agreed Tier-0/Tier-1 bandwidth targets and the site reliability measurement process were approved by the WLCG Management Board (7 March 2006).

The following milestones had to be delayed:

- *CAS-1: Castor2 Readiness Review.*

The review is postponed to early June. The review was initially re-scheduled for April, but the chair of the review resigned. A new chair has been appointed and new dates agreed.

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

The monitoring of the services has been implemented, a general tool checking site service availability is being deployed and the set of tests has been agreed, but completion of the milestone is postponed to end of April 2006.

3   Service Challenges Progress

The Service Challenge 3 throughput re-runs – both disk-disk and disk-tape – were successful in their objectives of demonstrating improved rate and stability. In individual tests, many of the sites demonstrated rates at or even above their nominal rates. The re-run of disk-disk transfers reached a sustained aggregate rate of close to 1GB/s in January. The subsequent disk-tape tests successfully demonstrated rates to tape exceeding 50MB/s at each of the 6 sites participating. The disk-tape tests did not go as smoothly as the disk-disk tests, and several sites did not participate.

Whilst the steps required to ramp-up these data rates to the full nominal values at all sites – adding the additional complexity of full Tier-0 and Tier-1 activities in parallel – should not be underestimated, this work nevertheless represents an important milestone in the preparations for full-scale LHC Computing services.

During the SC3 re-run and the preparation of SC4 several issues in the File Transfer Services were uncovered and addressed. The Gridview monitor was also improved and is now used regularly to monitor progress. In March several pre-SC4 tests have been performed with a new Castor-2 instance at CERN dedicated to this and several initial tests to most of the Tier-1sites have been run.

The SC4 Preparation Workshop took place in Mumbai on 10-12 February. There were detailed presentations and clarifications of the experiments' goals for SC4. Significant progress was made towards agreeing the details of functionality that would be available in the middleware to be deployed for SC4. In particular the services to be provided, the timetable, the experiment use cases and the priorities for data management developments were agreed.

Progress has been slower than planned in certifying the SC4 middleware distribution, gLite 3.0. The SC4 planning foresees a 6-week period for testing by experiments on the Pre-Production Service (starting 15 March) followed by one month for installation and set-up at the participating sites. By the end of the quarter the package was still not available on the Pre-Production Service for testing, and re-scheduling will be necessary if the start date of the SC4 service phase (milestone SC4-4: 1 June) is to be maintained.

4   Summary of Major Progress and Issues

4.1   Progress in Tier-1 Centres

The full 2006 capacity goals were originally scheduled for the beginning of April, the date by which the full capacity should be installed when LHC is running. In view of the slower build up of demand planned during 2006 it was agreed in the MB that this date should be moved to 1 September. Progress looks good for building up processor and tape system capacity at most sites, but the installation of disk capacity is going much more slowly. A review of experiment needs for disk storage at Tier-1 centres during the second half of the year is needed.

The connection via the OPN network is now operational on most sites. The configuration for Tier-0/Tier-1 transfers as been tested during the preparation of SC4. Some sites are also testing their network connectivity and configuration to the Tier-2 sites. In the next quarter the use cases of the experiments will define which network transfers are going to be established and optimized.

In the next quarters the sites will have to prepare themselves to become able to run in standard operations mode, preparing for stable operations, 24x7 site support and interventions scheduling. An internal review in early June will assess the situation and details will be provided in next quarter's reports.

4.2   VO Boxes Installations

VO Boxes were deployed at all Tier-1 sites accordingly to the experiments needs. Many sites continue to have major reservations about the long-term support of VO Boxes, particularly from security and scalability perspectives.

Workshops on VO Boxes took place in January (Amsterdam) and at the very beginning of April (Rome) and the agreement reached was that all Tier-1 sites should install the VO Boxes. Meanwhile the general middleware software would be adapted to address the needs that are currently solved by the specific VO Boxes. During this period the experiments agreed to freeze their requests for new features and changes regarding their VO Boxes configurations. Following the Rome workshop a report will be drawn up formalising these agreements and laying out a plan for middleware developments.

4.3   Castor 2

The Castor 2 system has performed well during intensive Tier-0 testing at CERN during the quarter, both in terms of performance and reliability, with the data recording milestones (CERN-DR-2, CERN-DR-5) easily reaching their targets. An ATLAS Tier-0 test demonstrated their nominal throughput, driven by a 320 MB/s input stream, though only sustained for a 24 hour period.

Production of three experiments using CASTOR2 (CERN DR-4), after having been postponed several times, is now completed. The migration of the experiments data to Castor2 (CERN DR-7) is now scheduled for April for the ALICE, ATLAS and CMS experiments. LHCb awaits simultaneous and efficient support of "durable" and "permanent" storage classes (as defined in Mumbai); this is expected during 2006Q2 (CERN DR-7).

Castor 2 with the SRM 1.1 implementation is available for installation on sites outside CERN. The SRM 2.1 interface for Castor2 was delivered on schedule by RAL. However, discussions at the Mumbai workshop showed that not all SRM 2.1 implementations were ready for production use and also that significant clarification was needed on the definition and use of different storage classes. Therefore it has been agreed to start SC4 using SRM 1 and aim to migrate to SRM 2.1, or a follow on, in October. The implementation of durable and permanent storage classes is nevertheless in progress (CERN DR-9).

4.4   SRM Interface Coordination

The process for coordinating the testing of the SRM 2.1 implementations for the three mass storage systems used in LCG (Castor, dCache, DPM) was agreed, with a target for production by the end of September. A coordination team has been established and should propose a program of work with milestones.

It was agreed at the Mumbai workshop that the introduction of SRM 2.1 would be delayed until after SC4, due to unavailability of the dCache implementation, the need for a common agreement on how to define storage classes, and to allow a thorough test plan including interoperability testing to be implemented.

4.5   Site Monitoring and Metrics

The metrics for operations reliability and availability monitoring were defined and implemented in order to allow the verifications of the criteria for success specified in the MoU.

A new version of the Site Functional Tests (SFT, now renamed SAME) has been developed and the majority of the metrics measurements are available and under use. The metrics data are stored in a central database and the results can be exported to Excel for initial measurements and display. The first full version of the metrics system is expected for next quarter and more tests will be added. Some of the experiments do not yet have standard test jobs integrated into SAME, and there is at present no consistent measure of failure rate for production jobs.

4.6   LCG Middleware Releases and gLite 3.0

The LCG-2.7.0 distribution was released at the end of January and deployed to the majority of the EGEE sites in February and March. A 64-bit version of LCG-2.7 was available in March. This version of the middleware is the fallback solution for SC4, assuming that FTS, storage manager, and data management tool upgrades can be provided.

The preparation of gLite-3.0 began at the end of January. This is intended as the first converged gLite and LCG release. The readiness for SC4 in June is possible but not certain. The basis of this version is LCG-2.7 and the gLite-1.5 code release. The latter was not released until the end of January; integration with LCG-2.7 began immediately and many issues of compatibility, dependencies, and different base versions were discovered (see details in the Deployment Area report). By the end of the quarter the majority of problems have been resolved or it has been decided that they were not showstoppers for this release. However, at the end of the quarter there are still major uncertainties in the behaviour under stress of bulk job submission.

A high job failure rate and instability was observed on the Pre-Production Service (PPS), with many underlying issues, not only of the middleware itself, but problems in configuring and running these new services at the sites.

4.7   Applications Area

The main activity during this quarter has been the preparation of the software releases that are going to be used in the various data challenges and combined test runs of the LHC experiments during this year. About half of the functionality of SEAL has been completely migrated to ROOT and the experiments and the AA projects had made considerable effort in adapting their software to use the packages that have been migrated.

A detailed plan has been started to be prepared for the migration of the second half of the functionality. During the quarter many different releases and software configurations have been produced to help the experiments for the preparation of their production releases.

POOL products such as CORAL and COOL are coming with new functionality requested by the experiments. ROOT has made one development release including new functionality plus several bug fix releases, Geant4 has made various patch releases to support the experiments, and new versions of GENSER and GDML have been released. The HepMC package was installed in the LCG external area and is maintained by FNAL effort.

Concerning architecture and platforms, the Applications Area software is being adapted for the AMD64 platform in the framework of the certification and preparation for the new Linux SLC4. The next releases of the AA software will also be made available for this new platform.

4.8   Distributed Database Deployment – 3D

During the last quarter the database and Frontier setup for the pre-production phase has been largely completed. Progress is monitored during the weekly 3D phone meetings and two database workshops (in February at CERN and RAL, the second dedicated to database administrators).

All Tier-1 sites have made available small database clusters for the next phase. The Tier-0 database - an increase of database cluster capacity by a factor of two and a redundant 3 node Frontier production system - is now setup and ready for larger scale deployment.

The experiments have continued to refine their database models for condition data and all software frameworks now integrate the required conditions, but it is worrying that condition data models for several important sub-detectors are still missing.

CMS has proposed that 3D should now take over the follow-up of the SQUID installations at the LCG sites. This will be done in a combination of 3D meetings and LCG service meetings as the responsibility for SQUID is on some sites with the database teams, on other sites with the system administration teams.

Several sites noted that the service review workshop (milestone DBS-7) may in fact come too late during their budgeting or acquisition cycles for any larger change in their database capacity for the October service.

4.9   ARDA

During this quarter ARDA had no specific new milestones but the new program of work was in preparation and is now being discussed with the experiments. There is an agreement on the content and the detailed

timetable, with milestones and decisions points, is in preparation. It will be presented to the LCG Management Board in the next quarter.

During the current quarter the main duties were "experiment integration and support" activities, which have no fixed milestones but required a continued effort. The ARDA team concentrated in integrating components in the experiments systems and supporting the evolution of existing ones. Finally wherever appropriate (ALICE, ATLAS and LHCb for their end-user analysis system) ARDA effort has been used to prepare good documentation and in tutorial efforts.

AMGA, the metadata system developed in ARDA is now part of the gLite software stack. The system is made available for GANGA, for evaluation as a bookkeeping system in LHCb and to other non-HEP users.

The work on GANGA continued (automatic job splitting, logical folders, etc) as well as the activity of tutorials and demos (for ATLAS and LHCb). In addition GANGA has been demoed at the EGEE User Forum and it will be part of the EGEE final review.

The database of the Experiments Dashboard, for the CMS prototype, runs now as an Oracle service at CERN and has a new optimized and improved schema. All the data collection and monitoring of data handling within the dashboard has been redesigned in order to avoid the limitation of the first version. This new CMS dashboard is now capable to collect data also from PheDex and is it expected to be in production in April 2006.

5 Experiments

5.1 ALICE

The AliRoot code is reaching completeness with the final implementation of the raw data format, calibration procedures and alignment procedures using the ROOT geometric modeller. Detailed studies on the impact of misaligned geometries on the tracking performances in the central tracking systems and the MUON arm are well advanced. The overall calibration procedures including online calculation of calibration parameters are examined detector by detector. The program, called SHUTTLE, for collecting online parameters to make them available in the offline Calibration Data Base has been implemented. A complete documentation of AliRoot including a primer for new users has been released.

The distributed computing environment has reached sufficient maturity to make it available to general users. It incorporates the services distributed through the LCG middleware and ALICE specific services not available elsewhere. VO boxes hosting the ALICE services have been installed in all Tier1 sites and all Tier2 sites pledging resources to ALICE. The production phase of the Physics Data Challenge 06 will start on 10 April with the simulation of pp events defined by the Physics Board. Although all the utilities exist, distributed analysis has not yet been tested in a realistic environment and at a realistic scale. This testing will be a major challenge until the end of the year. A complete documentation and primer of the usage of ALICE distributed computing environment and distributed analysis has been released.

Distributed analysis based on the Parallel ROOT Facility (PROOF) and on the xrootd I/O subsystem will be the baseline for performing calibrations and prompt analysis on the CERN Analysis Facility during data taking and could be adopted by the Tier2 sites dedicated to mainly analysis. Preliminary discussions have taken place on how to operate PROOF at CERN, and exploratory work is under way on possible ways of integrating xrootd and CASTOR. A prototype system will be deployed on medium-sized farms at CERN, CCIN2P3 and FZK to be exercised by general users.

The computing resources pledged so far by Funding Agencies for ALICE are well below the needs documented for and validated by the Computing Resources Review. New developments in the EGEE/LCG middleware services are planned to be released as late as October 2006. The adoption of new services or implementations at such a late time could have negative impacts on the availability of stable software in due time for the first LHC physics runs. The personnel situation for offline computing remains critical, stabilisation of key personnel is mandatory to ensure availability and quality of services to users.

5.2   ATLAS

Integration of new software infrastructure components that are necessary for Computing System Commissioning operations (mainly ROOT5 for schema evolution, COOL 1.3 for the Calibration Data Challenge) took much longer than expected.

The implementation of the realistic detector geometry (including miscalibrations and misalignments) is also in progress. Release 12, including all these features, will be used for the large-scale CSC simulation production from May onwards.

Tier-0 throughput tests were run successfully in January as last part of the ATLAS SC3 activities; nominal rates were reached and held for >24 hours. Preparations for SC4 tests (Tier-0 and data distribution) later in April are making progress.

The functional tests of data distribution that had been foreseen by ATLAS as preparation for SC4 starting in late March did not succeed as the production FTS service was used at that time for throughput tests and the service on the pre-production service was not stable enough. This lack of preliminary testing will generate a delay later on during the year as a larger part of the ATLAS software will have to be tested at the same time within a limited time window.

5.3   CMS

Details of interactions between new CMS Software Framework, Data Management and Production Workflow are being defined. A prototype now exists for Data Management catalogs and the Monte Carlo Production Tool.

The user's analysis tool, CRAB, is being ported to the new Software Framework. The Data Transfer component, PhEDEx, is being modified to use FTS as underlying file transfer tool. The CMS Tier-0 project has started.

SC4 middleware (gLite 3.0 stack) were not usable on the preproduction system during the quarter. SRM and CERN Castor2 instabilities during completion of Physics TDR Monte Carlo production caused significant operational load and delayed setup of tools for SC4 data transfer activities.

5.4   LHCb

The VELO analysis group produced a report detailing their alignment procedure; the other tracking detectors are now also studying their alignment strategies. Work has commenced understanding how to perform the global alignment; a discussion document is expected in April.

The LHCb Workload and Data Management System (DIRAC) was reviewed for the 2006 Data Challenge. The other tools are being finalised to allow automated processing to be triggered as data files become available.

Core developments in Gaudi now make use of the developments in the LCG application area, following the merging of ROOT and SEAL. LHCb has retired the use of the CLHEP library from the LHCb-specific parts of the applications (though it is still necessary for GEANT4) and LHCb has contributed to testing and improving the new ROOT geometry and Linear Algebra packages.