

CERN-RRB-2010-036

23 MARCH 2010

PRINCIPAL LHCC DELIBERATIONS

17TH MEETING OF THE COMPUTING RESOURCES REVIEW BOARD

20 APRIL 2010

EMMANUEL TSESMELIS

SCIENTIFIC SECRETARY, LHCC

GENERAL

This document summarises the principal LHCC deliberations concerning the World-wide LHC Computing Grid (W-LCG) Project at the Committee's session in February 2010.

The LHCC considers that the W-LCG has made excellent progress in all aspects and congratulates the W-LCG and the experiments on the performance of the Grid with first LHC collision data and analysis, while recognizing that it has yet to be stress-tested under a high LHC data rate.

CONCERNS FROM THE PREVIOUS COMPUTING RESOURCES REVIEW BOARD

| SUB-SYSTEM | CONCERN | STATUS |
|---------------------|--|--|
| Resources for ALICE | The pledged resources do not fully match the ALICE requirements. | Computing resource availability remains a long-standing problem for ALICE. |
| Job Submission | Low average job submission success rate. | Improved job submission success rates were reported. |

STATUS OF THE W-LCG

The LHCC referees and the Computing Resources Scrutiny Group (CRSG) held a joint review of the W-LCG project on 16 February 2010, and heard reports from the computing groups of the four main LHC experiments. This was the first combined LHCC-CRSG review, intended to make the process more efficient. The LHCC referees thank the CRSG, the experiments and the W-LCG groups for their cooperation in organizing this.

Peaks of 650k jobs/day and data transfer rates of 1 petabyte/week (during STEP'09) have been achieved, with increasing reliability levels at most sites. Overall installation of CPU and tape resources in 2009 kept pace with pledges, but delivery of disk storage was delayed at some sites. The usage of these resources is still well within the capacity, because the data available in 2009 was below that expected when the resources were planned. The average CPU efficiency (CPU/wall-clock ratio) has improved to about 70-80%.

Most sites met or exceeded performance targets set but with some minor problems and a few more serious ones. Improvement in processes for scheduling interventions have been implemented, especially at Tier-0. There were difficulties with two Tier-1 sites. One was already flagged following STEP'09, although two other Tier-1 sites discussed have since resolved their problems. There are also concerns about data access, support models for end-user analysis and networking.

The Committee supports recommendations made by the W-LCG Management for mitigating these problems. In particular, introduction of risk analyses (i.e. improved planning) as part of the scheduling for interventions, with monitoring afterwards; site visits by an external review panel following a major incident with follow-up 3-6 months later; and reviews of site storage, data access and user support.

W-LCG middleware has handled core tasks adequately and development is now targeted at improved control, recovery, simplification and stability. Applications software performed well, with quick responses to requests for patches from experiments. The project is entering the maintenance mode, while still evolving products as required. Reduced manpower is available for future work on middleware and applications.

The agreed pledges of resources for 2010 are mostly near to the requested levels, with the exception of the ALICE Tier-1 and Tier-2 site pledges, which are only about half of those requested. However, recent announcements of budget reductions from some funding agencies threaten to delay the delivery of these resources beyond the agreed deadline of 1 June 2010, which could impact the processing of the 2010 LHC data.

As the W-LCG project moves from development to its phase of regular operations, the monitoring of the project needs to change from formal milestones to metrics for tracking performance, reliability and scalability. Many of these exist already, but others need to be developed and reviewed regularly.

The proposed construction of a new Tier-0 Computer Centre has been suspended while various options are considered. Meanwhile, gains in power efficiency from aggressive replacement of older machines, and delays in the LHC schedule, have resulted in an additional year or so before the new facility will be needed, although a decision must be made later this year.

The transition from Enabling Grids for E-Science (EGEE) to the European Grid Initiative (EGI) is well advanced, but the outcome of several project submissions remains uncertain, with feedback not expected before March 2010. This is a concern for the level of funding of future operations, middleware and applications development, but does not represent a major risk for the W-LCG project.

Preliminary estimates of beam time, based on the new LHC schedules for 2010-2011, with a shutdown in 2012, were used by the experiments for planning requests for computing resources in 2010-2012. These have since been superseded by more realistic estimates, which will be used for more detailed planning. The new estimate for LHC beam time in 2010 exceeds that used in 2009 when planning requests for 2010 computing resources. It was agreed not to adjust the current numbers or to ask for any increases to the agreed 2010 pledges. That aim is that the final numbers to be presented to the Resource Review Boards in April 2010 will be agreed at the W-LCG Management Board meeting on 23 March, and with LHCC and CRSG representatives present.

ATLAS and CMS computing performed well on cosmic-ray data and Monte Carlo challenges and this preparation paid off during the data collection period in December 2009. Both experiments adhered to and exercised their full distributed computing models. All essential systems and processes are in place, including calibration processing and the application of data quality. There was good performance and utilization of all tiers. The metrics typically show improved performance and point to some systematic problems, which are being addressed. Both experiments observe that the 2009 data run was not a stress test but was a valuable learning experience. The experiments are well

positioned for sustained operations and higher rates and it is understood that new challenges will arise.

ATLAS and CMS are not currently resource limited, which enabled quick turn-around times for reprocessing and the ability to have extra copies of data sets distributed to facilitate access. Both experiments demonstrated participation in analysis at the level of several hundred users.

During the data collection and processing the Tier-1 resources were stable. While technically available, the ASGC Tier-1 site (~10% of ATLAS/CMS resources) continues to have difficulties with stable operations. CMS Tier-2 sites have achieved a plateau of 80% reliably stable sites.

To cope with a possible shortfall of resources in 2010, both experiments presented some belt-tightening measures and the intention to use resource management tools for disk and CPU. ATLAS showed a possible 5/8 reduction in the time per event for full simulation achieved with the new release of Geant4 and a reconsideration of storage of cosmic-ray data.

After intensive Monte Carlo production, allowing the full chain to be exercised, the LHCb effort moved to collision data in autumn 2009. The total resource usage in 2009 was as requested, but with less data reconstruction and fast analysis at Tier-0 and CERN Analysis Facility, and more Monte Carlo production at Tier-2 sites. First experience with real data was very positive, with fully automatic job creation, submission, and re-submission in case of failure. Seven sets of processing conditions were used. There has so far been no stripping, and DSTs were distributed to all Tier-1 sites. No problem with file transfers was encountered, with excellent Grid latency.

Over one million collision events were collected by ALICE and stored at Tier-0 on a large disk buffer. Only those passing quality requirements will be stored permanently. Two copies were spread over the Tier-1 sites. First reconstruction was completed one day after the end of data taking, and five re-reconstruction passes followed with regularly refined conditions. Some Tier-1 sites could not participate in that phase, because of excessive memory consumption of ALICE software, related to early non-optimal alignment and calibration corrections. This is expected to be resolved within a few weeks. Bulk analysis is organized into “trains”, with “wagons” under the responsibility of analysis groups. Twenty such train passes have taken place so far. End-user analysis is taking place at analysis facilities and in batch at Tier-1 and Tier-2 sites over the Grid.

Computing resource availability remains a long-standing problem for ALICE. The pledged resources for 2010 barely cover the needs, and some funding agencies have since announced further reductions. Even without these reductions, the deficit in terms of disk storage at the Tier-1 and Tier-2 sites amounts to 40-50% in 2010. In 2011, the deficit of computing capacity at Tier-1 sites is as much as 40%, caused essentially by the processing of the heavy-ion data to be collected at the end of 2010. Note that ALICE evaluates computing resources according to calendar years. The Collaboration is encouraged to move to the W-LCG accounting calendar to avoid confusion. If the deficit persists, the number of replicas will have to be reduced, with impact on the analysis efficiency, and Monte Carlo production must be reduced, which impairs the accuracy. The Committee hopes that means can be found to remedy this situation.