

CERN-RRB-2010-091
29 SEPTEMBER 2010

PRINCIPAL LHCC DELIBERATIONS

18TH MEETING OF THE COMPUTING RESOURCES REVIEW BOARD

12 OCTOBER 2010

EMMANUEL TSESMELIS
SCIENTIFIC SECRETARY, LHCC

GENERAL

This document summarizes the principal LHCC deliberations concerning the World-wide LHC Computing Grid (W-LCG) Project at the Committee's sessions in May, July and September 2010.

The W-LCG continues to operate very well. The LHCC congratulates the W-LCG project team and the experiments in their successes at processing and analyzing the data from the first months of the 2010 LHC run.

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 for the long term.	Computing resource availability remains a long-standing problem for ALICE.
Tier-2 Sites	The pledged resources are only about half those requested.	The distribution of W-LCG usage is roughly as planned, with over 50% of CPU delivered by Tier-2 sites.

STATUS OF THE W-LCG

GRID RESOURCES

Grid resources are used heavily, with over a million jobs and 100k CPU-days per day and several hundred users from each of the collaborations. Large simulation production continues in parallel with data taking and uses a large fraction of the CPU resources. The distribution of usage is roughly as planned, with over 50% of CPU delivered by Tier-2 sites. Tier-0 usage is a little below expectation, but is expected to increase with the delivered luminosity. Data transfer from Tier-0 to Tier-1 sites has mostly been at the expected rate, but has exceeded the target of 1.3 GB/day without any problem on many occasions. For all the experiments data has been available for analysis within hours of data taking. About 5 PB of data has been stored by the CASTOR mass-storage system so far this year, with 70 TB/day written to tape and 1 PB/day transferred in and out of storage.

Resources are being used in accordance with the planned computing models, with some variations in the light of experience. Networks have been used at the anticipated scale, and a working group has been established to plan for future needs. ATLAS has tested data caching, in addition to the data placement of their computing model. The number of interactions per bunch crossing (μ) is higher than had been expected in the first year of LHC running. The LHCb experiment and computing model were designed to handle $\mu=0.4$, but $\mu=1.4$ has been observed. This results in a larger event size and pressure on the available disk space in order to record data without loss of efficiency for the charm and beauty physics channels. In response the LHCb Collaboration has distributed the data across their Tier-1 sites, rather than keeping a full copy at each site as originally planned, and they have modified their event selection and analysis strategies to cope with the multiple interactions. For

all experiments, more analysis than anticipated has used Event Summary Data-level (ESD-level) data, but use of the Analysis Object Data (AOD) and derived data formats is increasing. All groups plan full reprocessing of the complete 2010 proton-proton data sample before the end of the year.

Installed resources meet the levels pledged for this year, although delays and hardware problems have occurred at some sites, especially for disks. Use of disk storage is good, while further CPU and tape capacity remain for the additional data expected in 2010-2011. Procurement cycles for the equipment are long and planning is required well in advance to ensure that pledged resources are available on schedule. The Computing Resources Review Board will receive for its meeting of October 2010 a report from the Computing Resources Scrutiny Group, which is reviewing the use made by each experiment of the resources provided. Requirements for 2012-2013 should also be discussed, but precise evaluations are impossible without details of running conditions and better analysis of the experience so far. Plans for 2013 running have been discussed by the Research Board and consistent set of assumptions will be provided and used by all experiments in planning their resources requests.

GRID OPERATIONS

The experiments report no major problems with data processing, and generally good performance from the W-LCG. They have coped with the peaks in analysis activity in advance of the major 2010 summer conferences.

Operational problems at sites remain the main stability issue. There are typically 5-6 service incidents per month, lasting a few hours or sometimes more extended, and caused by power, cooling or other hardware failures or by database problems, rather than by specific Grid middleware. In the last two weeks the necessary response to a general security alert reduced capacity by up to 50%. These incidents are largely unavoidable, but lessons are learnt to mitigate similar problems in the future. Computing models must evolve to become resilient to outages at some sites.

One major incident in April-May 2010 resulted in a small loss of ALICE data. A configuration error in the CASTOR mass-storage system at the Tier-0 centre resulted in raw data being distributed across all available tape pools instead of to dedicated pools as intended. This included a pool in which tapes were recycled after some time. ALICE, ATLAS and CMS were affected, but only ALICE data files were overwritten, including 1773 files of collision data at 900 GeV centre-of-mass energy that had not yet been copied to the Tier-1 centre. All but 56 of these files were eventually recovered from the tapes. Actions taken include a review of the software change procedure and improved monitoring of CASTOR. For the heavy-ion running a back-up copy of ALICE raw data will be stored temporarily at Tier-0 until replication at Tier-1 has been completed.

GRID INFRASTRUCTURE

As part of CERN's agreed Medium Term Plan, budget reductions of 8.25 MCHF over 5 years will come from W-LCG materials. Cutbacks include changing from a 3-year to a 4-year equipment replacement cycle and reducing the increase in Tier-0 resources from 30% each year to about 25% per year. The implications require further detailed planning, but there will be a strict limit to the growth of overall computing resources at Tier-0. This also affects the strategy for Tier-0 infrastructure. The planned new Tier-0 building has been cancelled and the need for containers is delayed by a year to 2014. There are plans to upgrade the existing Building 513 power from 2.9 MW to 3.5 MW, including up to 600 kW of diesel-backed power, together with the use of local hosting in Geneva (100 kW today, potentially more). A call for bids to host a remote Tier-0 data centre, in the longer term, has been issued and 6-7 countries have expressed an interest.

HEAVY-ION RUNNING

Planning for the heavy-ion (HI) running in November 2010 is in hand, but there are concerns over some very recent requests for increases in requirements. ALICE want to record 200 TB/day, at twice the rate originally planned, to compensate for the reduced length of the HI running period. CMS plan to record HI data without zero suppression in their Silicon Tracker and Calorimeter and with a minimum-bias trigger, which will increase their data rate to about 150 TB/day. The ATLAS rate is expected to be 10-20 TB/day and LHCb will not participate in HI running. ALICE will export 20% of their raw data during the HI run and the remainder during the technical stop. It has already been agreed to keep two copies of the ALICE HI data at CERN until the export has been completed, to reduce the risk of data loss. CMS will export all their HI data to FNAL after software zero suppression and ATLAS will export all raw data to Tier-1 sites. These requests significantly exceed the foreseen data rates and there has been no readiness test for combined HI operations at these rates. CASTOR has some headroom, but also needs to maintain a sustained and reliable service for other users, including non-LHC experiments. Some 2011 resources have been pre-purchased and can be implemented to help match the needs. A combined test including ALICE and CMS (and perhaps also ATLAS) must be arranged in good time before the start of the HI run. In case of problems during the run, archiving of data to tape must take priority and the experiments may be asked to make use of their data buffers, which can hold around a day's data, to spread the load over time. The LHCC encourages all collaborations to work in a cooperative and constructive way together with CASTOR staff to test the system thoroughly. The LHCC recognises the benefit to CMS of recording HI data without zero suppression, but given the impact of this late request on Tier-0 operations it encourages CMS in particular to develop contingency plans to ensure that HI data taking can continue in case problems arise during the run.

FUTURE PLANS

The resource planning for the coming years is a concern in view of the approved CERN Medium Term Plan and the low resources pledges for ALICE. The full effect of introducing non-expert users to the W-LCG will have to be seen. The transition from the Enabling Grids for E-Science (EGEE) to the European Grid Initiative (EGI) is in progress. The W-LCG must continue to address long-term sustainability of the available system, including the level of effort available for middleware support and the data access for analysis and also the long-term sustainability for some of the Tier-1 centres.