CERN-RRB-2012-009

Computing Resources Review Board

Minutes of the 20th Meeting held at CERN on 18 October 2011

Present:

Europe:

J. Sancton (FNRS, Belgium), J. Lemonne (FWO, Belgium).

M. Lokajicek (Institute of Physics AS CR, Czech Republic), O. Novak (Ministry of Education, Youth and Sports, Czech Republic), J. Ridky (Institute of Physics AS CR, Czech Republic), D. Adamova (Nuclear Physics Institute AS CR, Czech Republic).

J. D. Hansen (NBI, Denmark).

D.-O. Riska (HIP, Finland).

F. Malek (CNRS/IN2P3, France), P. Rebourgeard (CEA/IRFU, France), L. Serin (CNRS/IN2P3, France).

S. Bethke (MPI, Germany), K. Ehret (BMBF, Germany), M. Fleischer (DESY, Germany), V. Guelzow (DESY, Germany), H. Marten (BMBF, Germany), P. Mattig (BMBF, Germany),

G. Vesztergombi (KFKI-RMKI, Hungary).

F. Cervelli (INFN, Italy).

F. Linde (NIKHEF, Netherlands), A. van Rijn (NIKHEF, Netherlands).

B. Jacobsen (Norwegian Research Council), F. Ould-Saada (University of Oslo).

M. Turala (HNIN, Academy of Sciences, Poland).

G. Barreira (LIP, Portugal).

F.-D. Buzatu (Institute of Atomic Physics, Romania).

D. Filatov (Ministry of Education and Science, Russia), V. Ilyin (Moscow State University), M. Itkis (JINR, Russia), V. Savrin (Institute of Nuclear Physics, Russia).

N. Colino (CIEMAT, Spain), F. del Aguila (Universidad Granada, Spain).

P. Karlsson (Swedish Research Council), M. Nylen (Umea University).

T. Nakada (CHIPP, Switzerland).

A. Zagorodny (National Academy of Sciences of Ukraine), G. Zinovjev (National Academy of Sciences of Ukraine).

A. Medland (STFC, United Kingdom).

Americas:

W. Davidson (National Research Council of Canada), R. Teuscher (University of Toronto)

L. Bauerdick (FNAL, USA), J. Butler (FNAL, USA), S. Gonzalez (DOE, USA), H. Gordon (BNL, USA), T. Hallman (DOE, USA), P. Jacobsen (DOE, USA), S. Rajagopalan (BNL, USA), S. Rolli (DOE, USA), R. Ruchti (NSF, USA), J. Shank (Boston University, USA), M. Tuts (Columbia University, USA), T. Wenaus (BNL, USA),

Asia Pacific:

Q. Meng (NNSF, China), Y. Zhang (NNSF, China).

Y. Rozen (Technion, Israel).

H. Kamiyama (Permanent Mission of Japan), T. Kawamoto (University of Tokyo, ICEPP, Japan).

D.-J. Chung (MEST, Republic of Korea), H. Kim (National Research Foundation of Korea), U. - K. Yun (National Research Foundation of Korea).

S.-C. Lee (Academia Sinica, Taiwan).

CERN:

S. Bertolucci (Chair), J. de Groot (Scientific Secretary), F. Hemmer (IT Department Head), T. Lagrange (FP Department Head), J. Salicio Diez (PH Department), E. Tsesmelis (LHCC Scientific Secretary), E. van Hove (FP Department).

WLCG:	I. Bird, S. Foffano.
C-RSG:	D. Espriu.
SG:	B. Loehr.
ALICE:	P. Giubellino, Y. Schutz.
ATLAS:	F. Gianotti, G. Mornacchi.
CMS:	I. Fisk, D. Bonacorsi, J. Incandela, T. Rodrigo, and G. Tonelli.
LHCb:	P. Campana.

The agenda and presentations to the meeting are available from the INDICO web site:

https://indico.cern.ch/conferenceOtherViews.py?confId=149413&view=lhcrrb&showDate=all&showSession=all&detailLevel=contribution

---

Introduction

S. Bertolucci welcomes the participants to the meeting.

He notes that WLCG has been among the key elements for the success of the first phase of LHC running. The performance of the system has been impressive, in particular in the way it was able to adapt to the rapidly changing experimental conditions. Last year saw 16 PByte of data written to tape. To date this year, despite larger and more complex events, about 15 PBytes have been recorded. The heavy ion run planned towards the end of the year should add significantly to the volume of data recorded.

The minutes of the previous meeting are approved without correction.

WLCG Status (I. Bird)

I. Bird presents a status report of the WLCG project.

He presents some statistics about the data taking in 2011. Until the end of September the LHC has delivered about 4 fb-1, causing about 2 PBytes per months of data to be recorded. Until the end of the year the volume of data written to tape is expected to reach 20 PByte. Data is read from CASTOR at a rate of about 10 PByte per month.

The experiments' computing models are evolving with reduced data read from tape in favour of keeping data on disk. He shows statistics on cpu usage and the number of jobs. At present there are over a million of jobs a day. All of the Tier-1 and Tier-2 capacity is essentially used all the time with some Tier-2 centres delivering more than was pledged. The CERN Tier-0 is used heavily during the data taking but less so outside these periods. Efficiency is high for all experiments except for ALICE. The ALICE problem has been understood and is being addressed. 2011 appears to be the first nominal year of LHC computing with data taking on-going, (re)processing and analysis.

The high LHC luminosity results in up to 20 interactions per bunch crossing for ATLAS and CMS. This adds to the complexity of the events recorded and to increased cpu time requirements. Some data on event sizes and trigger rates are presented. ATLAS and CMS have put much effort into reducing event size in order to allow somewhat higher trigger rates. LHCb runs with a constant trigger rate under the 'luminosity levelling' scheme.

A lot of effort has gone into software improvements leading to reduced memory requirements and improvements in reconstruction time. The efficiency for scheduled ALICE

processing jobs is now good while the efficiency of analysis jobs is being studied. For the heavy ion run ALICE will run with data reduction, allowing doubling the readout frequency.

ATLAS is moving towards dynamic data placement with physics data sets copied to Tier-2s only when needed and some pre-placement of popular data sets. LHCb fully utilize their Tier-1s for data taking leaving no time for reprocessing. In a change of their computing model, LHCb has started using Tier-2s for reconstruction.

Based on the experience in 2011 there have been some changes in the requests for resources for 2012, mostly for disk space at CERN (ALICE, CMS), cpu at the Tier-0 (ATLAS) and global disk space (LHCb).

I. Bird then shows the reliability of CERN + Tier-1 and the Tier-2 sites. CERN + Tier-1 efficiency is mostly not an issue anymore. Most of the Tier-2 centres have reliabilities over 95% with the top 50% od the sites having almost 10% reliability.

The number of service incidents has decreased with response time well within targets. Most incidents are due to infrastructure problems (power, cooling, hardware failure) and are at a constant level. Other sources of incidents are associated with storage services and databases. The associated operations load is now viewed as sustainable.

Status of the Tier-0: There were many responses to the request for proposals for a remote Tier-0 extension. Following visits to the sites and discussion, tender documents have been prepared and sent out. Tests are planned for 2013 (perhaps already in 2012) with production expected in 2014.

Concerning LHCONE, the project to improve Tier-2 connectivity, an interesting working proto-infrastructure now exists. There are some doubts though about the scalability and manageability of the system. Bearing in mind also that the Tier-2 networking has improved, the decision has therefore been taken to slow down the development while investigating emerging solutions.

After almost tow years of data taking and 10 years after starting to design the system it has been agreed to prepare a documented vision for the future to be able to communicate our needs to the software and technology providers. This will also be an opportunity to rebuild a number of the common solutions. A few working groups have been created to address key topics and produce a strategy for the future. A first draft of the report is requested by late 2011 or early 2012.

Summarizing, I. Bird says that Grid operations have continued smoothly over 2011 with no major issues. The experiments are making good progress in data processing and analysis. Tier-1 and Tier-2 resources are fully utilized with the Tier-0 utilized fully during peak times. There are additional requests for resources for 2012, mostly for the Tier-0. Planning has started for a strategy document for the evolution of the WLCG infrastructure.

On a question from D. Adamova about the possible use of commercially available Cloud computing, I. Bird replies that this is being discussed by one of the working groups he mentioned earlier. Several experiments have performed tests. In his view it is certainly a possibility but today the cost is rather high.

D. Riska asks about the handling of the requests for increased resources. I. Bird replies that this will come up in the report of the S-RSG later during the meeting. The pledges vs. the requests will be shown during the presentation by S. Foffano.

A. Medland asks what I. Bird considers to be the main risk or constraint on the collaboration in the mid-term. I. Bird sees little risk associated with the collaboration, this works rather well. The main risk in his view is the question if the rate of increase of resources can match the increased need of the experiments.

Principal LHCC Deliberations (E. Tsesmelis)

E. Tsesmelis, Scientific Secretary of the LHCC, says that the LHCC has reviewed the WLCG in its June and September sessions. The Committee congratulates the WLCG team and the experiments for the success in recording, processing and analysing the data. Further details are given in a paper that has been submitted to the meeting (CERN-RRB-2011-096).

Status of Common Projects Accounts (T. Lagrange)

T. Lagrange says that there are no movements to be reported since the cut-off date for the report that was sent to the Board.

Status of Resources and Financial Plan (S. Foffano)

S. Foffano starts with an overview of her presentation. Supporting information may be found in the written report (CERN-RRB-2011-107).

Concerning the composition of the WLCG Collaboration she says that since the last meeting there have been two changes: LLNL (US), has been given full member access based on a Letter of Intent (LoI) pending signature of the MoU, and LPSC Grenoble, a former Tier-3 has become an ALICE and ATLAS Tier-2. Discussions are under way with three new countries that have expressed interest in becoming Tier-2 sites. In its present form the Collaboration includes one Tier-0, 11Tier-1 sites and 67 Tier-2 federations. The different sites, Funding Agencies and their representatives are documented in Annexes 1 and 2 of the WLCG MoU. She presents the up to date tables with the details of the Tier-1 and Tier-2 sites. She notes that the Tier-1 data are rather stable but that changes involving the Tier-2 sites are not infrequent. She asks the members of the RRB to keep the LCG Office informed of any changes.

Concerning funding and expenditure for WLCG at CERN S. Foffano presents a summary of the expected situation at the end of 2011 and projections for 2012 – 2016. For this year a slight overspend on personnel might occur whereas for materials there will be a significant underspend mainly due to delays in the Computer Centre upgrade and the tendering for the remote Tier-0.

Moving to Tier-1 and Tier-2 accounting, S. Foffano notes that monthly reports on Tier-1 and Tier-2 are published and distributed by the LCG Office. Sites are encouraged to consult the accounting portal to check their data. The 2011 Tier-1 and Tier-2 resources have been almost fully installed. All Tier-2 sites are now reporting data correctly. S. Foffano adds that at recent meetings of the LHCC and the LCG Overview Board the very positive resources usage at CERN and at the Tier-1 and Tier-2 sites was noted.

She shows in graphical form the CERN and Tier-1 accounting for 2011, showing the use of the resources by the experiments, the MoU commitment and the installed capacity.

The Tier-2 cpu accounting is shown, comparing the 2011 pledge with the actual usage in the months of May – September. The data show that in a number of cases the actual usage exceeds the pledged value by a significant amount.

Concerning the pledges of resources for 2012 and 2013, at the last meeting the sites were requested to enter the data via the REBUS tool using as input the requests from the experiments as approved by the C-RSG. Very few problems were experienced in the process and by the end of September most of the sites had entered the data with those remaining entered by the WLCG Office. A table with a summary of the requirements and the 2012 pledges is shown highlighting some shortfalls, in particular for ALICE, whereas other experiments are in a more comfortable situation. She cautions that the requirements are preliminary and have not yet been approved by the C-RSG.

Summarizing, S. Foffano recalls the LLNL as new ALICE Tier-2 site and the change of status of LPSC Grenoble. WLCG funding and expenditure: some delay in committing materials budget is foreseen for 2011. The Tier-0, Tier-1 and Tier-2 accounting show good resources use overall. Pledge data for 2012 and 2013 was collected using REBUS tool with complete data presented in CERN-RRB-2011-107, Annex 1.

Report from the C-RSG (D. Espriu)

D. Espriu, Chair of the Computing Resources Scrutiny Group (C-RSG) starts his presentation by congratulating the experiments and the WLCG Collaboration for the very successful operation in 2011.

He recalls the membership of the C-RSG. A number of members have changed, but others have been members from the very beginning and should, according to the terms of the MoU, be reappointed or replaced this year or latest during 2012.

He outlines the contents of his presentation, noting that it reflects the written report (CERN-RRB-2011-087). He cautions that there is an unfortunate error in the written report presented to the meeting, with cpus at the Tier-2s stated incorrectly. The correct numbers will be given in this presentation and a revised version of the document has been uploaded to INDICO. He requests the experiments to submit their reports before 1 March 2012.

D. Espriu shows the planned 2011 and 2012 LHC physics runs. The lifetime of about 20%has not unexpectedly been lower than the nominal value of 28%. On the other hand, the luminosity has been very high with values reaching over 10^{33} $cm^{-2}s^{-1}$. Thanks to the high luminosity and data taking rates higher than planned, the experiments have recorded more than 70% of the maximum possible number of events. Also, because of an average of about 15 events per bunch crossing for ATLAS and CMS, the events are more complex. This year has been a challenge for the experiments faced with rapidly changing experimental conditions, the need for reprocessing of the data and 'chaotic' computing for analysis.

Turning to the scrutiny of the 2011 resources utilization, D. Espriu shows a summary of the 2011 usage of resources (cpu, disk and tape) at CERN and the Tier-1 and Tier-2 centres. Compared to the same data for 2010 there has been a marked increase. The Tier-2 cpu utilization efficiency has similarly improved for all experiments, except for ALICE. In view of this, the C-RSG recommends to increase the assumed Tier-2 cpu efficiency to 67% for 2012 and 70% for 2013. The low efficiency for ALICE is due to the comparatively low efficiency for analysis jobs and should be improved. The fraction of 2011 CERN and Tier-1 resources used by experiment is shown, together with the fraction of cpu used at CERN. ATLAS and CMS appear to be converging to a fraction of around 18%. LHCb has reduced its use of CERN cpu to 28% whereas the value for ALICE is 59%. This is explained in the context of the shortfall of resources for ALICE, but D. Espriu notes that this value can probably not be sustained in the future. For the Tier-2 centres, the fraction of cpu used by experiment show stability for ALICE, ATLAS and CMS with a welcome increased use by LHCb. The situation of CERN and Tier-1 + Tier-2 delivered versus pledged resources is very satisfactory with no major problems reported. D. Espriu notes the 80% Tier-1 tape resources delivered and says that he considers this a positive development indicating the reduced reliance by the experiments on tape resources.

D. Espriu then reports on the usage of the resources by the experiments. Before examining the details he notes that the computing models have allowed the recording, distribution and analysis of substantial amounts of data. Performance has been smooth and without major difficulties. Peaks in usage associated with the seasonal conferences have been observed. Data placement policies have been different from what was envisaged in the computing models but reprocessing policies have quickly converged to what was planned. The available cpu resources have generally exceeded the experiments' needs leading the experiments to increase significantly the production of simulated data. Some inefficiencies are associated with large memory footprints.

For ALICE the use of disk at the Tier-1s and tape of CERN has improved whereas the use of tape at the Tier-1s is low. ALICE has used more than its share of cpu resources at CERN. Efficiency at CERN and the Tier-1s is still low.

ATLAS efficiencies at CERN and the Tier-1 and Tier-2 centres have been quite high. Use of disk at the Tier-2s is low, as is the use of tape at the Tier-1s. ATLAS will examine this and has agreed to submit a reduced request for Tier-1 tape.

CMS uses less CERN facilities than ATLAS. Tape usage at CERN and at the Tier-1s is low. CMS has agreed to consider a possible reduction of Tier-1 tape requests.

There are no particular problems for LHCb with usage of resources as expected. Efficiency at the Tier-2 centres is extremely high at 98%.

Concluding his report on the usage of resources by the experiments, D. Espriu says that the experiments have made significant efforts to mitigate the growth in resources by reducing event sizes, and processing times. This has to some degree compensated for the high pileup. Data distribution policies have been changed substantially, reducing the number of copies stored at the Tier-1 and Tier-2s and moving to more compact data formats. Reconstruction times have improved and fast Monte Carlo simulations have been implemented. In order to

optimize the usage of resources, experiments have redistributed tasks among CERN, the Tier-1 and Tier2 centres. Except for ALICE, user efficiency is better than planned.

The last part of D. Espriu's talk is dedicated to the scrutiny of the requests for 2012 1nd 2013 resources.

ATLAS has requested, endorsed by the C-RSG and accepted by CERN, an increase of cpu resources at the Tier-0. There is an increase of disk resources requested at the Tier-2s. The request for 2012 is endorsed by the C-RSG and the request for 2013 will be examined next year.

There are only small changes for CMS with respect to the data submitted in spring. The C-RSG endorses the CMS request. The request for tape resources is still being discussed with CMS.

For LHCb there has been a small reduction in the request for Tier-2 cpu resources.

D. Espriu recalls that ALICE was asked at the spring meeting to revise its request for resources in order to bring it in line with the expected resources. After reconstruction and analysis of the first heavy ion data is clear that ALICE will not have sufficient resources and that the expectation for additional resources is not good. ALICE nevertheless prefers to submit a request for resources based on the physics needs of the experiment. D. Espriu mentions that there have been some changes in the right direction in particular with respect to event sizes. ALICE expects to reduce further the event size of heavy ion events by a factor of 3.5. In addition there have been substantial improvements in processing times for reconstruction and simulation as well as the memory footprint. Rather than bringing the request more in line with these advances, ALICE prefers to keep the data acquisition saturated by increasing the data taking rate by a factor of two. The C-RSG is of the opinion that ALICE will not be able to compensate its lack of pledged resources by an increased reliance on CERN as has been done previously. After discussion with the LHCC and the C-RSG, ALICE has updated its request for resources. In the view of the C-RSG this request is still not realistic, even in the light of a possible future contributions of planned Tier-1 centres in Korea and Mexico. The C-RSG takes note of the ALICE request for resources confirming that the request corresponds to the physics needs of the experiment, but considers the request to be unrealistic.

Concluding his presentation, D. Espriu presents a number of recommendations and requests:

- The C-RSG is concerned about the medium term sustainability of the ALICE computing model and the low user efficiency. The C-RSG recommends a substantial revision of the model.
- There is room for improvement in staging strategies and/ or dynamic data placement.
- The C-RSG encourages close collaboration between the different centres and the experiments to continue the implementation of efficient and cost-effective access to the data.
- WLCG Tier-2 accounting is improving but is still insufficient. Installed cpu capacity compared to pledged and installed disk capacity at the Tier-2 centres are not centrally accounted. It would be useful to disentangle organized and chaotic processing activities.
- A balanced use of the worldwide LCG resources is essential for the long-term coherence of WLCG. A continuous increase in CERN resources is not sustainable.
- A discussion on memory footprint is needed. The C-RSG noted that lack of memory makes half the cores unusable.
- The R-CSG recommends that CERN's policies of resources sharing when allocations are not fully used are clearly stated.
- In view of the current statistics, the C-RSG proposes to make firm the April 2011 recommendation of assuming a Tier-2 efficiency of 67% for 2012 and 70% in 2013.
- The C-RSG welcomes the capability of the experiments to take data at increased rates but the C-RSG does not see how a substantial increase could be accommodated with existing resources and does not recommend a formal modification of the computing models in this direction.

- It is recommended to improve the mechanisms of communication between experiments and the Tier-1 and Tier-2 centres in order to optimize usage.

Discussion:

P. Giubellino acknowledges extensive discussions with the C-RSG. He notes that substantially increased efficiencies do not yet show in the data presented by D. Espriu because of the cutoff date of 31 August. ALICE is of course aware of the shortfall of resources, but P. Giubellino underlines that the cpu time per event has been reduced by a factor of two and the event size will be reduced by a factor of 3.5. He does not consider it a good idea to reduce the rate of events written to tape as this determines the mount of physics that can be done. All efforts should be made to reduce the amount of computing needed to treat these events. He notes that the fact that all ALICE analysis is done on the Grid has an impact on the overall efficiency. He hopes that by the time of the next RRB the overall situation will have been improved significantly.

F. Gianotti thanks D. Espriu and the C-RSG for their work and acknowledges the constructive interaction with the C-RSG. The current year is the first year that the ATLAS computing model has been stressed. ATLAS has been using the resources as expected. The fact that the Tier-2 disk usage appears low at 43% of the pledged value is due to a snapshot taken at the end of July. The number is small because at that time only about 1.6 fb^{-1} had been recorded compared to the present value of 5 fb^{-1}. Furthermore, this was just after an extensive 'cleaning campaign'. Today, a much larger fraction of about 60% is being occupied. A better snapshot of the occupancy will be available at the end of the year. If it turns out that at the end of the year the disk usage is less than expected ATLAS will reduce its request as was done previously for tape resources. F. Gianotti emphasizes the distributed nature of the ATLAS analysis. CERN resources are used for first pass reconstruction and time-critical tasks only. Care should be taken to ensure sufficient Tier-2 disk resources in order to allow the worldwide collaboration to participate fully in the analysis of the data.

G. Tonelli mentions that the snapshot at the end of July of this year presented by D. Espriu is not representative of the challenges facing the experiment. The situation in September or October is substantially different. With a pileup value around 15 the computing task becomes really challenging with respect to reconstruction times and memory footprint. CMS has decided to stick to the present software version that has been used for the entire year to avoid redoing part of the analysis. The price to pay is that this version of the software was optimized for much smaller pileup values. All this means that all resources, and even some additional ones, are presently being used. There is a continuous effort to optimize the use of resources while improving the software eventually. This has allowed a request for 2012 resources which is not very different from last year's plans, despite the larger than expected data volume and increased complexity of events. He acknowledges the positive interaction with the C-RSG. He pleads for some flexibility to be able to cope with unexpected conditions in 2012.

P. Campana thanks the C-RSG for the fruitful interactions. This year has been already a 'flat top' year of LHCb data taking with all computing resources being used fully. LHCb is concerned about disk resources because LHCb has increased the amount of physics data by increasing the trigger rate from 2 to 3 kHz. By reducing the number of copies of the data LHCb has managed to stay within the limits of the allocated resources. He notes that the resources requests from LHCb are comparatively modest with respect to the other experiments.

M. Turala remarks that all those present at the meeting are very appreciative of the work done by WLCG. But this message does not always pass to the people at the different participating computing centres. He suggests that a letter from CERN Management or WLCG could be sent to his ministry to recognize the effort. He suggests that other representatives who would similarly be interested could contact CERN Management or I. Bird. S. Bertolucci takes the suggestion on board. He notes that, in the particular case of Poland, he had recently the opportunity to personally congratulate the minister for the Polish contribution to WLCG computing.

Bertolucci asks for additional comments concerning the experiments' proposals. There being no further comments the experiments' requests are endorsed by the RRB with the caveats expressed by D. Espriu on behalf of the C-RSG.

Summary

S. Bertolucci summarizes the meeting. In his view, the combined efforts of the experiments and the C-RSG the system are causing the system to be as lean as possible. Care should be taken to be proactive because reacting to problems will most likely result in lower efficiency. Innovative technologies will need to be evaluated, not only what concerns hardware but also in the way these resources are being used.

He thanks the participants in the meeting for their contributions.
