Computing Resources Scrutiny Group

C.Bozzi (Italy), T. Cass (CERN), G.Lamanna (France), D.Espriu (Spain, *Chairman*), J.Flynn (UK), M.Gasthuber (Germany), D.Groep (The Netherlands), T.Schalk (USA), W.Trischuk (Canada), B.Vinter (Nordic Grid), H.Meinhard (CERN/IT, *Scientific Secretary*)

INTRODUCTION

The Computing Resources Scrutiny Group (CRSG) is an independent committee established by the WLCG Memorandum of Understanding whose members are selected by the Funding Agencies represented in the C-RRB. The purpose of the CRSG is to inform the decisions of the Computing Resources Review Board (C-RRB) for the LHC experiments.

According to the WLCG MoU, every year the CRSG shall scrutinize

- The resource accounting figures for the preceding year
- The use the experiments made of these resources
- The overall request for resources for every experiment for the following year and forecasts for the subsequent two years

The CRSG shall also examine the match between the refereed requests and the pledges from the institutions and make recommendations concerning apparent under-funding.

This report summarizes the deliberations of the CRSG regarding the usage of the computing resources by the four LHC experiments during the elapsed months of 2011. We consider the overall performance of the computing models and of the WLCG to be good. There are nevertheless some inefficiencies that have been identified and will be discussed in this report.

We have also examined some updated requests for 2012 and 2013. The modifications with respect to the preliminary scrutiny presented in the April 2011 C-RRB are small, except for the ALICE collaboration.

Part A of this report is concerned with the overall usage of the WLCG resources and with the scrutiny of the different experimental collaborations' use of these resources. Part B deals with the modifications with respect to the preliminary April 2011 scrutiny. A summary of Tier 2 usage kindly compiled by I.Fisk is appended as part C.

The elapsed months of 2011 have witnessed full usage of the WLCG resources. The computing models and the available resources have been subject to real stress just before the seasonal conferences and, with some exceptions that will be commented on below, have performed well. It now appears possible to draw rather firm conclusions and establish recommendations for the collaborations. These will be discussed in below and in the B section in connection with each experiment.

The LHC running conditions

After the Chamonix meeting early in 2011, it was agreed that, barring unforeseen circumstances, the LHC should run for the best part of 2011 and 2012 (the estimate is 8 months per year), with only a relatively short break at the end of 2011. The energy has continued to be 7 TeV (3.5 TeV + 3.5 TeV) during 2011 but it may change slightly during the coming year until the end of the current combined long run, which may also be subject to some uncertainty contingent on the physics output. About 10% of the time is expected to be dedicated to heavy ion (HI) physics. A long shutdown will follow to enable the machine to reach the design energy of 14 TeV (7 TeV + 7 TeV).

For the scrutiny the most relevant quantity is the total number of seconds when the beam is declared to be stable and good for physics. Following CERN management recommendations the following scheme has been adopted:

Live time: 30 days/month = **720** hours
Folding in efficiencies 720 x 0.7 x 0.4 = 201.6 effective hours/month = **725760** sec/month

RRB year	RRB year start	RRB year end	Months (max) Data taking	Total live time (in Ms)	pp	AA
2011	**April '11**	**March '12**	**8**	**5.9**	**5.2**	**0.7**
2012	**April '12**	**March '13**	**8**	**5.9**	**5.2**	**0.7**

For the period covered in the usage report presented here the estimates have generally been reflected in reality with about 2.8 Ms of pp beams being delivered to the experiments. This reveals the excellent performance of the LHC complex of accelerators.

From March 1st to July 31st 2011 the experiments have accumulated more than 70% of the total number of events they expected for this period. This substantial number of events was recorded despite the machine average live time being closer to 20% than to the nominal 28% used in the above table. This large number of recorded events has been possible thanks to experiments using all the available bandwidth and effectively recording events at rates larger than the nominal ones. Everything seems to indicate that this trend will continue until the end of the pp run and it will extend to the AA run. Barring surprises, the final number of recorded events will likely be of the order of 75% of the planned one.

While it was already clear in the April 2011 scrutiny that the experiments had been able to cope with rapidly changing running conditions (such as the luminosity increasing five orders of magnitude over a few months), the total volume of data was still limited and there was a moderate sense of urgency for physics results.

The situation has now evolved to steady running conditions and the statistics indicate a massive use of the available resources. Some aspects of the computing models such as large individual non-organized computing usage, format and distribution of the data sets, the flexibility to cope with increasingly challenging running conditions (pile up, bunch separation), and the urgency to reprocess and analyze large amounts of data in a short time have represented a real challenge for the computing models and for the WLCG as a whole.

In addition during the early months of 2011 the collaborations reconstructed and analyzed the events recorded during the first AA run. This run, during which the LHC performed quite well, has tested the core of the ALICE computing model.

Proton bunches are injected now with a minimal separation of 50ns (design value was 25ns). To compensate they contain more protons, squeezing them as much as possible to sustain a substantial luminosity ~ 10^33. The consequence is the appearance of events with many interactions (pile-up). The experiments have estimated that this may reach up to ~20 interactions per crossing except at the LHCb interaction point) at the beginning of fills. Pile-up that was lower than expected at the beginning of the run is now reaching ~15 interactions per crossing and up. This has had a substantial impact on reconstruction times and on the size of the data sets which are larger than expected for the same pile-up conditions as there is some non-linearity with the number of interactions.

Collaborations now appear to take this effect into account in their estimates in a homogeneous way. Pile-up indeed has become a major point of concern for ATLAS and CMS as it has a direct impact on the cost of computing. The eventual move to 25 ns crossings with shorter bunches would diminish multiple interactions in the same crossing but enhance the out-of-time pile-up, an effect that is already observed but whose impact on the event size is yet to be quantified precisely.

Interactions with the experiments

The recommendations of the previous C-RRB report in April 2011 urged the experimental collaborations to submit a detailed account of their use of the resources along with their revised requests by September 1st. ATLAS, CMS and LHCb complied with a few days after this deadline while ALICE informed us that they would need some two more weeks to submit their report.

Following the reception of the reports, referees were assigned to the different experiments and a number of interactions took place. As agreed with the ATLAS and CMS management in 2009 the scrutiny procedure for these two experiments is done by a single team of referees, using common techniques and methods. This ensures that a coherent set of principles is applied.

In the CRSG report CERN-RRB-2011-019 we asked for a revision of the ALICE request to make it commensurate with the available resources profile and in this way facilitate a realistic scrutiny. Following this request our referees met several times with ALICE representatives. ALICE submitted in September a revised estimate of the resources they needed for 2012 and 2013 that went in the opposite direction and in fact represented a substantial increase in the disk request. This was deemed unacceptable by the CRSG and the collaboration made adjustments in their computing model in order to bring the request more or less in line with the April 2011 figures. The CRSG fully understands that the request of the ALICE collaboration is driven by physics requirements and is reasonably estimated within the parameters of their computing model but is concerned by the mismatch with the expected level of pledges.

Generally speaking the interactions with the experiments are very fluid and we thank the respective managements for their openness and collaboration. Thanks are due in particular to Ian Fisk who compiled and summarized the Tier 1 and Tier 2 usage for 2011.

Interactions with the LHCC

The CRSG chair was invited to an informal LHCC computing mini-review on September 20 th that was very useful.

In this meeting the collaborations presented detailed discussions of the reprocessing times, event sizes, etc., that is, the basic parameters of their models, and compared with previous estimates. Modifications of the data placement policies were discussed too. The efficiency of user analysis was also reviewed and in general found to be better than expected, with the notable exception of ALICE where it was alarmingly low. WLCG resources are now commonly used by a large fraction of the collaborations, which represents very welcome news.

The collaborations also reviewed their updated requests for 2012 when applicable.

Since the last scrutiny no issues appeared for which we thought it was necessary to refer to the LHCC. At the moment of submitting this report the LHCC recommendations had not been received by the CRSG yet.

Mitigation of the resource growth

Experiments were urged to use the experience gained in the first years of running to modify the their computing models in ways that would make them sustainable in the long run and mitigate the growth in resources. In the reports submitted to us and in conversations with the collaborations we have seen substantial revisions and adaptations to meet this goal.

- Experiments have made an effort to reduce the raw event size (and the size of all subsequent derived formats) and event processing times. These efforts have mitigated the serious challenge that pile-up represents. This has allowed experiments to record events at a higher rate, indicating some margin of safety and redundancy in the resources available.
- The experiments have made substantial changes in their data distribution policies, reducing the number of copies stored in Tier 1 or Tier 2 and moved to more compact datasets such as AOD's.
- Reconstruction times have generally improved.
- Substantial progress in the implementation of fast MonteCarlo simulations has been made
- Experiments have been very active in redistributing tasks among CERN, Tier 1 and Tier2 to optimize the usage of resources. The CRSG welcomes an improved, more equilibrated distribution of the usage.
- The experimental collaborations have implemented aggressive data cleaning policies.
- The user efficiency is better than planned (with the exception of ALICE).

Overall assessment

Generally speaking, the experiments' computing models and the WLCG have demonstrated in a remarkably smooth way their capability to record, distribute and analyze the substantial amounts of data delivered to them by the LHC during the current year.

The performance of the WLCG throughout the year has been regular and without any noticeable difficulties, with periods where usage has been quite intensive corresponding to simulation and reprocessing campaigns and physics analysis activities jumping to high levels just before the seasonal winter and summer conferences such as the EPS conference, the Lepton-Photon conference or the Quark Matter conference. Noticeable peaks of individual user analysis have been detected in the weeks preceding these events but the efficiency of the system remained high throughout (with one exception already mentioned). The increased efficiency has merited a recommendation from the CRSG to revise the efficiency for Tier 2 CPU.

The GRID fabric works reasonably well, data distribution and network performance are excellent. A similar comment applies to the middleware.

The data placement policy and the detailed computing activities have been quite different from what was envisaged in the computing models, in the direction of making the some of the computing models less ambitious but more realistic by becoming more hierarchical and organized. The new policies represent a saving in the disk resources, but may impact on the network requirements in the future.

On the contrary, the reprocessing policy is quickly converging to the one indicated in the computing models as the number of events disfavours frequent reprocessing.

CPU resources are generally exceeding the experiments' needs at this point and the experimental collaborations have had substantial headroom that they have employed to increase simulation production; some collaborations have produced nearly three times more MonteCarlo events than indicated in the last revision of the computing models. Although less visible, there is also some headroom for disk, but this is not automatic now because it requires good data management policies.

Some collaborations and sites have experienced inefficiencies associated with large memory footprints for computing jobs. This is caused by events generally being larger than expected, which is in part due to the relevance of pile-up with 50ns bunches. This issue merits a more detailed study in the months to come in order to optimize resources.

Both ATLAS and CMS make a commensurate usage of CERN and Tier 1 CPU resources. CERN usage has represented for both major collaborations 18% of the total of the CERN+Tier 1. We note a smooth evolution of the two computing models in the direction of making them converge towards 20% usage at CERN, which seems reasonable. However, in the interest of the long term sustainability of the WLCG concept this figure cannot grow.

ALICE usage of CERN resources compared to external resources is very large. ALICE relies enormously on CERN because CERN has until now been the only provider to pledge resources matching the request. This model is now reaching its limits as now there is real pressure on the necessarily limited CERN resources. As emphasized elsewhere we recommend a revision of the ALICE computing model.

LHCb has decreased their previous inordinate reliance on CERN resources to a more reasonable level.

Recommendations

Specific questions related to the experiments' requests are deferred to the separate scrutinies of the different experiments.

- The CRSG is very concerned about the medium term sustainability of the ALICE computing model and about the low user efficiencies that have been reported to us. We recommend a substantial revision of this model. Strategies that have been proved to be successful elsewhere should be adopted. The collaboration is already working hard on this issue, but we suggest that it be a definite priority for action during the 2013 LHC stop if not already fixed by then.
- There is still room for improvement in the implementation of efficient staging strategies and/or dynamic data placement policies.
- The WLCG accounting of Tier 2 resources is improving steadily but is still insufficient: the Installed CPU compared to the pledged and installed disk capacity at the Tier 2 centres is not centrally accounted so far. It would be useful to disentangle the efficiency of organized/chaotic activities.
- Care should be taken that the worldwide LCG resources are used as much as possible as there may be a tendency by collaborations to place heavier demands on CERN resources or suggest that a larger than originally planned part of their analysis should be done at CERN. A balanced use of the resources is essential for the long

time coherence of the WLCG. A continuous increase in the request for additional resources at CERN is not sustainable in the medium term.

- The CRSG encourages close collaboration of the different centers with the experiments to continue the implementation of intelligent storage management policies to allow efficient and cost-effective access to data.
- We encourage that a discussion on the issue of the memory footprint is undertaken as soon as possible. We note that in some cases lack of memory makes half the cores unusable.
- The implications for best-use of resources of the interplay between improvements in network bandwidth and dynamical data placement policies should be evaluated.
- The CRSG recommends that CERN's policies of resource sharing when allocations are not fully used are clearly stated.
- In view of the current statistics we propose to make firm the April 2011 recommendation of assuming a Tier 2 efficiency of 67% for 2012 and increase this to 70% in 2013.
- While welcoming the experiments capability to record events at increased rates, the CRSG does not see how a substantial increase of the data taking rate could be accommodated with the existing computing resources and does not recommend a formal modification of the computing models in this direction.
- We encourage that the mechanisms of communication between experiments and the Tiers are continuously improved to optimize the usage.

On the scrutiny process

The CRSG is now satisfied with the quality and quantity of the information provided by the experimental collaborations. The experiments reports are well documented in the majority of cases. For future reviews we insist that the following good practices be maintained:

- All changes to the models compared to the previous review should be well documented
- All documents should be provided sufficiently early to allow time for the review, a deadline for the revised requirements should be agreed upon well ahead of the final report deadline. For the upcoming April 2012 C-RRB meeting this deadline is **March 1st 2012**.

CRSG workplan for 2012

1.- The CRSG plans to continue a close follow-up of the implementation of the ALICE computing model. Some substantial modifications are deemed absolutely necessary.

2.- The CRSG would like to follow closely the developments in dynamic data placement policies and have estimates of the eventual impact of these policies on network resources.

3.- In order to provide a better scrutiny, the CRSG would like to understand better the efficiency of disk management policies.

4 – We plan to monitor the consequences of the running conditions on pile-up, data sizes and reprocessing times.

5 – We plan to enter in conversations with the experiments and the LCG in order to address the memory footprint issue.

6- The CRSG shall continue to recommend a policy of mitigation of the request for new resources by optimizing the use of existing ones as much as possible.

On the CRSG membership

There have been no changes in the composition of the CRSG since the last C-RRB. During 2012 it would be necessary to start the process to renew or replace those members of the CRSG (including the chairman) that were not replaced during 2010 and 2011, in order to approximately comply with the MoU terms.

PART A

Scrutiny of the WLCG resources utilization in 2011

This report refers, unless otherwise stated, to the calendar year 2011, from January 1st to August 1st . In the April 2012 C-RRB a report for the full calendar year 2011 will be provided. The experiments are kindly asked to report before **March 1st** about the usage made during the previous calendar year

This report has used the following sources:

1.- Cumulative accounting from January to July for Tier 1s and CERN

https://espace.cern.ch/WLCG-document-repository/Accounting/Tier-1/2011/july-11/Master_accounting_summaries_July2011.pdf

2.- Month-by-month accounting of the CPU delivered by the Tier 2s

https://espace.cern.ch/WLCG-document-repository/Accounting/Tier-2/2011/july-11/Tier2_Accounting_Report_July2011.pdf

3.- The EGEE accounting portal at CESGA

http://www3.egee.cesga.es/gridsite/accounting/CESGA/tier1_view.html

4.- WLCG accounting reports for non-GRID CPU

5.- 2011 pledges as presented to the C-RRB

6.- The Tier-1 and Tier-2 Usage Reports kindly provided by Ian Fisk..

7.- The documents that the experiments have provided to the CRSG.

The following table describes the degree of usage of the different resources.

October 2011

Resource	**Site(s)**	**Used/Pledged** **Period average**	**Used/Pledged** **End of period**
CPU	CERN	**52 %**	**---**
	T1	**83 %**	**---**
	T2	**117 %**	**---**
Disk	CERN	**99 %**	**99 %**
	T1	**112 %**	**116 %**
	T2	Not available	Not available
Tape	CERN	**64 %**	**75 %**
	T1	**47 %**	**43 %**

For comparison we reproduce the analogous table presented in the April 2011 C-RRB that refers to the year 2010.

April 2011

Resource	**Site(s)**	**Used/Pledged** **Period average**	**Used/Pledged** **End of period**
CPU	CERN	32 %	---
	T1	62 %	---
	T2	122 %	---
Disk	CERN	75 %	110 %
	T1	89 %	110 %
	T2	Not available	Not available
Tape	CERN	52 %	69 %
	T1	52 %	60 %

The CPU figures correspond to a time average over the year obtained from averaging the monthly figures; those for disk or tape are usage relative to the installed capacity at the end of the accounting period.

The figures speak for themselves of the intensive use of the WLCG resources. The marked decrease in the use of tape at the Tier 1 is to be noted.

Efficiencies

The computing TDR estimated the efficiency to be 85% for CPU and 70% for disk in the case of organized (group driven) analysis, reducing to 60% in the case of chaotic (user-driven analysis).

The assumed efficiency is manifestly incorrect for the Tier 2. Even though a large fraction of the Tier 2 is already user/chaotic analysis (an exception is LHCb where only Monte Carlo production is carried out at the Tier 2), the statistics yield efficiencies that are much higher than 60% for three of the experiments.

The evolution of the numbers in the case of ALICE merits a separate discussion. While the overall CPU efficiency at Tier 2 is 50%, the one associated to chaotic analysis drops to 35% with a huge dispersion among users. This issue is elaborated further in the ALICE usage report below.

In view of these figures we recommend the revision of the official figure of 60% for Tier 2 CPU efficiency to 67% in 2012 and going up to 70% in 2013.

Efficiency of the utilization of the CPU at Tier 2s per experiment in January-August 2011 (left column) compared to the April 2011 report (right column)

ALICE	**50 %**	73%
ATLAS	**89 %**	85%
CMS	**80 %**	66%
LHCb	**98 %**	88%

Disk usage

Disk usage is difficult to analyse. A metric based exclusively on disk occupancy does not account for how frequency of access or how efficiently disks are managed.

We are aware of the technical difficulties involved and of the timescale needed to define and implement disk usage metrics. In the April 2010 C-RRB we required the experimental collaborations to provide the disk utilization in terms of the various data types involved (e.g. RAW, RECO, AOD, derived data, group data, user data and so on) and how frequently they were changed/replaced on disk in order to be able to assess the efficiency of disk usage.

So far, detailed information along these lines has been provided by LHCb. ATLAS provided the CRSG with a breakdown of different types of data only. No information has been received from CMS an ALICE.

From the available information the CRSG has been unable to verify the theoretical efficiency of disk, which is a very relevant ingredient of the total cost. We encourage the larger experiments to be more collaborative in this particular respect.

Sharing of the WLCG resources

The following tables give an idea of the use by the different experiments of the disk and CPU made available to them through the WLCG. The percentages refer to the fraction of the total mass storage, disk and CPU used per experiment (therefore all columns add up to 100% up to rounding errors). On the first (CERN+Tier 1) table the last column indicates which fraction of the total CPU that a given collaboration has used has been at CERN rather than using the T1's (and, consequently, does not add up to 100%). For comparison the percentages for April 2011 are shown in a separate table.

Percentage of use of the resources by experiment in January-July 2011 (CERN+Tier 1s)

Collaboration	**% of tape inT1+CERN used at end of period**	**% of disk inT1+CERN usedat end of period**	**% of CPU in T1+CERN used**	**% of which at** CERN
ALICE	**11 %**	**13 %**	**13 %**	**59 %**
ATLAS	**41 %**	**47 %**	**52 %**	**18 %**
CMS	**42 %**	**32 %**	**23 %**	**18 %**
LHCb	**7 %**	**8 %**	**12 %**	**28 %**

Percentage of use of the resources by experiment in 2010 (CERN+Tier 1s)

Collaboration	**% of tape inT1+CERN used at end of period**	**% of disk inT1+CERN usedat end of period**	**% of CPU in T1+CERN used**	**% of which at** CERN
ALICE	8%	6 %	16 %	33 %
ATLAS	35 %	57 %	59 %	14 %
CMS	52 %	29 %	17 %	20 %
LHCb	5 %	9 %	9 %	46%

The metrics show noticeable changes with respect to last April's report (that referred to the year 2010). The percentage of the total computing done at CERN has increased very substantially for ALICE. It has increased only slightly in the case of ATLAS, and it has decreased markedly below their original computing model in the case of CMS. LHCb also reports a substantial decrease of the percentage of CPU used at CERN compared to the total CPU effort. The latter is in line with previous CRSG recommendations.

Percentage of use of the resources by experiment in 2011 (Tier 2s)

Collaboration	% of CPU in T2 used (October 2011)	% of CPU in T2 used (All 2010)
ALICE	**6 %**	7 %
ATLAS	**54 %**	59 %
CMS	**31 %**	30 %
LHCb	**9 %**	4 %

Delivered versus pledged

The overall level of fulfilment of the pledges can be seen from the following table.

Resource	Site(s)	Installed / pledged
CPU	CERN	**100 %**
	T1	**99 %**
	T2	**117 %***
Disk	CERN	**100 %**
	T1	**101 %**
	T2	Not available
Tape	CERN	**100 %**
	T1	**80 %**

The figures refer in all cases to the end of the reporting period.

The turnout in CPU at the Tier 2 indicates that the percentage installed is actually above 100%. However automated accounting of this is still not in place. The figure indicated (*) is the delivered versus pledged.

It should be noted that several centres such as PIC, RAL and TRIUMF have installed disk capacities well above their nominal pledges.

The installed/pledged figures regarding tape indicate that there is relatively low use of this commodity.

Usage by the individual experimental collaborations

In what follows CPU usage refers to the average over the period. Disk and tape usage refers to the occupancy at the end of period. Units are kHS06 and PB for CPU and memory, respectively.

ALICE

		2011 request	2011 pledge	2011 usage	2012 request	2013 request
CPU/kHS06	T0+CAF	62	62	74 (82)	120	120
	T1	117	71	54 (57)	165	165
	T2	121	81	68 (78)	171	171
Disk/PB	T0+CAF	6.1	6.1	3.6	14.2	14.3
	T1	7.9	5.5	3.7	14	14
	T2	6.6	7.3		8	8
Tape/PB	T0+CAF	6.8	6.8	5.9	20	24
	T1	13.0	8.0	2.0	17	21

[1] “2011 request” is the one made in October 2010. Larger storage requests were made for the April 2011 RRB. Pledges are from the April 2011 RRB.

[2] ALICE usage for CERN and T1s from WLCG Master Accounting Summary for January to July 2011. T2 CPU usage is from EGI Accounting Portal, www3.egee.cesga.es for January to July 2011.

[3] CPU is total normalised elapsed wall time in kHS06 days from January to July, divided by the number of days (212). Disk and tape values are at the end of the period. Since the pledges were updated in April 2011, the values in parentheses are average CPU capacity used for the period April to August 2011.

[4] 2012 and 2013 requests are from ALICE, September 2011.

[5] We do not have independent access to accounting for T2 disk usage.

Comments on the ALICE usage report

So far in 2010 and 2011 ALICE has recorded 9e7 PbPb events for a total raw data size of 0.8 PB and 2.5e9 pp events with a total raw size of 1.9 PB. There is also 0.5 PB of calibration data and a relatively small set of cosmic events (2 10^8 events, 15 TB).

The average CPU power used at CERN, 74 kHS06, is larger than for the other three experiments. Moreover, for ALICE the ratio of CPU at CERN to that at T1s is 138% for the period January to July 2011 compared to 28%, 34% and 44% for ATLAS, CMS and LHCb respectively. ALICE continues to use a far greater proportion of CPU resources at CERN than the other experiments, mainly because CERN has until now been the only funding agency to meet the collaboration’s requests in full. Now that CERN resources are saturated, ALICE risks losing the relative enhancement at CERN that they have enjoyed hitherto.

Since the pledges were updated in April 2011, we also considered the average CPU power used for the period April to August. The T1 resource is apparently underused. However, ALICE's preparations for the 2011 Quark Matter conference had a large effect in April and May; T1 usage from June onwards (and prior to April) is very close to the pledge.

ALICE has suffered from declining efficiency of CPU usage during 2011. At the September LHCC Referees' meeting, the collaboration quoted average efficiencies of 90% for MC production, 66% for reconstruction and scheduled analysis, but only 35% for user analysis. There were noticeable drops in CPU efficiency at all tiers in the run-up to major conferences when user analysis peaked. The collaboration has identified some causes for this (including network transfer of many small files of calibration data) and noted an enormous variability in the efficiency of individual user analyses. Some measures have been taken to address the efficiency issue (upgrading the server infrastructure and changing the way calibration data is accessed) and indeed efficiency has been rising since July 2011.

The collaboration now saves to the T1s only that RAW data which passes quality tests for "physics" content. This has significantly reduced the T1 tape usage and is reflected in a reduced request for the rest of 2011.

ATLAS

Resource	**Site(s)**	**Pledged [1]**	**Used [2]**	**Used/ Pledged**	**Average CPU efficiency**
CPU (kHS06)	T0+CAF	75	64	85 %	66 %
	T1	250	224	90 %	88 %
	T2	281	324	115 %	89 %
Disk (PB)	T0+CAF	7.0	4.9	70 %	-
	T1	27	22	82 %	-
	T2	34	15	44 %	-
Tape (PB)	T0+CAF	12	13	108 %	-
	T1	32	15.5	48 %	-

[1] pledged resources from April 2010 RRB (.web.cern.ch/lcg/Resources/WLCGResources-2010-2012_15DEC2010.pdf).

[2] storage information from WLCG accounting summaries (https://espace.cern.ch/WLCG-document-repository/Accounting/Tier-1/2011/july-11/Master_accounting_summaries_July2011.pdf) and CMS. CPU usage from the EGI accounting portal http://www3.egee.cesga.es/gridsite/accounting/CESGA/egee_view.php

Comments on the ATLAS usage report

The status of the ATLAS offline computing, the resource usage and the 2012-2013 requests were shown at the LHCC referees meeting held at CERN on September 20th, 2011. ATLAS has provided a detailed report on the computing resource usage in 2011, compared with the predicted needs.

ATLAS has recorded data effectively at a rate of 275 Hz (nominal rate is 200 Hz) without difficulties. If calibration triggers are included, ATLAS rate goes above 500 Hz. Compression for RAW data is reported to work efficiently and AOD and ESD volumes are lower than expected. The collaboration reports a pile-up slightly below expectations (it is ~15 at the beginning of fills at present).

ATLAS is continuously improving their physics analysis model, toward using derived data (D3PD) which are Root N-tuple files. While these are currently small enough to fit on local resources, they expect an increasing volume that will thus require grid usage and therefore an increase in user analysis on T2 sites. This may increase to some extent the need of disk at these sites, so far substantially underused.

The ATLAS data distribution policy was modified last year for RAW and ESD data, and a dynamic data placement strategy at T2 resulted in aggressive clean-up of disk space. The bulk of the ATLAS collaborators are not using ESD for general analysis any more. Progress has been made in various parameters with respect to previous expectations. Compression for RAW data on disk allowed reducing their size by a factor of 1.5. Data size per event and reconstruction times have been reduced as well.

ATLAS reports good progress in the implementation of Fast MC simulations. In fact ATLAS has done about three times more simulated data than originally planned, thus indicating that CPU and disk resources are more than sufficient at the present stage.

The implementation of the ATLAS computing model seems satisfactory. We note however the low degree of usage of disk at the Tier 2 and tape at the Tier 1. The ATLAS computing management has been asked to reduce their request of these two resources in view of their historical record of lower than optimal usage.

CMS

Resource	**Site(s)**	**Pledged [1]**	**Used [2]**	**Used/ Pledged**	**Average CPU efficiency**
CPU (kHS06)	T0+CAF	106	34	32 % (22 %)*	55 %
	T1	132	86	65%	88 %
	T2	315	259	82%	80 %
Disk (PB)	T0+CAF	4.5	3.5	79%	-
	T1	16	14	87%	-
	T2	20	14	70 %	-
Tape (PB)	T0+CAF	22	10	46%	-
	T1	44	21	48%	-

[1] pledged resources from April 2010 RRB (.web.cern.ch/lcg/Resources/WLCGResources-2010-2012_15DEC2010.pdf)

[2] storage information from WLCG accounting summaries (https://espace.cern.ch/WLCG-document-repository/Accounting/Tier-1/2011/july-11/Master_accounting_summaries_July2011.pdf) and CMS. CPU usage from the EGI accounting portal http://www3.egee.cesga.es/gridsite/accounting/CESGA/egee_view.php

(*)Last year's % of usage.

Comments on the CMS usage report

The performance of the CMS computing system was generally smooth throughout the whole period.

CMS presented their current status at the LHCC Referees' meeting held at CERN on September 20-th. They showed their performance numbers for the 300Hz trigger (375 HZ if 25% overlap is accounted for) taken during 2011 running. During this period the machine's performance averaged to about what was planned and CMS software performed close to predictions. Fills were observed to start with a pileup of ~15 and average 7-8 over a run. Vertex reconstruction is still approximately linear with luminosity.

CMS provided the CRSG with a Resource Utilization document detailing their performance. Event sizes generally agree with estimates. CPU utilization @ the T0 was much lower than expected but improving with respect to 2010. The average Tier-1 CPU utilization was > 80% with 21PB of tape used and 14PB of disk used.

CMS has successfully made the transition to a 64 bit code base and a new application I/O. Their latest reconstruction code is > a factor of 2 faster and has a somewhat smaller memory footprint. Even with this speed up the reconstruction time for events with a pile-up of 12 take 11 sec/event which was expected to happen at a pile-up of 16.

Their Tier1 efficiency is high (87%) and the Tier2s are running with > 80%.efficiency. They appear to have successfully made a transition to AOD usage for their analysis.
In recent months they have been running > 200k analysis jobs/day and > 800 individual users/month.

In general the computing details are close to expectation from their computing model with a few exceptions.

CPU utilization at CERN was lower than desired on average with an CPU efficiency of 55 %. Much of this inefficiency seems to come from a large memory footprint of the RECO code that doesn't allow for multi-core utilization. The CMS collaboration have been working on this memory issue but more work looks to be needed.

MonteCarlo RECO event sizes were ~50% larger partly due to out of time pile-up too.

Resource	Site(s)	Pledged	Used	Used/ Pledged
CPU (kHS06)	T0+CAF	21	20	95 %
	T1	70	57	81 %
	T2	48	65	135 %
Disk (PB)	T0+CAF	1.5	0.9	60 %
	T1	3.8	2.0	53 %
	T2	--	--	--
Tape (PB)	T0+CAF	2.5	1.1	40 %
	T1	3.9	1.7	44 %

Comments on the LHCb usage report

The usage of the computing resources by LHCb displays a rather healthy situation. Also for 2011 the experiment is successful in using the resources and adapting to the contingent situation of higher interactions rate: LHCb has operated at constant luminosity ~$3*10^{32}$ which implies a constant trigger rate (3kHz as expected), and consequent data volumes proportional to time in colliding beams (pileup is as expected).

The Tier0/1s CPUs have been constantly used for Monte Carlo production and then used to absorb the peak requests of data processing during the LHC running. However to cover the apparent lack of resources obtained at the Tier0/1s, some extra resources from Tier2s, together with resources that LHCb has obtained from sites that are not official LHCb Tiers, were commissioned for the full reconstruction of the end of the 2011 data-taking period and to allow getting all the data reprocessed in time for the 2012 winter conferences. With these resources an “end of year” reprocessing is scheduled to start end of September 2011 and lasting 2-3 months.

The alternative use of the Online Farm, connected to the CERN Storage, potentially envisaged during the last April 2011 scrutiny process, was not finally considered for this campaign.

Here the CRSG wish to point out and appreciate the efforts done by the LHCb collaboration in: 1) re-addressing the quite extreme peak load for CERN T0 resources. Now the CERN share is of the order of ~ 20% and in line with the pledge; 2) providing information on the jobs failure, the improved overall efficiency which also demonstrate the efforts spent to further increase the resource utilization.

The use of storage resources is the part of the LHCb computing model where more changes, like the reduction in the number of replicas, have taken place. It is appreciated the detailed information provided by the LHCb collaboration explaining the major aspects of the storage reorganization as well as the critical analysis of previous choices turned out be impractical.

This year LHCb expects to collect raw data for 50kB/event * 3kHz trigger rate * $5*10^6$ seconds = 750 TB for a single copy. These data are not kept on disk. These data correspond to 450 TB size of reduced DST format of a single copy to be stored and an average ~12% (retention rate) "stripped" analysis DSTs corresponding to ~ 300 TB to be stored. At the end of the reprocessing

LHCb in synthesis will have 4 copies * 300TB plus 4 copies * 450TB = 3.0 PB storage foreseen, which will be then reduced the to just two copies, freeing 2*450TB.

After the total reprocessing will be completed a total 3.0 PB will be reduced to 2.1 PB by rigorous “clean-up”. However more storage is required to be ready by early 2012 data taking. A preliminary estimation for 2012 foresees a shortfall of about 2 PB in required disk resources. The LHCb collaboration presents three possible scenarios: a) reducing further the data copies living room to the next MC production; b) drastic reduction of existing MC disk copies; c) delaying the MC 2011 production to spring 2012. The CRSG suggests to perseverate in the approach of reorganizing the data storage and likely find room by critical clean-up (see options a) and b)) to be ready for new MC production. However, the updated resources estimate for 2012, is planned for the March 2012 RRB and there are still many unknowns that will affect parameters of the simulation. The LHCb collaboration let us know that in order to accommodate the foreseen MC production needs between now and March 2012 (generation of about 1 billion events with much improved physics tuning of GEANT4+Generators, under 2011 beam and detector conditions) they will have to drastically reduce the amount of existing MC kept on disk. Current disk usage of existing MC is about 1.8 PB, the new sample will require 1.2 PB in addition. LHCb does not expect any increase in pledged disk storage before April 2012, so this will have to fit into the ~2.5 PB of storage available for MC by March 2012 (after recovering the 0.9TB from the first pass real data reconstruction described above).

In March 2011 it was proposed to keep two RAW data tape copies at CERN and one distributed across the Tier1s. Following the discussion at the last WLCG RRB the collaboration has dropped this request.

It was reported that the number of jobs from LHCb executing at CCIN2P3 during the spring was lower than what would be expected from the pledged resources. Since July the usage of IN2P3 by LHCb is at the expected level, so the "shortfall" is gone. This inability took some time to resolve because of a communication issue. Now the problem is solved, so the LHCb collaboration considers this issue closed.

In conclusion, the LHCb computing model is robust and was able to incorporate parameter adjustments and a satisfactory usage of the resources.

PART B

Scrutiny of the requests for 2012 and 2013

ALICE

There have been changes to some of the computing model parameters used in calculating requests for 2012 and 2013.

The data size of a raw pp event has more than doubled since the April 2011 request because of increasing pileup in pp running and a change in the trigger mix. For PbPb events, increasing event complexity has been balanced by compression, leaving the raw data size unchanged. Compression by a factor of 2 is assumed for the 2012 and 2013 requests, but the collaboration is confident that they can achieve a factor of 3.5 (resulting from a factor of 5.5 compressions of their TPC data with the rest of the event record remaining about the same size) after replacing hits by clusters in the HLT. Algorithms are in place and being tested. ESD and AOD event sizes increase with the complexity of the events and so have increased for both pp and PbPb.

ALICE intends to take advantage of the heavy-ion event compression to allow an increase in trigger rate, maintaining saturation of the DAQ data-transfer rate and resulting in even more complex (biased by the trigger) events that will increase the aggregate reconstruction time.

CPU use per event for pp reconstruction and Monte Carlo is unchanged, but there have been significant reductions by factors 2.8 and 2.5 respectively for PbPb reconstruction and MC simulation, where the memory footprint has been decreased.

The September 2011 ALICE resources request assumes recording 2 10^8 PbPb events in the 2011 and 2012 heavy-ion runs, whereas in the computing model a standard data-taking year assumed 10^8 such events. The T0 CPU request is essentially fixed by the need to reconstruct the heavy-ion events in 4 months before the subsequent pp run starts.

The collaboration's desire to record and process more events offsets improvements in CPU use per event and leads to more ESDs on disk. A further increase in capacity would be needed to keep ESDs from previous years on disk. In combination this would produce disk requirements at all tiers which would be unsupportable in practice. In response to this the collaboration has re-evaluated its requirements, keeping multiple copies of ESDs for both real and MC events on disk only for the current reconstruction pass and one copy (or none) for earlier passes. This keeps the disk request closer to that made in April 2011, with some increase at T1s. However, the request still exceeds 2011 pledges by almost a factor of two at CERN and the T1s. We do not expect 2012 pledges to increase significantly (although ALICE should gain access to an additional T1 in Korea and possibly one in Mexico), so even the revised request looks unrealistic.

Recommendations

We encourage ALICE to review their overall data flow aiming to:

- Establish that the target compression ratio of 3.5 for raw PbPb events can be achieved and use this to reduce the amount of data stored for offline analysis. If they achieve this then they could record and analyse a correspondingly larger number of events within the same computing envelope.
- Classify their computing jobs to identify and improve those with the lowest CPU efficiency. User analyses is highlighted, but look for groupings within this set or enforce stricter use of analysis wagons in reconstructions, limiting user access to data for smaller, inefficient jobs if necessary.
- Continue to optimise all computing jobs to reduce CPU requirements (as has been done for PbPb reconstruction and simulation this year). It may be necessary to revisit the decision to rely on Geant3 simulation; parametric simulations, like those used by the other LHC experiments, may need to be developed.

If the requested resource requirements cannot be met, we note that ALICE has determined optimum disk/CPU ratios, 0.1 PB/kHS06 for T1 and 0.07 PB/kHS06 for T2, to enable the best use of what is available.

It seems imperative for a collaboration that is under-resourced to strive to match the efficiency of the other LHC experiments. ALICE is indeed already making strenuous efforts to address this, but is hampered because their computing model fully embraces the Grid and any computing task can run at any site. In a situation with ongoing under-resourcing, we think that ALICE may have to consider a retreat to a more hierarchical model, particularly for user analysis.

ATLAS

Generally speaking, the ATLAS forecast of resource usage by the end of 2011 and 2012 is unchanged with respect to April 2011. The mean number of interactions per beam crossing is expected to be as large as 25. Major reprocessing campaigns of the current and future datasets will also be performed, which imply a a significant increase in usage (both CPU and disk). ATLAS claims that "usage will not be a linear extrapolation from the usage Jan-July, but will have larger upward jumps."

The ATLAS requests are the same as in April 2011, with the notable exception of CERN CPU. The request for additional CPU at the Tier0 was motivated by expected improvements in the LHC duty cycle and increased luminosity and pile-up. The CRSG understands that this request is justified and has the support of the CERN management, with the caveat that the fraction of usage of resources at CERN should always be kept below certain limits.

Upon a request of the CRSG the ATLAS collaboration has agreed to reduce their tape request at the Tier 1s. The CRSG has also requested ATLAS to revise their request for disk at the Tier 2s. While ATLAS prefers not to reduce their disk request at this moment, agrees with the CRSG to revise the disk usage early in 2012.

The scrutinized requests for ATLAS are contained in the following table

CPU [kHS06]	2012	2013
CERN	*73 → 111*	*111*
Tier-1	*259*	*273*
Tier-2	*266*	*289→ 281*
Disk [PB]		
CERN	*9*	*10*
Tier-1	*27*	*30*
Tier-2	*47*	*53*
Tape [PB]		
CERN	*18*	*18*
Tier-1	*36 → 29*	*40 → 33*

CMS

CMS has made a request to increase their disk space at the Tier-0 to "improve the utilization of the CERN resources during non-peak periods". This would enhance their analysis job throughput. If this increase in disk space would allow more jobs to run on the available memory presumably this would help with their low CPU utilization.

Other than that the April 2011 recommendations remain valid; they are reproduced below for convenience. In view of the current level of usage the CRSG has asked the CMS collaboration to reduce their Tape at Tier 1 in the medium term. Conversations with the collaboration are proceeding and the C-RRB will be informed.

CPU [kHS06]	2012	2013
CERN	*120*	*120*
Tier-1	*145*	*145*
Tier-2	*315*	*315→306*
Disk [PB]		
CERN	*5 →7*	*7*
Tier-1	*22*	*27*
Tier-2	*26*	*26*
Tape [PB]		
CERN (including HI)	*23*	*23*
Tier-1	*51*	*59*

LHCb

In order to figure out the updating of resources estimate for 2012 (that is planned for the March 2012 RRB) the LHCb collaboration cannot make any precise prospects until the LHC running conditions for 2012 will not be fixed: there are still many unknowns, in particular the LHC beam energy and whether the experiment will run at 25 ns bunch spacing, that will affect the parameters of the MC simulation.

In the absence of additional information the April 2011 remains valid. It is reproduced below for convenience.

Date	Site	CPU (kHS06)	Disk (PB)	Tape (PB)
2012 period	CERN	*34*	*3.5*	*6.4*
	Tier-1	*113*	*9.5*	*6.2*
	Tier-2	*43*	*0*	*0*
2013 period	CERN	*33*	*4.0*	*7.7*
	Tier-1	*110*	*11.1*	*8.0*
	Tier-2	*42*	*0*	*0*

PART C

Usage of the Tier 2's