CERN-RRB-2012-116

Computing Resources Scrutiny Group

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

INTRODUCTION

This report summarizes the deliberations of the CRSG regarding the usage of the computing resources by the four main LHC experiments (ALICE, ATLAS, CMS and LHCb) during the first eight months of 2012.

We have also re-examined updated requests for 2013. Due to the extension of the pp run from 22 to 30 weeks the experiments have submitted revised estimates for the ‘computing year’ starting April 1st 2013. Tentative estimates for 2014 are included.

Finally, this report contains some comments for the period extending beyond 2014. These should be taken as very preliminary but they can be of some help both to experiments and to funding agencies for medium-term planning.

2012 is an exciting year for particle physics. Computing has been a crucial ingredient of the successful LHC run leading to the discovery of the Higgs-like particle announced on July 4th. The run now continues in order to establish the properties of this particle.

The CRSG wishes to express its praise for the outstanding performance of the LHC, of the four experiments under review at this RRB and especially of the WLCG.

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 reassessment of the 2013 request and a preliminary estimate for 2014. Part C contains some advances on the 2015 and beyond requests.

The LHC running conditions

The planning estimates assumed by the CRSG for 2012 have generally been reflected in reality as the running year has progressed.

At the Chamonix meeting held in February 2012 it was decided to increase the energy to 8 TeV (4 TeV + 4 TeV) providing slightly larger cross sections for the processes of interest. About 10% of the time was expected to be dedicated to heavy ion (HI) physics. This schedule was changed following the July 4th announcement adding about 30% more time for pp physics, increasing the total run time from 22 to 30 weeks into mid-December with the pPb run being postponed to January 2013 according to current plans.

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 ideal scheme displayed in the first row of the following table was assumed.

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	HI
2012	**April '12**	**March '13**	**8**	**5.9**	**5.2**	**0.7**
2012	***April ’12***	***March ‘13***	***10***	***7.2***	***6.5***	***0.7***
2013	**April ‘13**	**March ‘14**	**-**	**-**	**-**	**-**
2014	**April ‘14**	**March ‘15**	**-**	**-**	**-**	**-**

In the period March-August 2012 the total pp live time was 3.8 Ms well in line with the existing estimates that were used in the previous scrutiny. The total live time reflects the excellent performance of the machine close to the ideal maximum as well as the readiness of the collaborations to take large amounts of data.

After the extension of the pp run by a further two months the machine live time is expected to amount to about 7.2 Ms. This revised scheme is displayed in the second row of the table. No data taking is forecast until spring 2015 when the LHC will start operation with an energy close to the design value.

In addition to achieving a beam time in line with the most optimistic expectations, experiments have been recording at rates substantially larger than the nominal ones. The following are the average real/nominal rates for pp events:

ALICE: 400 for pp, 560 for pPb /100 Hz

ATLAS: 340 prompt + 150 delayed /200 Hz

CMS: 375 prompt + 300 delayed /300 Hz (includes 25% data set overlap)

LHCb: 4000 + 1000 deferred/ 2000 Hz

These increased trigger rates were adopted after validating extensions of their respective physics programmes by the LHCC, taking advantage of some headroom in CPU capacity and reduction in data sizes to increase their rate. The LHCC in 2011 warned against a very substantial increase in the trigger rates on the grounds that computing resources likely could not be increased in a matching proportion. The CRSG endorsed this recommendation, adding that sustainability of the WLCG in the medium and long run requires a smooth budgetary profile in the present circumstances. Deferred data is to be processed during the long shutdown starting in 2013.

After the excellent performance of the LHC, the limits of the existing resources appear to have been reached and some collaborations have limited further increases in data taking rate that were previously envisaged.

Pile-up has represented a renewed challenge for the experiments. Proton bunches are injected with a minimal separation of 50ns rather than the design value of 25ns. To compensate for the reduction in number each bunch contains more protons and they are squeezed as much as possible to sustain as large a luminosity as possible. The consequence is the appearance of events with many interactions (pile-up). This has a substantial impact both on reconstruction times and on the size of the data sets which are larger than expected because of the increased pile-up with respect to the design conditions. Out of time pile-up is also observed. During the elapsed months of 2012 pile-up has averaged to 20 interactions per crossing and can be as large as 40 just after a fill. At the LHCb IP the average multiplicity has been 1.7 (below the 2.5 peak value for 2010 but above the value of 1.5 averaged in 2011 and in any case above the design multiplicity). All things considered, pile-up now seems less of a threat to the efficient running of the experiments than it appeared in previous scrutinies but it has a definite impact on resources.

During the early months of 2012 the collaborations reconstructed and analysed the events recorded during the second PbPb run. This run impacts mostly on the ALICE collaboration which has revised some computing model assumptions in view of past experience.

A satisfactory trial pPb run took place in September in preparation of the HI run in early 2013. Unlike in 2011 LHCb plans to take data during the pPb period.

Interactions with the experiments

Documents from the experimental collaborations were received shortly after the September 1st deadline. The referees exchanged sets of Q & A with the computing representatives to clarify various points in the respective reports. Several meetings with the collaborations took place too to reach a reasonable consensus. The CRSG is satisfied with the information provided by the experiments and the WLCG collaboration and thanks in particular the computing coordinators for their availability. For the upcoming April 2013 C-RRB meeting the deadline to submit the relevant documents to the CRSG is **March 1st 2013**.

As agreed with the ATLAS and CMS management the scrutiny procedure for these two experiments is done by a single team of referees, using common techniques and methods, ensuring that a coherent set of principles is applied.

The CRSG had asked the ALICE collaboration in the past to submit requests more aligned with the expected resources and in this way facilitate a realistic scrutiny. Finally, this experiment submitted in April 2012 a request roughly in line with the expected pledges, implying a sizeable reduction in their Tier 1 request. Unfortunately this reduction had an unexpected side effect: ALICE found shortly afterwards that some pledges would be reduced in proportion. In particular the collaboration was extremely worried by a reduction in the contribution of GridKa. In June the CRSG expressed their concern to the GridKa director recommending that no reduction of the absolute pledge is adopted. While we applaud the realistic move of the ALICE collaboration, it is clear that if funding agencies reduce their contribution in the same proportion the gap between requests and pledges will never be bridged. We therefore ask for the funding agencies' understanding in this particular point.

Interactions with the LHCC

Since the last scrutiny no issues appeared which we thought necessary to refer to the LHCC.

On September 25th and in preparation of the LHCC review, the CRSG was invited to attend a series of presentations by the WLCG collaboration and the four experiments' computing representatives. The CRSG found this joint meeting quite productive and both the LHCC and the CRSG vowed to intensify their exchanges.

In this meeting the overall performance of the WLCG over the past months was reviewed. In total 19 PB of data have been written in 2012, with a monthly rate close to 4 PB. The scale of the CASTOR storage system approaches 100 PB, with data transfer rates increased to 3-4 GB/s input and ~15 GB/s output. When LHC is running, CERN exports around 2 GB/s data, and WLCG transfers often exceed 10 GB/s overall.

For the LHC experimental program as a whole, the extension of the 2012 run implies a need for some additional resources. The estimates for 2013 have been revised to take into account the effects of the extended run, and 2014 requests are close to those of the revised 2013 ones. These additional requests will be discussed latter in this document.

During the 2013-2014 shutdown computing activities will include a full reprocessing of the complete 2010-12 data sets, simulations for 13 TeV running, and physics analysis.

The experiments also discussed in some detail the expected needs for 2015 and beyond. This will be commented upon in part C.

Overall assessment

The CRSG sees now a massive use of all the available WLCG resources. The GRID fabric works well; data distribution and network performance are excellent.

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, and the urgency to analyse large amounts of data in a short time, that represented a real challenge for the computing models and for the WLCG as a whole, have been put to a test and the challenge has been passed very successfully.

The pressure on grid resources has been considerable, particularly in the weeks preceding the July 4th announcement but the system has responded adequately.

The collaborations have implemented more realistic and more organized data distribution policies. The number of reprocessing passes has decreased dramatically as the number of events disfavours frequent reprocessing. Not surprisingly, we detect a more efficient use of the resources in those collaborations where the computing model tends to favour organized analysis and a more hierarchical structure, but with one exception the efficiency is high overall.

We welcome the efforts of some collaborations to provide a more detailed information on their disk usage including details on the popularity of the various files and datasets. Armed with this information, several experiments are implementing dynamical data placement policies, taking advantage of the good connectivity. This represents a substantial change with respect the original provisions of the MONARC model and allows for a mitigation of the need for disk. The number of copies stored in Tier 1 or Tier 2 has been reduced and more compact datasets are now used for analysis. The collaborations have been very active in redistributing tasks among CERN, Tier 1 and Tier2 to optimize the usage of resources and an equilibrated distribution of the usage to the point that the boundaries between Tier 1’s and some large Tier 2’s are disappearing to some extent

This optimisation has allowed them not only to cope with the increasing amounts of data generated by the excellent LHC performance and the more complex events due to pile-up, but also allowed them to record at the increased rates indicated before.

As indicated, the experimental collaborations are now making full use of the resources made available to them by the WLCG participating centers. Experiments have set up task forces to reduce processing times and they have generally improved, partly under the pressure to deal with increasing values of pile-up and partly due to the desire to be able use more bandwidth within the existing resources. While there is surely an asymptotic regime to improve the reprocessing times with decreasing gains, there is probably considerable room for improvement in the other major ingredient in the CPU budget, namely Montecarlo production. As of now, experiments are able to simulate many more events than envisaged in their computing models, but the impact of simulation is growing rapidly and may become unsustainable. More manpower effort needs to be dedicated to optimize Montecarlo production.

Some experiments plan to use their HLT farms (or parts thereof) for reprocessing or MonteCarlo production during 2013 and 2014. We encourage this line of development aiming to the use of these resources not only during long shutdowns but in all periods when the beam is off.

We welcome the action taken by some experiments to take into account their non-WLCG resources making the scrutiny process more transparent.

Recommendations

Specific questions related to the experiments' requests are deferred to individual scrutinies (part B of this report). We reiterate the recommendations of actively pursuing the reduction of reprocessing times, reduction of file sizes, optimizing the number of copies distributed across the tiers, removal of unused data to tape, and using derived data sets for analysis. In addition

- We recommend intensive use of the HLT farms during 2013 and 2014 for reprocessing and simulated data production. The possibility of using the HLT farms for MC production or reprocessing during periods with no beam should be thoroughly investigated.
- We recommend an aggressive implementation of dynamical data placement policies and a close scrutiny of disk usage. The collaborations should provide data access statistics to demonstrate that the data placement policies are meaningful and effective. In this respect the information provided to date is not satisfactory and we ask the WLCG to compile statistics from the Tier1 and principal Tier 2 stating what fraction of the disk volume is seldom or never accessed, detailed by VO, and establish suitable metrics
- The CRSG is considering a revision of the disk efficiency for subsequent scrutinies. Some collaborations have suggested assuming a 100% efficiency if buffering for data is included in the disk budget. This concept needs precise quantification.
- If possible, we would like to ask the WLCG to automate the retrieval of the information that the CRSG uses in their reports.
- We anticipate the need for substantial progress in the implementation of fast MonteCarlo simulations to reduce significantly the relative weight of simulation in the computing budget. Experimental collaborations should be prepared to dedicate enough manpower to this task and report progress to the CRSG in subsequent scrutinies.
- We underline the importance of smooting out CPU needs throughout the year and consider the possibility of using external (cloud?) resources for very localized demands. The desire to present new results in winter and summer conferences has a substantial cost in terms of CPU peak demand.
- The CRSG encourages close collaboration of the different centres with the experiments to continue the implementation of intelligent storage management policies to allow efficient and cost-effective access to data. We consider this issue very relevant for the operation of the LHC experiments after 2014.
- We recommend that hardware upgrades at the T1 and T2 are in synchrony with the newly upgraded software to take full advantage of the technology available.

- We encourage the experimental collaborations to continue working on realistic estimates for the computing needs in 2014 and beyond, keeping the budgetary constraints in mind and working with the CRSG as necessary.

On the CRSG membership

Following the October RRB D. Espriu will leave the CRSG and a new chairman will be appointed. We remind the C-RRB that several CRSG posts are still pending replacement. In addition the CRSG recommends the inclusion of one or two additional members in order to have enough manpower to fulfill their mandate.

The chairman would like to thank the CRSG members for their dedication and the experiments' spokespersons and computing managers for their collaboration and understanding. Thanks are also due to the CERN management for the support provided.

Finally, the chairman would like to thank the funding agencies represented at the C-RRB for their generous support to the WLCG project.

PART A

Scrutiny of the WLCG resources utilization in 2012

This report refers, unless otherwise stated, to the calendar year 2012, from January 1st to August 31st.

This report has used the following sources:

1.- Cumulative accounting for Tier 1s and Tier 2s CERN

https://espace.cern.ch/WLCG-document-repository/Accounting/

2.- The WLCG accounting portal at CESGA.es

http://accounting.egi.eu/egi.php/

3.- WLCG accounting reports for non-GRID CPU

4.- The REBUS pledge portal

http://wlcg-rebus.cern.ch/apps/pledges/summary/

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

The following table describes the degree of usage of the different resources. The first set of tables (blue) compile general information, gotten from the WLCG accounting. The set of tables referring to specific experiments (yellow) use information obtained from the collaboration themselves. The latter have been cross-checked with the statistics from the accounting tools whenever possible.

October 2012

Resource	Site(s)	Used/Pledged Period average	Used/Pledged End of period
CPU	CERN	**58 %**	**---**
	T1	**95 %**	**---**
	T2	**167 %**	**---**
Disk	CERN	**102 %**	**99 %**
	T1	**131 %**	**129 %**
	T2	Not available	Not available
Tape	CERN	**84 %**	**86 %**
	T1	**63 %**	**72 %**

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 pledged capacity at the end of the accounting period.

For comparison we reproduce the analogous table presented during the April 2012 C-RRB that refers to the whole year 2011

April 2012

Resource	Site(s)	Used/Pledged Period average	Used/Pledged End of period
CPU	CERN	**55 %**	**---**
	T1	**93 %**	**---**
	T2	**166 %**	**---**
Disk	CERN	**105 %**	**119 %**
	T1	**121 %**	**137 %**
	T2	Not available	Not available
Tape	CERN	**75 %**	**97 %**
	T1	**47 %**	**51 %**

The figures show remarkable stability, indicate a full use of the existing resources and an increased use of tape storage. They also show the existence of resources beyond the pledged ones.

Efficiencies

The computing TDR estimates the efficiency to be 85% for CPU and 70% for disk in the case of organized (group driven) analysis (taking place mostly at Tier 1). The efficiency for user analysis is now estimated to be 70%.

The numbers in the case of ALICE merit some comments. The overall CPU efficiency is relatively low. This issue is elaborated further in the ALICE usage report below. Due to the implementation of the ALICE computing model the average efficiency for Tier 1 and Tier 2 is very similar.

Efficiency of the utilization of the CPU at Tier 2s per experiment during 2012 (left column) compared to the April 2012 report (right column)

ALICE	**60 %**	54%
ATLAS	**88 %**	88%
CMS	**84 %**	83%
LHCb	**95 %**	93%

Disk usage

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

The CRSG has reasons to believe that there is considerable room for improvement in disk efficiency (understood not as a mere percentage of occupancy).

During this scrutiny the CRSG has begun a dialogue with the experiments, aiming to find alternative or additional measures of disk usage. The experiments already record frequency and times of access to files and are using this information to guide their data replication and cleanup policies. Potentially a common synthetic metric for efficiency of disk use could be agreed on for future scrutinies.

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 reported in April 2012 are shown in a separate table.

Percentage of use of the resources by experiment in January-July 2012 (CERN+Tier 1s)

Collaboration	% of tape inT1+CERN used at end of period	% of disk inT1+CERN used at end of period	% of CPU in T1+CERN used	% of which at CERN
ALICE	11 %	14 %	11 %	48 %
ATLAS	38 %	45 %	56 %	15 %
CMS	42 %	33 %	26 %	29 %
LHCb	9 %	9 %	8 %	24 %

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

Collaboration	% of tape inT1+CERN used at end of period	% of disk inT1+CERN used at end of period	% of CPU in T1+CERN used	% of which at CERN
ALICE	12 %	14 %	15 %	52 %
ATLAS	39 %	46 %	51 %	17 %
CMS	41 %	33 %	23 %	21 %
LHCb	8 %	7 %	11 %	26 %

The metrics indicate stability in the implementation of the computing models. The dependence on CERN resources of ALICE has decreased slightly with respect to 2011. ATLAS has a large amount of resources outside CERN and, in addition, it was reported to the CRSG that the batch system at CERN gave a slow return for ATLAS jobs. CMS share of CERN resources has increased after some improvements (see April 2012 report) and is close to the fraction devised in the computing model. The LHCb fraction is likewise quite reasonable.

Percentage of use of the resources by experiment in 2012 (Tier 2s)

Collaboration	% of CPU in T2 used (October 2012)	% of CPU in T2 used (All 2011)
ALICE	**7 %**	11 %
ATLAS	**54 %**	54 %
CMS	**36 %**	33 %
LHCb	**3 %**	3%

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	**105 %**
	T2	**108 %***
Disk	CERN	**85 %**
	T1	**97 %**
	T2	Not available
Tape	CERN	**100 %**
	T1	**87 %**

The figures refer in all cases to the end of the reporting period. The Tier 2 CPU(*) percentage quoted is delivered/pledged. Concerning Tier 1's, KIT and NL-LHC were below 90% at the end of the accounted period.

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. Data are provided by the experiments and cross-checked with the WLCG accounting tools whenever possible.

ALICE usage

Resource	Site(s)	2012 request	2012 pledge	2012 usage	Efficiency
CPU/kHS06	T0+CAF	125	90	66	58 %
	T1	95	95	73	56 %
	T2	207	115(194)	136	60 %
Disk/PB	T0+CAF	7.8	8.1	8.4	
	T1	7.0	7.2	6.8	
	T2	12.4	9.1(12.9)	9.6	
Tape/PB	T0+CAF	17.1	20.0	9.7	
	T1	11.3	11.5	3.9	

The figures in parentheses for T2 cpu and T2 disk include non-WLCG resources (we thank ALICE for reporting these). The usage figures are supplied by ALICE and cover the period April to September 2012.

A new analysis facility came into use at KISTI, South Korea, as anticipated, but a South Korean T1 is not yet available. Negotiations are ongoing for T1 resources in Russia and Mexico, with India as a more distant possibility. We strongly encourage these efforts by ALICE to recruit additional resources.

CPU power at CERN and T1s is underused so far compared to the pledges but usage should increase with the extended pp run and the anticipated pPb run. Average shares of CPU use are 10% reconstruction, 11% analysis trains, 27% user analysis and 52% simulation. The collaboration is actively moving user analysis into the trains, which run more efficiently, and is continuing a switch to using smaller AOD formats instead of ESDs for analysis. However, the collaboration's efficiency of CPU use is still affected by large I/O requirements and a rather heterogeneous user community and it lags the other LHC experiments. CPU consumption per event has stabilized for RAW data and MC processing, but is still growing for analysis trains.

A new calibration strategy has been implemented, with a reconstruction pass on a small percentage of the new data. The first full reconstruction is now about as good as the previous second pass reconstruction, reducing CPU consumption.

ALICE has recently started to keep 2 instead of 3 copies of ESD's on disk but they are already essentially saturating the pledged capacity.

Tape use is expected to rise to meet the pledges with the remainder of the long pp run and, especially, the pPb run to come. Some reduction of tape use at T1s is achieved by copying to them only data of sufficient quality for further processing.

ALICE will be relatively less affected by the 2012 pp run extension than the other LHC experiments, given the relative importance of the heavy ion running in their computing needs.

ATLAS Usage

Resource	Site(s)	Pledged	Used	Used/ Pledged	Average CPU efficiency
CPU (kHS06)	T0+CAF	111	111	100 %	89 %
	T1	285	436	153 %	91 %
	T2	328	612	187 %	87 %
Disk (PB)	T0+CAF	9	6	67 %	
	T1	30	27	90 %	
	T2	45	30	67 %	
Tape (PB)	T0+CAF	18	20	111 %	
	T1	38	24	63 %	

ATLAS has been very successful in responding to heavy simulation requests from the physics groups by taking advantage of unpledged and opportunistic resources. This has allowed them to greatly exceed their pledged resources. Since these resources were available they ran substantially more full simulations than fast ones and hence the large CPU numbers in the table.

The full simulation time has improved to 3100 HS06sec/event. As noted in the spring scrutiny ATLAS is able to make substantial usage of the fast simulation for many of their studies and are developing an integrated simulation framework, which will allow the fine-grained choice of fast simulation toos..

Their T1's and T2's have been performing nicely and they have made substantial improvements in their ability to distribute and monitor jobs between the T1's and T2's. They are also exploring wide area data access and caching.

ATLAS has updated their data placement and replication policy by reducing the disk space occupied by the ESD data and reduced the real DESD copies from 2 to 1. They have also reduced the number of reprocessing in 2012 from 1.5 to 1.

They have had a 350Hz prompt reco trigger and a 150Hz delayed reconstruction trigger so far in 2012 and will increase that to 400Hz and 200Hz for the 8 week pp extension. They are reporting an average pileup of ~20 in the 2012 data.

The ATLAS reporting of disk usage numbers is somewhat different than the other LHC big experiments. Rather than report usage numbers assuming the agreed-to 70% disk efficiency usage factor they report actual file usage (which they believe has an efficiency is close to 1) and the buffers needed for movement of data. However the needs are requested including the 70% efficiency factor. As this has led to some confusion we have had explicit discussions on this point with the experiment. As long as the usage numbers including buffer space and the requested numbers can be directly compared this is acceptable but will have to continue to be explicit in future reports.

CMS usage

Resource	Site(s)	Pledged	Used	Used/ Pledged	Average CPU efficiency
CPU (kHS06)	T0+CAF	121	75	62%	86%
	T1	137	138	101%	89%
	T2	320	368	115%	84%
Disk (PB)	T0+CAF	7	5.9	84%	
	T1	21	20.2	96%	
	T2	27	24	89%	
Tape (PB)	T0+CAF	23	18.9	82%	
	T1	47	36.4	77%	

CMS resource estimates have been generally accurate thus far in 2012. They ran a prompt reconstruction trigger of ~300Hz with a 25% overlap for a total rate of 375Hz.They have also been taking ~ 300hz of “parked” data to be processed during 2013 and abandoned the idea of an additional 300Hz due to lack of matching Tier 1 resources. Their observed pileup has been ~ 25 interactions per crossing. Raw event records have been smaller than predicted and reconstructed event sizes larger.

The T0 farm has had ~90% CPU utilization; an improvement over the 70% seen in 2011 due to a smaller memory footprint. The T1 centers have had an average pledge utilization slightly over 100% and have been doing a somewhat larger fraction of the simulation production. This freed up T2 resources for more analysis activity. Taking advantage of unpledged/ opportunistic resources the T2 centers have delivered ~ 115% of pledges averaged over the last 12 months with a 84% CPU efficiency. The T2 disk usage has benefited from a successful analysis move away from usage of the reco formats to the smaller AOD records.

Simulation production in the first half of 2012 has been twice the previous estimates.

LHCb usage

	Pledges			**Usage**		
2012	**CPU (kHS06)**	**Disk (PB)**	**Tape (PB)**	**CPU (kHS06) Efficiency (%)**[1]	**Disk (PB)** [2]	**Tape (PB)**
Tier 0	34	3.5	6.4	12.0 / 92	2.46	3.9
Tier 1	91	7.3	5.5	43.2 / 90	6.21	5.53
Tier 2	48			27.9 / 92		

[1] CPU efficiency based on (CPU time of successful jobs)/(Total Wall Clock Time)

[2] The pledge values includes tape cache. The usage numbers are the SLS values.

In comparison to the expectations for 2012 the following changes with relevant impacts for computing and storage has to be noticed:

1. slight increase in luminosity to $4x10^{32}$ $cm^{-2}s^{-1}$ (was 3.7)
2. extension of the run by 2 months (walltime – in effective beamtime is relatively higher)
3. due to 'deferred trigger' techniques the effective trigger rate reaches 5kHz
4. participate in the HI run (beginning 2013)

LHCb has taken several actions to fit these changed demands and the resources effectively provided by the supporting sites (in contrast to the pledges for that period).

The observed CPU usage throughout the period until (and including) August 2012 fits very well within the expectation (~50% used so far) and the changed 2012 running parameters/conditions listed in the previous section. The extended LHC pp period, the scheduled full reprocessing (2011 and 2012 data) and the ongoing MC activities, will consume the remaining 50% for this period. Throughout this period LHCb has changed the prompt reconstruction to 50% of the new data and extend the scheduled reprocessing periods by two months to stay within the CPU resource limits.

The Tie r2 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 by using them principally for MC production while only some selected Tier2s used in addition for data reprocessing

In spite of all the recent changes, leading to further adaption of the storage usage (within the existing computing model), LHCb shows a good resource usage for the disk storage. The situation is clearly not as comfortable as in previous years but, according to LHCb, sufficient. Further adjustments at the number of copies (i.e. DST/MDST from prompt 2012 reco) and stretching the timeline lead to matching resource numbers for this period.

Due to the special 2012 situation (discussed above) the situation for tape resources is clearly different and required short term adaptions to allow further storing of raw data until March 2013 as 'parked/locked' data (including HI data). The current estimates showing ~800TB tape is missing for RAW@T1 and RAW@T0 (each). To cover also the new data format FULL.DST (RAW+SDST together) the extra demand is ~3.5PB@T1 and ~1.2PB@T0 - not to forget the slight increase in tape resources for archive purposes of ~500TB@T1 and ~100TB@T0.

Looking at the 'active data management' tasks to match the effective disks resources, one aspect is new – the observed imbalance between pledged disk and CPU resources over time, leading to unused CPU resources just because the 'matching amount of data' is not local to that CPU resources. This will require further discussions between the experiment and those Tier 1.

The overall resource situation is less comfortable as the years before, requiring more and more frequent changes in the usage models. LHCb has, again, demonstrated very healthy and adaptable computing activities, leading to suitable resource usages and demands for this period. The obvious shortfall in tape resources (largely dominated by the special 2012 situation) should be addressed by all relevant boards and/or individuals to allow continuing storing of all possible physics data up to March 2013 and their deferred (detailed) analysis until LHC re-start.

PART B

Revised requests for 2013 and tentative estimates for 2014

ALICE

Resource	**Site(s)**	2013 ALICE	2013 CRSG	2014 ALICE	2014 CRSG
CPU/kHS06	T0+CAF	*126*	**126**	*135*	**135**
	T1	*160*	**120**	*160*	**130**
	T2	*145*	**145**	*211*	**200**
Disk/PB	T0+CAF	*11.0*	**11.0**	*11.0*	**11.0**
	T1	*10.8*	**10.8**	*10.8*	**10.8**
	T2	*15.8*	**15.8**	*15.8*	**15.8**
Tape/PB	T0+CAF	*22.8*	**22.8**	*26.1*	**26.1**
	T1	*21.0*	**21.0**	*23.9*	**23.9**

The table shows the latest 2013 and 2014 resources requests from ALICE together with the CRSG's recommendations.

ALICE has historically faced a lack of computing resources. Partly because of insistence by the CRSG, ALICE lowered its requests in order to better reflect the anticipated resources available in practice. However, in 2012 the reduced request led to an unanticipated corresponding reduction in the pledge at a T1 site. This situation may arise at further T1 sites for the 2013 requests and has led the collaboration, in particular, to raise the T1 CPU request in response. The CRSG deplores a situation where the collaboration derives a request from the computing model, reduces it to accommodate expected overall resource availability, but subsequently feels forced to modify it in anticipation of the way pledges are calculated. We hope this can be addressed satisfactorily in future and make a plea for the funding agencies to revise their contributions to ALICE resources so that 100% of the scrutinized needs can be met. For now we propose a change of the T1 CPU request which maintains the sum of T1 and T2 CPU power at the level ALICE would be requesting without the pledge calculations taken into account.

The jump in T2 CPU for 2014 reflects an accumulation of postponed analysis and MC simulation, especially from the p-Pb run at the end of 2012/13. We anticipate improvements in efficiency for CPU use and therefore have applied a cut of just over 5% to the T2 CPU request for 2014.

Disk requirements have been revised down, presumably helped by the reduction in the number of stored ESD copies and the switch to smaller AODs for analysis. The increases in tape capacity in 2014 arise partly from reconstruction delayed from 2013 (to lower CPU usage) and partly from MC simulations.

ALICE has previously achieved a reduction in the size of raw events by a factor of 3.5 by “clustering”. In future, clusters not assigned to relevant tracks may be discarded, giving a further large reduction in event size. We observe, in contrast, that the collaboration does not anticipate being able to use their online farm for other processing during the long shutdown and does not have a fast MC simulation. Both points are of concern to the CRSG. However ALICE has plans to use its DAQ farm for analysis during the second long LHC shutdown and is playing a founding role in a fast simulation project to be launched at CERN in 2013.

Improving the efficiency of CPU usage is an absolute must for ALICE. The collaboration is actively moving user analysis into trains, which run more efficiently, and is continuing a switch to analysis using smaller AOD formats instead of ESDs. However, the CPU efficiency is still affected by large I/O requirements and a rather heterogeneous user community and it lags the other LHC experiments. We strongly encourage efforts to improve this situation.

The CRSG encourages ALICE in its efforts to recruit new resources, especially T1 sites, which, in combination with improved efficiency, would mitigate under-pledging of CPU resources.

ATLAS

Resource	**Site(s)**	2013 ATLAS	2013 CRSG	2014 ATLAS	2014 CRSG
CPU/kHS06	T0+CAF	*111*	**111**	*111*	**111**
	T1	*319*	**319**	*373*	**355**
	T2	*355*	**350**	*408*	**350**
Disk/PB	T0+CAF	*11*	**11**	*11*	**11**
	T1	*35*	**33**	*36*	**33**
	T2	*53*	**49**	*56*	**49**
Tape/PB	T0+CAF	*27*	**23**	*31*	**23**
	T1	*43*	**40**	*53*	**44**

The table shows the latest 2013 and 2014 resources requests from ATLAS together with the CRSG's recommendations. ATLAS plans to reprocess all real data toward the end of 2012. In 2014 they are planning a full reprocessing including the simulation data, a large MC sample for the 13 TeV run and an equivalent amount of new simulation for the < 2013 data. They plan to use the HTL farm in 2013 and partly in 2014.

In 2013 and 2014 the event data model will evolve to accommodate modern computer architecture. They will start with the tracking algorithms. They intend to reduce the scatter of data over the memory address space to help with cache usage as the memory footprint is still too large for many-core machines. They are working on making the code multithreaded as well in an algorithm by algorithm approach.

Their CPU requests for 2013 and 2014 have used the existing processing time/event as a worst case proposal and hope that their code improvement pays off before the end of the long shutdown.

In spite of the vast amount of resources already used for MonteCarlo, ATLAS states to have more requests from the physics groups for simulations than they can satisfy. They are projecting to satisfy these requests by some combination of more usage of the fast simulation and finding more unpledged or opportunistic resources on the grid. While we commend ATLAS for their past success in finding resources we recommend they do not plan on it and that they make much heavier usage of the fast simulation and on MC code improvements. Progress in this direction will require devoting substantially more manpower than at present.

The CRSG thinks that the request in CPU at T1's in 2013 is justified based on the extended running period. The amount of CPU recommended in 2014 is based on a 20% increase with respect to our recommendation for 2013 in April 2012. In the CPU at Tier 2 centers we recommend no increase in 2014 over 2013. We note that ATLAS appears to have no problem in obtaining ample resources in Tier 2's.

Regarding disk, we can only recommend a small increase in 2014 with respect to the April 2012 approved request for 2013 . We assume an improved usage of disk by aggressively implementing monitoring techniques based on data access and popularity. This recommendation could be revised in spring 2013 based on new evidence. We recommend an early installation of the 2014 resources.

CMS

Resource	**Site(s)**	2013	2013	2014	2014
		CMS	CRSG	CMS	CRSG
CPU/kHS06	T0+CAF	*121*	**121**	*121*	**121**
	T1	*175*	**165**	*175*	**175**
	T2	*350*	**350**	*350*	**350**
Disk/PB	T0+CAF	*7*	**7**	*7*	**7**
	T1	*26*	**26**	*26*	**26**
	T2	*28*	**26**	*29*	**27**
Tape/PB	T0+CAF	*26*	**26**	*26*	**26**
	T1	*50*	**50**	*60*	**55**

The table shows the latest 2013 and 2014 resources requests from CMS together with the CRSG's recommendations.

The run extension has a strong impact on the needs for 2013 and 2014. Some of the 2013 resources may have to come via an early installation of 2014 resources as the 2013 pledges are more or less in place and not a lot more will be available by April 2013. The projected availability of the HLT farm for simulation and reconstruction by early 2013 should help with this.

For the rest of the 2012 running CMS has decided to increase the prompt trigger rate to 400Hz (including dataset overlap) and keep the parked trigger @ 300Hz.

CMS's archival storage needs at the T1's are projected to increase by 11% (5PB) and they projects an increased CPU need of ~ 20% to be located at the T1 centers to handle the additional 8 weeks of pp running. They also project the ability to use their HLT farm (with ~10% of the T1 capacity) by early 2013 for reconstruction and simulation production. They are also prototyping the usage of commercial cloud computing to handle peak CPU demands in the future.

They are asking for no additional CPU resources in 2014. This is partly due to anticipated code improvements. They have a software project in progress to allow efficient usage of multicore/multithreaded CPU's by the end of the long shut down.

The reduction in CPU at T1's with respect to CMS requests is justified by assuming a partial use of the HTL farm. The recommendation for disk at T2's is based on assuming an improved usage of disk using monitoring techniques based on data access and popularity. We recommend an early installation of the 2014 resources.

LHCb

Resource	**Site(s)**	**2012** **Pledged**	**2012** **Needed**	**2013** **Needed**	**2014** **Needed**
CPU (kHS06)	T0 (CERN)	34	34	**34**	**34**
	T1	91	110	**110**	**110**
	T2 + others	47	46	**46**	**46**
	HLT farm			**20**	**20**
Disk (PB)	T0	3.5	3.5	**4.4**	**5.5**
	T1	7.2	6.3	**8.6**	**10.4**
Tape (PB)	T0	6.4	6.2	**6.5**	**7.3**
	T1	5.3	10.0	**10.8**	**11.9**

The update of the LHCb computing resource usage estimates is based on the latest changes in the LHC schedule as well as recent modifications to the LHCb running parameters which imply over 40% extra more data in 2012: the rate was increased in 2012 by approximately 10%, up to 5 kHz; additionally CERN decided to extend the pp running period for 2 extra months in order to increase the total integrated luminosity of the experiments before the 2013/14 shutdown. Finally, LHCb has decided to take part in the pPb run scheduled for January/February.

In April 2012 the LHCb collaboration declared that the expansion of the LHCb physics program would have required additional CPU and storage resources while during the first part of 2012

LHCb have made all the efforts to fit within the pledge at the cost of delay to the physics results. LHCb did not request an increase in the CPU capacity in 2012 and they did a lot of efforts to stay within the pledges for storage resources. They ask to review the 2013-14 pledges to fit better within the new physics requirements. Today the perspective is definitely clearer:

- Data “stripping” process has implied the need to produce a new data format after a reconstruction pass, FULL.DST that includes in a single file the reconstructed SDST and the corresponding RAW data with a consequent increase of data on tape. In order to make space for the extra size of the new FULL.DST format, reconstructed data samples (SDST and FULL.DST) produced by the prompt reconstruction of 2012 data are currently being removed from the Tape systems at the sites. This implies that no further stripping of the 2012 data will be possible until the reprocessed data becomes available.
- The new changes in running conditions and scientific cases (as mentioned in the introduction) will require additional resources for LHCb. The expected increased size of the 2012 RAW and Reconstructed formats (tape) is as follows: 1) RAW data: 1.7 PB, 43% increase with respect to previous estimates (1.2 PB). 2) Reconstructed data: 3.1 PB, 120 % increase with respect the previous estimate (1.4 PB). 3) Heavy Ion RAW data: 100 TB. 4) The disk resident formats (DST and MDST) used for physics analysis are expected to increase by 43 %.

Prompt reconstruction will be allocated to CERN but only 50% of the collected data will be promptly reconstructed. In order to free a fraction of the Tier0 capacity for analysis about 20% of this activity will be sub-contracted to Tier2s (downloading the RAW data from CERN). Full reprocessing of data will be allocated to the Tier1 and to CERN. 20% of the reconstruction will be sub-contracted to Tier2s in order to free some CPU resources at the Tier1s for analysis. A new full reprocessing of 2011 data is planned starting in March 2013, once the full reprocessing of 2012 data is over. Another stripping pass of 2012 data is foreseen in spring 2013 to provide samples for new analyses not included in the reprocessing. The rest of the schedule for further processing of 2011 and 2012 is maintained. In the Autumn of 2013 it is planned a new stripping of the samples and during 2014 a full reprocessing of both data samples should provide the ultimate version of these data. Given the significant increase of the physics samples, a 40% increase for the simulation needs has been included in the model.

For 2013 and beyond the extra CPU work needed for the new MC simulation production will be accommodated by using the High Level Trigger (HLT) farm and by using the Tier0/Tier1 resources outside the periods in which they are not dedicated to real data processing.

The CRSG endorses LHCb requests for 2013-14 but is concerned by the load on networking that the tasks assigned to the Tier 2 represent.

PART C

Preliminary forecast for the period 2015-2017

So far computing has not been a limiting factor in the performance of the LHC experiments. The experimental collaborations have had enough resources to analyze vast amounts of data, provided by the excellent accelerator performance, in short periods of time. This has been possible in spite of a priori adverse conditions such a pile-up much larger than anticipated at this stage (due to the LHC running with a 50 ns interval between bunches, forcing the machine operators to pack many more protons per bunch at the interaction point) and by the experiments' own desires to take data at rates much larger than originally envisaged in the respective computing models (this has been possible thanks to the magnificent performance of the detectors and subdetectors and the maturity of the software and analysis tools, compounded with the natural desire to accumulate as much new physics data as possible).

For this scrutiny, the experiments have made available to us and the LHCC first estimates of the potential requirements for the period 2015-2017 when data taking resumes in full at a planned energy of 13 TeV, close to design. Both ATLAS and CMS think that they are capable of increasing their trigger rate to 1 kHz (the TDR value was 200Hz and 300 Hz, respectively), and LHCb has the potential to double its current rate to 10 kHz. Only ALICE would expect a modest increase of data volumes by about 20%. In total, this translates into a significant step up in CPU, disk and tape requirements that the CRSG has not quantified with precision at this point.

Up to now, it has been possible to accommodate the increasing computing demands within a roughly flat budget thanks to a decided effort of the experimental collaborations to continuously improve the efficiency and consequently mitigate the natural growth of the requirements and the hardware progress of technology. The speed of the reconstruction code has been improved by a factor of 8 since 2010, and memory reduced by about 40%. Although still lagging behind in efficient use, disk is now much better used; in order to manage Tier-2 disk space, popularity information is now used to make reasonable choices of what should be cleaned up and how many copies of relevant data should be kept and where should they be placed. The CRSG applauds the efforts that the experiments are making.

It clear that the LHC re-start after the long shutdown will imply larger computing resources. It is so far unknown whether the real jump will take place in the RRB year 2015 or will be postponed to 2016 (experiments seem to differ slightly in their estimates to this respect– this issue will have to be settled in the 2014 scrutiny), but it will come at some point. It is doubtful whether, even taking Moore’s law into account, the required rise in 2015 or 2016 will prove possible within a roughly flat budget. This growth may require significant annual investments, difficult to get given the ongoing financial problems. In addition the EU Framework 7 program is coming to an end without a definite alternative yet in place, implying a significant reduction in the number of supported personnel. The remaining resources will have to be prioritized for operations, leaving little or no room for further optimization efforts.

The experiments have to invest heavily in software improvements to be able to accommodate the needs within reasonable resources. Monte Carlo simulation is a major resource consumer. Experiments are urged to invest heavily in making large scale simulations sustainable by e.g. devoting the necessary manpower to optimize the processing and make ample use of fast simulations. The resources spent in simulation could be potentially reduced by a large factor by the end of the long shutdown if enough efforts are devoted to this task. Likewise, in reprocessing times there is surely still room for improvement too.

Many of these proposed code improvements are based on taking advantage of the latest machine technology. This will only pay off if this new hardware is made readily available at all of the tier centers.

A flat spending profile may lead, taking into account technological progress, to a ~20% increase in CPU and disk availability. Of particular importance is to adapt the existing software and analysis tools to the new generation of CPU hardware. This may bring in large gains that will not come for free in any case as it will require adapting all existing processes.

Disk volumes are harder to improve - event size and further reduction in the number of effective copies thanks to improved network access and data caching below file level can help in the mitigation of requests.

A substantial investment in software manpower will be needed to achieve all the previous objectives.

The CRSG is concerned by the existing uncertainties and even at the risk of going slightly beyond our mandate would like to make several statements.

The funding agencies should be prepared to continue their support to computing of the LHC experiments at least at the same level than is done at present or even contemplate the possibility of a moderate overall increase. A fraction of this increase could undoubtedly come via the incorporation of new Tier 1 and Tier 2 centers into the WLCG, i.e. new contributors. In any case we insist on the need to provide 100% of the scrutinized pledges and in that there is no real room for a budget reduction at this moment, lest computing become a limiting factor.

The LHC machine developers and experts may also play their part. We would like to strongly argue in favour of a 25ns bunch distance after the LHC re-start. Averaged over the different Tiers, the experiments have estimated that the present 50 ns bunch spacing represents up to ~40% more resources in CPU as well as disk compared to 25ns, because of the higher pileup. The implications on computing cost are thus very important.

Finally the advantages of taking data at rates much higher than contemplated in the TDR's should be balanced by the costs in manpower and computing that this represents. Probably a refinement of the triggers having the associated costs in mind could help in selecting the real physics objectives without implying unsustainable computing costs. Thus the experiments should carefully evaluate the physics need/return for the various trigger streams that sum to the high trigger rate. While the costs associated to detector upgrades are discussed in great detail, those associated with the computing resources are no less important.