Computing Resources Scrutiny Group

C.Bozzi (Italy), C.Diaconu (France), D.Espriu (Spain, *Chairman*), J.Flynn (UK), D.Groep (The Netherlands), J.Knobloch (CERN), A.Lazzarini (USA), H.Marten (Germany), W.Trischuk (Canada), B.Vinter (Nordic Grid), H.Renshall (CERN/IT, *Scientific Secretary*)

Introduction

This report summarizes the deliberations of the Computing Resources Scrutiny Group (CRSG) established by the WLCG Memorandum of Understanding regarding the computing requests by the four LHC experiments for 2009 and 2010. According to the new schedule of the LHC restart following the 19 September 2008 incident, 2009 and 2010 (from October to October) will formally constitute as a single run, with a short break over the end of 2009 period but for the purposes of WLCG procurement and availability it counts as two separate periods: one extending from October 2009 to the end of March of 2010. We shall refer to this period as "2009". The second period extends from April 1st 2010 to October 2010. We shall refer to this period as "2010". The resources for 2009 have to be in place by September 1st 2009 and the resources for 2010 by April 1st 2010.

The requests from the experiments for 2009 and 2010 (as defined above) were made public on April 7th. It is clear that to analyze the requests of four complex computing models in such a short time is a formidable task that would have been impossible without the experience gained from last year and the generous effort made by the referees. Yet, in view of the scarcity of the time available the present scrutiny should be considered as preliminary.

The purpose of the CRSG is to inform the decisions of the Computing Resources Review Board (C-RRB) for the LHC experiments. The starting point is the resource request information presented to the C-RRB by the different experiments and the guidance that the C-RRB cares to give. The CRSG then enters into a sustained dialogue with each experiment seeking to understand to what extent the computing resource requests are well motivated, the usage made of these resources and the accounting figures regarding usage and availability of the pledged resources.

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
- The CRSG shall make recommendations concerning apparent under-fundings.

In carrying out the scrutiny of the experiments requests the scope of this group is largely limited to the implementation of the respective computing models whose TDRs were reviewed by the LHCC some time ago. The evolution of the commissioning of the experiments as well as the implementation of the computing models in successive tests along with a better understanding of their needs have motivated a number of changes, sometimes representing limitations in the original model or assumptions. Obviously there is a gray area where the respective competences of the LHCC and the CRSG overlap and it is not so clear what would represent a change of the computing model or what would just be a natural adaptation of it to the changing circumstances.

In our 2008 report we brought to the attention of the LHCC several issues that we believed were relevant enough and they were dealt in a specific session devoted to reviewing the WLCG project and the four experiments computing status in February 2009.

LHC beam time

A new set of running conditions was approved in March 2009 following the Chamonix meeting. They should be adhered to by all four experiments and are as follows:

Year	Pp	AA
	Beam time (seconds/year)	Beam time (seconds/year)
2009	1.7×10^6	0
2010	4.3×10^6	5×10^5 (*)

The AA estimate (*) corresponds to, at most, one month of data taking with an assumed efficiency x availability of around 20% (our estimate). The above numbers correspond to an effective beam time that differs considerably from the ones used in the 2008 report. Throughout the 2008 report to the RRB an efficiency of 50% was assumed, although already there it was mentioned there that it was probably more realistic to assume at most a 40% efficiency. In fact the in the revised LHC schedule the efficiency factor is conservatively taken to begin at a modest 10% in 2009 and ramp up to 32% only at the end of 2010. This represents an important reduction in effective beam time and is the single most important ingredient in the revision of the experiments computing requirements as it will be discussed below case by case.

Usage of the resources

In 2008 our report was entirely dedicated to scrutinizing the resource requests by the experiments for the years 2008 and 2009 and no attention was paid to the usage the experiments made of the resources available to them. The WLCG project keeps an accounting of this at http://lcg.web.cern.ch/LCG/accounts.htm. In 2008 for Tier1's and CERN the CPU usage was roughly at the 40% level of the pledges, whereas disk and tape are at the 60% level approximately. In the first two months of 2009 CPU usage was approximately 60% of the pledged resources, disk at a 75% and massive storage at the 60% level. The figures show a clear improvement of the usage of the GRID resources, increasing numbers of users are submitting jobs as the middleware is becoming progressively more mature and the commissioning of the LHC for physics runs approaches. Yet, the level of usage of the resources by the experiments of the CERN and Tier1 resources made available to them does not seem totally optimal to us and this is a cause of concern to some extent.

The usage of Tier2 resources is higher in general, although extremely irregular in distribution.

The 2009 and 2010 scrutiny of the experiments request

It was not until April 7th 2009 that the four experiments produced a revised set of estimates for their computing needs for 2009 and 2010 (as defined above). The CRSG appointed referees to review the experiments requests; the same ones as in 2008 who were already familiar with the respective computing models and their evolution. The referees interacted as quickly as was feasible with the respective computing management of ALICE, ATLAS, CMS and LHCb. The

CRSG wishes to thank the four experiments for the collaboration offered to clarify the unclear issues in such a short time.

The CRSG held two plenary meetings; in the first one there was a general discussion to set up the framework of the reviewing process and a template was later circulated. In the second one, which held two weeks later, the drafts of the requests reports were reviewed and further questions posed. Revised drafts were circulated by email up to the point where convergence was reached.

Uniformity in the scrutiny for all experiments was a high priority but the CRSG acknowledges that it could not be implemented to the same degree as in the 2008 report due to the lack of time and the different speed of response of the respective computing managements.

Even if some experiments produced estimates for 2011 and beyond, we have not considered at all any projections beyond 2010. For a proper scrutiny we have yet to see real collisions and real data with the computing models going through a reality check and he CRSG prefers not to commit itself to any specific forecast beyond 2010. Experience gained once real data-taking is underway should reduce remaining uncertainties considerably allowing better estimates for 2011 and beyond.

The group also plans to look at the quality and effectiveness of the monitoring and accounting tools in the immediate future.

While in the 2008 reports to the Plenary and C-RRB we stated that the experiments requests and the scrutiny for 2008 and 2009 were coincident in general terms (differences typically ran at the 10% level although in some identified items discrepancies were larger) this is not the case in the present scrutiny. Experiments compare their requests for 2009 and 2010 to the current pledges, while it would probably be more meaningful to compare 2009 to 2008, and 2010 to 2009 in view of the delay in the LHC commissioning for physics and the reduced beam time. We will substantiate this statement below.

Note that the expected beam time in 2009 (1.7×10^6 s) is approximately half the beam time assumed in our scrutiny 2008 (3×10^6 s). Likewise for 2010 (4.3×10^6 s versus 9×10^6 s considered in the 2008 assessment for 2009). Of course this direct comparison may be far too simple as experiments have been collecting cosmics over the last months and will continue to do so and this data needs to be stored at least partially. There may also be updates of the computing models or modifications of their parameters that the experiments find advisable. Finally, it might be argued that more Monte Carlo simulated data may be required. Yet we take note that the GRID resources are far from being saturated at present, as we have seen.

However, it is clear that in order to substantiate requests at the level of the pre-existing 2009 and 2010 pledges even if the actual number of events to process will be roughly half of those considered for 2008 and 2009, respectively, one needs to make considerable modifications to the computing models. We fear that these some of these modifications represent changes that, if agreed, make the some of computing models unsustainable.

The CRSG has also been left with the impression that some of the modifications proposed do not have a real bottom-up justification, and/or the assumptions behind have not been clearly spelled out or are not completely transparent to us.

Thus, taking as starting point our baseline report submitted to the C-RRB in November 2008, we have made those modifications that we understand justified, rejecting others that appear to us unjustified, unrealistic or unsustainable. Taking into account the new efficiency and availability conditions, the convenience of some modifications for the first year or running, and the necessity of additional Monte Carlo simulated data in some cases, we conclude that the experimental needs for 2009 and 2010 should be roughly satisfied by granting the resources originally planned for 2008 and 2009, respectively. In fact probably a small reduction with respect to the resources approved for one year earlier (at the 10% level) may even be justified. In the case of massive storage a substantially larger reduction may be advisable as the original plan envisaged data taking already in 2007.

General recommendations

We list here several items affecting more than one experiment. A scrutiny and recommendations for each individual experiment shall be provided latter in this report.

- It seems prudent to scrutinise the experiments' use of resources after the first months of data taking in 2009. The CRSG commits itself to provide such a report in the shortest delay which is feasible, hopefully for the April 2010 RRB.
- It is important, given the long resource acquisition cycle, to inform the Tier1 and Tier2 computing centres of the resources required for calendar year 2011 as soon as possible. We would suggest the C-RRB to modify the timing of their meetings to facilitate the procurement process.
- The WLCG represents a computing effort of an unprecedented scale. In spite of increasingly demanding tests being passed uncertainties remain. We recommend that the different collaborations undertake a proper risk analysis and take stock of their results in future requests in order to cope with the most likely failures or shortfalls.
- In the case of ATLAS and CMS the information provided to us about their AA program remains all too sketchy.
- As running conditions may vary in the future (with the presence of 50ns or 75ns bunch crossings leading to pile-up) the collaborations should be aware that this has to be accommodated within the existing resources by decreasing the event rate or similar measures.
- The experiments are asked to actively pursue the policy of reducing the size of their raw events, and other derived formats, in future years as much as possible as detectors become better understood. No additional progress has been found along these lines in the present requests.
- We welcome efforts to remove `dark' or `orphaned' data and encourage the experiments to pursue this vigorously. Likewise we commend them to pursue a vigorous programme of purging data that is no longer required.
- We recommend the experiments make maximal use of the distributed resources in the GRID avoiding as much as possible the use of CERN facilities.
- In the case of CERN resources, we advocate for a very clear separation between the contributions used for calibration and first pass reconstruction and central analysis ('express stream' or similar), and those used to perform physics analysis by the CERN based physicists.

The CRSG wishes to state that the recommendations contained in this scrutiny are to the best of our knowledge rigorous. They correspond to the real needs of the experiments for a given LHC live time in the present stage of the commissioning and of their computing model implementation. There is no contingency for late delivery or failure to meet the pledges included in our estimates.

LHCC matters

Our scrutiny has identified some new issues that need to be brought to the attention of the LHCC. Below we provide a list of them along with some older ones that are still pending

- The LHCC recommend that ALICE undertakes a full assessment of how their physics reach might be affected by requested computing resources not materializing. We are waiting for this statement.
- The event size has a very direct impact on the computing requirements. Some experiments made an effort to reduce the raw event size (and the size of all subsequent derived formats) by establishing a reduction profile after startup. We believe that this effort should be followed by the experiments with the largest computing needs without unduly jeopardizing the physics.
- The potential proliferation of different data formats serving the same purposes should be watched closely as it seems a matter or concern. Strong overlap between different data sets as claimed by some now is a new matter of concern.
- More than one collaboration seems now to place heavier demands on CERN resources or suggest that a larger than normal part of their analysis should be done at CERN. This is hardly compatible with the worldwide LCG effort.
- Cosmic data taking is now much emphasized by experiments; while it is clear that cosmics are extremely useful in commissioning for calibration, this data is by nature transient and it seems somewhat questionable to us to support substantial requests based on cosmic runs.
- We recommend that the issue of the data analysis strategies is reviewed by the LHCC very soon, in order to ensure both, a reliable start-up and a coherent long term strategy. In particular it is recommended that the strategy for official reprocessing as a part of the computing model are better specified and then reviewed by the LHCC
- We recommend that a clearly defined sharing among the experiments of the resources installed at CERN is considered, in particular for the analysis facilities, such that an optimal usage of these resources is achieved.

Preliminary scrutiny of the ALICE request for 2009 and 2010

Overview

ALICE has presented its updated 2009 and 2010 requests to the LCG management board in March and April 2009. The experiment states that it has not changed its computing model (including data sizes and processing requirements) and that changes in resources requested stem from changes to the LHC operation schedule.

The ALICE computing model has a 100Hz event rate for pp data in the LHC steady state, leading to 10^9 pp events recorded annually. To maximise the physics impact of the initial 10 TeV data the collaboration plans to use a higher event rate in the 2009 and 2010 pp runs. In the 2009.04.09 requirement update ALICE says that they will record 1.5 x 10^9 events from 11.1 x 10^6 s of pp running and in the 2009.03.13 request the same number of events is expected from 6.1 x 10^6 s. For our scrutiny we have split this fixed number of events using the ratio of running times, 5/2, to give 0.43 x 10^9 events from 1.7 x 10^6 s in 2009 and 1.07 x 10^9 events from 4.3 x 10^6 s in 2010.

ALICE assumes a 1-month heavy-ion run in October 2010 leading a standard year's worth of events (10^8). We follow this for our scrutiny.

ALICE now has a larger disk buffer (of order 1.5PB) in front of the mass storage system at CERN. The claimed advantage is that data can be checked and possibly discarded before going to mass storage. There is also an automatic means to purge data that is no longer required. ALICE has not suffered from dark or orphaned data to date and expects that their bookkeeping systems should not let such data accumulate.

Real data (including cosmics)

Based on previous experience, ALICE expects to retain up to 400TB of cosmic events. Only one copy of the cosmics data is kept in permanent storage. This data will probably not continue to be processed once collisions start: if it is, it will be for calibration/alignment purposes.

Simulated data

ALICE aims to simulate pp data corresponding to a standard year's worth of events in 2009 and 2010. Simulated AA data is taken to be 50% and 75% of a standard year's worth of events in 2009 and 2010 (calendar years) respectively.

Comments and recommendations

In the summary tables below we have compared our estimates (CRSG) of the resource requirements with the ALICE requests and the. For CPU capacities, we have made two estimates: (a) based on a "minimal" processing scenario with scheduled and chaotic analysis of reconstructed collision and MC data, (b) assumes additional analysis is needed (factor of 3 in 2009 and 2 in 2010) in the early years while algorithms are being checked and debugged (this additional analysis is also a surrogate for additional reconstruction processing needed during early running). Accepting that some additional processing is needed during early running, a middle ground between estimates (a) and (b) is appropriate and this is adopted in the following

2009

		CERN	T1ext	T2ext	Total
CPU/kHS06	Request	42.4	42.8	36.0	**121.2**
	CRSG	2.6	33	59	**94.6**
Disk/PB	Request	2.7	4.3	4.4	**11.4**
	CRSG	2.6	3.1	6.4	**12.1**
MS/PB	Request	3.7	5.9		**9.6**
	CRSG	1.7	2.8		**4.5**

2010

Resource		CERN	T1ext	T2ext	Total
CPU/kHS06	Request	46.8	102.4	80.8	**230.0**
	CRSG	38	85	83	**206**
Disk/PB	Request	4.9	9.9	12.4	**27.2**
	CRSG	4.0	9.5	8.4	**21.9**
MS/PB	Request	6.7	11.6		**18.3**
	CRSG	6.0	9.1		**15.1**

.

The large difference between the CRSG figure for CERN CPU in 2009 and the requested amount merits comment. The CRSG value accounts for first reconstruction (at T0) and calibration/alignment tasks (CAF) only. In practice the ALICE computing model redistributes processing tasks to the CERN T0 when there is spare capacity and this has been done in for their 2009 and 2010 requests. Reconstructing AA data, as happens near the end of the 2010 period, puts a much greater minimal load on the T0 resources.

Our mass storage estimates include 0.4PB at CERN for saved cosmics data and the CERN disk estimate includes 1.5PB for the cache fronting the T0 storage.

We believe the assumptions of recording 1.5 x 10^9 pp events and a standard year's worth of HI events are optimistic and hence expect our estimates to be conservative.

We note ALICE's recent purge of data and encourage the experiment to maintain this good practice.

The following tables summarize the overall resources endorsed by the CRSG (the detailed distribution is give above). For comparison the estimates submitted to the C-RRB of November 2008 are given (the assumed effective beam time was almost twice then)

	Old report (October 2008)		**This report (April 2009)**	
	2008	**2009**	**2009**	**2010**
CPU (kHS06)	**98**	**172**	**94.6**	**206**
Disk (PB)	**14.4**	**22.1**	**12.1**	**21.9**
Tapes (PB)	**6.7**	**17.1**	**4.5**	**15.1**

Preliminary 2009 Scrutiny of the ATLAS experiment request

Overview

The information available to the ATLAS reviewers included the summary of ATLAS preliminary resource requirements presented to the WLCG Management Board at the end of March, 2009; Ian Bird's overview of updated resource requirements for all experiments which was presented to the WLCG Management Board on 7 April 2009; the summary of ATLAS resource requests for 2009-2010, presented to the WLCG Management Board on 7 April 2009 and a subsequent update shared with the reviewers on 20 April 2009, and a brief note, entitled "Factors influencing the current ATLAS Computing Resource Request," provided to the reviewers on 20 April 2009. In the time available, one of the ATLAS review team was able to meet once with the ATLAS computing liaisons and several other ATLAS principals. Prior to the meeting, ATLAS received a set of written questions designed to help elucidate a number of issues that had been identified at the 8 April 2009 CRSG meeting.

This year's scrutiny began with a top-level consideration of how the recently published run 2009-2010 LHC plans affected the scrutiny provisions and recommendations made last year. As a zero-th order estimate, it might be expected that the 2009/2010 resource request relative to last year's 2008/2009 request would scale as the LHC run live times which were assumed for the two requests. Last year, the LHC guidance was that 2008+2009 would yield an integrated run time of (3+9) x 10^6 s; the recently released plan now calls for an expected 2009+2010 run live time of only (1.7+4.3) x 10^6 s.

The ATLAS Computing Model

At the highest level, the model they use today is essentially unchanged from the one appearing in their 2005 TDR. ATLAS points out that many of the parameters driving the resource requirements predicted by their model have changed as the model has been confronted with (and calibrated by) the reality of data taking and data analysis on cosmics in 2008 and 2009Q1.

As noted in the 2008 scrutiny, their event sizes are larger now than they were in the original model assumptions. These have not changed since 2008, and there is no immediately obvious way to reduce them without compromising the experiment quality according to ATLAS.

Experience with calibration and reconstruction of cosmic ray data has shown that recall speed of raw data from tape limits ATLAS ability to keep up with the data. As a consequence, the experiment has begun to rely much more heavily on disk storage for their reprocessing.

Last year, they were aware that event and simulation processing times and event sizes were greater than originally projected; these were factored into their 2008 T0/CAF requests, but they did not have time to propagate the impact of this to their T1 and T2 sites. This has now been done for their 2009 requests.

Changes since the last scrutiny

ATLAS resource requests this year reflect a number of changes since last year. The LHC run is expected to encompass three different energies (0.9 TeV, 8 TeV, and 10 TeV), and additional bunch microstructures (now to include 50 ns on the way to 25 ns). These changes have allegedly caused the need for a proportionate increase in the number of different scenarios ATLAS now needs to simulate. Altogether, ATLAS estimates that the number of simulations is effectively 2.5X greater due to these factors relative to the real data volume, as compared to what it was expected to be last year. While last year they were projecting simulating only 25% of the data volume, they are now expecting to simulate as much as 75%, which is to be expected since the amount of MC events needed does not scale linearly with the amount of real data.

The change of LHC run mode to one with a very long run followed by a very long shut down has also caused ATLAS to re-plan its approach to data analysis. The long unbroken run now means that their earlier plan to take advantage of regularly spaced intervals of down time to reprocess and analyze is not longer possible. Instead, they will need to reprocess the already acquired data while at the same time continuing to acquire, calibrate, distribute, and analyze new data concurrently. They need to be able to provide immediate feedback to the detector operations or risk losing live seconds due to a mis-calibration, which drives the near real time processing requirements in the new scenario without breaks. This has increased their CPU and storage/throughput requirements by effectively a factor of ~ 1.6X-1.7X.

The long run may experience detector performance changes that must be tracked and continuously accommodated throughout the run. ATLAS now foresees the need to agilely perform new simulations on the fly as needed, in parallel with all the other activities.

Cosmic data taking in 2008, along with a few days of live beam, have allowed ATLAS to improve their understanding of the detector. As mentioned above, they now realize that they will need a bigger sample size of simulated data to adequately model the detector. In addition, calibration is now better understood and has become a more involved and refined process in order to fully exploit the detector’s capabilities.

Real data 2008/2009

ATLAS now has a little more than 1 PB of data from 2008. Most of this is from cosmic ray data taking, although they have a few days of live beam data taken just prior to the magnet incident last September.

ATLAS has made an effort to clean up orphan or dark data, because they realize this is an inefficiency factor for their resource utilization. Last year, they were experiencing as high as 50% dark data volume; they have been able to reduce it to less than 10%.

ATLAS has now developed a plan to recycle cosmic ray data. By 2010 they will have deleted 80% of the 2008 cosmic data volume (acquired at a rate of ~1 PB/yr). The other 20% they will archive indefinitely because it provides their only means to track secular changes in the detector over time, since they have no real beam data yet. In the longer term, they will implement a rolling cosmic ray data sample window, deleting the rest.

Simulated data

As discussed earlier, ATLAS has increased the ratio of simulated to real data from 25% - 30% to 75%. This increases the resource requirements by ~2.5X for that portion allocated to simulations. Still, they point out that the total amount of simulated data is smaller than what was foreseen in the (2005) C-TDR for 2010.

ATLAS has begun to use a simplified, much faster simulation tool (ATLFast) to provide the larger data volumes now needed.

Comments and recommendations

The ATLAS liaisons provided the CRSG team with a 20 April 2009 revision of their 3 April 2009 resource request for 2009+2010. The T1 resources have been revised downward by ~ 10% - 15%. All other resource requests remain the same as in the 3 April version.

They project simulations to account for 35% of the T1 CPU utilization in 2009/2010, reconstruction will account for 45%, and the remaining 20% will be allocated to production analyses organized by physics and performance groups. About 40% of the disk capacity is will be for production analyses and 35% will be dedicated to simulated events. At the T2s, about 50% of the capacity will be dedicated to simulations and the remainder to user and group investigations.

The AA running has a different model that is at the present time less well understood than the pp model. ATLAS described their AA model to be similar to the one developed by ALICE. They estimate ~ 10% of the resources will be used for AA analysis.

The projected needs across all categories for this year's 2009 are consistent with last year's request for 2009. However as noted above, the 2009 live time that is now expect is much less than that project last year: it is actually half of it taking 2009 and 2010 as a single run.

In part, presumably, processing of and storage of cosmics taken last year is the basis of this need, stored like real data in absence of beam. The other elements include the new approach to real-real time analysis to mitigate the risk of running with a mis-calibrated detector (previously addressed by reprocessing during the shut-down periods), as well as more MC requirements due to the multitude of run conditions. Lastly, the propagation of the event size effect for T1s and T2s (which ATLAS only did for the CERN-based resources in the 2008 requests), has now also been applied to the T1s and T2s.

In 2008, the CRSG reviewers were able to drill down in detail to confirm that the basic elements of the ATLAS 2008 request were understandable and that the request growth at that time reflected identifiable developments derived from improved experience-based performance data. This year, our understanding is more qualitative at this point. The changes noted above all tend to increase CPU requirements and disk storage. However, based on the information provided and given the uncertainties in the model and the input parameters used by ATLAS, the latest requests are not understood at the same level of detail and depth as was possible last year. In particular, exploring the sensitivity of the model to changes in parameters, needed for a comprehensive requirements analysis, was not done. A quantitative analysis for the requested increments (relative to the beam live seconds and the previous requests for 2008 and 2009 and the beam live time) is not feasible at this point. It will take further discussion with ATLAS to understand at a more quantitative level how the factors discussed above translate into the latest request.

Preliminary scrutiny of the CMS requests for 2009 and 2010

Overview

The CMS requests for 2009 and 2010 present a number of changes in the CMS computing model. The requests have been made assuming an LHC live time of 1.7×10^6 s until March 2010 and 4.3×10^6 s from April 2010 until the shutdown.

We examined the revised CMS computing requests for Tier0, Tier1 and Tier2 resources for the coming 2009 and 2010 combined run. Given the LHC expected live time in 2009 and 2010 and the fact they have very little experience that could be used to justify changes in their computing model with respect to the fall of 2008, we expected their 2009 requests to be similar to the ones that were already in place for 2008 and that the 2010 requests would look similar to their former requests for 2009. This was not the case. The CMS computing model has changed in ways we did not expect. We asked the computing management of CMS for additional explanations and we have investigated the motivations for such an increase with respect to older plans.

Comments concerning the proposed changes

First of all, their new reprocessing strategy involves one further pass per year (3+3 in 2009+2010 instead of 2+2) and, most importantly, they require a full reprocessing in one month instead of spreading it over several months. This is a significant change and we do not believe that CMS have any real experimental justification for it at this point. Neither were we able to assess how little re-reconstruction would be necessary in the 2009-2010 period as a result of the low-luminosity and low-duty cycle. Our general impression was that their original model expected 200 (going down to 150) Hz of data to be taken independent of luminosity. So we do not think a slow turn-on for the LHC has really been accounted for in their current requests. In conclusion, we do not recommend adopting this change in reprocessing strategy.

We also noted their increased request for MC production (1.5x raw data, which represents a 50% increase over the previous estimates). We feel that the Monte Carlo production during first 6 months of data taking will involve primarily the generation of small samples and their subsequent comparison with experimental data in order to tune the generators, detector response, etc. This will be an iterative process, and it is difficult to predict how many events will be generated. However, given that there will be different running conditions for the LHC in terms of energy and beam structure, we think that the CMS request is reasonable.

CMS now assumes a 40% overlap between their primary datasets (up from 10% last fall) which directly translates into increased requirements for disk space of the same order of magnitude. While motivated (presumably) by simulations, we think that this 40% overlap can be largely reduced by appropriate strategies, and we note that the overlap in current experiments in hadron colliders (e.g. the Tevatron) is of the order of 10%.

We have scrutinized the impact of the above changes on the resources requested at the various Tiers, by using the same (simplified) version of the CMS computing model which was used last fall. We performed several evaluations by allowing more and more changes to the computing model. In all cases, we find that the CRSG recommendations of fall 2008, shifted by one year, reasonably covers any scenario which has been examined, are robust enough against unforeseen circumstances and flexible enough to allow CMS to make reasonable changes in their computing strategies in case of need.

Generally speaking, the tape recommendations reflect the reduced live time of the LHC and scale down by roughly a factor 2. We also corrected for a previous estimate for the T0 CPU, which was not taking into account the LHC efficiency factor. We assumed 50%, following the CMS model.

Concerning cosmics, we were concerned that a significant amount of cosmic ray data has already been taken and re-processed. We wondered whether this had an impact on the GRID resources CMS is requesting for 2009 and beyond. We suggested that older MC sets and cosmic data – once analysed two or three times could be removed from Tier1 and Tier0 disks. The CMS reply indicated that they plan not to delete any cosmic ray data from tape. This decision seems unjustified to us.

Summary

We have investigated the motivations for such an increased request with respect to older plans. We have seen no compelling evidence that such an increase is really needed. Due to the reduced live time of the machine, the new 2009 (2010) requirements should be even smaller than the old 2008 (2009) requirements.

Nevertheless, due to the uncertainties which are still present in the models and in running the accelerator, we think that we should allow for some contingency. We think that the recommendations for CMS made by the CRSG to the C-RRB last fall, corresponding to an LHC live time of about 50% more than what is foreseen now, offers enough contingency to cover any aspects of the CMS computing model which might have been overlooked.

In conclusion, we reaffirm the CRSG recommendations of fall 2008, shifted by one year, for the data taking of 2009 and 2010, except for the above changes on tape and T0 CPU requirements.

The CRSG recommendations for the fall 2008 were

2008 Summary

Resource	Tier0	CAF	Tier1	Tier2	Total
CPU (kSI2k)	11000	2000	9090	11405	**33495**
Disk (TB)	285	1670	7010	7010	**15975**
Tape (TB)	3870		6120	-	**9990**

2009 Summary

Resource	Tier0	CAF	Tier1	Tier2	Total
CPU (kSI2k)	14700	1700	16430	17270	**50100**
Disk (TB)	285	2580	7200	5300	**15365**
Tape (TB)	8370		16500	-	**24870**

For the present scrutiny we make the following recommendations

2009 Summary

Resource	Tier0	CAF	Tier1	Tier2	Total	Fraction of 2008
CPU(kHS06)	22.0	8.0	36.4	45.6	**111.9**	**0.83**
Disk (TB)	286	1670	7010	7010	**15975**	**1.00**
Tape (TB)	2200		5000	-	**7200**	**0.72**

2010 Summary

Resource	Tier0	CAF	Tier1	Tier2	Total	Fraction of 2009
CPU(kHS06)	29.2	6.8	46.8	65.6	**148.8**	**0.85**
Disk (TB)	286	2580	7200	5300	**15365**	**1.00**
Tape (TB)	4300		12000	-	**16300**	**0.66**

Review of the LHCb experiment computing model and requests for 2009 and 2010

Introduction

In September 2008 the computing requirements of the LHCb experiment for 2008 and 2009 were analysed in detail. The current report provides estimates for requirements according to the new accelerator schedule.

The scrutiny performed by the referees include a full re-simulation of the model applied and validated for the previous scrutiny, an analysis of the model parameter changes proposed for 2009 and 2010 and an attempt to take into account the parameter changes in the model simulation. A number of remarks and recommendations are also given.

Besides the LHCb resources request, the scrutiny was based on three documents provided by the computing responsibles of the LHCb experiment: a detailed answer file to an initial number of questions, a second answer file including further details to some remaining points and an extra spreadsheet explaining the proportionality with the LHC running time. We used also a presentation in the WLCG MB of 31st of March. We also had a phone discussion with the LHCb computing coordinators. We would like to thank them for their collaborative attitude.

The model applied for 2008 scrutiny is repeated here for the scenario of operations in 2009 and 2010. The following two tables compare respective model assumptions in the old and the new report:

Year	pp operations **old**		pp operations **new**	
	Physics beam time (seconds/yr)	Luminosity (cm^{-2} s^{-1})	Physics beam time (seconds/yr)	Luminosity (cm^{-2} s^{-1})
2008	**$3 \bullet 10^{6}$**	2×10^{32}	**--**	--
2009	**$9 \bullet 10^{6}$**	2×10^{32}	**$1.7 \bullet 10^{6}$**	2×10^{32}
2010	$1 \bullet 10^{7}$	2×10^{32}	**$4.3 \bullet 10^{6}$**	2×10^{32}

Year	pp operations **old**		pp operations **new**	
	Expected DAQ events	Assumed MC events simmulated	Expected DAQ events	Assumed MC events simmulated
2008	**$6.0 \bullet 10^{9}$**	**$1.6 \bullet 10^{9}$**	--	--
2009	**$1.8 \bullet 10^{10}$**	**$4.0 \bullet 10^{9}$**	**$3.4 \bullet 10^{9}$**	**$1.6 \bullet 10^{9}$**
2010	$2.0 \bullet 10^{10}$	$4.0 \bullet 10^{9}$	**$8.6 \bullet 10^{9}$**	**$4.0 \bullet 10^{9}$**

The following table would provide the total requirements in 2009 and 2010 assuming no changes at all in the model

	Old report (October 2008)		**Now (April 2009)**	
	2008	**2009**	**2009**	**2010**
CPU (kSH06*yr)	**26,24**	**71,24**	**23,64**	**61,56**
Disk (TB)	**1162**	**3501**	**870**	**2385**
Tapes (TB)	**1150**	**4626**	**707**	**2545**

It should be stressed again that, except for the new beam time and hence deduced expected data acquisition rates (i.e., number of events per year), the model applied here was exactly the one used in the 2008 scrutiny, which resulted in a good agreement with the LHCb requests. Except for detector data that have already been taken in the absence of the accelerator beam (e.g. cosmics) and simulation results that have already been stored, the new running strategy for 2009 and 2010 should roughly correspond to a shift in the resource requirements estimated in 2008. However they disagree by large factors with the computing requirements announced by LHCb for 2009. This raises the question of the modifications of the computing model and its implementation.

Analysis of the Computing Model Evolution

The LHCb computing model used as an input for the ressources request in April 2009 underwent a number of modifications that are summarised in the following tables:

Data Format (kB/event)	Nov. 2008	2009	2010
RAW	35	50	40
rDST	20	35	25
DST	110	100	100
Monte Carlo DST	400	400	400

CPU/event (HS06)	Nov. 2008	2009	2010
Reconstruction	9.6	12	12
Stripping	0.8	0.8	0.8
Analysis	1.2	1.2	1.2

The contingency factors assumed for the RAW and DST data are large and are not substantiated at present. More details are needed in order to support this increase, although a safety factor must be taken into account.

Assuming 1.7 10^6 s DAQ time 3.4 10^9 events will be collected in the 2009 data taking period. Four reprocessings are envisaged using 2009 data and no stripping is foreseen. The data reduction from DST to mDST is now introduced into the model. This feature will not lead to a reduction of the storage since no stripping is foreseen. It is not clear why this large number of reprocessings is needed nor how this be possibly done within the running, since the online farm canot be used during the data taking. It seems more reasonable to limit the full scale reprocessings to two over this period, while most of the improvements could be done on dedicated samples

Assuming 4.3 10^6 s DAQ time 8.6 10^9 events will be collected in the 2009 data taking period, and three reprocessing *and* stripping passes are foreseen (one during the winter shutdown 2010/11).

The LHCb requirement in terms of compute power is dominated by Monte Carlo simulations at the Tier-2s. According to the presentation in the WLCG MB LHCb plans to produce 2.6 10^9 events in 2009 (0.6 b-events and 2.0 non-b-events) and 3 10^9 MC events in 2010 (1.0 b-events and 2 non-b-events). It is difficult to understand why, compared with the original strategy in September 2008, the number of MC events increased from 1.6 10^9 to 2.6 10^9 for the first phase of data taking, while the number of expected events decreases by about a factor of two.

The start of the analysis strategy is to increase the resources available centrally at CERN. This request is motivated by the need to increase the capacity of accessing the first data by more physicists and also justified by the insufficient reliability of the GRID computing in particular for the analysis jobs. Such a strategy may lead to an inefficient usage of resources and can endanger the distributed computing model.

LHCb added several kHEP-SPEC06 to the requirements for dedicated servers at the T0 and T1s. This is a change in the accounting philosophy since kHEP-SPEC were always only used to express the pure compute power for simulation, reconstruction, analysis etc. Although we seem to understand this as an attempt to express the full cost of LHCb computing as good as possible, the cost for additional services should not be taken into account in the models.

LHCb added requirements for a disk cache of 70 TB in front of tape systems per T0 and T1, which increases the total requirement by 490 TB. This is in fact a kind of temporary storage that is not available for effective storage of real physics or simulation data which might produce a significant cost that has not been accounted for in the past. We recommend including this into the resource estimates.

Given the changes of the computing model we have done a new requirements simulation with the following changes

- Increase size of RAW data format ⇨ applied; 50 kB/evt in 2009, 40 kB/evt in 2010
- Increase size of rDST data format ⇨ applied; 35 kB/evt in 2009, 25 kB/evt in 2010
- Change size of DST format ⇨ not applied
- Increase processing power per event for reconstruction ⇨ applied; 12 HS06/evt
- 4 reprocessing passes instead of two ⇨ not applied
- Increase number of MC events in 2009 from 1.6e9 to 2.6e9 ⇨ applied

- Decrease number of MC events in 2010 from 4e9 to 3e9 ➪ applied
- Increase local analysis effort at CERN at the expense of Grid usage ➪ not applied
- Add CPU power for experiment specific servers to kSH06 ➪ not applied
- Add 70 TB of disk space per T0 and T1 for T1D0 storage ➪ applied
- Assume 20 TB of disk space at T2s in 2009 and 2010 ➪ applied

CPU (kSH06*year)	**LHCb request**		**New RSG model**	
	2009	**2010**	**2009**	**2010**
Online farm	--	--	1,44	3,6
CERN T0 + T1	11,37	19,19	0,88	2,84
Tier1s	16,00	33,99	3,8	11,56
Tier2s	21,86	31,48	29,56	34,12
Total	**49,23**	**84,66**	**35,68**	**52,12**

Disk (TB)	**LHCb request**		**New RSG model**	
	2009	**2010**	**2009**	**2010**
CERN T0 + T1	780	1470	389	750
Tier1s	2800	4400	1329	2248
Tier2s	20	20	20	20
Total	**3600**	**5890**	**1738**	**3018**

Tape (TB)	**LHCb request**		**New RSG model**	
	2008	**2009**	**2009**	**2010**
CERN T0 + T1	1200	2300	451	1482
Tier1s	1300	2900	621	2082
Tier2s	--	--	--	--
Total	**2500**	**5200**	**1072**	**3564**

While the accepted modifications of the model lead to a sensible increase in the resources, the modelling attempted by the CRSG still undershoots the LHCb requests by 30% in the case of CPU and by a factor of 2 for storage. This is partially due to the different strategy of MC simulation (now also done at CERN and T1's for 2009), to the change in the analysis concept, (proceeds now also in T2's for specific analysis) and to changed reconstruction and stripping strategies. These issues could not be clarified completely in the conditions of this exercise

.

	Old report (October 2008)		**This report (April 2009)**	
	2008	**2009**	**2009**	**2010**
CPU (kSH06)	**26.24**	**71.24**	**35,68**	**52,12**
Disk (TB)	**1162**	**3501**	**1738**	**3018**
Tapes (TB)	**1150**	**4626**	**1072**	**3564**

Conclusions and Recommendations

The resources requirements of the LHCb experiment present significant increases compared to the previous estimations, justified by a more realistic accounting of the needs related to the analysis of the first data delivered by the LHC. While the gross features of the new model can be linked to a reasonable strategy to plan for the unknown features of the data taking, the quantitative analysis has not been completed in the present scrutiny. A number of features have been however identified:

- The changes of the computing model ingredients are largely dictated by concerns related to the LHC startup. Safety factors are applied which lead sometimes to significant increases. The scrutiny group could not evaluate the fundamentals of such requirements nor could analyse the assumptions considered in this modifications.
- The CPU resources do not directly scale with the LHC running time since the resources are dominated by the Monte Carlo simulation which proceeds largely independently of the data taking. The amount of scaling for CPU is 50% at CERN and 25% in Tier 1. The situation is different for disk and tapes where the amount of proportionality is 75% (90%) for CERN (T1) resources.
- The increased request for analysis power at CERN is related to the need of quick turn-around in the first analysis of the data. This seems not to be imposed by the lack of computing resources elsewhere. We could not get any evidence that this increase in the central resources is really devoted to the central tasks that should be executed at CERN. Indeed, the new computing power seem to be be open to normal analysis jobs, that would be better served in an centralised environment given the observed imperfect GRID computing reliability. LHCb presented however a clear plan to reduce in the next years the computing power at CERN to 25%, as expected from the TDR.
- The reprocessing strategy is at present unclear and several “restarts” of the processing (undefined number) are assumed. The number of reconstruction full passes seems to be covering also the dedicated calibration and tuning analyses, which do not need usually a full reprocessing of the data. The 2009 data in particular is not aimed at physics results, and it is therefore difficult to assume that a period of intense development for reconstruction tuning can be accommodated while full reconstruction campaigns are planned. The resources allocated to this activity amount to about 10 to 15% at CERN and the resources necessary to full two passes over the data sets can be considered as more realistic for the given time scales. A general reprocessing at the end of the 2010 data taking can also be justified, using this time the online farm.
- We recommend not including the CPU power for dedicated servers (VO-Boxes, etc.) at the T0 and T1s in future requirements tables as this imposes a change in the general accounting philosophy of the whole collaboration.
- We recommend considering an average disk cache size per site in front T1D0 storage in future requirements tables as this might be a significant cost factor (depending on the data handling strategies) at least at LHC start-up when remaining background storage is still comparatively small.