CERN-C-RRB-2005-006
CERN-LHCC-2005-006
LHCC-G-089
9 March 2005

LARGE HADRON COLLIDER COMMITTEE

Review of Computing Resources for the LHC Experiments

January 2005

Executive Summary

The computing resources required by the LHC experiments will be of a much larger scale than the LEP experiments. The sheer size of the collaborations establishes a need for global analysis facilities, which present a challenge well beyond the current experimental HEP collaborations. The total volume of data recorded by each experiment will be larger than current experiments and the anticipated CPU power required to process and analyze the data will be much larger and more distributed than currently required at the Tevatron at Fermilab or the B factories. In fact, the majority of the computing power for analysis will not be at CERN, but will be distributed around the world.

The committee evaluated the computing resource requirements of the four LHC experiments at a meeting in January. The computing model documents submitted in December 2004 were used for the core of the review.

The committee was very impressed with the overall quality of the material presented in the computing model reports and the presentations at the review sessions.

The committee examined the computing resource numbers presented in the computing model reports and reviewed the methodology used in determining the CPU, disk, tape and network capacity estimates. The committee also compared the estimates to those presented in the Hoffmann Review Report. It was noted that the computing models, in some cases, had evolved significantly from the time of the Hoffmann Review.

The most striking change, that of increasing disk requirements, almost certainly makes the models more credible than in 2000; particularly when considering the support of a large number of simultaneous users/processes. The increases of 500-800% have not been fully justified which reflects the overall uncertainty in the analysis models.

The experience gained from the recent data challenges has provided a foundation for testing the validity of the computing models. These tests, however, are incomplete. The upcoming data challenges and service challenges are essential to test key features such as data analysis and network reliability.

The data analysis models in all four experiments are essentially untested. The risk is that distributed user analysis will not be achievable on a large scale. There is a risk that disk and CPU resource requirements will increase if key GRID functionality is not used. There is also a risk that additional manpower will be required for development, operations and support.

The committee was concerned about the dependence on precise scheduling required by some of the computing models. Calibration schemes and use of conditions data have not yet been tested. These are expected to have an impact of only about 10% in the resource requirements but may have a more significant impact on timing and scheduling.

Based on the examination of the Computing Models for the four experiments, the committee made several global recommendations. These recommendations are addressed on Page 10. The recommendations and comments for the individual experiments are given in separate sections at the end of the report.

The committee expresses their congratulations to the collaborations for producing the detailed computing models that will form the foundation for the physics analysis at the LHC. While many of the concepts have been validated, much work remains to put in place a truly distributed analysis structure based on GRID tools.

Aside from issues of peak capacity, the committee finds that the computing models presented are robust enough to handle the demands of LHC production computing during early running (through 2010.) There remains a concern about the validity of the data analysis components of the models.

Introduction

Starting in 2007 the LHC is expected to produce proton-proton collisions at a center of mass energy of 14,000 GeV with an initial luminosity of approximately $2 \times 10^{32} cm^{-2}s^{-1}$. The luminosity is expected to rise steadily until it reaches the design luminosity of $10^{34} cm^{-2}s^{-1}$ by 2010. The collaborations push to process the data quickly so that first physics results are available as quickly as possible. It is expected that the trigger rates will not vary much with luminosity, which implies that the expected amount of data output rate will remain constant.

During the commissioning period of the accelerator complex and the 4 large detectors, the large computing systems required for data processing will be acquired and deployed. The purchase and deployment of the storage elements and CPU will be staged to meet their needs as the overall volume of data steadily increases and the event complexity increases. This staging will allow the experiments and computer centers to take advantage of technology advances and anticipated price decreases.

Once data taking commences, it is expected that LHC will operate as proton-proton collider for seven months during each year followed by approximately a month of Heavy Ion(HI) running. The recording rate for the events expected during proton-proton running is of the order of 100 MB/s. Since Heavy Ion collisions could possibly produce up to 8000 charged particles per unit of rapidity in each event, this number is expected to increase to 1.25GB/s for Heavy Ion collisions in ALICE. The event data sizes and raw data rates anticipated for the four experiments are given in Table 1 for proton-proton collisions and in Table 2 for Heavy Ion collisions.

P-P	SIMU	Sim ESD	RAW	Trigger Rate	RAW Rate	RECO	AOD	TAG
	MB	MB	MB	Hz	MB/s	kB	kB	kB
ALICE	0.4	0.040	1.0	100	100	200	50	10
ATLAS	2	0.5	1.6	200	320	500	100	1
CMS	2	0.4	1.5	150	225	250	50	10
LHCb	--	0.4	0.025	2000	50	75	25	1

Table 1: The anticipated event size and raw data rates for the four experiments during proton-proton collisions at the LHC are presented in the Table. SIMU: Simulated RAW data, RAW: data from the experiment to be recorded in mass storage, RECO (or ESD): output from the reconstruction code, AOD: reduced data format for analysis, TAG: summary for event section.

Heavy Ion	SIMU	Sim ESD	RAW	Trigger Rate	RAW Rate	RECO	AOD	TAG
	MB	MB	MB	Hz	MB/s	kB	kB	kB
ALICE	300	2.1	12.5	100	1250	2500	250	10
ATLAS			5	50	250			
CMS			7	50	350	1000	200	10

Table 2: The anticipated event size and raw data rates for the three experiments during Heavy Ion collisions at the LHC are presented in the Table. LHCb will not participate in Heavy Ion running. Some information on event sizes for Heavy Ion running was not available at the time of the review. SIMU: Simulated RAW data, RAW: data from the experiment to be recorded in mass storage, RECO (or ESD): output from the reconstruction code, AOD: reduced data format for analysis, TAG: summary for event section.

The Computing Resource Review

In 2001 the Steering Group of the LHC Computing Review released a report[1] containing an evaluation of the plans and prospects for data management and computing at the LHC. As a part of this process three independent panels reviewed the software and computing plans for each of the four experiments. In the final report (CERN/LHCC/2001-004) the panel made a number of recommendations including the following: 1) "After critical assessment, the review accepts the scale of the resource requirements as submitted by the four experiments…", and 2) "A multi-Tier hierarchical model similar to that developed by the MONARC[2] project should be the key element of the LHC computing model."

During the past year, in preparation for the Computing MoUs and TDRs, each of the four LHC experiments has produced updated estimates for their anticipated computing capacity requirements in terms of disks, tapes, CPUs and networks for the Tier-0, Tier-1 and Tier-2 centers. These numbers were documented in Computing Model documents that were submitted to the LHCC in mid-December 2004. These new estimates varied from those submitted to the LHC Computing Review in 2001. This change was found to be significant enough to warrant an independent review of the new resource estimates. A committee made up of several LHCC referees and three external referees was formed to review the new computing models and the revised resource estimates.

The Committee was charged with examining the current computing resource requirements, and to check their validity in the light of the presently understood characteristics of the LHC experimental program. The charge was neither to review the underlying computing architecture, nor the overall cost. Costing of the CERN elements will be done by CERN and the LCG. Detailed costing of the distributed regional centers will be done by the Tier-1 and Tier-2 centers and together with their funding agencies.

Each of the four experiments submitted documents outlining their computing models in December 2004. (ALICE[3], ATLAS[4], CMS[5], LHCb[6]) The committee received these materials and then listened to a series of presentations from each of the four experiments. The committee met at CERN in mid-January to listen to presentations from each experiment during a two and a half day review session. At the end of the review, feedback was given to the representatives of the experiments in the form of a closeout session attended by the computing coordinators and the spokespeople. In addition, to some general comments and recommendations, specific comments and recommendations for each of the four experiments were presented. These comments and recommendations are included in this report.

The committee was very impressed with the overall quality of the material presented in the computing model reports and the presentations at the review sessions.

The Computing Resource Estimates

The committee examined the computing resource numbers presented in the computing model reports and reviewed the methodology used in determining the CPU, disk, tape and network capacity estimates. The committee also compared the estimates to those presented in the Hoffmann Review Report. It was noted that the computing models, in some cases, had evolved significantly from the time of the Hoffmann Review.

In general, the reviewers concentrated on examining the validity of the assumptions used in estimating the capacity of the essential elements of computing model, including such inputs as the expected number of reconstruction passes and the general strategy for simulation, analysis and data storage. The reviewers were able to examine in detail the resources required for reconstruction and reprocessing of the physics data and simulated data. The physics analysis requirements for the collaborations were not as well known at the time of the review and could not be reviewed in as much detail.

In the preparation of their computing models and requirements for early running the experiments made some common assumptions about the early running conditions and plans for the LHC commissioning and operations period from 2007-2010. The assumptions are in agreement with recent machine planning.[7] These assumptions are shown in Table 3.

Year	pp operations			Heavy Ion operations		
	Beam time		Luminosity ($cm^{-2}s^{-1}$) @ATLAS/CMS	Beam time		Luminosity ($cm^{-2}s^{-1}$)
	(sec/year)	days		(sec/year)	(days)	
2007	5×10^6	100	5×10^{32}	--	--	--
2008	10^7	200	2×10^{33}	10^6	20	5×10^{26}
2009	10^7	200	2×10^{33}	10^6	20	5×10^{26}
2010	10^7	200	10^{34}	10^6	20	5×10^{26}

Table 3 A summary of the common assumptions for the LHC running schedule used for the Computing Model planning for the period from 2007 to 2010. Proton-proton running for LHC-b and ALICE will be at a reduced luminosity throughout the period.

The experiments also made common assumptions about the efficiencies for utilization of the CPU, disk and mass storage resources. These assumptions are listed in Table 4.

Efficiency for scheduled CPU	85%
Efficiency for "chaotic" CPU usage	60-75%
Disk utilization efficiency	70%
Mass Storage utilization efficiency	100%

Table 4: The experiments used common assumptions for the efficiency of utilization of the various computing resources in their current computing models.

Each experiment made an assessment of the computing required for reconstruction, simulation and analysis. The reconstruction and simulation packages are now available and have been tested during the data challenges of the past year, so the estimates were mostly based on experience gained from tests with simulated Monte Carlo data, though in some cases large improvements in performance were assumed. The estimates are given in Table 5. For CMS and ATLAS, it is expected that the reconstruction time will be about a factor of ten larger for HI events. Calibration and alignment strategies have not been well tested yet, but are expected to have about a 10% impact on the computing resource requirements.

	Units	ALICE		ATLAS	CMS	LHCb
		p-p	Pb-Pb	p-p	p-p	
Time to reconstruct 1 event	kSI2k sec	5.4	675	15	25	2.4
Time to simulate 1 event	kSI2k sec	35	15000	100	45	50

Table 5: The estimates used in the computing resource estimates for the time to reconstruct and simulate 1 event for p-p collisions (in kSI2k sec) . The estimates for Pb-Pb events in ALICE are also given. The reconstruction time for Heavy Ion events in ATLAS and CMS are expected to be about a factor of ten larger than the estimates for p-p events.

The computing resource requirements for a typical early year of running, in this case 2008, are summarized in Table 6. Each experiment expects reprocess their data sample 2-3 times per year. The amount of simulated data varies among the experiments with ATLAS planning to generate and fully process only about 20% of their event data sample. CMS plans to generate a simulated data sample equal to their physics data set.

In the Table, the Tier-0 resources are sometimes separated from the total resources required at CERN, since several experiments requested additional Tier-1 and Tier-2 type functionality at CERN. ALICE gave peak and average capacity for their CPU requirements. We have included the average capacity for their Tier-0 center in Table 6. Note that the computing model for ALICE has been updated since the time of the review and can be found in the updated version of the Computing Model Report[3] dated 04 February 2005.

		2008	HR	2008	HR	2008	HR	2009	HR
		ATLAS	ATLAS	CMS	CMS	LHCb	LHCb	ALICE	ALICE
Tier 0 CPU	*MSI2k*	4.1	3.915	4.6				4.5	
CPU at CERN	*MSI2k*	6.3	6.21	7.5	7.38	0.9	2.025		7.416
Tier 0 disk	*PB*	0.35		0.41				0.5	
disk CERN	*PB*	1.95	0.41	1.71	0.796	0.8	0.33		0.53
Tier 0 tape	*PB*	4.2		3.8				2.7	
tape CERN	*PB*	4.6	9	5.6	4.172	1.4	1.22		3.23
Tier 1 cpu	*MSI2k*	18	11.286	12.8	9.18	4.4	6.3	10.6	8.424
Tier 1 disk	*PB*	12.3	2.16	6.7	1.565	2.4	0.75	6.3	1.08
Tier 1 tape	*PB*	6.5	10.8	11.1	5.115	2.1	1.6	8.7	1.48
Tier 2 cpu	*MSI2k*	16.2		19.9	9.675	7.6		10.9	
Tier 2 disk	*PB*	6.9		5.3	1.75	0.02		1.7	
Tier 2 tape	*PB*	0	0	0	1.25	0		0	
total CPU	*MSI2k*	40.5	17.496	40.2	26.235	12.9	8.325	26	15.84
total disk	*PB*	21.2	2.57	13.7	4.11	3.22	1.08	8.5	1.61
total tape	*PB*	11.1	19.8	16.6	10.537	3.5	2.82	11.4	4.71
CPU Increase	*now/HR*	2.3		1.5		1.5		1.6	
Disk Increase	*now/HR*	8.2		3.3		3.0		5.3	
Tape Increase	*now/HR*	0.6		1.6		1.2		2.4	
WAN IN Tier0	*Gb/s*	4.5		7.2					
WAN OUT Tier0	*Gb/s*	2.25	1.5	9.5	1.5	0.53	0.31		1.5
WAN IN per Tier1	*Gb/s*			5.7					
WAN OUT per Tier1	*Gb/s*			3.5					

Table 6: A summary of the computing requirements (CPU, disk, mass storage and network bandwidth) for the four experiments. The projected requirements are given for the year 2008, which is expected to be the first full year of data taking, except for ALICE where the numbers are for a standard year of running (2009→). Also included are summaries of the computing requirement for each experiment from the Hoffmann Review (HR) – in grey. The HR numbers for CMS have been corrected for efficiencies. The current CMS estimates have been adjusted by moving the estimates of resources required for one Tier-1 and one Tier-2 center to the total for CMS resources at CERN.

In comparison of the current figures with those from the Hoffmann Review of 2000 the committee observes the following trends:

1. Overall there has been a significant increase in the required computing resources. This increase is approximately a factor of two but is not uniform across the experiments, Tier sites, or components (tape/disk/CPU).

2. The most significant increase for all experiments is in the amount of disk storage required. This is true across all Tier levels (except ALICE Tier-0). The total disk requirements have increased by ~300% for CMS and LHCb, ~500% for ALICE and ~850% for ATLAS.

3. The total CPU requirement has increased by 50% for CMS, ALICE and LHCb, and by 130% for ATLAS.

4. The total tape requirement has increased for CMS (60%), ALICE (140%) and LHCb(20%), but has fallen by (40%) for ATLAS.

5. The increase in resources is much smaller at the Tier-0 than at the rest of the centers. For the Tier-0, the required CPU has remained the roughly the same and the tape requirements have grown by up to 30% (CMS and LHCb) or even fallen (ALICE and ATLAS). In three cases, however (CMS, ATLAS and LHCb) significantly more disk is required at the Tier-0 (210%, 240%, 480%). The greater figure for ATLAS, in this case, probably accounts in part for the decrease in tape requirements.

6. In all cases, the biggest increase in CPU requirement is in the Tier-2 centers. This in part reflects the immaturity of Tier-2 planning at the time of the Hoffmann Review. The Tier-1 resource requirements have also all increased significantly. The largest increases are again in the required disk storage, where already the already large storage capacity has increased by 300 – 600%.

It is difficult to draw detailed conclusions; the changes are almost certainly the result of a combination of factors. It appears plausible that the increased in resources come from assuming a constrained financial envelope that is now being spent at a later date. An increased number of Tier-1 and Tier-2's compounds the increase.

The most striking change, that of increasing disk requirements, almost certainly makes the models more credible than in 2000; particularly when considering the support of a large number of simultaneous users/processes. The increases of 500-800% have not been fully justified which reflects the overall uncertainty in the analysis models.

Global and General Comments

At the time of the Hoffman review, it was anticipated that the computing resources would be nearly divided equally among the three Tiers with CERN provided about 1/3 of the total computing services. In the update resource estimates, there has been a significant change in the balance. As noted above, the requirements for Tier-0 CPU capacity have not significantly increased. Overall, there has been a large increase in the amount of disk storage required. There has also been an increase in overall CPU power with respect to the Hoffmann Review. This increase is primarily at Tier-1's and Tier-2's. In addition, the overall number of Tier-1 and Tier-2 centers has increased.

Many tests aimed at determining the validity of the computing models have been done since the time of the Hoffmann review. Each experiment launched a series of data challenges to test their software and computing infrastructures. These data challenges are ongoing while the collaborations continually incorporate new software features and work to improve the reliability of the production service. Last year, the LCG project deployed a production version of grid middleware, although with reduced functionality with respect to early planning. The experiments have tested distributed Monte Carlo production over the past year using the LCG-2 and other grid and non-grid tools.

Large parts of the production and reconstruction models have been tested and validated during the past year. Physicists inside the collaborations can now use the simulated Monte Carlo data generated and reconstructed at a large number of sites all over the world for their physics studies. Data analysis of these large data sets at distributed sites is only just beginning.

The experience gained from the recent data challenges has provided a foundation for testing the validity of the computing models. These tests, however, are incomplete. The upcoming data challenges and service challenges are essential to test key features such as data analysis and network reliability.

The LHC user community is large and will be distributed on several continents. In addition, data analysis is inherently "chaotic" and difficult to schedule in advance. Among the most difficult challenges of the models is the deployment of grid-based distributed data analysis. The grid tools needed to achieve distributed user analysis are not yet available. Therefore, the distributed analysis models have not yet been validated. The physicist user community inside the collaborations will ultimately perform the best tests of the analysis chain. The users will have a chance to exercise the available analysis software during the upcoming studies of physics potential using large data samples produced during recent Monte Carlo data challenges. These important tests are now planned to begin within the next year.

The data analysis models in all four experiments are essentially untested. The risk is that distributed user analysis will not be achievable on a large scale.

The computing models for all four experiments rely on the GRID. The functionality has been tested in part through the recent 2004 production data challenges. However, the reliance on the complete functionality of GRID tools varies from one experiment to another. It is not at all clear that the full functionality required for distributed analysis will be available and tested in time for the first physics data at the LHC. This lack of functionality could have an impact on the resources required for analysis. The experiments may need to increase the number of copies of data and also the amount of local processing power at the regional centers. The decreased automation in job submission may also result in an increase of manpower to direct the computing resources effectively.

There is a risk that disk and CPU resource requirements will increase if key GRID functionality is not used. There is also a risk that additional manpower will be required for development, operations and support.

The quasi-real time requirement for the first pass of reconstruction sets constraints on the schedule for processing at the Tier-0 center. During proton-proton running, the experiments plan to operate their first pass reconstruction while the RAW data remains on disks at the Tier-0 centers so that the data will not have to be pulled back from the mass storage systems. There are also constraints set by the availability of the necessary calibration and alignment information required for accurate reconstruction. The availability of production code and calibration constants will limit the production schedule, particularly in the early commissioning phase. Unforeseen delays in the production schedule will be difficult to recover, particularly if tied to specific calendar dates such as the shutdown period.

The committee was concerned about the dependence on precise scheduling required by some of the computing models.

The calibration schemes for the experiments have not been worked out in detail. There have been some preliminary tests as part of the test beam programs, but in general, the deployment of calibration and alignment software and the use of conditions data is behind schedule. The experiments plan to implement calibration software and conditions databases in the upcoming year. The required computing resources are not expected to be large – about 10%, however, the work is manpower intensive, particularly in the commissioning. Therefore, calibration and alignment can have a big impact on the processing schedule.

Calibration schemes and use of conditions data have not yet been tested. These are expected to have an impact of only about 10% in the resource requirements but may have a more significant impact on timing and scheduling.

In determining the required capacity, the experiments used a variety of "safety factors" to mitigate schedule risks and to guarantee adequate storage capacity at the Tier-0 center. These factors include addition of CPU cycles to keep up with data processing. While some schedule contingency is certainly justified, these contingency factors were not always well motivated. In addition, the contingency factors on processing times and RAW data size vary significantly among the experiments.

The committee did not review the manpower demands imposed by the new computing models, nor was there an examination of the cost of equipment or operations. This work will be done in the context of the LCG and through the MOU process. For any future cost comparisons, it would be helpful if the experiments could agree on a methodology for the equipment costs. It was noted that a new PASTA report is planned for this spring to help with these cost estimates.

The ATLAS request for a Central Analysis Facility that will be used for debugging and calibration and will serve as an analysis facility for CERN based users received strong support from the committee. It was felt that all experiments will find a need for local analysis at CERN and Tier-1/Tier2 capability at CERN should be included in their computing plans.

Since this facility will inherently have a special status due to its access to the large storage facilities at CERN, careful planning for the use of this facility will be required so that this facility does not become the default user analysis facility. The distributed analysis model is essential to obtain the required computing capacity that guarantees access to the data for the global LHC user community and will greatly enhance the physics potential of the experiments.

Global Recommendations:

Based on the examination of the Computing Models for the four experiments, the committee makes the following global recommendations:

- The committee recommends that the average and the peak computing requirements of the 4 experiments be studied in more detail. A month-by-month analysis of the CPU, disk, tape access and network needs for all 4 experiments is required. A clear statement on computing resources required to support Heavy Ion running in CMS and ATLAS is also required. Would it be possible to smooth or reduce the demand that seems to peak at the Tier-0 centers during the shutdown period?

- Detailed plans for distributed analysis during the initial period should be worked out.

- Dependence of the computing model on raw event size, reconstruction time, etc. should be addressed for each experiment. This should include an analysis of safety factors and risks.

- Details of the ramp up (2006-2008) should be determined and a plan for the evolution of required resources should be worked out.

- A complete accounting of the offline computing resources required at CERN is needed from (2006-2010). In addition to the production demands, the resource planning for calibration, monitoring, analysis and code testing and development should be included - even though the resources may seem small.

- The committee supports the requests for Tier-1/Tier-2 functionality at CERN. This planning should be refined for the 4 experiments.

The committee recommendations should be addressed by the collaborations in preparation for their Computing TDRs. It is presumed that the Computing Model documents will form a strong foundation for these documents. The Computing TDRs will provide the most detailed estimates of the computing resources requirements, which should again be expressed in terms of the required capacity.

Conclusions

The committee expresses their congratulations to the collaborations for producing the detailed computing models that will form the foundation for the physics analysis at the LHC. While many of the concepts have been validated, much work remains to put in place a truly distributed analysis structure based on GRID tools.

Aside from issues of peak capacity, the committee finds that the computing models presented are robust enough to handle the demands of LHC production computing during early running (through 2010.) There remains a concern about the validity of the data analysis components of the models.

Comments and Recommendations for ATLAS

The committee was impressed by the level of thought and planning that have gone into the formulation of the ATLAS computing model, which is in general fairly specific and detailed. It is based on a hierarchy of Tier sites, where the Tier-0 at CERN has the responsibility of first reconstruction pass as well as serving as the custodian of one copy of the raw data; ten distributed Tier-1 centers share a second copy of the raw data as well as the responsibility for reprocessing and for organized analysis-related activities. The 25 Tier-2 centers provide the computing power for end-user analysis and simulation production.

A rapid but accurate calibration on 10% of the data will be performed at the Tier-0 center prior to the prompt reconstruction, which is required to keep up with real time data taking within the 5 day limit of the input disk buffer.

Significant thought is currently being devoted to the process of obtaining calibration constants as well as to the issues related to populating and distributing the conditions database. This activity is very much welcomed by the committee. The committee notes that the ability to properly calibrate the detector before the prompt reconstruction pass is a crucial component of the computing model that needs to be fully validated during the upcoming Data Challenge 3.

The committee finds that key elements of the computing model such as the event size and the reconstruction processing time have a relatively weak basis of evidence, with large reduction factors with respect to current benchmarks being assumed in the model. In addition, the variations of processing time and event size with backgrounds and luminosity have not been investigated, leading to large uncertainties in the overall estimates of tape, disk and CPU.

The committee recognizes the merits of the distributed analysis model that should make optimal use of the available resources, but notes the lack of practical experience with the model. There is also a need to provide resources to develop the managerial software based on Grid middleware that will handle the distributed environment. The lack of a clear understanding of the analysis patterns also makes the motivation for having multiple copies of ESD and AOD formats weaker, as well as leaving a large uncertainty on the amount of CPU required for analysis. The impact on the overall cost is significant since quite appropriately analysis represents a large fraction of the required resources.

The usage of the TAG database to create event collections for analysis is a key feature of the computing model that is still largely without design and remains untested. The committee was unclear about how this mechanism will work and considered that, since the creation and use of event collections will be the primary interface to physicists, it will require a much attention in the upcoming data challenges and could possibly several iterations before it has the needed functionality and ease of use.

The model requires that events can be processed in quasi-real time as they get transferred from the online event filter farm, thus at a rate of 200Hz. This assumption introduces roughly a factor of two in the CPU power for prompt reconstruction, which the committee notes is not fully justified and could be somewhat relaxed by requiring the ability to keep up with data only on average.

The Heavy Ion physics program should be fully included in the planning, considering that the usage of computing resources during the shutdown period must be optimized at a global Tier-0 level.

The committee notes that the foreseen amount of fully simulated events (20%) is rather small and presents a risk for physics results. A larger number would be certainly safer if it can be accommodated in the available resources.

Comments and Recommendations for ALICE:

The ALICE experiment gave a detailed presentation of the computing model including the specific assumptions on the running scenario in the years 2007 – 2009. This model is very briefly summarized. A standard year of data taking will consist of 7 months of pp collisions (10^7s) at a rate of 100 Hz, (100MB/s), 1 month of heavy-ion collisions (10^6s) at a rate of 100 Hz, (1250MB/s) followed by a 4 month shutdown period.

Data from pp running will be processed immediately at the Tier-0. The RAW and summary ESD are archived at the Tier-0 as well as exported to the Tier-1's. Data from AA running are processed only during the shutdown. Archiving and exporting happens as for the pp data. Simulation (MC) data is generated at Tier-1 and Tier-2 centers.

The ramp up of resources is foreseen as follows:

- 20% of total installed in 2007,
- 40% of total installed in 2008,
- 100% of total installed in 2009.

No further upgrades are planned; and the collaboration only requests resources for replacements and repairs.

The committee raised a number of issues with the computing models as presented at the time of the review and made the corresponding recommendations:

The process for obtaining calibration*s* and the data quality monitoring process was not described in the computing model document. It appeared that these issues had not yet been thoroughly examined by the collaboration. The committee recommends that the process of obtaining calibrations and performing Data Quality Monitoring should be defined in more detail.

There was no request for a facility with Tier-1/Tier2-level capabilities at CERN. The committee felt that these facilities would be needed, particularly during early commissioning. The committee urges the collaboration to plan for a Tier-1/2 – level facility at CERN to facilitate commissioning and calibration efforts.

The resource planning extends through 2009 with no provisions for system upgrades thereafter. This did not seem realistic. Sufficient resources for early 2007 data taking and commissioning should be planned.

The processing strategy generates large peaks in demand for CPU, network and tape. The scheduled peaks in CPU capacity required at the Tier-0 were of particular concern to the committee. In the case of the Tier-0, only average and not peak numbers were requested. The peak demand should be clarified since it is unlikely that extra resources will become available during the shutdown when other experiments also may have peak demands.

The specific split of resources among the Tier-0, Tier-1 and Tier-2 centers appears ad hoc. It is not clear that the values chosen are indeed optimal for the experiment. The specific choice of split up of resources between Tier-0/1/2 should be better motivated and possibly optimized to the needs of the experiment.

The *peak demands* may be smoothed if a different time profile for processing steps is chosen. In particular, the choice to process the HI data during the 4 months shutdown introduces a high risk. The current planning for processing and reprocessing should be validated and possibly reiterated. It should be investigated whether the pronounced peaks in requirements for cpu, tapes and networks cannot be smoothed out. There is not currently a plan to use the HLT farm during down times for reprocessing. It was not clear if this added capacity could be used to ease the peak demand for CPU or for another type of data processing during the shutdown period.

Finally, it was noted that the experiment relies on a very high but not very well defined additional functionality of grid middleware. How the resource needs change if the required functionality should not be fully available was not clear and should be examined in the context of the Computing TDR.

The ALICE collaboration faces a large computing challenge due to the large event data size during HI running. Their recent update of the computing model[3] addresses many of the issues raised in this review.

Comments and Recommendations for CMS

As is the case for the other experiments, it is inevitable that current estimates of the resources needed for CMS computing are subject to a significant uncertainty. To illustrate this observation we quote a few examples of estimates taken as input to the model, each of which are subject to an uncertainly of order a factor of two:

- The raw event size currently produced by the CMS full Monte Carlo simulation is 0.3 MB. This has been increased to 1.5 MB as an estimate of the event size expected at turn on; this estimate is based on reasonable assumptions and the experience from previous experiments. For example, a factor 2.5 increase has been assumed to account for conservative setting of thresholds/zero suppression at startup.
- A safety factor of two in the Tier-0 resources for event reconstruction has been included. CMS will have a disk buffer to allow temporary storage of approximately five days' raw data prior to reconstruction. It is likely that during scheduled running periods around 50% of the time will be spent in physics stores, with the remainder being taken up with machine studies, preparing for stores, accelerator equipment failure, etc. Nevertheless, the resources estimate assumes that Tier-0 should be able reconstruct events at 150 Hz: the currently assumed rate to tape for CMS.
- CMS has assumed that it needs to produce fully simulated Monte Carlo events at the same rate at which it will be collecting real data from running at the LHC. This seems to us to be a reasonable target, although if resources were tight a somewhat smaller number of fully simulated MC events would probably also be perfectly adequate.
- For the reconstruction of Heavy Ion data CMS assumes that it will achieve a factor of ten improvement in speed with respect to the current performance. This probably achievable, but will require significant effort.

CMS does not yet have a fully developed calibration and alignment scheme. The committee notes that the ability to calibrate the detector before the prompt reconstruction pass is a critical component of the computing model that needs to be fully validated as soon as possible.

We find clear justification for a sizable CERN-based facility in addition to the Tier-0 center. This will be especially important for detector-related (time critical) activities such as monitoring, calibration, and alignment.

The CMS plan calls for each year's p-p dataset to be reprocessed on the Tier-0 facility during the 4-month shutdown. However, it is not clear to us that this will be affordable, given the heavy demands for other activities, such as the processing of the heavy ion data of ALICE and possibly the other experiments.

The CMS plan for analysis computing in the early years of LHC running attempts to minimize reliance on some of the currently least mature aspects of the Grid, such as, global data catalogues, resource brokers, and tools for distributed analysis. This is achieved by, for example:

- Streaming output dataset by reconstructed physics objects,
- Placing specific data streams at specific Tier-1 sites,
- RECO+RAW (FEVT full event) is taken for planning purposes as the basic data format for analysis in the first few years of LHC running.

The committee considers this approach to be conservative, but in our view not unreasonably so, given the current uncertainties in the delivery of the relevant Grid middleware, etc. This approach does, however, lead to some potential concerns:

- It may be more difficult to balance the load across all Tier-1 sites.
- Some political problems may result from the need to decide which Tier-1 sites host the most frequently accessed (most interesting?) streams.

The plan to focus "analysis" activities at Tier-1 sites on organized production activities appeared to be very sensible (e.g., AOD production, dataset skimming, calibration/alignment jobs). This seems to make optimum use of the high quality production computing environments that will be available at the Tier-1 sites. The committee notes that there may be one or two Tier-1 sites that may prove to be exceptions to this general practice.

We note that although a specific baseline model and estimate of resources have been presented, a lot of thought has gone into considering possible alternatives. Although "real life" in the early years of LHC running will probably differ in significantly from the assumptions currently made, we have confidence that the CMS computing model has some flexibility to respond to the actual demands that will be placed upon it.

In conclusion, the committee finds that within the necessarily large uncertainties inherent in attempting such planning two-three years before the first data, the input assumptions made by CMS and the subsequent calculation of needed resources seem to be reasonable.

Comments and recommendations for LHCb

During this review, the LHCb group presented a computing model and resource assessment based on an output rate of 2kHz of raw data from the high level trigger (HLT). This rate is currently thought to have 4 components: 200 kHz exclusive decay modes, 900 Hz for inclusive $B \rightarrow \mu X$, 300 Hz for events containing D*, and 600 Hz for high-mass $\mu^+\mu^-$ candidates. This trigger plan represents a significant change from the group's previous scheme, which had a total output rate of 200 Hz. It is notable that computing resources required for their new plan are similar (within 50% except for disk) to those in the Hoffman report, even though event rate is increased by an order of magnitude; the savings in the new scheme result from less Monte Carlo production, which is the dominant use of CPU, and a much smaller raw event size.

In their computing model, a full copy of the raw data is stored at CERN and also distributed among the ~6 Tier-1 centers. The initial reconstruction is performed at the CERN Tier-0 and the Tier-1 centers during the 7 months of data taking. A second reconstruction pass is planned to be done during 2 months of the shutdown, using CERN, Tier-1, and the LHCb online farm. The output of the event reconstruction is a reduced DST (rDST), hoped to be 50 kB including raw data, which is to be used as the basis for selecting particular events for further analysis (stripping). They plan for 4 stripping passes through the rDSTs per year, mostly using CERN and Tier-1 facilities, but with one pass also using the online farm. The stripping reads in both the rDST and raw data, and writes out at least four streams of selected events, corresponding to the 4 components of the trigger identified above. The output DSTs are stored on disk at CERN and Tier-1 centers, and are also archived in mass storage. In the LHCb model, user analysis is performed at the Tier-1 centers. The Tier-2 centers are used almost exclusively for MC production.

The Table 7 below summarizes LHCb's stated requirements for CPU, DISK, and MSS for 2006-2010 of LHC running. Note that these requirements are averages for the year.

LHCb	2006	2007	2008	2009	2010
CPU(MSI2k year)	3.89	7.78	12.97	14.45	17.87
MSS (TB)	1030	2060	3433	7144	11632
Disk (TB)	984	1969	3281	4015	4749

Table 7 LHCb requirements for CPU, DISK and MSS for LHC running in the years 2006-2010. These requirements are averages for the year.

The committee was impressed by the level of planning that has gone into the LHCb computing model, and by the clarity and detail of the presentations. In general, the committee believes that LHCb presented a well reasoned plan with appropriate resources for their proposed computing model.

We have several comments on specific issues.

<u>Time variation of resource requirements.</u> In the LHCb computing plan, the peak cpu and network needs exceed the average by a factor of 2. This variation must be considered together with expected resource use patterns of other experiments. LHCb (and the other experiments) should consider scenarios to smooth out peaks in resource requirements.

Monte Carlo. Monte Carlo production consumes more than 50% of the projected CPU resources. Any improvement in performance of the MC or reduction in MC requirements would therefore have a significant impact on CPU needs. The group's current MC estimates, while difficult to justify in detail, seem reasonable for planning.

Event size. The committee was concerned about the LHCb computing model's reliance on the small expected event size (25 kB). The main concern is I/O during reconstruction and stripping. LHCb believes that a factor of 2 larger event size would still be manageable.

rDST size. The rDST size has almost as large an impact on computing resources as the raw event size. The committee recommends that LHCb develop an implementation of the rDST as soon as possible to understand whether the goal of 50kB (including raw) can be achieved.

Event reconstruction and stripping strategy. The multi-year plan of event reconstruction and stripping seems reasonable, although 4 strippings per year may be ambitious. If more than 4 streams are written, there may be additional storage requirements.

User analysis strategy. The committee was concerned about the use of Tier-1 centers as the primary user analysis facility. Are Tier-1 centers prepared to provide this level of individual user support? Will LHCb's planned analysis activities interfere with Tier-1 production activities?

Calibration. Although it is not likely to have a large impact on computing resources, we recommend that details of the calibration plan be worked out as soon as possible.

Data challenges. Future data challenges should include detector calibration and user analysis to validate those parts of the computing model.

Safety factors. We note that LHCb has included no explicit safety factors (other than prescribed efficiency factors) in their projections for computing resources. This issue should be addressed in a uniform way among the experiments.

Appendix A

Committee Membership

Chair: P. McBride
CERN Chief Scientific Officer: J. Engelen
Representatives from the LHCC: F. Forti, T. Wyatt
External: E. Blucher (Univ. Chicago), N. Geddes (RAL), T. Haas (DESY)
LHCC Chairman and Secretary: S. Bertolucci, E. Tsesmelis
PH Department: J.-J. Blaising

In addition, the following people were in attendance:
Computing MoU Task Force Chair: D. Jacobs
IT Department: J. Knobloch
LCG Project Leader: L. Robertson

The Charge to the Committee

During the past year, in preparation for the Computing MoUs and TDRs, the LHC experiments produced updates to the estimated computing capacity requirements in terms of disks, tapes, CPUs and networks for the Tier-0, Tier-1 and Tier-2 centers. The numbers were seen to vary from those submitted to the LHC Computing Review in 2001. Such a movement, though not surprising, was significant enough to warrant an independent review of the new estimates.

Starting from the present requirements documents, the task of this Review is thus to examine critically, in close discussion with the computing managements of the experiments, the current estimates and report on their validity in the light of the presently understood characteristics of the LHC experimental program. The exercise will, therefore, not be a review of the underlying computing architecture.

---

[1] Report of the Steering Group of the LHC Computing Review CERN/LHCC/2001-004
http://lhc-computing-review-public.web.cern.ch/lhc-computing-review-public/Public/Report_final.PDF
[2] MONARC Phase 2 Report CERN/LCB 2000-001, March 2000
http://monarc.web.cern.ch/MONARC/docs/phase2report/Phase2Report.pdf

[3] ALICE Computing Model CERN-LHCC-2004-038/G-086, draft: 05-Jan-05, updated 04-Feb-05
[4] The ATLAS Computing Model CERN-LHCC-2004-037/G-085
[5] The CMS Computing Model CERN-LHCC-2004-035/G-083
[6] LHCb Computing Model CERN-LHCC-2004-036/G-084 (CERN-LHCb-2004-119)
[7] R. Bailey presentation at the Chamonix Workshop XIV:
http://indico.cern.ch/contributionDisplay.py?contribId=39&confId=044