CERN-RRB-2003-131
CERN/LHCC-2003-047
LHCC-G-054
24 September 2003

LARGE HADRON COLLIDER COMMITTEE

Review of Computing Resources for the LHC Experiments
The CMS Experiment

September 2003

On 2-5 September 2003 the LHCC held a review of the software manpower requirements of the four LHC experiments. The Review Committee concentrated on the details of their planned software activities and work packages and examined how software development has been organized in each experiment. The leaders of the software effort in each of the experiments made presentations to the Committee that were primarily focused on the effort required for core software activities, simulations, high-level trigger infrastructure and for the experiments' interaction with the LCG projects. They also described the level of effort required for the development of sub-detector application software and algorithms.

The Review covered the level of effort required to develop, validate and process Monte Carlo simulations (physics processes as well as the detector description and response) and for processing the LHC data so that it can be made available for analysis. For the Monte Carlo simulations, the collaborations presented their plans, including those for the Data Challenges as well as information on validating and improving the simulations, for example by benchmarking against test beam data.

The Committee was charged with the task of examining the required and available manpower for the core software and sub-detector software inside the experiments and explore their relation to the LCG Project in order to evaluate whether the experiments are putting the proper effort into these computing tasks. The LCG projects were not reviewed per se but participated in the review in order to provide a link to the experiments in terms of the overall required manpower and to provide details of the deliverables to the experiments. Moreover, the review was not a technical evaluation of the computing but remained focused on issues related to manpower resources, the scope of the work and the progress evaluation mechanisms.

Membership of the Review Committee:

Chairman: R. Cashmore - CERN Director for Collider Beam Programmes

Representatives from the LHCC: F. Ferroni, J.-J. Gomez-Cadenas,
Y. Karyotakis, P. McBride (rapporteur), P. Seyboth

External: J.-J. Blaising, N. Geddes, R. Gerhards, S. Wolbers

CERN Director for Technology Transfer and Scientific Computing: H. Hoffmann

LHCC Chairman and Secretary: M. Calvetti, E. Tsesmelis

EP Division: I. Videau

IT Division: F. Hemmer

1. INTRODUCTION

Much has been written about the unprecedented challenge of computing for the LHC experiments. The scale of the software and computing projects for the LHC experiments is far beyond that seen previously, to the extent that the LHC endeavors have attracted much publicity outside the field, in areas facing data challenges on similar scales, albeit on longer timescales. The excitement (hype) created by this interest and the recognition of the problems exemplified by the extensive funding available to "GRID computing" should not be allowed to obscure the underlying problems, which remain very real. The volume of data expected is enormous on the scales seen at LEP and the size and complexity of the events is much higher than at the Tevatron. A year of LHC data equates to over a thousand years of LEP data. Careful planning and testing of all aspects of the data processing for the LHC experiments is required to meet this challenging task, including the basic software, the computing infrastructure, the data handling and bookkeeping, data calibration and Quality Assurance (QA). The large data challenges foreseen by the experiments that are scheduled to take place over the next few years are essential for the experiments to test their software and their computing models as they prepare for the flood of both data and users. Proper preparation of the software and computing is essential for timely physics results from the LHC after machine turn-on in 2007.

2. GENERAL COMMENTS

The Review Committee congratulates the experiments, the LCG Project and the software and computing coordinators on the large amount of work they had completed in preparation for the review. Their careful planning and thoughtful presentations were extremely helpful for the committee and should aid the experiments in their pursuit of well-tested and robust software and computing for their experiments before the arrival of the first LHC data in 2007.

Physicists and physicist-programmers have traditionally developed the software required for reconstruction, simulation and calibration of each sub-detector. A large effort is required to develop efficient and robust algorithms in these areas. The software managers of the four experiments indicated that although the task is large and there is at present a shortage of manpower working in these areas, the shortfall can be addressed by appropriate prioritisation of the work of the large number of physicists in the collaboration. This is not to understate the real nature of the deficits. These tasks presented must be, and will be, performed at some point.

There is, however, another part of the required software, the core software for which the development is critical at this moment and the shortfall, although not large, already puts the experiments at risk of not having their software infrastructure ready for the beginning of data taking in 2007. The core software includes the framework, the event data model, the software management system, the data management infrastructure and other important components that form the backbone of the reconstruction, simulation, calibration and high-level trigger software. The core software team is also responsible for testing, QA, verification of code, documentation and user support of the production releases. Software engineers, computer professionals and highly skilled physicist-programmers are required to do most of the development work in these areas. **The committee estimates the missing level of effort in the core software is about 20-25 FTEs adding the missing manpower of the four experiments together.**

Table 1. Required number of FTEs for software activities for 2003 for each experiment. The tasks included in the lines 1-4 require experienced software engineers, computer professional and skilled physicist-programmers. The numbers required for Production/Grid Integration for ATLAS include not only tool development and GRID integration as for the other experiments but also physicists doing production validation during their Data Challenge (DC).

Year 2003	ATLAS		CMS		LHCb		ALICE	
(Units = # FTE)	*Needed*	*Actual*	*Needed*	*Actual*	*Needed*	*Actual*	*Needed*	*Actual*
1 Coordination/ Management	**3.9**	**3.9**	**7.5**	**4.3**	**1.8**	**1.8**	**2.0**	**1.7**
2 Core software and infrastructure	**27**	**15.4**	**20.8**	**13.5**	**16.1**	**9.8**	**17.9**	**19.1**
3 Grid integration, DC production	**26.3**	**18.3**	**20.5**	**15.9**	**6.6**	**4.9**	**10.2**	**10.5**
4 LCG contributions	**4.8**	**4.8**	**6.9**	**6.9**	**2.6**	**2.6**	**4.9**	**4.9**
5 Sub-detector Software	**70.8**	**49.3**	**70.0**	**47.0**	**32.4**	**26.2**	**48.3**	**39.7**

The experiments presented the effort required in terms of their current needs, the actual number of FTEs available, and the projected needs for the next few years. In the four days of the review, the committee examined the numbers that appeared to be significantly different among the four experiments. ALICE had taken a somewhat different view of their manpower planning and planned the scope of their projects to fit the manpower available. CMS separated the management of various subtasks from the tasks themselves and also counted the coordination effort required for producing a TDR in the category of Coordination/Management. ATLAS did a bottoms up estimate of the effort required based on the estimates of the current subtask managers. Their projections indicated that they require a somewhat larger number of core software developers than the other experiments. In the end, the committee found a relatively consistent picture of the number of core software developers required by the four experiments.

In addition to the core software, there are specialized software tools needed for enabling distributed production, reconstruction and analysis of large amounts of Monte Carlo generated simulated data for the upcoming Data Challenges (DC). The teams producing these tools are often funded by outside sources including the GRID projects. The collaborations have taken advantage of these opportunities in this critical area and have used the available manpower to make progress. In the end all experiments will depend on the deployment of functional and tested GRID components. Over the past year, each experiment has been engaged in evaluating middleware components provided by European projects (EDG, NorduGrid) as well as US projects (for example iVDGL, GriPhyN, PPDG), and at this moment they all look forward to the first deployment of LCG grid components.

The LCG projects are charged with developing tools and infrastructure software for the four LHC experiments. It is important to note that it is not a fifth experiment, but rather an integrated part of the whole program and as such an essential component of each experiment. The Committee finds that the linkage between the experiments and the LCG projects is generally good, but that the commonality of purpose should be reinforced. The process that has determined the needs of the experiments and launched the projects has been effective in identifying the highest priority projects. We encourage the LCG Project leadership and the experiments to find ways of facilitating the integration of LCG components into the experiments' software. A flexible mechanism that can periodically optimise the combined LCG and experiment work plan is needed to assure that the experiments receive the tools they have requested.

The Committee finds that there is a clear shortage of manpower for core software development inside the experiments that should be addressed. Examining the larger picture, it seems that this is an opportune time to review the priorities and the manpower inside LCG and the experiments. This evaluation may not be straight forward given the constraints of experiment construction/computing schedules and the external commitments of LCG. However, the Committee believes that relatively minor changes could produce significant improvements in the overall project. In addition to this global rebalancing, **the management of experiments must monitor their own core software development projects and coordinate their planning in order to avoid unnecessary duplication of effort.**

The committee was concerned that the LHC collaborations appeared to have such difficulty in addressing apparent shortfalls in high priority areas of computing and software. One possible tool to help manage this problem would be to include software and computing shortfalls in the experiment common funds, allowing funding agencies to directly reassign LHC funding to address this shortfall.

The software teams in the experiments contain a large number of people on temporary contracts (1 to 3 years). These fractions range from approximately 35% for CMS to almost 70% in ALICE. A fraction larger than about 50% in the core software team inevitably puts the experiment in a fragile state, since too much valuable expertise may be lost. It is imperative that the **people currently in these positions be replaced once their contracts expire**, as the manpower requirements in this area will not fall in the next 5 years or so. It is **highly advisable that some be given longer-term contracts in order to maintain the expertise and guarantee the continuity** and stability of the software effort.

The experiments that have established a strong core group at CERN are more advanced in their software development and have been able to establish a good working interaction with the LCG projects. **Given the existence of LCG, and the fact that there are 4 experiments, the experiments <u>must</u> have a core software group working at CERN to manage the project effectively and to interact efficiently with the LCG development team.** These core teams cannot function effectively if they are built from a large number of people working only a small fraction of the time on development tasks.

Given the size of the task and the distributed nature of the collaborations, the experiments also need to take advantage of expertise elsewhere. In the absence of a formal MoU process, Software Agreements with collaborating institutes are a way of recognizing the work contributed in the software effort and help to ensure the delivery and maintenance of software components. This process formalizes and acknowledges the contribution of a university group or institute and the responsibility for a software component lies with the institutes rather than individual collaborators. On the other hand, formal software agreements may not always be the proper tool since they could constrain the contributions of members of the collaboration and could keep the collaboration from taking advantage of software contributions from talented individuals at other institutions. Software and computing for the LHC experiment requires a global effort. The European contribution to the software and computing effort for the LHC comes primarily through the direct participation of CERN staff, fellows and associates, through contributions to the LCG projects and through development of the Regional Centres. The European contributions in these areas have been substantial. The direct participation of European institutes to the core software development inside the experiments, however, seemed surprisingly small. A contribution to the core software development coming from a broad group including a number of European institutes seems essential to the success of the LHC program.

The effort required to develop and maintain the core software in the experiment is likely to increase over the next few years. The experiments are entering a phase where the **collaborations' management needs to place computing and software development into a priority position** and to put it on an equal footing with detector construction and installation in order to be ready for the start-up of the LHC. **The committee, therefore, recommends that there be a regular review of the software effort during the period while the experiments build up their software and computing infrastructure.** The committee found that it was helpful to have open presentations with the experiments and the leadership of the LCG present at the review and recommends a similar format is used at future reviews.

3. SPECIFIC COMMENTS TO THE PRESENTATIONS AND ANSWERS

3.1 CMS

CMS has a very effective core group of software developers based at CERN. The CMS software and computing management has done a good job with their manpower estimates and the committee congratulates them on the efforts.

CMS has provided valuable support for development in the LCG applications area. They have been active in the implementation and led the effort in integration of POOL. The work has been very constructive for LCG and the committee commends them for their efforts and contributions. **We encourage them to continue to support LCG, to continue their development efforts and to maintain their full participation. We encourage them to participate in inventing a mechanism for optimising the level of effort inside the LCG.**

The committee reviewed the level of effort estimates from CMS and found them reasonable. **We find they are missing about eight (8) FTEs at this time in the area described as core software**. It would be most useful to address these shortages on the timescale of the upcoming Data Challenge and the follow-up preparation of the computing TDR. The CMS report to the committee also indicated that they would have shortfalls in other areas of the software and we remind the collaboration that they will need to act aggressively to solve these shortfalls as well.

3.2 LCG

The committee was impressed with the overall LCG project and the process for initiating projects seems to be working well. We welcome the proposal for enabling the involvement of the LCG developers in the integration of new LCG products into the experiments' infrastructure. The Application Area (AA) projects are well organized and are developing within the scope set out by the SC2 and the RTAGs. The first production release coming from the POOL project has been delivered to the experiments on time. The integration with the experiments is the critical issue at the moment for projects such as POOL and SEAL.

Although the committee did not review the LCG in detail, the level of effort seemed appropriate. There did appear to be a significant shortage of manpower beyond 2004. Now that the project is well underway, it is time to reconsider the optimisation of manpower across the experiments and the LCG projects. This review process should examine the manpower assigned to each project and the scope of the overall LCG projects.

The committee notes that there is a Comprehensive Review of the LCG projects planned for November 2003. This will provide a good opportunity for the LCG to review the project plan and scope.

The experiments and LHC need the GRID to succeed. The delivery of the promised tools is now late. It is urgent that these tools exist and are robust. The experiments eagerly await the opportunity to put them to the test through their next round of data challenges that begin early next year.

4. CONCLUSION

The Review Committee was impressed by work that has gone into the preparation of the presentations and reports presented at the review. The software teams in the experiments all have ambitious work plans for the coming years and will need the full support of their collaborations to meet their goals before the start of LHC data taking in Spring 2007.

The committee reviewed the manpower estimates from the four experiments and finds that the level of effort that is currently missing from the teams that develop core software is at a level of 20-25 FTEs. **The Review Committee finds that CMS is missing about 8 FTEs at this time in the area described as core software.** Every effort must be made to find this manpower to ensure that the upcoming Data Challenges are successful. The manpower that is missing is not easy to find inside the collaboration since it is primarily software engineers and computer professionals and skilled physicist programmers that are needed.

Between 30%-70% of the core software developers on the experiments are on temporary contracts (3 years or less). Since the demand for core software developers is unlikely to decrease before the start of data taking, **these temporary positions must be refilled as they become vacant. In addition, a few longer-term positions should be created in order to allow the experiments to maintain required software expertise.**

Given the existence of LCG, and the fact that there are 4 experiments, **the experiments <u>must</u> have a core software group working at CERN** to manage the project effectively and to interact efficiently with the LCG development team.

Given the size of the task and the distributed nature of the collaborations, the experiments also need to take advantage of expertise elsewhere. **In the absence of a formal MoU process, Software Agreements with collaborating institutes are a way of recognizing the work contributed in the software effort and help to ensure the delivery and maintenance of software components.**

The experiments depend on a variety of products being developed by the LCG in order to succeed. Many of these LCG projects are already underway and some are already being deployed (POOL). The LCG projects will undergo a Comprehensive Review by the LHCC in November. This upcoming review could serve as an opportunity to review the scope, priority and level of effort assigned to each project.

The experiments are entering a phase where **the collaborations' management needs to place computing and software development into a priority position** and to put it on an equal footing with detector construction and installation in order to be ready for the start-up of the LHC.