HL-LHC Software and Computing Review Committee Report

July 20, 2020

Authors:

Amber Boehnlein (Jefferson Lab) Chair

Volker Beckman (LHCC, IN2P3)
Catherine Biscarat (LHCC, IN2P2)
David Britton (Glasgow)
Rosie Bolton (Cambridge, SKA)
Claudio Grandi (INFN)
Fabio Hernadez (IN2P3)
Thomas Kulr (LMU)
Gonzalo Merino CIEMAT)
Gordon Watts (UWashington)
Jack Wells (ORNL)

HL-LHC Software and Computing Review High Level Summary

Meeting the needs for Computing and Software in the HL-LHC era is an active area of research and development on the part of WLCG, ATLAS and CMS. The HEP Software Foundation (HSF) has become an influential coordination forum for activities common across HEP extending beyond the LHC activities. Ultimately, Technical Design Reports will be essential for the computing and software activities. At this point, the focus is appropriately centered on R&D activities. To gauge the state of the current R&D activities, the HL-LHC Software and Computing Review Panel requested five documents. The two from ATLAS and CMS were requested to document a baseline computing model under current assumptions and to explore the elements of the models that dominate the costs and to outline R&D programs to mitigate risks. WLCG, HSF and the DOMA activity were also asked to prepare documents. On May 19, 2020, the computing coordinators of the ATLAS and CMS experiments, the leadership of WLCG, and the coordinators of HSF and the DOMA activity made summary presentations and took questions from the review panel. The review format was abbreviated due to the Coronavirus pandemic.

The review panel appreciated the preparation of the documents and the presentations and found the interactions valuable. The baseline computing models are being used by the experiments to understand the gaps between the projected needs and estimates based on budgetary considerations.

For this initial review, a detailed assessment of the baseline computing models was outside of the review scope. The experiments, WLCG, the DOMA project and HSF presented promising lists of R&D activities intended to close the resource gap, using the past experience that many changes can add up to a significant total. One active area of investigation is to reduce the requirements for storage and compute, for example, by reducing the size of the analysis formats. R&D activities include activities designed to develop and improve code performance on hardware architectures with accelerators (such as GPUs) and to undertake infrastructure projects to integrate in High Performance Computing (HPC) centres. At this early stage, a multi-prong approach seems prudent. For the next review, there will be a more formal assessment of the gains expected compared against a standard set of HL-LHC parameters. These parameters will be common across experiments to facilitate the assessments. They can evolve in time, and must be agreed upon by the LHC Machine Group, by WLCG and by the experiments.

WLCG: The review committee recognizes and appreciates the central role of WLCG in bringing together experiments, computing and storage providers, and the community in the effort to provide the best possible solution for the upcoming computing and data treatment challenge. WLCG has already established all relevant links and is engaged in a number of promising R&D studies. WLCG is the appropriate body for coordination of LHC specific efforts and to focus R&D on the most promising tasks, with special attention to common solutions across the different experiments, while working with the partner facilities to ensure the common solutions can be implemented. Several facilities issues were discussed in the review, including demands on the

Wide Area Network provisioned through LHCONE, and the implications of the data lake model on the system administration.

ATLAS and CMS: The review committee appreciates all the work that lay behind the inputs produced by both experiments to this review. The experiments should be congratulated on their excellent track record of computing innovation over many years, which forms a solid basis for tackling the future challenges of HL-LHC. ATLAS is using a system of baseline, conservative and aggressive R&D projects to assess progress. In addition to R&D, CMS plans to continue to investigate strategies used successfully in the past, such as data parking. Taken as a whole, ATLAS and CMS have a comprehensive list of activities. One area of concern shared by the experiments and WLCG is finding means to ensure that the highly skilled personnel essential for R&D in computing and storage have meaningful career paths within the LHC community to provide for sustainability and the need for continual evolution over the lifetime of HL-LHC.

Hep Software Foundation and Software: The review committee congratulates the HSF for establishing a forum where common software developments and techniques are discussed, especially for common software that extends beyond the LHC experiments. The value of this is recognized by the experiments and the community, including the work in foundational tools such as ROOT and Geant4. Common software has played an essential role for the community in the past and will do so, perhaps even more, in the future. We note particularly that effort on generators is needed as one of the components to solve the HL-LHC computing challenge, however the required work does not fit into the established funding schemes.

DOMA: The DOMA activity is tasked by WLCG with implementing the data lake concept of aggregating 'stateful' storage at relatively few sites, providing data access at the others through caching or streaming. Proof of concept activities have been promising. The three DOMA working groups (Third Party Copy, Data Access and Quality of Service) are addressing important issues. They should be commended for the progress accomplished since the start of the DOMA project two years ago.

The HL-LHC Computing and Software Review committee anticipates a second review in 9-12 months which will focus on detailed R&D roadmaps.

The following sections include detailed comments compiled during the review.

The ATLAS Experiment:

Summary

ATLAS described the main components of the model and presented three evolution scenarios: baseline, conservative and aggressive.

The results of the model were presented in both an ATLAS and a common WLCG set of parameters.

Results

The baseline scenario is clearly defined. The conservative and aggressive R&D scenarios are useful tools to encapsulate future evolution information. It will be good that ATLAS works towards quantifying better the potential impact of each of the R&D projects in terms of costs and outcome. Risk mitigation strategies should be elaborated.

Recommendations to ATLAS:

- The ATLAS data carousel is presented as a central part of the R&D strategy to optimize the usage of resources. We understand a model like this has deep implications on sites' operations, so we encourage ATLAS to develop a clear roadmap and engage sites in the process. Also, we think it will be good that ATLAS further quantifies the potential savings of the model, including mitigation strategies for the inherent risks such as the overall uncertainty on the future of tape technology.
- Full adoption of DAOD_PHYSLITE appears to be strategic in order to keep ATLAS resource requests under control. We encourage ATLAS to be sure that this activity receives the needed effort.

The CMS Experiment:

Summary:

The CMS input to the review presented:

A description of the CMS computing model, including an extensive list of past achievements and the current plans to move towards smaller analysis datasets.

The results of modelling HL-LHC requirements based on both a CMS and a common WLCG set of parameters.

Results

CMS' plots suggest that, using the CMS model assumptions, flat-cash projections would result in a factor of 4x shortage of CPU throughout Run-4; a shortage of disk that grows to a factor of 2x by the end of Run-4; and a shortage of tape which grows to a factor of a little more than 2x at the end of Run-4. Using the WLCG common parameters, these shortfalls grow to 7x (CPU); 4x (Disk); and a little over 3x (Tape), as far as can be judged from eyeballing the curves.

Recommendations to CMS:

- Parking data was outlined as a possible mechanism to enable CMS to balance the overall cost of resources needed at the Tier-0. The documents were inconsistent on how and when data parking would be used and we encourage CMS to clarify this point.
- CMS noted concerns about the future of Cold Storage (Tape) and the need for R&D on how to best use Fast Storage. However, no related plans were presented. We encourage CMS to formulate a plan.

Comments to ATLAS and CMS:

The reviewers did not feel there was sufficient information, at this point, to accurately assess the true magnitude of the challenges facing the collaborations as it was not uniformly clear how much of the expected future improvements, if any, were included in the modelling.

The reviewers felt that whilst ATLAS and CMS discussed a number of ways future requirements could be reduced, this stopped short of presenting a roadmap for the next few years. Many of the potential gains depend (partly or entirely) on projects external and/or collaborative projects (eg DOMA and Event Generator/Detector Simulation improvements). In some of these cases, there is a responsibility for ATLAS and CMS to provide input/requirements; to deploy and test; and to possibly adapt the computing model to incorporate. However, in all these areas, ATLAS and CMS will need to understand the expected benefits, current status, and level of risk contributed to their overall strategy. It is understood that the outcome of R&D cannot be known; on the other hand R&D is not undertaken unless there is some expectation of its benefit and criticality.

The reviewers noted that between them ATLAS and CMS raised a number of areas of concern, including future hardware costs, long-term availability of cold storage; software development manpower; HLT rates; and networking. These are shared risks across experiments and WLCG. As such, any crisis would be managed in a coordinated way. Other approaches are partly or entirely under an experiment's control and should be examined and, if necessary, mitigated.

In particular, projections for networking in general seems to be reasonable (albeit with a few possible pinch-points). However, ATLAS and CMS should examine whether their future computing model overly exposes them to a risk from networking compared to other experiments.

For the next review, ATLAS and CMS should:

- **Recommendation-1**: ATLAS, CMS need to iterate with WLCG and with the LHC Machine Group on a set (or sets?) of common and realistic parameters that are appropriate for baselining the computing models. Once agreed, these parameters should be used consistently. In addition to common machine and running parameters, these agreed

parameters should include (a) an agreed expected cost evolution of resources and (b) the expected lifetime of hardware so that hardware replacement costs can be properly included.

- **Recommendation-2**: Starting with the performance of the current software stack and existing computing model, ATLAS and CMS should produce tables showing how they expect CPU and storage requirements to be mitigated by current and future developments, within and outside of ATLAS and CMS, and assess the status and risks associated with each of these developments. These tables would then underpin a credible strategy.

- **Recommendation-3**: ATLAS and CMS should develop R&D roadmaps that are informed by strategy developed in the previous recommendation.

WLCG

The review committee recognizes and appreciates the central role WLCG has in bringing together experiments, computing and storage providers, and the community in the effort to provide the best possible solution for the upcoming computing and data treatment challenge. WLCG has already established all relevant links and is engaged in a number of promising R&D studies. WLCG should stay at the centre of these efforts and help to focus R&D on the most promising tasks, with special attention to common solutions across the different experiments.

Common parameters: A continuous effort is needed to work on a common and realistic set of parameters across all experiments in the preparation of the TDR. These parameters can evolve in time, but need to be agreed upon by the LHC Machine Group, by WLCG and by the experiments.

Implementing a computing model in the (political) reality: In view of strong preferences of some of the WLCG partners to invest in certain infrastructures or software developments, the solutions investigated cannot always follow the approach of using the best possible infrastructure/software for a given task. Thus, the approach to find a computing model for HL-LHC has to take into account road map decisions in the various partner countries. WLCG is thus encouraged to closely monitor the development of the computing landscape and political background in the participating countries in order to propose a computing model that can also be financed and implemented. Planning should also consider the evolution of the international computing landscape, such as the European Open Science Cloud (EOSC).

Network: The availability of a strong and reliable network as a backbone for the data lake model to work is of great importance. Close collaboration with LHC partners and with NRENs is required to ensure that the evolution of the networks is sufficient on the international, trans-atlantic, and national level.

The network has to be shared with other experiments and it is crucial to continue to work on having a common monitoring framework (marking the traffic by experiments or activities). We further encourage simulation of the new model to understand how well it will accommodate the experiment's production and analysis workflows.

Budgetary considerations: The flat budget approach in planning is certainly appreciated. However, there might be some possibility to on-board new partners into WLCG. In addition, we might be able to convince funding agencies to a slightly stronger commitment, if the arguments from the physics side are strong. Therefore, WLCG might also consider a scenario with a moderate budget increase of +10% particular if this can be shown to benefit a broader community.

Coordination with other large experiments: Many WLCG facilities also support other large experiments, from the particle physics domain but also from other scientific fields. WLCG is encouraged to continue on the path of integrating other experiments in their discussion and converging on the same technical solutions when possible. These are, at the time being, Belle 2 and DUNE, but might include further experiments from other domains in the future.
The advantages and risks of a common "Scientific Computing Infrastructure" (as described in the Granada paper) should be further investigated.

Improve coordination: In the course of this review, we noticed a lot of work is ongoing and led by different bodies, experiments, DOMA, WLCG, and HSF. The coordination between these has to be verified and investigated for potential improvement between the different bodies.

Data lake approach and load on parties involved: The data lake scenario requires substantial caching and also efforts on the data providing sites. It has to be evaluated what this means for the partners involved, and whether this scenario can be implemented in all major participating countries.

The DOMA Activity

The three DOMA working groups are all addressing important issues in data management. They should be commended for the progress accomplished since the start of the DOMA project two years ago.

The third party copy working group is focusing on a somewhat shorter term but still important issue of modernization of the software tools for inter-site, bulk, asynchronous data movement. The progressive roll-out of these tools, foreseen in Run3, will require modifications of the experiments' data replication infrastructure. The usage of common tools such as the new mechanisms for identification and authentication of users and services, as well as Rucio and FTS for driving the data movement for both ATLAS and CMS will likely help keeping this process under control. The reviewers felt that a robust, standards-based foundation is being prepared by the DOMA project in this respect and the work must continue as planned.

The access working group is addressing a very important topic of taking advantage of the capacity of the network interconnecting WLCG sites to cache data near compute nodes and to stream data to compute nodes from remote sites, as opposed to statically replicating data, as is currently done. The network will become an even more important resource that needs to be provisioned, monitored and exploited as such. The delays associated with network capacity provisioning in some countries and in some regions of the world need to be factored in so that the connectivity of the WLCG sites will be ready on time and at the appropriate level of capacity for the data lake model to work. This issue seems to be well understood by the DOMA project.

The reviewers understand that as a result of implementing this way of delivering data to the applications, the roles of the WLCG sites may be further specialized. Specifically, it is expected that most of the work of delivering data to remote, cache-only or even storage-less sites will be shared by the current Tier-1s and big Tier-2s. Those sites would need to deploy storage and network capacity for delivering data to remote sites in addition to the capacity required to deliver data to their local applications. This potential additional cost would need to be planned for by those sites and their individual funding agencies.

In addition, it is understood that one of the goals of implementing this data delivery model is to reduce the costs related to the effort needed to operate the storage infrastructure in small- and medium-sized WLCG sites. However, from the provided information it is not possible to conclude whether the result will be a net reduction in effort for operating the WLCG infrastructure as a whole or rather a shift of the required manpower to the sites expected to power the foreseen WLCG content delivery network.

The Quality of Service working group is addressing another area that is expected to help reduce equipment costs for storage, by specializing types of hardware according to specific usage by the experiments (beyond disk and tape, as it is currently the case). Introduction of this differentiation in usage will as well require adaptation of both the experiments' workflows and the storage infrastructure of each individual site. The impact of the latter is not clear, in particular for the many WLCG sites which operate storage services shared by several experiments and even several sciences.

It is not clear for the reviewers to what extent those necessary activities will be enough to close the gap between the storage requirements as expressed by both ATLAS and CMS (which are understood as not including the potential benefits of the DOMA project results) and the storage capacity that is foreseen to be available within the expected flat budget scenario.

The data lake model being developed by the DOMA working groups makes even more central the role of Rucio and FTS for both ATLAS and CMS data movement. Although not a technical barrier, this poses the problem of the sustainability of Rucio and FTS in the long term. The reviewers welcome the fact that both ATLAS and CMS have agreed on using common tools for data movement which is likely to help in the long term.

The constraints imposed by several HPC sites, often specific to each one of them (e.g. no outbound network, no local storage, etc.), could make difficult the implementation of the foreseen data lake model for effectively and efficiently delivering data to applications executing at those sites. This could have a negative impact on the attractivity of the WLCG platform for those sites and therefore the possibility of exploiting their computing capacity for the experiment's needs.

Recommendations:

- The DOMA project should include in its programme of work specific tasks and their associated milestones for quantifying the potential reductions in equipment costs and manpower savings resulting from the data caching & streaming strategy for delivering data, as well as from the usage of different kinds of hardware according to the expected usage by the experiments.
- The DOMA project should make sure that sites expected to be at the core of the foreseen WLCG data delivery network are engaged and play an active role in the project so that their inputs are taken into account early on in the design and testing phases of the system.
- For the purposes of this review leading up to TDR it would be very beneficial to see the various tasks and targets mentioned in the document placed onto a timeline showing how they relate to each other and to broader milestones in the WLCG and HL-LHC projects.
- Special attention should be devoted to keep the effort needed for the software development sub-projects (e.g. RUCIO, FTS, AAI, XRootD & Xcache, dCache, etc.) at the right level to reach the DOMA goals within the time constraints.

Software/Hep Software Foundation:

We congratulate the HSF for establishing a forum where common software developments and techniques are discussed. The value of this is recognized by the experiments and the community. Common software has played an essential role for the community in the past and will do so, perhaps even more, in the future. The cornerstone of the common projects are ROOT, Geant4 and the event generators.

The activities described in the document are perceived to be essential by both the ATLAS and CMS experiments. HSF is a community-based activity with dedicated leadership effort and contributed effort on a per-project basis. HSF is playing a strategic role in coordinating developments that are vital in keeping the resource request for the experiments under control. It would be beneficial to formalize a role by which the priorities of WLCG can be transmitted to HSF. This would provide a mechanism for more explicit connections from the experiments to HSF as well as some of the larger external national efforts. Recognition is an important way to garner further funding, and to recognize a group's efforts and their value in the field.

The lack of career perspectives is identified as a risk for attracting and keeping people with the necessary skills and guaranteeing sustainability. Training and education are an important part of the strategy.

Common themes came out of the ATLAS, CMS, and HSF documents. We encourage the groups to work together to extract common projects that make sense. A few examples follow.

- Reconstruction algorithms
- Common effort on analysis preservation
- Encouraging a common approach and best practice for HEP software in accelerators.
- (Conditions) databases?

Effort on the generators and detector simulation is needed as one of the components to solve the HL-LHC computing challenge, and particularly for work on generators, funding mechanisms will need to be established.

Comments:

- The effort on common software has been quite successful to date starting with the initial focus on ROOT and Geant4. These efforts should continue and be strengthened and extended with other common projects in consultation with WLCG and all of the LHC experiments.
- Structures should be considered that would allow institutes and funding agencies to make more formal commitments to international projects on common software.
- We recommend that members of the HSF community continue to participate in International standardization bodies to be sure that what comes out of them is usable by HEP experiments.
- There is a CERN process looking at Open Data and Analysis Preservation. The committee looks forward to hearing in the future the roles that HSF will play in realizing the strategy, particularly for analysis preservation.

Appendix: Charge for the 2020 HL-LHC Computing review

Technical Design Reports (TDR) will be developed by ATLAS and CMS to document the plans for Computing and Software in the HL-LHC era. The timeline for delivery of these documents are currently understood to be in the second half of 2023 for a review in early 2024. Under the auspices of the LHCC, a review panel will assess the TDRs, similar to the process for reviewing detector TDRs.

Prior to the review of the Computing TDRs, several preparatory meetings with the review panel will take place, with the first meeting scheduled for May 18-20, 2020. This initial meeting will focus on experiment specific issues and on long range R&D including Data Organization, Management and Access (DOMA). The outcome of the meeting will be a short summary assessing the state of the plans.

The documents and presentations should cover the following areas:

- Establish a baseline computing model. Using the best available knowledge concerning data rates, the size of output formats and the processing time of production tasks, scenarios for computing and storage projections should be presented. Where applicable, estimated resources for user analysis can be included. The baseline computing model should outline basic assumptions about data access and placement, and hardware evolution.

- From the baseline computing model, establish anticipated cost drivers and infrastructure assumptions. The roles of the WLCG data centres should be outlined, including the expected technical specifications such as network connectivity.

- The major technological risks and uncertainties should be outlined as well as any necessary programs of research and development that are underway to mitigate the risks and/or introduce new methodologies and technical advances.

In addition, the May 2020 meeting will include short presentations on common tools and community software (such as ROOT, Geant4 and event generators and the broader aspects of heterogeneous architectures) to assess the plans and timelines for HL-LHC functionality.

Supporting documentation (each 20 pages or less) will be made available by May 1, 2020 and will include:

- overviews from the perspective of ATLAS and CMS,
- outline of DOMA and necessary technology R&D
- common tools and community software
- facility issues, including the role of the Tiers, non-Grid resources such as HPC Centres and critical policy issues, e.g. authentication and authorization