Snowmass2021 - Letter of Interest

HPC Facilities for Large Experiments: Opportunities and Challenges

Thematic Areas: (check all that apply □/■)
■ (CompF1) Experimental Algorithm Parallelization
□ (CompF2) Theoretical Calculations and Simulation
□ (CompF3) Machine Learning
■ (CompF4) Storage and processing resource access (Facility and Infrastructure R&D)
■ (CompF5) End user analysis
□ (CompF6) Quantum computing
□ (CompF7) Reinterpretation and long-term preservation of data and code

Contact Information:
Katrin Heitmann (Argonne National Laboratory) [heitmann@anl.gov]

Authors: Stephen Bailey (Lawrence Berkeley National Laboratory), Debbie Bard (Lawrence Berkeley National Laboratory), Matthew Becker (Argonne National Laboratory), Rahul Biswas (Stockholm University), Ian Foster (Argonne National Laboratory/University of Chicago), Salman Habib (Argonne National Laboratory), Zelimir Djurcic (Argonne National Laboratory), Katrin Heitmann (Argonne National Laboratory), Evangelos Kourlitis (Argonne National Laboratory), Tom LeCompte (Argonne National Laboratory), Peter Nugent (Lawrence Berkeley National Laboratory)

Abstract:
Experiments in High Energy Physics have major computing demands across the Cosmic, Energy, and Intensity Frontiers. High Performance Computing (HPC) facilities offer opportunities to the experiments to access centralized large-scale resources, potentially fulfilling essential parts of their computational demands. Given the massive computational needs of the HEP community, efficient use of the facilities is a crucial element. However, such efficient use, given the complex computing models for the experiments, comes with its own challenges. In this LOI we highlight some of the opportunities and challenges that HPC facilities offer and pose for HEP experiments. To develop optimal solutions, HPC facilities and the HEP community will have to work closely together and continue to invest in the development of approaches that will allow the experiments to take advantage of the unique capabilities that HPC facilities have to offer. The Snowmass Computational Frontier provides an important vehicle to foster discussions across the science frontiers and the facilities to enable significantly increased use of HPC resources in the future.

1 Introduction

Computational demands across the three frontiers in High Energy Physics (Energy, Intensity, and Cosmic Frontier) are growing rapidly. The Energy Frontier has had to face this growth for a long time already, with the HL-LHC being a significant new challenge on the horizon. In the Intensity and Cosmic Frontiers, the sizes of the detectors and the associated experimental datasets have been constantly increasing with time – the need for large, well-managed computing resources has also become evident.

The computing demands and ecosystem for HEP experiments are complex, from needs to process the data from the experiment directly, to predictions from large-scale simulations, to end user analysis needs. This nontrivial structure is mirrored in part by the diverse Snowmass Computational Working Groups as well. In turn, this complex computing landscape leads to nontrivial computing models for the experiments across all the frontiers. In the Energy Frontier, this challenge has been addressed by building a tiered approach that includes computing resources at a range of scales. For the Intensity and Cosmic Frontier, the computing models for the new large experiments are still somewhat under development and thus the experiments have the opportunity to argue for and construct a more flexible approach tailored to their needs.

In this LOI, we focus on the role that High Performance Computing (HPC) Centers can play in this environment. As an example, in the Cosmic Frontier, NERSC has taken on the role of primary computing center for collaborations like the Dark Energy Spectroscopic Instrument (DESI), LSST Dark Energy Science Collaboration (LSST DESC), LZ collaboration, IceCube Neutrino Observatory, and CMB-S4. Both data processing and collaboration-wide analysis tasks are supported. The LSST DESC has also taken advantage of the ALCF and resources from international partners for targeted computing needs, such as large-scale simulations and data processing. The Energy Frontier experiments, like ATLAS and CMS, have also been encouraged to use HPC facilities for some of their computational needs, and have carried out focused work at the three DOE/SC facilities (NERSC, ALCF and OLCF) in the past. Among the Intensity Frontier experiments, NOvA has been able to take advantage of NERSC. These examples clearly show that HPC centers and major HEP experiments can work together successfully. They have also revealed challenges, many of which can be addressed by close collaboration between the centers and the experiments. In this LOI, we wish to highlight the opportunities and the challenges that HPC facilities provide for large experiments and how, with close collaboration between the facilities and the experiments, some of these challenges can be addressed.

2 Opportunities

The integration of HPC centers in the computing model for major experiments offers many opportunities that are difficult to realize with computing resources owned and maintained by the experiments themselves. This is in particular true for the Intensity and Cosmic Frontier experiments that have not yet built computing facilities as elaborate as those maintained for the Energy Frontier.

The HPC facilities provide well-maintained, centralized computing resources. The hardware is replaced on a regular cycle, is usually reliable, and does not suffer from problems old hardware often exhibits. This provides an opportunity for the experiments to partially (or even substantially) offset the investments they have to make in their own hardware infrastructure or in the development of a state-of-the-art software environment. The cost of replacing and installing hardware on a regular cycle is often a major burden for the experiments and this problem can be made more manageable, provided certain conditions are met by the HPC centers. In addition, the scale of the available computing at the HPC centers enables very large computing jobs and can therefore provide very fast turn-around on computing tasks such as the creation of large simulations or data processing at scale. However, use of this capability requires that the experiments optimize their software pipelines to run at these large scales, a major challenge in of itself.

The HPC facilities also offer a range of training opportunities that help users at all levels. Specific programs, like the NESAP program at NERSC or Early Science Programs at the LCFs provide resources (training as well as people effort) to optimize codes for the available platforms. In addition, the large user community and strong engagement of the facilities staff enable efficient sharing of expertise and therefore more rapid progress on all fronts.

The HPC facilities also provide large-scale storage resources and straightforward access for many collaboration members. For example, LSST DESC has been able to build a tailored environment for their collaboration members and provide them an easy entry to using NERSC computing facilities.

3 Challenges

As hinted above, many of the opportunities provided by HPC facilities also pose associated challenges for the experiments. The regular replacement of the computing resources often comes with a change in computational architectures and programming models. For example, at both ALCF and NERSC the move from KNL architectures to GPU-enhanced systems caused a major disruption for the experiments; it is not clear how they will handle this new challenge. The computing architectures at the facilities can potentially evolve on a faster timescale than the lifetimes of the experiments. Unfortunately, many of the code bases and computing models employed by experiments have not been amenable to such changes on these short timescales. In addition, the experiments might wish to use more than one HPC facility. The facilities all have sufficiently different systems that portability issues can often become a critical bottleneck. In addition, software environments vary across the facilities and many users carry out development work on their own resources adding another layer to the portability challenge.

Another challenge is posed by the HPC centers' allocation and access processes and policies. HEP experiments are large international collaborations and access to the HPC systems is sometimes not granted to all members of the collaboration. If a center is supposed to function as the major analysis hub, this can cause a major problem. In addition, stable allocations are not provided over a long period of time. Instead, yearly proposals have to be written and approved by the centers. Exceptions to this policy exist already, but in order to establish the HPC facilities as major resources for the experiments, this problem has to be solved in a more coherent way, with long-term arrangements becoming the norm.

Many experiments rely on the uninterrupted availability of computing resources for certain tasks. No current HPC facility provides 24/7 access to filesystems and computing resources; indeed, all have regular maintenance periods when resources are not accessible. Further, catastrophic machine failures or external events can create downtime periods longer than the typical timescales for maintenance. This challenge could be overcome by either only shutting down systems at one facility partially or by better coordination of the centers to pick up critical, time-sensitive computing loads during these events. In addition, the allocation demands for the experiments are often not evenly distributed throughout the year but have time-critical demands due to a range of reasons. Given that the HPC facilities support many users across many fields, these time-critical needs might not be easy to fulfill. This list of challenges is clearly not exhaustive but provides a few important examples.

4 Path Forward

As part of the Snowmass Process we hope to develop a comprehensive view of the integration of HPC facilities into the computing models of major HEP experiments. We strongly believe that many of the current challenges can be solved via a close collaboration between the experiments and the facilities. Many of the challenges will need continued investments to develop tools and approaches that will enable efficient use of the powerful resources available. At the same time, HEP experiments might help the facilities to develop solutions to some of the challenges that will then enable other experiments across the DOE complex to integrate these resources efficiently in their computing models.