HEP Computing Challenges

Christopher D. Jones, Adam L. Lyon[1] and James F. Amundson
Fermilab Scientific Computing Division
August 18, 2020

Introduction

Experimental Elementary Particle Physics is approaching major changes to how it handles computing. This change is driven by two different factors: future experiment's computing needs and the changing computing landscape. In the white paper we will motivate both points and stress the need for additional resources to meet these needs.

The computing needs for upcoming experiments such as ATLAS and CMS during the High Luminosity LHC, HL-LHC, era as well as the neutrino experiment DUNE are well beyond those of the previous generation of experiments. The need comes both from a large increase in the amount of data to be taken per year as well as the increased complexity of processing the data. In a white paper we intend to address the scale of the data processing computing challenge facing these experiments as well as the challenge storing and retrieving the data.

In order to meet the computing needs for the upcoming experiments, the field will need to go beyond its present reliance on CPU resources provided by the world wide computing grid. To meet this need the field will have to look into new computing hardware, such as GPUs and/or FPGAs as well as new computing facilities such as commercial cloud computing and High Performance Computing centers. In this letter of intent (LOI), we will examine areas that will require attentional effort. Specific details and options will be described in other focused LOIs and later white papers. This LOI serves as an introduction and a “big picture” description.

Computation

As mentioned above, the computational resources are evolving to even more massively parallel systems and specialized processors. All of the leadership class DOE Exascale High Performance Computing centers (NERSC, ALCF, OLCF) are focusing on machines with a large fraction of their computing power realized by GPUs. Furthermore, specialized machines, such as those for AI, will feature highly flexible FPGAs. Making efficient use of these systems will require adopting heterogeneous computing techniques in our code bases. The compute power these machines offer, however, will be much more than what HEP has had access to in the past. Of course gaining access to computing time at these centers is not trivial. Paradigms for working within the centers' award processes or negotiating special access will need to be

[1] Corresponding author email lyon_at_fnal.gov

determined. We will also require interfaces (e.g. HEPCloud) to access these machines and submit workflows. HEP is used to running on sites where we or partner institutions select and own the hardware. Running on HPC sites means having to adapt to their computer and software choices. One major area of research involves how to effectively use the storage systems on HPCs.

Not everything can be run effectively on an HPC machine and some applications and workflows need to run on-site. Databases and their servers will be hosted on host lab systems. Submission system infrastructure also needs to be managed and run at the host lab. There are some workflows that are very unique or are under test that should happen on local computing resources. Therefore, compute sites at local institutions and accessible over the Grid need to be maintained and upgraded accordingly as needed.

Commercial clouds such as AWS and Google serve as extra resources to be brought to bear when necessary. Furthermore they may also offer specialized and useful hardware, such as TPUs and FPGAs. Interfaces to such resources must be maintained and updated as new services are offered.

Many institutions are exploring high capacity and high throughput analysis facilities. Such facilities will offer analysis tools (e.g. Jupyter Notebooks), computing and storage systems such that extremely large datasets can be analyzed with time-to-results compatible with interactive users. The demands on these systems may be extremely high, making scheduling and handling many multiple users a challenge.

Long Term Storage

Being prodigious producers of experiment data and simulation and the need to share those data with collaborators around the globe, several HEP institutions, and Fermilab in particular, have extensive experience in long term storage of massive volumes of data. The technology for such storage and disk cache will need to evolve to keep up with the huge demands from the HL-LHC and DUNE along with the software for data handling, movement, caching, and cataloging. High throughput wide-area-networking will be required. Making this cost effective will be a substantial challenge.

Frameworks

HEP data processing frameworks have been used by experiments to handle the processing of online data filtering, data simulation and reconstruction. To meet the computing needs for experiments in the future will require the frameworks to evolve in order to handle the changing computing hardware landscape, i.e.. many-core system, GPUs, FPGAs, as well as a diverse set of computational sites, i.e. traditional HEP grid sites, commercial cloud computing and the High Performance Computing centers.

R&D in Post-Moore Computing Programming Models

Looking beyond the Exascale area, we expect “Post-Moore” computing and very specialized processors to become more prevalent. Such technologies include neuromorphic (chips that mimic biological structures or “brain inspired”) and quantum computing. To be prepared for this era, R&D will be required to determine programming models for such radically different means of computation. Frameworks will need to evolve to take advantage of the new computing paradigms that such technology could offer.

Conclusion

The intention of this LOI is to introduce the challenges facing HEP computing in the HL-LHC and DUNE eras and beyond. Support from funding agencies is required in order for our field to evolve with new computing technology and paradigms. It is unlikely that a single experiment on its own can solve the whole challenge and it is beneficial to find a common solution for all experiments for a large fraction of the new developments. We will need a substantial and specialized workforce with expertise in these cutting-edge areas. We know it will pay off. Past experience proves that adapting to new technologies and to more powerful and efficient computing leads to exciting scientific discoveries that were previously thought to be impossible to achieve.