On the need for a closer collaboration between HEP Experiments and HPC centers

High Energy Physics (HEP) experiments deal with real or simulated particle collisions, which can be processed and analyzed individually without any correlation. The distributed data-intensive computing of the HEP community is referred to as High Throughput Computing (HTC): it heavily relies on in-house custom-built computing farms, glued by the GRID middleware as developed in the past two decades.

Experiments computing models are constantly evolving, with a slow but steady transition towards a smaller list of requirements for such systems, helped in this transition, by system virtualization. As of today, even if not predominant, Cloud access is already a reality for all medium-large size experiments.

At the same time, larger experiments have prototyped the utilization of High Performance Computing (HPC) systems. This comes from a variety of factors:

- The availability of large-sized research grants from HPC centers for HEP use cases. A typical PRACE HPC system deploys resources comparable to the full HEP scale; this is even more true when considering the top systems from the US and China.
- The push by funding agencies and countries to have HEP using the HPC infrastructure for at least a fraction of the processing. HPC systems are built with the best technology available, even before it hits the shelves, in order to optimize global system sizes and power consumption. HEP systems are instead built to be economically affordable, and hence usually behind the former by one, if not two, generations, privileging solutions which give the best performance over money. The utilization of HPC systems for HEP provides access to more advanced resources, with all the added benefits.
- HPC systems are funded via specific programs (PRACE in EU, and government programs in Americas and Asia), while HTC systems must be at least partially deployed by using the standard budget of Research Institutions, with extrapolated needs currently expected beyond sustainable levels. Even recent estimates for High Luminosity LHC (HL-LHC) exceed current budget levels; hence the need to access new opportunities and collaborations is felt of primary importance.
- On the other hand, HEP has been up to very recently the only Science with data-intensive workflows at the scale of hundreds of Petabytes or more, rapidly reaching the exabyte scale. Our today's use cases are a showcase of what is expected to be largely common in the next decade, with Astroparticle, Genomics and other sciences expected to face a similar situation. HEP can thus offer a technical ground for experimentation of the typical use cases HPC systems will have to face in the future.

Efficient use of HPC system by High Energy Physics experiments is not trivial since these systems are custom-built having in mind use cases different from HEP ones. Such systems feature performant node-to-node interconnections, needed for large scale MPI tasks, scarce local scratch disk, and are not meant for accessing data residing outside the facility. Recently, HPC systems include node per node accelerator cards, in order to boost total performance at lower power consumption and hence global ranking. Storage systems are optimized for latency and speed, and not for total size or large overall event throughput as in HTC systems.
Current HEP HTC systems are built using different technical solution: note-to-node connectivity is scarcely relevant, large on-node scratch areas are important, and global connectivity is needed in order to access remote datasets and calibrations. The use of accelerator cards is marginal if not absent. The HEP workflows are typically data-intensive, and thus current implementations deploy large storage systems close to the computing farms. Almost 100% of the experiments software stack is written for Intel x86_64 architecture, definitely the best choice for affordable computing in the last decade.

Given the highlighted differences of the two approaches, how can we make sure current and future experiments software can be adapted to tap the enormous computing power HPC systems will deploy in the next decade, to a mutual interest?

From the technical point of view, given an HPC architectural design driven by the best performance, the utilization of such system for HEP, strongly depends on how flexible and general purpose the systems will be built. If the Funding Agencies want HEP to use efficiently the HPC infrastructure it is necessary that we reach mutual technical understanding, in order to be able and shape our workflows for an HPC utilization; on the other hand, we expect that HPC systems will need to evaluate a few key aspects in the design to better match HEP requirements.

In the document we are presenting here, CMS offers its technical view on the requirements and “desiderata” when planning to run on HPC systems. It should serve as a help for the first handshaking with sites, in order to be able and understand the best directions towards a future more ubiquitous utilization of such systems for HEP processing.