Ian Bird
30 January 2017

Evolution of Scientific Computing

INTRODUCTION

High Energy Physics (HEP) has demonstrated a unique capability with the global computing infrastructure for LHC, achieving the management of data at the many-hundred-Petabyte scale, and providing access to the entire community in a straightforward manner. This is still a unique facility in science, but aspects of it are more and more needed for other scientific endeavours. Fifteen years ago, when the work started, there were no examples of how to build such a system, and no experience from industry or others. If we were to design the system today, of course, we would benefit from the tools and expertise of the global internet companies.

HEP has a challenge for the foreseeable future – which is how to achieve a scale of computing and data management that is orders of magnitude greater than that of today, while maintaining a reasonable cost envelope. HL-LHC is the most immediate such challenge, but we also have other high-rate experiments, and future potential facilities that must be considered. The desire is to build on the existing global structures built for LHC, leverage the experience and capabilities now available in the internet sector, and plan to evolve a HEP-wide scientific data and computing environment for the future of our field. Importantly, in addition, we observe similar needs arising in related fields (astronomy, astro-particle) with many of the HEP facilities often directly involved. In planning for the future, we must take into account compatibility and synergy at the facility level.

Taking the success of the Worldwide LHC Computing Grid (WLCG) as a starting point it could be envisaged to evolve the infrastructure and tools as a basis for computing for HEP for the coming years, while challenging the concerns of cost (both in terms of equipment and operationally), organization, and community needs. Can we build on what we have today, evolving and providing tools and infrastructure services to be used or adapted to future needs of the HEP community?

In considering how to do this, we can think of essentially four key high level areas to address. These areas are discussed below.

General Infrastructure

The fundamental components of a HEP infrastructure are the essential building blocks of the computing, and probably today one of the major successes of WLCG. We have in place global networking infrastructures, not only those provided by the National Research and Education Networks (NRENs) and their coordinating bodies, but HEP-specific structures such as the private LHC Optical Private Network (LHCOPN), and the very successful LHC Open Network Exchange (LHCONE) overlay network, which provide the ease of management and connectivity that will be essential for the future. Today this is already used by more than the LHC experiments. While the LHCOPN is a private network, it is nevertheless a good model for specific situations in the future – as will be discussed below.

We also have a global Authentication, Authorisation, Accounting (AAA) service, and associated trust and policy networks. This is extremely valuable, and very unique. However, it is clear that the X.509[i] underpinnings are probably not the best for the future and other federated identity mechanisms (e.g. eduGAIN[ii]) are being investigated. These base services are supported by mature monitoring, operational and support processes and teams, including worldwide collaboration on security and incident response. All of the above are recognized as being of high value, and we have seen other HEP experiments (e.g. Belle-II, and others, such as the CERN recognised experiments) asking to be able to benefit from them already.

For LHC we have the computing resources at close to 200 sites using this infrastructure. The "grid" that enables the coherent use of those resources must evolve over the coming years, and be capable of supporting continually evolving computing models, and being agile to technology changes. This is already the case with the adoption of clouds, private and external, use of opportunistic resources (HPC facilities, volunteer computing), etc. In addition, significant effort has been invested in developing sophisticated data management tools to deal with the processing and distribution of the huge volumes of data. One of the concerns today for LHC is the cost of the computing resources, and in particular the cost of storage which accounts for close to 70% of the overall hardware cost. That cost is in large part driven by the need to distribute data globally, with consequently many, many, copies of the data, and consequent costs both in storage and in operations. For the future, we must consider an alternative model to reduce those costs.

A HEP Data Cloud

An idea that has been discussed recently, builds on the experience of large commercial cloud providers, as well as the LHC expertise in many-hundred-Petabyte scale data management. The idea is to connect the large HEP data centres between themselves with a dedicated and private multi-Tb/s network. This "virtual data centre" would store all of an experiment's data, and by policy replicate it between the data centres. This can be thought of as a RAID across data centres. This network is probably managed with Software Defined Networking (SDN) to ensure the data flows and reliability. In this way, we achieve reliability, and availability. Into this data cloud we would plug compute resources. These resources may be co-located at the data centres, or may be other facilities – such as commercial centres, other large-scale HEP-owned resources, etc. The model also allows for inclusion of commercially procured storage. Policy would prevent reliance on those for non-reproducible data sets, and should be redundant enough that a commercial centre could "unplug" without loss of data. This clearly relies on a very strong collaboration with the networking community, with adequate policies and capabilities to agilely connect to commercial partners.

A key concept in this vision is that data does not leave this "data cloud". Rather, clients access data within the cloud from externally and arrange for the processing in that infrastructure. Clearly "ntuple" scale data sets will eventually be downloaded for analysis locally, but all organized and large scale analysis would be done in the data cloud. In the LHC case all of the AOD/mini-AOD level data would be kept in the cloud. Having all of the data virtually co-located in this way may open the way to radically new analysis models, and strongly supports todays models (e.g. analysis trains). It would also permit the increased use of tape as an active store for organized analysis, again helping with cost.

This type of model also allows cost optimization through the use of hybrid centres: HEP owning compute resources at a level that is guaranteed to be fully used is very cost effective, and supplementing this with elastically provisioned resources as needed (presumably with some form of cheap spot-market pricing). This would allow an agile control of the cost, and can evolve as the commercial markets evolve. Of course, this may require new funding models, and management methods. The data centres we address here are at the scale of a few MW, much less than large scale commercial centres, but larger than most University solutions.

Today we have many Tier 2 centres that also provide significant resources, together with other opportunistic resources. We should distinguish in terms of scale rather than role. Large scale centres could participate as "compute plug ins" in the above model, while others primarily provide simulation resources. It should be remembered that some 50% of LHC compute loads is simulation, and the same will be true for forthcoming large experiments. Depending on the type of resource, some centres may be best suited to specific types of workload (e.g. HPC for event generation).

This model will also easily accommodate the anticipated case of a funding agency moving academic computing internally to cloud credits rather than in-house facilities. The data cloud model also may be very interesting for other sciences, e.g. SKA regional centres, as it provides resiliency and long term preservation capabilities. Scale-out is also inherent to the concept, although for practical reasons we might imagine loosely coupled US, European, and Asia-Pacific instances.

The model does require potential changes in funding models: can we easily procure commercial resources at large enough scale to get economy? Do politics allow us to benefit from the largest cloud vendors? At the moment, it is clear that the real cost efficiency and elasticity require the use of spot-market style pricing. What are the implications of that for procurements?

Software

Software is as important as the infrastructure itself. It is essential that this is recognised and supported. We have made significant progress in this area by setting up the HEP Software Foundation (HSF). It must be understood that there is no one-size-fits-all, but rather we need a community wide coordination of available tools covering the full stack from general workflow and data management tools to the application level. Common tools and libraries that can be used to build up the needs of an experiment are required.

The HSF has made steps in this direction, covering many of the aspects of software for the HEP community, with the aim of collecting a set of tools contributed, developed, maintained and evolved by the community. It is also a mechanism for pursuing common R&D efforts in software, and for coordinating things like technology tracking, and developing software tooling for development and performance analysis.

While the HSF is a good framework, this does not remove the need to engage appropriate levels of investment in this area. We (HEP) must recognise that software efficiency and performance will be key to maintaining an affordable infrastructure. We must get ourselves into the position of being able to evolve our codes to make efficient and best use of the evolving computer architectures. This is not a one-off effort but will require sufficient and on-going investment in people and skill development and retention.

Software and infrastructure must be treated together; the separation is a source of inefficiency (and thus cost).

Organization and Steering

If HEP should evolve the existing infrastructure to one for the entire community, for the next decades, it clearly needs the buy-in of all of the major stakeholders, and the community itself. What works best is a lightweight steering mechanism rather than strong "governance". The experience in setting up the HSF is very clear.

A simple steering group should be set up, for example from the heads of IT of the major High Energy and Nuclear Physics laboratories worldwide, as well as computing project leaders from the major projects, facilities, experiments. This group should have a mandate from a body such as the International Committee for Future Accelerators (ICFA), which is key to obtaining recognition in many areas.

The governance should be very lightweight, and through consensus of the community informing the steering group. The role would be to ensure that the physical and software infrastructure evolve in the direction that is suitable for the community and its projects. It should also be a mechanism to obtain or encourage funding and contributions of effort, through direct feedback to the Funding Agencies and labs. There are many mechanisms already in existence through which working groups or effort can be coordinated: HSF, HEPiX, projects, etc.

Finally, such a steering group would be an ideal forum within which to broker community-wide needs, such as licensing, joint procurements, agreements, and policies. For example, HEP-wide agreements with cloud vendors to get scale economies. It should also address political concerns, for example how to evolve the funding models.

Experiment or Facility-Specific Aspects

Since the above describes a long-term evolution for the HEP community, it is clear that experiment facility/project specific processes should be factored out. For example, the Memorandum of Understanding and resource management processes of WLCG would be specific to LHC, while other projects may well have their own such resource policies. Clearly new projects will need to fund the computing resources they will need.

We should note that this paper is not proposing to use the same resources for all experiments, but rather to try and use the same infrastructure, tools, software, and support as far as possible so that new projects are easier to support on existing facilities. Of course, this helps opportunistic use and sharing, but does not impose it.

Conclusion

We have described a potential model for the evolution of HEP computing over the coming years. Many of the ideas here are building on things that already exist or that have been discussed in various forums. Importantly, we are capable of actually building prototypes of this now. We already have active storage federations as well as many of the fundamental tools we will need. HEP should organise itself urgently to build some scale prototypes and demonstrate its leadership in preparation for Exabyte-scale data facilities, and offer the experience to upcoming scientific endeavours. HEP should show itself to be interested in engaging with other scientific communities, and that we plan to invest in the most cost-effective solutions.

---

[i] X.509:Public Key Infrastructure: https://tools.ietf.org/html/rfc5280
[ii] eduGAIN: Interconnection of global identity federations: http://www.geant.org/Services/Trust_indentity_and_security/Pages/eduGAIN.aspx