21/01/21
Jamie.Shiers@cern.ch
https://orcid.org/0000-0001-8404-3750

LHC Computing – the First 3 Decades

Introduction

Computing for the Large Hadron Collider (LHC) at CERN arguably started shortly after the commencement of data taking at the previous machine – LEP[1] – some would argue it was even before. Without specifying an exact date, it was certainly prior to when today's large(st) collaborations, namely ATLAS and CMS, had formed and been approved and before the LHC itself was given the official go-ahead at the 100th meeting of the CERN Council in 1995.

Approximately the first decade was spent doing research and development; the second – from the beginning of the new millennium – on grid exploration and hardening; and the third providing support to LHC data taking, production, analysis and most importantly obtaining results.

Just as with LEP computing, the challenges looked at some (many) points both insurmountable and the costs unaffordable. At the time of writing, we have perhaps only reached a false summit but we are now joined by other disciplines facing similar issues. Thus, we can look forward to the future – both of the LHC and any eventual "successor" machines, with cautious optimism.

Whereas some of this story is told in the book "From the Web to the Grid[2]", the purpose of this document is not to repeat what is written there. Rather, it attempts to capture a few additional elements of the story and to provide pointers to documents that still exist recording the path, on occasions painful, to what can only be considered (great) success – not just in terms of physics but also computing. Thus, it is far from being a complete record of LHC computing but at least part of what entered the consciousness of this particular being.

Whilst many may consider that it is (far) too early to even contemplate a "history" of LHC computing, a significant amount of information has already been lost; more still is hard to find and the non-silicon based memories are fading even faster. Thus, this document maybe considered a "stub", to eventually be completed with material from relevant CHEP presentations, CERN School of Computing lectures, as well as that from documents referenced below.

Just don't leave it too late…

---

[1] The story of computing for LEP is at least partially told in
https://zenodo.org/record/4139583.
[2] See https://cds.cern.ch/record/1141328?ln=en.

21/01/21
Jamie.Shiers@cern.ch
https://orcid.org/0000-0001-8404-3750

Mass Storage

Whereas the issue of storage was always a major concern for LEP, it was close to a nightmare when considering the data volumes and rates foreseen at the Superconducting Super Collider (aka "Clyde") and / or the LHC. Recall that it was originally foreseen that LEP data taking would at least start with 6250 bpi round tapes, with the move to "3480" cartridges with a capacity of 200MB – only a very small increase over that of the open reel volumes – being a major step forward. Given that of the order of 1 million such cartridges would have been needed per LEP experiment, the LHC would need something like 1 billion! Therefore, it was not surprising that other technologies, such as high capacity video devices, were investigated in the early 1990s. Whilst these did good service for some high data volume experiments at that time, such as the NA49 experiment at CERN, fortunately technology moved on significantly by the end of the decade and beyond. Of course, the startup date of either of the then-future hadron colliders was unknown at the time but predicting the hardware of choice 1 – 2 decades into the future remains essentially impossible.

From about the same time, i.e. the early 1990s, HEP began to look at the IEEE Reference Model for Mass Storage Systems that had been developed through and around a series of conferences and workshops that continue to this day: https://www.storageconference.us/history.html.

The model as it stood at the time, together with some systems that claimed to be based on it, can be found in the lecture notes from the 1993 CERN School of Computing. These have been extracted and converted back to machine-readable form at https://zenodo.org/record/4139255.

Some of today's WLCG sites still run systems based on those described nearly three decades ago and / or claim to have been designed according to the reference model.

The Rise and Fall of Object Databases

Whereas, at the time of LEP, the programming language of choice was expected to remain "Fortran[3] forever", the possibility of different languages, and in particular those supporting the object-oriented paradigm, were increasingly explored (and the rest, as they say, is history).

A series of R&D projects were kicked off in the mid-1990s, including MOOSE (RD41), RD44 (aka GEANT-4) and RD45 – the latter to study object persistency,

---

[3] Fortran continues to be developed, with the most recent version of the standard dating from 2018. Apart from support for object orientation that came many years ago, additional features include "co-arrays" that could prove useful for exploiting parallel architectures.

that many took to imply databases, and their potential applicability to the storage of event data in HEP[4].

The proposal that led to the RD45 project, as well as the regular status reports and other documents, can be readily found via the CERN Document Server (using search string "RD45").

The Experiments' Computing Technical Proposals

Towards the end of 1996, ATLAS[5] and CMS[6] published Computing Technical Proposals[7]. Both of these are relatively slim documents of just under 100 pages each. Written at a relatively early stage of the above-mentioned R&D project, they capture very much thinking of the time, as well as a possible model of regional centres that was carried over into later projects.

Whilst reference is made to technology tracking working groups that went under the collective name of PASTA[8] (the acronym did not cover the networking WG), the reports (1996, 1999) cannot be found, although a reference to them, from 2003, is so far still available[9].

It is surely not surprising that many of the explicitly named candidate "solutions" did not make it to the data-taking phase of the LHC – yet another reminder about the challenges of predicting the future.

Can this really be different for any future fcc-*xx* machine?

---

[4] The issue of databases in HEP has featured regular in Conferences on Computing in HEP, is covered in the book "From the Web to the Grid", and can also be found in a slightly shorter version in https://cds.cern.ch/record/1056959?ln=en.

[5] See https://cds.cern.ch/record/322322?ln=en.

[6] See https://cds.cern.ch/record/322321?ln=en.

[7] The ALICE experiment was approved in early 1997 and LHCb in the 2nd half of 1998. All experiments produced more detailed Computing Technical Design Reports "in the 3rd millennium".

[8] Possibly (again) some pieces at pcbunn.cithep.caltech.edu/pubs/PASTA.html but currently not responding.

[9] See https://it-dep-fio-ds.web.cern.ch/Documentation/tapedrive/pasta_tape02.html.

21/01/21
Jamie.Shiers@cern.ch
https://orcid.org/0000-0001-8404-3750

Towards a (Network) Model of Regional Centres

A project that influenced significantly thinking regarding LHC computing models was MONARC – Models of Networked Analysis at Regional Centres for LHC Experiments[10]. Although there was clearly a preference for a network-centric model – it some ways not too dissimilar to that deployed for LEP and other experiments – it was far from clear whether the necessary bandwidth would be either available or affordable. Hence, a "fallback" solution with a very limited number of such ("continental?") sites was considered with "tapes in a suitcase" as distributed mechanism.

But you have to get the data in and out too!

Enter The Grid

Around the time that the MONARC project was wrapping up and coinciding with the dawn of the new millennium, the potential of grid computing greatly enthused many in HEP. This led to a series of projects, such as the EU DataGrid in Europe and others elsewhere, complemented by an LHC-oriented one, initially the LHC Computing Grid that is now known as WLCG – https://wlcg.web.cern.ch/. For this project, a wealth of documents[11] can be found, from design reports, through memoranda of understanding to on-going service reports – as well as plans for future LHC runs and upgrades[12].

Given how well the grid has performed, it is easy to overlook the considerable effort that was required to get to this point. So called "Service Challenges" were first proposed in 2004[13], and these were followed by Common Computing Readiness Challenges as well as Scale Tests of Experiment production – the latter taking place in two phases in 2009[14] – "just in time".

"I think we have it"

… was how the Director General "summarized" the presentations by ATLAS and CMS on their searches for "the" Higgs boson. This continues to be studied in greater detail with ever increasing amounts of data, requiring more and more computational resources. However, the task of "getting the Grid to work", as opposed to keeping it running and extending it, could be considered to be done.

---

[10] See https://cds.cern.ch/record/510694?ln=en, for example.
[11] See https://wlcg-docs.web.cern.ch/. A safe place?
[12] The TDRs, including those of the LHC experiments and their upgrades, can be found via https://cds.cern.ch/collection/LHCC%20Public%20Documents.
[13] See, for example, https://cds.cern.ch/record/853112/files/cer-002536189.pdf.
[14] See, for example, https://cds.cern.ch/record/1322146/files/CR2011_015.pdf.

21/01/21
Jamie.Shiers@cern.ch
https://orcid.org/0000-0001-8404-3750

This was the conclusion of the referees of the EGI InSPIRE project that had a work-package[15] devoted to supporting HEP and other "heavy users" of the grid, as well as senior physicists (well, one at least) from the LHC experiments.

Job done – time to move on!

Whatever Next(?)

Future colliders – both lepton and hadron – are under consideration that will likely be natural successors to LEP and the LHC respectively. Starting possibly half a century after each of their forebears, technology for computing, networking and storage will have all advanced so significantly as to "seem like magic" to those working on the former machines.

The challenges that will be faced will surely be different but if there are lessons that can learned from the past you need to start by recording – and preserving – as much information and "knowledge" as possible. Sounds easy, but it isn't, as the LEP "experience" tells us.

If the past is indeed another country, what is the future?

---

[15] See https://documents.egi.eu/public/ShowDocument?docid=742.