WLCG/HSF Workshop Summary

Benedikt Hegner

EP-SFT Group Meeting 9.4.2018

Based on final session's slides of Graeme Stewart

Goals of the Workshop

- Identify projects to help with the challenges faced in HEP Software and Computing in the next decade
  - HL-LHC, LBNF, Belle II, Juno, ...
- We will be resource constrained - how do we do the best with the resource envelope we have?
- Our strong belief, reinforced by the CWP, is that we must work together to solve these problems
  - Not enough development effort in the experiments (a constant refrain)
  - Very hard challenges coming up
    - Huge increases in data taken by the detectors
    - Ambitious physics program to support
    - Increasingly challenging hardware environment to run on
  - Working with partner organisations too - WLCG evidently high on that list
- Importance of these projects is measured in their impact on these challenges

Workshop Organization

- Mix of plenary and (technical) parallel sessions
  - Covering most of the topics of the Community White Paper

Topics covered:

- Technology Watch
- HEP Use Cases
- Simulation
- Analysis Facilities and Use Cases
- Common Data Management and Data Lakes
- Frameworks and Infrastructure
- Training & Careers
- Workload Management
- Programming for Concurrency & Co-Processors
- Performance and Cost Modeling
- Visualization
- Software Development
- Security

https://indico.cern.ch/event/658060/

L. Sexton-Kennedy - Opening

Highlights - Participation!

In the end we had **211 participants!**

About a quarter from CERN

Highlights - very attentive and active audience!

Highlights - very productive discussions!

To get an idea what actually happened during the conference, have a look at the live notes:

https://docs.google.com/document/d/1QSkvwRK_2HENuxYXcs9Op1dTUK824KddQ1Tfan-P0WU/

55 pages of documentation of discussions and outcomes

Has not been fully cleaned up yet, but parts like

- Software Development
- Simulation

are in good shape already

Leading to quite some nice results

- Data Management R&D starting
- Training group identified core skills as a prime first target
- Common software libraries seriously discussed (VecCore, TrickTrack, Matriplex?)
- Packaging group moving ahead with real tests
- Software developers focus on performance and optimisation
- Frameworks - take on the challenge of heterogeneity and organise workshop follow ups

There is lots to do - we advanced our thinking of how to do it

Some presentations I really liked I/III

“Dask: distributed computing for scientific python” - Martin Durant

- ➔ a very interesting technology to look at!

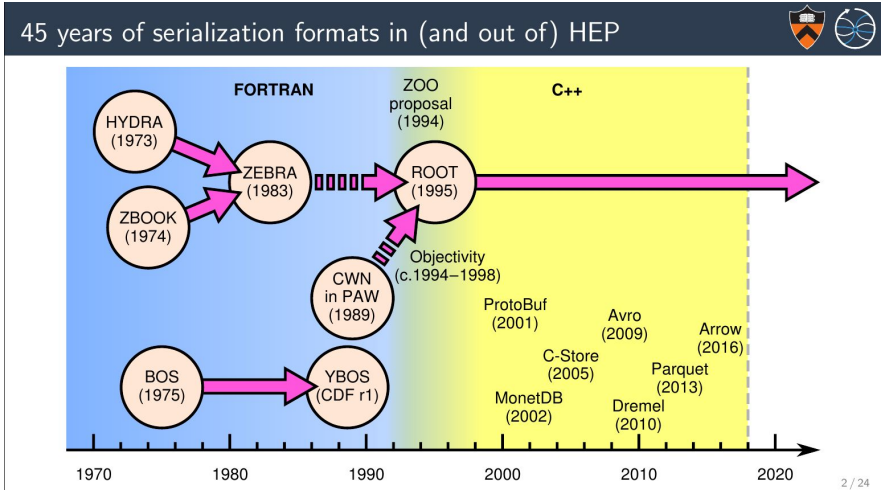
“Programming languages for frameworks: is C++ still the best choice?” - Radu Popescu

- ➔ C++ is a powerful and generic language
- ➔ ERLANG is exactly modeling what we are aiming for - actors and messages
- ➔ Rust is very powerful and strict (no mem leaks!)

Some presentations I really liked II/III

[“Overview of Serialization Technologies”](#) - Jim Pivarski

- Explaining history, features and HEP specifics
- ROOT covers most of our archiving needs, but lacks documentation
- Other solutions are better for interprocess communication



Some presentations I really liked III/III

“Simulation and reconstruction challenges for DUNE / LAr TPCs” - Erica Snider

- ➔ Very nice summary from underlying physics to technical challenges
- ➔ They seem to have a working *community-wide software suite* !

“SKA and Aeneas” - Ana Scaife

- ➔ square kilometre array
- ➔ huge data to deal with

(RAW data for one (!) picture is 936 TB)

- ➔ Lots of ML effort in object classification needed

Ah... another highlight - Naples is great for food!

Graeme Steward - Concluding

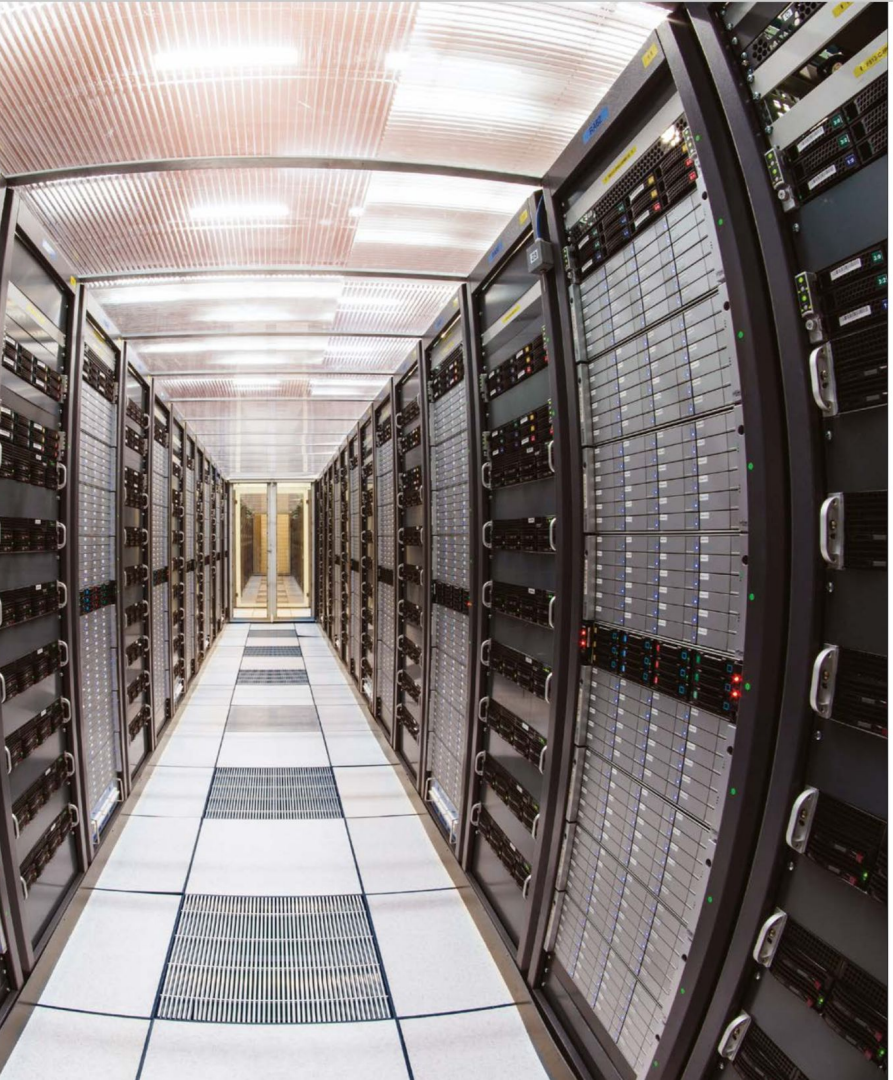
Feedback after the workshop

People found it very useful

A bit too much of presentations vs. discussions

This setup of a joint workshop should be repeated

And I agree with all these statements!



Software and computing

Time to adapt for big data

Radical changes in computing and software are required to ensure the success of the LHC and other high-energy physics experiments into the 2020s, argues a new report.

It would be impossible for anyone to conceive of carrying out a particle-physics experiment today without the use of computers and software. Since the 1960s, high-energy physicists have pioneered the use of computers for data acquisition, simulation and analysis. This hasn't just accelerated progress in the field, but driven computing technology generally – from the development of the World Wide Web at CERN to the massive distributed resources of the Worldwide LHC Computing Grid (WLCG) that supports the LHC experiments. For many years these developments and the increasing complexity of data analysis rode a wave of hardware improvements that saw computers get faster every year. However, those blissful days of relying on Moore's law are now well behind us (see panel overleaf), and this has major ramifications for our field.

The high-luminosity upgrade of the LHC (HL-LHC), due to enter operation in the mid-2020s, will push the frontiers of accelerator and detector technology, bringing enormous challenges to software and computing (*CERN Courier* October 2017 p5). The scale of the HL-LHC data challenge is staggering: the machine will collect almost 25 times more data than the LHC has produced up to now, and the total LHC dataset (which already stands at almost 1 exabyte) will grow many times larger. If the LHC's ATLAS and CMS experiments project their current computing models to Run 4 of the LHC in 2026, the CPU and disk space required will jump by between a factor of 20 to 40 (figures 1 and 2).

Even with optimistic projections of technological improvements there would be a huge shortfall in computing resources. The WLCG hardware budget is already around 100 million Swiss francs per year and, given the changing nature of computing hardware and slowing technological gains, it is out of the question to simply throw

Inside the CERN computer centre in 2017. (Image credit: J Ordan/CERN.)

Community White Paper article published in April
CERN Courier

(Remember, you can still sign the CWP for a little while more)

The HSF Role

- Improve communication in the community
  - We can't work together if we don't know what's going on
  - We have the hsf-forum mailing list - do please use it
    - Along with other specialist lists, e.g. the technical forum list
- Description and inventory of community activities and projects
  - **You** can put this onto the HSF website
  - New working groups are really welcome
    - HSF can help marshal interested people in the community
- Organise Community Reviews
  - Great way to get feedback from a dedicated group (GeantV, Analysis Ecosystem)
- Help improve education, training and recognition of software
  - Publish your software, cite the software of others (DOIs)
- New projects should be building cooperation into their core
  - Visit, discuss, **collaborate** - funding agencies really want this
  - HSF can help to advise on how to do this

We made good progress in identifying some of these this week, but no means did we exhaust the possibilities - we need to follow up on this

The HSF moving forward

- Links between software and computing communities are very important
  - So this workshop was a real step forwards
  - We have to solve both software and computing challenges - isolation will not work
  - The joint workshop was a real step for forward - should be repeated!
- The HSF: It's you!
  - Please contribute to the organisation
  - We can help find solutions and new answers to questions
  - Don't be afraid to constructively criticise
    - But better to fix and improve things
  - HSF work is successful and provides a natural way for working together
  - PyHEP and an opportunity for other WG/BOF meetings before CHEP
- Thanks very much for coming