Future Analysis Software Input for the EPPSU

Danilo Piparo[1]

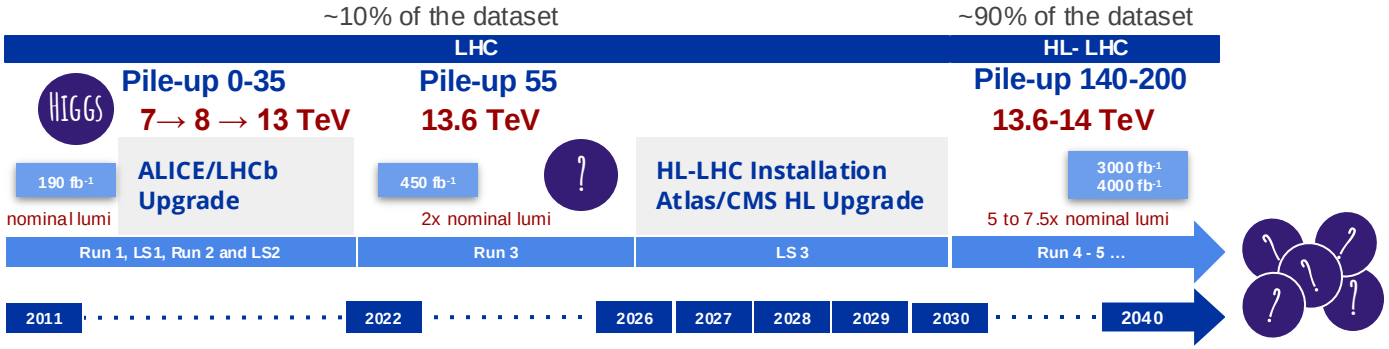
16-12-2024

[1] CERN EP-SFT

Provide a unified software package for the storage, processing, visualisation and analysis of scientific data that is reliable, performant, supported and sustainable, that is easy to use and obtain, and that minimises computing resources and scientists' time needed to achieve results.

The success of experiments and all ROOT users at large is our priority

The Need of Strategic Thinking

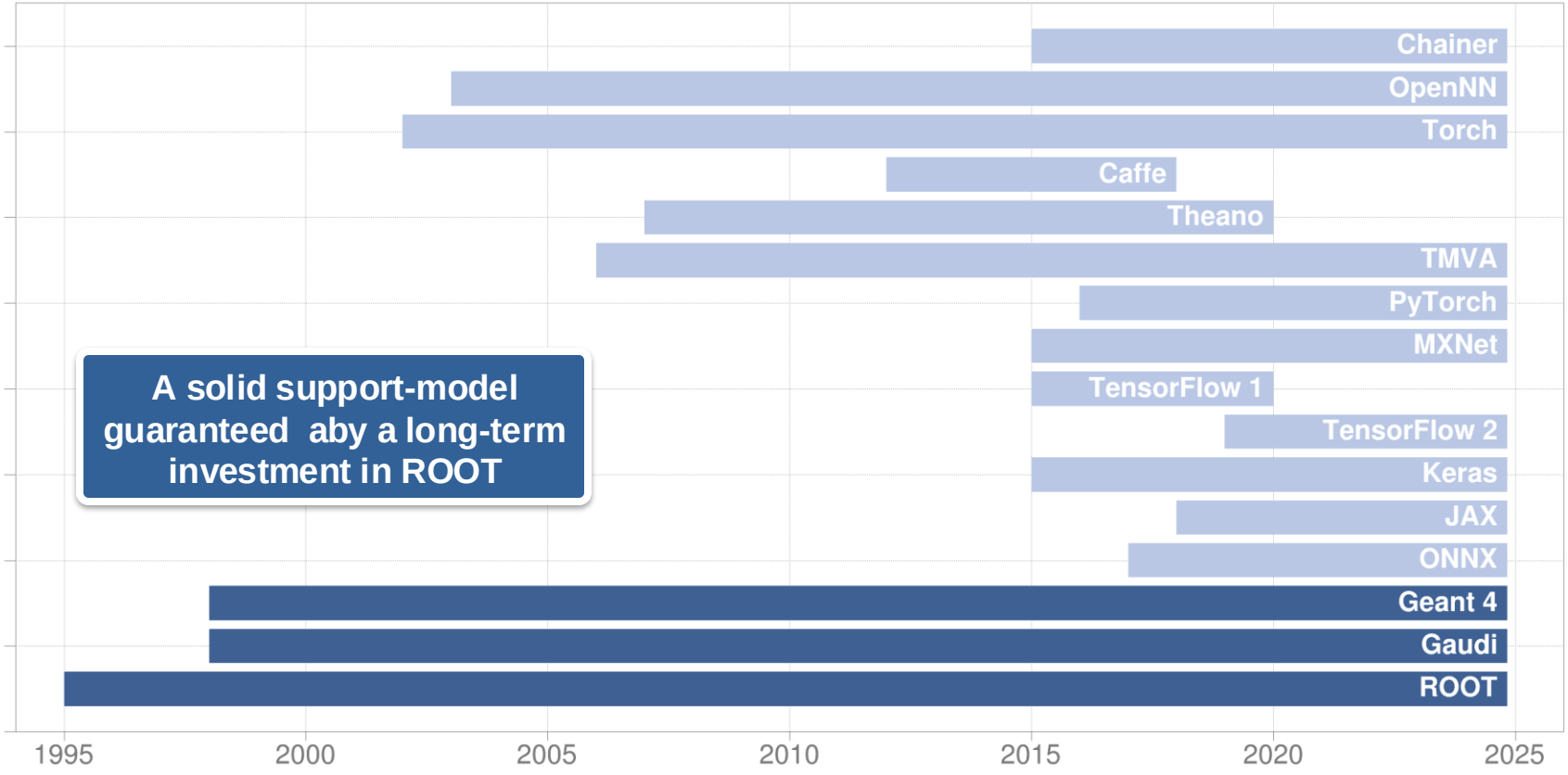


We often discuss future needs of future (analysis) scientific and HEP software because:

- **ROOT is part of the HEP C++ and Python software ecosystem, including analysis**
- The HL-LHC program alone might well produce about 30 EB of data in ROOT format
- Other future experiments might also benefit from this format

The full exploitation of the physics potential of present and future accelerators also passes through ROOT, including its analysis interface.

HEP Common Software Support Timeline



A solid support-model guaranteed aby a long-term investment in ROOT

Plot inspired by M. Mazurek

Processing Large Datasets in a Reasonable Time

- **Needs also dictated by the evolving computing infrastructure**
  - Compute, storage and network – real hardware, with its strong features to exploit and limitations, to avoid.
- Keywords for our future software: **flexibility in hardware exploitation and high runtime performance**
- **Compute: Fully leverage individual nodes (high core count, ~1k?), and distributed systems (interactive usage and batch/Grid)**
  - "Statistics accumulation mode": all 3 approaches will be available to us, the proportions are not not yet fully known
  - Parallelism, in all forms, will have to be there
- Storage and network: **hardware aware data access patterns and optimised retrieval of information** useful to preserve these two crucial resources

- **Software lifecycle and technical debt management** will be key
  - Complexity of our software systems will increase, not necessarily effort at disposal for their maintenance
- **Development, maintenance and support of advanced scientific software has to be substantiated** with career opportunities open to experts in the field
- Poor funding for scientific software could affect quality and quantity of scientific results, and might ultimately undermine the full exploitation of the investments done in accelerators, infrastructure and detectors

- Sustainability does not only mean development and support, but **also environmental sustainability**, e.g. enforced by policy by FAs supporting HEP computing (and therefore analysis)
- **Opportunity: environmental sustainability principles embedded in our analysis software**, e.g. through conserving energy (sound interfaces that prevent mistakes? Effective but less CPU-intense compression approaches?)

- Some uncertainty surrounds the future: e.g. languages, hw, infra
- **"Interoperability aware" design and implementations could help us adapting to that landscape**
  - E.g. Interoperability schemes for end stage experiments software with data science and ML frameworks

An approach which also could help:

- Research and look around for the right tool for the job
- Always consider cost-benefit ratios when looking to industry alternatives: are solutions of others always cheaper, more influenceable, supported and better than ours for solving our problems?
  - Also in the long term and considering all risks involved?

R&D(&Integration) to Production and Support

- **Investments R&D for analysis software are vital** to meet the challenges posed to us by hardware trends, infrastructure, and ambition of HEP scientific programs
- The return in terms of cost reduction and increase of quality/quantity of scientific results can be substantial provided that integration in production systems is also accounted in the funding
- The overall improvement gained by a successful software R&D activity is limited by the effort available to integrate the product in production so that it is actually used