

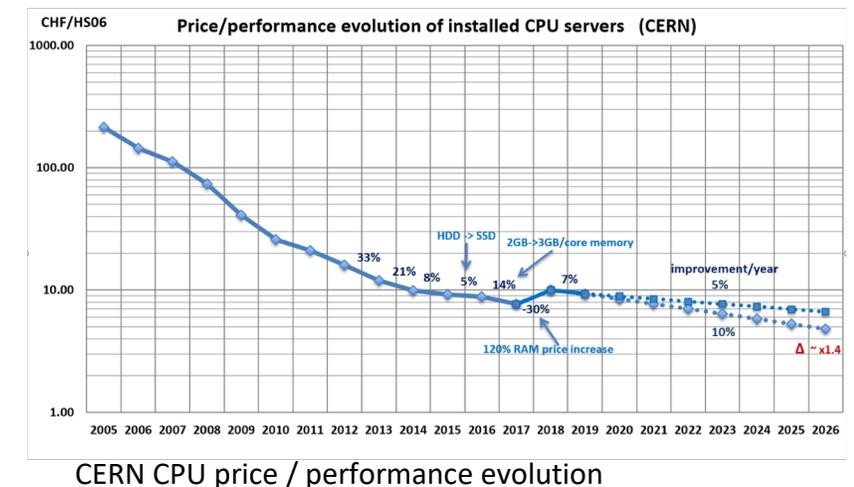
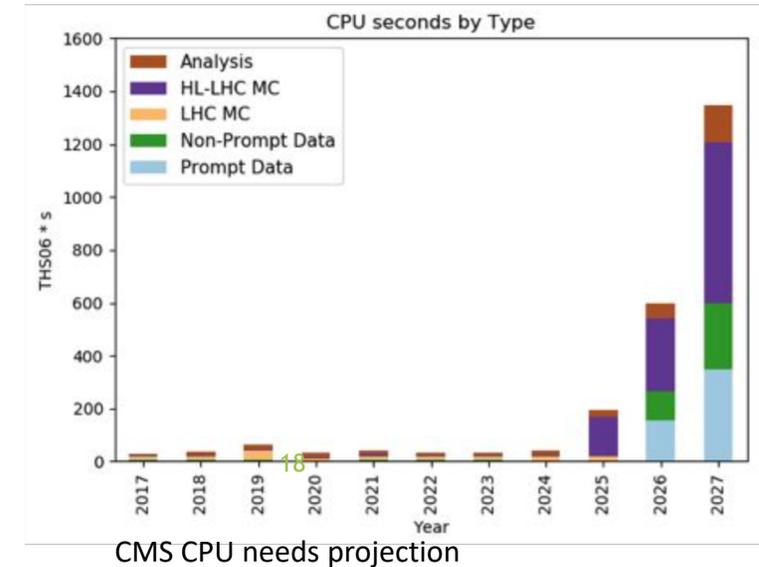
A proposal for a Scientific Software Institute

Scientific Computing Forum
25 February 2019

Ian Bird, Simone Campana, Pere Mato, Stefan Roiser, Markus Schulz, Graeme Stewart
CERN EP/SFT & IT/DI

There are challenges ahead ...

- The future data taking periods, e.g. at the LHC, will pose major challenges on the data processing infrastructure of experiments
 - New programming paradigms and development techniques are needed to exploit advances and changes in hardware
 - high level parallelism, vector instruction sets, non x86 architectures, ...
- At the same time not clear if the “flat cash” scenario, i.e. 20 % increase in performance / year, still holds in the future
 - Need to invest also into software development



The Who and the What

- Propose a strong collaboration among European research institutions & labs, European Universities and physics experiments on software R&D, engineering and sustainability to tackle those future challenges
 - Initiative in addition to already ongoing efforts in different countries
- R&D areas to concentrate on are
 - Application software (trigger, reconstruction, simulation and analysis)
 - Distributed computing (data management and data processing frameworks, data analysis facilities, exploiting HPC facilities)

The Who and the What

- Propose a strong collaboration among European research institutions & labs, European Universities and physics experiments on software R&D, engineering and sustainability to tackle those future challenges
 - Initiative in addition to already ongoing efforts in different countries
- R&D areas to concentrate on are
 - Application software (trigger, reconstruction, simulation and analysis)
 - **?** Distributed computing (data management and data processing frameworks, data analysis facilities, exploiting HPC facilities)

The What (ctd)

- Access, preservation and dissemination of knowledge
 - A lot of expertise available within experiments
 - Many successful initiatives which can benefit the community as a whole



Advanced C++ course
in LHCb

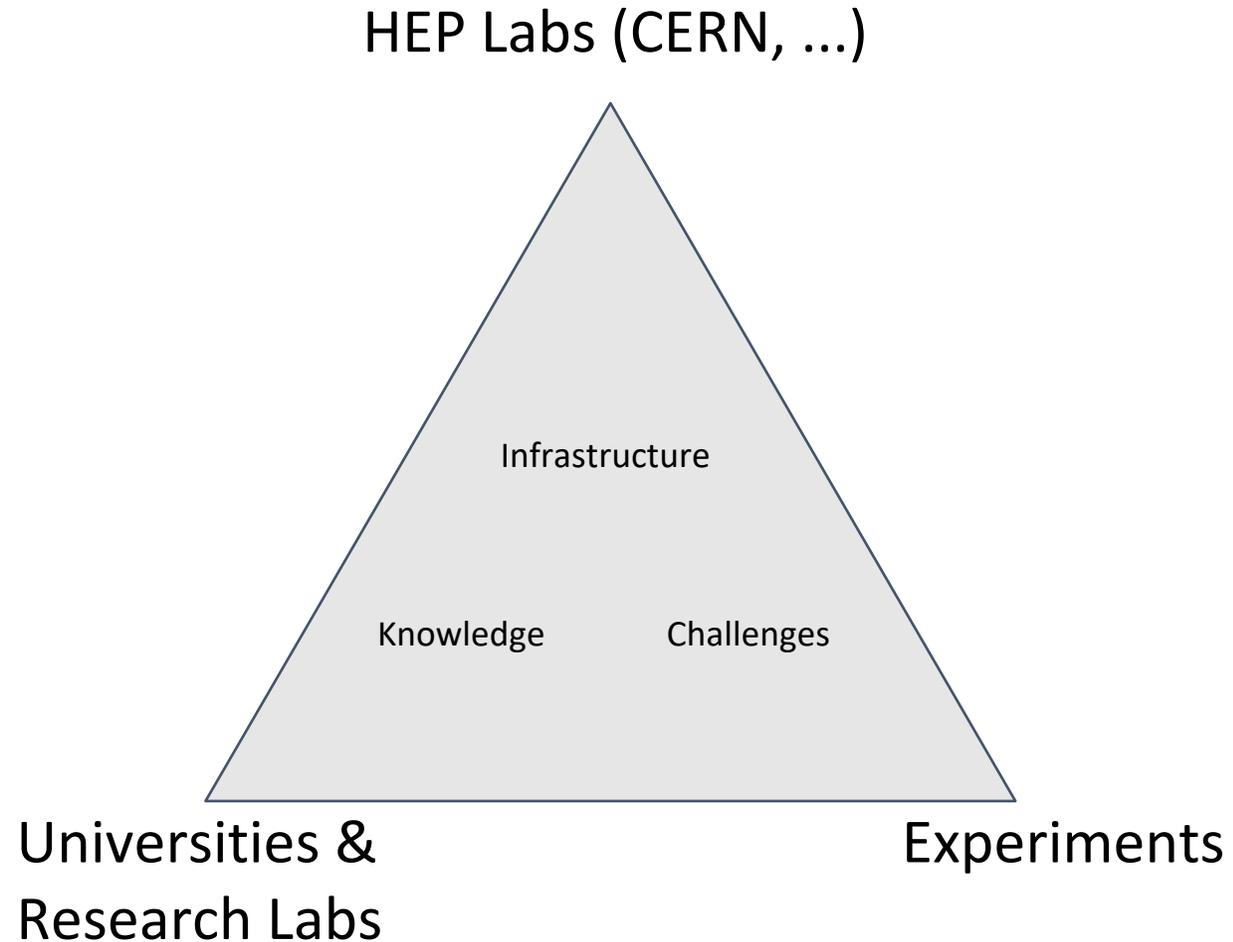


1 week hackathon on upgrade R&D

Mandate of the institute

- Establish a group of researchers to lead and engage in R&D and engineering activities
 - Placement of people at universities, research institutions & labs
- Enable training and qualification for students and young engineers
 - Leverage on already existing activities (UK/"Data Scientist", DE/ERUM, ...)
- Develop a career path for data science engineers in physics
 - Being in contact with possible candidates can facilitate future hirings

Interplay of stakeholders



Benefits for the various parties

- A central hub for R&D and exchange of ideas and people
- Preservation and access of engineering knowledge and experience
- Students receive training and career opportunities
- Teaching opportunities and visibility for lab employees
- Dual career opportunities for computing oriented physicists
- Experiments get solutions for their future challenges
- Additional international collaborations for universities
- Framework for leveraging national funding opportunities
- Recruitment pool for future hirings via working with students

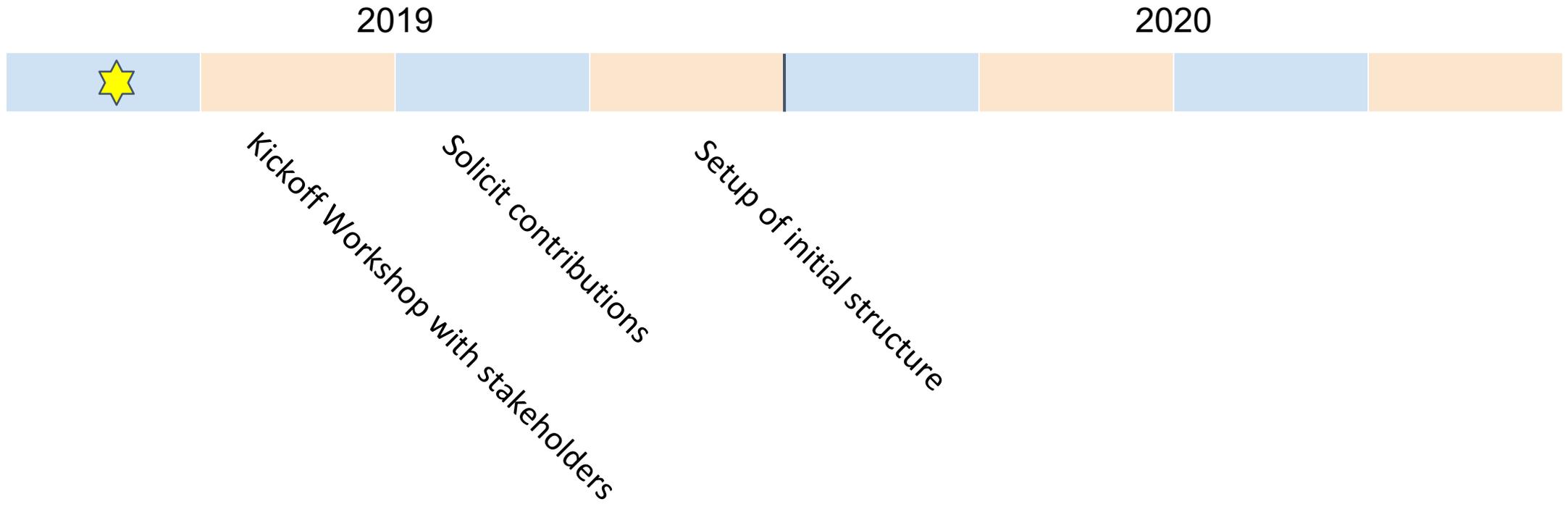
How to organize?

- Shall experiments drive the scope of the institute?
 - Tackle R&D work which is of common interest for multiple beneficiaries
- Few “areas” to concentrate on: application software, training, more?
- Governance?
 - “Institutional board” with representatives of stakeholders, takes major decisions, e.g. appoint head of institute

Staffing Levels?

- Probably a good idea to have a small core team organizing the institute (head + “area leaders”)
 - Spending good part of their time in those roles (30 - 50 % each)
- How much time should people spend when joining the R&D efforts?
Up to 30 %, more?
 - Co-funding of the work of members possible by countries and/or CERN?
 - Apply for EU money to kickstart the activity?
 - Line management for those activities lies within the institute?

Timeline to setup the infrastructure?



Interplay with other parties?

- HSF, WLCG and others are key players to establish discussions among the experiments
 - Can those fora serve as input for common R&D work?
- Collaboration with IRIS-HEP
 - Can this institute act as the European part of the HEP software initiative?
- Openlab, EU projects (BEST4HEP, ...)
 - Strong coordination needed. How to interact?

How to reach out further?

- Engage also applied computer science institutes?
 - Get in contact with software engineers who know about physics, rather than the inverse
 - How can this be done?
- Can we engage with other data intensive sciences?
 - From the start, in a second step?

Many thanks !

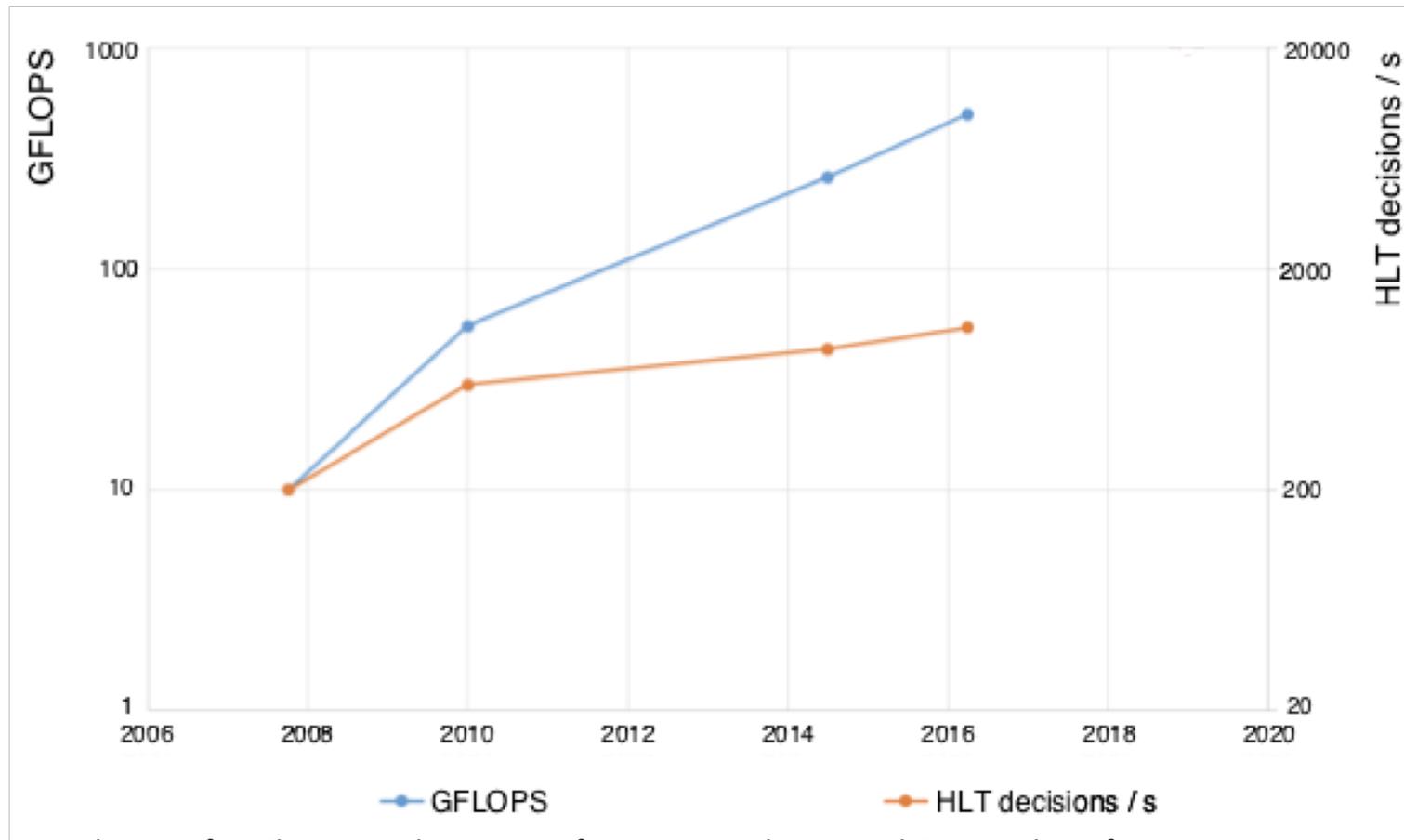
- For a lot of feedback and input from
 - Tommaso Boccali
 - Concezio Bozzi
 - Pete Clarke
 - Davide Costanzo
 - Michel Jouvin
 - Thomas Kuhr
 - Gonzalo Merino
 - Andrea Valassi

Summary

We propose a scientific software institute which aims at

- R&D on future software challenges in physics experiments
- Training and qualification of next generation data science engineers
- Foster collaboration between universities, research labs, experiments and CERN
- Promote and establish a career path for applied software engineering

Backup



Evolution of LHCb HLT application performance vs theoretical CPU peak performance