Challenges in (federated) Computing in Particle Physics

Concezio Bozzi
INFN Sezione di Ferrara
JENA Computing Workshop
Bologna, June 12th 2023

In a nutshell:

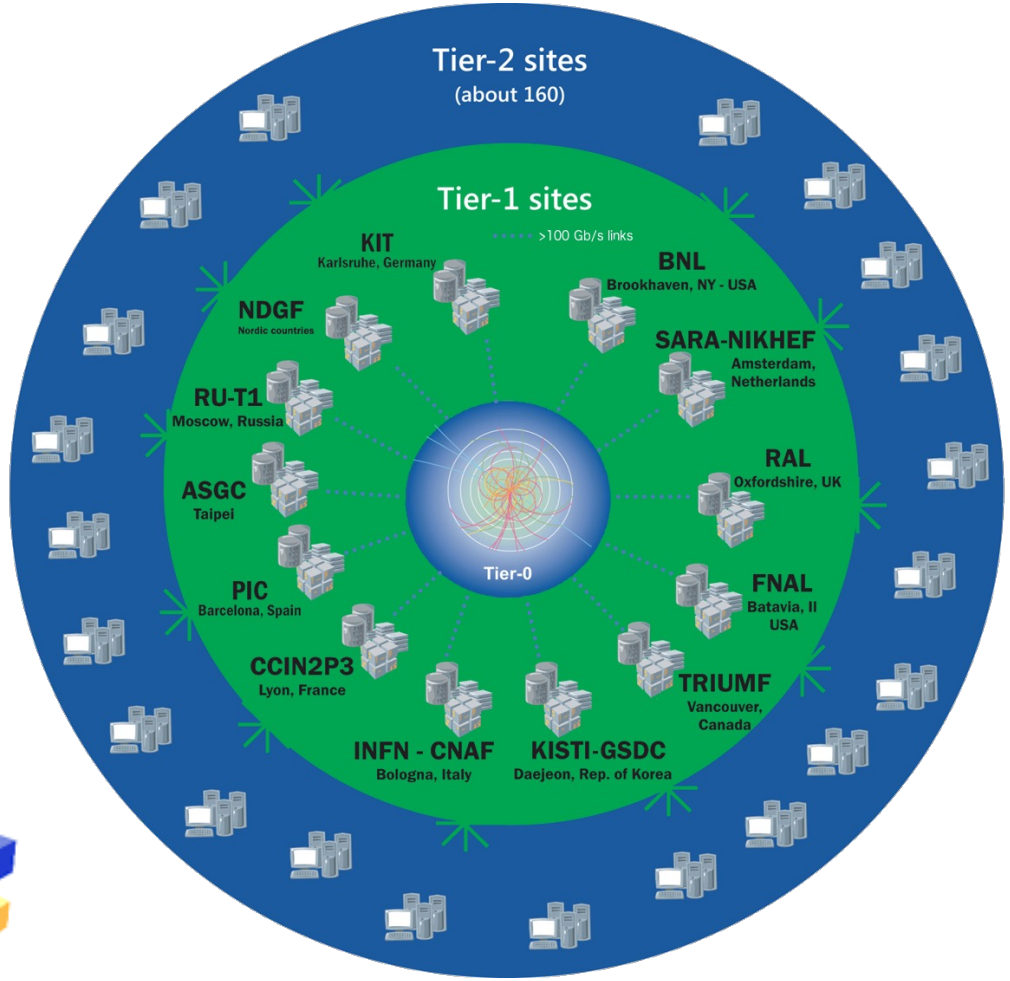
Disclaimer:

This presentation is not a comprehensive review; it is meant to set the scene and give context for further discussions during the workshop.

Acknowledgments are due to a number of colleagues for fruitful discussions and input; all errors and mistakes are mine

HEP computing is federated (since long)

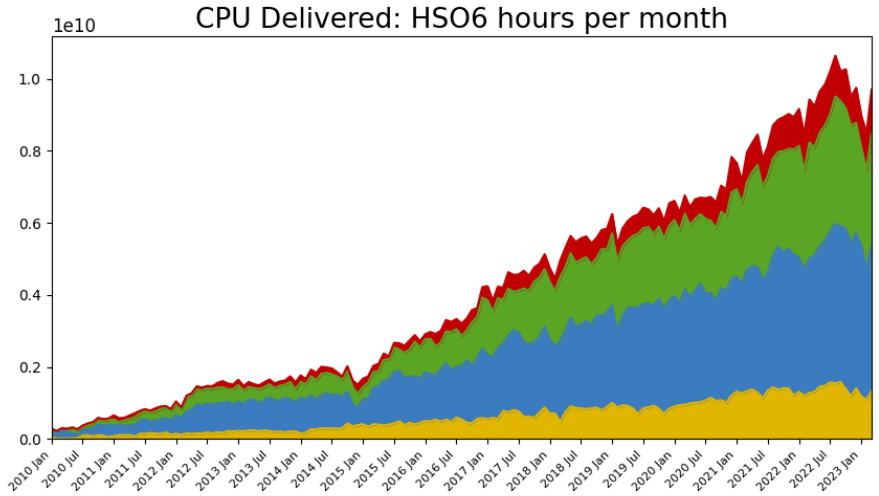
- HEP computing embraced a large-scale distributed model since early 2000s
- Based on grid technologies, federating national and international grid initiatives
- WLCG: an international collaboration to distribute and analyse LHC data
- Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all physicists.
- Belle-II, DUNE, JUNO, and Virgo/LIGO as observers

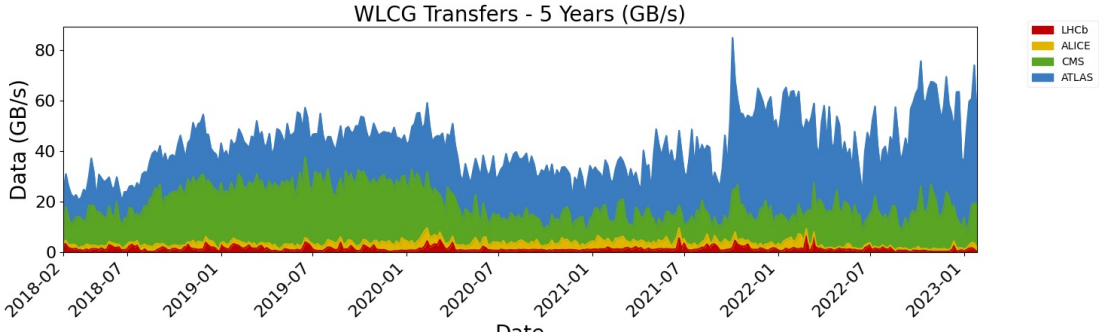


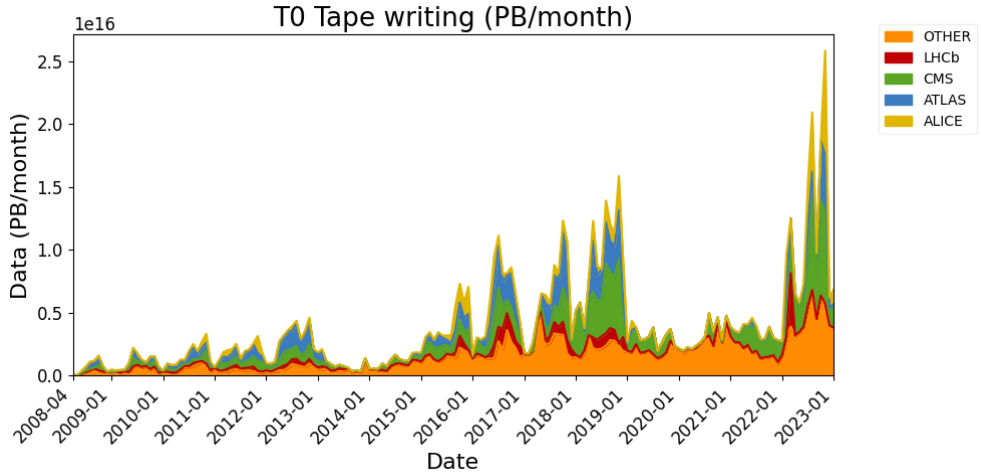
WLCG
Worldwide LHC Computing Grid

HEP computing is federated

- HEP computing embraced a large-scale distributed model since early 2000s
- Based on grid technologies, federating national and international grid initiatives
- WLCG: an international collaboration to distribute and analyse LHC data
- Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all physicists.
- Belle-II, DUNE, JUNO, and Virgo/LIGO as observers

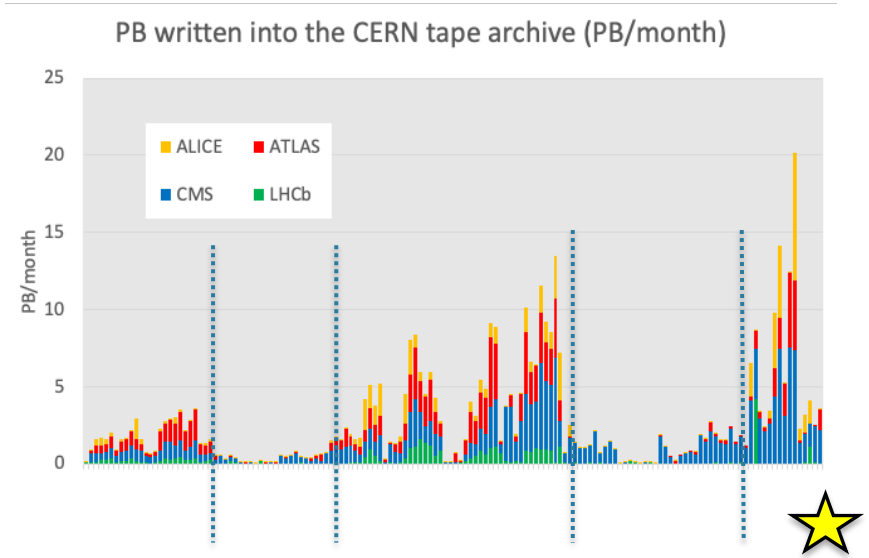




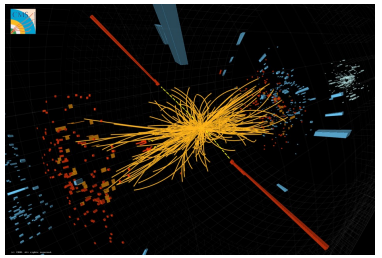


WLCG
Worldwide LHC Computing Grid

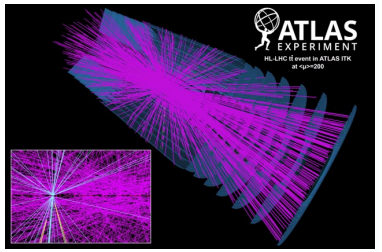
LHC Luminosity and Run conditions

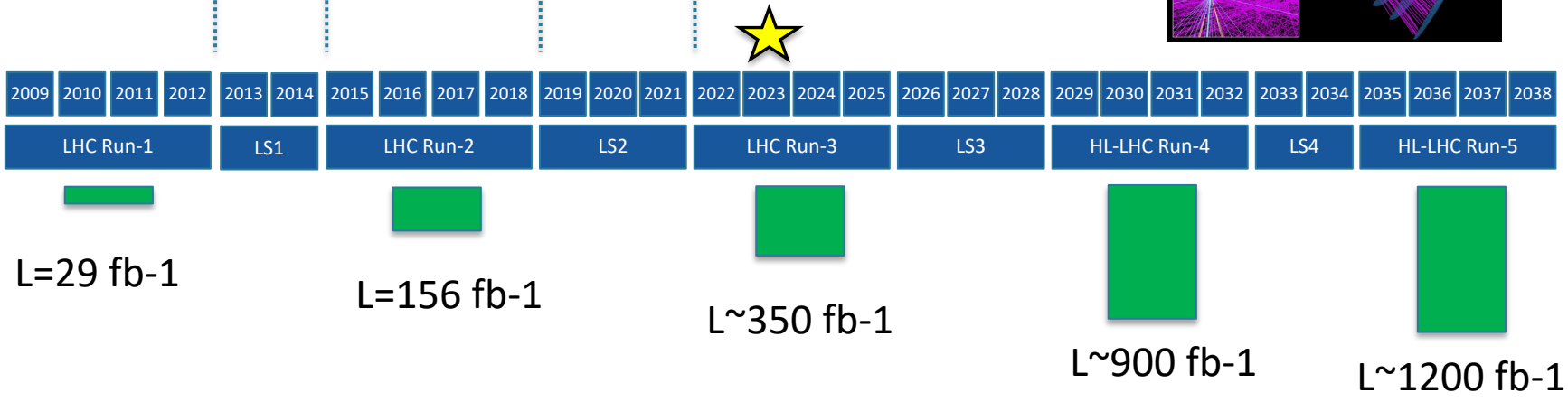


2011 Higgs candidate



HL-LHC simulated event





THE Challenge

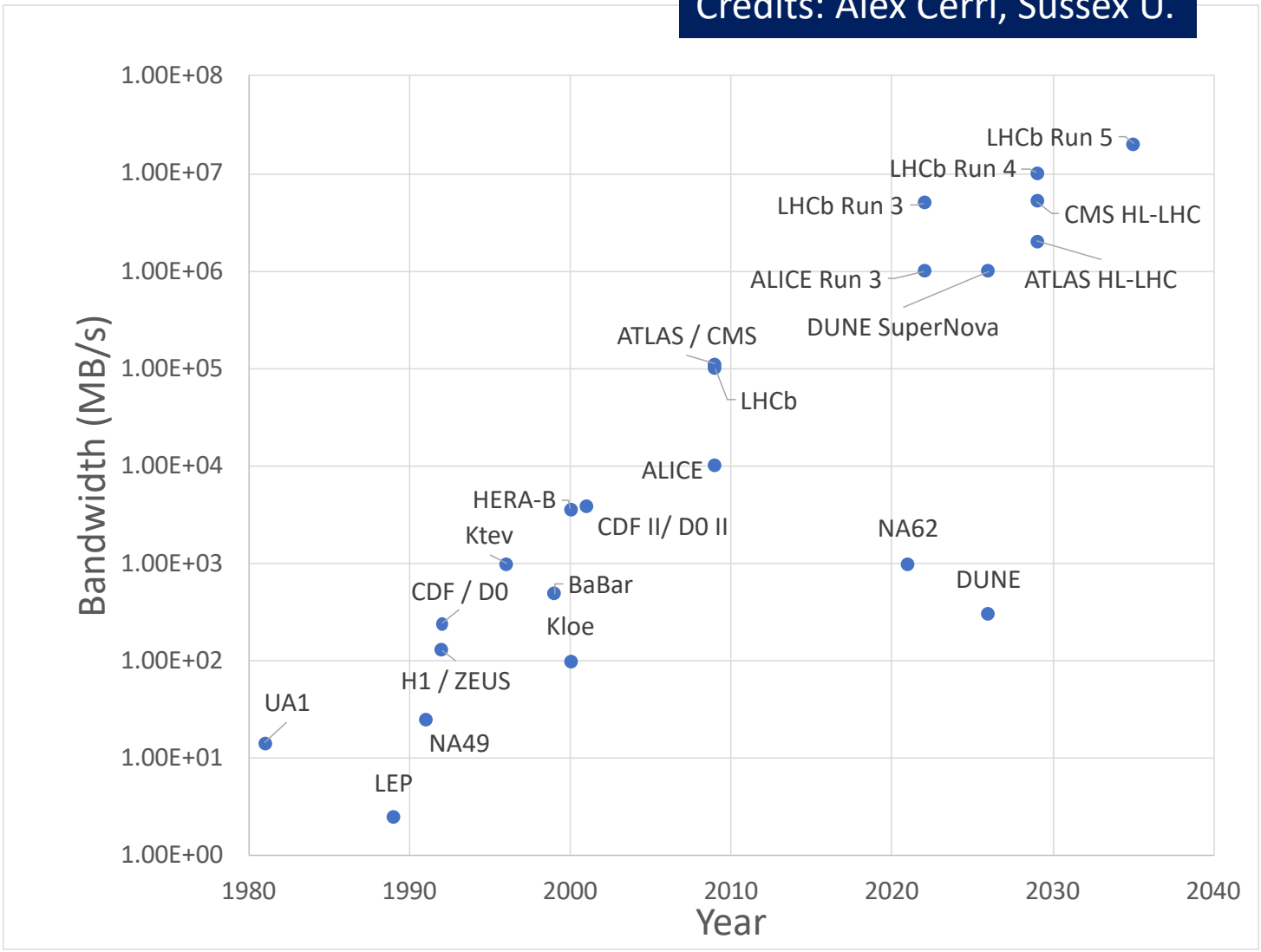
Data throughput from detector back-ends: 1-10TB/s

Typical LHC “live-time”: 5Ms/year

→Data volumes: 5-50EB/year

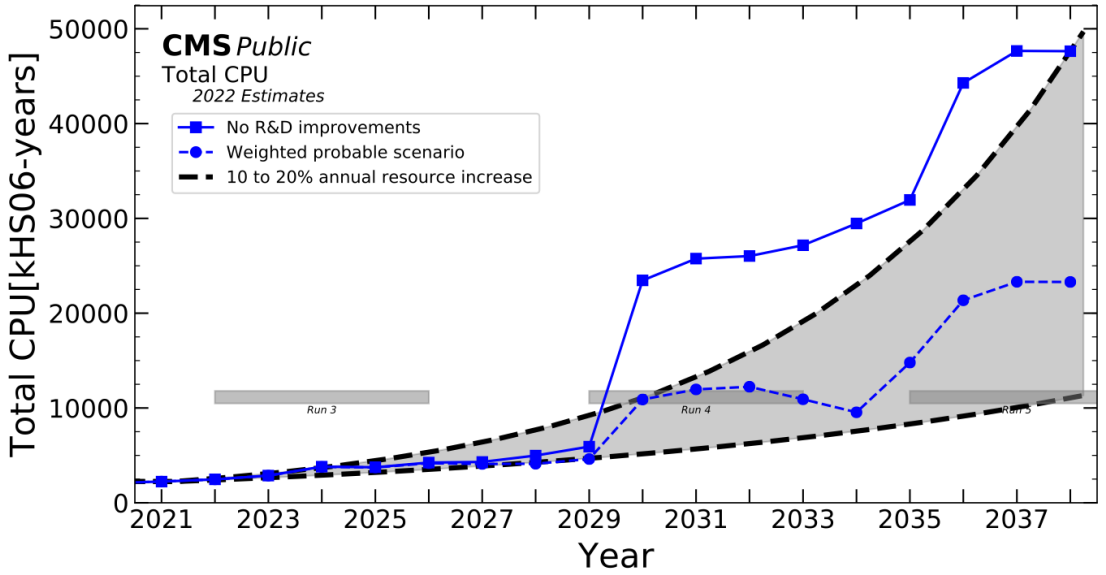
(triggers significantly reduce offline volumes)

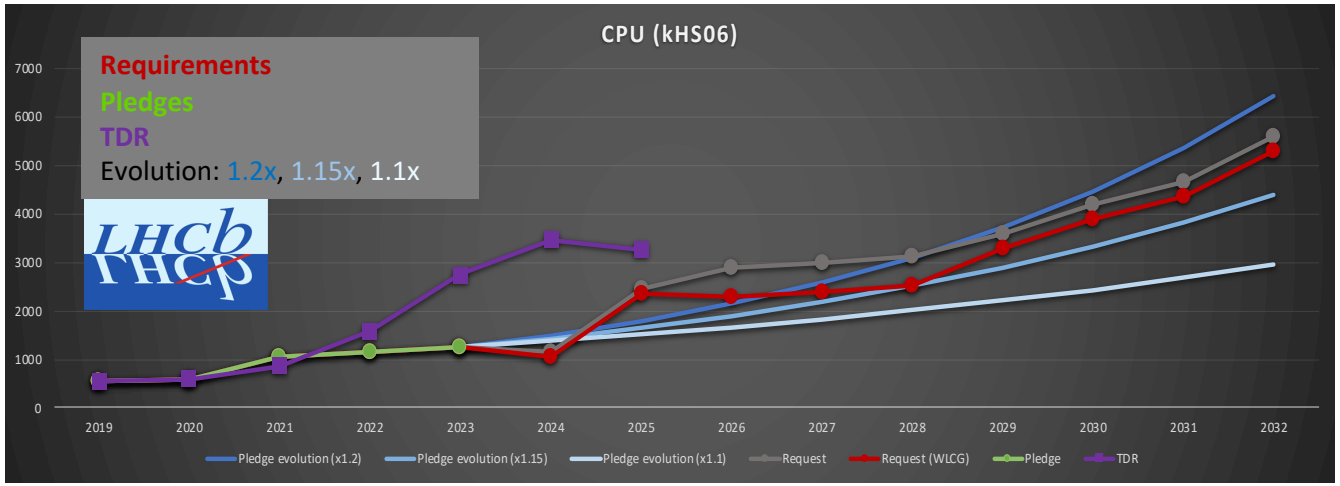
Credits: Alex Cerri, Sussex U.

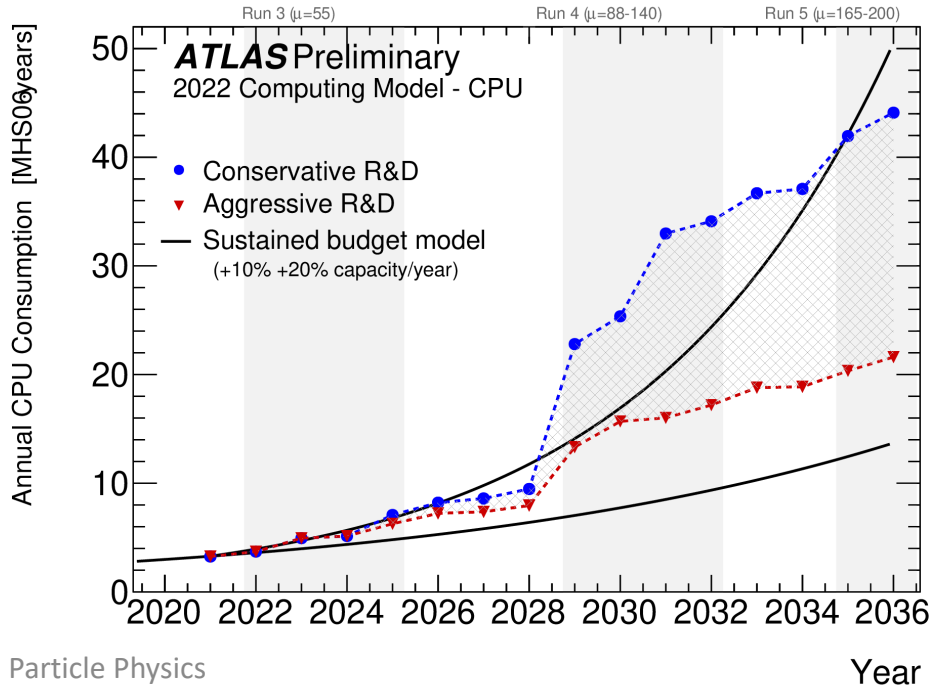


Impact on offline resources

- Intensive software and computing R&D ongoing to match available resources
  - "sustainable budget model": +[10-20]% per year

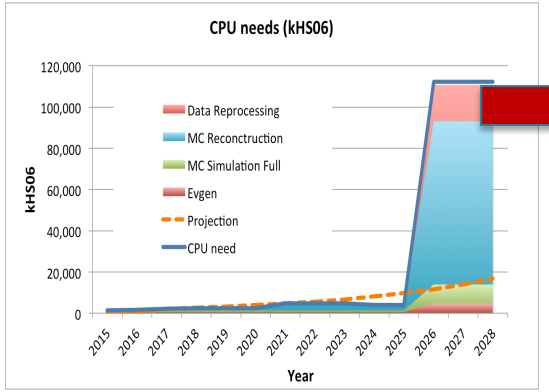




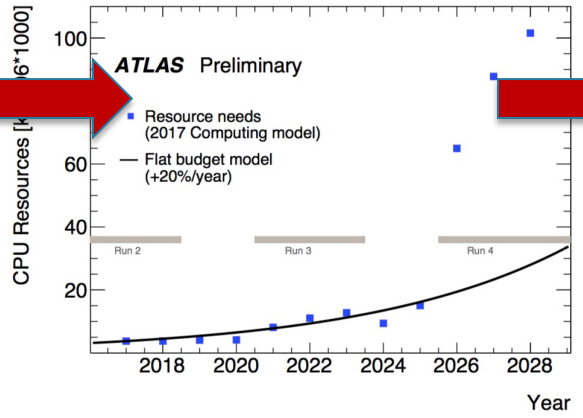


HL-LHC computing resource needs evolution

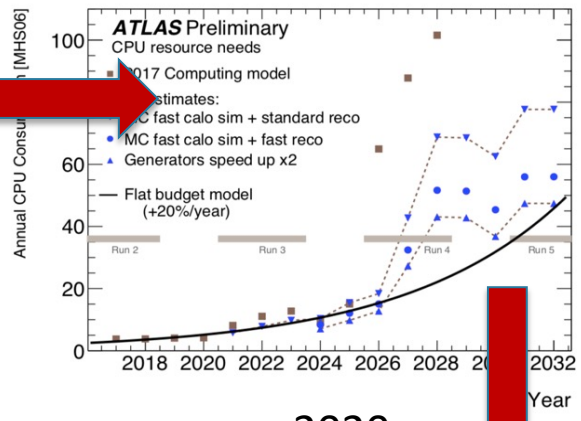
2015



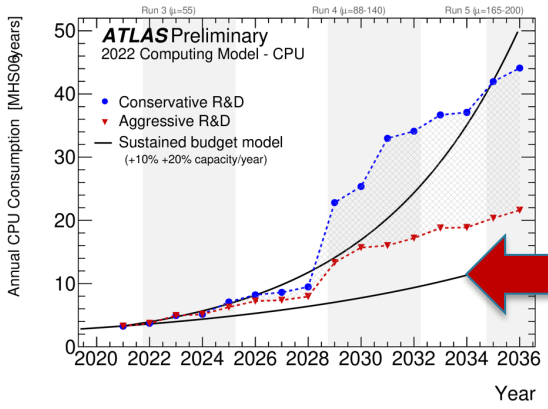
2017



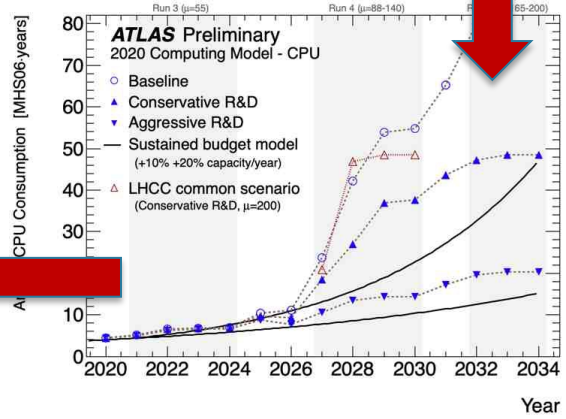
2018



2022



2020

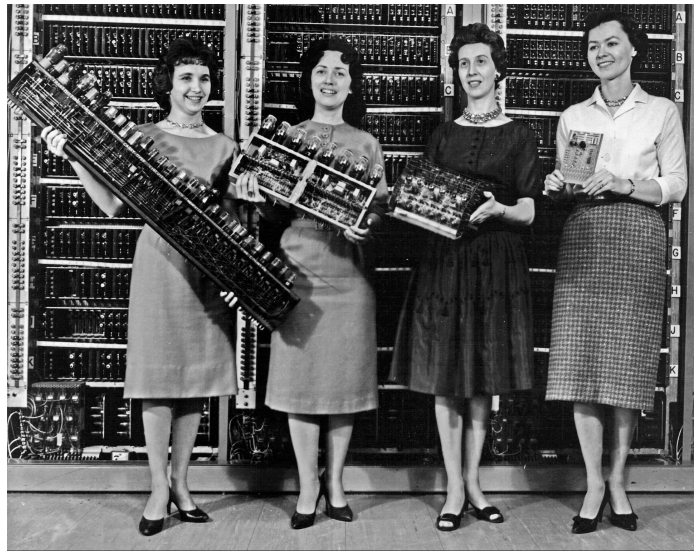


2015 projections: resource needs = 10x more than budget allows

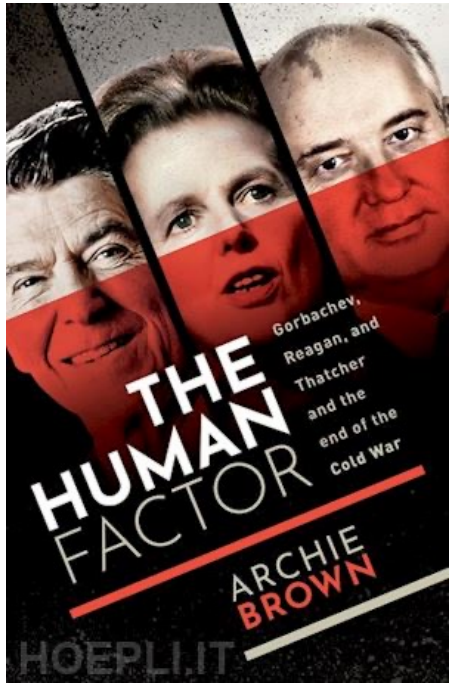
2022 projections: resource needs compatible with budget (optimistic scenario)

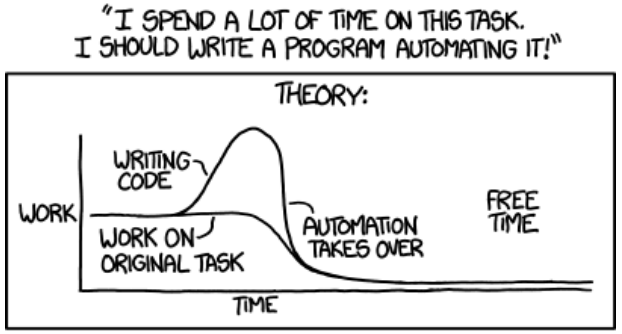
Main themes

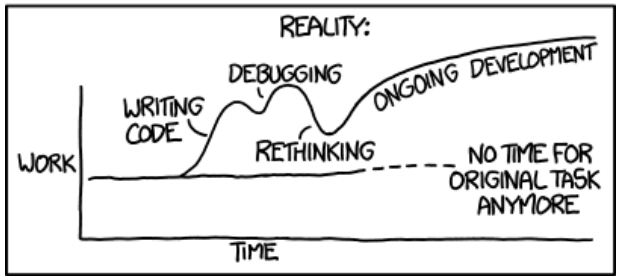
- Evolution of hardware and software
- Evolution of the distributed computing infrastructure
- Training and recognition of human effort



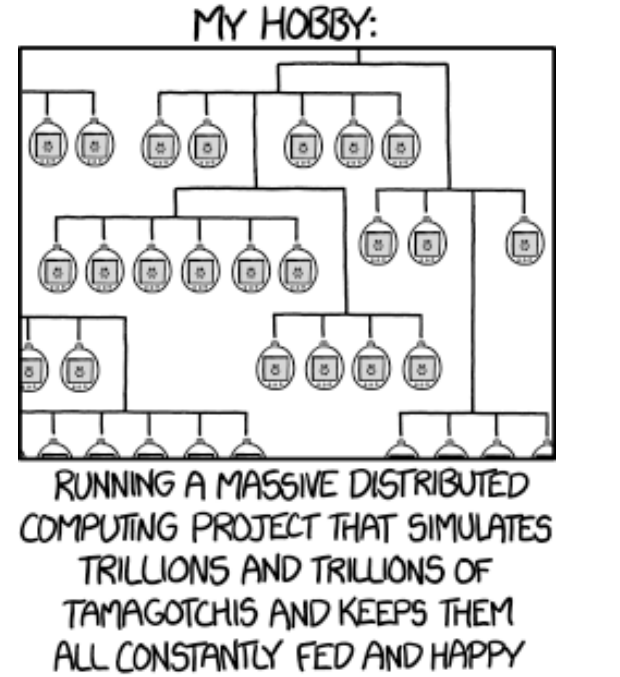
[credits]







[credits]



[credits]

Hardware and software evolution

Paradigm change in computing architecture: break-down of

- Dennard's scaling

(power used by silicon device/volume independent on the number of transistors)

- Moore's Law

transistor density doubles every two years

- Clock speed

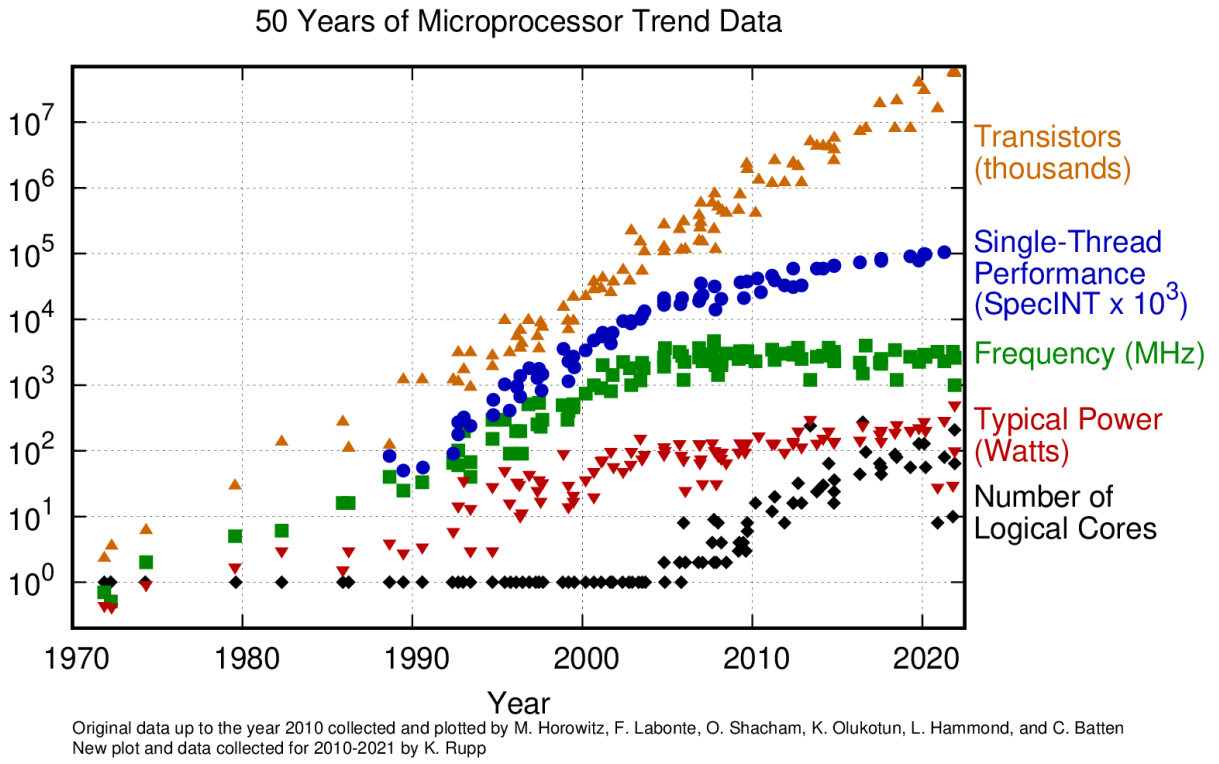
increased 1,000 times in 1970-2000

Evolution towards heterogeneous systems

- Multi-core servers using co-processors (e.g. GPUs) and complex memory configuration

This is a challenge for HEP software

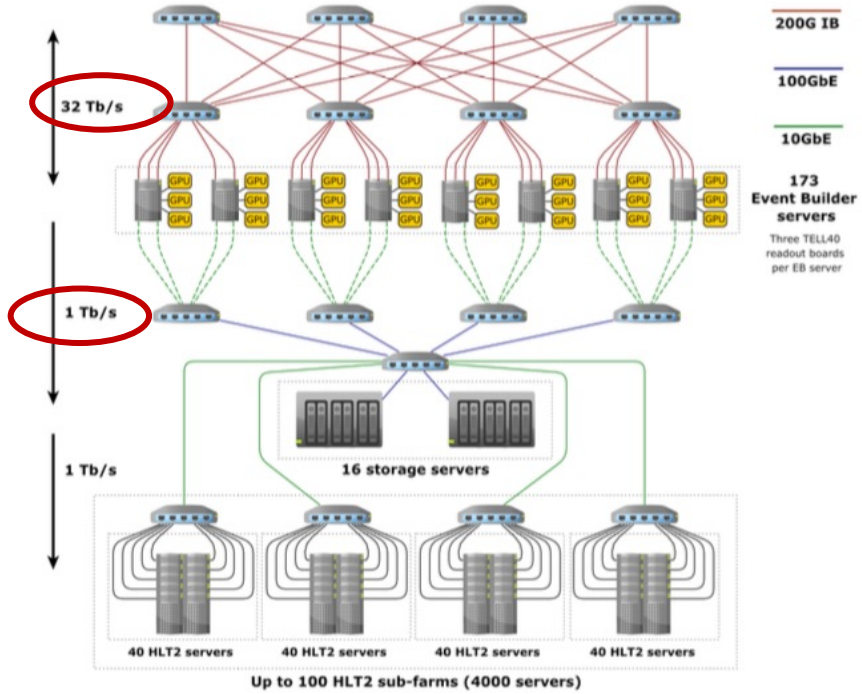
© Karl Rupp

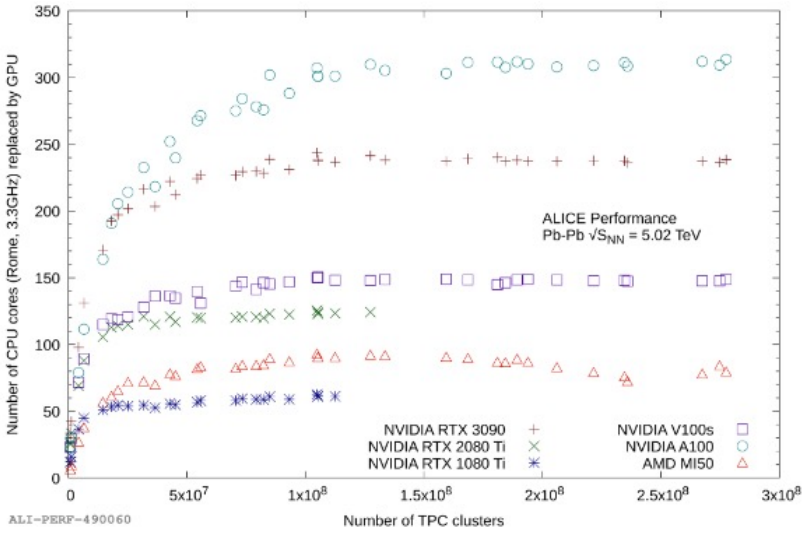


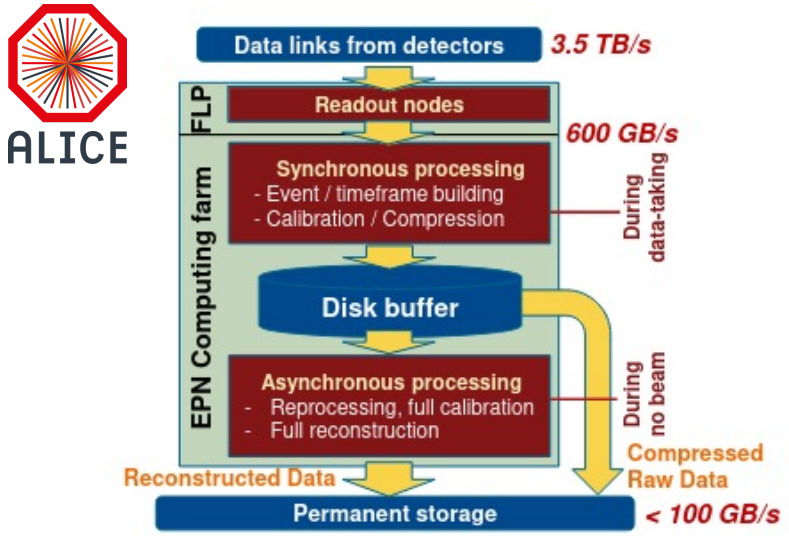
Challenge accepted

Software triggers:

- Moving high-level reconstruction and event selections closer and closer to where data are generated
- ALICE and LHCb implemented full SW trigger in Run3







C. Fitzpatrick, CHEP2023

TD. Rohr, G. Eulisse, CHEP2023

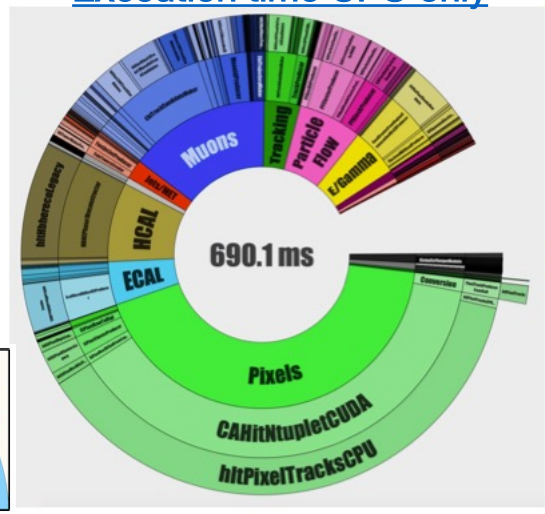
F. Pantaleo, CHEP2023
G. Parida, CHEP2023
ATL-SOFT-PUB-2021-002
A. Krasznahorkay, CHEP2023

Challenge accepted

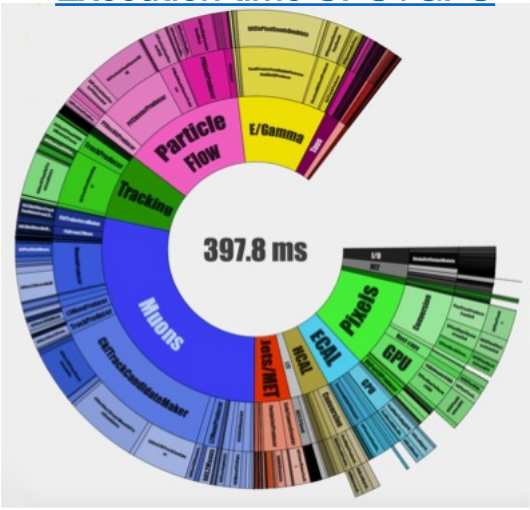
Software triggers:

- Moving high-level reconstruction and event selections closer and closer to where data are generated
- ALICE and LHCb implemented full SW trigger in Run3
- CMS also running an HLT equipped with GPUs since 2021
- ATLAS making excellent progress
- Heterogeneity is key
  - current workhorse: GPUs
  - Working also on other platforms (e.g. FPGAs) and paradigms (ML/AI)

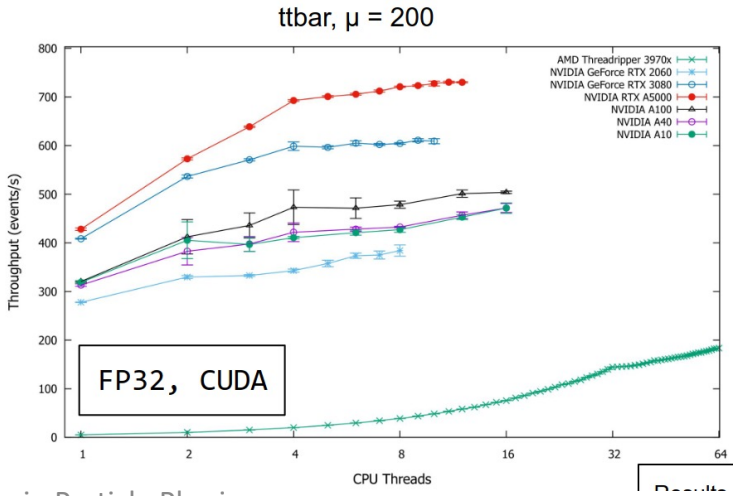
Execution time CPU only



Execution time CPU+GPU

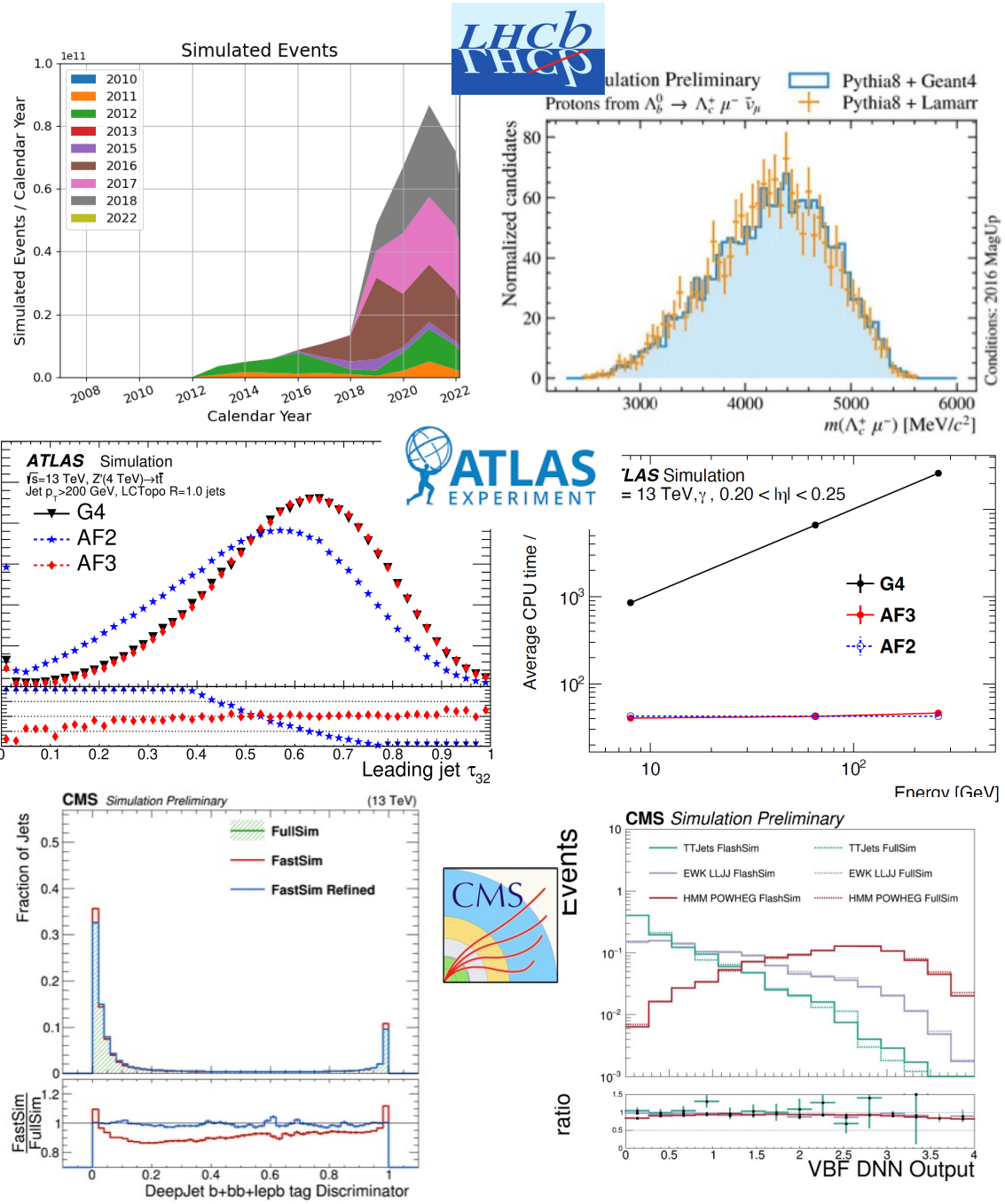


Processing throughput: +70%, performance per Watt: +50%, performance per initial cost: +20%



Simulation

- A large fraction of compute work is and will be spent on simulation
  - ~50%, LHCb: >90%
- GEANT4 detailed simulation is our workhorse
- Exploiting fast simulation techniques
  - Re-use underlying events (LHCb ReDecay)
  - Parameterised simulations (LHCb Lamarr)
  - ML techniques (ATLAS AtlFast3, CMS FastSim and FlashSim)
- Exploiting full GPU simulation
  - AdePT and Celeritas
  - Beware of intrinsical branching in Monte Carlo codes
- Porting fast simulations on heterogeneous resources, exploiting portability models, e.g. ATLAS



Interlude: AI / QC

- Multivariate analysis commonplace in in HEP since 30 years
- ”Modern” ML now making paradigm-shifting contributions
  - Driven by industry, dedicated solutions needed for HEP, e.g.
    - Generative models in simulation, generation, lattice gauge theory
    - Unsupervised classification for anomaly detection (BSM searches)
    - Ultra low-latency inference for control of particle accelerators
  - Software and hardware needs might not be aligned with industry standard – partnership / direct contributions beneficial
- Quantum Computing is also a paradigm shift
  - Currently in the Noisy Intermediate-Scale Quantum (NISQ) era
    - Rapid development of (open-source) software and hardware on several platforms
  - HEP use cases: lattice gauge theory, event generation, data analysis

Heterogeneous resources: HPC / clouds

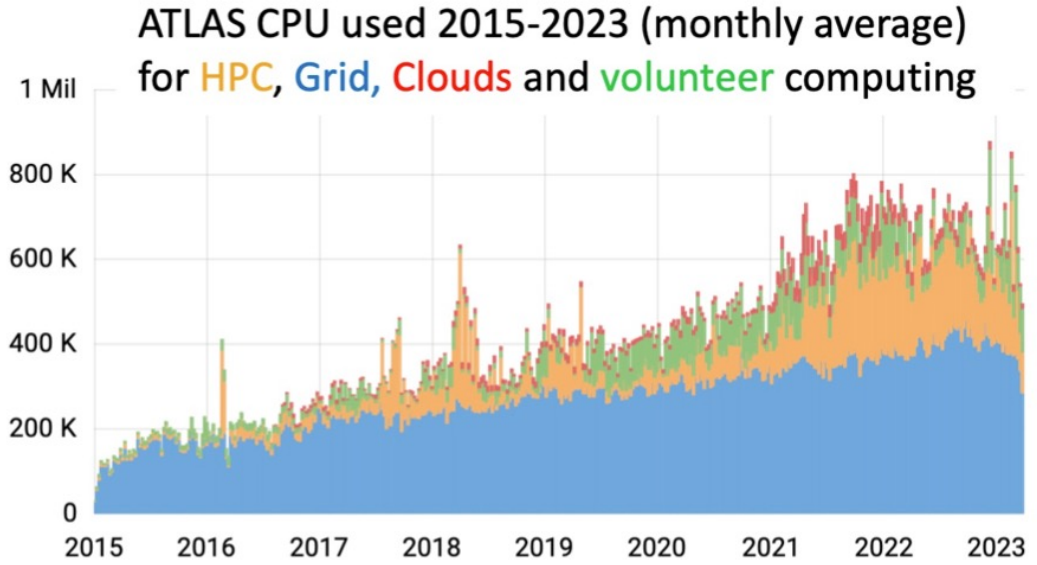
- HPC centers strategic in the agenda of several countries/regions
  - Very heterogeneous in hardware and policies
    - GPU offering is now the default
  - Ever-increasing pool of resources
  - Challenge is that they are not generally suited for data-intensive processing

[credits]

- Clouds are virtually infinite and flexible
  - Cost effectiveness to be demonstrated
  - Challenges: interfaces, vendor locking, networking, procurement, economic model

Heterogeneous resources: HPC / clouds

- Exploiting HPCs / clouds that have policies similar to those of HEP gives excellent results
- In other cases, workload management systems had to be creatively adapted in order to use these resources
- Typical issues:
  - Absence of network connectivity on the worker nodes
  - Absence of standard software repositories (CVMFS)
- Typical workflows: mainly Monte Carlo, but other as well
  - Ingress/egress requirements

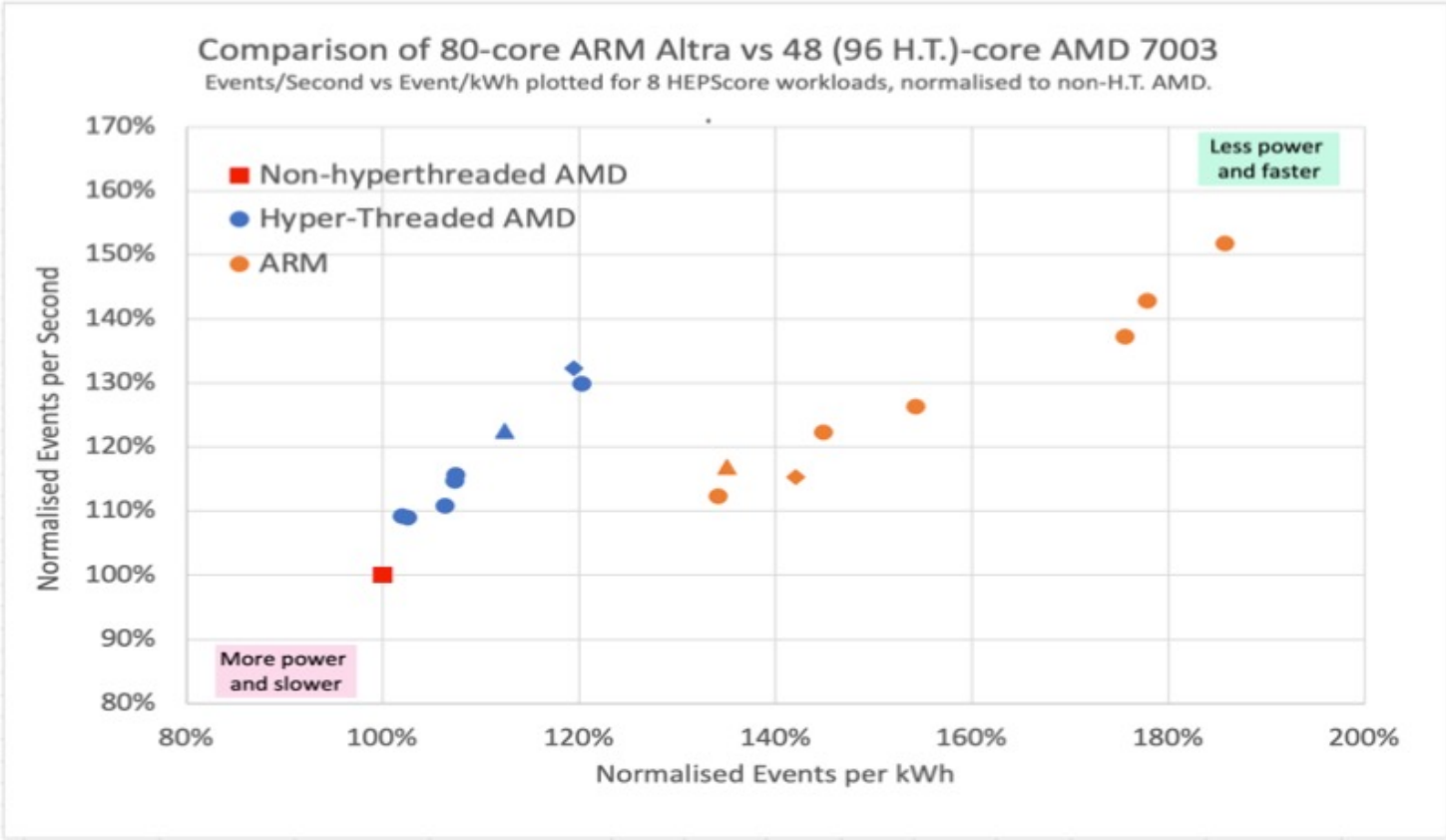


A. Klimentov, CHEP2023

ARM / PPC

D. Britton, S. Campana, B. Panzer (CHEP 2023)

- ARM is the most promising non-X86 CPU architecture, with high “bang per Watt”
  - Motivates effort in porting and validating experiment software for a large number of workflows
  - some sites planning to provide resources
  - “bang per buck” varies in different countries
- Power8/9: available at some HPC but no future prospects



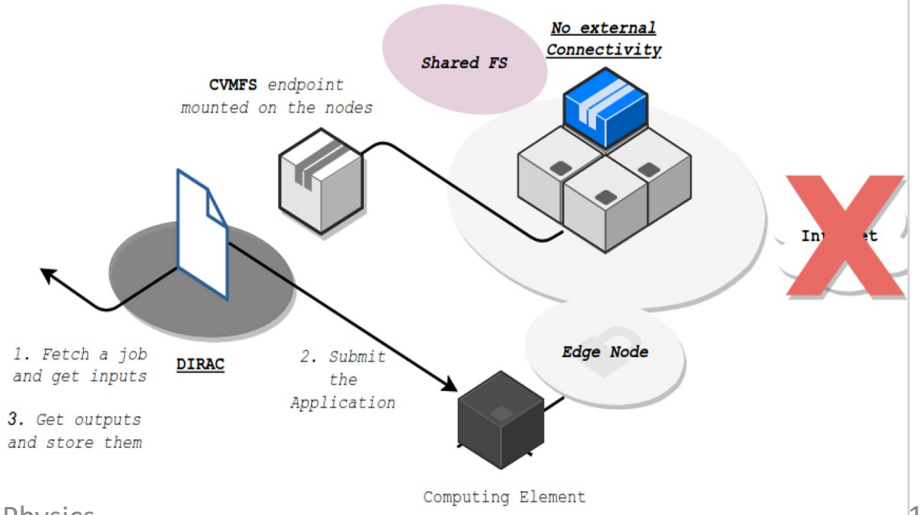
Heterogeneous resources: big challenges!

- "monolithic" architecture (x86 CPU) is no more
  - ARM / PPC (?) / accelerators
  - Rapidly evolving landscape
- Software portability must be dealt with
  - GPUs are becoming dominant source of computing power in HPCs
  - Multiple competing architectures: NVIDIA, AMD, Intel
  - Different programming languages for each architecture
  - Experiments lack human resources to re-code for each architecture
  - CMS early adopter of Alpaka portability layer
    - 1 codebase for GPU/CPU, replacing CUDA with Alpaka soon
- Distributed infrastructure must cope as well
  - Sometimes by "rediscovering" old paradigms

C. Leggett, CHEP2023

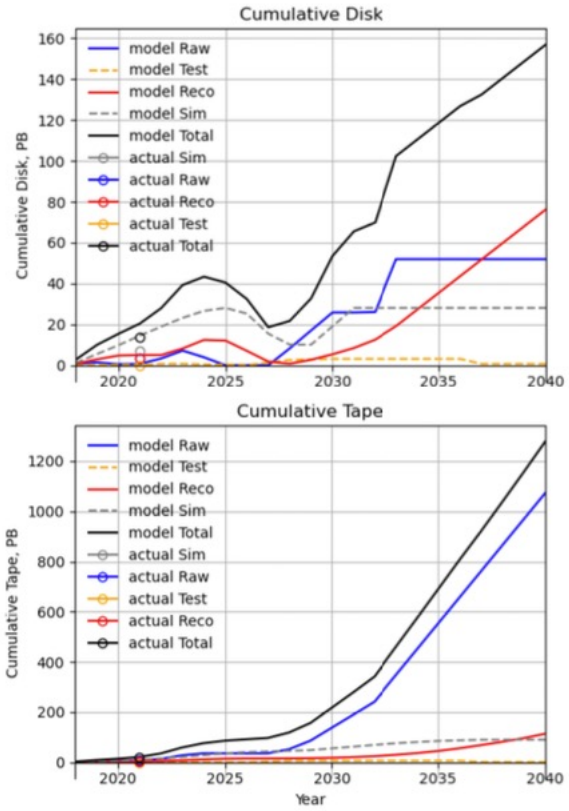
	CUDA	Kokkos	SYCL	HIP	OpenMP	alpaka	std::par
NVIDIA GPU			intel/llvm compute-cpp	hipcc	nvc++ LLVM, Cray GCC, XL		nvc++
AMD GPU			openSYCL intel/llvm	hipcc	AOMP LLVM Cray		
Intel GPU			oneAPI intel/llvm	CHIP-SPV: early prototype	Intel OneAPI compiler	prototype	oneapi::dpl
x86 CPU			oneAPI intel/llvm computecpp	via HIP-CPU Runtime	nvc++ LLVM, CCE, GCC, XL		
FPGA				via Xilinx Runtime	prototype compilers (OpenArc, Intel, etc.)	prototype via SYCL	

A. Boyer, CHEP2023

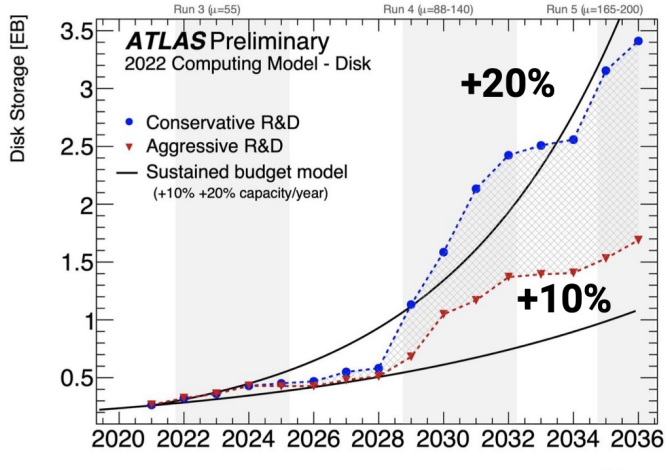


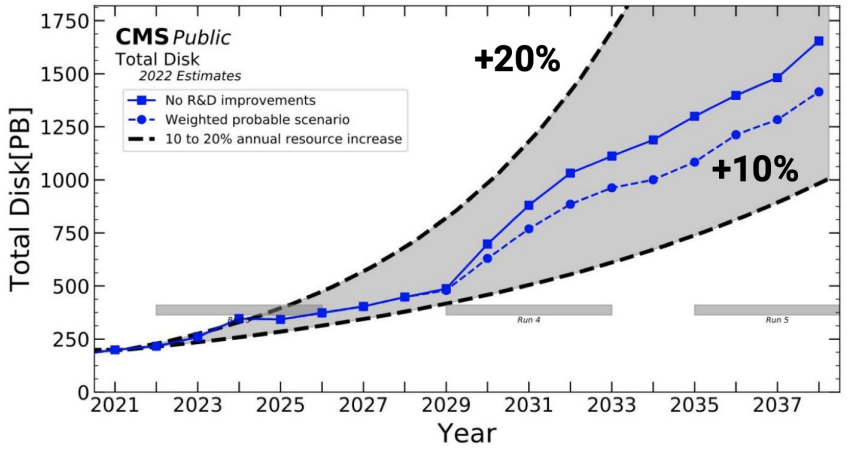
Storage and data management

- Storage is perhaps the main challenge
  - No opportunistic storage
  - Data is the main asset of HEP, but storage needs are hard to fulfill
  - Storage services are hard to operate
- …but also a great opportunity
  - Decades of experience in deploying and operating large storage solutions and managing large data volumes
- The objective: build a common HEP data cloud



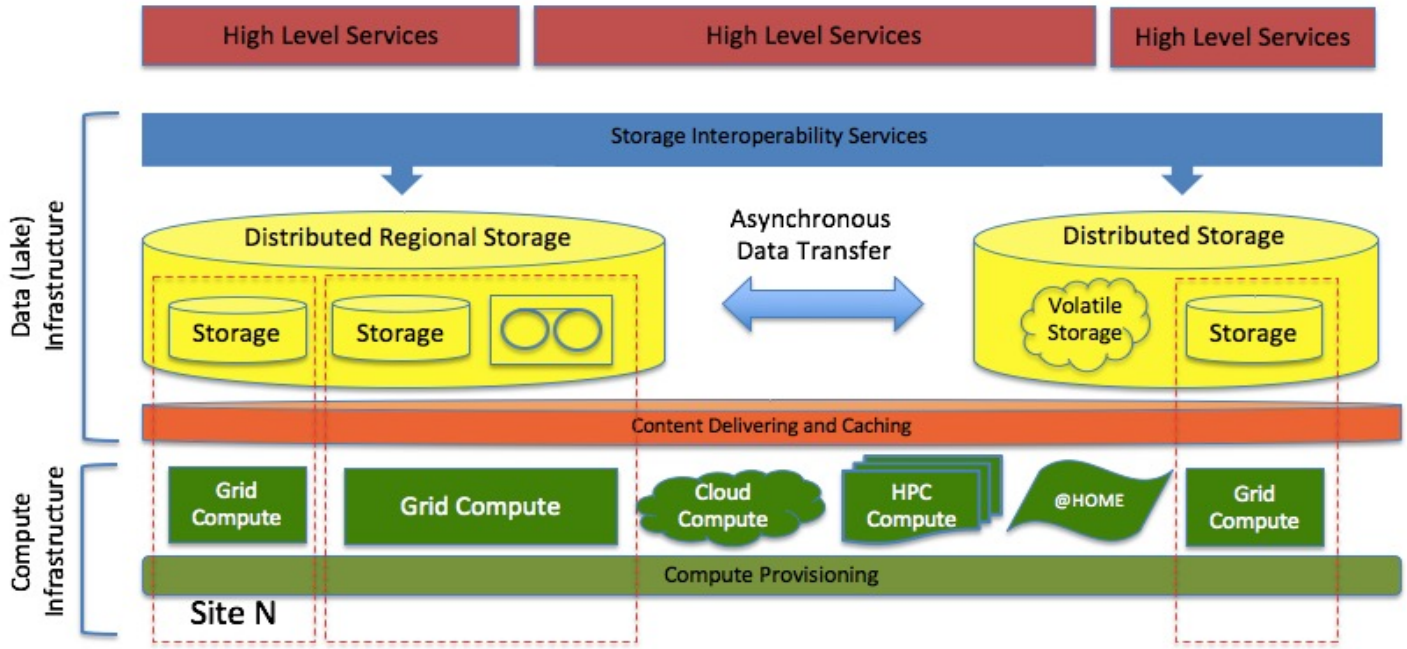
DUNE Computing CDR





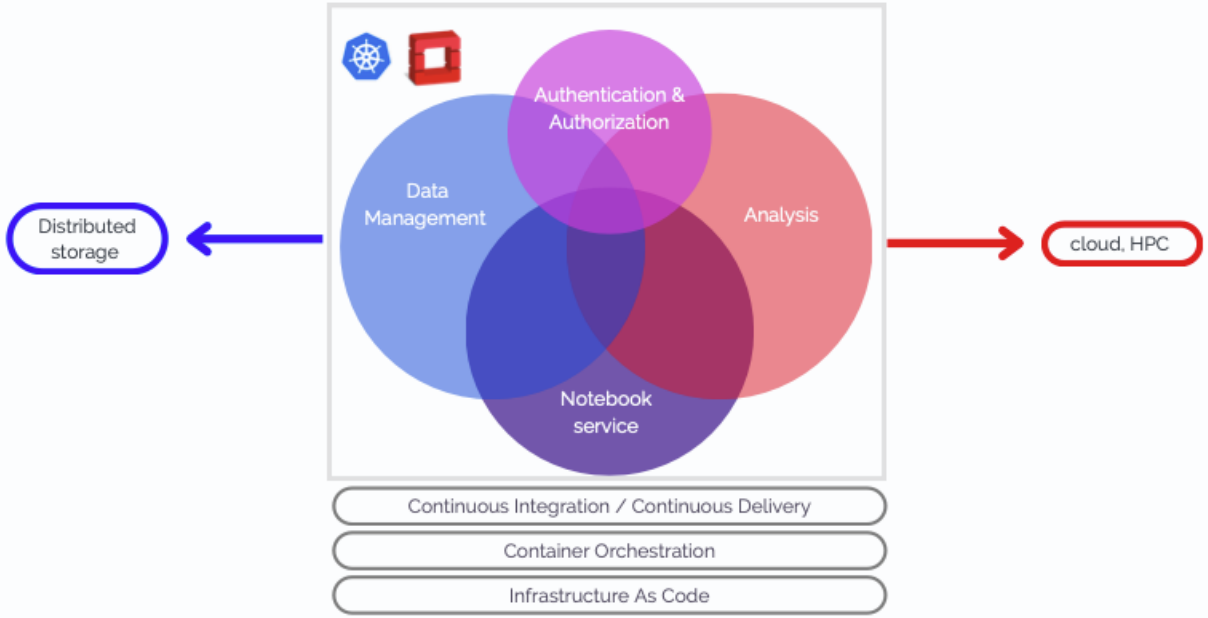
Storage and data management

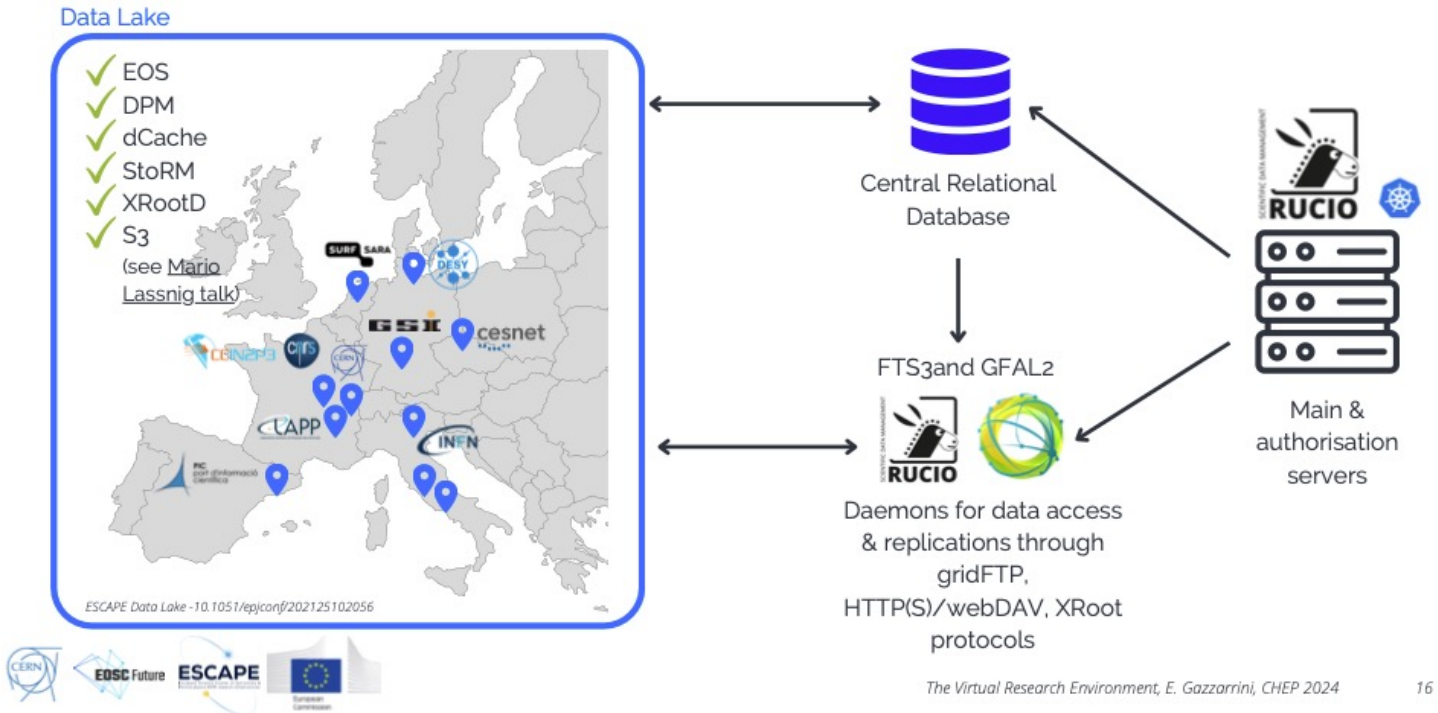
- Localize bulk data in a cloud service (Tier 1's → data lake): minimize replication, assure availability
- Serve data to remote (or local) compute – grid, cloud, HPC, ???
- Simple caching is all that is needed at compute site (or none, if fast network)
- Federate data at national, regional, global scales



Concrete examples

- ESCAPE data lake as part of EOSC-Future Virtual Research Environment (E. Gazzarrini, CHEP2023)
- Based on CERN File Transfer Service (FTS), and the Rucio data management and orchestration open-source tools
- Exploited successfully within the ESCAPE Dark Matter Science Project
- Another data lake has been implemented in the Nordic countries, see M. Wadenstein, CHEP 2023



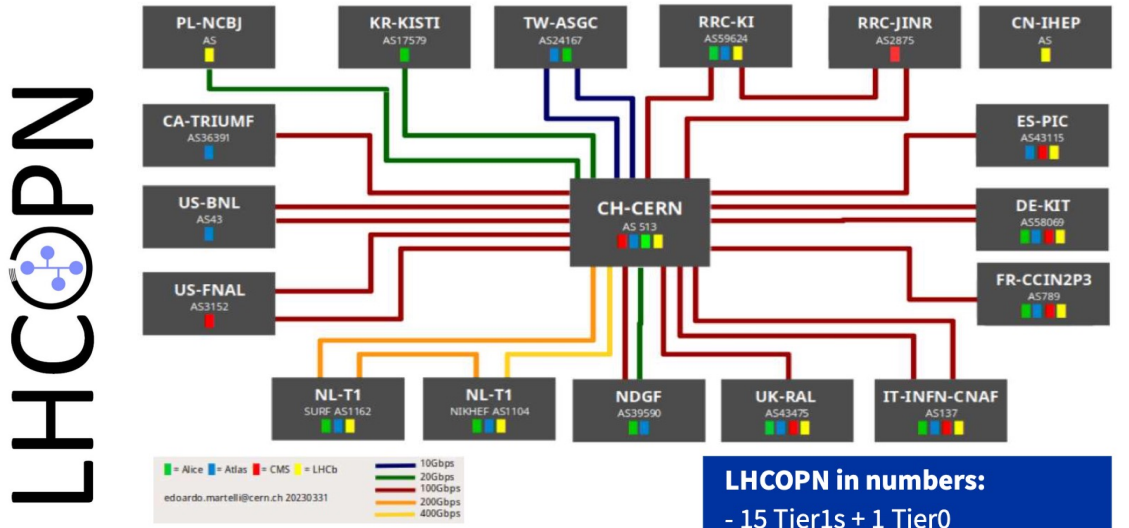


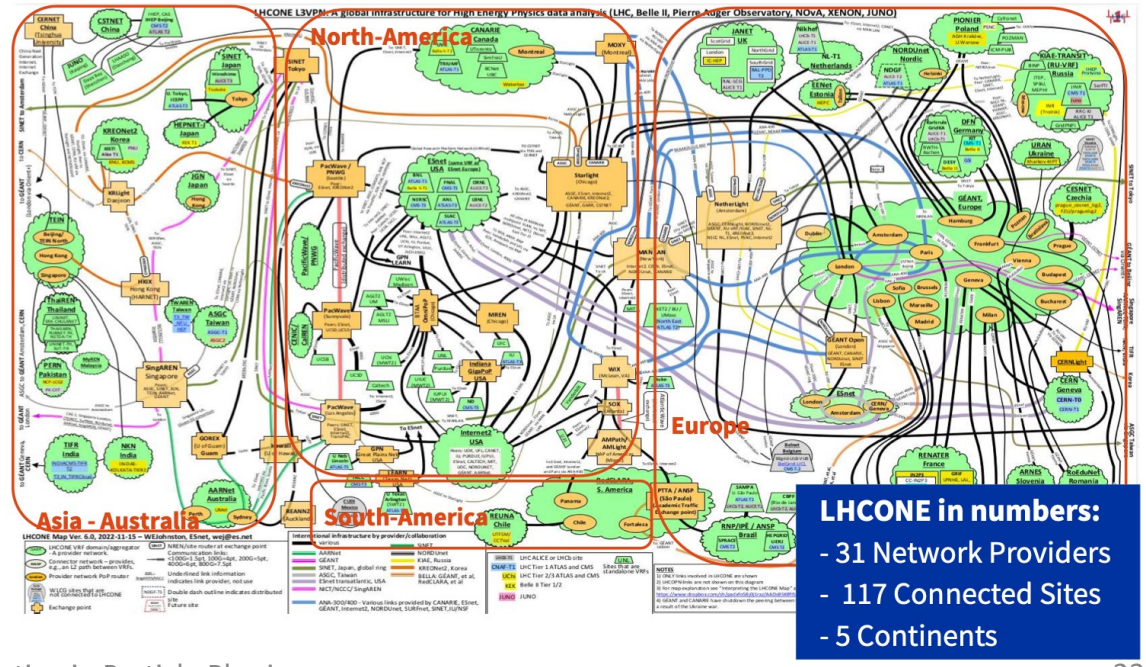
A distributed computing ecosystem

- Tools that enable us to operate the distributed computing infrastructure are mature and consolidating
- Commonalities emerge between experiments / communities
- Still far from ”turnkey” solution though
  - Some level of expertise needed for commissioning / configuration / operations

Network

- Heavily relying on network for much of the federated infrastructure
- Network evolution is challenging given the expected increase in number of sites, the addition of HPC and clouds, the implementation of data lakes
- Technological R&D
  - Software-defined network tools can help to meet transfer deadlines in overloaded networks
  - Data Center Interconnect (CERN-CNAF @1Tbps)
- The expected requirements (4.8Tbps from CERN to Tier1s) are being exercised with multi-year data challenges
  - October 2021: 10% test
  - Further tests in 2024 and later

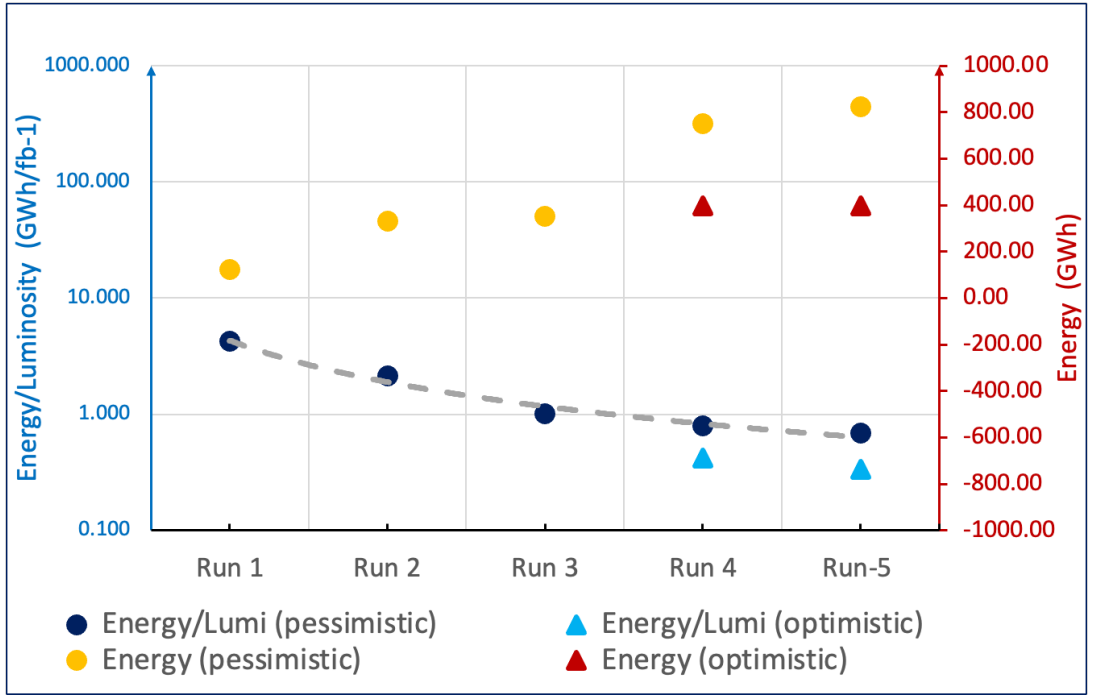




Sustainability

- The contribution of data-centers to greenhouse-gas emission is sizeable and growing, HEP is no exception
- Ongoing initiatives to quantify and improve carbon footprint
- Recent study shows that the energy needs of HEP computing can be kept under control by
  - modernizing the facilities towards higher energy efficiency.
    - major capital investment
  - Improving the software and computing models
    - gradual process bringing early benefits
  - improving the hardware technologies and optimizing the hardware lifecycle strategy.
    - investment in software portability

D. Britton, S. Campana, B. Panzer (CHEP 2023)



(BLUE) GWh/fb^{-1} = energy needed to **analyse** the data
(RED) = energy

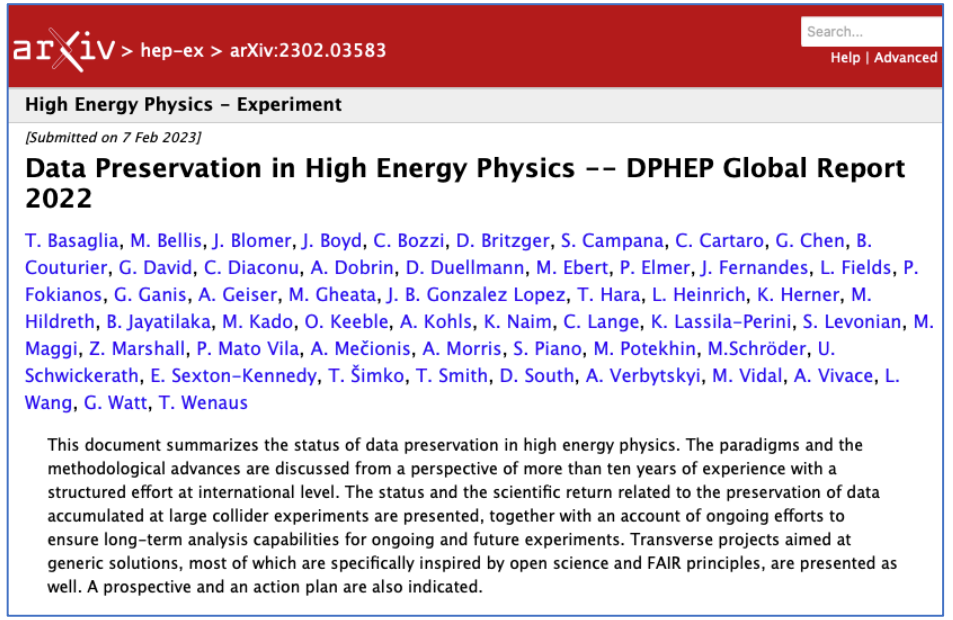
Energy needs in Run-4 and Run-5: +100% (+10% only) compared to Run-2 in a **pessimistic** **(optimistic)** scenario

GWh/fb^{-1} a factor 10 lower between Run-1 and Run-5
In Run-5, GWh/fb^{-1} in the **optimistic** scenario is half compared to the **pessimistic** scenario

Not to be forgotten

- Analysis: the “last mile”
  - ROOT Data Analysis Framework and the “pythonic ecosystem”
    - Very active area, prototyping analysis facilities (see e.g. pre-CHEP WLCG workshop)
    - Use cases and boundaries with distributed computing infrastructure to be sharpened further

- Data preservation, open data, (FAIR) data management plans, EOSC
  - Status, prospects and action plan detailed in recent report
  - Reana as “Workflow as a service”
    - re-executes analyses on a workflow-aware, container-based computing backend



arXiv > hep-ex > arXiv:2302.03583

Search... Help | Advanced

High Energy Physics - Experiment

[Submitted on 7 Feb 2023]

Data Preservation in High Energy Physics -- DPHEP Global Report 2022

T. Basaglia, M. Bellis, J. Blomer, J. Boyd, C. Bozzi, D. Britzger, S. Campana, C. Cartaro, G. Chen, B. Couturier, G. David, C. Diaconu, A. Dobrin, D. Duellmann, M. Ebert, P. Elmer, J. Fernandes, L. Fields, P. Fokianos, G. Ganis, A. Geiser, M. Gheata, J. B. Gonzalez Lopez, T. Hara, L. Heinrich, K. Herner, M. Hildreth, B. Jayatilaka, M. Kado, O. Keeble, A. Kohls, K. Naim, C. Lange, K. Lassila-Perini, S. Levonian, M. Maggi, Z. Marshall, P. Mato Vila, A. Mečionis, A. Morris, S. Piano, M. Potekhin, M.Schröder, U. Schwickerath, E. Sexton-Kennedy, T. Šimko, T. Smith, D. South, A. Verbytskyi, M. Vidal, A. Vivace, L. Wang, G. Watt, T. Wenaus

This document summarizes the status of data preservation in high energy physics. The paradigms and the methodological advances are discussed from a perspective of more than ten years of experience with a structured effort at international level. The status and the scientific return related to the preservation of data accumulated at large collider experiments are presented, together with an account of ongoing efforts to ensure long-term analysis capabilities for ongoing and future experiments. Transverse projects aimed at generic solutions, most of which are specifically inspired by open science and FAIR principles, are presented as well. A prospective and an action plan are also indicated.

reana

Reproducible research data analysis platform

Human factor

Training:

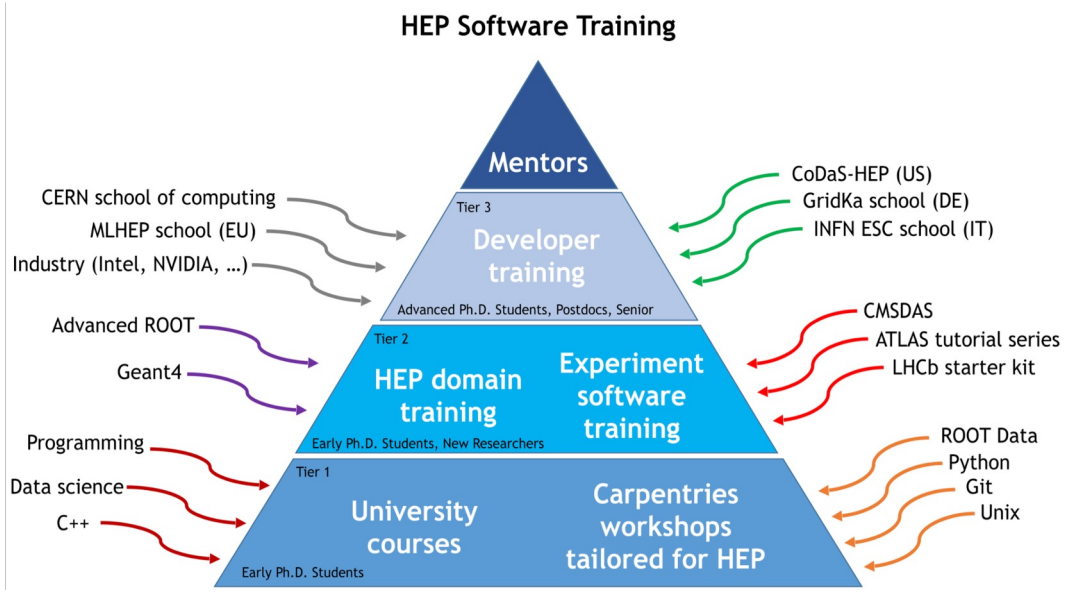
- Students/postdocs often lack basic knowledge (software and computing skills)
- University courses are often insufficiently oriented towards these technical aspects

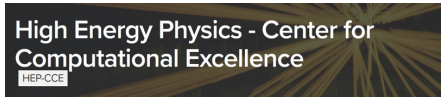
Situation definitely improving in the last few years

- Several tutorials organised by HSF and others
- Funded projects, in combination with experiments and existing projects, are helping HEP software to advance
- The HEP Software Foundation (HSF) as forum to build the community and share knowledge

Career recognition/opportunities:

- Computing & Software activities need to be recognized as fundamental to research activities and bearing a large impact on the final physics results.





Conclusion

- Challenges ahead!
- Rapidly evolving situation in all domains
  - Hardware & Software: parallelization & heterogeneity in reconstruction and simulation; we need to continue to modernize our applications to take advantage of hardware evolution
  - Infrastructure: utilization of (heterogeneous) resources that have not been natively designed by us; flexibility is key
  - Infrastructure: consolidation and scalability of tools and practices
- People:
  - Training is fundamental
  - R&D, deployment and operations in software and computing should have proper recognition
- Software & Computing is no longer a ”service” but an “element” of the “scientific apparatus” in HEP experiments