Tackling the Computing Challenges at CERN

CERN OpenLab Summer Student lecture

4th July 2023

Alessandro Di Girolamo - CERN IT


… a brief “random walk” through
CERN

What does *CERN* stand for?

Conseil

Européen pour la

Recherche

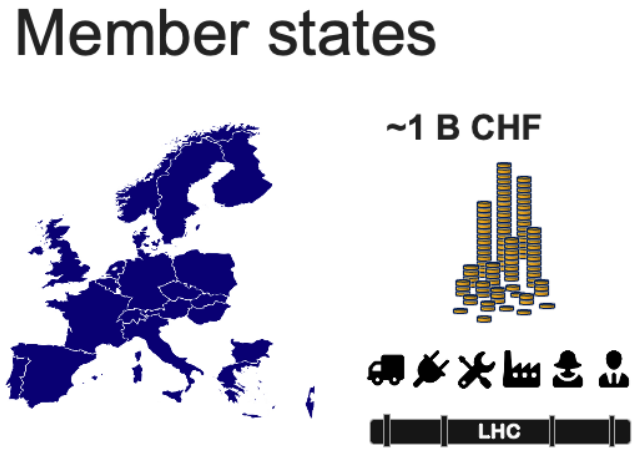
Nucléaire

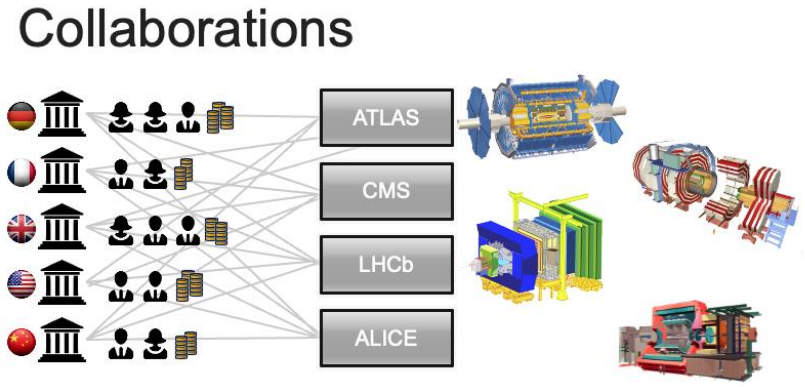
Nuclear?

CERN

Who is it?

- The Council
  - Member states
- The Collaborations
  - experiments





CERN: a worldwide collaboration

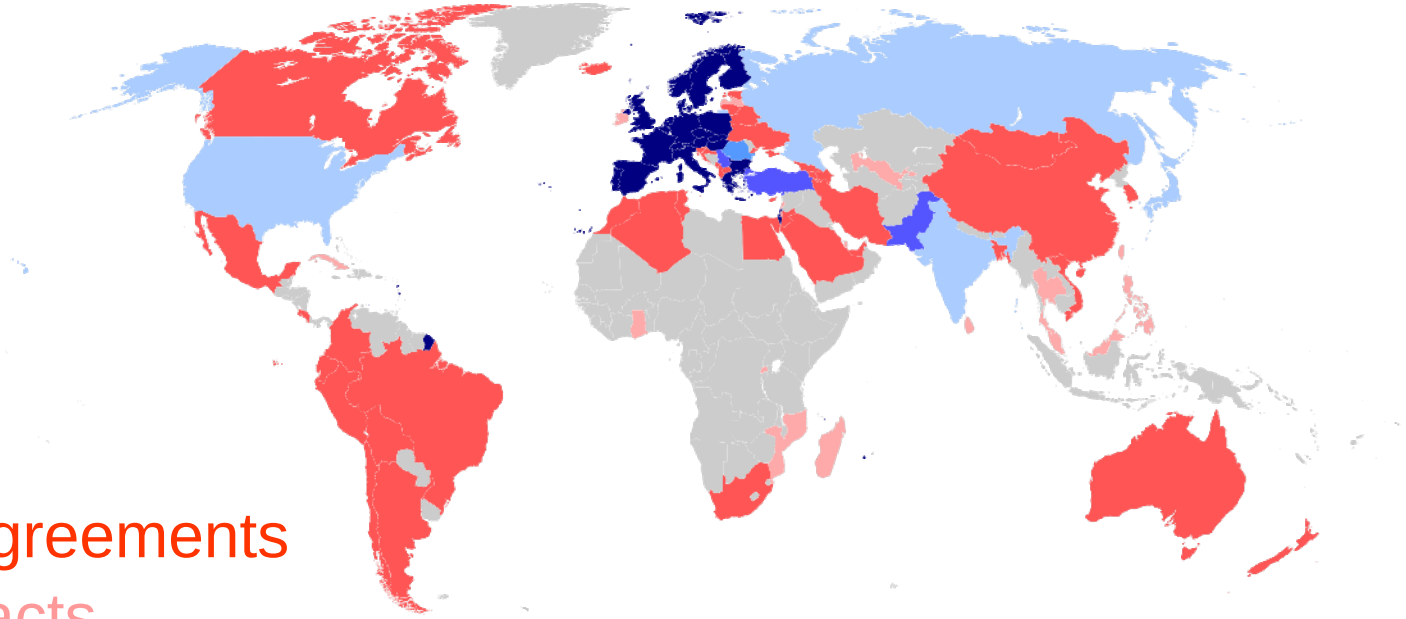
23 members

+ Associates

+ Observers

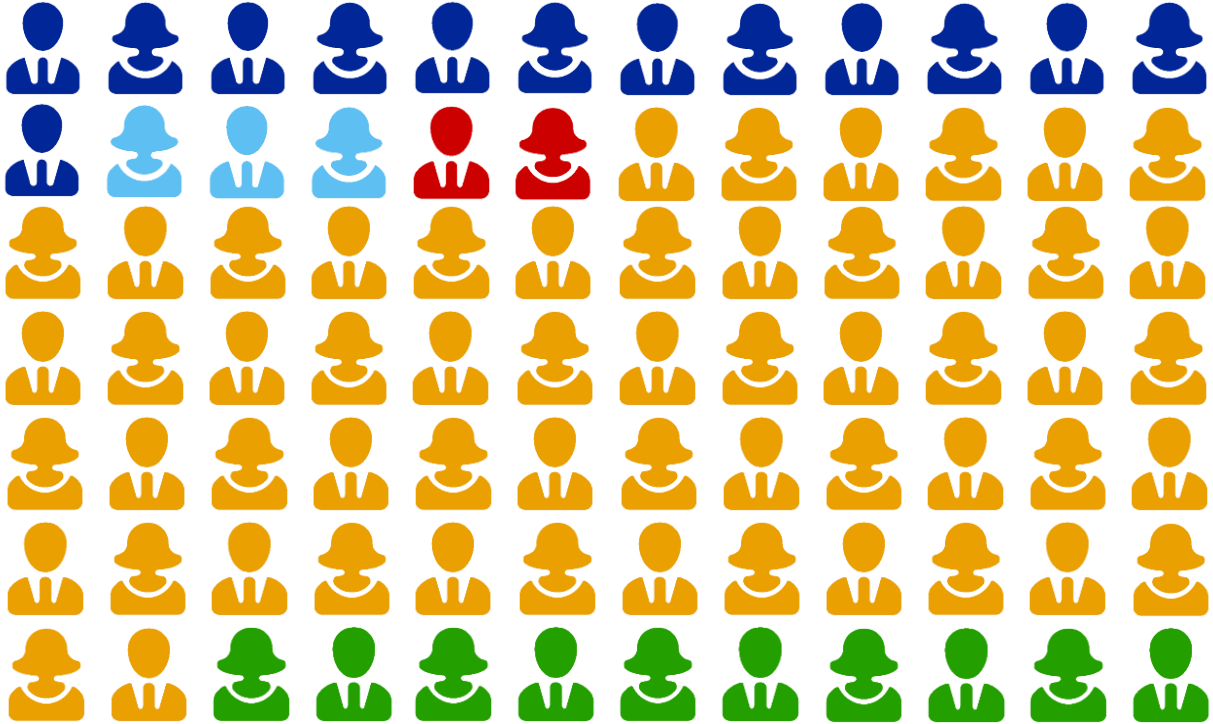
Cooperation agreements

Scientific contacts



How many persons?

15000+!



2500 staff

600 fellows & apprentices

500 students

10000+ users

2000 external companies

Our Mission

The CERN Mission

- perform world-class research in fundamental physics.
- provide a unique range of particle accelerator facilities that enable research at the forefront of human knowledge, in an environmentally responsible and sustainable way.
- unite people from all over the world to push the frontiers of science and technology, for the benefit of all.
- train new generations of physicists, engineers and technicians, and engage all citizens in research and in the values of science

CERN

How does it work ?

Accelerating and colliding

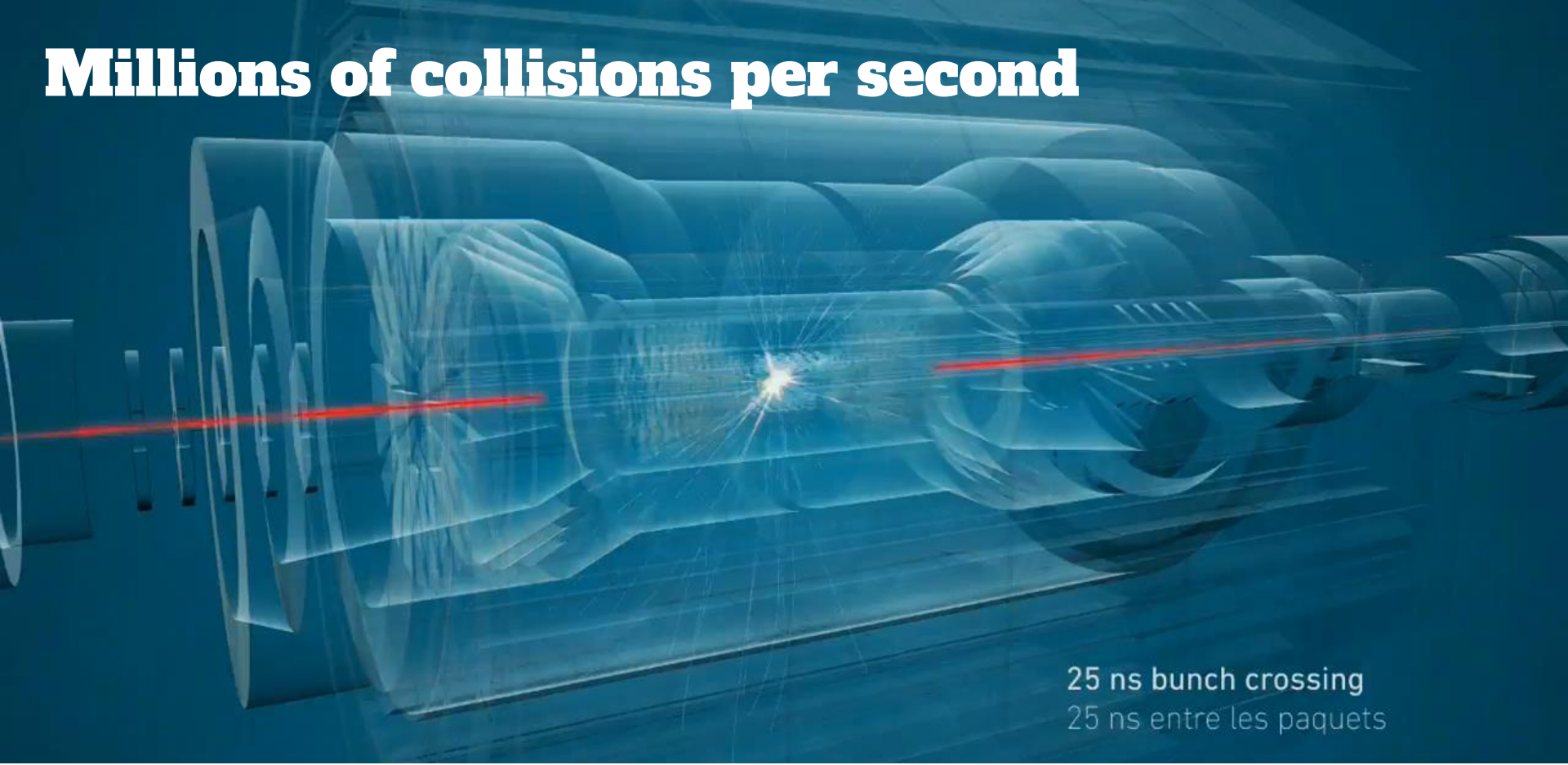
Accelerator chain

Incredible levels of energy

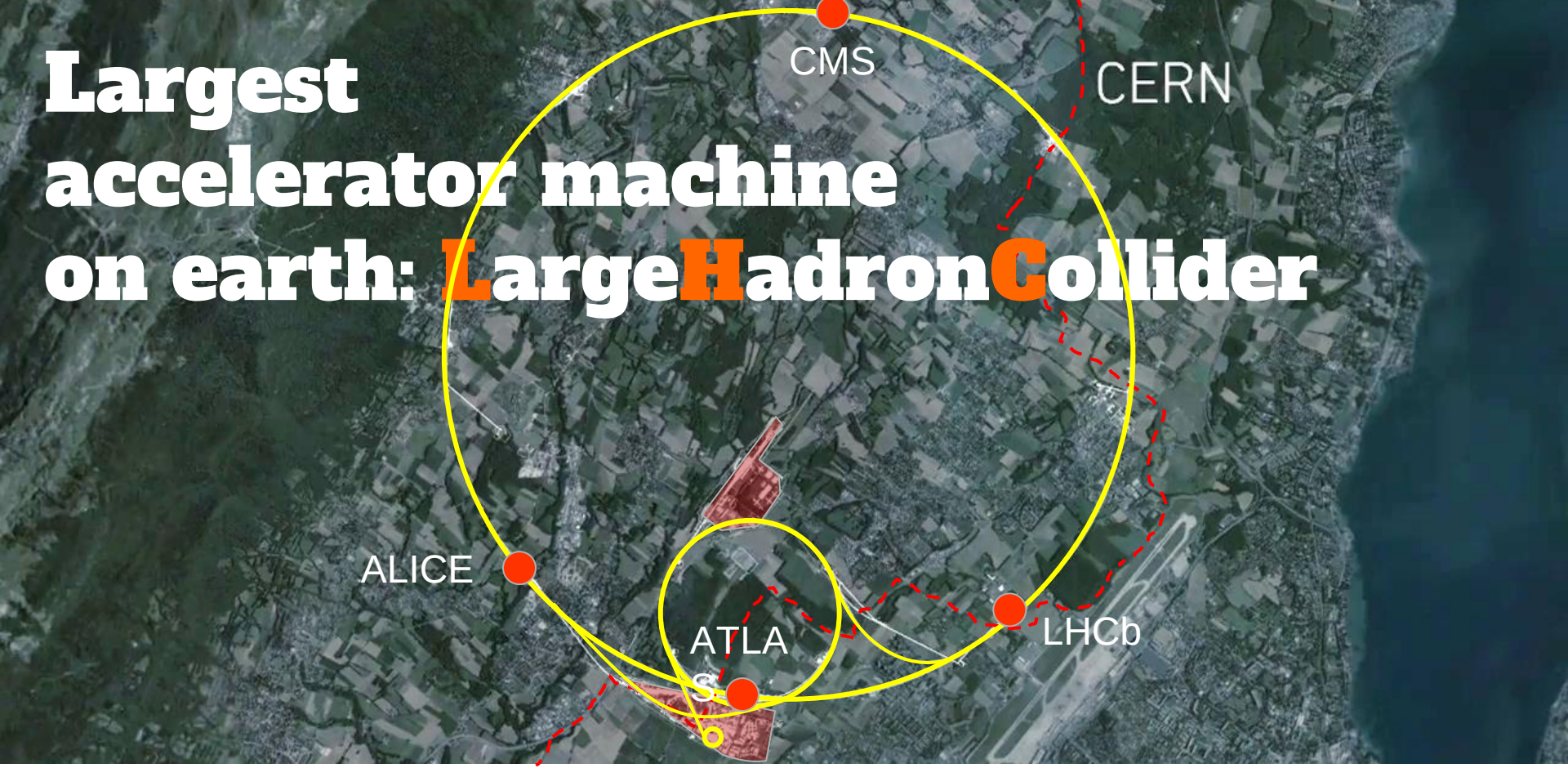
100’000’000’000’000’000’000’000

7+7 TeV

Millions of collisions per second



Largest accelerator machine on earth: LargeHadronCollider



ATLAS and CMS

- Two general-purpose detectors cross-confirm discoveries, such as the Higgs boson.

46m long, 25m diameter
weights 7'000 tonnes
100 million electronic channels , 3 000 km of cables

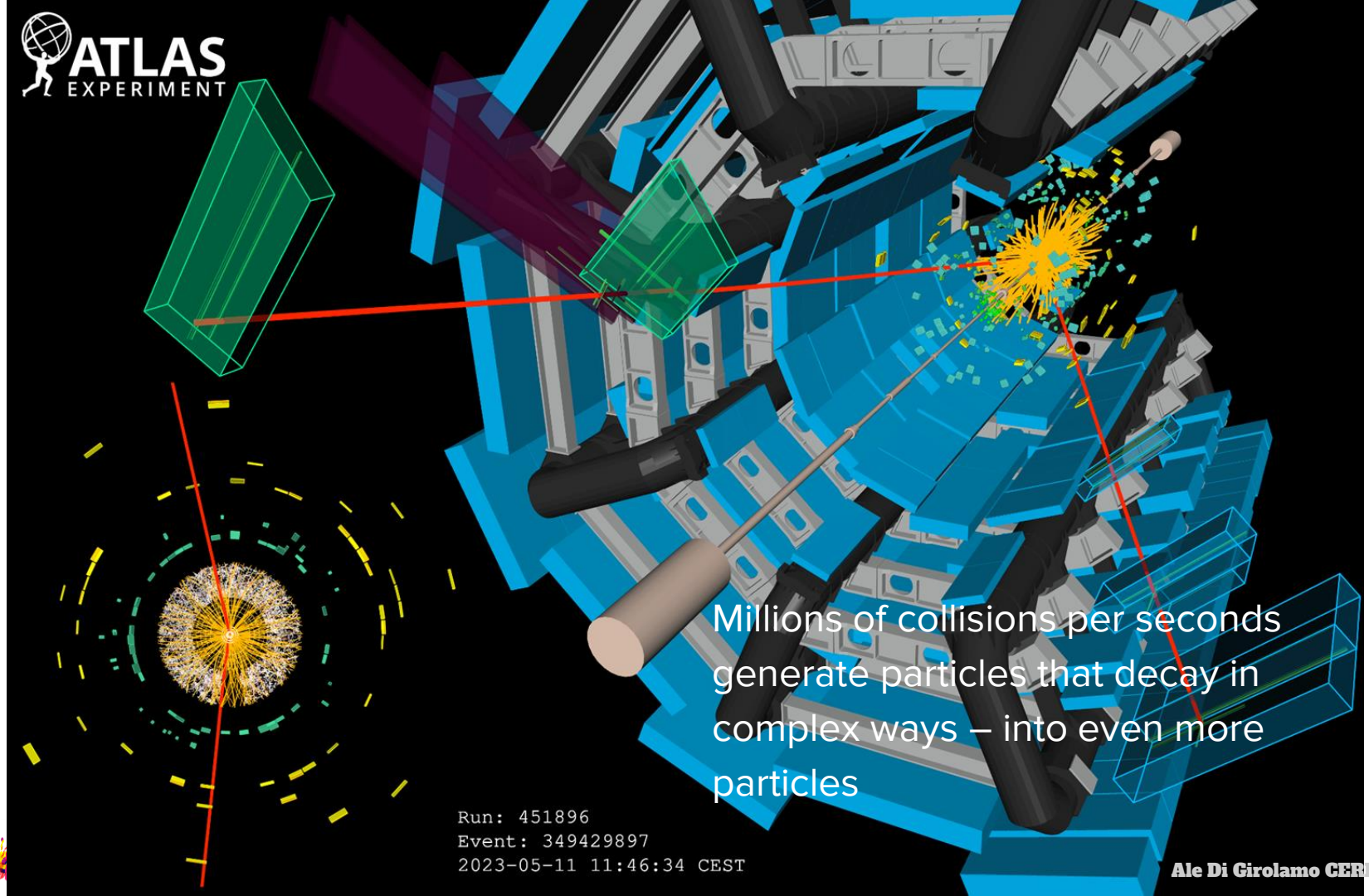
22m long, 15m diameter
weights 14'000 tonnes
Most powerful superconducting solenoid ever built

ALICE and LHCb

- ALICE and LHCb experiments have detectors specialised on studying specific phenomena.

Studies the «Quark Gluon Plasma», state of matter which existed moments after the Big Bang.

Studies the behaviour difference between the b quark and the anti-b quark to explain the matter-antimatter asymmetry in the Universe.



Millions of collisions per seconds generate particles that decay in complex ways – into even more particles

Data filtering (Trigger)

Data generated 40M times per second. Each "event" ~ 1.5MB

PB/s

100k selections per second

TB/s

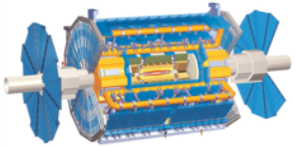
2-5 k selections per second

GB/s

PetaBytes of data per second:

The "trigger" is filtering data in real time, selecting potentially interesting events

From Data to Paper(s): data processing chain



Trigger

Raw data (RAW)

Reconstruction

Analysis Object Data (AOD)

Derivation

Organized production

Generation

Event generator output (EVNT)

Simulation

Simulated interaction with detector (HITS)

Digitization+Trigger

Simulated detector output (RDO)

Reconstruction

Analysis Object Data (AOD)

Derivation

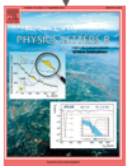
Derived AOD (DAOD)

Chaotic analysis

Detector data

Analysis

Simulated data



Online

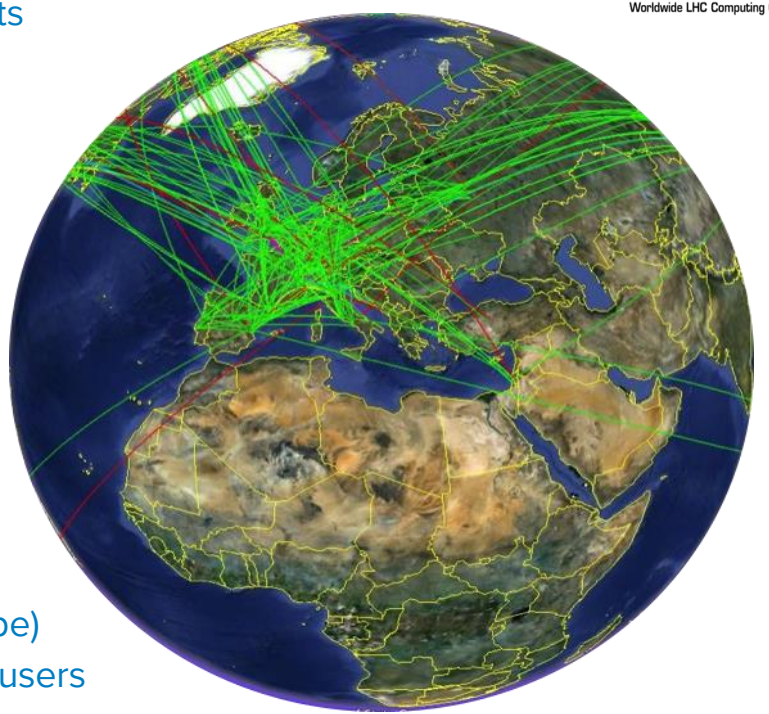
Tier-0

Grid

The Worldwide LHC Computing Grid - WLCG

- International collaboration to distribute and analyse LHC data
- Integrates computing centres worldwide that provide **computing** and **storage** resource into a single infrastructure accessible by all LHC physicists
- Global **network** connectivity

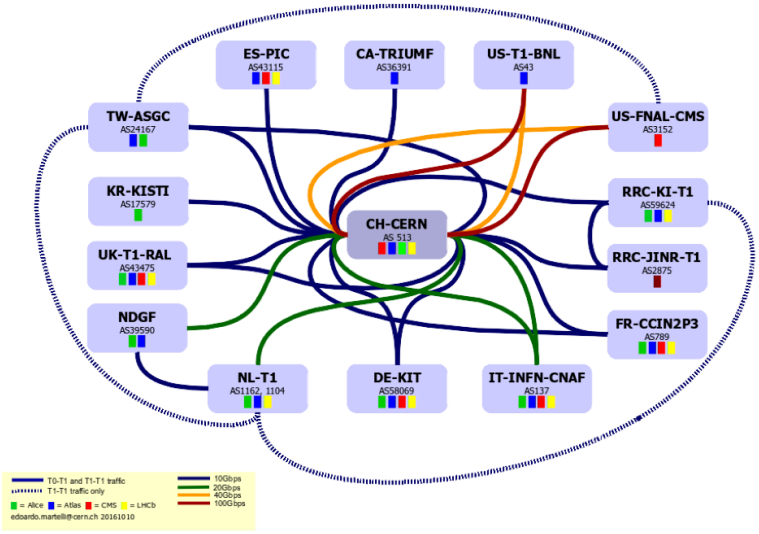
  - **Tier-0 (at CERN):** Large resource, recording and custodial archival of collision data, prompt reconstruction
  - **Tier-1s:** Memory and CPU intensive tasks second tape copy of detector data, occasional Tier-0 overspilling
  - **Tier-2s:** Processing centres, nowadays many are similar to the Tier-1s



- Around 1.5 million CPU cores fully occupied 24/7
- Around 1.5 EB data (~600 PB on disk and >800 PB on tape)
- More than 100 PB moved every month, accessed by 10k users

Worldwide Networking

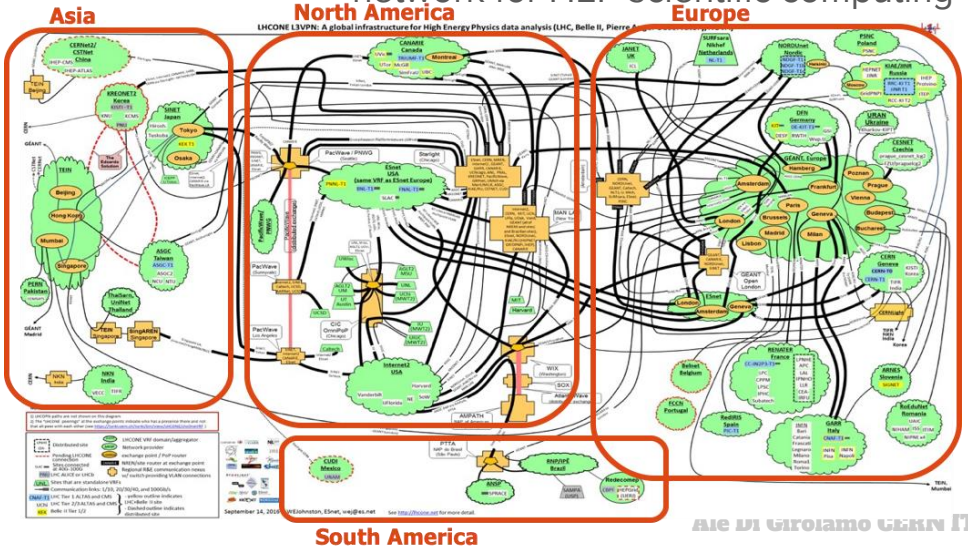
LHCOPN



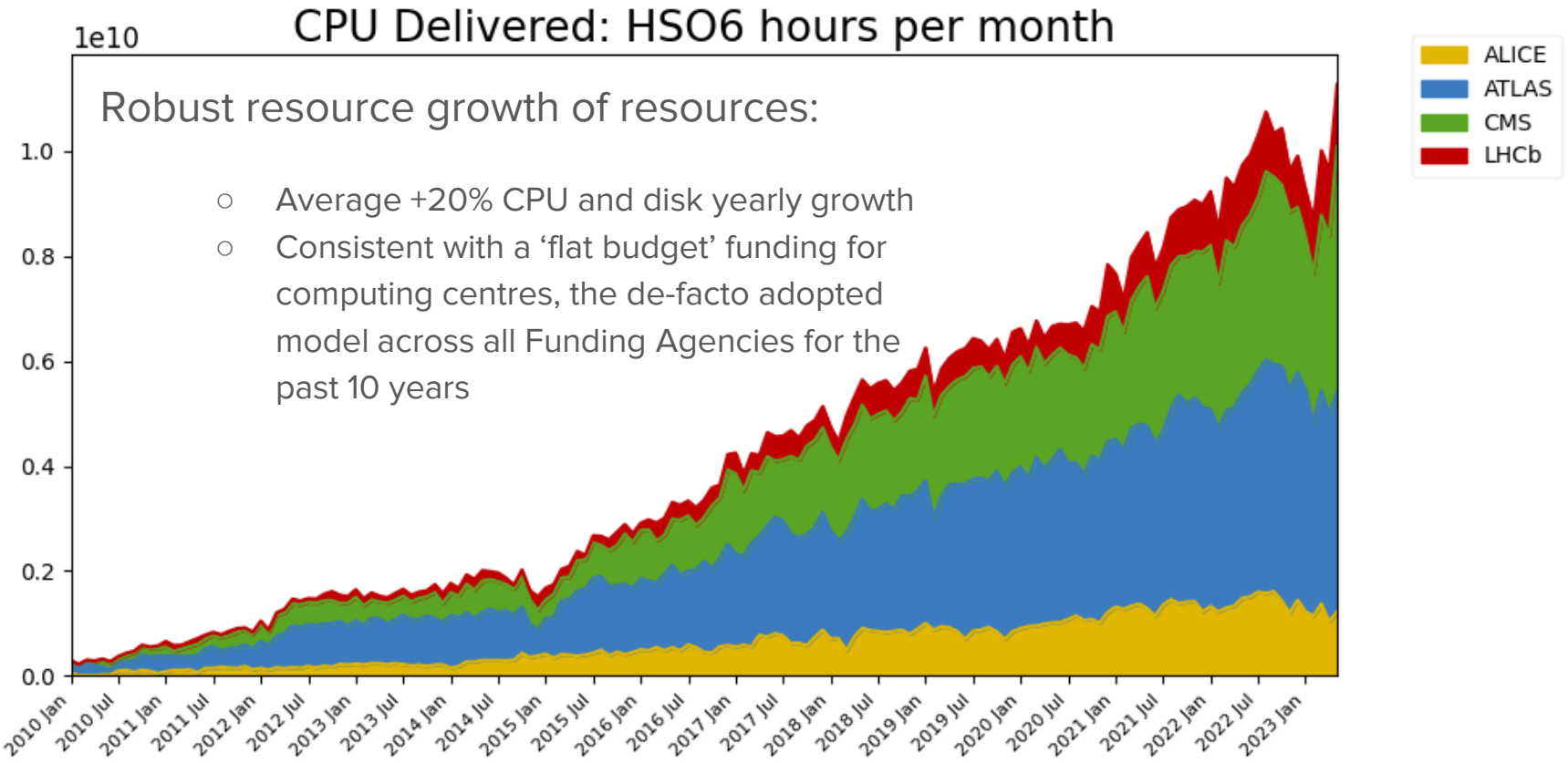
WLCG relies on powerful network infrastructure which connects the whole Grid

- LHCOPN: LHC Optical Private Network
  - For Tier0 ↔ Tier1s traffic,
  - managed by Tier0 and Tier1s

- LHCOne: overlay network
  - Enable NREN to manage general public network for HEP scientific computing



WLCG resources: compute

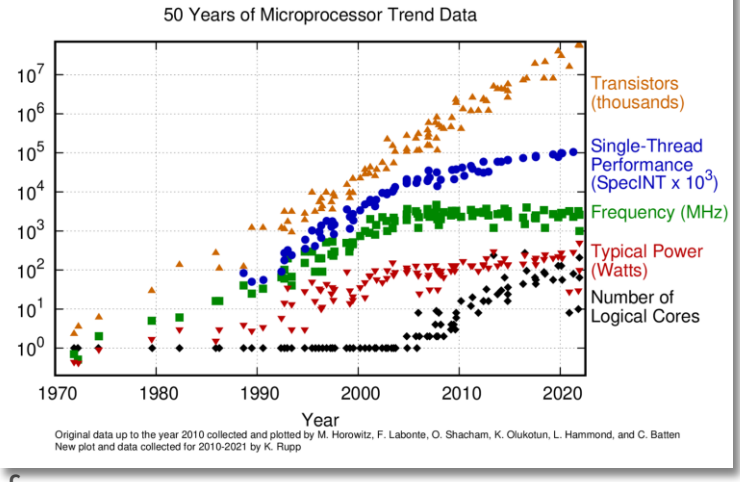


Robust resource growth of resources:

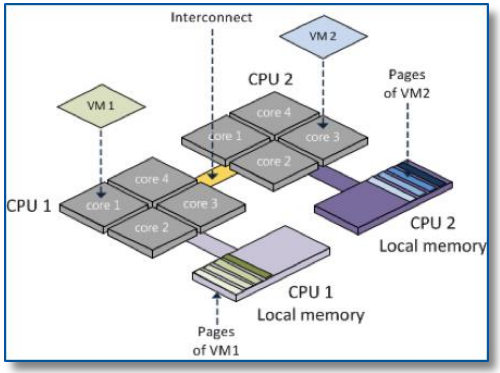
- Average +20% CPU and disk yearly growth
- Consistent with a ‘flat budget’ funding for computing centres, the de-facto adopted model across all Funding Agencies for the past 10 years

Hardware evolution

- A challenge not only for the HEP computing:
  - Transistors go into many cores, co-processors, vector units…
  - Clock speed stalled since ~2005
    - Single core performance is essentially also stalled
  - Mobile devices and data centres are the key volume markets
  - Memory consumption is a huge driver now
  - Memory access and I/O paths also become problematic

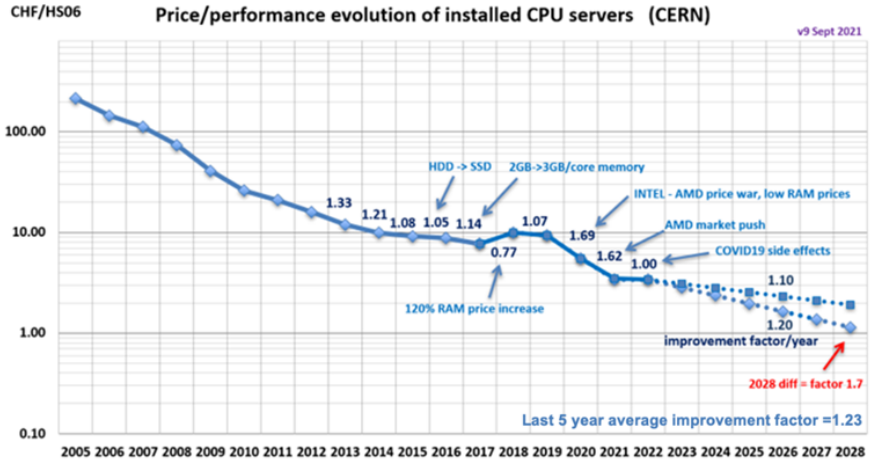


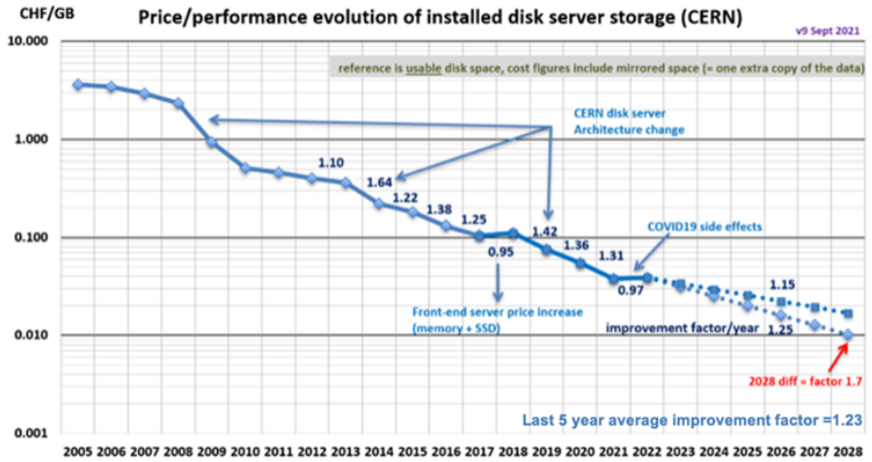
- Modernizing code will play a vital role for the upgrades of experiments and LHC
- Increase software performance by adopting modern coding techniques and tools
- Fully exploit the features offered by modern HW architectures
  - Many-cores GPU platforms, acceleration co-processors and innovative hybrid combinations of CPUs and FPGAs
- **Getting performant software requires significant investment in programming skills**
  - Also, it does not happen overnight…



Hardware cost evolution

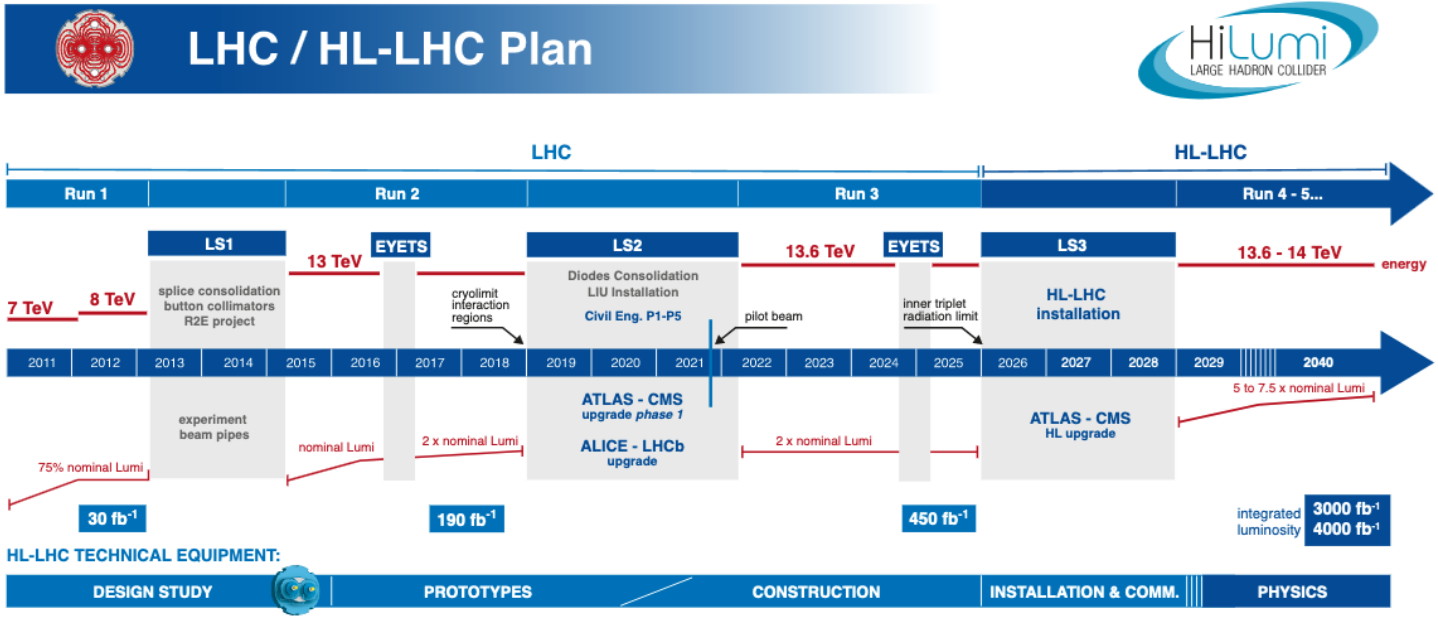
- Hardware cost is more and more dominated by market trends rather than technology
  - and science has no influence on these markets
- For the budget we have been assuming +20%/year in storage and CPU capacity, but we have to deal with large fluctuations and a “flat” budget





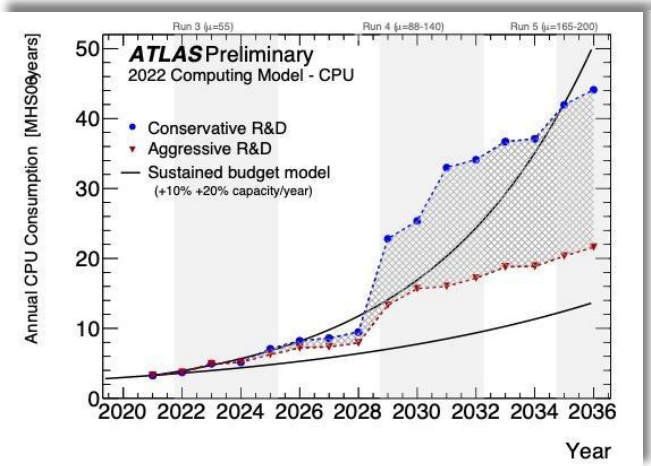
HL-LHC plans

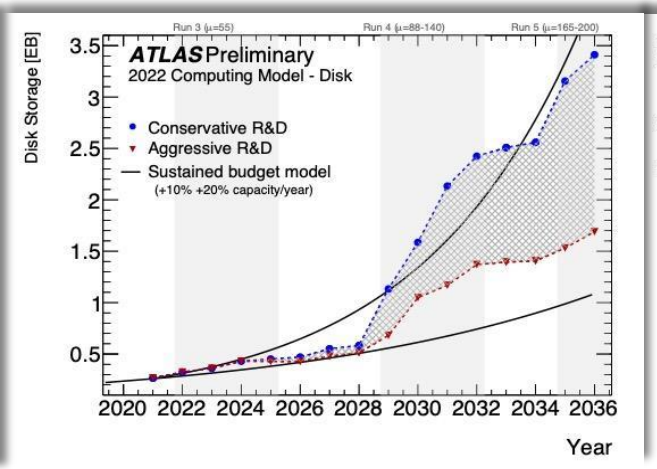
- Expected to be operational from 2029
  - Objective is to increase the integrated luminosity by a factor of 10
- ATLAS and CMS will face several challenges wrt Run3 (now): ~3x instantaneous luminosity, 3-4x pileup (mu ~60 -> 200)

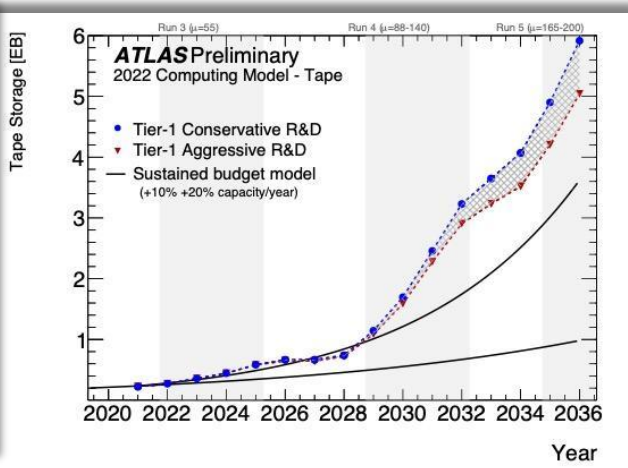


HL-LHC Software and Computing Challenges

- Computing Resource Estimates show need for major R&D (or budgetary) effort to achieve HL-LHC physics potential.
  - We have defined *Conservative R&D* and *Aggressive R&D* scenarios
    - Conservative should be achieved with today's effort and people
    - Aggressive requires more people, more R&D projects bringing resource savings
  - The black lines indicate the "flat budget" of 10% (lower line) and 20% (upper line).







Investing in brainpower is paramount!

Tackling the HL-LHC Computing Challenges

COMPUTATION	STORAGE
Make full use of all available resources, including new hardware architectures and facilities	Adapt computing model to maximise flexibility and to be able to operate with a wide range of storage types
Make physics choices that save computation without sacrificing experimental reach	Reduce the number of analysis formats and their size
Produce fewer simulated events with very high fidelity and more events with reduced fidelity	Make more use of cheaper storage technologies for data products that are not frequently accessed
Take advantage of 10 years of operating experience and redesign workflows to be more efficient without sacrificing quality	Write smaller/fewer events
Adapt the analysis model to maximise efficiency, especially by re-thinking user analysis	The more aggressive the development, the more efficient the use of resources → aggressive development plans = investment in people

James Catmore

Adapt to latest trends, more heterogeneity

Reduce the disk footprint

Use what we have more efficiently

Time to act: NOW

Experiments defined their roadmaps toward HL-LHC:
ATLAS S&C HL-LHC Roadmap ; CMS Phase 2 Computing Model

- Based on the HL-LHC schedule, it is clear that R&D projects need to start now
  - from R&D to demonstrators, with measurable impact
  - We see already lot of engagement!
- Integration and validation will require lot of tim
  - Late arriving R&D is risky

Some caveats: lots of “business as usual”

We have to keep on running the experiments while we are planning for major upgrades

- “we are building a new ATLAS while we are running ATLAS” (Karl Jakobs, ex ATLAS spokesperson)
- Failing is not an option

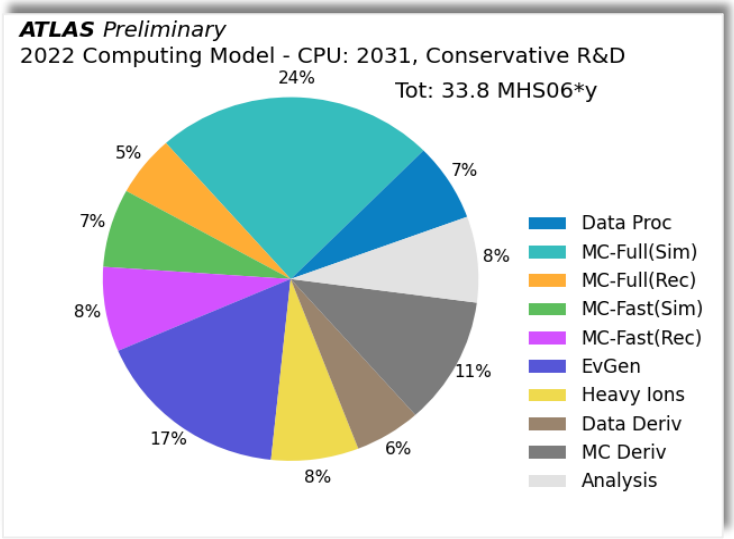
Lot of efforts that need to go into non-R&D work (or at the boundaries)

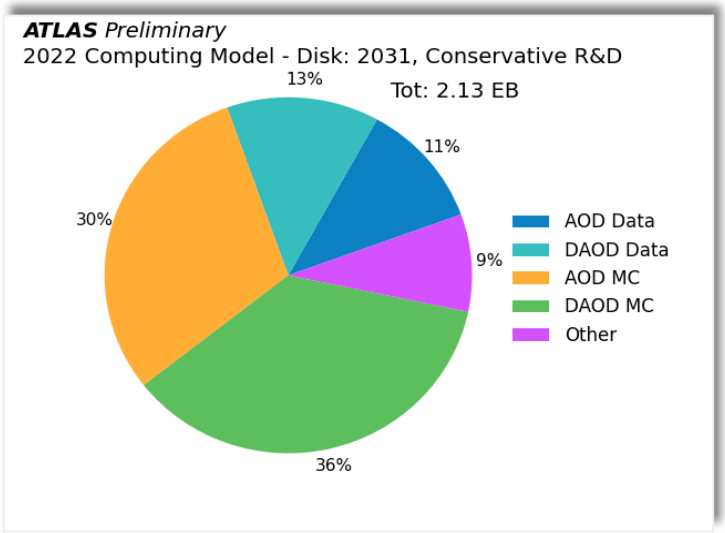
- Maintaining our current software
- Updates to the distributed computing infrastructure
- Upgrade geometry, digitization, re-tuning simulation…
- SW performance improvements

R&D projects are on top of this:

- balance (between R&D and “business as usual”) is key
- **We need a strong focus on “impact”**

HL-LHC - Resource Estimates for 2031

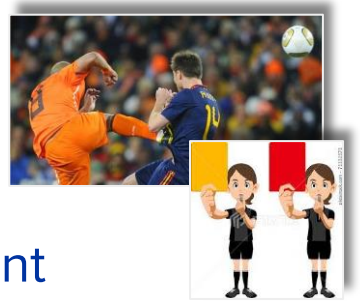




- There is no “silver bullet”:
  - it’s not that we do have 1 (one) single problem and if we solve that “we are done”.
- we need to work on each and every part of our system
- The opportunities are “everywhere”: we need to work hard to scavenge a few % here and there
  - An “evergreen” motivational speech, Any Given Sunday : true we do not play *against* anyone, but we still want to bring out the best from ourselves - while having fun!

Tackling the challenges

- There is nothing bad in being “two” in tackling difficult challenges!
  - Important we work together and we organize ourselves
  - We don’t compete
- There are no shortcuts and not so many “low hanging fruits”
  - It requires disciplines from all of us
  - Others might suffer from our “being late”



- Expectations management is paramount
  - We do not dictate! We do bring people onboard, we (try to) motivate them (and people motivate us!)
  - Crucial to clearly define our commitments taking into account all the “other” stuff we have to do
  - It is critical that we are able to estimate *by when* “that something” will be done.

Tackling the challenges: infrastructure evolution

- HPC and Clouds (and Industrial partners)
  - More and more often we are seeing them discussed together
  - Not really much overlap, except that they “both” are not standard Grid sites.
- They really deserve two different “set of slides” → next

High performance computing refers to computing systems with extremely high computational power that are able to solve hugely complex and demanding problems.

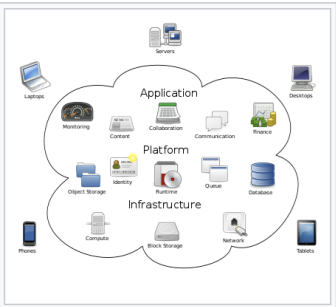
European Commission - Exscalate4Cov project : supercomputing to identify new therapies for COVID-19

In the digital decade, high performance computing (HPC) is at the core of major advances and innovation, and a strategic resource for Europe's future.

Cloud computing[1] is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.[2] Large clouds often have functions distributed over multiple locations, each of which is a data center. Cloud computing relies on sharing of resources to achieve coherence and typically uses a "pay as you go" model, which can help in reducing capital expenses but may also lead to unexpected operating expenses for users.[3]

Contents [hide]

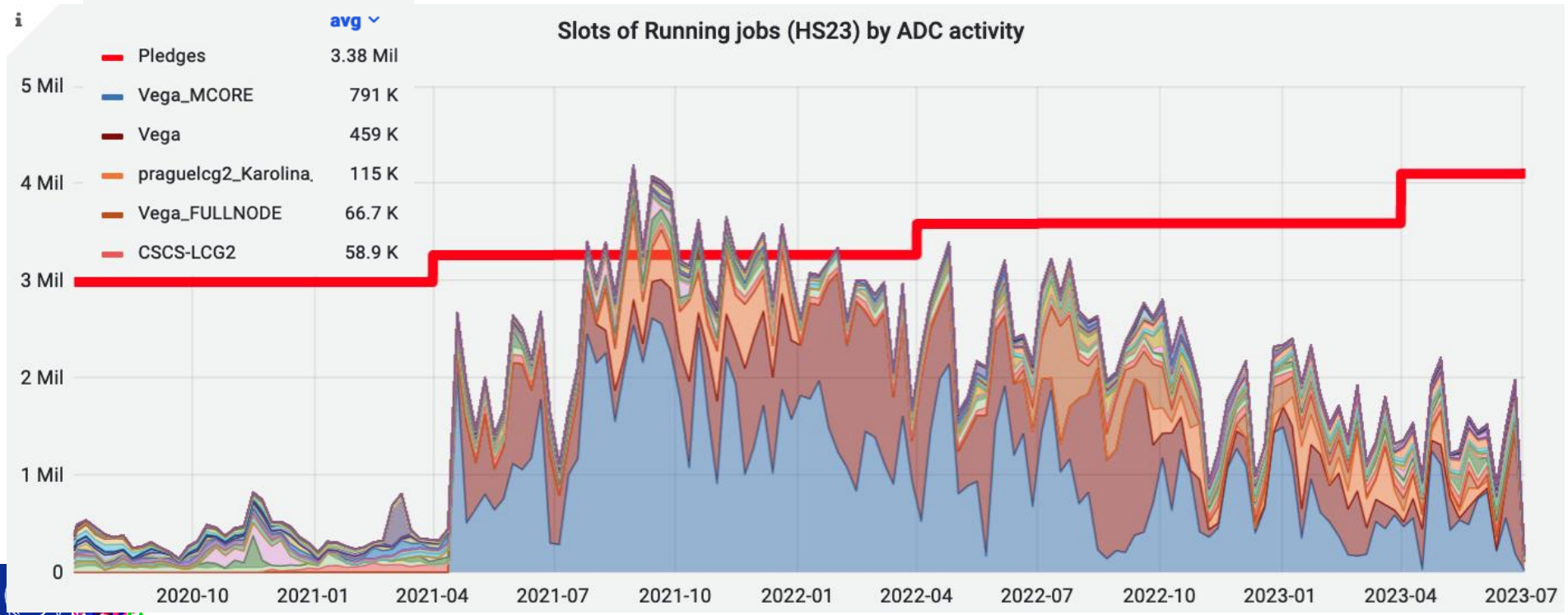
1 Value proposition
2 Market
3 History
  3.1 Early history
  3.2 2000s



Cloud computing metaphor: the group of networked elements providing services need not be individually addressed or managed by users; instead, the entire provider-managed suite of hardware and software can be thought of as an amorphous cloud.

HPC and ATLAS - last 3 years - HPC only

- A few HPCs make a big difference
  - But still it's important to highlight that many HPCs are integrated!
  - Looking forward to Leonardo(n4) and Perlmutter(n8)



A strategy towards opportunistic HPC

- Investment in proportion to opportunity
  - At the moment, e.g., we see ARM as a higher priority than Power
  - Not trying to get in on every machine!!
    - We don't need to – getting the easy/with internal good support ones gives ample CPU.
  - All the "easy" HPCs transparently integrated, running all workflows, are much welcome!
- We have a good (and growing) toolkit for edge services
  - With several available solutions, one is likely to fit a new machine
- Work towards several big HPCs is ongoing
  - E.g. Perlmutter in the US; hope to capture some of Fugaku(2)/CSCS with ARM development; Getting close to using Toubkal (UM6P, largest HPC in Africa).
- The details can make the difference:
  - HPC are not easy, need expertise on I/O, shared file systems, scheduling, AAI…
- Without forgetting:
  - The risk that the opportunistic CPU resources will disappear at any time is very real
  - The main physics program of the experiment should be supported through pledged resources
  - A constant yearly increment for both CPU and Disk is reasonable to minimize the risks that, if these resources disappear, the computing centers providing pledged resources won't be able to provide what is needed

R&D Projects with Industry

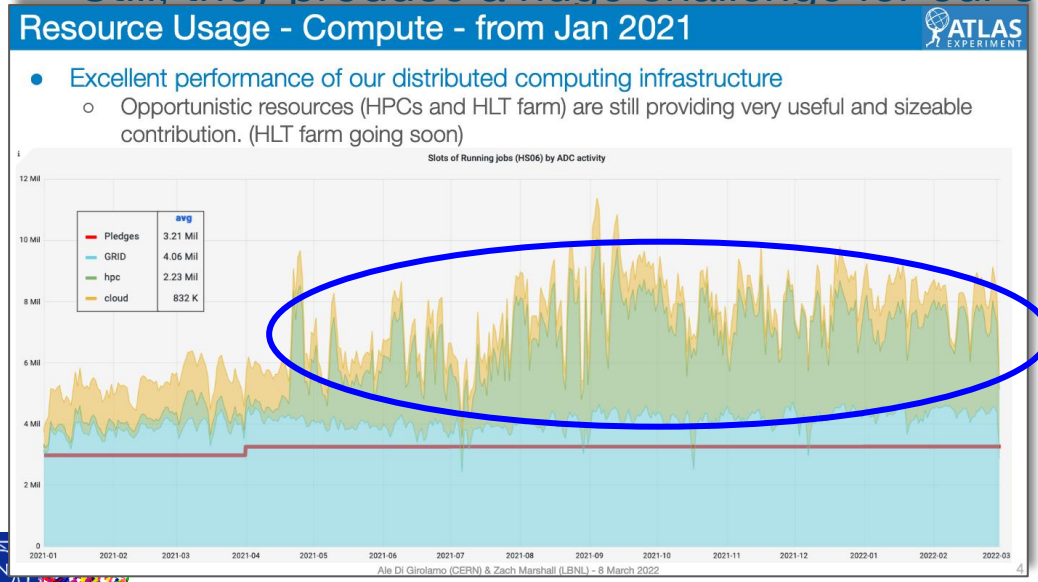
- The experiments are engaging with a number of industrial partners on R&D projects, e.g.:
  - SEAL, a decentralized cloud storage start-up
  - Google Cloud and Amazon Web Services

- These are valuable opportunities to evaluate technologies and to define a resilient long term strategy
  - Use of cloud compute for *all* workflows, and a comparison to traditional resources of the total cost of ownership (TCO)
    - Google already used for interactive analysis
  - Evaluation of test technologies (e.g. TPU, ARM) using cloud resources without requiring large-scale purchases
    - AWS extensively used in setup and validation of ARM Athena nightlies builds
  - Evaluation of remote cold storage as a complement / supplement to tape systems

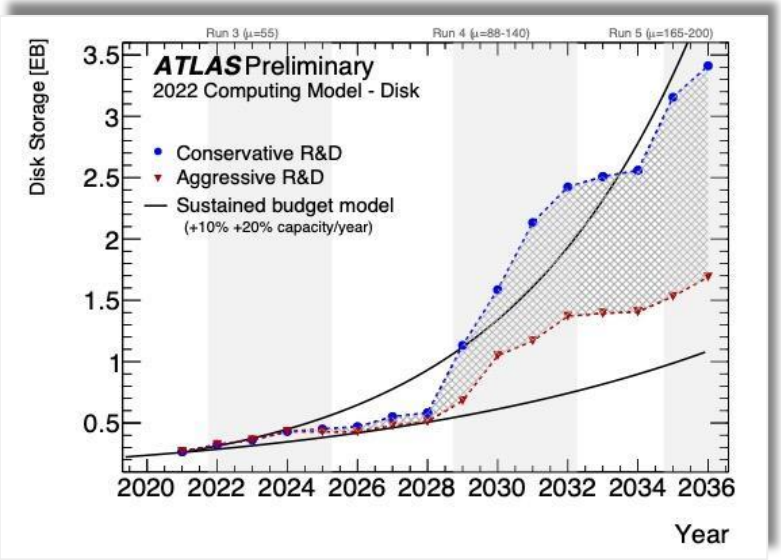
- These projects also present great connections with strong partners
  - Amazon and Google are major players that many groups already collaborate with
  - Being able to speak the same language and learn how they operate is invaluable

Opportunistic CPU but NO opportunistic Disk

- Resources extremely useful for the physics community
  - We have developed a detailed programme for secondary MC simulation samples
  - E.g.: large number of diphoton background events simulated → significant reduction of spurious signal systematic uncertainty → enabled a low diphoton-mass analysis; V+jets w both new Sherpa and MadGraph → primary setup for upcoming V+jets modelling pub, used in Z+jets measurements paper
- Still, they produce a huge challenge for our storage: we need to store all these



Storage challenges



- Our data sample is going to be 10x bigger!
  - today trigger rate is ~4kHz, will be 10kHz in HL-LHC,
  - RAW size from 1.5MB/event to ~4MB in HL-LHC
  - MC simulated events from 30B/year to 150B/year
- We are working on several R&D (and lots of “business as usual”) to make sure we make the best use of our Disk and Tape:
  - New analysis models: smaller data types
  - Data Carousel: trading Tape (cheaper, but slower) for Disk
  - Constant review of the content and the objects stored in each datatypes

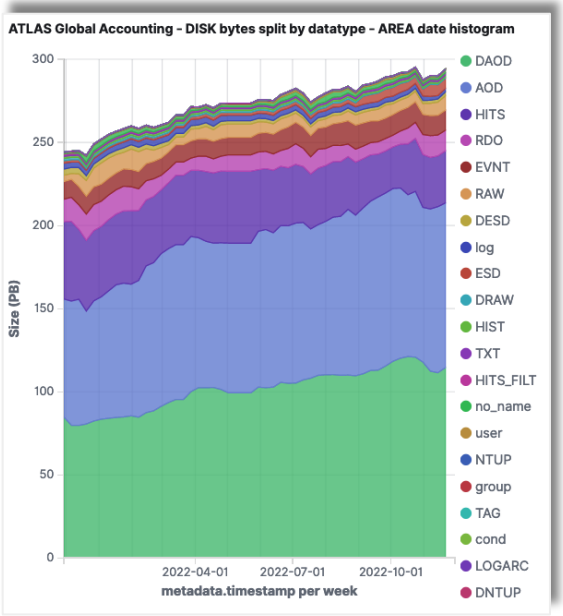
Storage challenges: more

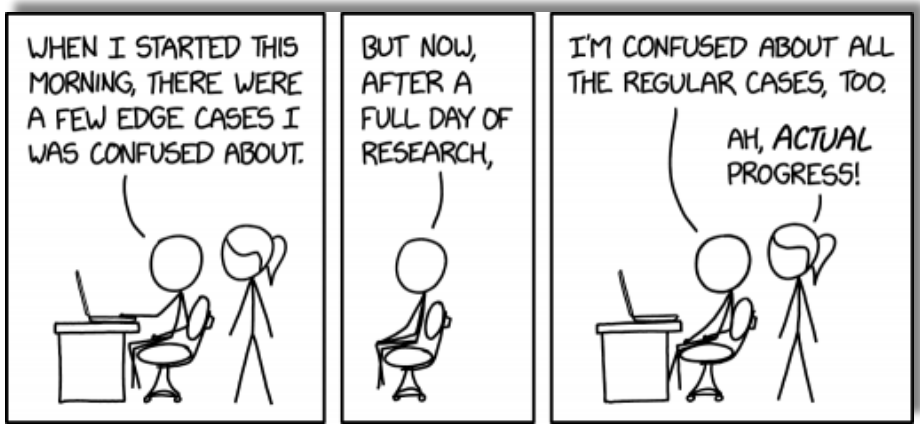
- Lot of analytics:
  - Too much, never enough - and always complicated to keep all the chain working! - help welcome!
  - understand what we use, how often, how many replicas …

Some more blue-sky R&D are fascinating, e.g.:

- fastchain: trading CPU for Disk
  - A fast end-2-end simulation, we could choose to save less (or zero) intermediate data - potentially important impact on disk and tape savings
- More aggressive deletion - and recreation on demand
  - Again trading CPU for Disk, one of the complexity is on the timely reproduction (with proper metadata)

Data Scientists: WANTED



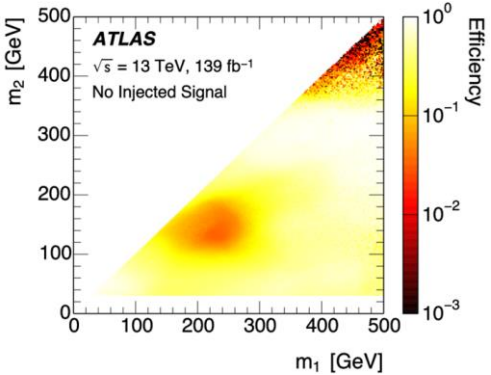


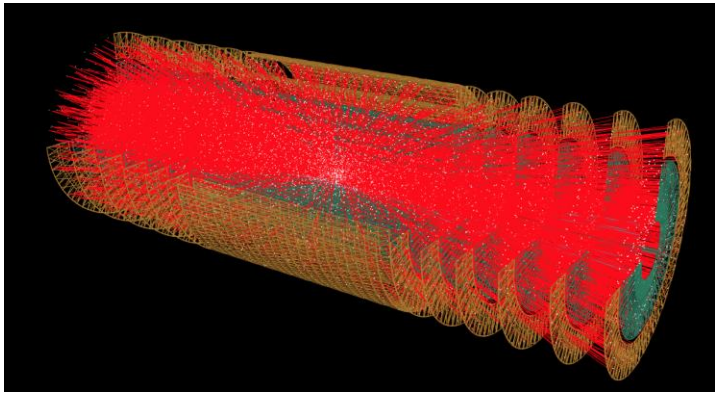
Tackling challenges: considerations

- HPCs are very useful to mitigate the risk of being short in CPU resource
  - With some caveats - as we discussed
- Clouds and in general project with industry help us in improving our frameworks, make it more sustainable and “community standard”
  - Some interesting “opportunistic storage” R&D
  - And TCO evaluation is important
- Many more R&D to mitigate the HL-LHC not-enough-CPU risks:
  - We are following with great interest the Geant4 collaboration (+Adept and Celeritas) and Event Generator groups’ R&D efforts
- Many R&D to mitigate the storage challenges’ risks:
  - Our systems are complex: need lots of brains and analytics expertise to take the right informed decisions
- Note that not all our R&D will reduce our resource consumption!
  - We want to do *great* physics, some of the R&D will help us on this too
  - And some others will push us to rewrite our (reco and tracking) code, to be able to be accelerators (e.g. GPU) friendly
    - → independently on the usage or not of accelerators, we will for sure gain having better written code: sustainability and capability to engage more people

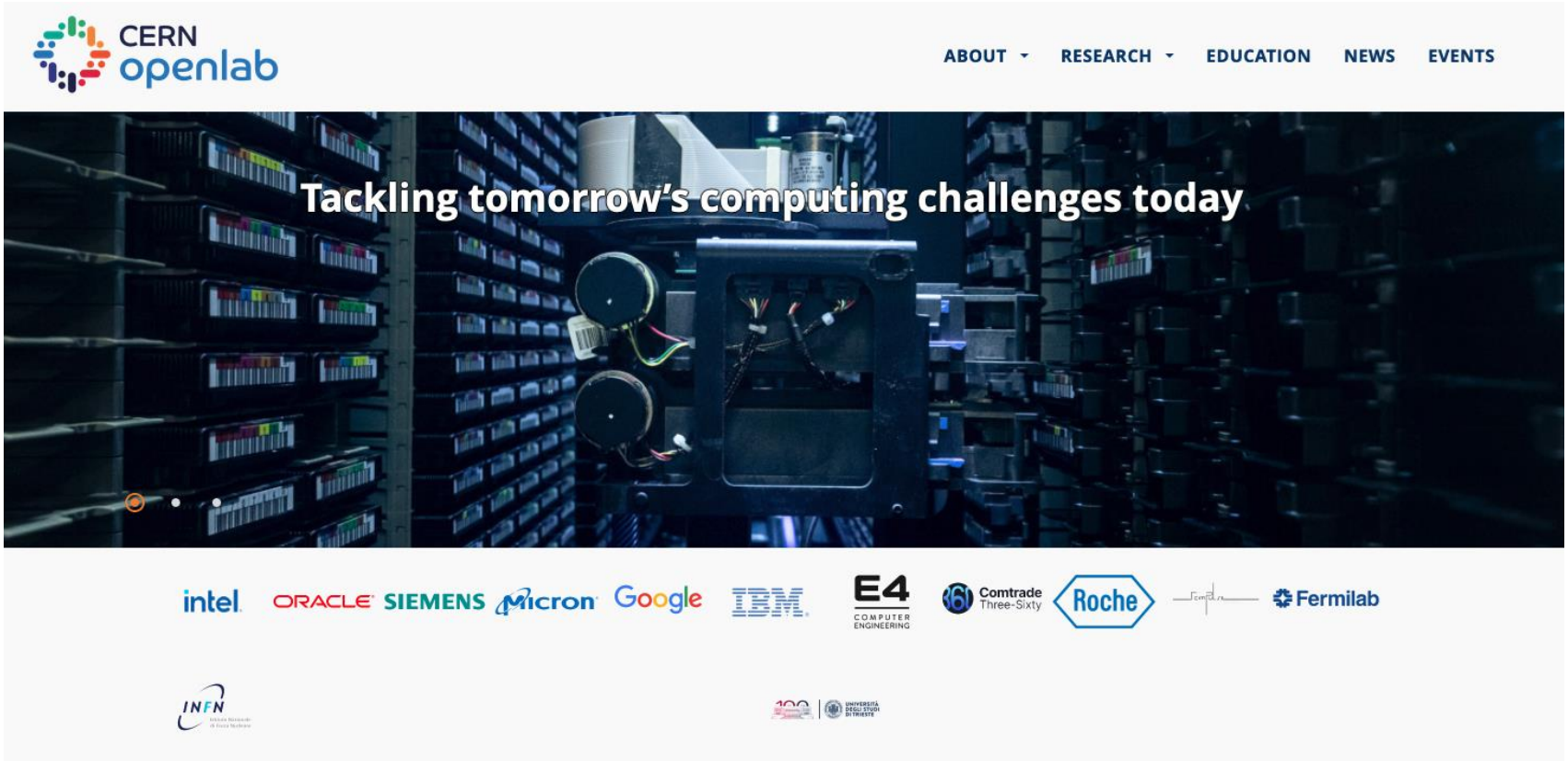
Machine Learning and Artificial Intelligence

- Machine Learning (and AI) is pervasive in the experiments
  - For deciding whether an event will be recorded (“triggering”)
  - For reconstruction of detector signals (e.g. pixel clusters)
  - For reconstruction of particles (e.g. electrons and heavy hadrons)
  - For high-level analysis (e.g. signal / background discrimination) and physical interpretation
  - For fast detector simulation (e.g. generative networks for calorimeter simulation)
  - Not only inside the experiments, but also work in collaboration with external machine learning experts (e.g. TrackML challenge, Higgs ML challenge)
- AI for computing: Operational Intelligence





CERN OpenLab



“Welcome”!

- The experiments are facing interesting, difficult, but solvable software and computing challenges for the HL-LHC
  - One of the biggest challenges not mentioned here is supporting and retaining skilled developers – YOU!
- Time to act is *now*
  - By experience we know how long and painful integration in the our frameworks and full physics validation are: we should take this into consideration to manage our expectations!
- Focusing our efforts to common shared objectives is paramount
  - The *way* in which we work can make the difference between success and failure!
  - Fragmenting efforts is going to be lethal - not effective
  - An interesting read: Socio economic impact of big science center

Success is no accident.
It is hard work, perseverance,
learning, studying, ***sacrifice***,
and MOST of all,
love *of what you are doing.*
-Pele

Thank you

... a special thanks for the many very useful discussions and providing material for this talk, to: Zach Marshall, David South, Concezio Bozzi, Stefano Piano, Fernando Barreiro Megino, Maria Girone and many more!

Plan B