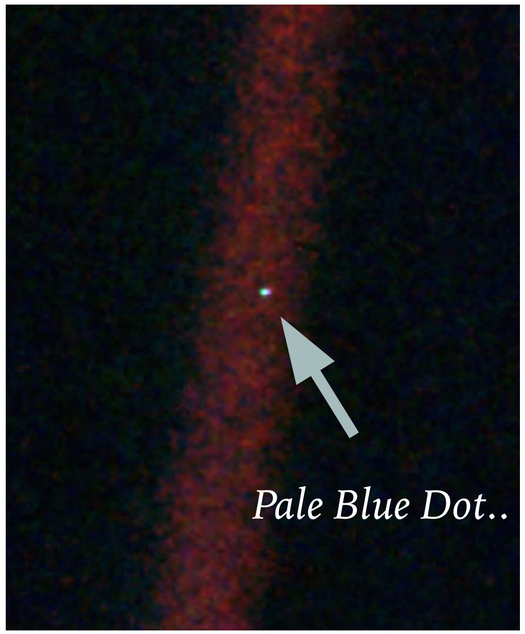
PHYSICS FROM AN END TO END SYSTEM

Giulio Eulisse (EP-AIP)

CHALLENGE: CONVERT MEGAHERTZ TO PAPERS AND PEOPLE

How do we convert 40+ MHz collision rate…

…for a population of a few 10K physicists across the globe…

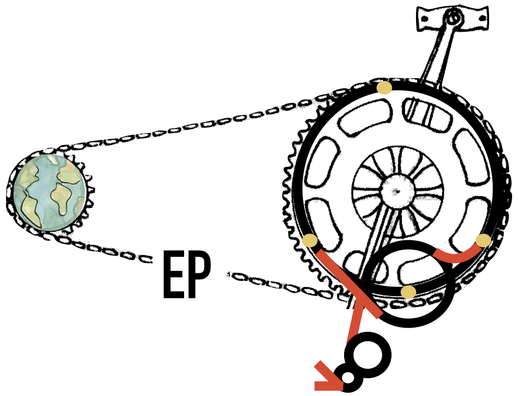


…to expand human knowledge…

WHY R&D FOR THE WHOLE END-TO-END CHAIN?

EP is end-to-end provider

Not only algorithms or plots: data acquisition, hardware / software integration, data & workflow management, software frameworks and toolkits.



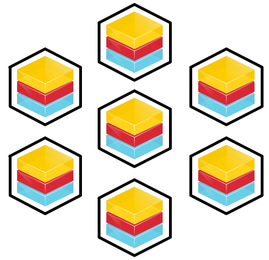
Cooperation

Being part of EP is all about collaboration with others. Worrying about end-to-end means worrying about integration with the rest of the world.

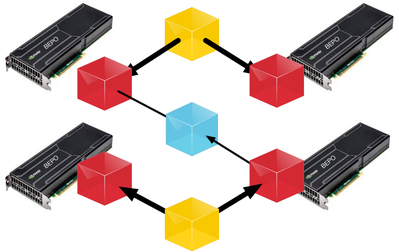
Modular solution(s)

Different design choices imply different trade-offs and might need different solutions. No "silver bullet", but modular ecosystems of interacting products.

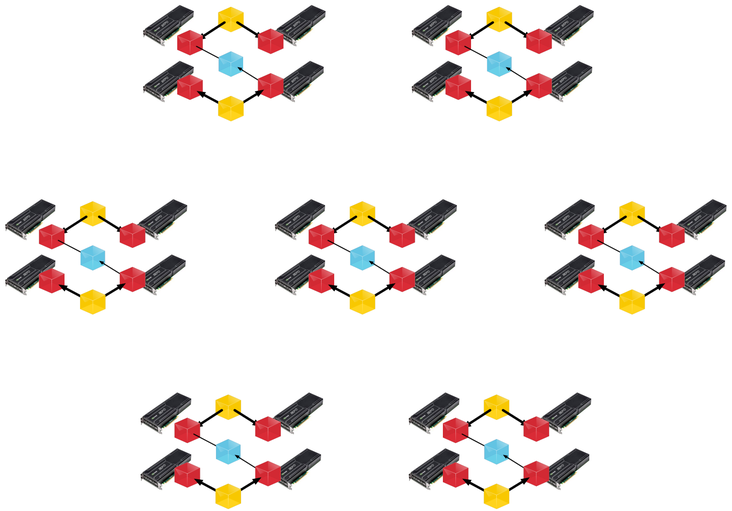
CHALLENGE: TACKLING EVOLVING COMPUTING INFRASTRUCTURE



From homogeneous, standalone resources…



…to heterogeneous datacenters. Blending of traditional Online and Offline roles (e.g. ALICE O2, LHCb)…



…actually a few of them, requiring negotiations with our WLCG partners…

UNKNOWN FUTURE: STAYING AGILE

Common trends among experiments:

- **Heterogeneous systems:** *different hardware depending on performed tasks (e.g.: GPUs, Tensor Units, low-power CPUs, FPGAs).*
- **Analysis facilities:** *few, well connected datacenters with dedicated general purpose clusters with high throughput interconnections between nodes.*
- **HPC-like resources:** *highly interconnected nodes which get most of their FLOPs from GPU-like hardware.*

Something completely different?

- **Opportunistic (commercial) clouds:** *cheaper computational resources, cost shifted to expensive connectivity / storage price. Could provide resources on demand.*
- **Distributed volunteer computing:** *unreliable in the past, can this be fixed by novel algorithms and an adequate business model?*

DATA MANAGEMENT

HEP computing is about data

*HEP is at the **forefront of scalability** needs for data management due to size and world wide collaborations.*
***Future experiments far more challenging** – increase in both data volume and number of objects to be stored.*

Data Lakes & Analysis Facilities

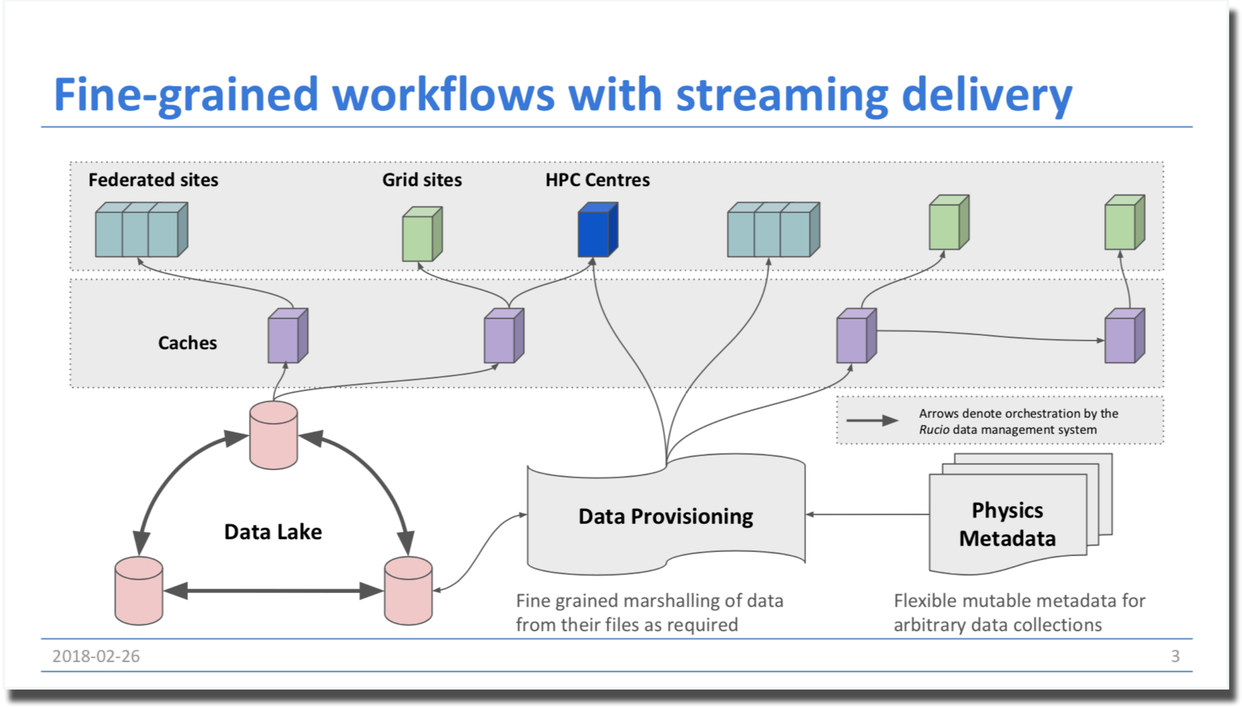
Fewer, well connected sites which act as authoritative source for caching layer seem to be a common trend for future designs.

Rucio

ATLAS solution for data management system should scale to Run3 needs. Looking ahead to Run4.

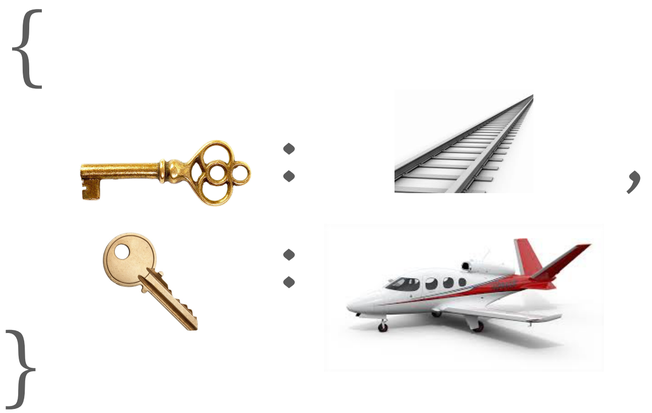
A collaborative effort

Championed, but not unique, to ATLAS. Other experiments expressing interest in it.



See M.Lassnig talk

ALTERNATIVES TO FILE-BASED ANALYSIS



HEP computing is about going through data, fast

File-based analysis has served us well. However, many hints we will be moving away from the operational sweet spot soon.

Exploring alternatives to scale further

*File-less alternatives, like key-value **object stores**, are a common solution to scale out data processing while **keeping system complexity under control**.*

Many applications

Not only event data, but also applicable to calibrations, quality control plots, monitoring.

Bridge technology to cloud ecosystems?

See D.Piparo & ROOT team lighting talk

INTEGRATION AND DEPLOYMENT OF MACHINE LEARNING EFFORTS

Technique of the future?

Machine Learning is a key problem-solving skill for the years to come. Optimised hardware could provide a factor 100x in performance.

Heterogeneous by design

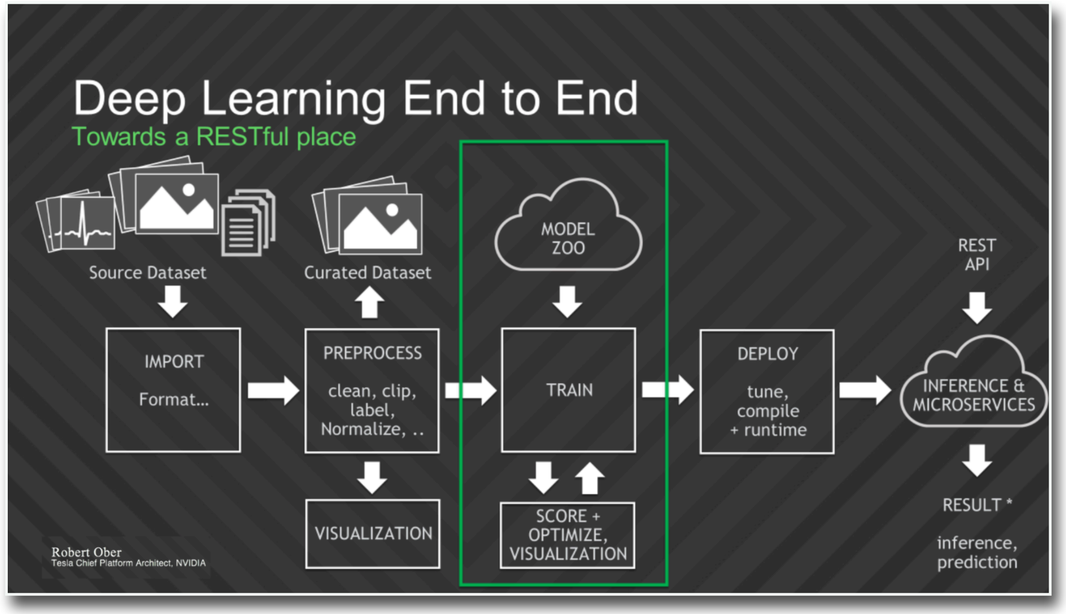
Once again, current ML / DL toolkit play extremely well with GPUs and custom accelerators.

Impedance mismatch

Address integration of our software frameworks with DL models in production. Not only data scientists but also data engineers!

Rapidly moving field

One of the challenges highlighted by previous discussions is that the field is suffering a "precambrian explosion" of tools and techniques.



See V. Innocente talk.

THANK YOU!

Performance optimisation

Data Management

Data Lakes

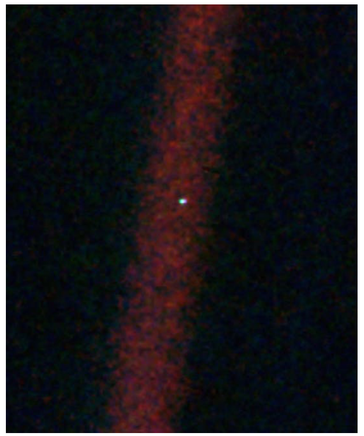
Novel Outreach Solutions

GPUs integration

TPUs

Key-value stores

Novel network fabrics



Heterogeneous infrastructure

Analysis Facilities

ML / DL Integration

Novel collaboration tools

Workload Management

Cloud computing

Improved analysis & simulation toolkits