Kubernetes as a Substrate for ATLAS Compute

Fernando Barreiro Megino (University of Texas at Arlington)
Lukas Heinrich (TU München)

KubeCon + CloudNativeCon Europe 2022, 18 May 2022, Valencia, Spain

ATLAS EXPERIMENT

UNIVERSITY OF TEXAS ARLINGTON

MDSI

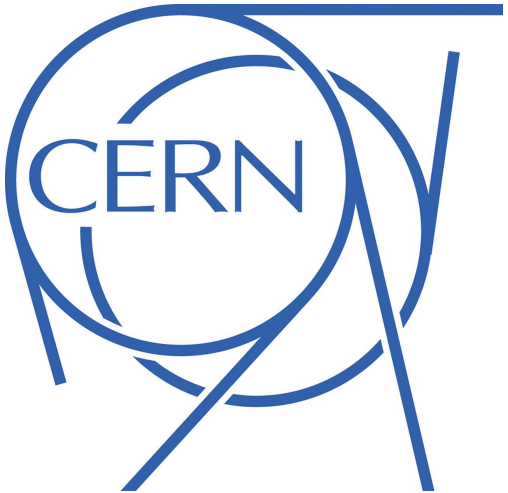
ODSL ORIGINS Data Science Lab

Presenters

Fernando Barreiro works as a Computing Engineer at the ATLAS Experiment at CERN. He is part of ATLAS' core team for workload management. In recent years he has led the effort to integrate public and private cloud resources through native usage of Kubernetes

Lukas Heinrich is a particle physicist working on the ATLAS Experiment at the Large Hadron Collider. He focuses on introducing modern cloud computing and data science tools to more systematically search for phenomena beyond the Standard Model of Particle Physics.

CERN: Where the Web was born



CERN DD/OC
Information Management: A Proposal

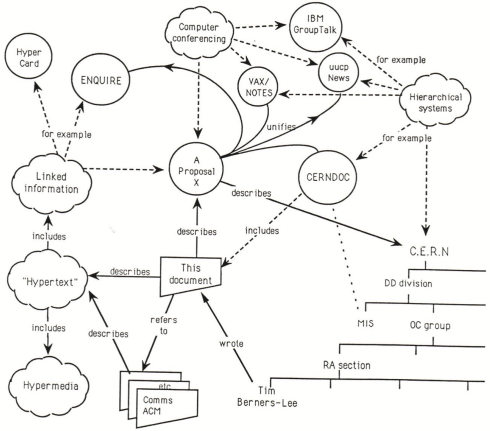
Tim Berners-Lee, CERN/DD
March 1989

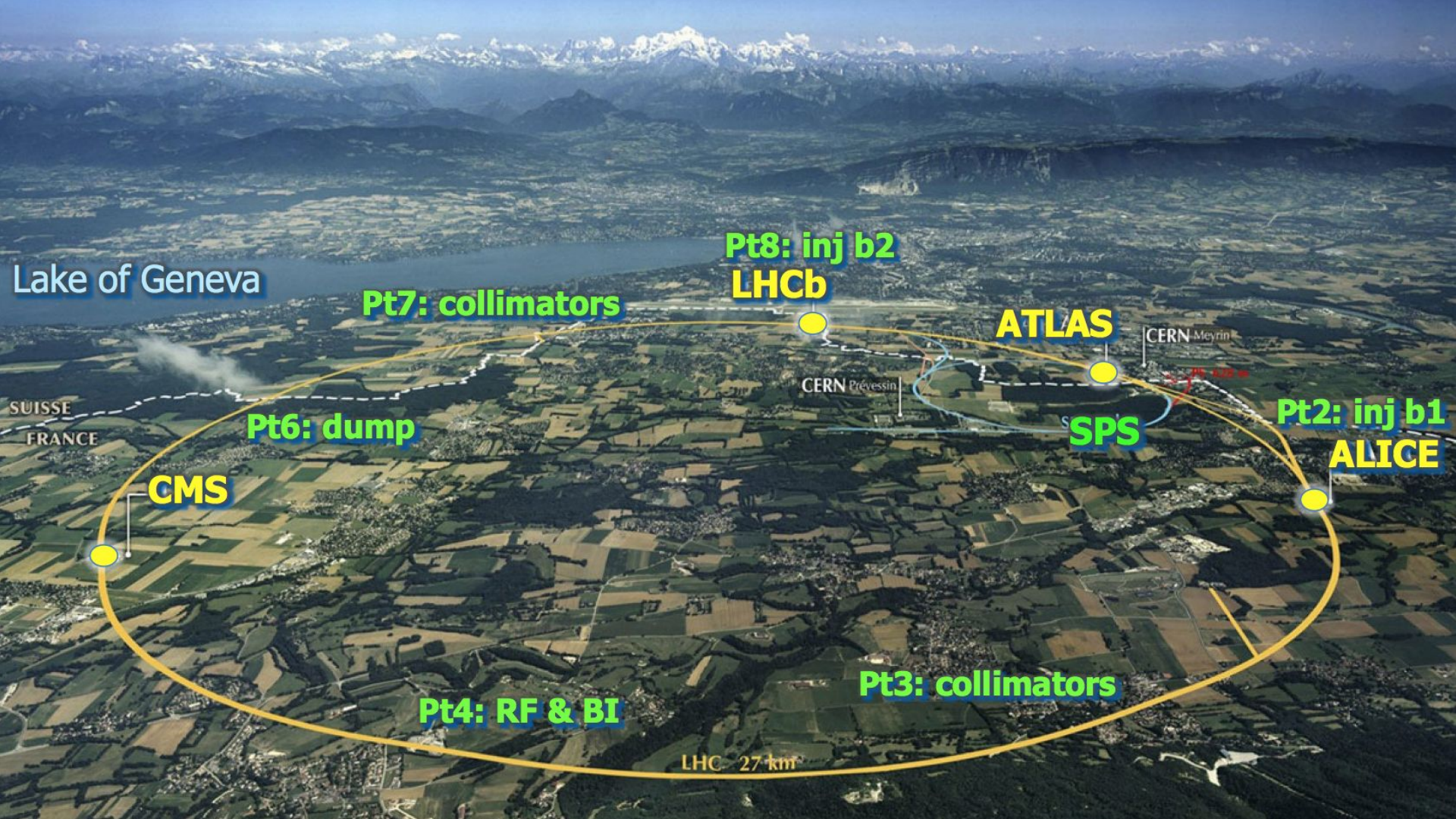
Information Management: A Proposal

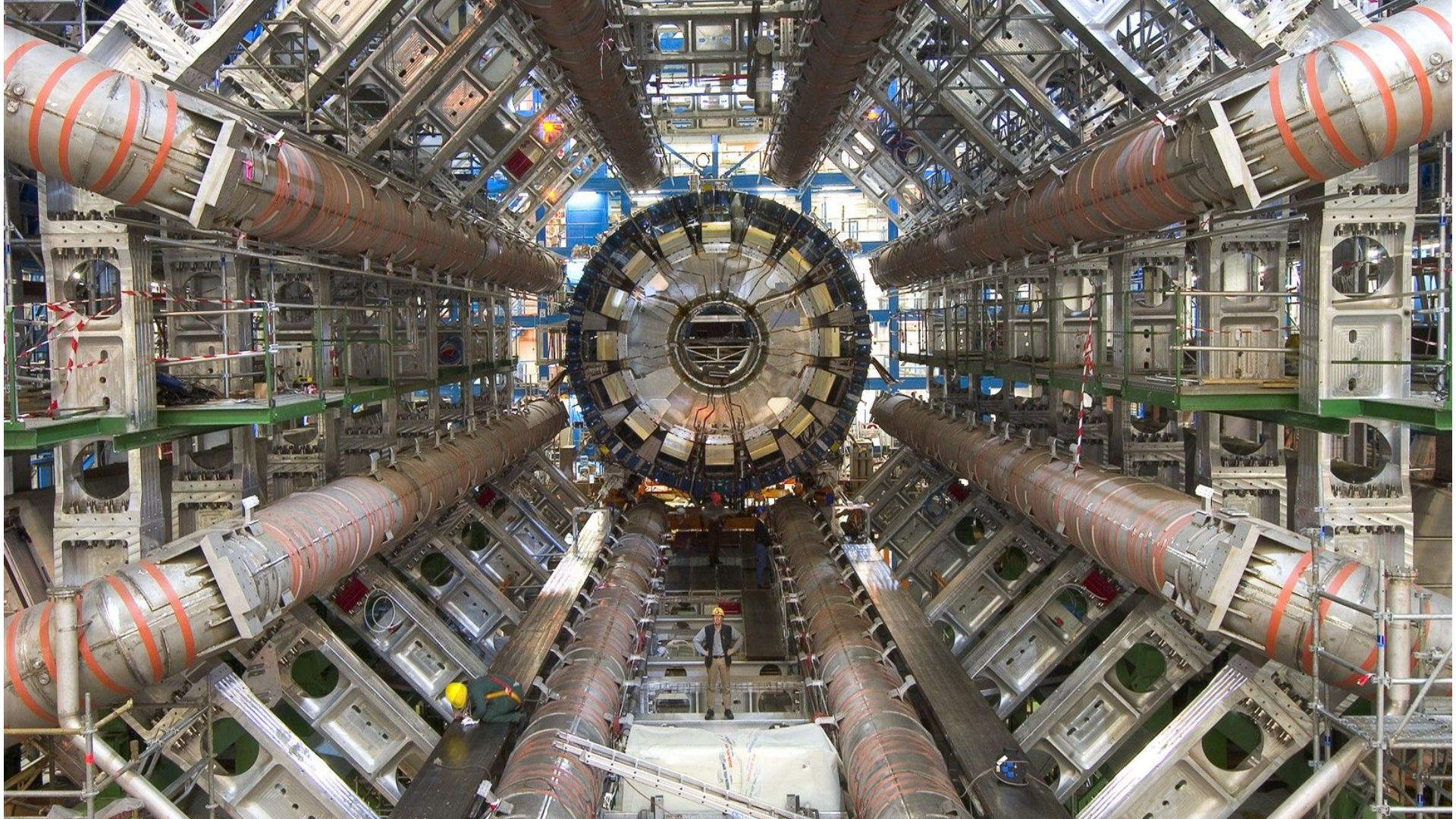
Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control




Lake of Geneva
Pt8: inj b2
LHCb
Pt7: collimators
ATLAS
CERN Meyrin
CERN Prévessin
SUISSE
FRANCE
Pt6: dump
SPS
Pt2: inj b1
ALICE
CMS
Pt3: collimators
Pt4: RF & BI
LHC 27 km



The New York Times

Physicists Find Elusive Particle Seen as Key to Universe

Scientists in Geneva on Wednesday applauded the discovery of a subatomic particle that looks like the Higgs boson. Pool photo by Denis Balibouse

By Dennis Overbye
July 4, 2012

← 10 years ago…

ASPEN, Colo. — Signaling a likely end to one of the longest, most expensive searches in the history of science, physicists said Wednesday that they had discovered a new subatomic particle that looks for all the world like the Higgs boson, a key to understanding

NEWS

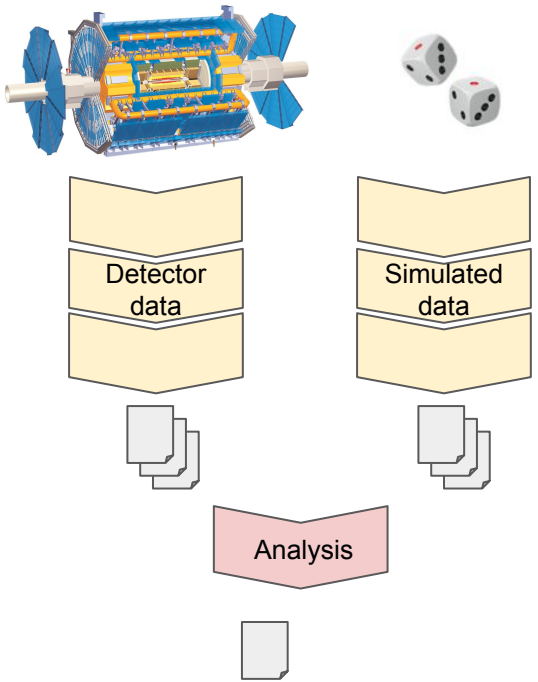
Science

Higgs boson scientists win Nobel prize in physics

By James Morgan
Science reporter, BBC News

8 October 2013 | Comments

ATLAS computing flow



“Production”

Exabyte scale

$O(10^5)$-$O(10^6)$ vCPUs

O(10) groups following organised campaigns

Turnaround expectation: weeks

Batch processing

“Analysis”

Petabyte scale

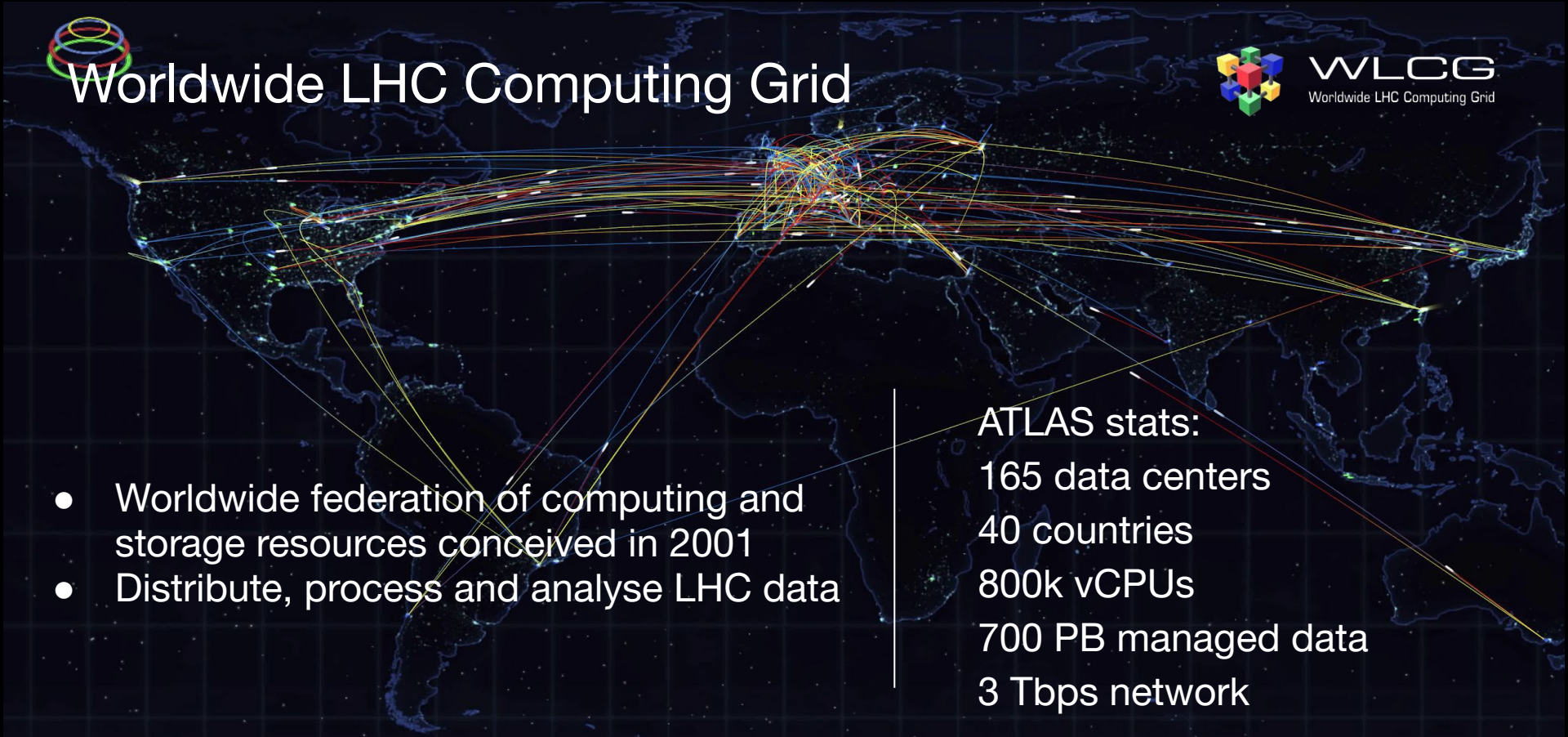
$O(10^5)$ vCPUs

O(100) simultaneous individual users

Turnaround expectation: ASAP / days

Batch and interactive processing

Worldwide LHC Computing Grid

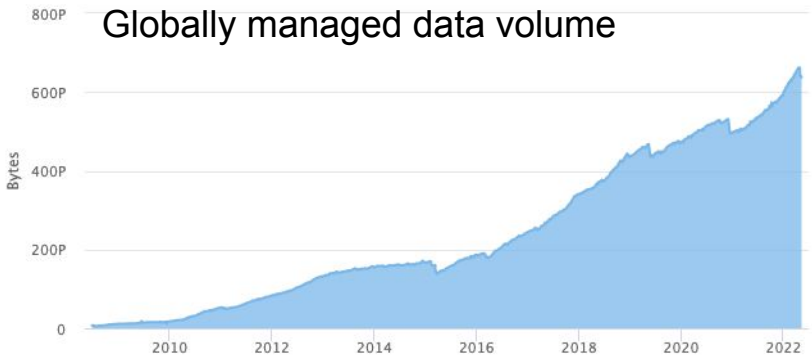


- Worldwide federation of computing and storage resources conceived in 2001
- Distribute, process and analyse LHC data

ATLAS stats:
165 data centers
40 countries
800k vCPUs
700 PB managed data
3 Tbps network

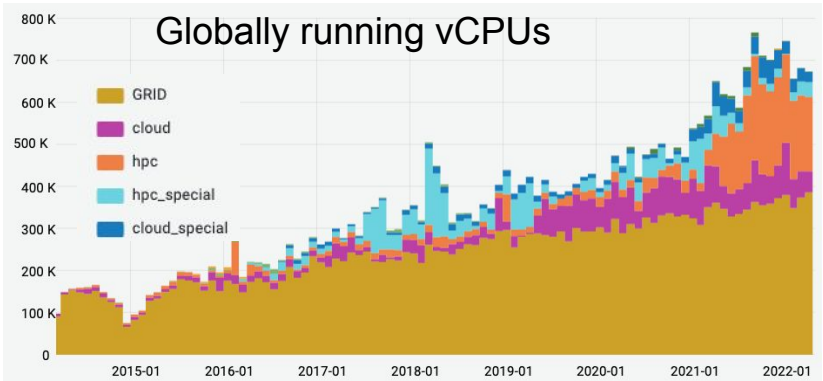
Image source: EGL application https://videos.cern.ch/record/2640380

ATLAS Distributed Computing



Data management: Rucio

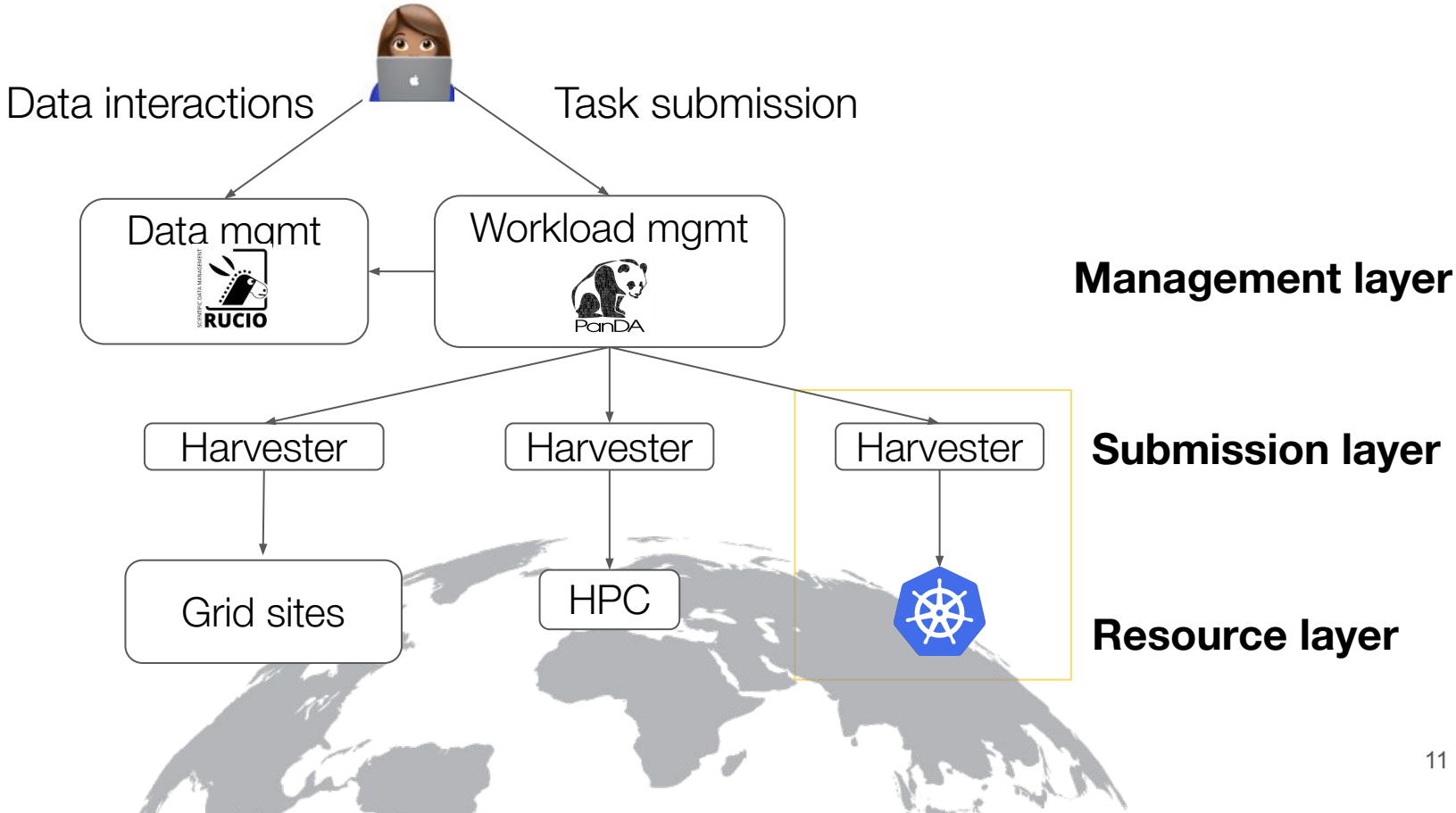
- Manage experiment data
- Interact with storage systems



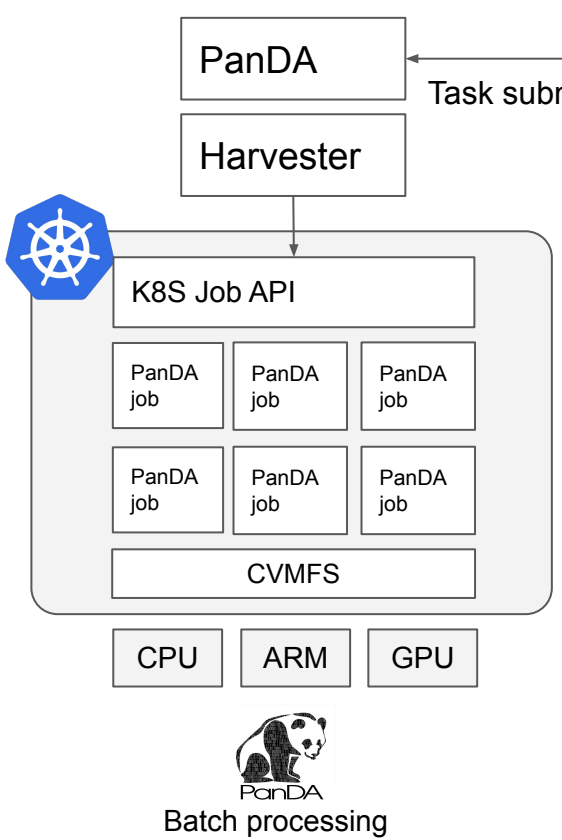
Workload management: PanDA

- Broker computational tasks
- Interact with compute systems (Harvester)

ATLAS job submission through Harvester

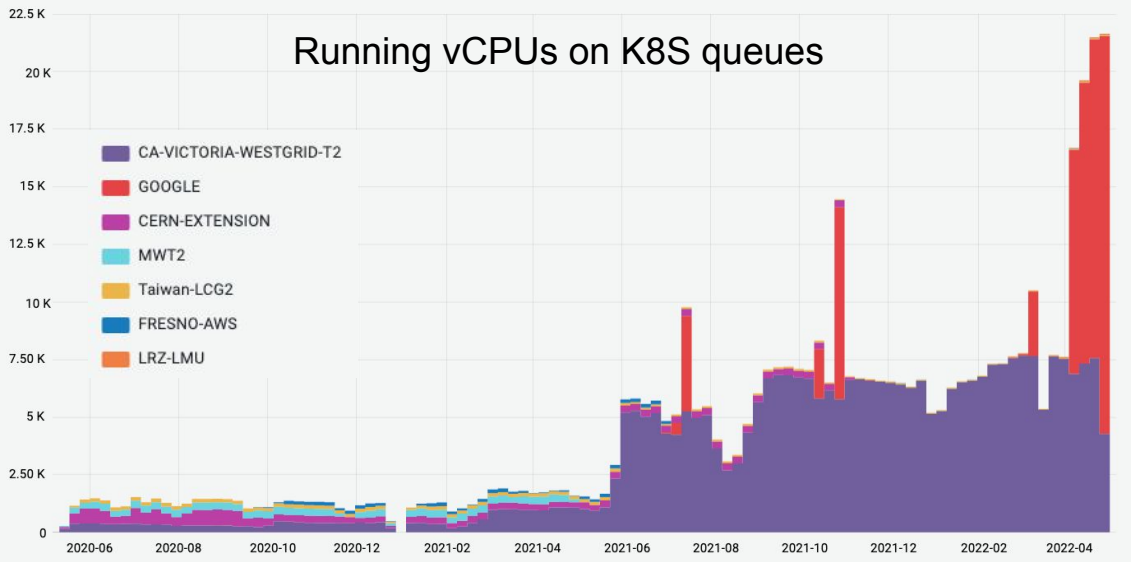


Grid processing: Harvester-Kubernetes integration



- Integrate K8S to work as any Grid site, while keeping setup as simple as possible
- Purely based on native K8S Job controller
- Harvester plug-ins for job submission & monitoring
- Usage of K8S common options
  - Limits for CPU, memory, disk, duration
  - Pod affinity to pack nodes
  - Priority classes
- High Energy Physics filesystem: CVMFS daemonset

Onward & upward: on prem and cloud K8S queues

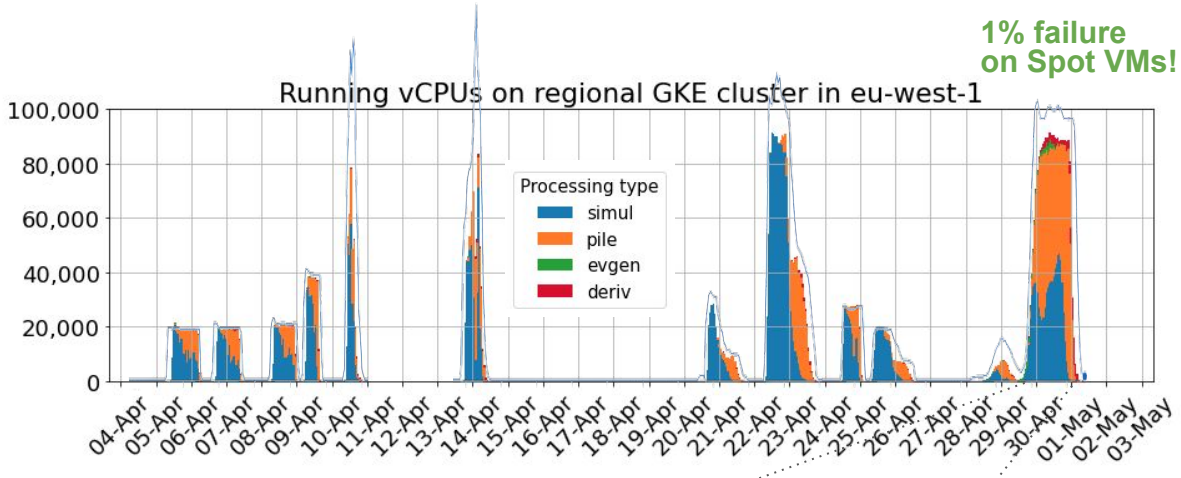


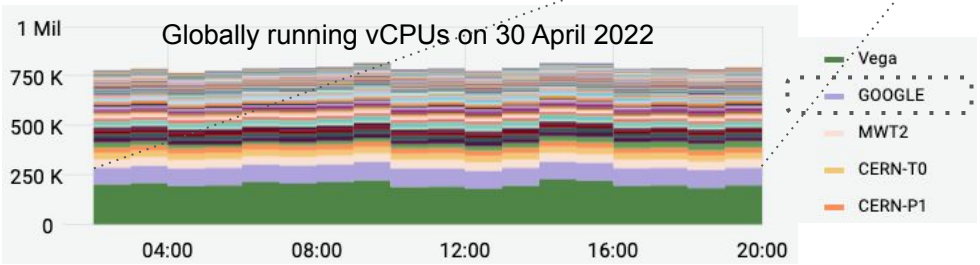
Mini K8S Grid scaled from few hundred vCPUs to many thousands vCPUs

Provider independent integration

University of Victoria, arbutus cloud, Google Cloud, CERN, THE UNIVERSITY OF CHICAGO, ASGC, FRESNO STATE, aws, lrz Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

Elastic cloud scale out





- O(10^8) events
- O(10^5) vCPUs
- O(10^4) Pods
- O(10^3) Nodes
- O(1) day
- O(1) Harvester instance
- <1 Engineer

Image credits: ICT Network News

Integration of heterogeneous architectures

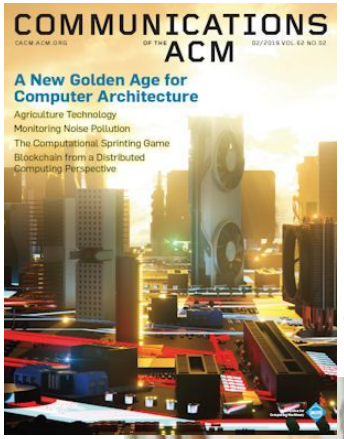
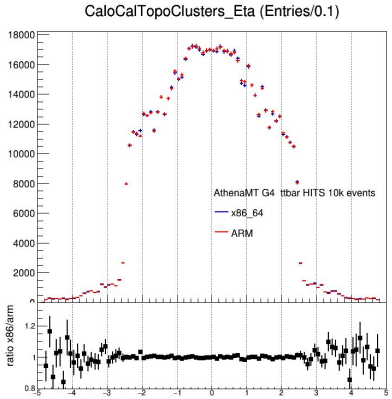
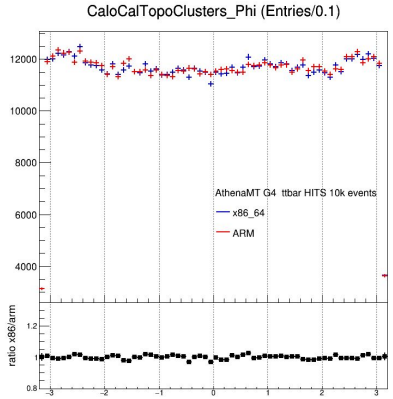


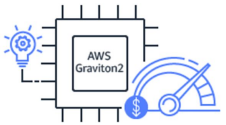
Image credits: Communications of the ACM, February 2019

Image credits: NVIDIA GTC May 2020 Keynote

- Cloud queues backed by resources not commonly available on prem
- Straightforward to integrate different architectures, e.g. ARM, GPU
- Multi-arch Docker images doing the heavy lifting

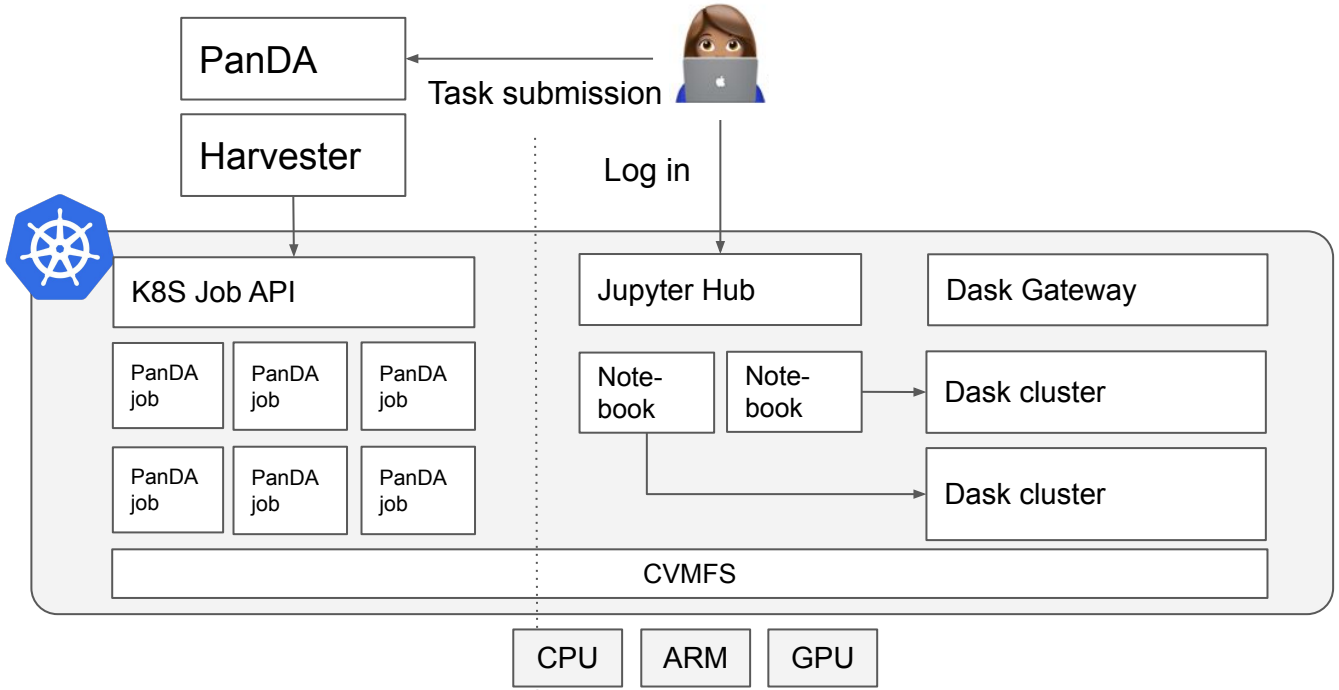






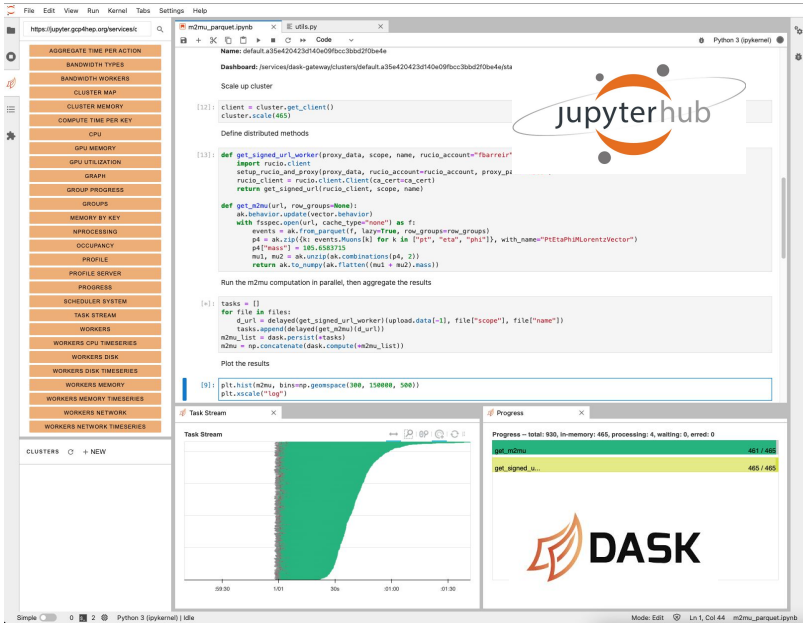
First ATLAS simulation task on ARM processors. Currently under physics validation. Generated on Amazon EKS cluster backed by Graviton 2 nodes.

Adding interactive analysis facilities

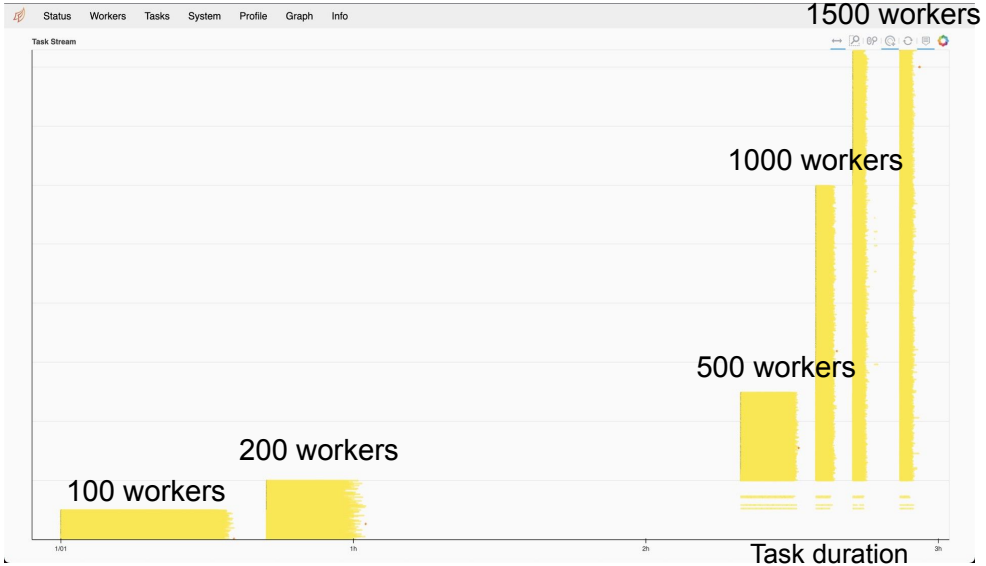


Distributed, interactive analysis

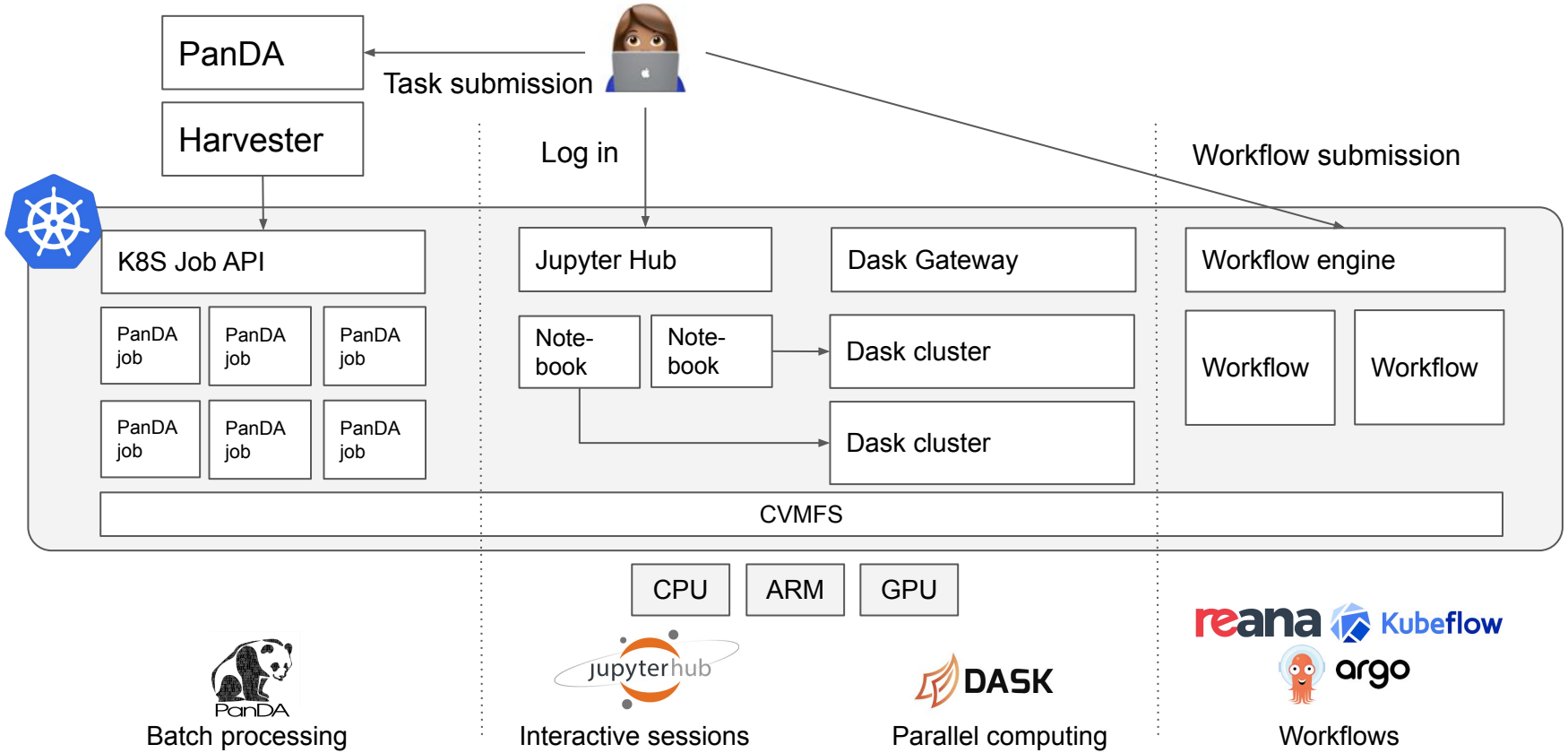
- Data science community technologies
- Helm charts available: DaskHub
- Configured for scalability and cost effectiveness on GKE



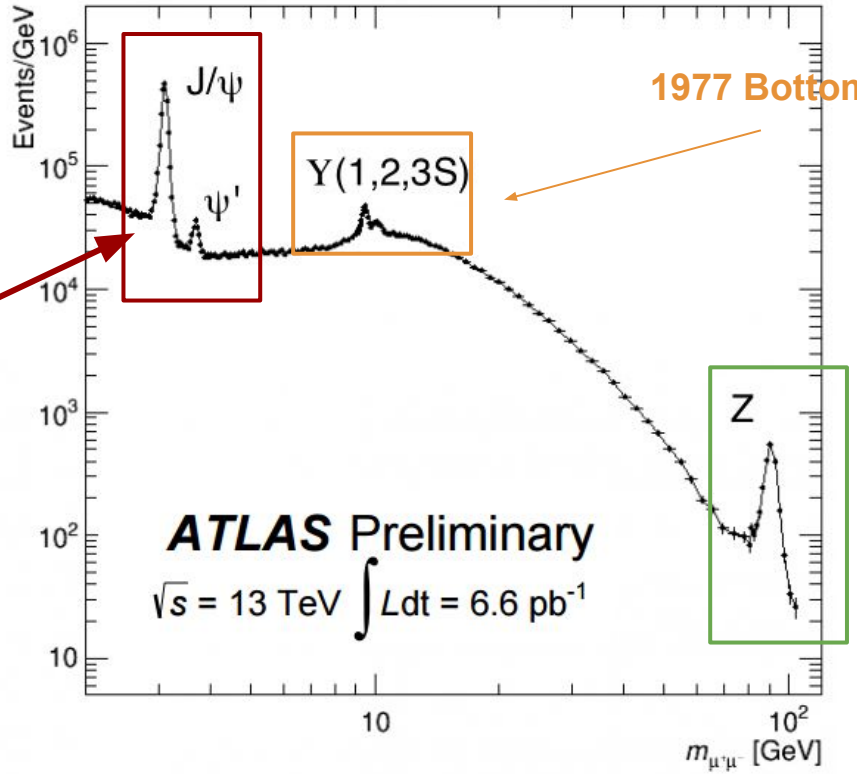
N workers



K8S substrate for ATLAS compute: idealistic view



Demo: History of Particle Physics in a single Graph

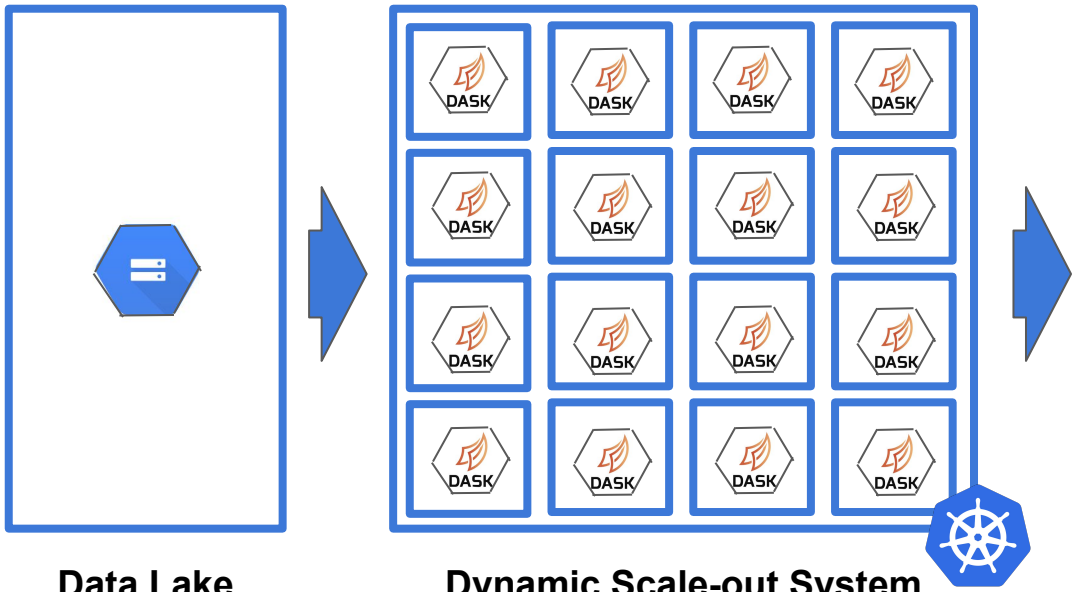


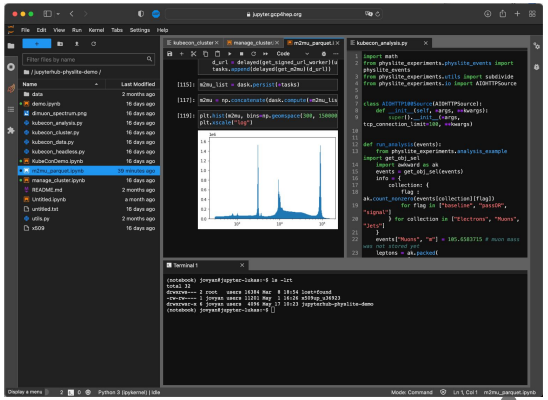
1976 Nobel Prize

1977 Bottom Quark Discovery

Z-Boson
1984 Nobel Prize

Demo: History of Particle Physics in a single Graph





Data Lake

100 TB of Data

Dynamic Scale-out System

1-10k vCPU/User

User-facing UI

Conclusions

- Kubernetes goes beyond service management
  - Demonstrated K8S native ATLAS batch processing at 100k cores
    - Compatibility to existing infrastructure
    - Reaching scale further than originally planned
  - Next generation, interactive services with high elasticity
  - Other functionalities that can be added
- Next years will determine K8S integration
  - Zero-to-grid-site via GitOps & Helm?
  - Elasticity of resources and availability of “exotic” resources
  - Managed K8S clusters simply work
  - Deploy K8s inside of HPC & university settings
- A lot of work still to be done, but it is a very promising environment

Acknowledgements

Karan Bhatia [4], Misha Borodin [6], Kaushik De [7], Johannes Elmsheuser [2], Miles Euell [4], Nikolai Hartmann [5], Alexei Klimentov [2], FaHui Lin [7], Tadashi Maeno [2], Usman Qureshi [4], Ricardo Rocha [3], Ming-Jyuan Yang [1]

1. Academia Sinica
2. Brookhaven National Laboratory
3. CERN
4. Google
5. Ludwig Maximilian University
6. University of Iowa
7. University of Texas at Arlington

Links

CVMFS plugin: https://github.com/sfiligoi/prp-osg-cvmfs, https://github.com/PanDAWMS/prp-osg-cvmfs
DaskHub: https://github.com/dask/helm-chart/tree/main/daskhub
Dask Gateway: https://gateway.dask.org/
Harvester: https://github.com/HSF/harvester/
PanDA: https://panda-wms.readthedocs.io/
Rucio: https://rucio.cern.ch/