Coffea-Casa: building composable analysis facilities for the HL-LHC

Sam Albin, Ken Bloom, Oksana Shadura,
Garhan Attebury, Carl Lundstedt, John Thiltges
University of Nebraska, Lincoln, USA

Brian Bockelman
Morgridge Institute, Madison, USA

26th International Conference on Computing in High Energy & Nuclear Physics

Reshaping physics analysis for HL-LHC

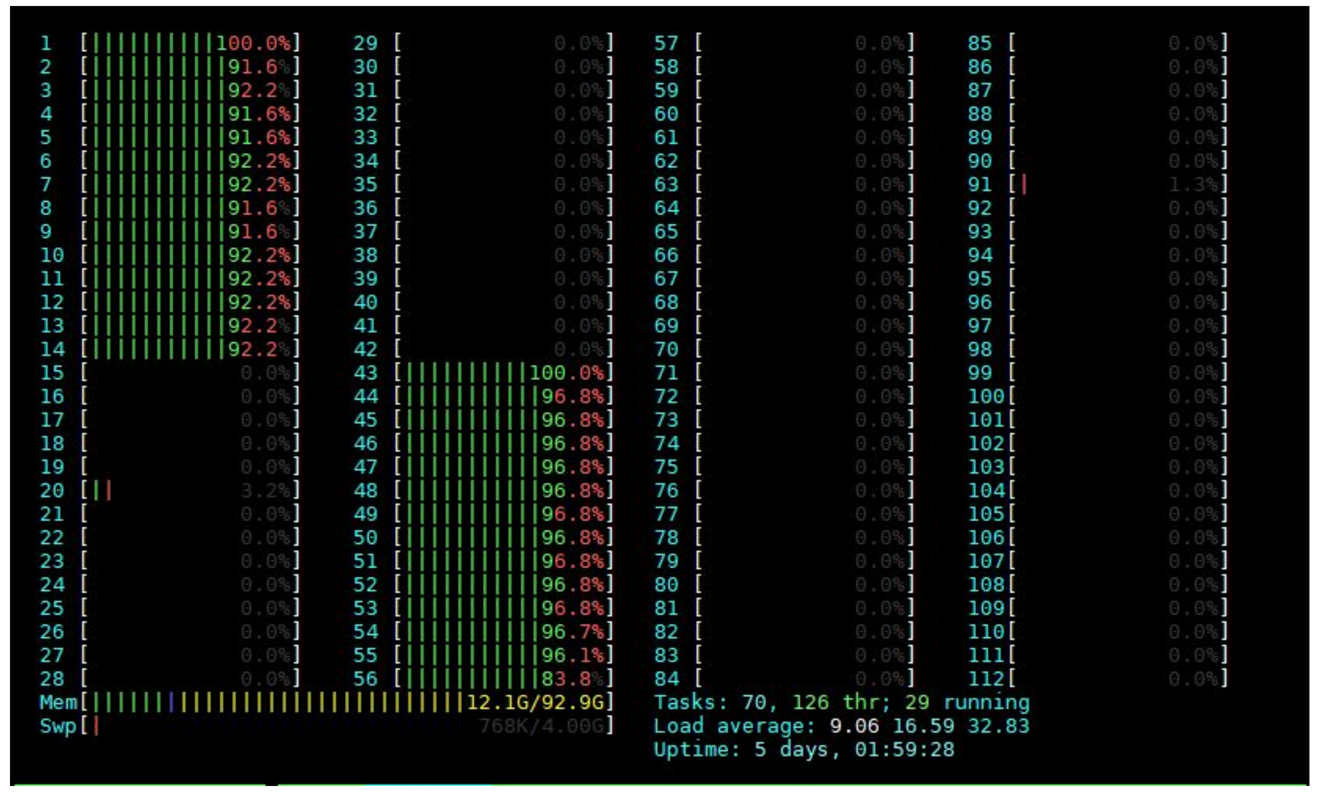
	LHC (Run 1 and Run 2)		HL-LHC (Run 4+)
Analysis dataset size	**10 TB**		**1000 TB**
Target scan turnaround time	Weeks		Hours
Minimum analysis team size	5-10	Run 3	< 5
Typical analysis resource	**….**		**Analysis Facilities**
Analysis specifics	• Search & Precision Physics • Simple ML techniques (BDT) • Basic analysis reproducibility		• Very High Precision Physics • Modern ML (Deep Learning) • Reproducible and Open Data

HEP Analysis Facilities

How the physicists see "Analysis Facility":

HEP Analysis Facilities are usually used for end-user analysis

Homelab (https://domalab.com)



"Analysis facility" could be any type of resource from laptop to Tier-2

- HEP data access
- Number of cores to scale
- Recipe how to run code
- Disk space
- Favorite analysis framework already available

HEP Analysis facilities: what the physicist dreams about

- **Quick interactive analysis turnaround:** *"I want to get my preliminary plots to be ready over coffee break"*
- **User improvement experiences (UX):** let's help physicists focus on the physics
- **Methods for efficient data scaling, caching at AFs:** more challenges with data-intensive analysis pipeline
- **Data reusability:** AF should support extraction of user defined experiment data formats to migrate them onto other facility, laptops or workstations at home institutions or at home

HEP Analysis Facilities

How the resource managers see it:

“Analysis facility” could be any type of managed computing / storage resources shared between multiple users used for end-user analysis

Fast network connection	Orchestration, Provisioning & Configuration Management	Multitenant VO support	Distributed disk storage system	Batch
		…		

HEP Analysis facilities: what resource manager dreams about

- **Easy deployment and reproducible setup:** Kubernetes can help to facilitate an easy AF deployment within Tier-X facilities (e.g. co-locating next to existing computing resources)
- **Modularity:** Kubernetes is ideal for demanding applications that require complex configurations (focusing on modular orchestration)
- **“Self-healing”**: easy rollback with Kubernetes

Building blocks used for designing AFs

- Columnar analysis and support new pythonic ecosystem
- Efficient data delivery and data management technologies
- Machine learning services and tools
- Efficient data caching solutions
- Support for object storage
- Easy integration with scalable computing resources
- Modern authentication (IAM/OIDC), tokens, macaroons
- Modern deployment and integration techniques

Computing resources available for end-user analysis during Run-2: are they all Analysis Facilities?

Community supported resources

Grid

Lxplus

LPC CAF

University Tier3

University Tier3

University Tier3

University Tier3

University cluster

University cluster

University cluster

Sometimes complex for user:
different configurations, different way to access, cannot easily move from one facility to other, different interfaces, different scaling mechanisms, lack of documentation, not suitable for interactive analysis

Personal computing resources

Laptop

Doesn't scale easily

Tier-1

Very rarely / not available for end users

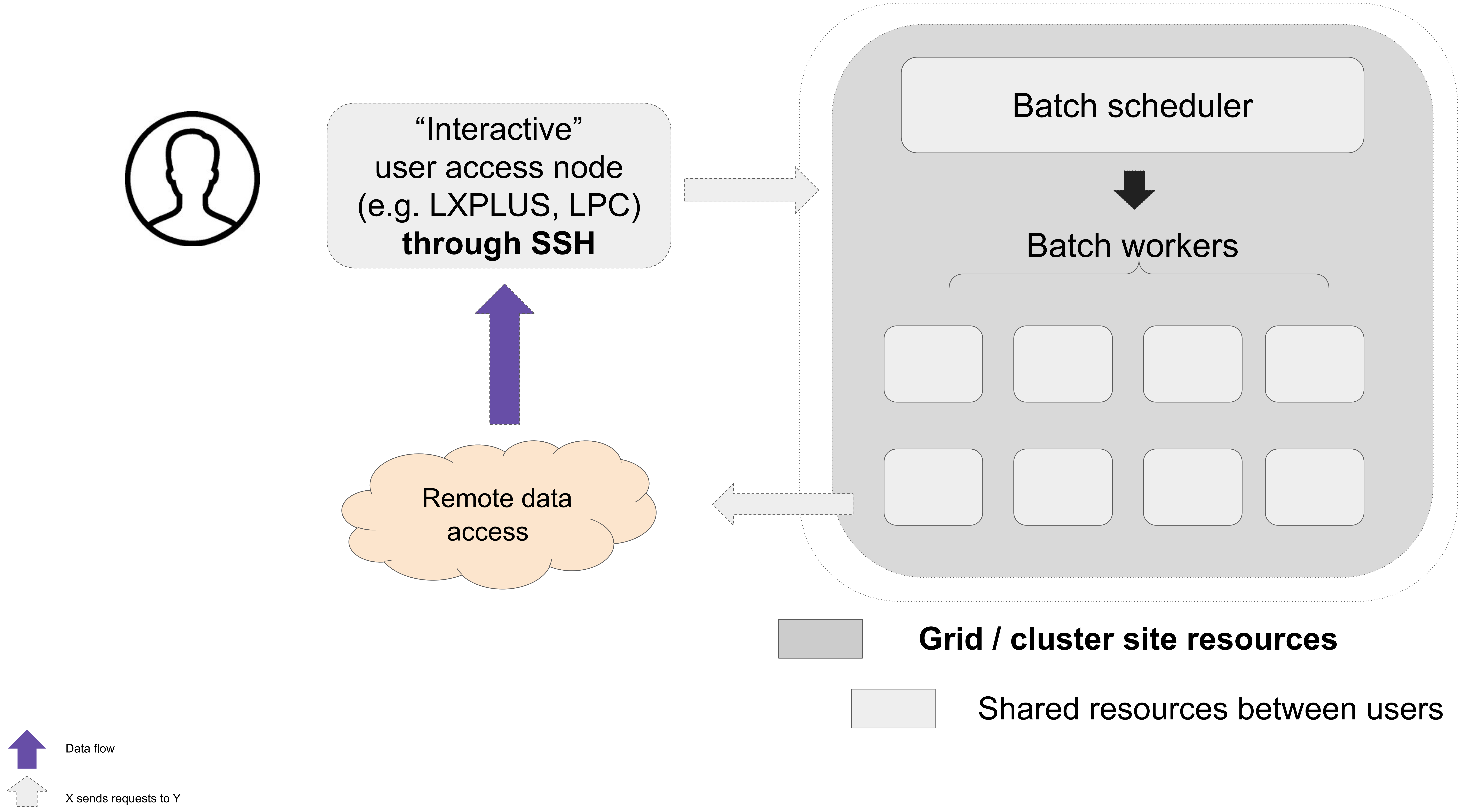
Tier-2

Very rarely / not available for end users

AWS, Google Cloud, Kubernetes (private and public clouds)

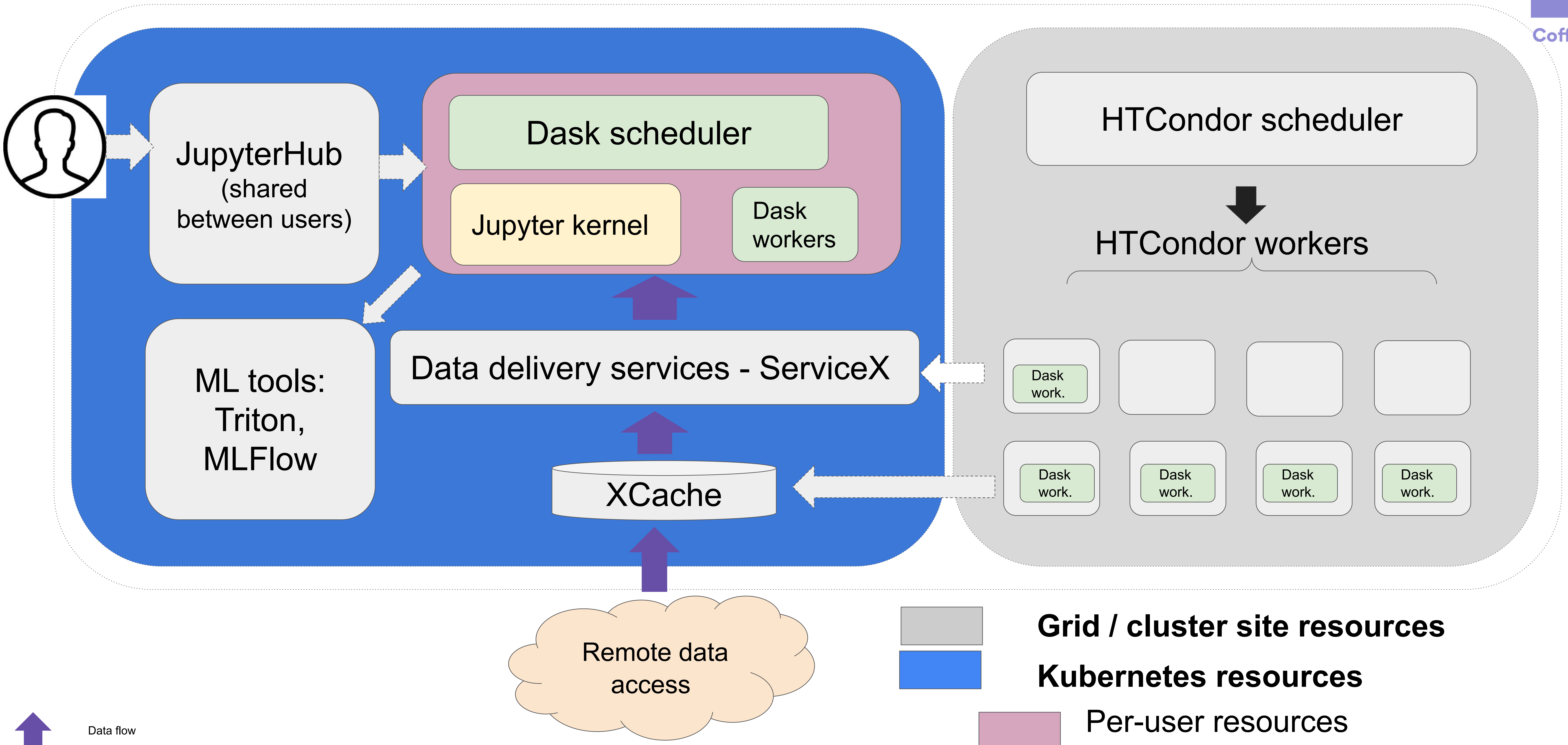
Very rarely available for end users ($$$)

Simplified diagram of hypothetical Analysis Facility currently used by user



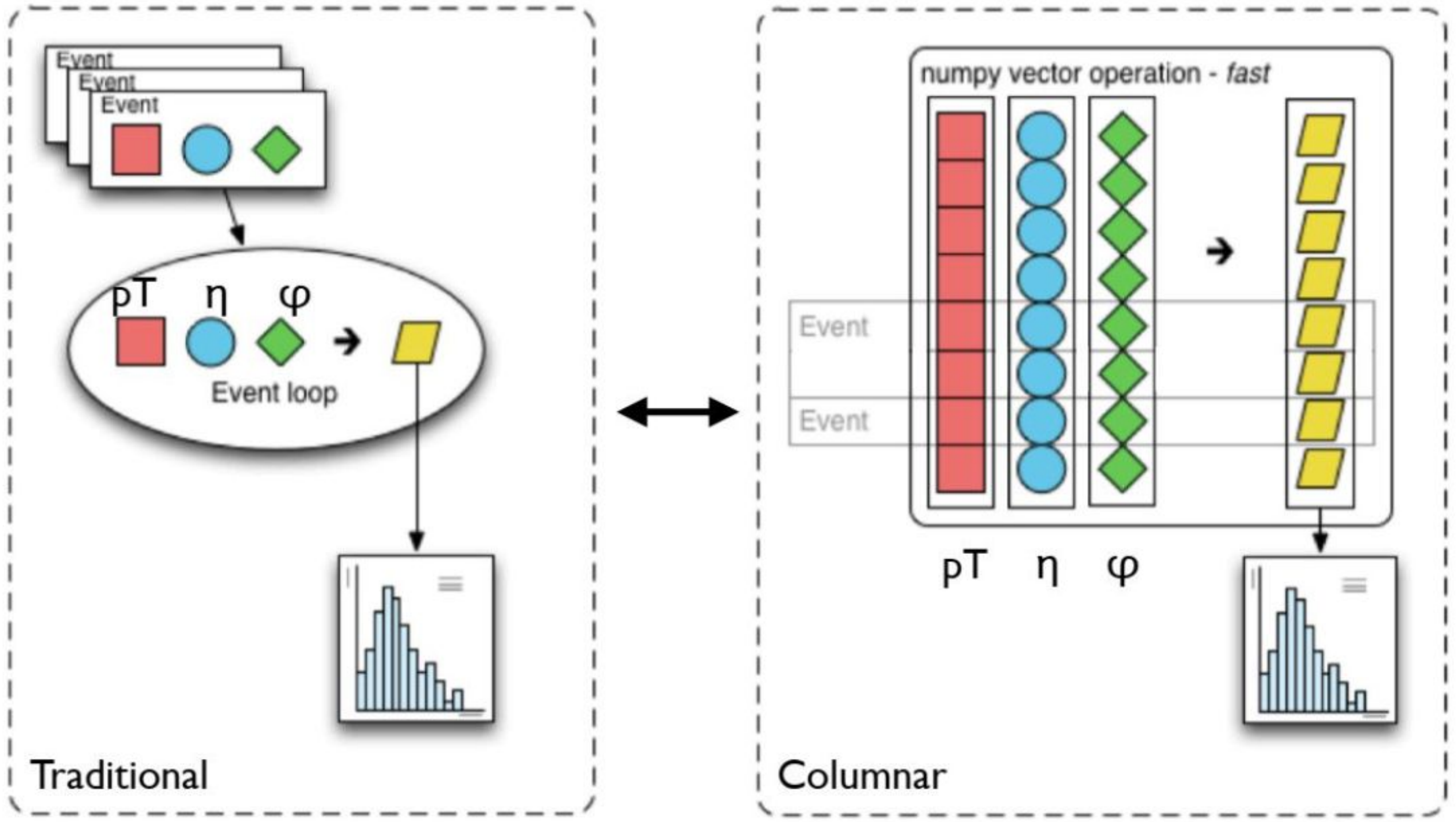
Coffea-casa Analysis Facility

Coffea-casa facility @ UNL is co-located at U.S.CMS Tier-2 at University Nebraska-Lincoln and other instance is co-located at U.S.ATLAS Tier-3 at University UChicago



https://iris-hep.org/projects/coffea-casa.html

Building blocks: columnar analysis and support new pythonic ecosystem



Coffea Analysis Framework

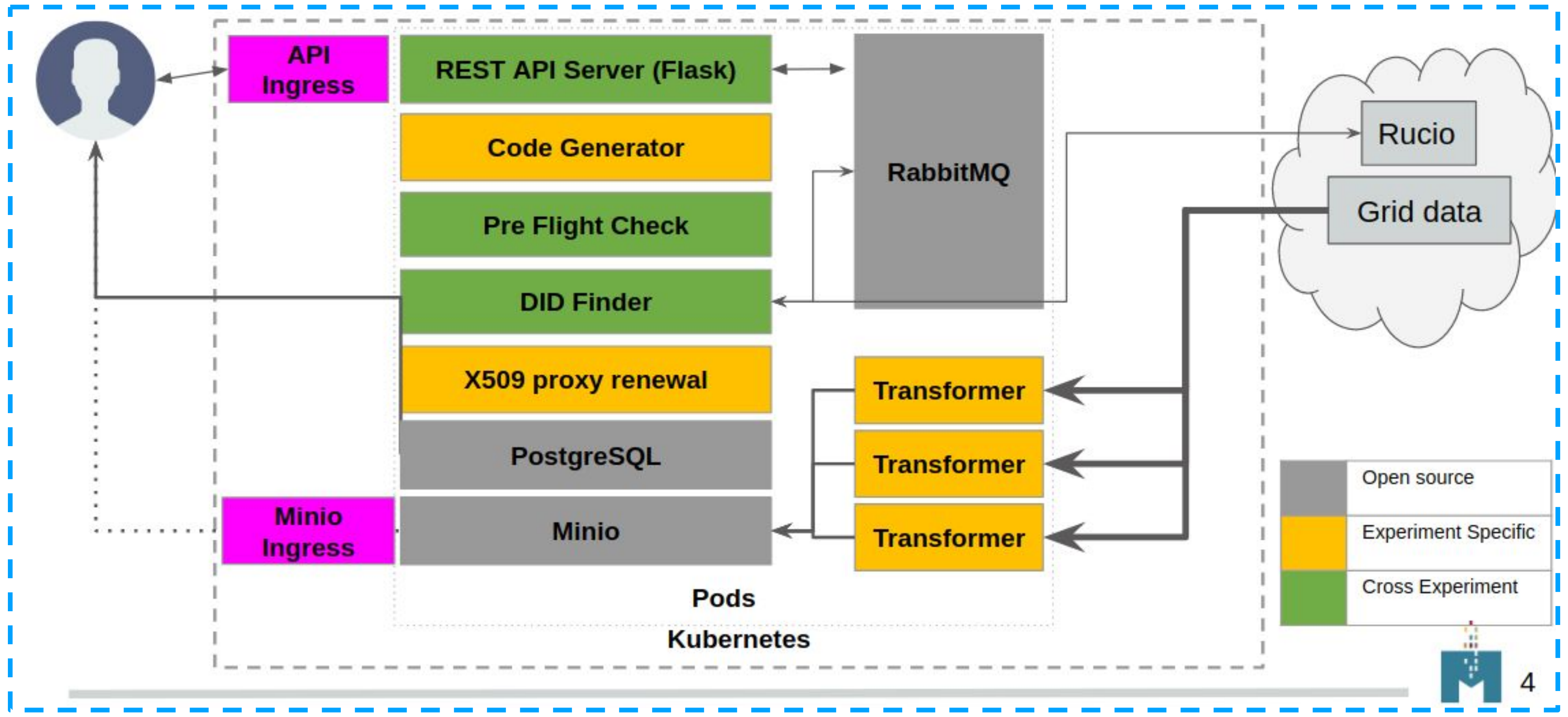
ROOT RDataFrame

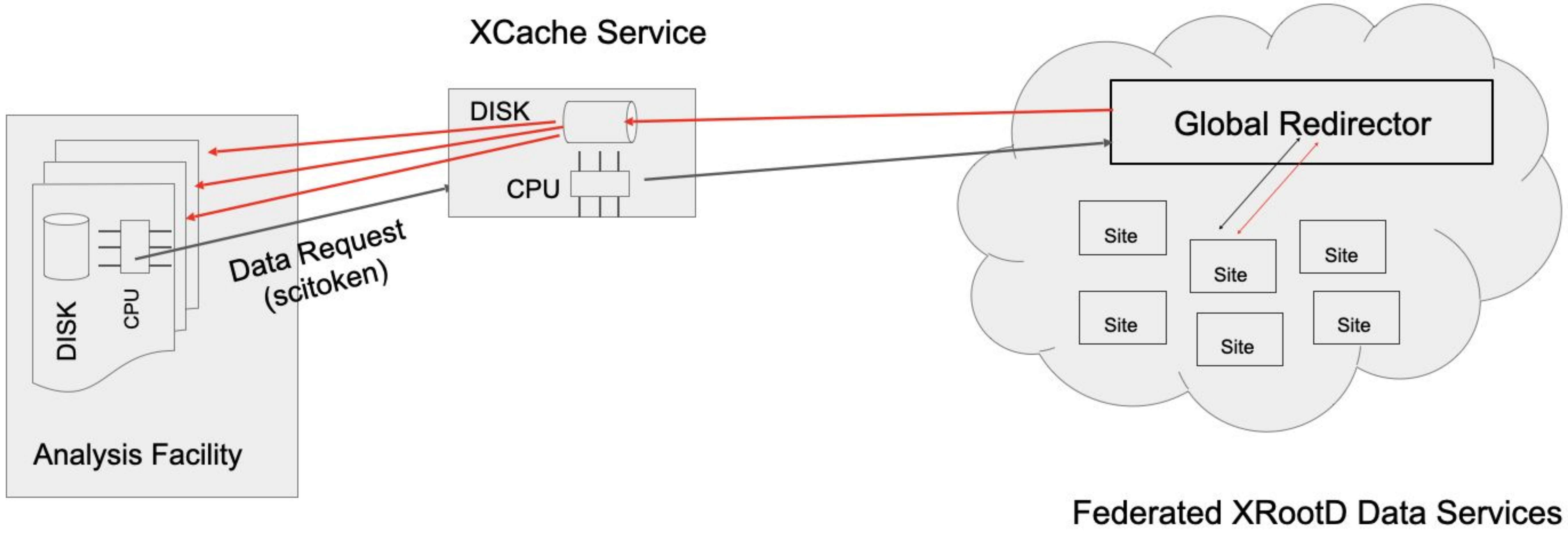
New columnar data analysis concepts!

New analysis frameworks!

Distributed executors!

Building blocks: data delivery and data management technologies

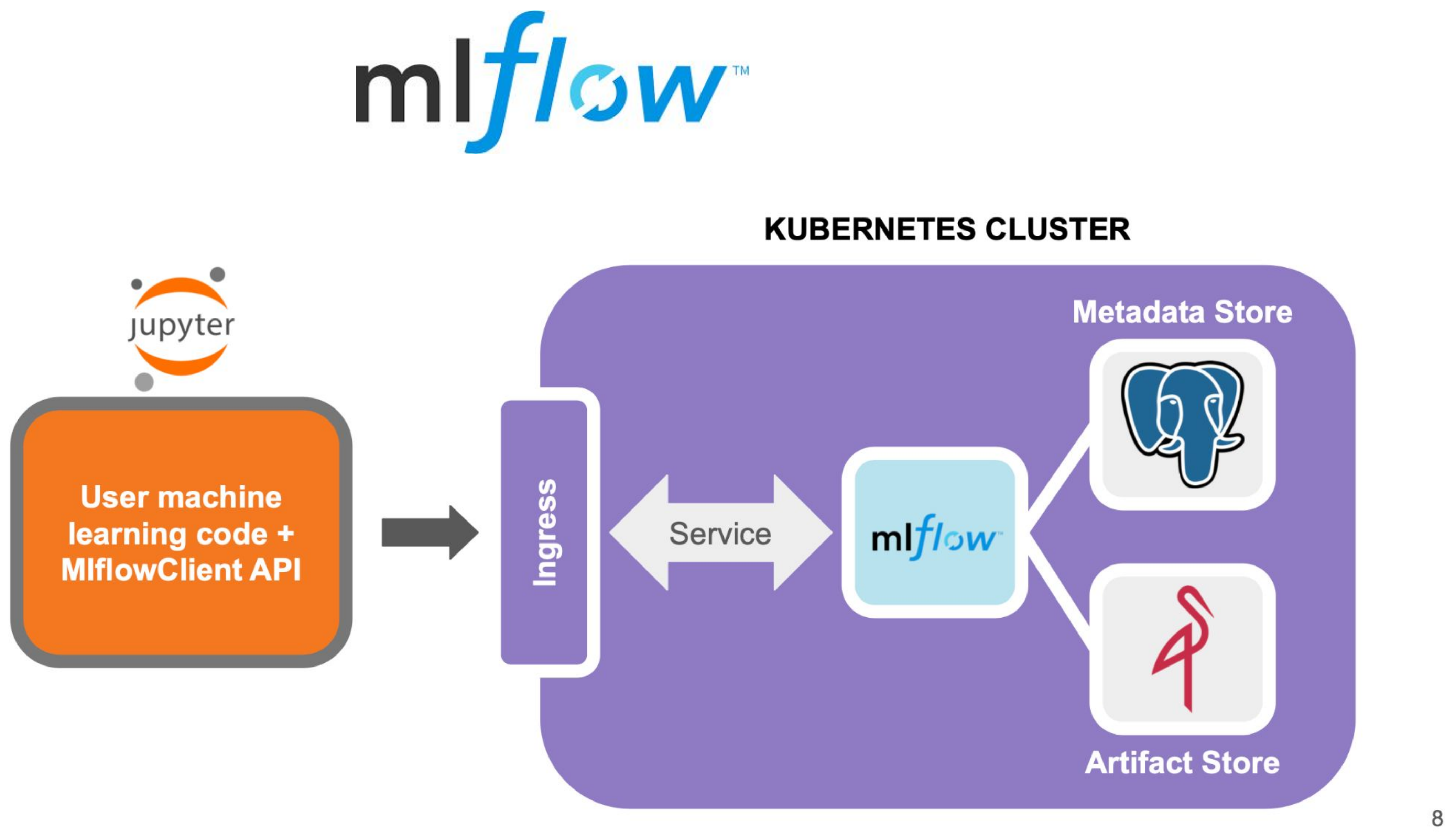


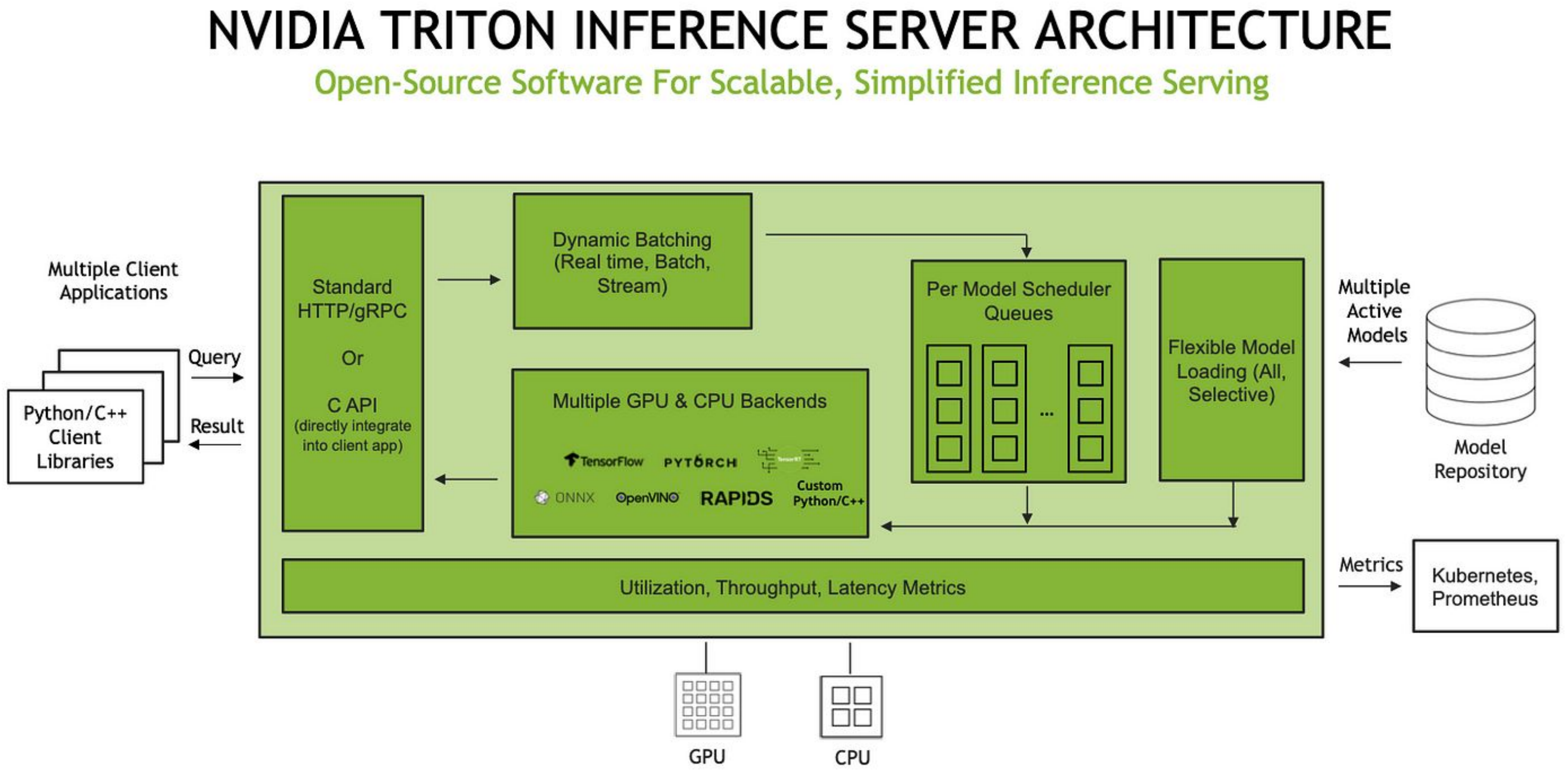


ServiceX - data extraction and data delivery service for columnar analysis (developed by IRIS-HEP DOMA))

XCache - cached-based placement of analysis datasets

Building blocks: machine learning services and tools





https://developer.nvidia.com/nvidia-triton-inference-server

- Provide a central store to manage models (their versions and stage transitions)
- Allow packaging and re-deploying models
- Allows easily to reproduce code
- Provides easy tracking of ML experiments

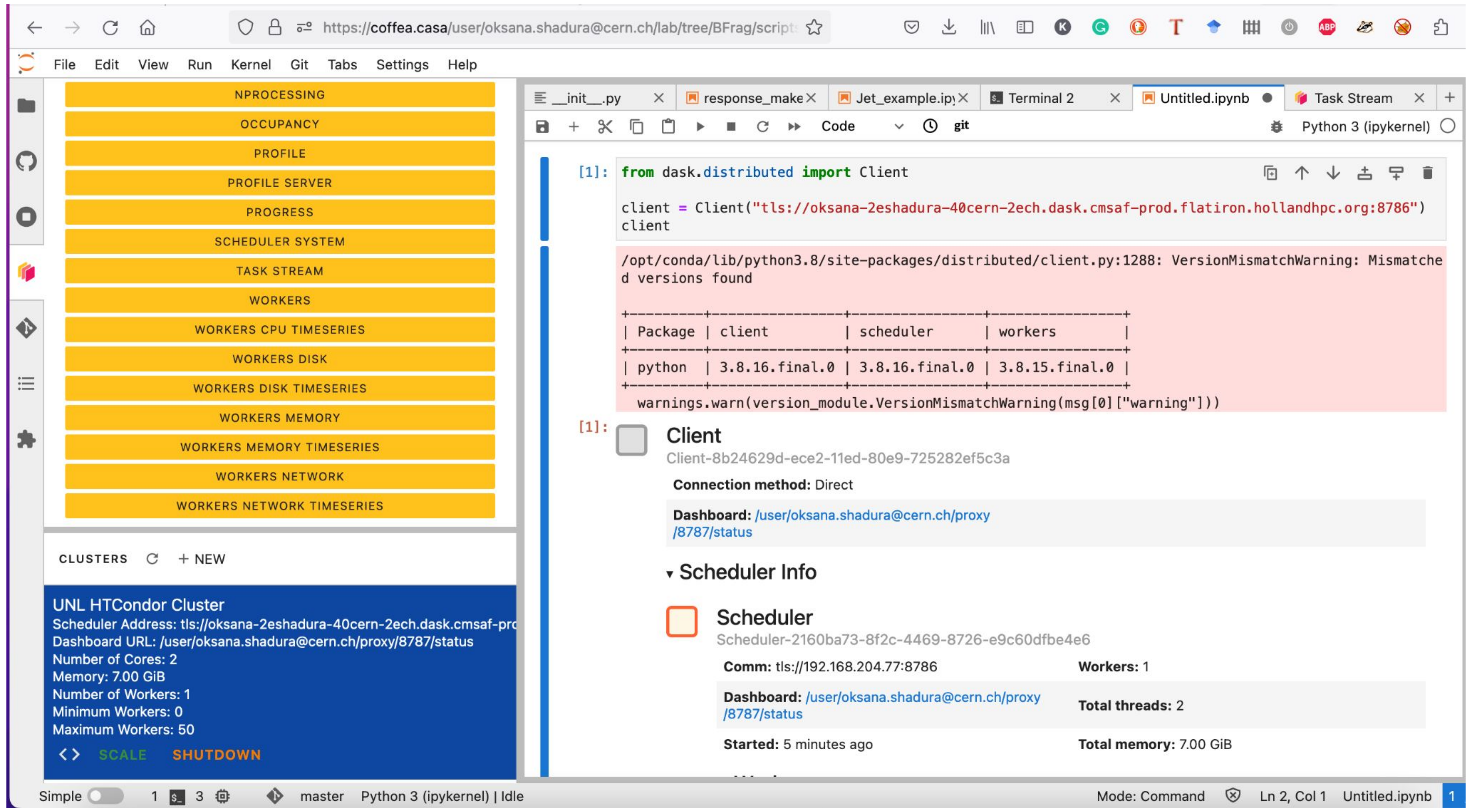
- **Support for various deep-learning (DL) frameworks**
- **Simultaneous execution** - Triton can run multiple instances of a model, or multiple models, concurrently, either on multiple GPUs or on a single GPU.
- **Dynamic scheduling and batching**

Check the talk from Elliott Kauffman "Machine Learning for Columnar High Energy Physics Analysis"

Building blocks: easy integration with scalable computing resources

Dask provides a task-management computational framework in Python (based on the manager-worker paradigm)

- Integrates with HPC clusters, running a variety of schedulers including SLURM, LSF, SGE and HTCondor via ***"dask-jobqueue"***
- **This allows us to create a user-level interactive system via queueing up in the batch system**



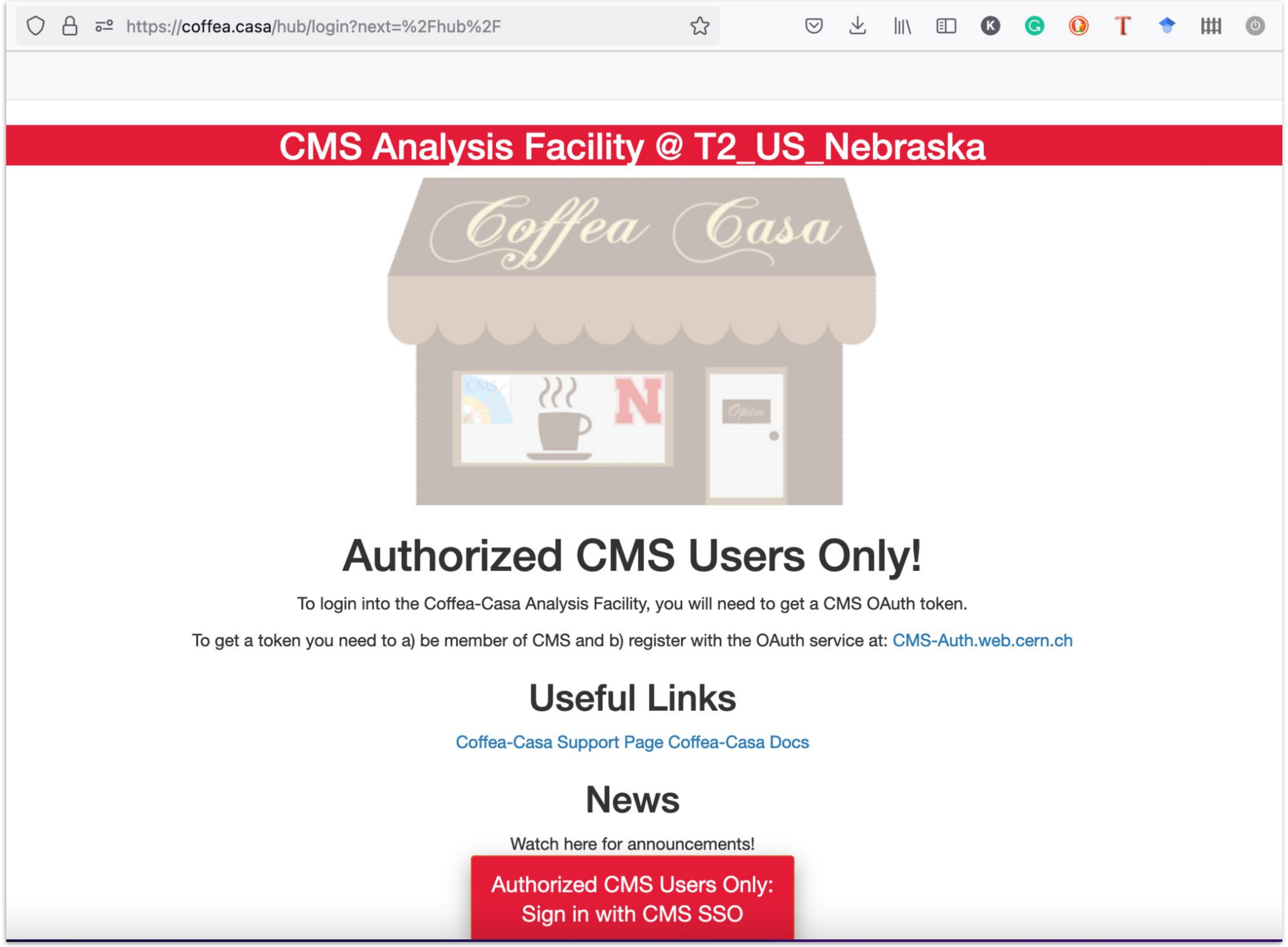
Dask can be used inside Jupyter or you can simply launch it through Jupyter and connect directly from your laptop

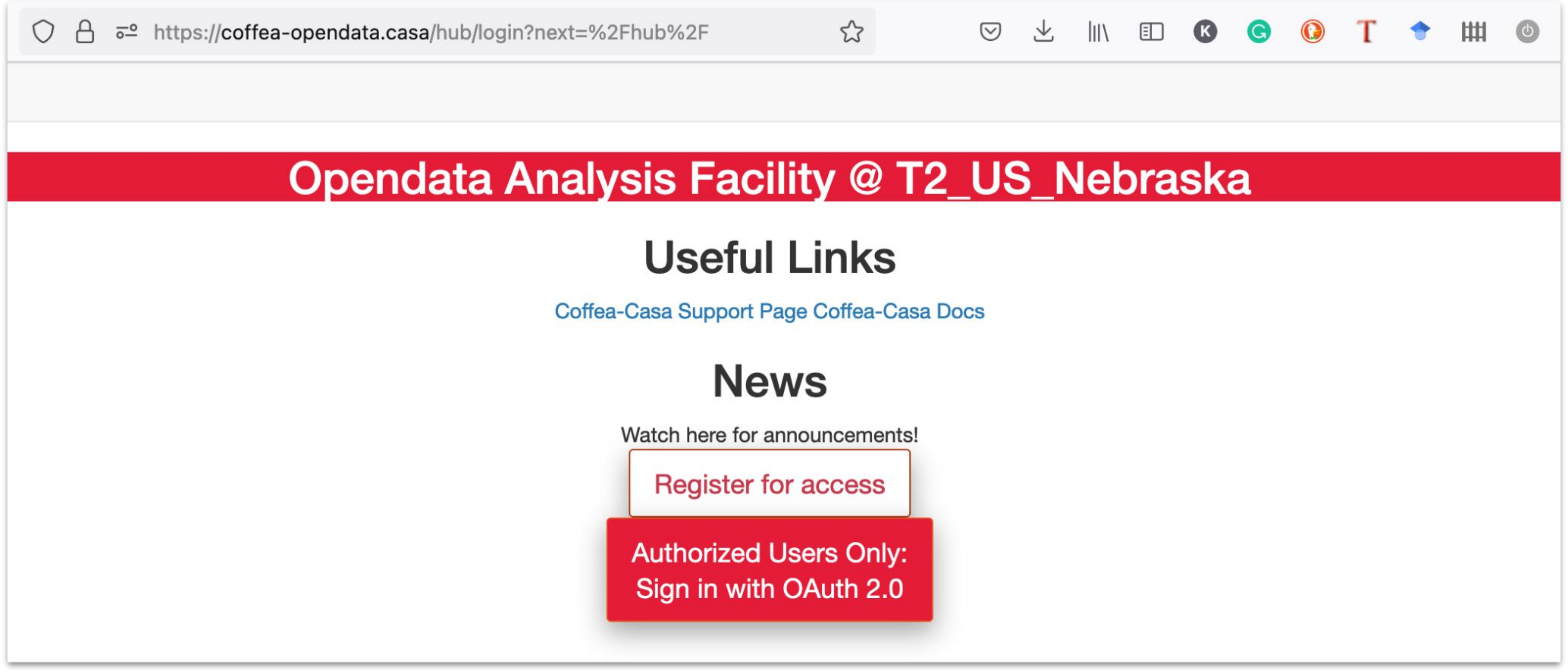
Building blocks: modern authentication (IAM/OIDC)

Authentication inside the system is independent of grid credentials

CMS Coffea-Casa Analysis Facility: https://coffea.casa

Opendata Coffea-Casa Analysis Facility: https://coffea-opendata.casa





Powered by CILogon COmanage™

Powered by CMS IAM instance

Building blocks: tokens

Token authentication (WLCG Bearer JWT)

JWT + XCache ServiceX

Other credentials

- Generated X.509 credentials (including a CA, host certificate, and user certificate) for use in Dask for TLS communication
- Enables also user communication to Dask scheduler endpoint

Building blocks: modern deployment and integration techniques - orchestration

For users:

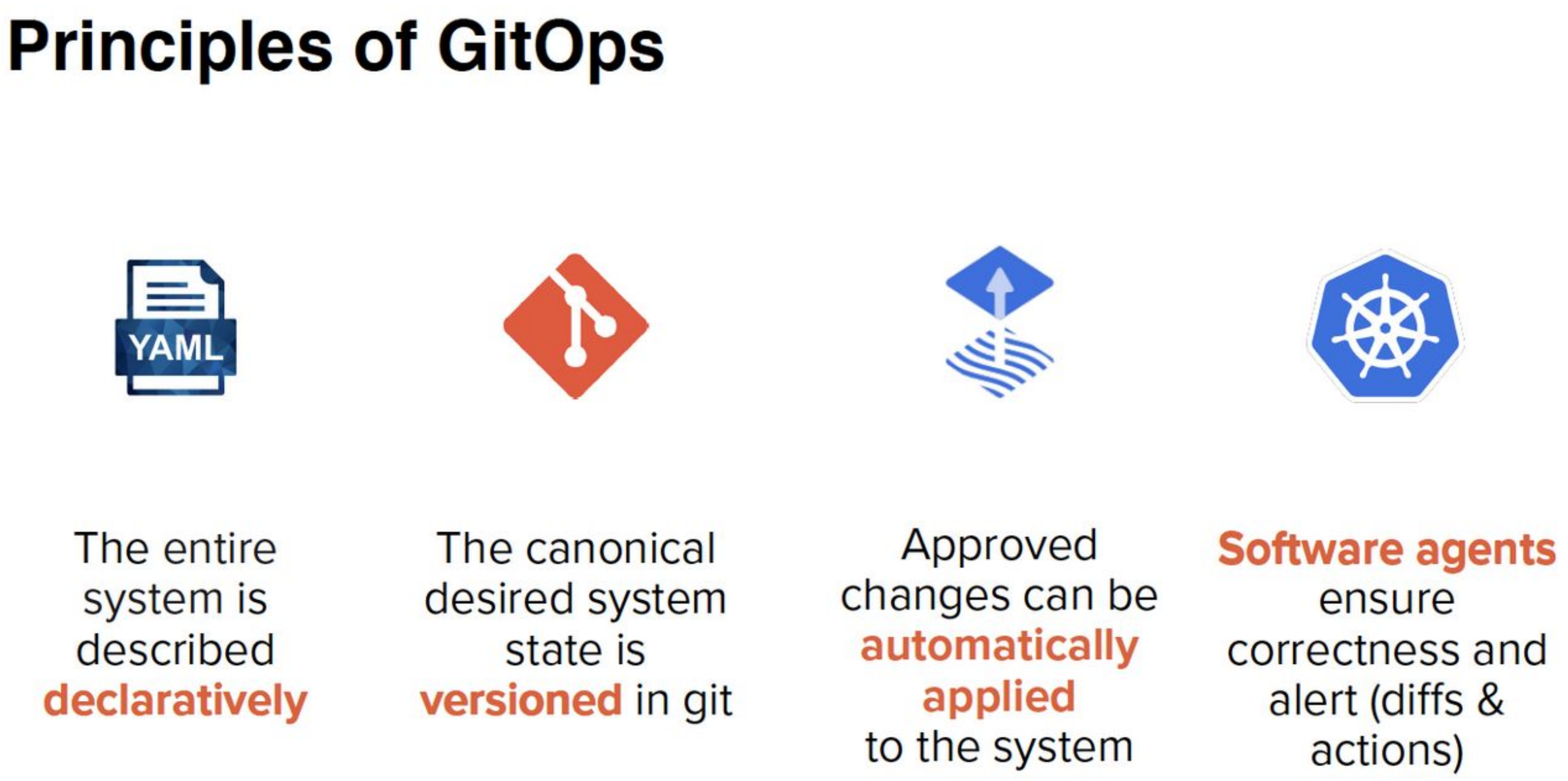
- Highly customized **“analysis” Docker container(s)**
- Investigating **Binderhub support**
  - It will allow users to share reproducible interactive computing environments from their code repositories at coffea-casa

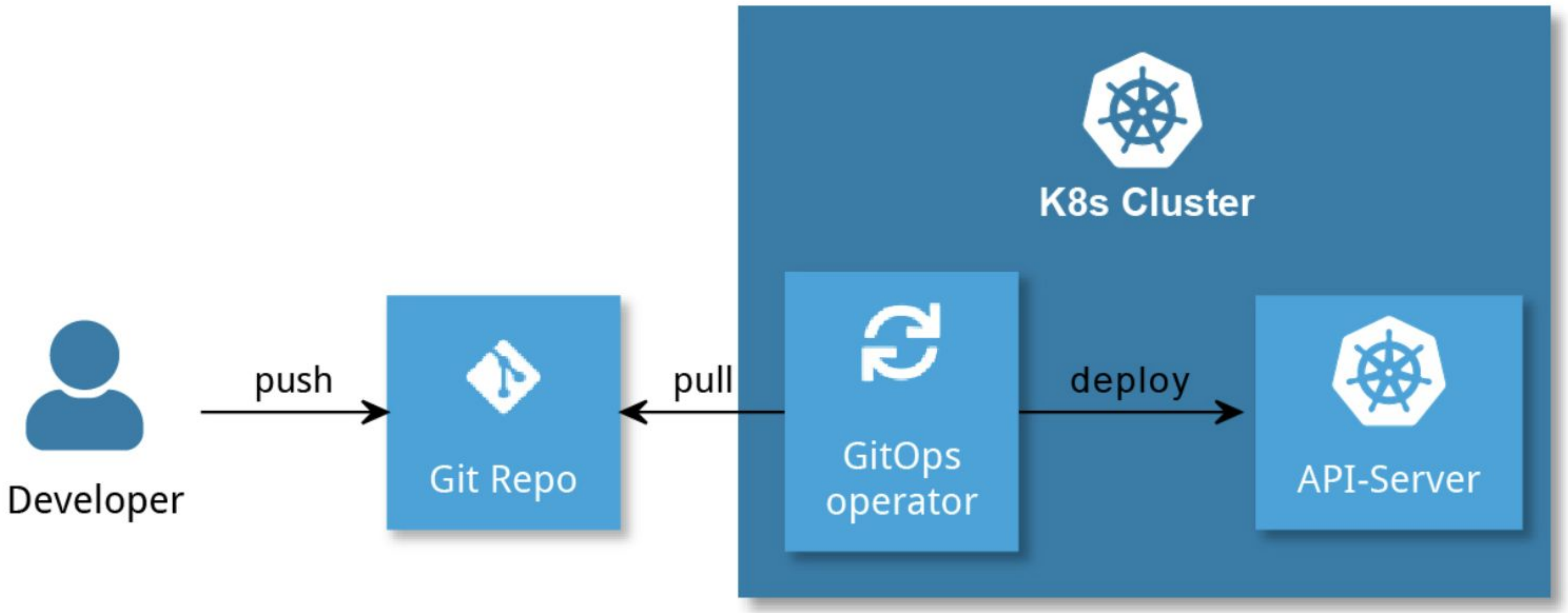
For developers:

- All features are **incorporated into a Helm chart** (Kubernetes packaging format)

Building blocks: modern deployment and integration techniques - GitOps

- **GitOps** defined as a model for operating Kubernetes clusters or cloud-native applications (e.g. coffea-casa AF)
- ***Concept: “infrastructure-as-a-code”***
  - Allow for rapid collaboration, better quality control, and automation (CD/CI)
  - AF is easily handled via a collaborative group of administrators in a deterministic manner
  - Allows easily packages the core infrastructure as a Helm chart

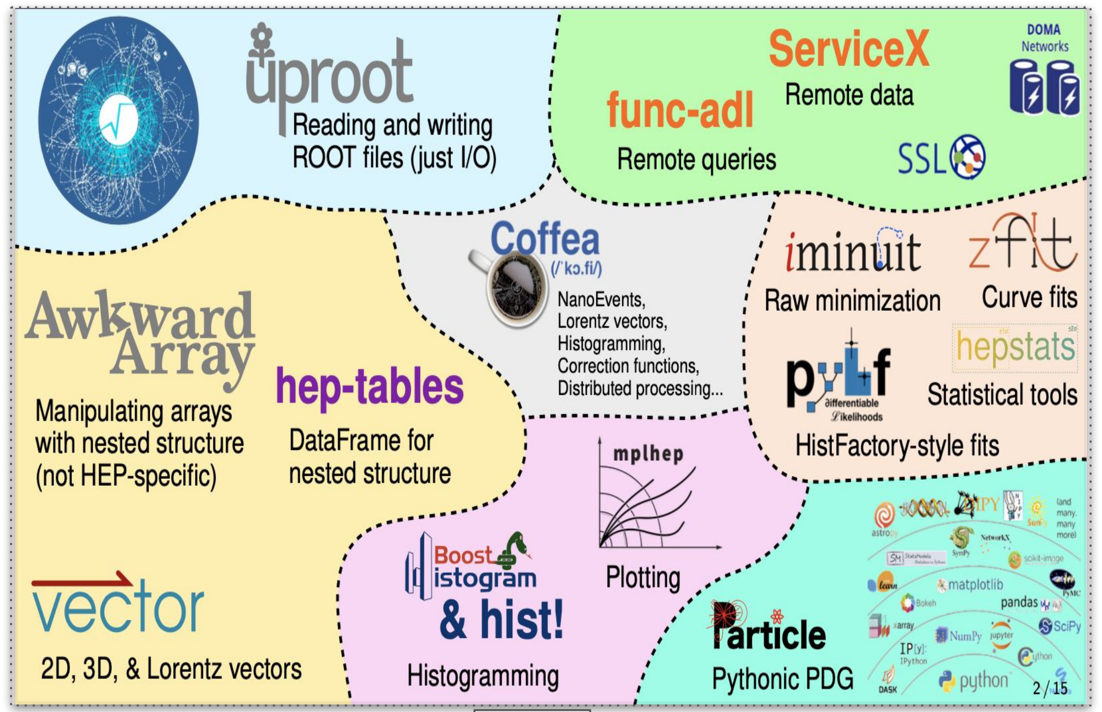


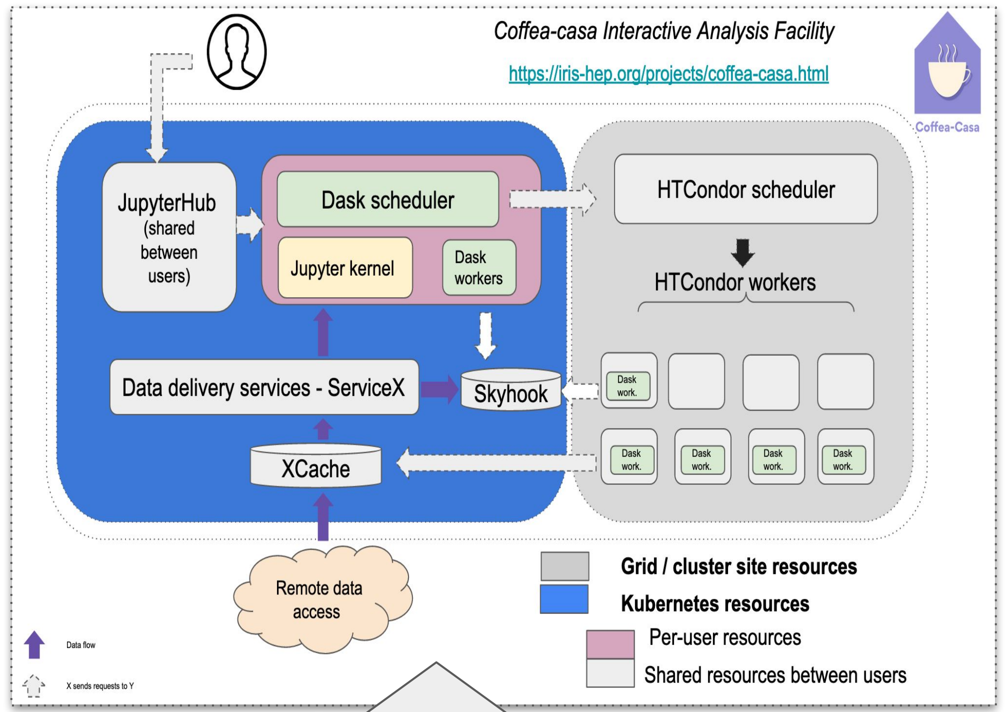


Analysis Grand Challenge

Motivation:

- Allow coping with HL-LHC data sizes by rethinking data pipeline
  - Evaluating the new Python analysis ecosystem and integrating a differentiable analysis pipeline
- Provide flexible, easy-to-use, low latency analysis facilities





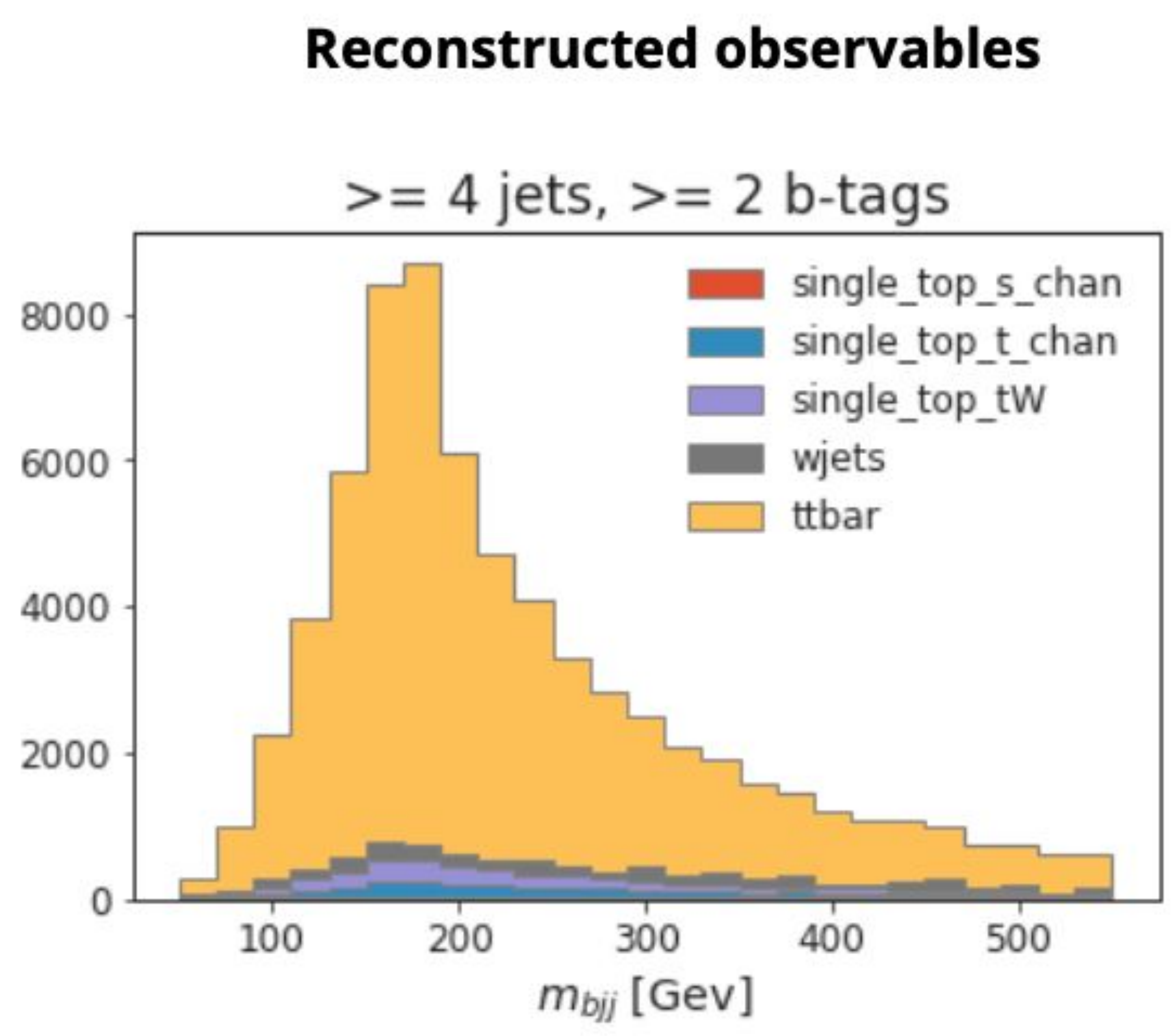
Analysis Grand Challenge (AGC)

Check the talk from Alexander Held "Physics analysis for the HL-LHC: concepts and pipelines in practice with the Analysis Grand Challenge"

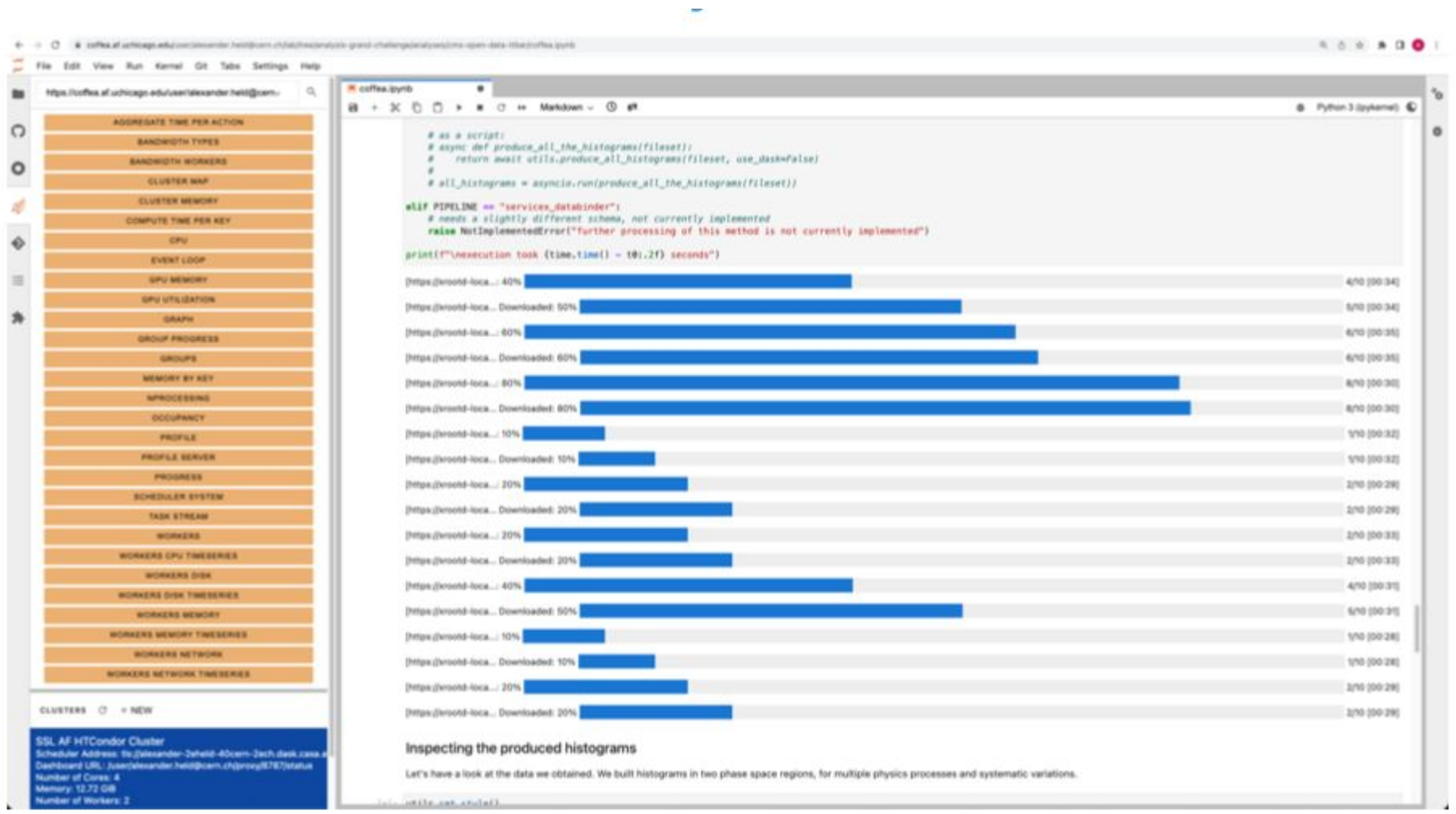
Analysis Grand Challenge **will be conducted during next years leaving enough time for tuning software tools and services developed as a part of the IRIS-HEP ecosystem before the start-up of the HL-LHC and** ***organized together with the US LHC Operations programs, the LHC experiments and other partners.***

Analysis Grand Challenge

- *Now:* defined a **physics analysis task** and developed **multiple implementations**
- *Next steps:* **plan in place** for how to **bridge remaining gap** towards HL-LHC
  - Two new flagship analyses, closer connections to LHC experiments
  - Extended functionality tested (data preservation, differentiable pipeline, …)
  - Incremental data rate goals for throughput



Output histogram from AGC analysis



Interactive analysis in a notebook

Conclusions

- *Coffea-casa* is a prototype analysis facility delivering extra functionality needed for **improved UX**
- Rethinking established design patterns and integrating new advanced services in traditional facilities enables possibility of **quick interactive analysis turnaround, allowing end-users to worry only about physics**
- We believe focusing on enabling ML-based analysis for facilities together with ability to handle HL-LHC data volumes is the right path to future analysis facilities

This work was supported by the U.S. National Science Foundation (NSF) Cooperative Agreement OAC-1836650 (IRIS-HEP).

Thank you!
Q&A

Backup

Coffea-casa AF components