Analysis Grand Challenge:
Coffea-Casa analysis facility as a test environment for packages and services

Coffea-casa contributors:

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

Brian Bockelman
Morgridge Institute, Madison, USA

Intro

- Adjunct Software developer @ University of Nebraska-Lincoln
  - Coffea-casa analysis facility developer
  - IRIS-HEP Analysis Grand Challenge co-coordinator (together with Alex Held)

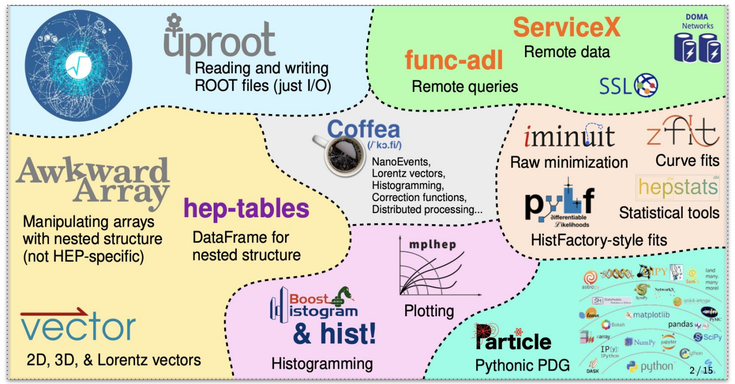
  My background:

  - Infrastructure engineer at BITP (ALICE)
  - Doctoral Student at GeantV simulation project
  - ROOT team member:
    - IO (compression algorithms) & build system
    - C++ modules in ROOT/CMSSW
    - Performance studies in ROOT

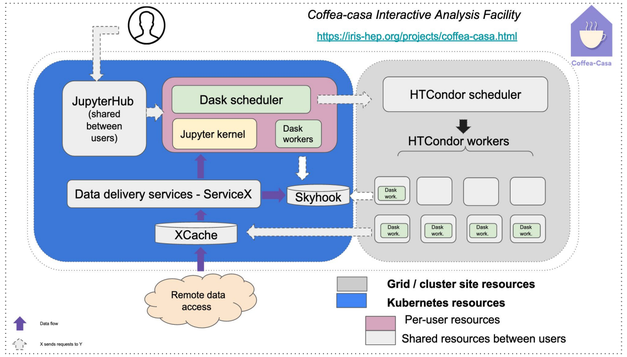
Analysis Grand Challenge (AGC)

- Analysis Grand Challenge (AGC): ***execute series of increasingly realistic exercises toward HL-LHC***
- The AGC is **an integration exercise** for IRIS-HEP
- Also **community project**: focus discussion, bring together different groups and experiments

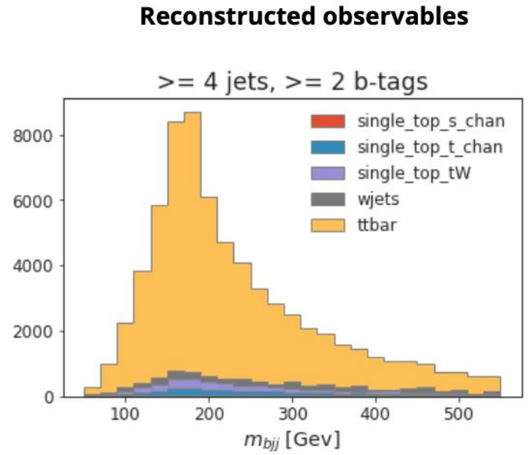
Analysis Tools and Services



Analysis Facilities



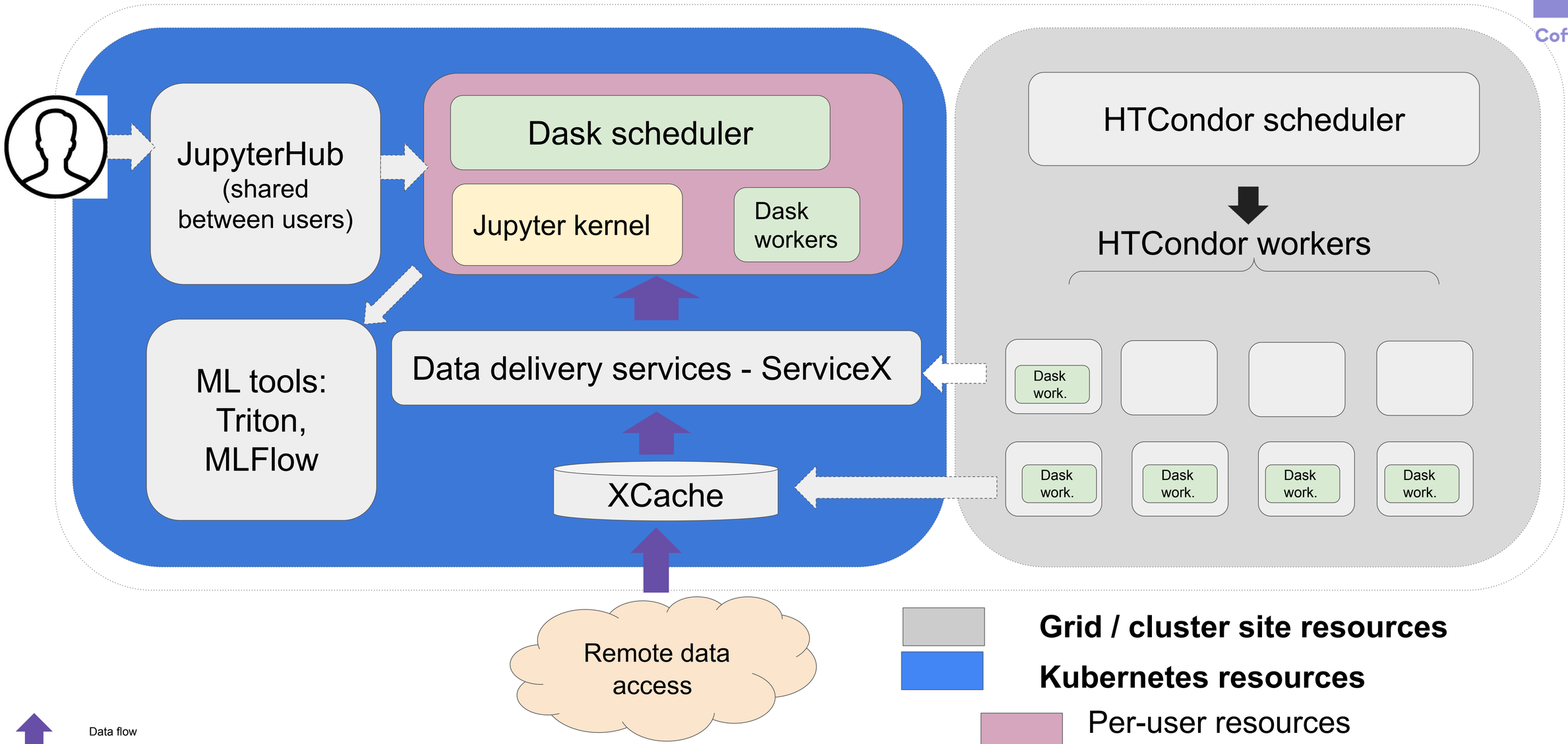
Execution of AGC analysis benchmark



example output of [analysis notebook]

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



Grid / cluster site resources

Kubernetes resources

Per-user resources

Shared resources between users

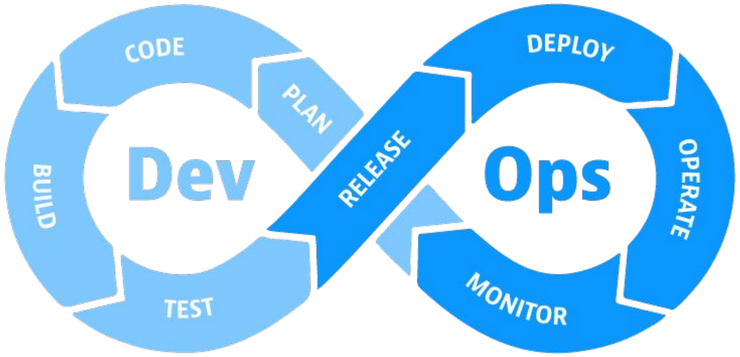
Data flow

X sends requests to Y

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

My interests

analysis workflows - UX - ML (MLOps) - Infrastructure as a Code



- In AGC we ensure that whole analysis pipeline (e.g. packages) are tested in the different aspects: scalability, UX, performance bottlenecks and make sure packages are available to early adopters and redeployable anywhere else
  - I would like to encourage more collaborative work between the facility architects and package / service developers: it definitely makes things work faster ;-)
  - The facility design should be treated as an additional “layer” while thinking about vertical and horizontal layering of analysis pipeline
  - Are there specific services that would be useful to provide?
- HEP analysis lifecycle is getting more and more complex (ML model training, inference services, autodiff etc..)
  - How to make sure that it scale?
  - Can we design all-in-one solution for HEP?