IRIS-HEP retreat

AGC focus area talk (10’)

Alexander Held (University of Wisconsin–Madison)
Oksana Shadura (University Nebraska–Lincoln)

IRIS-HEP Institute Retreat 2023, Sep 11–13 2023, https://indico.cern.ch/event/1288444/

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

Analysis Grand Challenge (AGC):

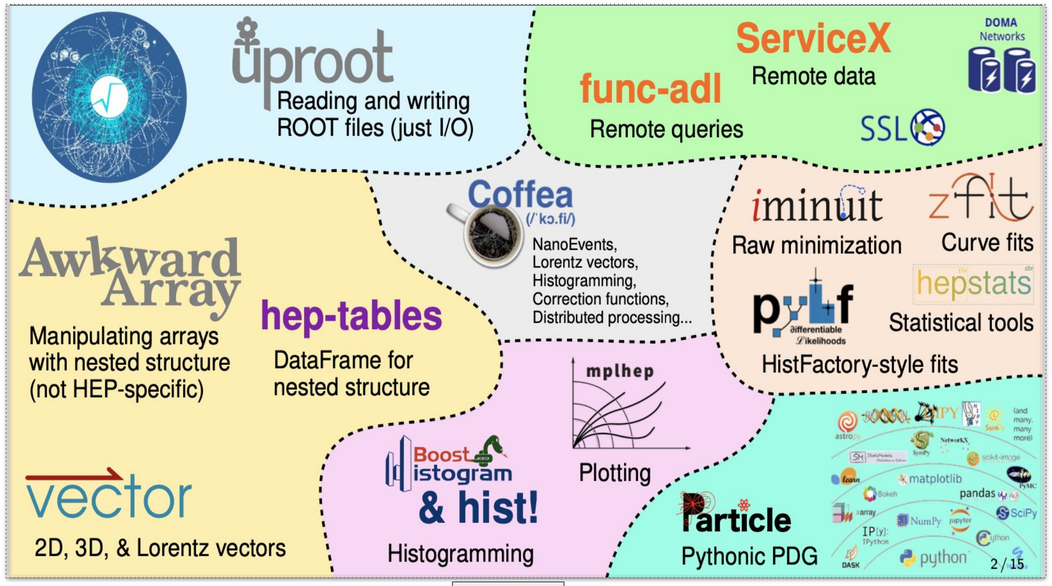
execute series of increasingly realistic exercises toward HL-LHC

The AGC is about executing an analysis to test workflows designed for the HL-LHC.
This includes:

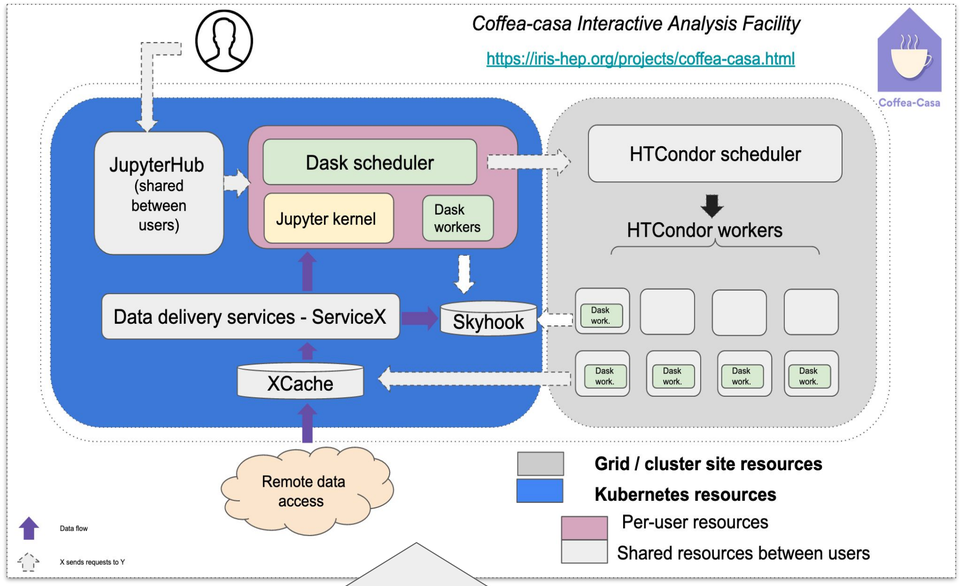
- **columnar data extraction** from large datasets,
- **data processing** (event filtering, construction of observables, evaluation of systematic uncertainties) into histograms,
- **statistical model construction and statistical inference**,
- relevant **visualizations** for these steps

The AGC is **an integration exercise** for IRIS-HEP towards reaching the goal to implement and deploy such a pipeline

Analysis Tools and Services



Analysis Facilities

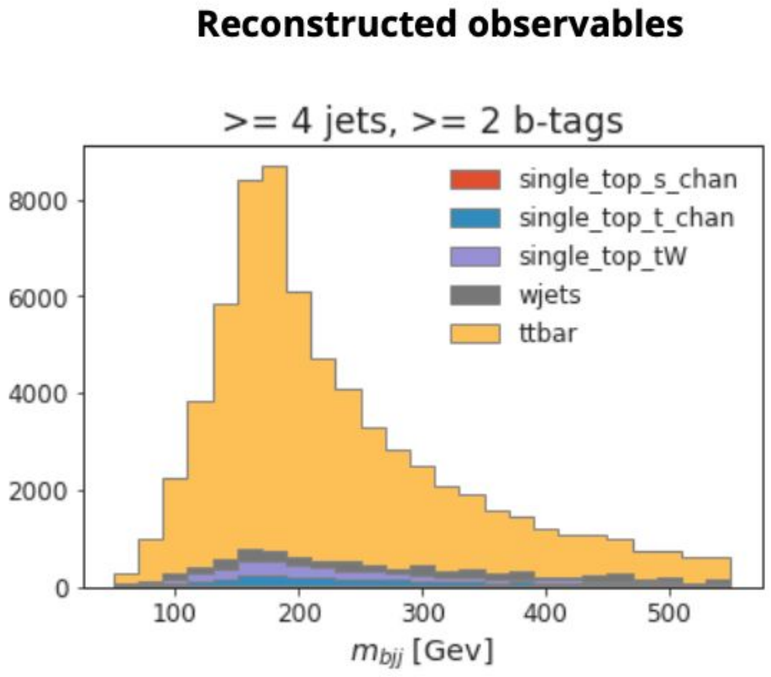


Analysis Grand Challenge (AGC)

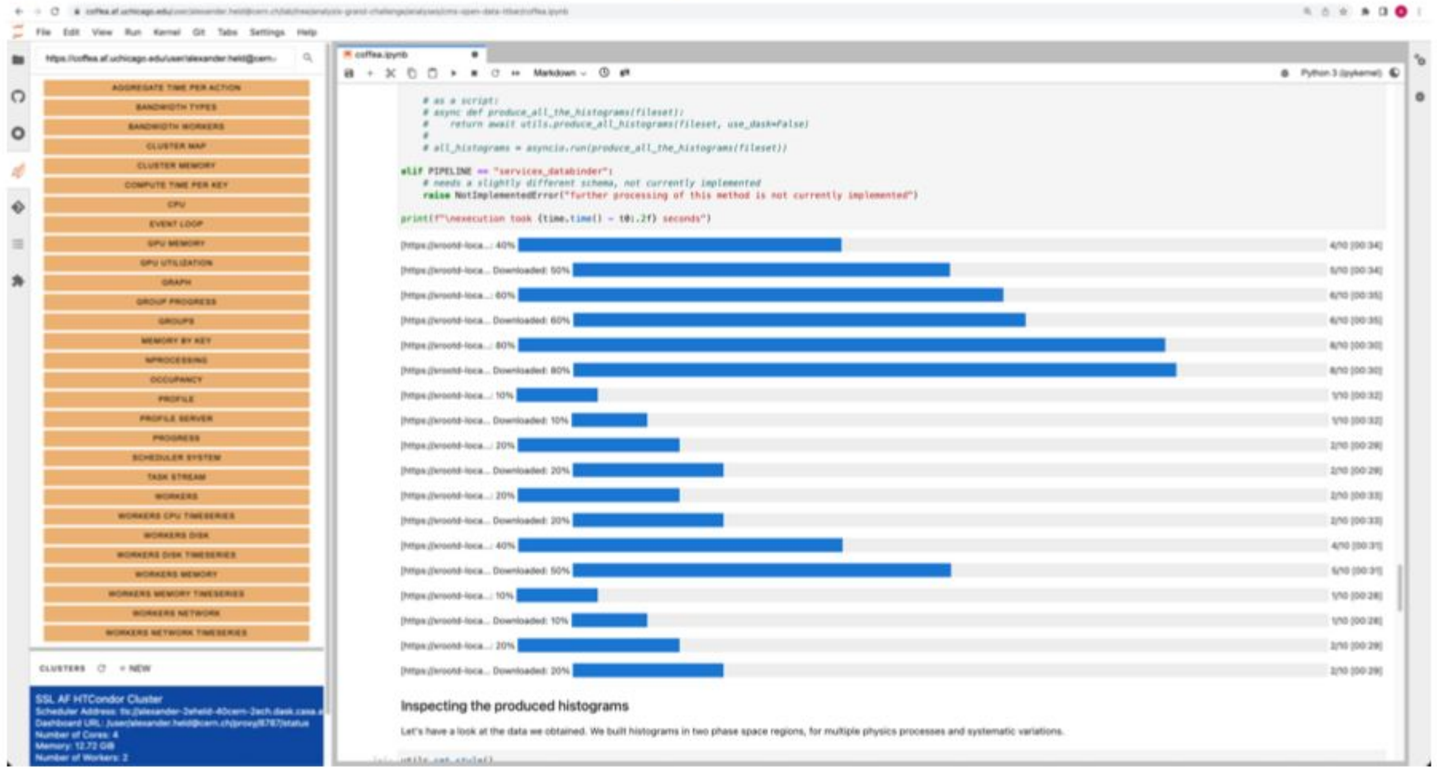
Current status

Defined a **physics analysis task** and developed **multiple implementations**

- Have done benchmarking, iterated with tooling & services involved
- Presented results at various conferences: ICHEP 2022, ACAT 2022, CHEP 2023 (+ AGC demonstration event!)



Output histogram from AGC analysis

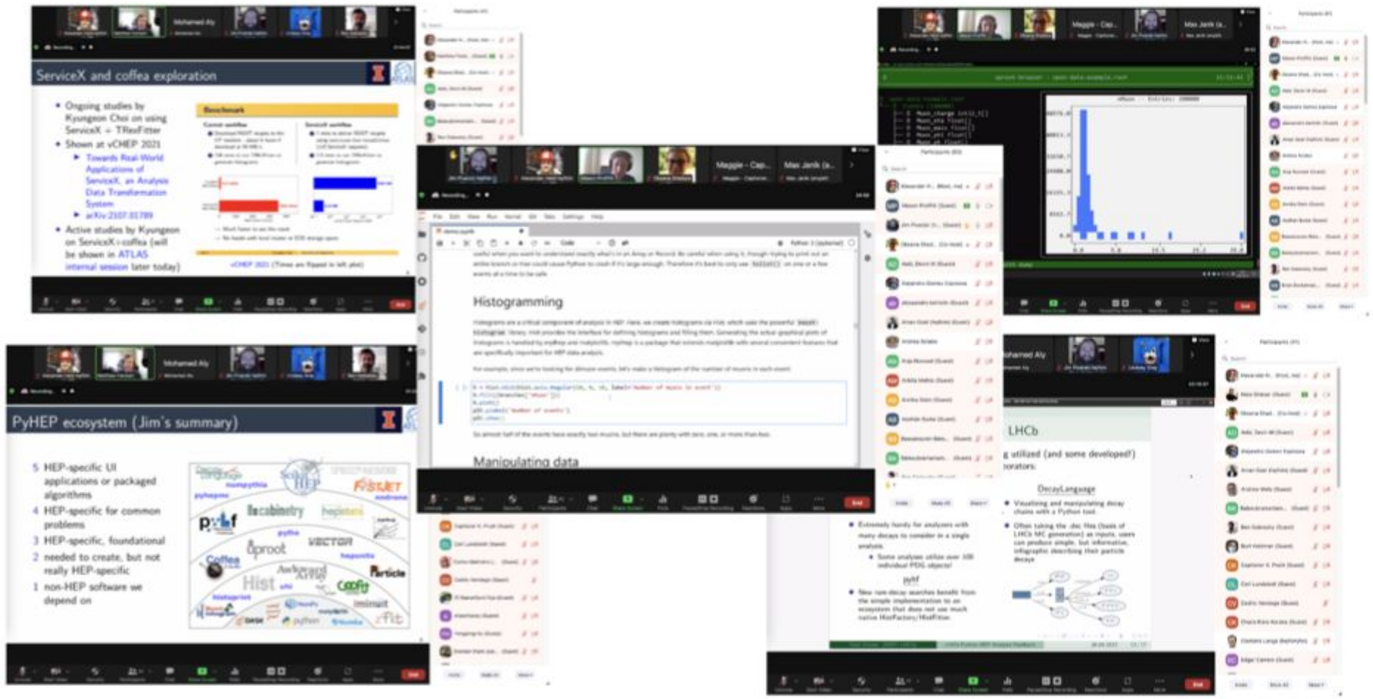


Interactive analysis in a notebook

Engaging & building communities

We use a multitude of avenues to **disseminate methodology** and engage in **feedback cycles**

- Annual AGC workshops (last iteration: 124 participants)
- Collaboration with US ATLAS/CMS Ops programs via dedicated meetings
- “Demo days” showcasing technical progress
- Collaboration with ROOT RDF group, JuliaHEP, SWIFT-HEP, US-CMS PURSUE & IRIS-HEP fellow programs



AGC tools 2022 workshop

Connecting projects and resources through AGC

- IRIS-HEP: integration exercise for **AS, DOMA, SSL areas**
- US collaborators: US Operation programs providing facility support and resources
  - **Coffea-casa analysis facility** as an example of joint effort with US Operation programs (US-CMS / US-ATLAS)
- Experiments: CMS and ATLAS
- International collaborators
  - Cooperation with **ROOT team on AGC** RDF implementation

Challenges & limits of current practices: Scaling & turnaround

- **Scaling to HL-LHC data volumes** with available computing resources
  - Need for new methods for efficient data scaling, caching at AFs to handle more data-intensive analysis pipeline

- **Analysis turnaround time**
  - Reaching interactive analysis turnaround times requires efficient analysis facility (AF) usage

Challenges & limits of current practices: UX & sustainability

- **User experience (UX)** for complex analyses: **increase scientific reach** of result
  - User improvement experiences allow physicists to focus on the physics
  - Need expanded Machine Learning (ML) tooling with good user experience + performance
  - Leverage ML technology to automatically optimize analysis sensitivity

- **Sustainability**
  - Limited person power to develop & maintain full stack -> rely on industry solutions & external developments (e.g. tokens)
  - Limited of analysis reproducibility & reusability
  - Need for central gathering point for community to discuss & develop analysis approaches

Roadmap for this week

- **Monday 11:00: brief look** at AGC (you are here)
- **Tuesday 9:00: previous plans** for AGC (as outlined in the strategic plan)
- **Tuesday 13:30: AGC during IRIS-HEP v2** (evolving the previous planning)
  - 90’ discussion session
  - Important inputs:
    - AS plans (Monday 13:30) & AS autodiff (Tuesday 11:00)
    - ServiceX plans (Monday 15:30)
    - AF plans (Tuesday 9:30)
- **Wednesday 9:30: summary**
- **Thursday 9:30: AGC demonstration event**
  - https://indico.cern.ch/e/agc-demonstration

Backup