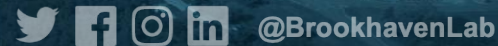# Future Challenges for the US Analysis Facilities – Looking Towards Run 4

Ofer Rind

US ATLAS WBS 2.3 F2F Meeting
November 30th, 2022

@BrookhavenLab

Looking Ahead…

As you've seen, AF efforts have ramped up significantly over the past year and much progress has been made, but the ultimate focus is really on Run 4.

- While the 2029 target date is clear(ish), the timeline is not well-defined as the technology is evolving, the tools are being developed, and the user needs are still being determined.
- This is an **opportunity** for US ATLAS, to identify and target particular needs, set our priorities, and forge connections with advanced users who can help keep development on the right track.

There are numerous efforts ongoing in parallel under the umbrella of various organizations: HSF, WLCG, IRIS-HEP, … all useful, but each with a different focus.

Brookhaven
National Laboratory

The IRIS-HEP Analysis Grand Challenge

A current effort of focus, with two components:

- Defining a physics analysis task of realistic HL-LHC scope & scale (e.g. ttbar cross-section measurement in single lepton channel)
  - Large data volumes, reduced data formats (e.g. PHYSLITE)
  - Add an ML component to the workflow
- Developing an analysis pipeline that implements this task
  - Finding & addressing performance bottlenecks & usability concerns

"Integration Exercise" for IRIS-HEP tools on AFs

- Interactive analysis with turnaround times of O(minutes) - parallel "bursty" execution, low latencies, data caching
- Reimplementation and benchmarking (?)

Brookhaven National Laboratory

AGC Tools

- Execution Environment: Coffea-Casa
  - Built on Dask with potential for an array of different site implementations
- Data Management: ServiceX + XCache
  - In active development, with security and other implementation issues
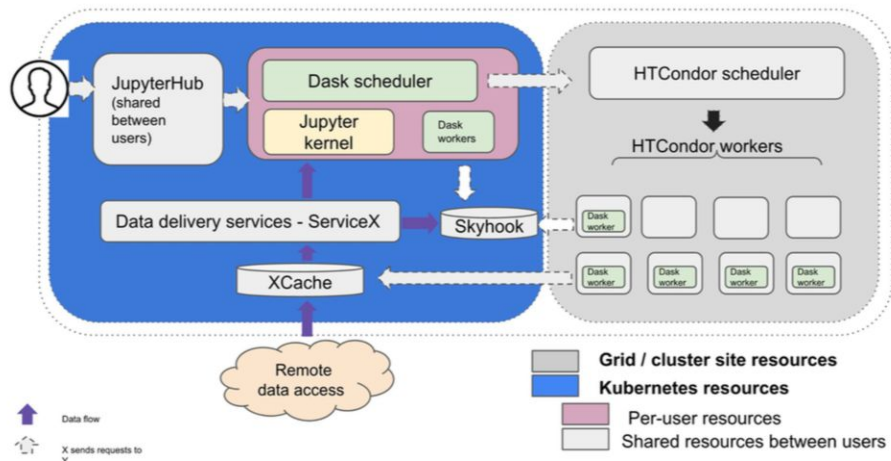
Upcoming AGC Events

- [IRIS-HEP Demo Day](#) 12/16/22, 10:00 CT
  - First in a series (Bimonthly) with short progress reports
- AGC Execution Workshop some time in late May, after CHEP 2023
  - Site results and information exchange
  - Promoting sites to the analysis community
- "Execution of full AGC pipeline(s) next Summer on various facilities (including diverse hardware configurations)"

Brookhaven National Laboratory

Discussion: US ATLAS Questions & Priorities

- The AGC is a very useful effort, but focus tends to be on IRIS-HEP tools
  - Are these technologies aligned with ATLAS analysis development efforts?
  - Do these tools fit in well with plans to develop benchmarks for ATLAS analyses? Should ServiceX be a key component at this stage? ML workflows?
  - To what extent should this drive our current efforts? There are other priorities not addressed as part of this (e.g. container/ environment management, user file sharing)
- Current IRIS-HEP focus is on open data
  - Should demonstrated use of ATLAS data, in particular physlite, be prioritized (FY23 milestone)?
- Need to keep a healthy feedback loop going with ATLAS users
  - Need to foster close collaboration with AMG (Analysis Model Group)
  - We are thinking about analysis in 7 years, but want to grow our services for these users now

Brookhaven National Laboratory

Extra Slides

An AGC implementation: software stack

Involves large number of packages from IRIS-HEP and partners

Analysis specific frameworks and packages (available in Docker container)

Data delivery service (k8s)

Optional services (k8s)

Credit: Alex Held

8