IRIS-HEP
Analysis Grand Challenge
Tools Workshop

Alex Held (UW-Madison)
Oksana Shadura (UNL)

April 25–26, 2022
https://indico.cern.ch/e/agc-tools-2

This work was supported by the U.S. National Science Foundation (NSF) Cooperative Agreement OAC-1836650 (IRIS-HEP).
Analysis Grand Challenge will be conducted during 2021–2023, 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.
The AGC is connecting different working groups and IRIS-HEP partners

Data Organization, Management and Access (DOMA):

- Data delivery
  - Data Factory/source (e.g., T0 or sim)
  - Data Store/Lake
  - Intelligent Data Delivery Service (IDDSS)
  - Data Cache
  - Compute Nodes/ Data Sinks

Analysis Systems (AS): tools

Scalable Systems Laboratory (SSL): deployment techniques and resources
Software and services requirements

Analysis specific frameworks and packages (available in Docker container)

Data delivery service (k8s)

Optional services (k8s)
Towards a benchmark analysis

- Main AGC analysis example is based on **Run-2 CMS Open Data**
  - **Open Data is crucial**: allows everyone to participate!
  - Currently using **PhysObjectExtractorTool** to convert existing **miniAOD** datasets into ntuple format
  - In close contact with CMS to make conversion to **nanoAOD** format possible in the medium term
    - Will switch to **nanoAODs** when available to more closely mirror **PHYSLITE / nanoAOD** workflows
  - Intention of demonstrator is to show functionality, **not** to discover new physics in the released Open Data!
  - **Many thanks** to the **CMS DPOA team** for all the help with the CMS Open Data!

- **Categorized datasets** in terms of role in AGC demonstrator ([AGC repository](#))

- **Everything is openly developed** ([AGC repository](#))
  - We encourage you to **re-implement** (parts of) the pipeline with other tools!
  - We hope this exercise can be useful broadly to the field for **benchmarking** purposes as well
AGC: The timeline

- **Nov 2021:** first demonstration of toolchain at AGC workshop ([agenda](#))
- **April 2022:** second iteration of AGC workshop happening now ([agenda](#))
- **May 2022:** related event: Analysis Ecosystem workshop ([agenda](#)), consider joining us in Paris!
- **Summer 2022:** benchmarking of system components in the AGC context
- **Spring 2023:** execution of AGC at full scale

Stay in touch via [analysis-grand-challenge@iris-hep.org](mailto:analysis-grand-challenge@iris-hep.org) (sign up: [google group link](#))
Day 1:
a tour through the ecosystem and pipeline demonstrations

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:00</td>
<td>Introduction</td>
<td>Alexander Held et al.</td>
</tr>
<tr>
<td>16:00</td>
<td>Foundation libraries (uproot, awkward, hist, mphep)</td>
<td>Mason Profitt</td>
</tr>
<tr>
<td>17:00</td>
<td>Queries with func_adl and data delivery with ServiceX</td>
<td>Gordon Watts</td>
</tr>
<tr>
<td>18:00</td>
<td>Columnar analysis with coffea</td>
<td>Lindsey Gray et al.</td>
</tr>
<tr>
<td>19:00</td>
<td>From data delivery to statistical inference with CMS Open Data</td>
<td>Alexander Held</td>
</tr>
</tbody>
</table>

Day 2:
Skyhook, scale-out on coffea-casa, user experience & experiment-specific sessions

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16:00</td>
<td>Data management with Skyhook</td>
<td>Carlos Maltzahn et al.</td>
</tr>
<tr>
<td>16:00</td>
<td>Scale-out with coffea: coffea-casa analysis facility</td>
<td>Carl Lundstedt et al.</td>
</tr>
<tr>
<td>16:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>17:00</td>
<td>Analysis user experience with Python HEP data science tools in CMS</td>
<td>Lindsey Gray</td>
</tr>
<tr>
<td>17:00</td>
<td>Analysis user experience with Python HEP data science tools in ATLAS</td>
<td>Matthew Foulds</td>
</tr>
<tr>
<td>17:00</td>
<td>Analysis user experience with Python HEP data science tools in LHCb</td>
<td>Nathan Allen Green</td>
</tr>
<tr>
<td>18:00</td>
<td>Discussion</td>
<td></td>
</tr>
<tr>
<td>18:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>19:00</td>
<td>CMS session</td>
<td>Lindsey Gray et al.</td>
</tr>
<tr>
<td>19:00</td>
<td>ATLAS session</td>
<td>Lukas Alexander Heinrich</td>
</tr>
</tbody>
</table>

For the last Tuesday block, we are hosting experiment-specific sessions: **If you are affiliated with CMS or ATLAS, please join!**
Communication channels

- Zoom link is attached to the agenda:
  - https://indico.cern.ch/event/1126109/videoconference/

- This meeting is being recorded, recordings will be made available afterwards

- We are encouraging you to use Slido for questions:
  - Join at slido.com with #152901 or with this direct link
  - Ask Questions, vote questions you’d like to see answered to the top
  - Feel free to also use Zoom chat for follow-up discussion

As a participant in this workshop, you are expected to follow our code of conduct: https://indico.cern.ch/event/1126109/page/24856-edi-statement
Generic AGC Tools sessions:

Opendata Coffea-casa AF @ UNL
Opendata Coffea-casa AF @ UNL

- For all generic sessions, you can follow tutorials on the prototype of the Open Data Coffea Analysis Facility @ UNL: [https://coffea-opendata.casa](https://coffea-opendata.casa)
- Instance has all AGC tools & packages including Open Data instance of ServiceX
- To access it, you will need to register (it is only two minutes, we promise!)
- More documentation how to register is available [here](https://github.com/CoffeaTeam/coffea-casa/discussions)
- Any questions & suggestions: [https://github.com/CoffeaTeam/coffea-casa/discussions](https://github.com/CoffeaTeam/coffea-casa/discussions)
If you already have this repository from previous workshop, don’t forget to pull updates from terminal (or JH plugin):

```
git pull origin master
```
ATLAS AGC Tools sessions:

ATLAS coffea-casa AF @ UChicago
ATLAS Coffea-casa @ UChicago

- As a member of the ATLAS experiment, you can use a prototype ATLAS Coffea Analysis Facility @ UChicago: https://coffea.af.uchicago.edu

- Instance has all AGC tools & packages and we tested the ATLAS instance of ServiceX: https://uproot-atlas.servicex.af.uchicago.edu/
  (documentation how to get ServiceX access is here)

- To access it, you just need to be a member of the ATLAS experiment, access there will be granted based on your credentials.

- More documentation how to register is available here

- Any questions & suggestions: https://github.com/CoffeaTeam/coffea-casa/discussions
Generic & CMS AGC Tools sessions:

EAF @ Fermilab
EAF @ FNAL

- For the generic and CMS experiment session, we would like to promote Elastic Analysis Facility @ FNAL: [https://analytics-hub.fnal.gov](https://analytics-hub.fnal.gov)
- Instance has all AGC tools & packages
- *To access it, you just need to have FNAL account (FERMI SERVICES domain credentials)*, access there will be granted based on your credentials.
- In the analysis notebook, please select in global configuration section: AF = “EAF”
CMS AGC Tools sessions:

CMS coffea-casa AF @ UNL
CMS Coffea-casa @ UNL

- For the CMS experiment session, we would like to promote *(it is not required for session)* a prototype of the CMS Coffea Analysis Facility @ UNL: https://coffea.casa
- Instance has all AGC tools & packages
- *To access it, you just need to be a member of CMS experiment*, access there will be granted based on your credentials.
- More documentation how to register is available [here](#)
- Any questions & suggestions: [https://github.com/CoffeaTeam/coffea-casa/discussions](https://github.com/CoffeaTeam/coffea-casa/discussions)
We hope you will enjoy the workshop!
We expect to have more such events in the future.

Big thanks to all speakers!

- Coffea team
- FuncADL & ServiceX teams
- pyhf & cabinetry teams
- many more people in the surrounding ecosystem who have helped us!

Many thanks to the team providing resources & support to make this workshop happen:
- UNL “coffea-casa” team: Ken Bloom, Garhan Attebury, Carl Lundstedt, John Thiltges
- SSL: Lincoln Bryant, Fengping Hu, Rob Gartner, Ilia Vukotic, Suchandra Thapa