



Analysis Grand Challenge

Alex Held (NYU)
Oksana Shadura (UNL)

Nov 30, 2021

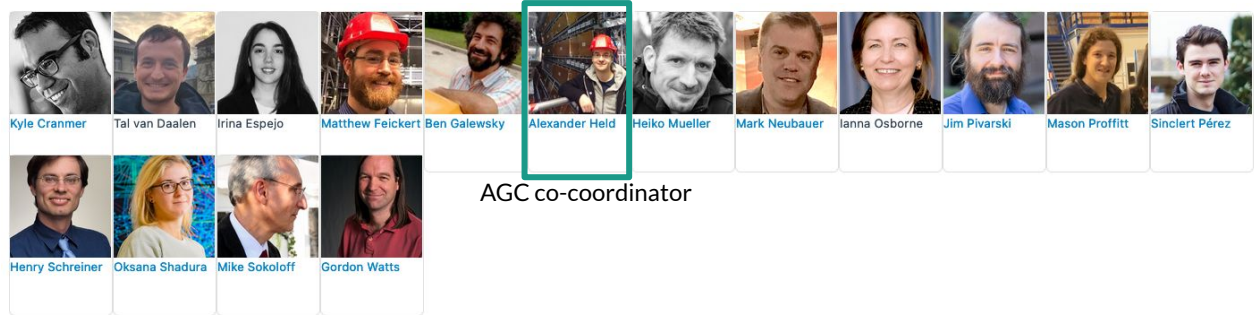
<https://indico.cern.ch/event/985527/>

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



Team

AS



DOMA



SSL



Analysis Grand Challenge

Motivation:

- Allow coping with HL-LHC data sizes by rethinking data pipeline
- Provide flexible, easy-to-use, low latency analysis facilities



Looking for new ideas for Analysis Facilities

- **New pythonic ecosystem**
- Discovering the benefits of **column-oriented (columnar) data analysis**
- **Interactivity** for user data analysis
- Deliver the needed data to the processing workflow in a fine-grained approach (**data delivery services**) and **efficient storage technologies** (e.g. object stores)
- **Kubernetes (k8s)** and new concept of "**infrastructure as code**"
- **Portability** and flexibility across different environments
- Integration with existing resources: current infrastructure is not going to be replaced in one day








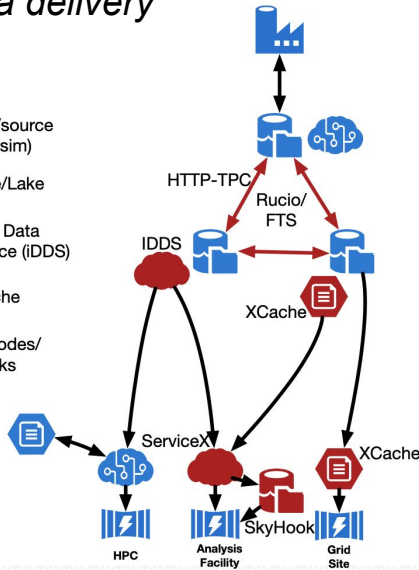
Coffea-casa vCHEP 2021 plenary talk

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.*

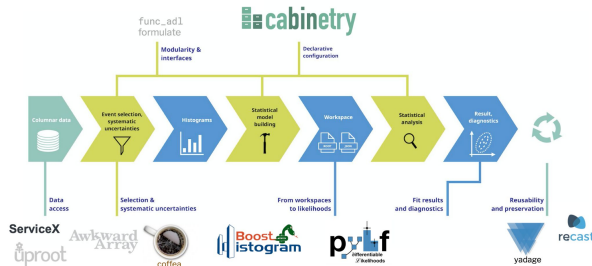
AGC is connecting IRIS-HEP focus areas

DOMA: Data delivery

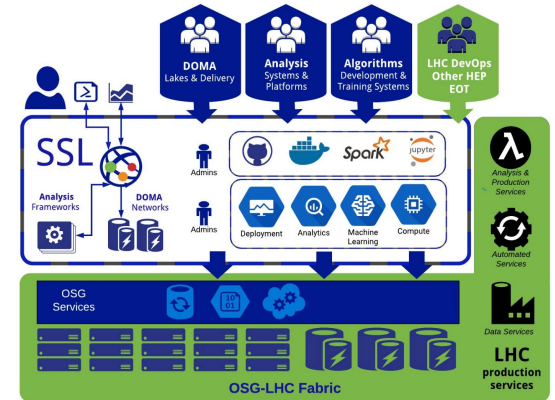
-  Data Factory/source (e.g. T0 or sim)
-  Data Store/Lake
-  Intelligent Data Delivery Service (iDDS)
-  Data Cache
-  Compute Nodes/Data Sinks



AS: tools



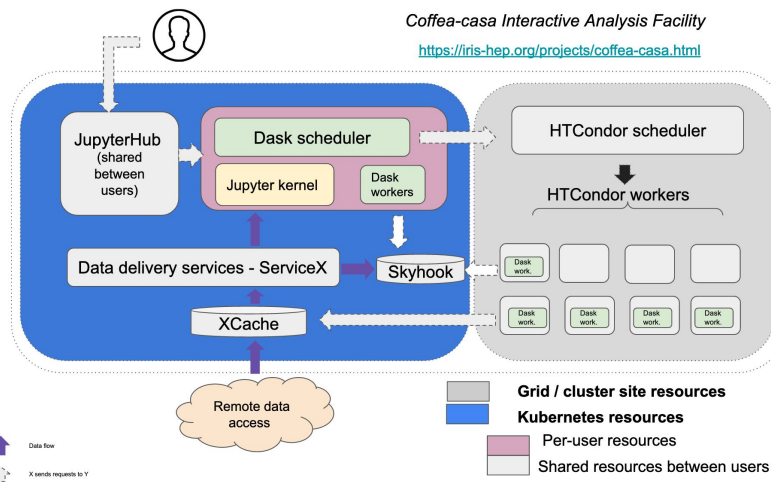
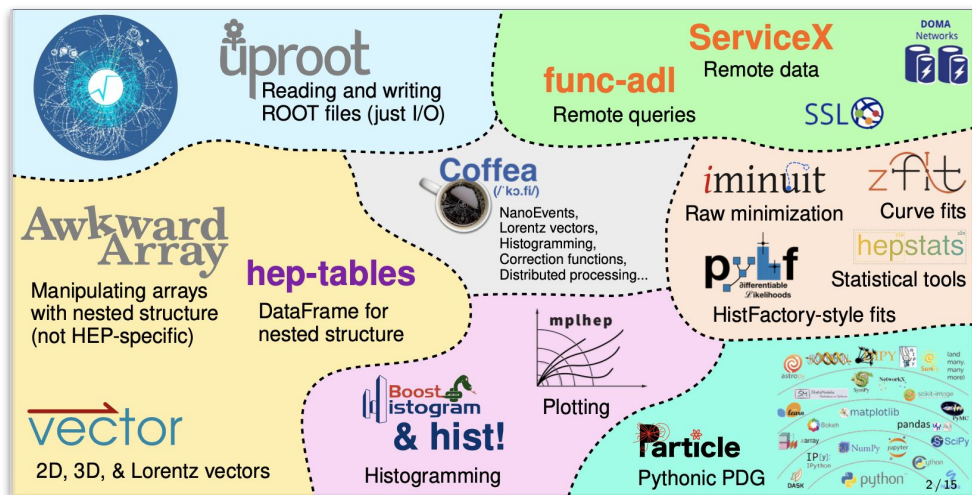
SSL: deployment techniques and resources



The idea behind the Analysis Grand Challenge

Analysis Tools

Analysis Facilities



Analysis Grand Challenge Analysis requirements

- **Analysis needs to include features encountered in realistic ATLAS/CMS analyses**
 - Handling of **large data volumes** (we have in mind: ~200 TB)
 - Handling of **different types of systematic uncertainties**
 - Ideally use of **modern formats** *NanoAOD / PHYS / PHYSLITE*
 - make it easier for ATLAS/CMS analyzers to adopt to their use case
- **Not intended to send physics message with analysis**
 - Want to show **realistic workflow**, not make physics claims
 - No need for real data, **simulation fully sufficient** (ideally many samples to simulate book-keeping)
- **Want to demonstrate enhanced functionality**
 - Possibility to **end-to-end optimize physics analysis**, potentially via automatic differentiation
 - Analysis needs to run on analysis facility
- **Analysis needs to be sufficiently specified for others to re-implement**
 - **Ideally: data is open and available to everyone (or scheduled to become public in the near future)**
 - Hoping to learn from comparing to implementations developed by others outside IRIS-HEP
 - Want to turn parts of analysis into mini-benchmarks for facility and tool benchmarking

Value of Open Data for the AGC



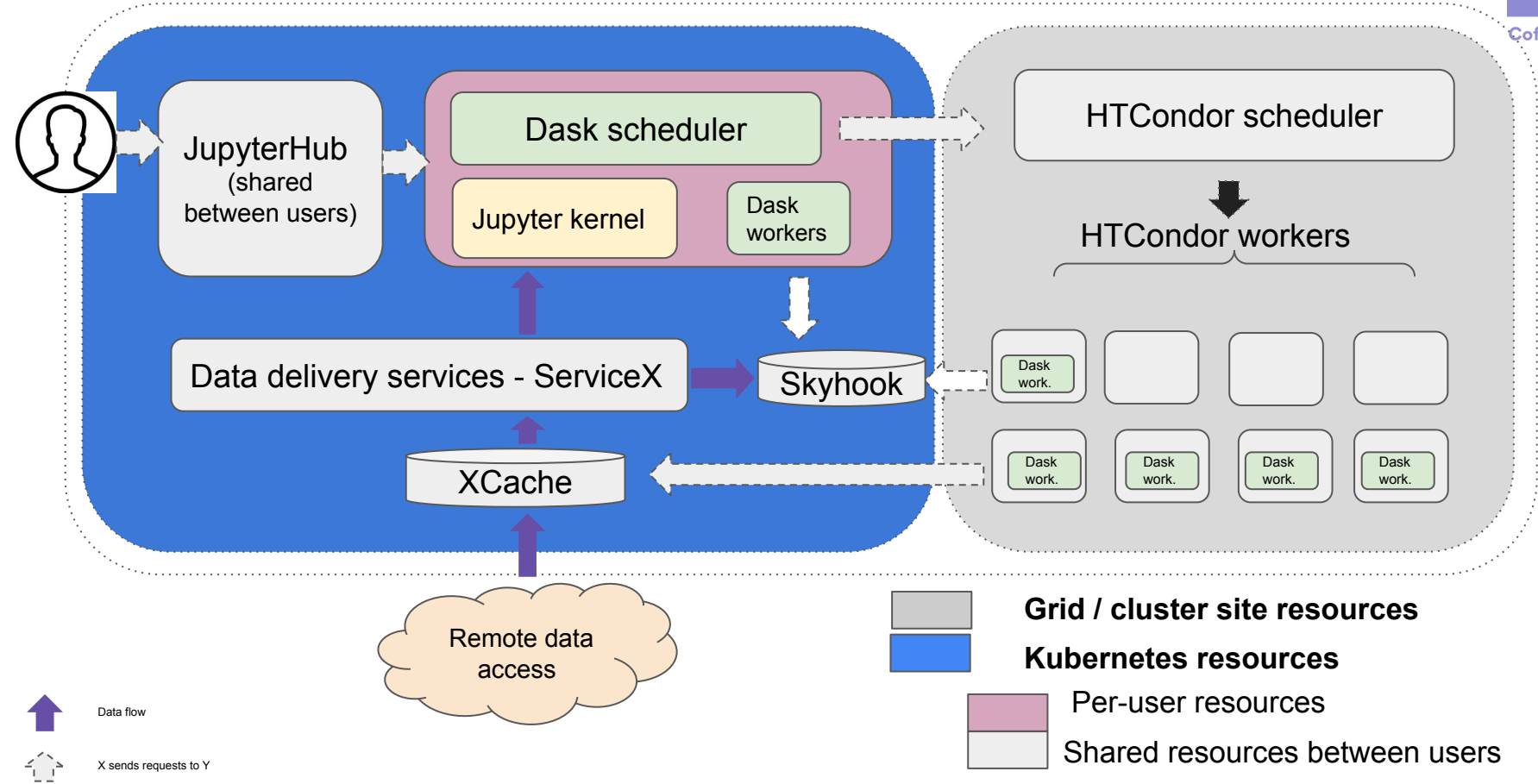
- Existing large datasets in Open Data (so far) restricted to CMS Run-1
- Would prefer to use **modern ATLAS PHYS/PHYSLITE or CMS NanoAOD formats**
 - Makes AGC implementation more relevant to current/future analyses & ideally re-usable
- Following up with ATLAS & CMS to understand whether we may be allowed to use (a) new dataset(s) for technical demonstration
 - Interest also from [ESCAPE](#) (following talk at [SwiftHep workshop](#)) / ROOT team
 - Upcoming CMS open data release expected to provide ~tens of TB within in two stages within ~next year
- **For now, building analysis pipeline based on available Open Data samples**

Building the AGC around Open Data may also help other experiments to join our efforts.

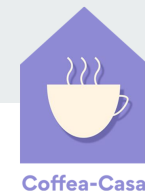
We would like to have follow-up events to the AGC workshop, where we could understand needs and discuss ideas.

Coffea-casa Analysis Facility

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



Coffea-casa deployment: existing coffea-casa AF and collaborations



- *Coffea-casa style AF facilities*, goal of adding more sites as soon as we gain experience



CMS AF @T2 Nebraska
“Coffea-casa”

<https://cmsaf-jh.unl.edu>

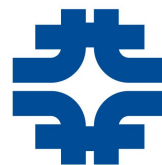
OpenData AF @T2 Nebraska
“Coffea-casa”

<https://coffea-opendata.casa>



ATLAS AF @Scalable System Lab
(UChicago)
“Coffea-casa”

New facility with ATLAS IAM, setting this up generated valuable feedback for future coffea-casa developments.



Elastic AF @ Fermilab

FNAL team is helping to test coffea-casa AF and we collaborate on FNAL EAF facility

In contact with the BNL team to evaluate the possibility to use coffea-casa experience at BNL.

IRIS-HEP AGC Tools 2021 Workshop, Nov 3–4th 2021

- Workshop showing **IRIS-HEP toolchain at coffea-casa instances**, aimed at PhD / postdoc level
 - <https://indico.cern.ch/e/agc-tools-workshop>
 - 2 afternoons CERN time (15:30 - 19:30) on **Nov 3/4**
 - Brief introductions to individual packages, notebook talks focusing on interfaces between tools
 - Using **Open Data** examples, then splitting into **ATLAS / CMS** - specific tracks
 - Interest from **LHCb & smaller experiments**: encouraged follow-up meetings to understand their needs better
- **102 registered** participants
 - *Closed registration because we were not sure if available AF resources would be able to host more participants*
 - **81 people connected** to Zoom on first day
 - Event recorded & to be shared on Youtube

Timetable

< Wed 03/11 Thu 04/11 All days >

Print PDF Full screen Detailed view Filter
Session legend

Tutorials

15:00

15:00 - 15:30	Introduction	Alexander Held et al.
15:30 - 15:50		
15:50 - 16:30	Data handling: uproot, awkward & vector	Mason Proffitt
16:30 - 16:50	Histogramming & visualization: hist & mplhep	Andrzej Novak
16:50 - 17:35	Columnar analysis with coffea	Lindsey Gray et al.
17:35 - 17:45	mini Q&A / overflow	
17:45 - 18:15	Break	
18:15 - 18:45	Queries with func_adl	Mason Proffitt
18:45 - 19:15	Data delivery with ServiceX	Kyungeon Choi
19:15 - 20:00	From data delivery to statistical inference: ServiceX, coffea, cabinetry & pyhf	Alexander Held

20:00

Timetable

< Wed 03/11 Thu 04/11 All days >

Print PDF Full screen Detailed view Filter
Session legend

Experiment specific session Tutorials

15:00

15:00 - 16:00	Data delivery and analysis pipeline with ServiceX, Skyhook & coffea		Carlos Maltahn et al.
16:00 - 16:30	Scale-out with coffea		Oksana Shadura
16:30 - 17:00	Discussion		
17:00 - 17:30	Break		
17:30 - 20:00	Experiment specific session: ATLAS	Experiment specific session: CMS	Lindsey G
17:30 - 20:00	Lukas Alexander Heinrich		

20:00

Thursday afternoon: split into ATLAS / CMS sessions

Recording... 10:46

Participants (63)

Alexander Held (he/hi/...)

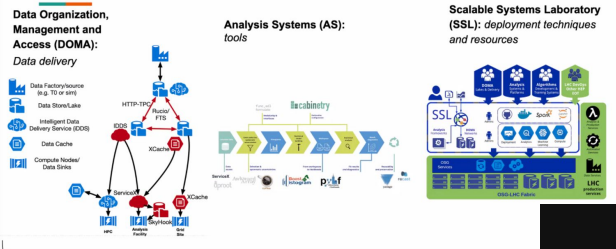
Oksana Shadura (Host)

Lindsay Gray

ram

Carl Lundstedt (Web)

The AGC is connecting different working groups and IRIS-HEP partners



Search

Alexand... (Co-host, me)

Oksana Sha... (Host)

AG Abhijith Gandrakota

AW Aj Wilridge (Guest)

AN Andrzej Novak

AM Anika Mehta (Guest)

AL Antoine Lesauvage (Guest)

AT Austin Townsend (Guest)

Ben Galewsky (Guest)

BY Brent Yates

Recording... 33:08

demo.py:nb

```
tree_arrays()
[1]: <Array [["Muon": 2, Muon_pt: [10.8, ... -1, 1]] type="100000" * ("muon": uint32,...]>
```

You can see some numbers in there, which indeed are from the data in the tree.

Branches

Now we assign this object (which contains both the names and contents of the branches) to another variable (branches):

```
branches = tree_arrays()
```

Next let's just look at each branch individually. You can access a single branch from branches in a similar way to getting an item from a ROOT file object (array-like notation):

Search

Alexand... (Co-host, me)

Oksana Shadura (Host)

Mason Proffitt (Guest)

Abdulazem Fadol Moham...

AG Abhijith Gandrakota

AW Aj Wilridge (Guest)

AE Andrew Eckart (he... (Guest)

AN Andrzej Novak

AP Anil Panta (Guest)

AM Anika Mehta (Guest)

AS Annika Stein (Guest)

AL Antoine Lesauvage (Guest)

AT Austin Townsend (Guest)

Ben Galewsky (Guest)

Brent Yates

Unmute Start Video Security Participants Chat Share Screen Pause/Stop Recording Live Transcript Reactions

Recording... 13:27

Lindsay Gray

ram

Alexander Held (he/hi/...)

Carlos Maltzahn (UC San...)

Jim Pivarski

What is Skyhook Data Management?

- Also known as SkyhookDM or just "Skyhook"
- Since 10/22/21 part of Apache Arrow (will be part of 7.0.0)
 - Columnar memory format for flat and hierarchical data
 - Large ecosystem of mapping Arrow data to storage, GPUs, FPGAs
- Offloads Apache Arrow scans into a storage system
 - Embeds the Apache Arrow library with minimal changes
 - Reduces client-side resource utilization (CPU, memory, network)
 - Faster networks → more CPU and memory BW for data movement
 - Particularly good for data-intensive selection operations
- Storage systems can optimize dataset operations based on local info
 - Fewer "magic numbers" applications have to worry about

Participants (42)

Alexand... (Co-host, me)

Oksana Shadura (Host)

Carlos Maltzahn... (Guest)

AJ Wilridge (Guest)

Andrew Wightman (Guest)

abbreviated with an ellipsis (...) so that it doesn't take up the

means that there are 100,000 entries and that each entry is a 32-bit

Recording... 01:55:56

Mason Proffitt

ram

Alexander Held (he/hi/...)

KyungEon Choi

Lindsay Gray

Oksana Shadura

Example ServiceX workflow

1. A user makes a ServiceX delivery request from Jupyter notebook via a REST interface
2. ServiceX backend looks for input datasets and retrieves an input file list
3. A relevant code is generated based on the input data format, query in func-adj, and so on
4. Transformer pods (workers) are generated to process each file (10 pods at first and scale up if necessary)
5. Outputs are streamed into the object store inside the Kubernetes cluster
6. Download outputs asynchronously

Where is ServiceX?

Recording... 01:05:56

ram

Alexander Held (he/hi/...)

Oksana Shadura

Carlos Maltzahn (UC...)

Jim Pivarski

Processing CPU Occupancy

Tasks Processing

Progress - total 100% to memory 50% processing 91% writing 0% error 0

Downloaded 202 / 172

Processed 584 / 584

Index 30 / 30

Participants (41)

Alexand... (Co-host, me)

Oksana Sha... (Host)

AG Abhijith Gandrakota

AW Aj Wilridge (Guest)

AW Andrew Wightman (Guest)

AL Antoine Lesauvage (Guest)

Ben Galewsky (Guest)

Burt Holzman (Guest)

Carl Lundstedt (Neb) (Guest)

Carlos Maltzahn (... (Guest)

CA Charlotte Ann Cooke

CM Cristina Mantilla S... (Guest)

DO Daniel Ocampo He... (Guest)

DP Danil Ponomarenko (Guest)

DK David Koch (Guest)

Unmute Start Video Security Participants Chat Share Screen Pause/Stop Recording Live Transcript Reactions Apps Leave

Participants (63)

Burt Holzman (Guest)

Carl Lundstedt (Neb) (Guest)

Charlotte Ann Cooke

Daniel Ocampo He... (Guest)

Danil Ponomarenko (Guest)

David Koch (Guest)

Invite Mute All More

Unmute Start Video Security Participants Chat Share Screen Pause/Stop Recording Live Transcript Reactions Apps Leave

Participants (41)

Alexand... (Co-host, me)

Oksana Sha... (Host)

AG Abhijith Gandrakota

AW Aj Wilridge (Guest)

AW Andrew Wightman (Guest)

AL Antoine Lesauvage (Guest)

Ben Galewsky (Guest)

Burt Holzman (Guest)

Carl Lundstedt (Neb) (Guest)

Carlos Maltzahn (... (Guest)

CA Charlotte Ann Cooke

CM Cristina Mantilla S... (Guest)

DO Daniel Ocampo He... (Guest)

DP Danil Ponomarenko (Guest)

DK David Koch (Guest)

Invite Mute All More

Analysis pipeline demonstration

- Showed ecosystem integration with **analysis pipeline example**
- Interfaced many packages, deployed on **Open Data coffea-casa**
- Analysis example is small (~500 MB input, ~1 M events), but:
 - This **approach scales!** To be shown at future events.
 - **Systematic uncertainties** included – captures realistic features

ATLAS Open Data $H \rightarrow ZZ^*$ with ServiceX, coffea, cabinetry & pyhf

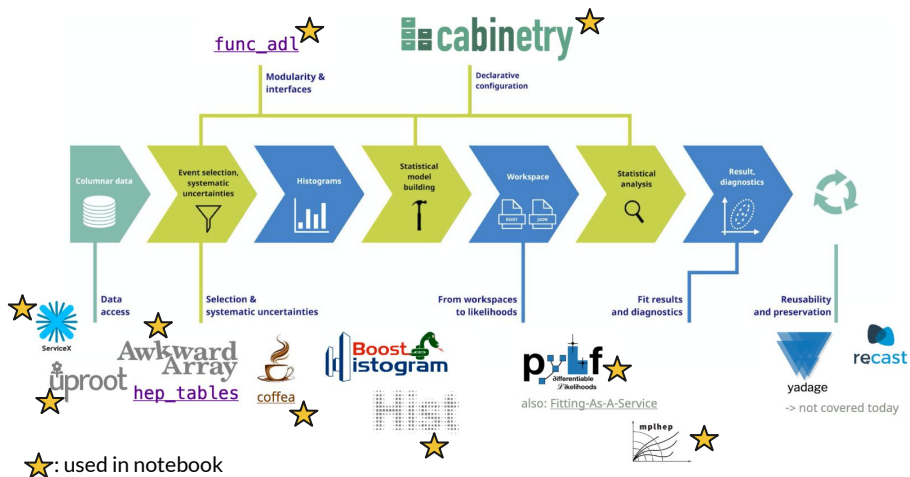
```
In [1]: import asyncio
import os
import re
import time

import awkward as ak
import cabinetry
from coffea.processor import ServiceX
from func_adl import ObjectStream
from func_adl_servicex import ServiceXSourceUpROOT
import hist
import mplhep
import numpy as np
import pyhf
import uproot

import utils
from utils import infofile # contains cross-section information

utils.clean_up() # delete output from previous runs of notebook (optional)
utils.set_logging() # configure logging output
```

full notebook: [HZZ analysis pipeline.ipynb](#)



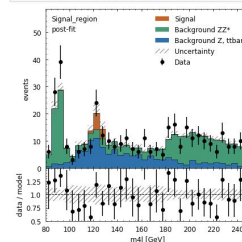
```
In [17]: # create post-fit model prediction
postfit_model = cabinetry.model_utils.prediction(model, fit_results=fit_results)

# binning to use in plot
plot_config = {
    "Regions": [
        {
            "Name": "Signal_region",
            "Binning": list(np.linspace(bin_edge_low, bin_edge_high, num_bins + 1)),
        }
    ]
}

figure_dict = cabinetry.visualize.data_mc(
    postfit_model, data, config=plot_config, save_figure=False
)

# modify x-axis label
fig = FigureDict()['figure']
fig.axes[0].set_xlabel("m[1] [GeV]")

# let's also save the figure
utils.save_figure("Signal_region_postfit")
```



AGC tools workshop: involved computing resources

Opendata Coffea-casa @ UNL

Used during generic session: 29 users connected and following tutorials / talks

ATLAS Coffea-casa @ SSL

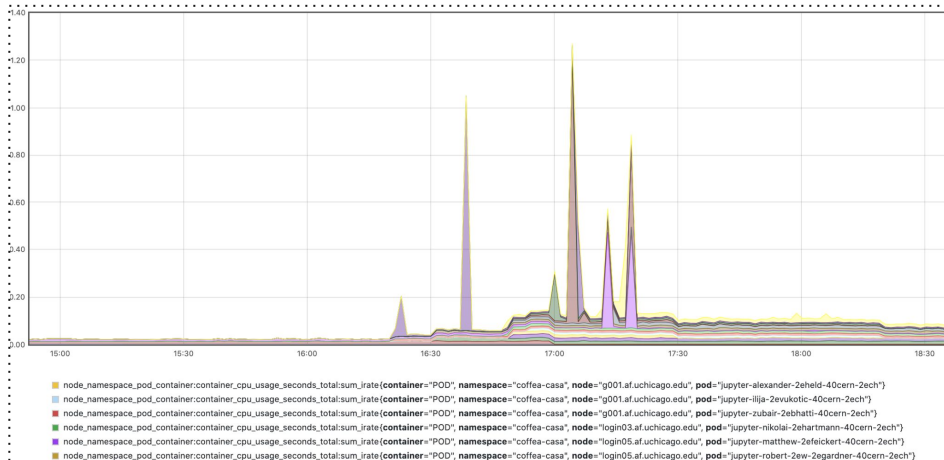
Used during ATLAS experiment session



Binderized repositories (except last two talks in agenda)

jupyterhub Home Token Admin okana.shadura@cern.ch

User	Admin	Last Activity	Running (29)	Shutdown Hub
Add Users				
[Stop All] [Shutdown Hub]				
sukeerthi.dharani@desy.de		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
lanna.osborne@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
daniil.ponomarenko@cern.ch		9 minutes ago	[stop server] [access server] [edit user] [delete user]	
engin.eren@desy.de		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
miroo.tracolli@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
douglas.benjamin@cern.ch		an hour ago	[stop server] [access server] [edit user] [delete user]	
rind@bnl.gov		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
lindsey.gray@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
matteo.marcegiani@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
clundstedt@unl.edu		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
matthew.feickert@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
nicholas.james.manganelli@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
kilian.liret@mu.de		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
msri@illinois.edu		a few seconds ago	[stop server] [access server] [edit user] [delete user]	
nikolai.hartmann@cern.ch		a few seconds ago	[stop server] [access server] [edit user] [delete user]	



ATLAS and CMS sessions

ATLAS

- Peak: 24 people connected
- Mix of topics:
 - UChicago AF
 - Integrating new techniques into existing workflows (ServiceX usage in TReXFinder)
 - Columnar analysis with PHYSLITE (including demonstration with coffea)
 - Brainstorming: towards usage of ATLAS Combined Performance (CP) tools with columnar analysis

The image displays three overlapping screenshots from a Zoom meeting. The top-left screenshot shows a slide titled "Event data model using awkward array" with bullet points: "Included PHYSLITE schema into the coffea NanoEvents module", "TODO List: CoffeaTeam/coffea#40", and "see interactive tutorial after this presentation". Below the text is a code snippet for creating a NanoEventFactory and generating events. The top-right screenshot shows a slide titled "Example analysis workflows with ServiceX" listing workflows like "Lupoid ServiceX - ROOT based Analysis" and "Uproot ServiceX - TReXFinder". It also includes a section for "Uproot ServiceX + ROOT-based Analysis" with a "Story" and a "Note". The bottom-right screenshot shows a Zoom meeting agenda for "ATLAS-Internal Meeting" on Thursday 4 Nov 2021, 17:00 - 22:00. The agenda includes: "Introduction" (17:30-17:45) by Lukas Alexander Heinrich; "U Chicago AF" (17:45-18:00) by Alexander Held; "ServiceX + TReXFinder" (18:00-18:30) by Kyungeon Choi; "Columnar Analysis with PHYSLITE" (18:30-19:00) by Nikolai Hartmann; and "CP Tools with Columnar Analysis" (19:00-19:30) by Matthew Feickert.

ATLAS and CMS sessions

CMS

- Peak: 22 people connected
- Mix of topics:
 - H $\rightarrow\gamma\gamma$ BDT executed on GPUs @ FNAL Elastic AF
 - SUEP analysis framework (using some of AGC tools)
 - Topcoffea analysis framework (using some of AGC tools)
 - Discussion: new interesting features and needed functionality to be implemented in coffea

Conclusion

Coffea based framework:

1. **We use PF candidates matched to tracks**
 - <https://github.com/SUEPPhysics/SUEPNano> (PFNano)
 - Use coffea framework to make hdf5 files with pandas dataframes
 - Recluster using fastjet (awkward array input)
2. **Significant increase in speed compared to NanoAOD-tools**
 - 40 Hz \rightarrow 1 kHz
 - Run through coffea-dask singularity image on Condor
3. **Possible set up Dask to submit condor jobs**
 - Dask works on MIT T3
 - Bookkeeping to handle xsec and gunsumweights
 - Currently controlled by python script
4. **Limit code set up to match and run on python3**
 - Still some dependence on ROOT (only histograms, no trees)
 - Works well with JupyterHub setups

IRIS-HEP AGC Experiment specific session: CMS

Thursday 4 Nov 2021, 17:30 \rightarrow 20:00 Europe/Zurich

Lindsay Gray (Fermi National Accelerator Lab. (US))

Videoconference

IRIS-HEP AGC Experiment specific session: CMS

17:30 \rightarrow 18:00 **Hgg BDT executed on GPUs @ FNAL Elastic AF** (30m)

Speaker: Lindsay Gray (Fermi National Accelerator Lab. (US))

Chained Quantile Re... code ElasticAF (needs FN... nVidia Triton

18:00 \rightarrow 18:30 **SUEP analysis framework** (30m)

Speaker: Chad Wells Freer (Massachusetts Inst. of Technology (US))

MIT_coffea.pdf

18:30 \rightarrow 19:00 **Topcoffea analysis framework** (30m)

Speaker: Andrew Wightman (University of Nebraska Lincoln (US))

topcoffea_framework...

19:00 \rightarrow 20:00 **Discussion** (1h)

Next AGC milestones

- **Dec 1, 2021:** Demonstrate ServiceX -> coffea -> cabinetry -> pyhf differentiable programming roadmap (analysis-grand-challenge/issues/1)
 - Demonstrated at AGC workshop:
<https://indico.cern.ch/event/1076231/contributions/4560405/>
- **Dec 1, 2021:** Execute IRIS-HEP AGC tools soft-launch event (analysis-grand-challenge/issues/2)
 - Done on Nov 3/4: <https://indico.cern.ch/event/1076231/>
- **June 1, 2022:** Coordinate with AS, DOMA, SSL, and operations programs to benchmark performance of prototype system components to be used for Analysis Grand Challenge (analysis-grand-challenge/issues/5)
 - In progress

Towards the next major milestone: June 1, 2022

- Improve **experiment-related coffea-casa setups** (e.g. improve experiment specific data access and other features)
 - Test integration of **SkyHook in coffea-casa@UNL and SSL (as a testbed)**:
 - **Target: ready before end of year**
- **Deploy and test** all packages and services (e.g. related to AGC) at various **analysis facilities**
- **Benchmark performance** of prototype system components for AGC
- Work with HSF DAWG group about specification of new **sub-benchmarks** as a potential new milestone for AGC
- Develop **analysis example** used for next round of demonstration (possibly based on new CMS Open Data)

Follow-up events after the AGC workshop

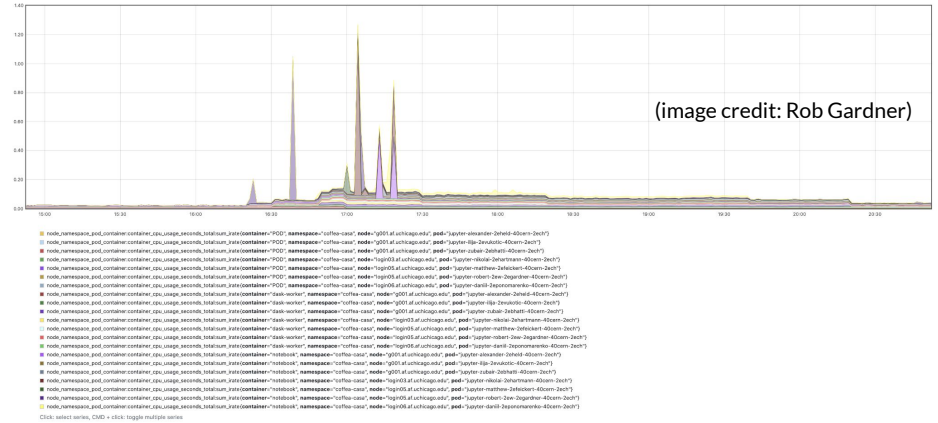
- Identified need for **technical meeting** to **coordinate analysis facility developments**
 - Many similar efforts ongoing at different sites / groups
 - Aiming for meeting in January
- Dedicated **tutorial / training event**: lower entry level, aimed at PhD students new to tools
 - Synergies with **HSF** training efforts & **IRIS-HEP training grand challenge**
- Another **showcase event** around spring / summer, possibly aligned with next milestone
 - Could be combined with training event
- Event(s) to address needs of **experiments beyond ATLAS / CMS**
 - Synergies with **HSF**, we are particularly interested in how our tools / workflows may map to other usecases



Backup slides

Latest addition: ATLAS coffea-casa AF @ UChicago

- Thanks to efforts of Lincoln Bryant, Fengping Hu, Rob Gardner, Ilija Vukotic, Suchandra Thapa, we managed to deploy coffea-casa AF @UChicago
 - Using ATLAS IAM
 - A lot of very valuable feedback and suggestion how to improve coffea-casa AF
 - *Next step is to enable efficient data access for ATLAS data*



<https://coffa.af.uchicago.edu>

Approach for AGC analysis definition

In 2021

- stick to **ATLAS H>ZZ*** example for demonstrations, allows testing interfaces and is ready
- identify possibility of using datasets with new ATLAS / CMS formats in parallel
 - May involve re-formatting existing Open Data if no new datasets become available

June 2022 milestone (“benchmark performance of prototype system components for AGC”)

- If **modern ATLAS / CMS formats** are available: design analysis around what samples we get
- Otherwise: extend **CMS H>tautau Open Data** analysis ([Open Data record](#)) with systematic uncertainties covering all uncertainty types identified in taxonomy (different types requiring different approaches)

March 2023 milestone (execute AGC)

- If **new ATLAS / CMS data** is unavailable: (reformatted) **CMS Run-1 Open Data**-based analysis

Benchmarks



- **HSF DAWG** interested in expanding existing [ADL benchmarks](#)
- **HSF DAWG and AGC** identified several potential **directions for extensions**
 - *Testing interfaces between different tools in analysis pipeline*
 - *Handling of systematic uncertainties*
- **Idea:** to specify **AGC** sufficiently well so that it can be used as **very large benchmark**
 - Also want to split into **sub-tasks** that can be used for benchmarks
 - Detailed specification may attract other users to write new implementations
- Specification of N new benchmarks potential new **milestone for AGC**

Past relevant milestones (selection)



- **Mar 1, 2021 (AS): G2.6** Scoping and specification of the target analysis to be used in the Analysis Grand Challenge
- **Mar 1, 2021 (AS): G2.7** Baseline programming Interfaces between components like ServiceX, func_ADL, HEP_tables, Coffea, cabinetry, and pyhf

Next relevant milestones related to AGC

- **Dec 1, 2021 (AS): G2.9 Differentiable programming roadmap** across services needed for analysis challenge
 - *Thinking about the possibility of a blueprint meeting in November to address this*
- **June 1, 2022 (AS, DOMA, SSL): G2.15** Coordinate with DOMA, SSL, and operations programs to benchmark **performance of prototype system components** to be used for Analysis Grand Challenge
 - *Define and test interoperability of software components (services and packages) to be used for AGC and show the demonstrator(s)*
 - *Select dataset(s) and examples (CMS Run1 AOD, NanoAOD or DAOD) applicable to be used for AGC (from G2.6)*
 - *Deploy multiple ATLAS, CMS and Open Data coffea-casa instances ready for benchmarking*
- **Sep 1, 2022 (management): G1.13 Blueprint workshops** (Grand Challenges Planning)
 - *Falls in between prototype phase (due June 1, 2022) and execution (March 1, 2023), could be used to identify issues based on experience with the prototype and figure out a plan to address them*

Next relevant milestones related to AGC

- **Dec 1, 2022 (DOMA): G3.11** Demonstrate ability to filter / process data at rate necessary for analysis challenge using SkyHook
 - *Prepare coffea-casa setup ready for benchmarking (related to G2.15)*
- **Mar 1, 2023 (AS, DOMA, SSL) G2.16** Coordinate with DOMA, SSL, and operations programs to execute the Analysis Grand Challenge
- **Dec 1, 2023 (management) G1.22** Blueprint workshops (Grand Challenges Progress)