



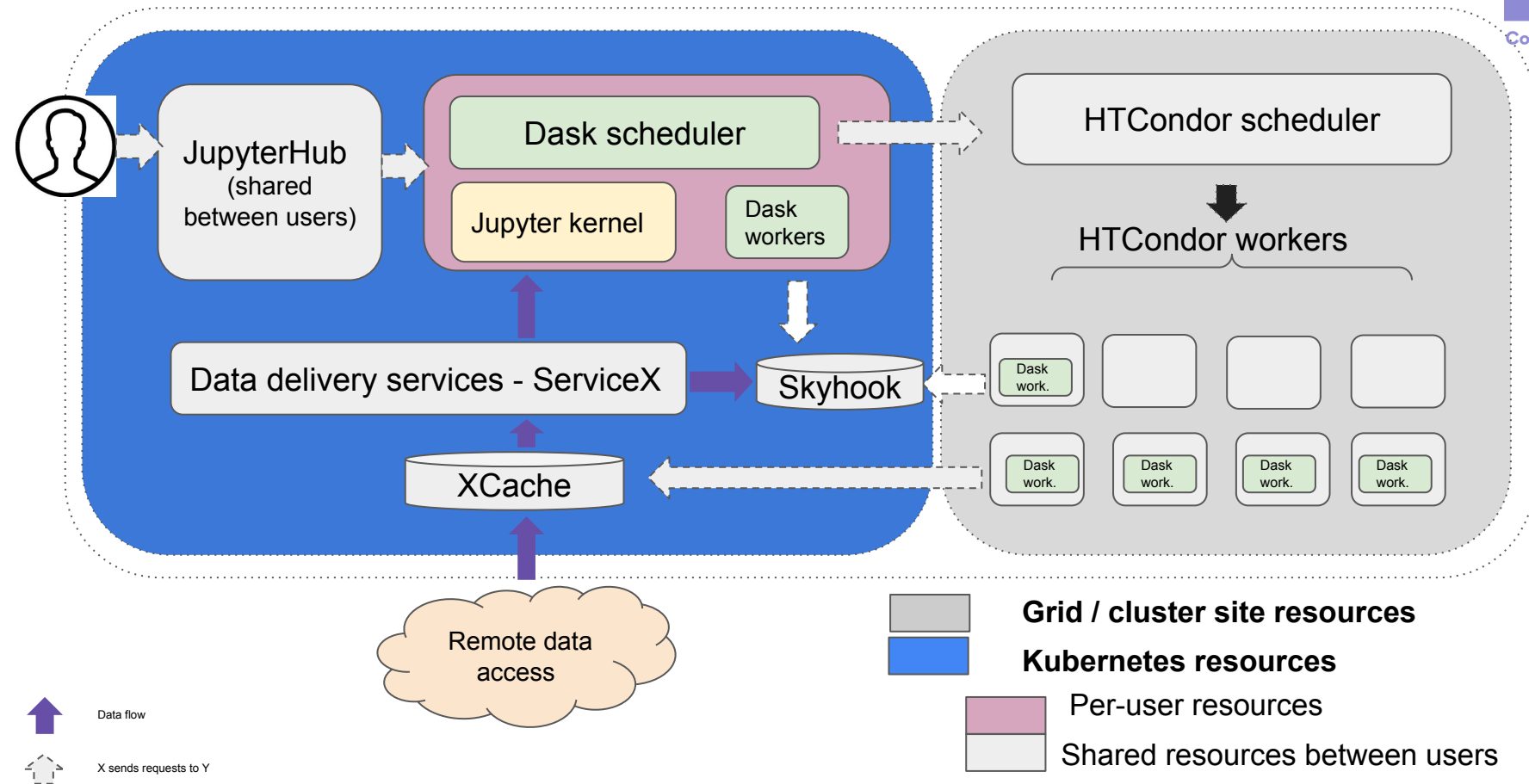
Coffea-casa news and development

Oksana Shadura
On behalf of coffea-casa development team

<https://indico.cern.ch/event/1098937/>

Coffea-casa Analysis Facility

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



Coffea-casa deployment: existing coffea-casa AF

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



CMSAF @T2 Nebraska
“Coffea-casa”
<https://coffea.casa>

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.

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

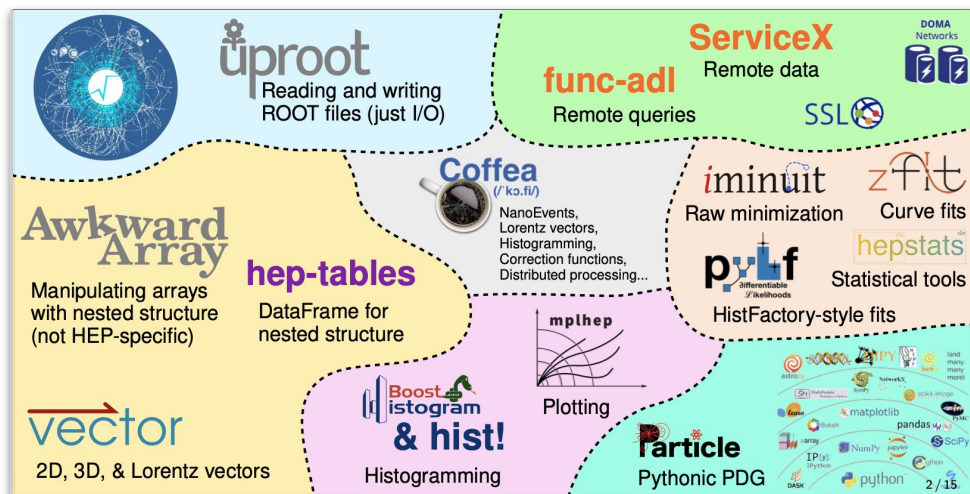


Coffea-casa vCHEP 2021 plenary talk

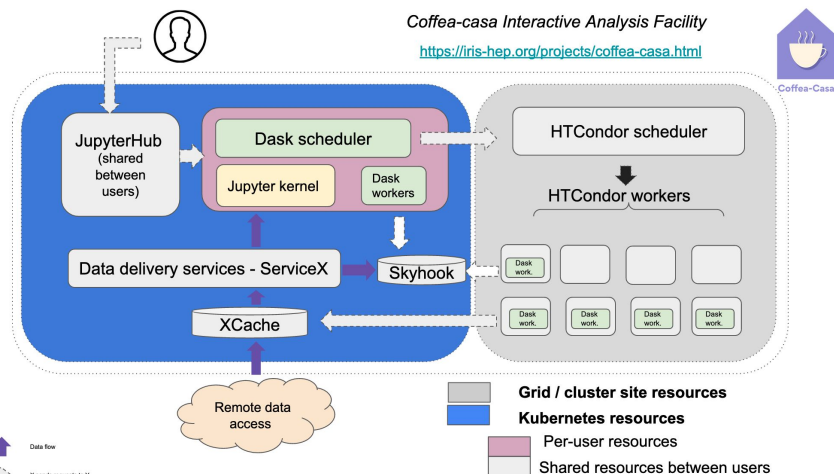
Analysis Grand Challenge will be conducted during **2021–2023**, leaving enough time for tuning software tools and services developed as a part of the IRIS-HEP ecosystem before the start-up of the HL-LHC and *organized together with the US LHC Operations programs, the LHC experiments and other partners.*

The idea behind the Analysis Grand Challenge

Analysis Tools



Analysis Facilities



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



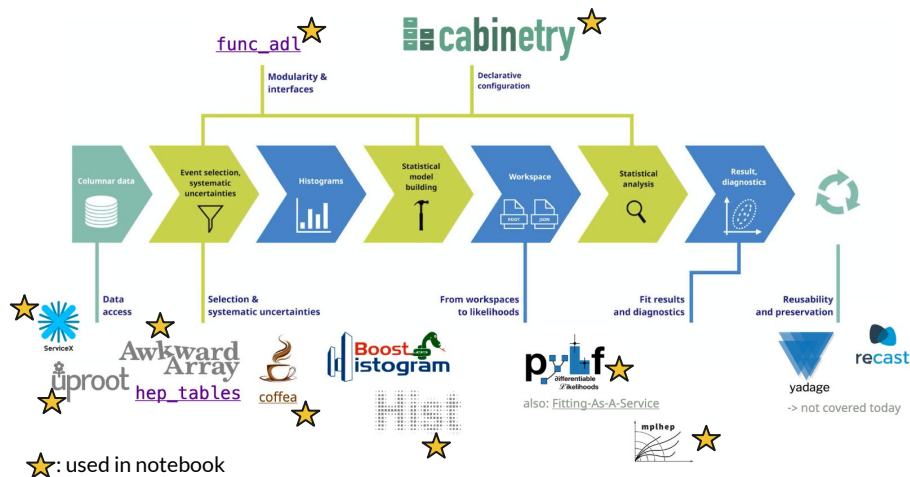
- Workshop showing **IRIS-HEP toolchain at coffea-casa instances**, aimed at PhD / postdoc level
 - <https://indico.cern.ch/e/agc-tools-workshop>
 - 2 afternoons CERN time (15:30 - 19:30) on **Nov 3/4**
 - Brief introductions to individual packages, notebook talks focusing on interfaces between tools
 - 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

Materials:

<https://github.com/iris-hep/analysis-grand-challenge/tree/main/workshops/agctools2021>

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](#)

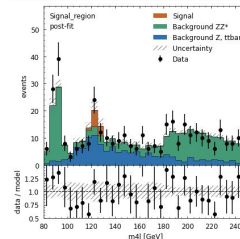
```
In [17]: # create post-fit model prediction
postfit_model = cabinetry.model_utils.prediction(model, fit_results=fit_results)

# 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 = figure_dict[0]["figure"]
fig.axes[0].set_xlabel("m1 [GeV]")

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



Next milestones



2022-02-01

No due date ⌚ Last updated 10 days ago

0% complete 18 open 0 closed

[Edit](#) [Close](#) [Delete](#)

- [Dask dependency management plugin for coffea-casa](#)
- [Add CVMFS to be available from notebook](#)
- [Create mailing lists for OD and CMS AFs @ UNL](#)
- [Preparation for ACG Tools: Skyhook](#)
- [Deploy monitoring for coffea-casa @ UNL: prometheus/elasticsearch instances](#)
- [Provide unique HTCondor users for each coffea-casa user](#)
- Add deployment documentation
- And many more!



Coffea-casa AF statistics

CMS instance:

- 27 active* users (some of them ~ 14 are using AF connected in the last 24 hours)
 - We had in total 100 users who ever used coffea-casa
- 8-9 git repositories with analysis tried on coffea-casa on a different stage of development (I found in Github)

* - *total stats over last month*

Please help us to become better:

<https://github.com/CoffeaTeam/coffea-casa/discussions>