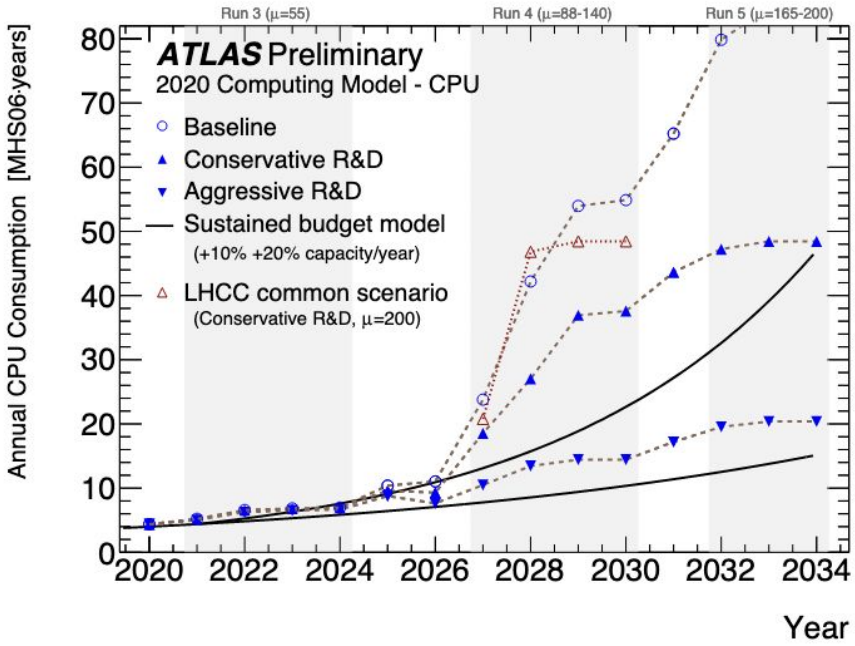
Analysis Facilities

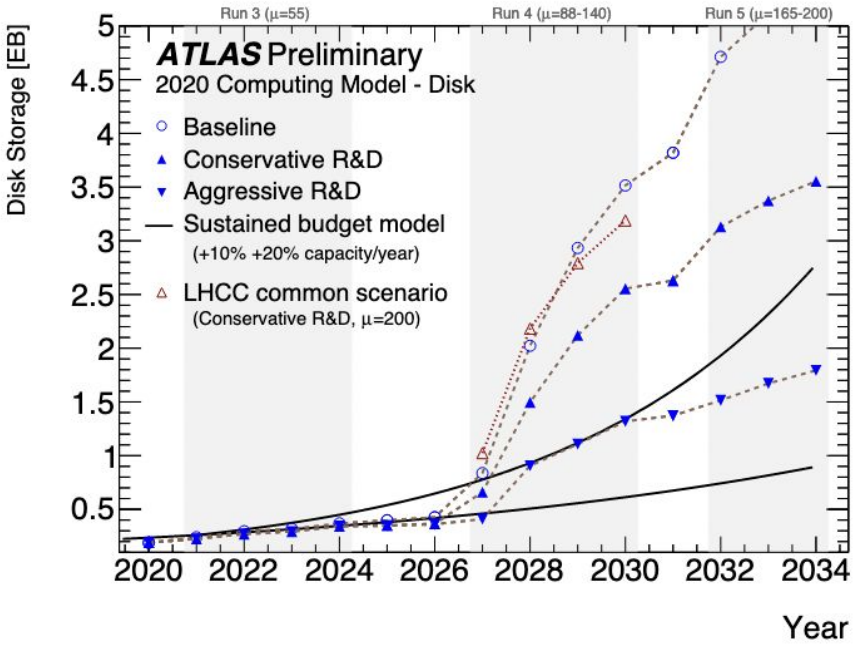
A. Forti

GridPP47

24 March 2022

HL-LHC explosion of Disk and CPU requirements





- Even restricting to users only
  - With current run2 data model ~ 100PB data set
    - Without redundancy
    - T3 don't have the storage/computing capabilities
  - Processing time with sequential algorithms also doesn't scale
    - Need to find other methods

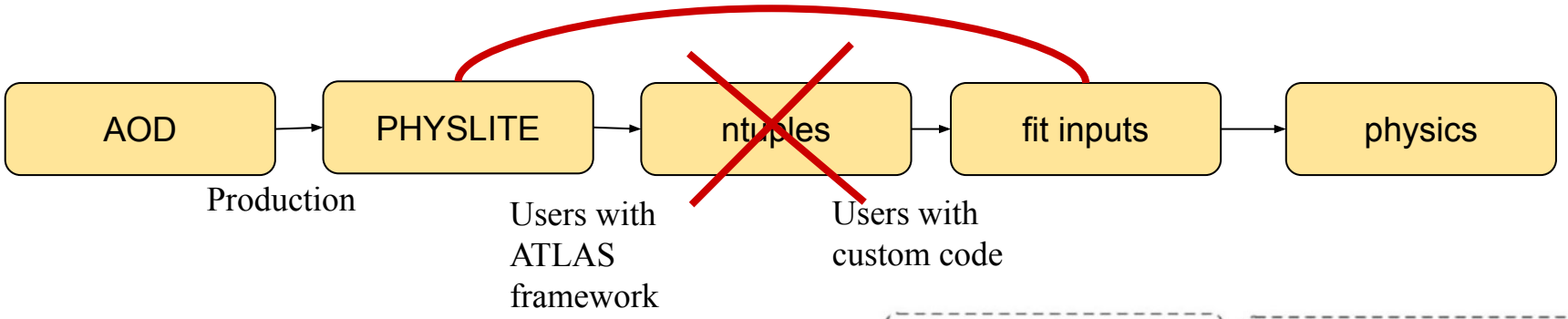
New formats in ATLAS

- DAOD_PHYS
  - Run3 format smaller, more compact but no changes in analysis
  - All events no duplication
- DAOD_PHYSLITE:
  - pre-calculated, calibrated objects
  - 10 kB/event
  - For Run4 but to be picked up **during run3**
    - First production
  - Users dataset size 2 PB/year
  - Still a root format (data structure TTree -> RNtuple)
  - 80% of analysis expected to use this format

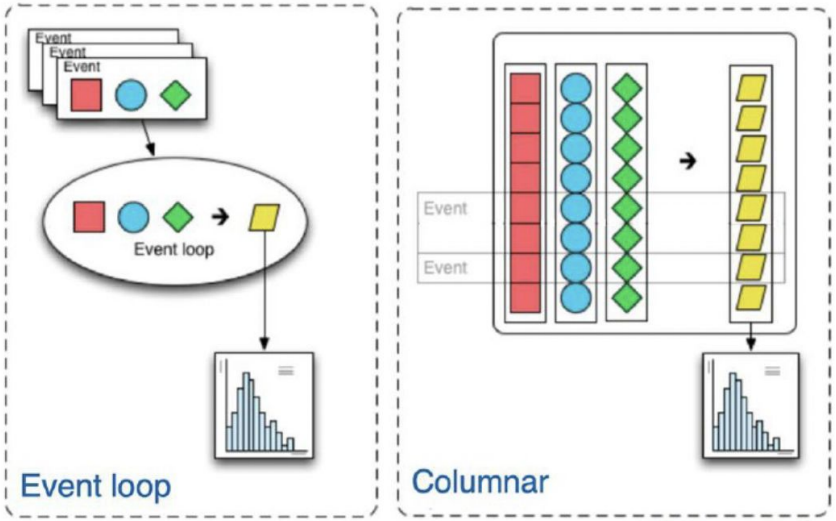
xAOD Type	Size per event
AOD	600 kB
DAOD	40 – 450 kB
DAOD_PHYS	50 kB
DAOD_PHYSLITE	10 kB

AIM

- Aim is to produce Histograms directly from PHYSLITE skipping the creation of ntuples which is a painful step for the users and requires a non negligible amount of space for intermediate data

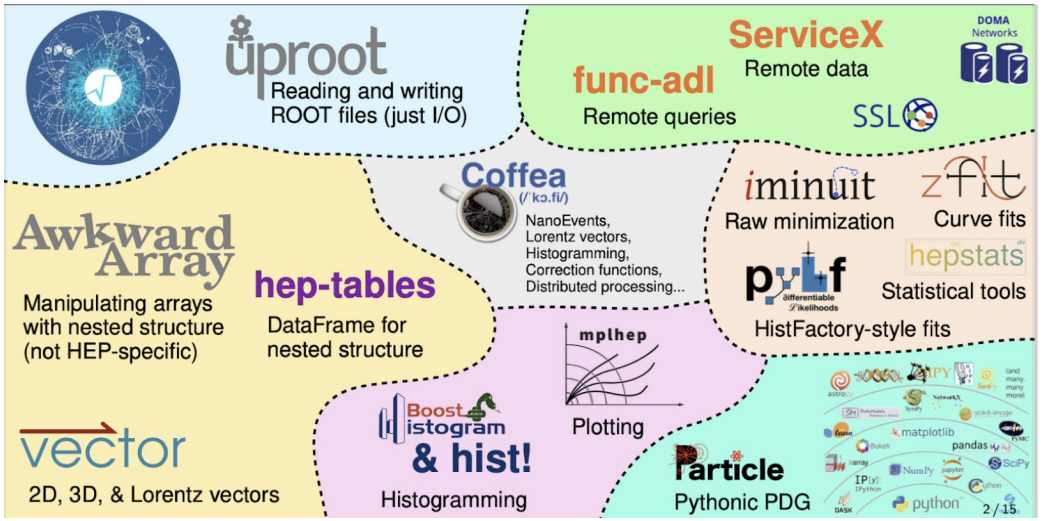


- NOT use the ATLAS framework after PHYSLITE is created
  - Not true at the moment due to systematic variations
- Columnar analysis
- Parallelization
- Analytics techniques



Analysis Evolution

- Tools evolution
  - More lightweight, decoupled from experiment frameworks
  - Ecosystem developed inside and outside of HEP
    - pandas, uproot, Dask, HDF5, scikit-hep, matplotlib, awkward array
  - ROOT ecosystem also evolving to allow for more efficient storage and I/O and parallelism
  - Data transformation and delivery services (idds, serviceX)
  - Columnar analysis services (coffea-casa, dask)
  - Machine Learning workflows (active learning, deep learning, HPO,...) multi layer
  - k8s
- More light weight for users means more complicated infrastructure



Analysis Facilities

- IRIS-HEP/HSF AF definition
  - *Infrastructure and services that provide integrated data, software and computational resources to execute one or more elements of an analysis workflow. These resources are shared among members of a virtual organization and supported by that organization.*
- This is quite generic because we don't know what the final implementation will be or if there will only be one size fits all yet
- A lot of discussion about what is needed but the only way to assess if a solution is valid is **testing**
- So what to test? and where? and who should learn about all this

User testing

- Example of tests made to better understand file formats
  - python tools: pandas, uproot, dask, awkward,....
  - file formats: parquet, HDF5, npz, (Up)root,....
- Some tests with ServiceX
- Full Dask PHYSLITE demo

Alternative storage formats

Loading times for all columns (≈1000) of 10k PHYSLITE events:

Format	Compression	Dedup. offsets	Size on disk	Execution time
(Up)root	zlib	No	117 MB	6.0 s
(Up)root (large baskets)	zlib	No	116 MB	5.0 s
Parquet	snappy	No	121 MB	0.6 s
Parquet	snappy	Yes	118 MB	0.6 s
HDF5	gzip	No	101 MB	2.0 s
HDF5	gzip	Yes	89 MB	1.6 s
HDF5	lzf	No	137 MI	
HDF5	lzf	Yes	113 MI	
npz	zip	No	92 MB	
npz	zip	Yes	82 MB	

Parquet seems especially promising, but all tested formats f

(Note: constant overhead for Uproot (≈2s), will be less significant for lar

Example analysis workflows with ServiceX

- Uproot ServiceX + ROOT-based Analysis
- Uproot ServiceX + coffea analysis
- Uproot ServiceX + TRExFitter

Demo: `DAOD_PHYSLITE` analysis with uproot/awkward on jupyterhub on GCP

Note: This tutorial is targeted at users interested in R&D and technical details. Much of this is still in early development/prototyping.

The image that runs on jupyterhub and the dask workers is defined by the following Dockerfile:

https://github.com/gcp4hep/analysis-cluster/blob/16fb374fe26948081cf3f3b02117d05366d96520/daskhub/docker/jupyter-physlite/Dockerfile

Read and process PHYSLITE using uproot/awkward

First, let's start with some general notes on reading `DAOD_PHYSLITE`

The PHYSLITE ROOT files currently follow a similar structure as regular ATLAS xAODs

They containing several trees, where the one holding the actual data is called `CollectionTree`. The others contain various forms of Metadata.

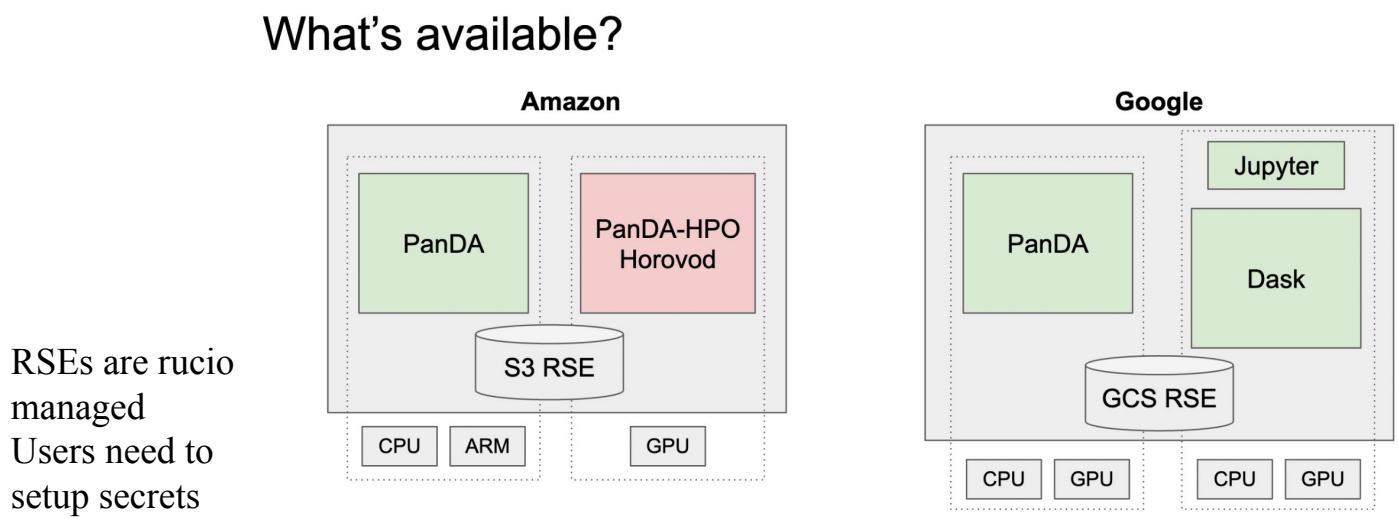
```
import os

import uproot
import awkward as ak
import numpy as np
import matplotlib.pyplot as plt
```

Jupyter notebooks

N. Hartmann, K Choi

Cloud

- Since panda works on kubernetes ATLAS is using heavily commercial clouds for all the testing
  - ARM nightlies, analysis evolution, "scaling out" testing, DAG type workflows
- Submission via panda or direct access both work



RSEs are rucio managed
Users need to setup secrets

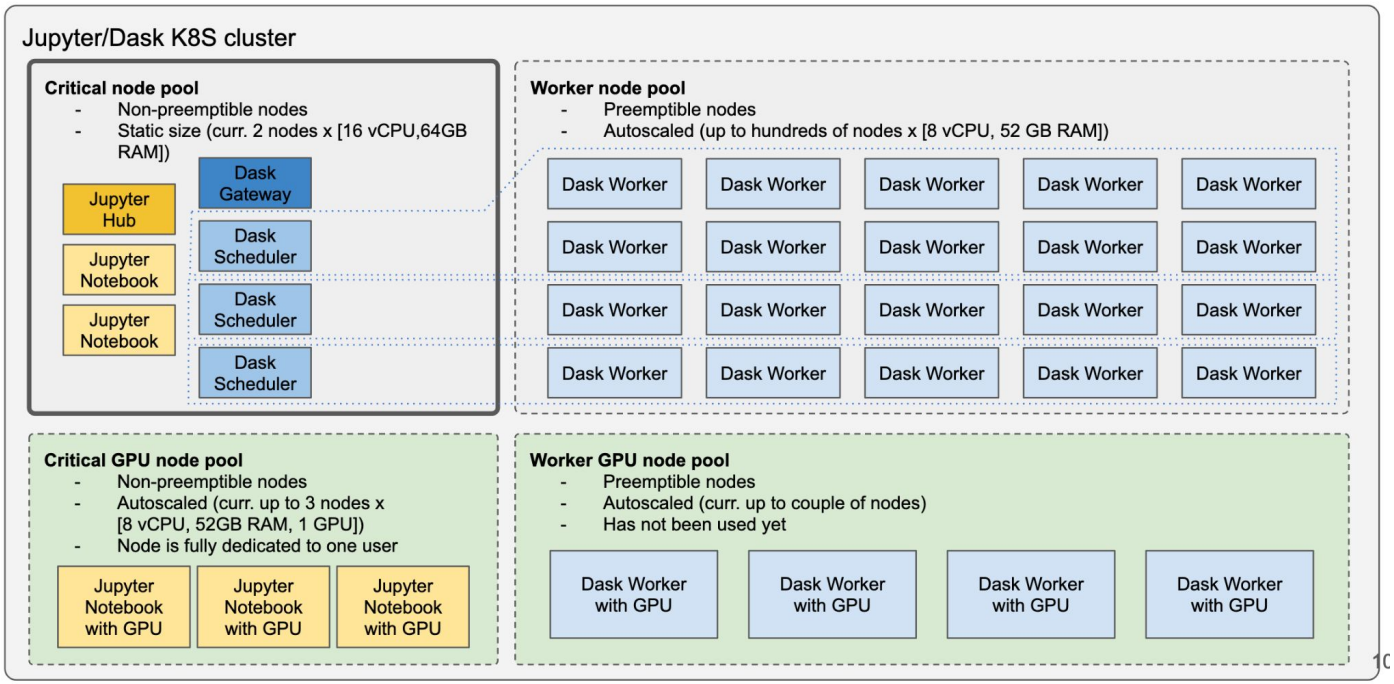
\+ Users can explore Amazon/Google infrastructure and services on their own: FPGA/GPU/ARM/XXL nodes, cloud AI platforms…

More details

- Any ATLAS user with CERN SSO

Sign in with ATLAS IAM

- kubernetes/dask/jupyter setup



- PHYSLITE environment setup

PHYSLITE environment
environment for experiments with DAOD_PHYSLITE with uproot, awkward etc.

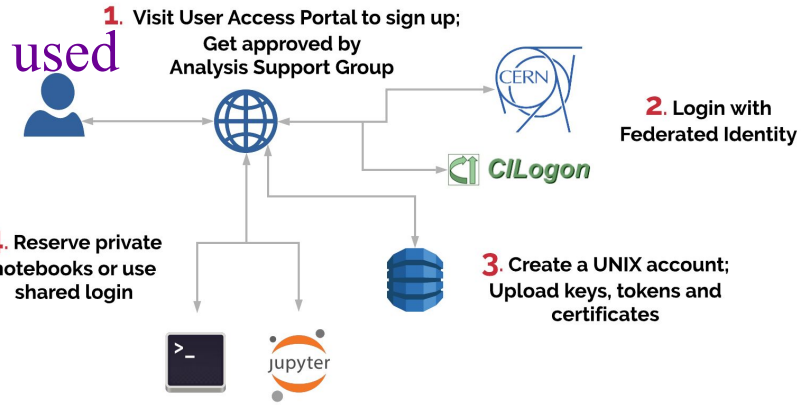
Evolution US/Chicago

- Three modes of access:
  - "Traditional" Tier3 (ssh access & HTCondor batch)
  - Jupyter hub scheduling to CPU and GPU
  - Enhanced notebook access with parallel columnar processing (Coffea)
  - Data Transformation services ServiceX
  - cephFS

ServiceX endpoints for ATLAS

Type	Collaboration	Input data format	Location	Endpoint	Purpose
Stand-alone	ATLAS	ROOT Ntuple	UC Analysis Facility	https://uproot-atlas.servicex.af.uchicago.edu/	Production
Stand-alone	ATLAS	xAOD	UC Analysis Facility	https://xaod.servicex.af.uchicago.edu/	Production
Stand-alone	ATLAS	ROOT Ntuple	SSL-River	https://uproot-atlas.servicex.ssl-hep.org/	Development
Stand-alone	ATLAS	xAOD	SSL-River	https://xaod.servicex.ssl-hep.org/	Development

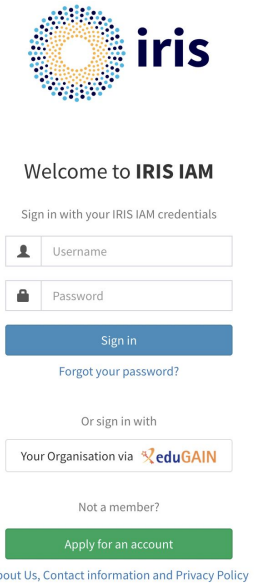
- Accessible to all ATLAS users via CERN SSO
  - Federated identity
  - >170 registered users but ATM only lightly used
- Working on a federated US T3
  - Accessible via Jupyterhub
  - Software via CVMFS
  - No common storage unfortunately

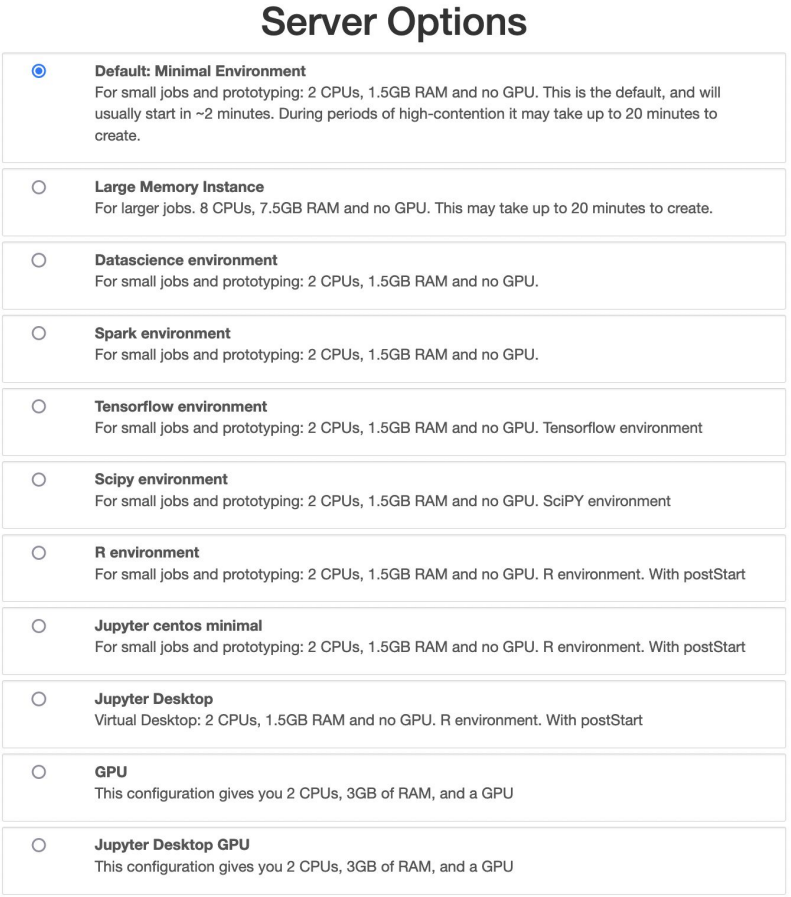


L. Bryant F. Hu I. Vukotic S. Thapa J Stephen D Jordan

UK/RAL

- Large OpenStack Cloud
  - Experiments can start their own services
    - Requires system administration skills
    - Smart users? but who maintains?
- Jupyterhub access
  - UK IRIS IAM / edugain credentials
  - (ATLAS) users not aware of it
  - Questions about software availability and data access
  - Does the UK need to build expertise on some of the tools?
  - Can we use it as a starting point?
  - Even without building the whole infra how could users use it?





HSF AF Forum

- Proposed by IRIS-HEP and discussed at the February HSF Coordination Meeting
  - Approved with 4 co-coordinators
    - Diego Ciangottini (CMS, INFN)
    - Lukas Heinrich (ATLAS, Munich)
    - Nicole Skidmore (LHCb, Manchester)
    - Alessandra Forti (WLCG/ATLAS, Manchester)
- Forum for discussion on the development of analysis facilities among different communities
  - Users
  - Workflows developers
  - Data centres infrastructure people
  - Experiments computing

Why another forum

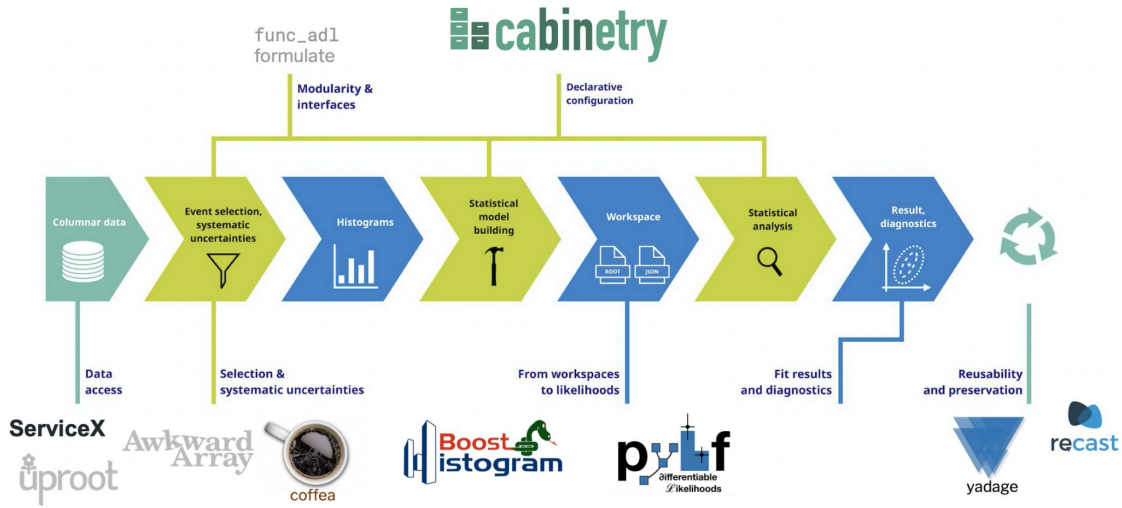
- Analysis Facilities is a hot and heated topic
  - Analysis for HL-LHC is going to change
  - Solutions might be eventually experiment specific but there is no forum where to exchange ideas
- We have the GDB
  - Monthly non dedicated meeting
- HSF Analysis Group
  - Dedicated to discuss analysis workflows analysis software not infrastructure
- Various experiment activities
  - Tend to be isolated different projects going in parallel
- Hope is to attract technical people who are willing to test new things
  - Aim is to help build/prune all these solutions
  - White paper at the end of the year

HSF AF/DOMA

- For DOMA the important part is of course the storage foundation of an analysis facility
  - EOS at CERN is the best example of what users want in analysis facility uniform name space, POSIX compliant, accessible from all resources
    - Can this be replicated?
    - Do we need other solutions?
    - We need to make sure it is integrated with the experiments data management?
- Object Stores (in UK definitely something to look into
  - Interesting Graeme's slide about RNtuple on DAOS
- Something else to look at CS3MESH4EOSC aiming at optimizing jupyter notebook access to data
  - *For effective collaboration, we need to allow scientists to easily access resources on different institutions.*

IRIS-HEP AGC

- Analysis Grand Challenge can give focus to these discussions and testing with the goal of building a **prototype**.
  - *The Analysis Grand Challenge includes both integration of software components for analyzing the data as well as the deployment of the analysis software at analysis facilities.*
- A bit like the Data Challenges may be repeated every 2 years to test incrementally



O. Shadura, A. Held

Kick-off meeting

- 25 March 2022 15:00-20:00 Tomorrow!

Analysis Facilities Forum Kick-off Meeting

25 March 2022
Europe/Zurich timezone

Enter your search term

Overview
Timetable
Participant List
Code of Conduct

We are pleased to announce a kick-off meeting of the HSF/WLCG "Analysis Facilities Forum" organised in the collaboration with IRIS-HEP, which will be held virtually on March 25, 2022.

The Analysis Facilities (AF) Forum provides a community platform for those interested in contributing to the development of analysis facilities for use within HEP experiments to develop and exchange ideas. One should interpret "development" in the broad sense, including contribution of ideas from potential end users for functionality to support the analysis of HEP data, specification and planning of the facilities themselves, and technical developments needed to realize AFs. Of course, HEP experiments have their own internal processes for developing and deploying AFs. This AF Forum is intended to support and strengthen those efforts by sharing among a broader community the key ideas and developments. The AF Forum also collaborates with related HSF Working Groups, such as the Data Analysis Working Group and the PyHEP Working Group, with strong connections to WLCG.

The development of future analysis facilities is of great interest to the HEP community, with much recent progress having been achieved. There are numerous ongoing efforts to stand up AFs which utilize new tools and techniques to help make data analysis tasks easier, more performant and more reproducible.

All participants in this workshop must abide to the Code-of-Conduct

Communication

- hsf-af-forum@googlegroups.com
  - https://groups.google.com/g/hsf-af-forum
  - for group members discussion
- hs-af-forum-convenors@googlegroups.com
  - to contact the coordinators for organisation
- Mattermost Team

Conclusions

- Analysis landscape is changing
  - Techniques tools and infrastructure
- Even if we don't want to look at the more custom services like serviceX we should be careful that we don't remain back on more industry standard tools
  - k8s
  - Jupyter
  - Oauth2.0
  - Containers
- How to use RAL cloud for testing is also important
  - It might be easier to start than finding local hardware and we don't give money to google….
  - It also may help understanding the blockers if AF will be on similar resources
    - Italians also looking at testing on cloud resources at their T1
    - Germans experimenting on cloud resources on their NAF@DESY

Backup

Infrastructure

- Whatever we end up with we have few fixed points that need to be solidified and will require work on the infrastructure
  - Containers
    - Standalone containers can run everywhere
      - Not yet a widespread practice
      - Don’t have yet a solid distribution
  - AAI - x509 -> tokens
    - One of the most important aspects for users facing different type of services and to simplify integration
    - Transition has started but it is still a huge amount of work that is needed
  - Data
    - Reduction
    - Smart placement
  - Jupyter Notebooks
    - Not going anywhere and need to be integrated
      - See Tadashi slides