Analysis Facility - US ATLAS Perspective

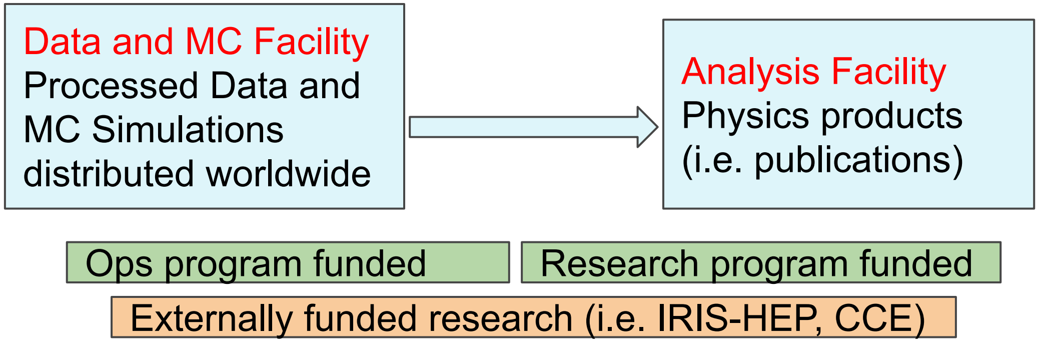
Kaushik De (UTA), Jahred Adelman (NIU),
Paolo Calafiura (LBNL), Mike Hance (UCSC),
Verena Martinez Outschoorn (UMass)

Future Analysis Systems and Facilities Workshop

October 27, 2020

What is an Analysis Facility

- Much has been said already - however, Analysis Facility (AF) probably means something different for every participant
- From US ATLAS Ops (Operations Program) perspective:



Data and MC Facility
Processed Data and MC Simulations distributed worldwide

Analysis Facility
Physics products (i.e. publications)

Ops program funded

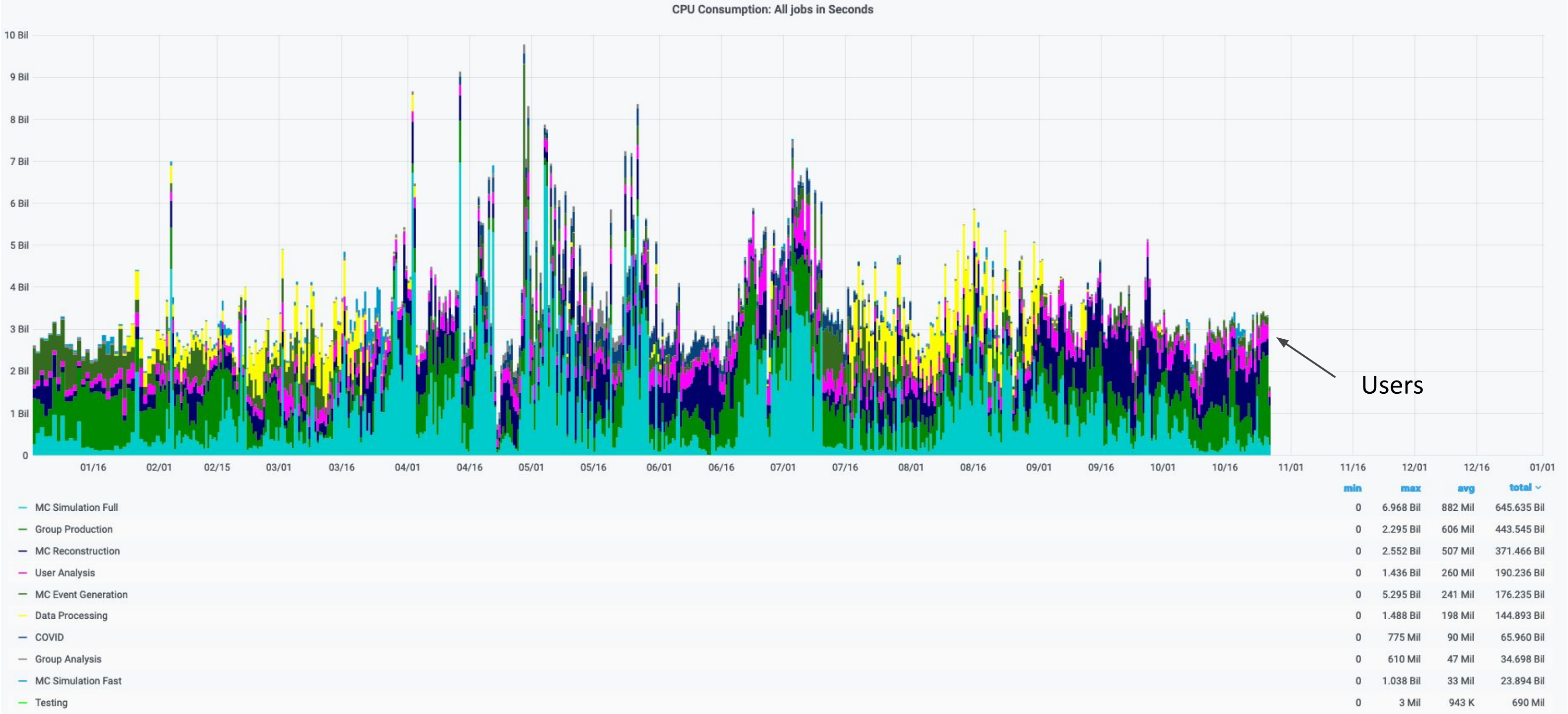
Research program funded

Externally funded research (i.e. IRIS-HEP, CCE)

- The boundary is fuzzy - often the focus of many discussions
  - Who pays for what - at the most basic level?
  - What is the functionality in each box - constantly evolving?
  - What infrastructure is needed for each box - this talk

Data and MC Facility

- US ATLAS Tier 1 and Tier 2 sites
  - We add HPC’s and commercial clouds as allocated
  - Also some opportunistic (non-tiered, non-pledged) resources
- About 10% of usage comes from user analysis jobs on the grid
  - Simulations, derivations, streaming, slimming, merging, HPO...



	min	max	avg	total ⌄
— MC Simulation Full	0	6.968 Bil	882 Mil	645.635 Bil
— Group Production	0	2.295 Bil	606 Mil	443.545 Bil
— MC Reconstruction	0	2.552 Bil	507 Mil	371.466 Bil
— User Analysis	0	1.436 Bil	260 Mil	190.236 Bil
— MC Event Generation	0	5.295 Bil	241 Mil	176.235 Bil
— Data Processing	0	1.488 Bil	198 Mil	144.893 Bil
— COVID	0	775 Mil	90 Mil	65.960 Bil
— Group Analysis	0	610 Mil	47 Mil	34.698 Bil
— MC Simulation Fast	0	1.038 Bil	33 Mil	23.894 Bil
— Testing	0	3 Mil	943 K	690 Mil

Analysis Facility

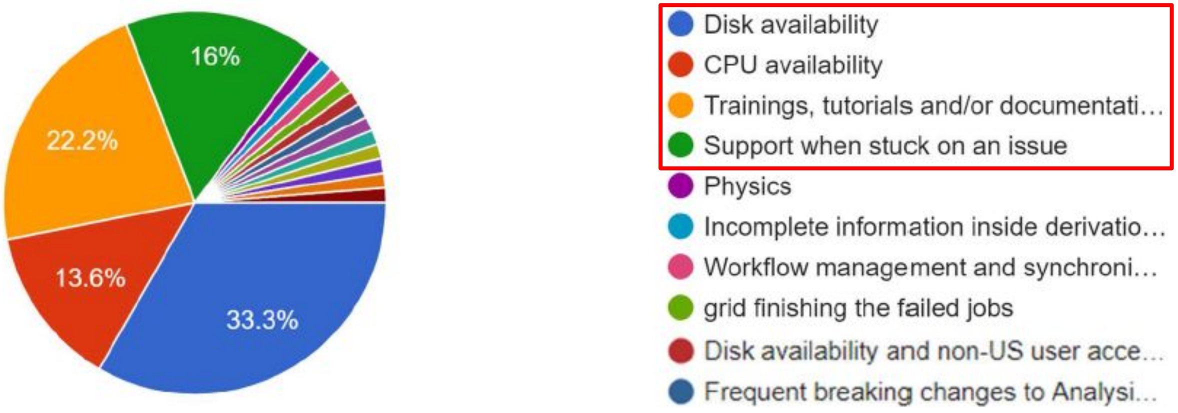
- Terminology in US ATLAS - Tier 3 or AF is interchangeable
  - They are very different from distributed T1/T2 grid facilities
- For many users, AF is their laptop/desktop
  - In addition to being the portal to ATLAS apps and services, many users actually do their computation on laptops/desktops
- For most users, AF is a small batch system
  - This is still the most popular user analysis facility - a local Tier 3 site
  - However, it is getting harder and harder for small groups to find funding for, maintain, and operate local Analysis Facilities
- For many users, AF is a big shared batch system
  - All US ATLAS users have access to two Shared T3s, a 3rd is being built
  - These AFs have been battle hardened during Run 2 (~4 years)
  - They provide usability apps, grid tools, ATLAS software and apps, derived data access, batch slots, local storage, Jupyter etc
  - These shared AFs are funded by US ATLAS - not research groups

Lessons Learned so far

- We have >10 years experience with AFs in US ATLAS
  - Every user needs something different to be productive
  - We provide the facility, training and support - users do analysis
  - Users vote with their feet - top down seldom works
  - We need robust, flexible, and easy to use systems that can evolve
  - All of the above are obvious - but worth repeating

What is your biggest limitation when you are trying to do analysis?

81 responses



Recent US ATLAS user survey

Ideal US ATLAS AF

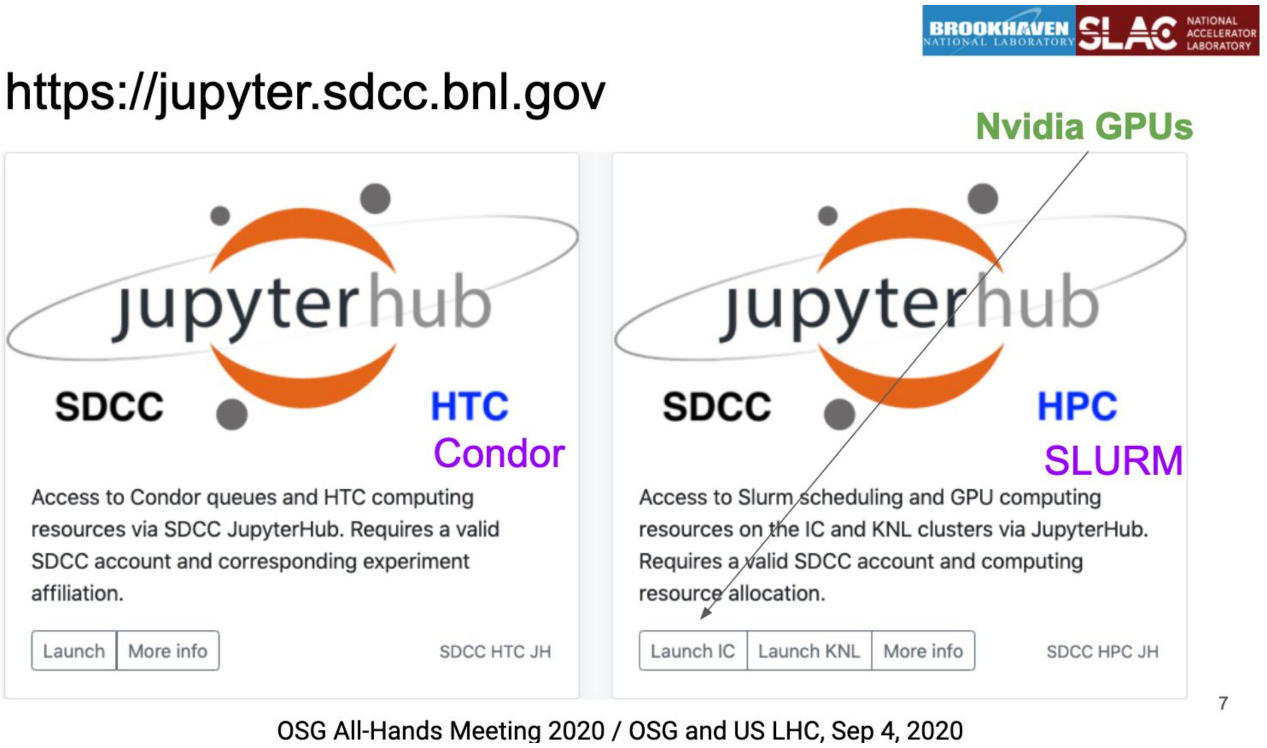
- 24x7 uptime, accessible worldwide, support ~300 active physicists, storage for all derived data, easy collaborative access, provide access to a wild world of apps and tools
  - Current level of derived storage ~100 PB - grow ~20% per year
  - Current analysis CPU+GPU needs ~50k cores - grow ~20% per year
  - Current users ~300 physicists - not expected to change a lot
  - Allow sharing of workspace with ~2000 international collaborators
  - Scalable, sustainable, interactive, (new) user friendly...
- Build at BNL T1 - roughly x2-x4 current facility cost
- At a dedicated large facility - see Paolo’s talk next
- On a cloud - see Johannes’ & Harinder’s talks this afternoon
- Hybrid facility - rest of this talk
- Choice is often driven by the source(s) of funding

Hybrid AF

- Current US ATLAS AF model
  - If gold plated AF not affordable, assemble many nuggets
- US ATLAS is supporting the following - on best effort basis
  - Enable the use of small local systems - limited support
  - Support most common workflows at T1/T2 facilities
    - Scalable and affordable for all users - not limited to US sites
  - Fund and support shared T3 facilities for end-stage analysis
    - At BNL, SLAC and UChicago (new - operational 2021)
    - Provide some support for new users (documentation, tutorials)
  - Support widely used tools that are scalable and sustainable
  - US ATLAS Ops funding is limited - x5 smaller than external sources of funding for analysis tools, services and facilities

BNL Shared Tier 3

- See talks by William Strecker-Kellogg (BNL) at OSG All Hands
- Easy-to-use interactive nodes, batch systems, local storage…
  - ~2200 CPUs, access to GPUs, scratch space, local storage…
  - Traditional batch systems and Jupyter



https://jupyter.sdcc.bnl.gov

Nvidia GPUs

BROOKHAVEN NATIONAL LABORATORY
SLAC NATIONAL ACCELERATOR LABORATORY
jupyterhub
SDCC
HTC
Condor

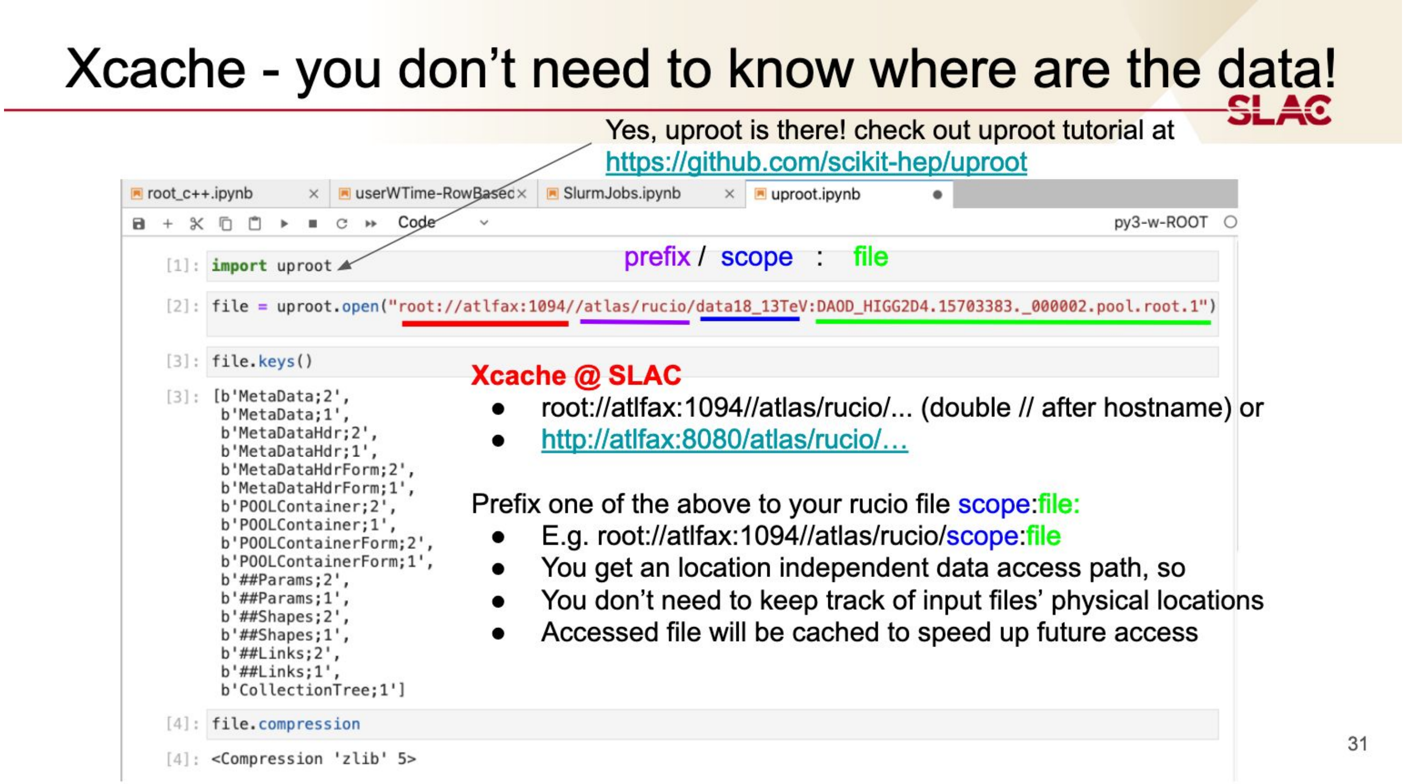
Access to Condor queues and HTC computing resources via SDCC JupyterHub. Requires a valid SDCC account and corresponding experiment affiliation.

Access to Slurm scheduling and GPU computing resources on the IC and KNL clusters via JupyterHub. Requires a valid SDCC account and computing resource allocation.

OSG All-Hands Meeting 2020 / OSG and US LHC, Sep 4, 2020

SLAC Shared Tier 3

- See talks by Wei Yang (SLAC) at OSG All Hands/WLCG
  - ~3400 CPUs, some GPUs, few PB storage
  - Interactive logins, batch jobs, Jupyter, xcache, containers...



Xcache - you don’t need to know where are the data!

Yes, uproot is there! check out uproot tutorial at https://github.com/scikit-hep/uproot

prefix / scope : file

```
[1]: import uproot

[2]: file = uproot.open("root://atlfax:1094//atlas/rucio/data18_13TeV:DAOD_HIGG2D4.15703383._000002.pool.root.1")

[3]: file.keys()

[3]: [b'MetaData;2',
      b'MetaData;1',
      b'MetaDataHdr;2',
      b'MetaDataHdr;1',
      b'MetaDataHdrForm;2',
      b'MetaDataHdrForm;1',
      b'POOLContainer;2',
      b'POOLContainer;1',
      b'POOLContainerForm;2',
      b'POOLContainerForm;1',
      b'##Params;2',
      b'##Params;1',
      b'##Shapes;2',
      b'##Shapes;1',
      b'##Links;2',
      b'##Links;1',
      b'CollectionTree;1']

[4]: file.compression

[4]: <Compression 'zlib' 5>
```

Xcache @ SLAC

- root://atlfax:1094//atlas/rucio/... (double // after hostname) or
- http://atlfax:8080/atlas/rucio/...

Prefix one of the above to your rucio file scope:file:

- E.g. root://atlfax:1094//atlas/rucio/scope:file
- You get an location independent data access path, so
- You don’t need to keep track of input files’ physical locations
- Accessed file will be cached to speed up future access

31

Towards a Gold Plated AF

- Can we do better than Hybrid/Tier 3 model
  - Yes - user surveys show some shortcomings of current model
  - A large well staffed high performance data center would be better
  - But need additional funding - current budgets will pay <10% of cost
- Build AF at BNL (colocated with US ATLAS Tier 1)
  - Gold plated == full featured, full scale
  - Interactive, easy to use, well connected, all derived data local…
  - Need to be x3-x4 current T1 capacity, large support staff
- Build AF at large ASCR or CISE funded facility
  - Similar concept to BNL AF
- Provision AF on commercial cloud
  - AWS, Google, MS, Oracle - many capable vendors
  - Limited only by funding
  - Simple list price based cost model shows x5-x10 cost of BNL T1

Summary and Speculations

- Many things will change during Run 3
- And again at the HL-LHC
- Analysis Systems being developed now will drive this change
    - Looking forward to many of the tools described in talks yesterday moving to production quality services
    - US ATLAS is/will partner in most of these tools
        - We also continue to support some limited ATLAS specific tools
    - Important metric - benefit and scalability on our infrastructure
        - Ease of installation and use in our Shared T3s
    - Additional metric - support cost
        - Sustainability is critical to long term success
- US ATLAS will continue to support the development and deployment of software and computing systems that our users need to accelerate their physics products