ATLAS Analysis Model &
the HL-LHC Conceptual Design Report

March 17 2020

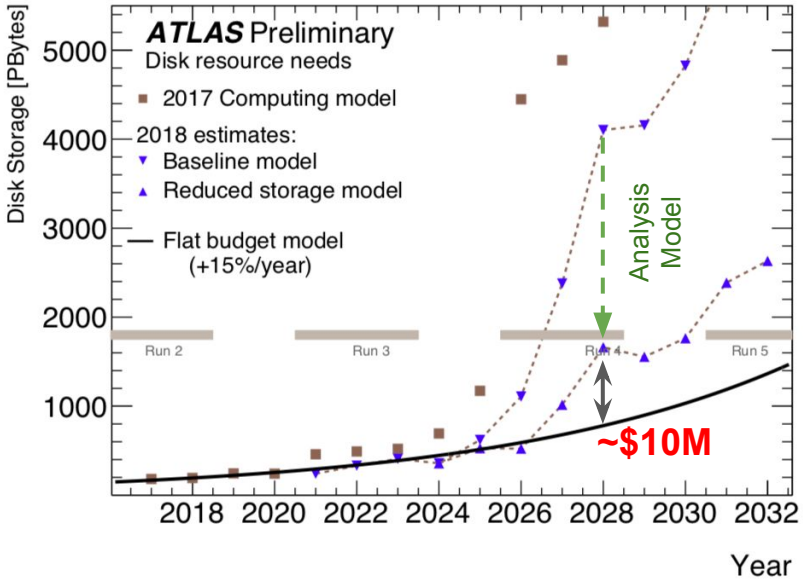
Paolo Calafiura (LBNL)
With lots of input from Davide Costanzo (Sheffield) and Wei Yang (SLAC)

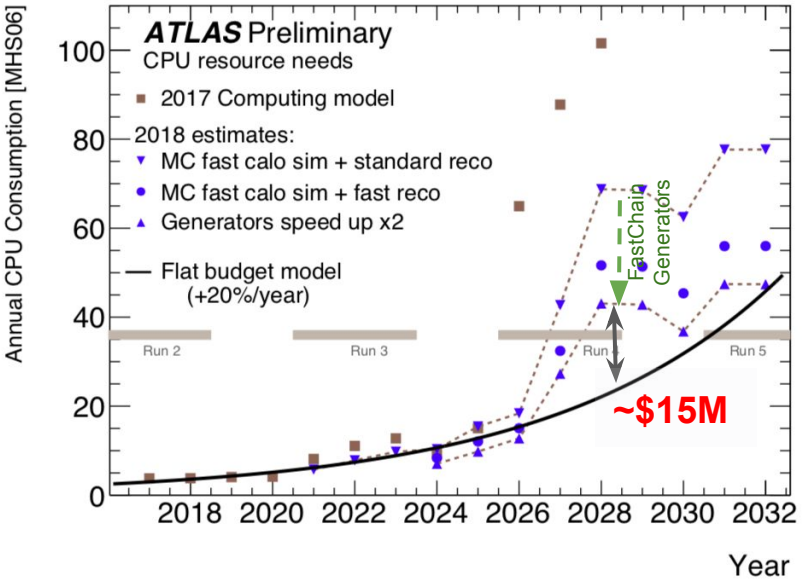
Overview

1. ATLAS HL-LHC Computing Model and the Conceptual Design Report
2. US ATLAS Analysis Facilities (a.k.a. Shared T3s)
3. Personal musings on the evolution of Analysis Facilities

The *Infamous* Resource Plots

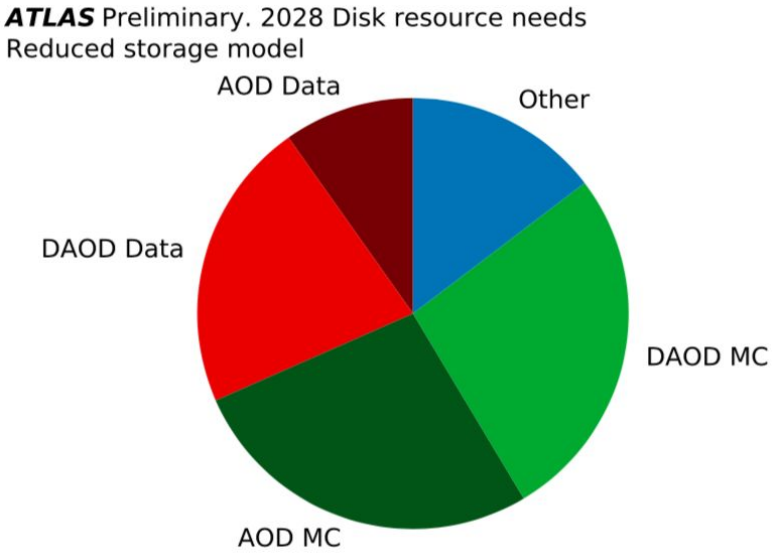
- LHC Phase-2 computing was advertised as a difficult problem to solve
  - The CPU/disk needed is above what we expect (link to plots)
  - Current model presented on 4 Dec 2018 at ATLAS weekly (link)
- Update of the model (and plots) will come this Spring.
  - **Several ideas for the CPU problem, fewer ideas for the disk problem (harder!)**
  - **Tradeoffs: memory-cpu, disk-network, cpu-disk.**

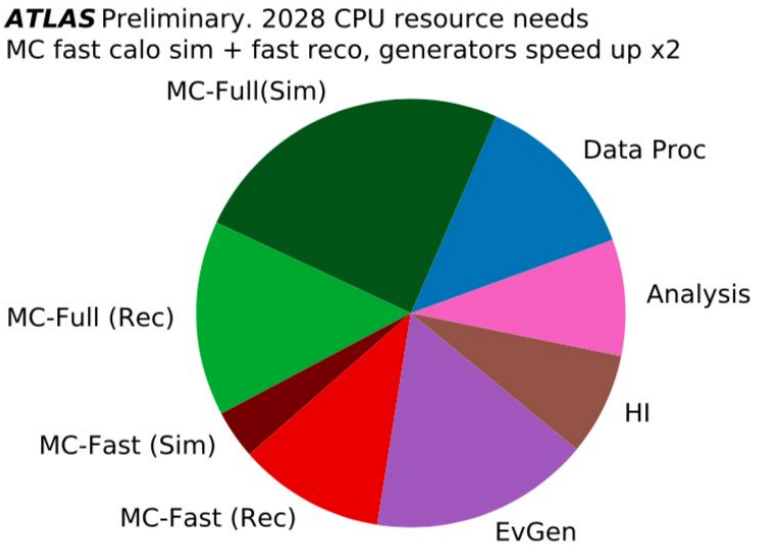




Analysis Matters

- Almost 90% of projected Run 4 disk usage
  - Over 50% of disk dedicated to MC
- Non-negligible CPU impact

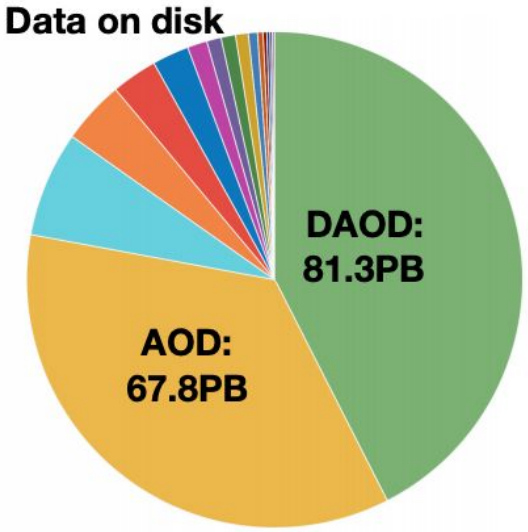




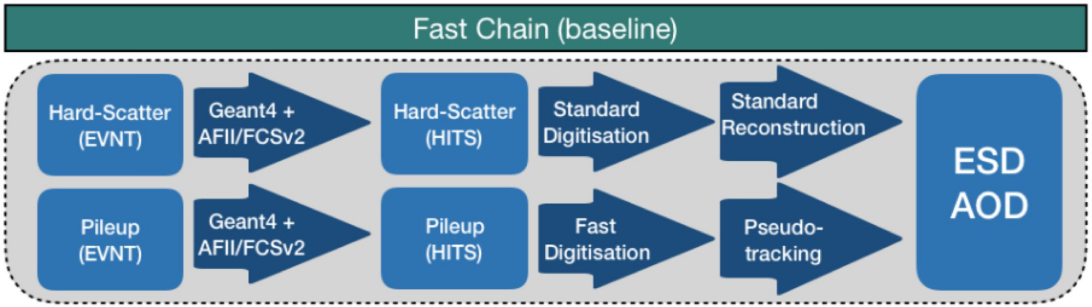
Analysis Model Evolution

- Most of the disk space is used by analysis formats (AOD and DAOD)
  - Analysis model has a direct impact on disk resources
  - Run-2 analysis model was too “expensive”
  - Analysis Model Study Group for Run-3 (aka AMSG-R3) provided recommendations for Run-3 (CHEP 2019)
  - Most analysis to move to a new DAOD_Phys format

- Further evolution for Run-4
  - Use a smaller pre-calibrated format - DAOD_PHYSLITE
  - Large use of FastChain. Are AODs needed as persistent MC format f



Run 2



Idea: Data carousel. Store AOD on tape

‘data carousel’

- Use tape as storage medium for AODs
- Recall AODs from tape in a rolling buffer
  - Retrieve only data used by workflow
- Process data (or MC) from disk once available there
- Reduce number of AOD copies needed on disk

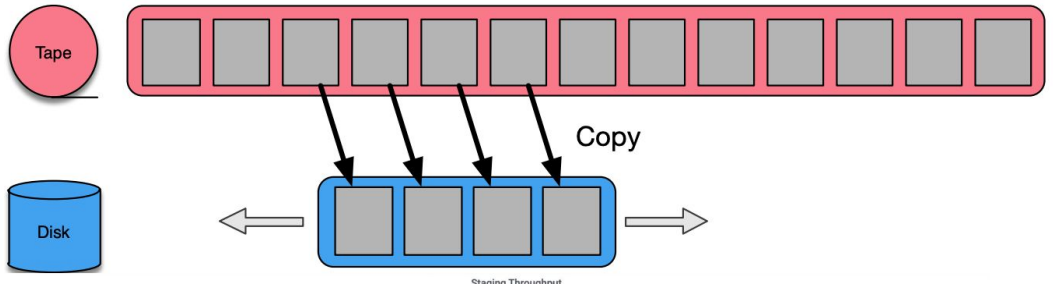
Tested tape performance in production For AOD Derivations

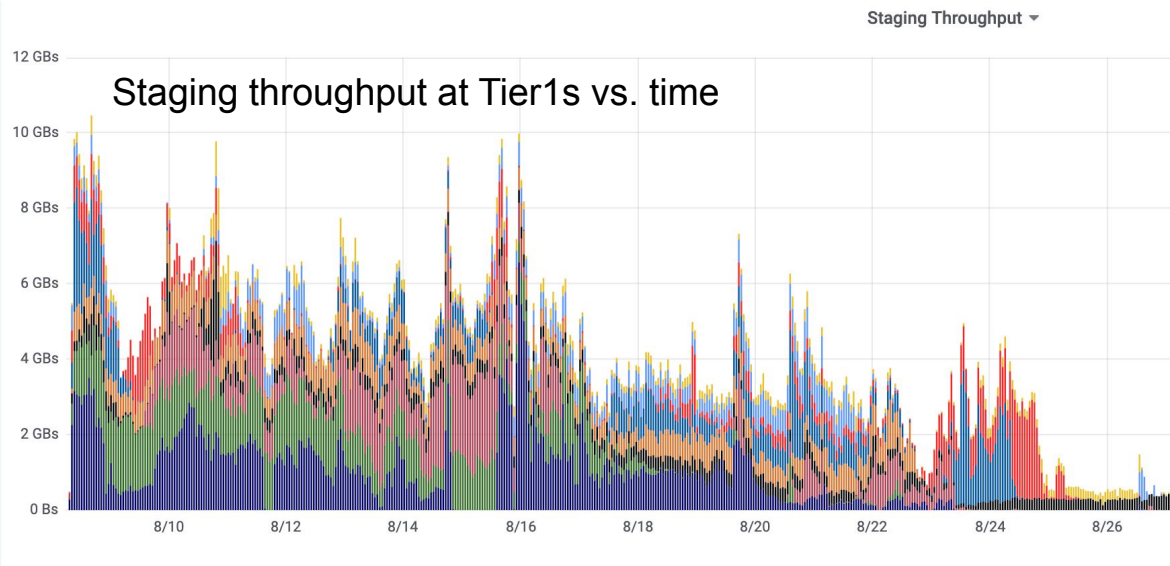
- Encouraging results: it worked

AOD access from disk is 100 PB/month

- For tape access we have to schedule access orderly
- Ongoing R&D and technical work to automate this on the data and job management sides

Data Carousel “commended” at last CRSG





Next Steps

2020

Release 22: Run 3 feature Freeze

LHCC Review of HL-LHC Computing

C&S HL-LHC **Conceptual** Design Report

~2023

C&S HL-LHC **Technical** Design Report

WLCG HL-LHC TDR

~2026

Run 4 Feature Freeze

US ATLAS Analysis Facilities

In Production:

- Two Analysis Facilities (a.k.a Shared Tier 3s) in operation @ SLAC and BNL for several years
- So far mostly use retired hardware from BNL T1 and former WT2 (SLAC Tier 2)
  - Direct FTE funding from US ATLAS
- The two T3s are part of US ATLAS Computing Facility and provide Physics Analysis Support
- No clear division between these two types of tasks and personnels.
  - All work together. Coordination meetings across WBSs

Pre-production Facilities:

- Cal State Fresno Virtual T3
  - An all-AWS Analysis Facility
- Univ. of Chicago ML(+Analytics) Platform
  - GPU resources for all ATLAS users https://www.atlas-ml.org/

Role of the Analysis Facilities

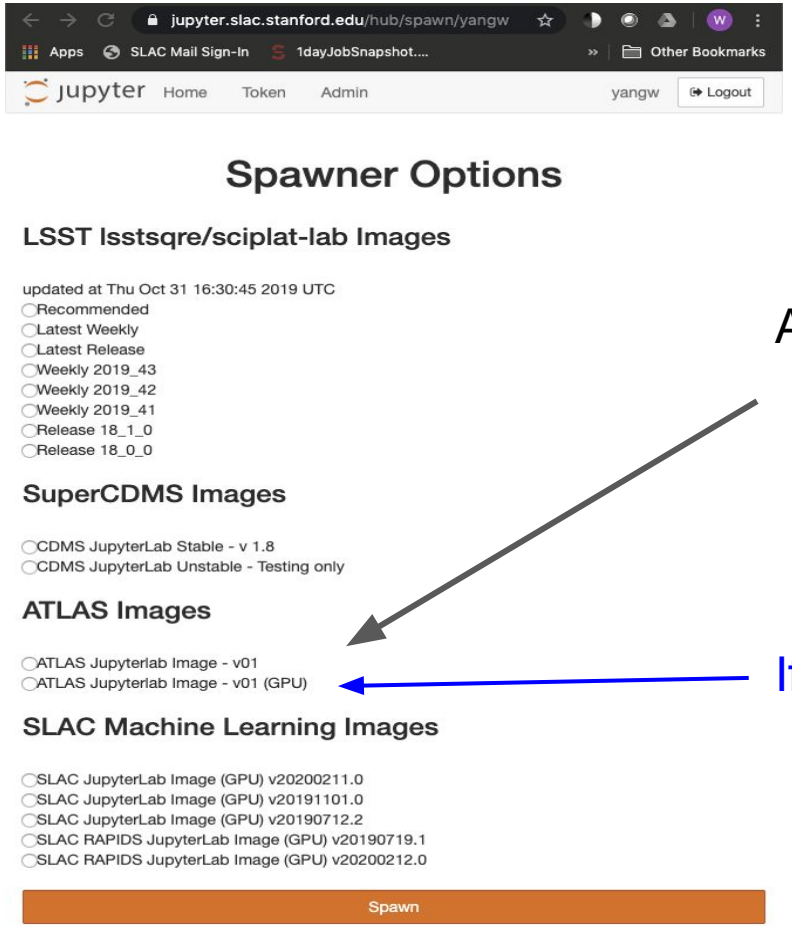
Directly serve US ATLAS physicists.

- Login, local batch, have home directory and data directory at AFs.
  - Similar to familiar lxplus environment
- **Users need support**
  - from how to use unix, to how to use batch, to how to run rucio and Athena

Complement grid based analysis capability.

- **Aim at easy**-to-use, easy-to-understand
- Optimized for **late stage data analysis**, batch and/or interactive
- **Quick turnaround** time
- Support both traditional (root,prooflite) and modern (scipy, jupyter) data analysis platform
- Support rucio and PanDA to scale out big tasks to grid

An Example: jupyter @ SLAC



Login via SLAC unix account

A list of ATLAS images are added at here

- Notebook with PyROOT (C++ or Python)
- uproot (pure python)
- root_numpy

If you want GPUs

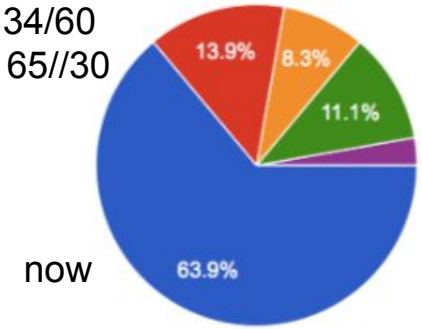
Start your Jupyter container

Recent US AFs survey and other data

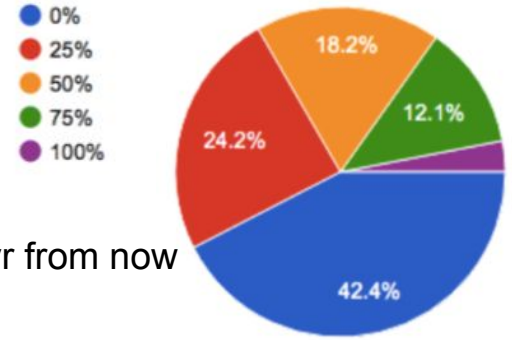
36 responses

users / active users
BNL: 134/60
SLAC: 65//30

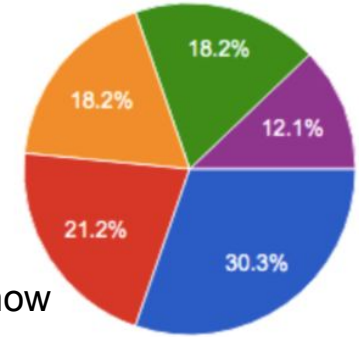
Survey Question: (excluding desktop and Grid usage)
What percentage of your group's computing do you expect to do at the US shared T3s now, 1yr, 3yr, 5yr from now



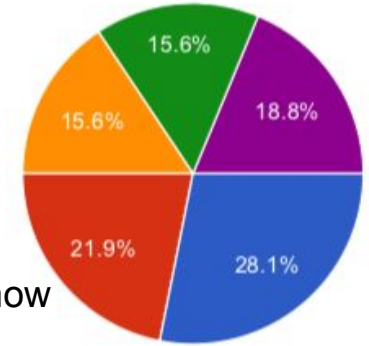
now



1yr from now



3yr from now



5yr from now

BNL and SLAC active user list show that almost no user actively use both AFs.

- Plenty of resources available
- Migration overhead?

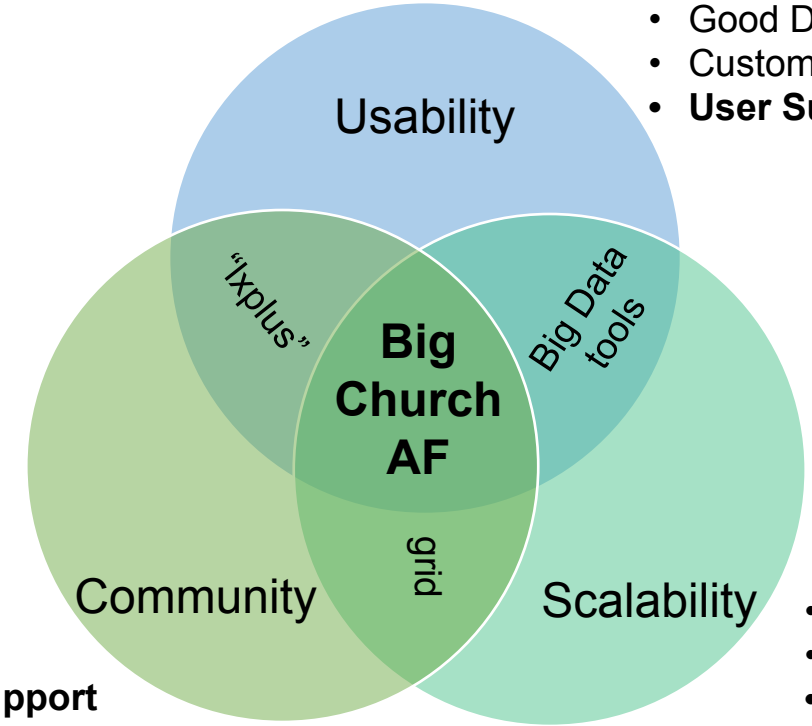
Geographic locations seems to play a role - most of the SLAC users are on the West Coast.

- Looking for personalized support?

Personal Musings in lieu of a proper conclusion

- Why aren't US ATLAS AFs as crowded as lxplus?
- Why isn't everyone using the (clearly superior) scientific python stack?
  - *My guess:* Users don't have time to learn how to be efficient: they **stick to defaults**
    - Migrating a community (e.g. an analysis group) is even harder: **shared knowledge**
- If we want AFs to succeed we need to overwhelm users with support:
  - User-level support:
    - From beginner defaults, to cluster configurability
      - We are doing this, need more effort, particularly in entry-level tutorials
  - Embedded support:
    - Analysis groups are high-value customers, embed one "**account executive**"
      - Ideally a **50/50 physics/tools** person with knowledge & contacts on both sides
        - These people are precious but do exist, we need a **career path** for them
  - DevOps support:
    - Cutting edge analysis tools are hard to deploy at scale
      - Fledgling **DevOps community**
        - Highly marketable skills, need to create a **pipeline** and a rewarding career path for those who chose to remain
- Challenge: can we provide this level of support across a Distributed Facility?

The Big Church Analysis Facility



- Good Defaults
- Customization
- **User Support**

- Reusability
- Reproducibility
- **Embedded Support**

- Data Parallelism
- Task Parallelism
- **DevOps Support**

Backup

Physics Community Feedback

We asked many questions to the ATLAS Upgrade Physics community:

(My) highlights from six groups replies:

- Precision measurements will likely require **full sim** (egamma)
- Systematics ofter **limited by MC accuracy** not statistics (egamma, taus)
- HH(bbtt) currently **limited by MC statistics**
- **ML** classifiers expensive to train. **Full sim** needed (taus, HDBS)
- **MC needs** expected to **scale linearly** with lumi. (Exotics)
- **NNLO and NNNLO** need will increase (Exotics)
- **TLA** for photons and b-jets (Exotics)
- **DAOD_PHYS_LITE** good enough for tau **systematics**