# US ATLAS and IRIS-HEP

Paolo Calafiura for US ATLAS Ops 11/24/2020

- Update on ATLAS HL-LHC activities
- IRIS-HEP blueprint activities
- Analysis Services and IRIS-HEP data challenges
- Analysis Facilities, SSL and OSG-LHC
- IRIS-HEP contributions to DOMA and Reconstruction
- IRIS-HEP and HEP-CCE

# New: HL-LHC Computing Conceptual Design Report

- Define R&D needs for HL-LHC, and high-level planning
  - Model two resource growth scenarios (+10%/yr, +20%/yr)
  - Model impact of 13 Conservative R&D and 15 Aggressive R&D goals
    - Show possible paths to meeting HL-LHC requirements without impacting physics reach
      - Access to derived analysis data looks like main concern
- CERN-LHCC-2020-015, presented to LHCC Review of HL-LHC Computing
  - Concern about **lack of qualified effort**

# Updated: (US) ATLAS HL-LHC Planning

- US ATLAS Ops planning includes 22 HL-LHC milestones to FY24 (HL-LHC TDR)
  - Prioritized based on US ATLAS commitments, expertise and physics impact. Reviewed internally and by BNL directorate *(PEMP notable)*
    - major influence on ATLAS computing CDR & vice versa
- ATLAS Road to Run 4 (R2R4). Starting from CDR, identify
  - R&D milestones to 2024
  - **Highlight key personnel, areas with effort shortage**
    - First iteration with Core, Simulation

# Key IRIS-HEP Contributions to US ATLAS R&D

DOMA

Data and Software Preservation

Machine Learning

Software Trigger & Event Reconstruction

Data Analysis & Interpretation

Software Development, Deployment, Validation, Verification

Training

# Key IRIS-HEP Contributions to US ATLAS R&D

DOMA

Machine Learning

Software Trigger & Event Reconstruction

Data Interoperability

Data and

- **Sustainable Software**
- **Future Analysis Systems and Facilities**
- **Fast ML**
- **…**

Software Development, Deployment, Validation, Verification

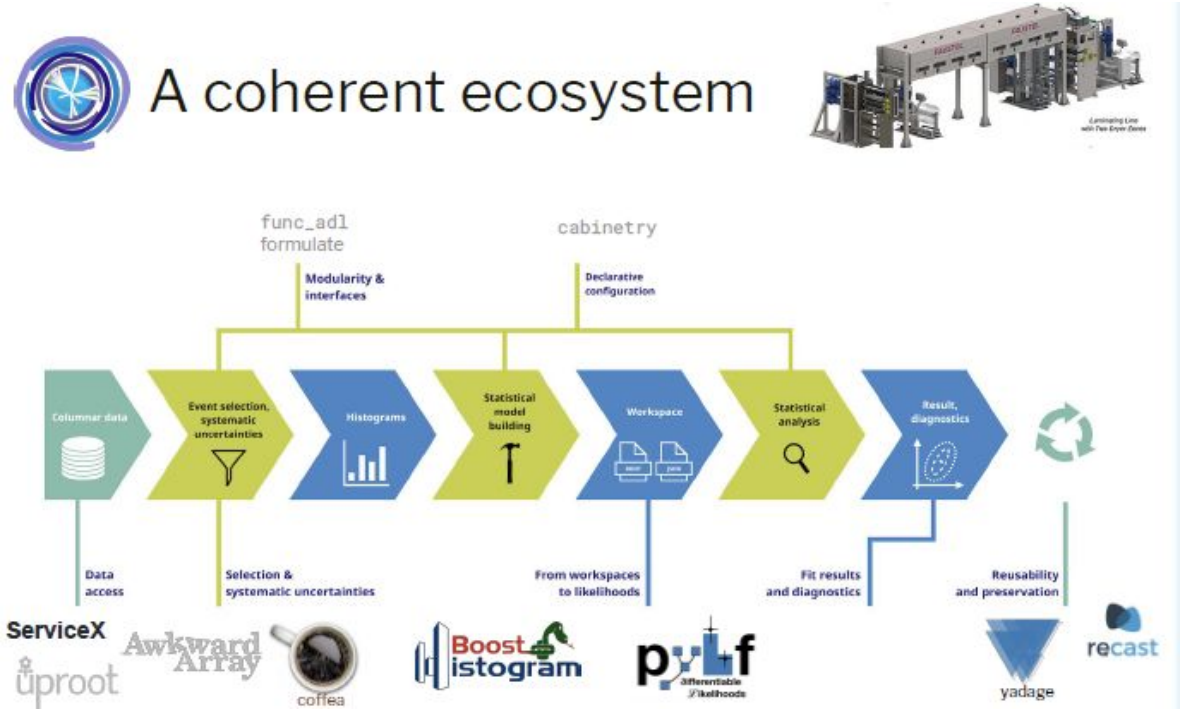
Training

# Impact of Future Analysis Systems Blueprint

Timely and well-organized workshop

- Nine contributions from ATLAS folks
- Multiple, lively follow-up discussions within US ATLAS
- US ATLAS/US CMS follow-up on AF for HL-LHC (organized by Brian)
  - Proposed joint AF R&D Projects
    i. Jupyter, Dask, Parsl **AS**
    ii. Federated Authorization **SSL** and **OSG**
    iii. Deployment - apps, infrastructure
- Other US ATLAS R&D priorities
  - Spillover of AF workflows to “cloud” resources via distributed scheduler **DOMA**
  - **Storage systems** to deliver 200TB in 25’ (IRIS-HEP GC), scaling to PBs by Run 5
  - Analysis re-interpretation **AS**

# Analysis Systems Challenge




**Not trivial** to build a **coherent** ecosystem out of 10 tools

- Engineering challenges:
  - Integrated testing/release
- Organizational challenges:
  - Long-term M&O
  - Shared goals
- Adoption challenges:
  - Demonstrate measurable advantages
  - Provide evolution path (from ROOT)
  - Support 1000s users

IRIS-HEP **200TB Grand Challenge as catalyzer** to address them

Shawn McKee

# Data, Analysis and Network Challenges

We need to prepare our infrastructure to deliver on our requirements for the HL-LHC era. This is going to require a series of steps (challenges) to move us toward the required scale

- Principles for architecting network, data and analysis challenges:
  - *We can't afford to do network (data) challenges at scale that don't also contribute to required operations.*
  - Needs to involve both production and prototyping networks (FABRIC/FAB, trans-oceanic research and education links, ESnet, GEANT, Internet2, etc)
  - Include big sites (as well as HPCs?) to understand operational modes and network access
  - Critical to define **production** milestones associated with each challenge.
  - Should be incremental, building every few years: **2021, 2023, 2025, 2027**
- We need to test in the context of what we are using or plan to use for analysis and production

  One goal: scale to ~Tbps networks for HL-LHC

# DOMA

Shared vision: Data Access through caches and intelligent delivery

- Comment: co-development of iDDS, ServiceX, and AF data infrastructure (Ceph, …) key to ultimate GC success
  - Deliver to every AF user 1 Tbit/s of filtered and reformatted data
  - Enable new workflows (i.e. HPO, ML training…)
  - Solve the HL-LHC storage challenge (i.e. Data Carousel)
  - Streaming as confluence of network, storage, and CPU
- iDDS good model of experimental priorities driving development

# iDDS in one slide

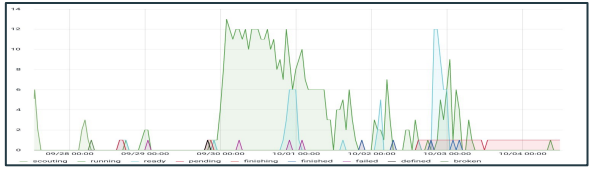
Wen, Tadashi, LHCC Nov 17

## Achievements

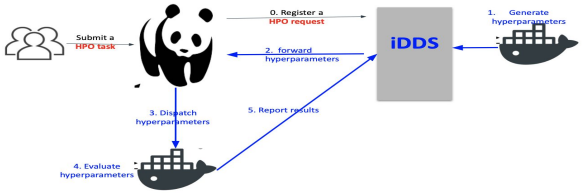
- Data Carousel in production since May
- HPO (Hyper Parameter Optimization)
  - Platform for HPO on GPU on the grid, HPC, and clouds
  - Advertised to ATLAS ML users, not specific to ATLAS
- DAG based workflow management
  - DOMA PanDA instance for Rubin Observatory (LSST) demo

## Near-term plans

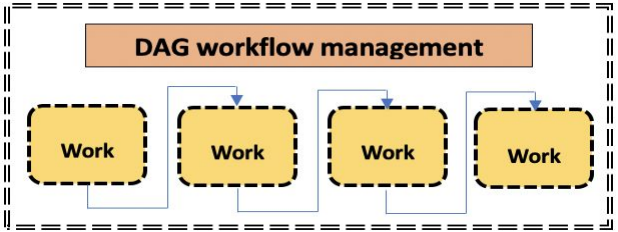
- Improvements of user experience
- More use cases in multiple experiments
  - DAG based Active Learning is ongoing
  - Dynamic transformation and placement on demand, for example Derivation on Demand
  - Fine-grained data transformation and delivery, for example Event Streaming Service



iDDS tasks accounting (by status)




iDDS HPO




iDDS DAG

# Innovative Algorithms

- Good balance of short, medium, and long-term projects
  - From mkFit deployment in CMS HLT, to ACTS serial and parallel implementation of critical algorithms (seeding, filtering, ambiguity resolution), to ML models for track finding and jet reconstruction
- Happy to see efforts to collaborate within and beyond IA
  - For example in the cross-cutting problem of parallel track finding
- For Y3+
  - Appreciate focus on delivering quantifiable physics benefits to experiments
    - Encourage IA to also explore “far out” algorithms perhaps through the fellowship program
      - BTW, thanks to Gordon, Heather, et al US ATLAS is increasing its participation to the fellowship program

# OSG-LHC and SSL

Important in developing a roadmap from Run 3 to Run 4 that provides sustainable and scalable solutions for distributed computing.

As mentioned before can play key role in addressing **security** issues for analysis facilities (and not only).

Development and distribution of **containerized services** focusing on what is needed for the grand challenges (AS, DOMA, network monitoring, …).

Encourage OSG-LHC to help experiments add **opportunistic** resources to their WLCG **pledges** in order to reduce the HL-LHC “gap”.

Important to incorporate experiment goals/milestones into IRIS-HEP planning.

# IRIS-HEP and HEP-CCE

Same goals: algorithm and I/O optimization

IRIS-HEP focused on new methods, HEP-CCE on new platforms

HEP-CCE **PPS** can help IA develop and optimize **portable** parallel algorithms

HEP-CCE **IOS** can help DOMA and AS meet their ambitious Tbit/s end-to-end requirements through their of **HDF5** and object stores (**HEPnOS**) expertise

# Conclusions

IRIS-HEP is an example for the HEP community on how to organize a multi-experiment research project.

Playing an important role in community building, training.

Analysis Grand Challenge key to coalesce efforts in different areas, and to demonstrate the tools IRIS-HEP will deliver to the community in Years 3-5.

US ATLAS has incorporated a number of IRIS-HEP milestones into HL-LHC planning - need to track them together, and iterate to add more.