

# Discussion

# User Experience At AF

G. Watts (UW/Seattle)

# Three Discussion Sessions Now...



**User Experience at Analysis Facilities**

Data Access at Analysis Facilities

User Monitoring at Analysis Facilities



# The Physicist

What does the physicist want?

What is there experience at an AF?

# What Does UX Mean? Some Examples...

User arrives at your portal and can't find Jupyter Notebook access they want

Bad Experience!

User tries to access a GRID dataset but can't download it due to poor local networking

Bad Experience! But perhaps discuss in the Data Access session next!

User tries to run interactive job on 500 TB of data, but it will take hours to complete a histogram

Bad Experience! But we also should keep in mind “not designed for this” limits!

This should run on k8s – it is the only solution!

Not about user Experience! True or not – user won't care as long as they get what they need

“People don’t know what they want until you show it to them” - Jobs

---





# Questions...

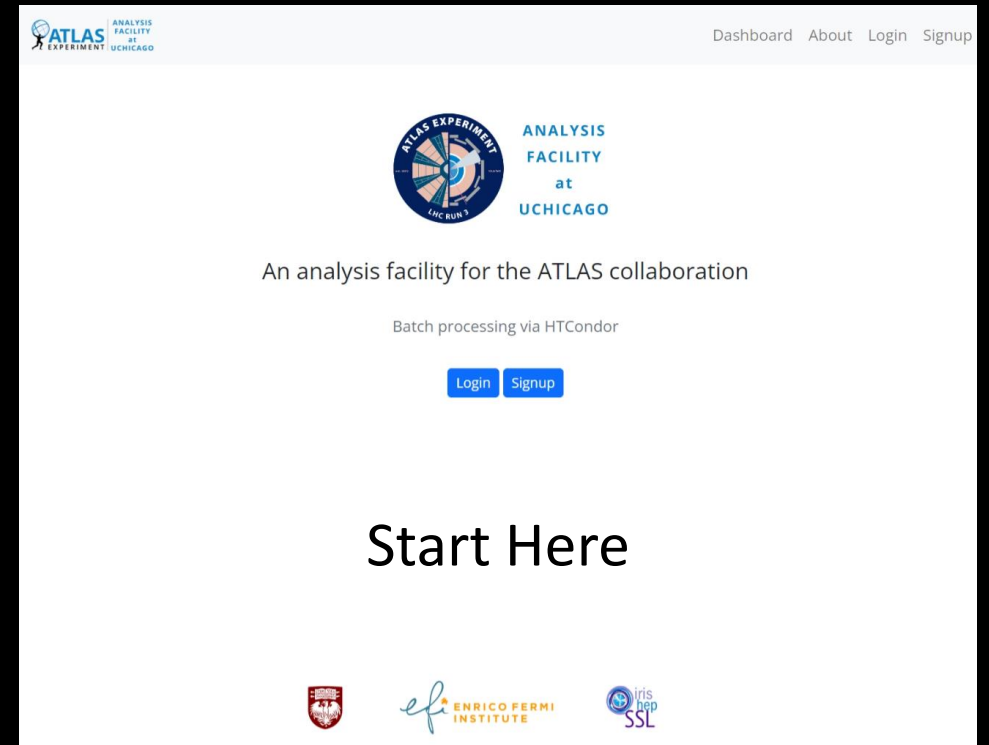
- What is impossible?
  - When should a job be done on the GRID rather than an AF?
- Can we (the experiments) collaborate on building these?
- What work gives us the most users?
- Are there technical decision we make that the user will care about?
  - But are unaware of?



I am going to go very fast through this talk!

# Access

- Landing Page
  - If you have multiple services, how does someone new know where to go?
- Minimize Security gateways
  - Don't make people authenticate twice
  - Auto-generate certificates/tokens?
- Unified Services
  - Ssh and Jupyter notebook going to the "same place"
- Access modality
  - Web
  - Ssh
  - Keeping it "local": ssh-mount, vscode

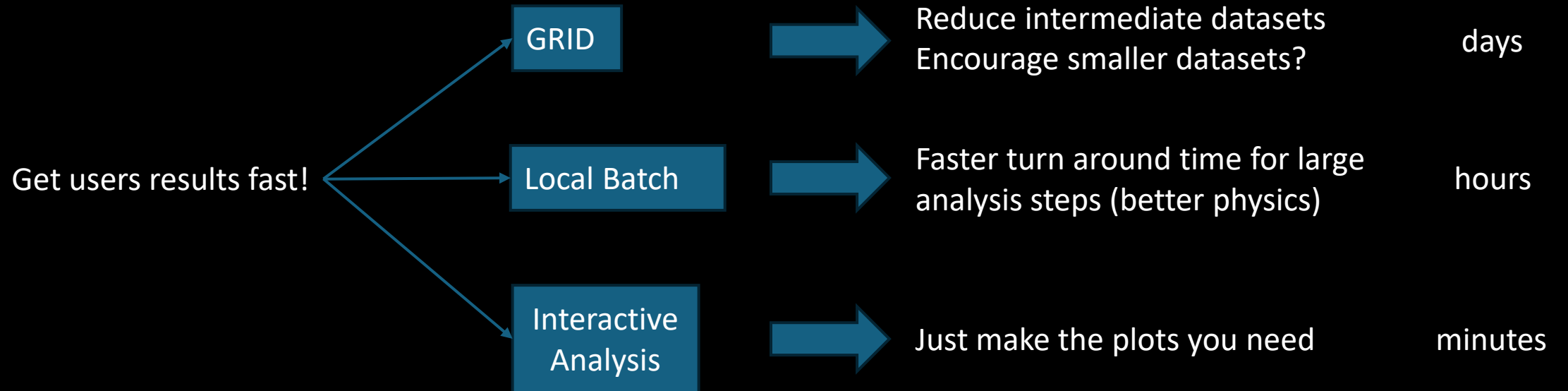


Unify across an experiment?



# Scaling

Q: “What would make the GRID useful?” – if I could shrink my 15 day effort to 1 day (-internal meeting)



How much scaling?

- CMS and ATLAS both are talking about analysis dataset sizes between 200 GB and O(100TB).
- Need something more careful from the experiments for benchmarking (LHCC questions?)
- Scaling should “just work” (minimal tuning by user?)



# Tools & Environments

Installed Tools: notebooks, ssh, etc.?

What should the user bring with them?

Method to share container environments in an analysis group  
(e.g. docker, [dev-containers](#), binderhub)



Scaling Impact?

Current Analysis Facilities tend to specialize – is this the right way forward?

- Services provided – ServiceX, REANA, etc.
- Locally installed tools like snakemake...
- Should coffea be there?

Machine Learning



Toolset and workflows – plethora of workflows

- Access to GPU's (efficient!)
- Workflow support: don't make a choice?

# Analysis Workflows Supported

- Traditional batch style analysis
- Columnar Analysis (RDF & python)
- Machine Learning

Scaling!!

## Sizes

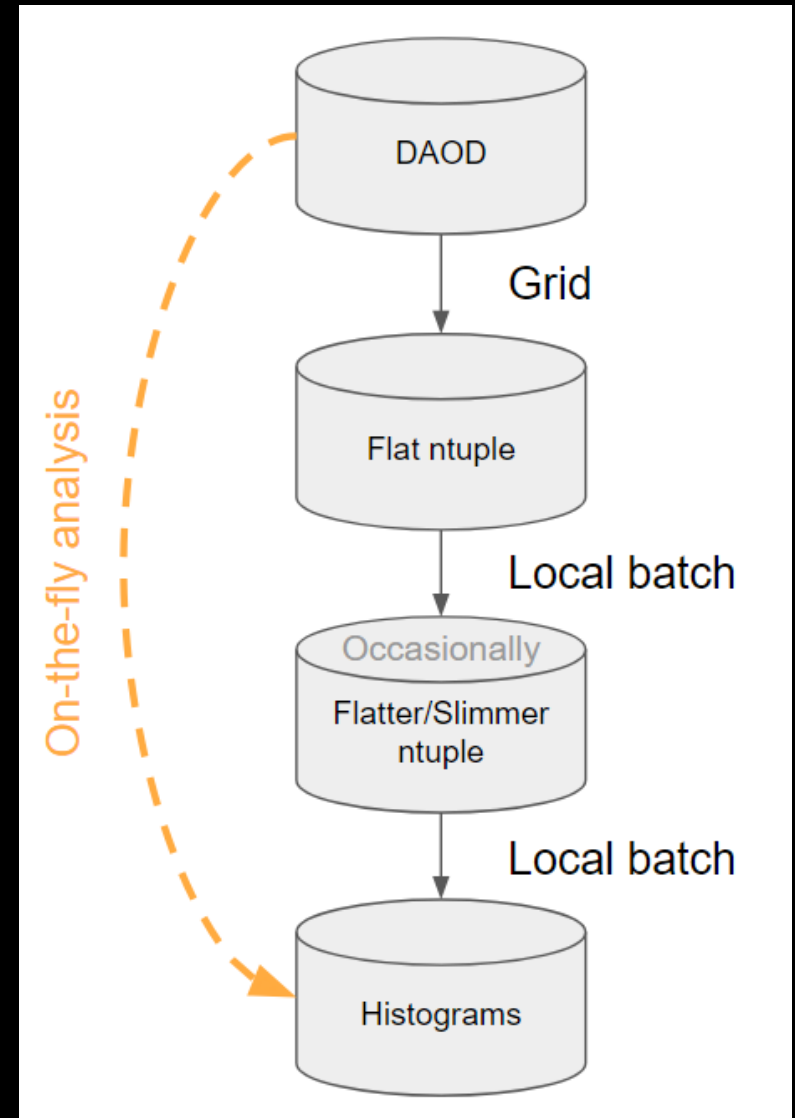
- Likely datasets between O(100 GB) to O(100 TB)
- High speed (PHSLITE/nanoAOD)
- Low level (detector level ML inputs)

## Data Formats

- Experiment Issues
- Augmented Data

## Tools/Services

- Snakemake, REANA, ServiceX, etc.



(old skool)

# Help!

- Documentation (s.g. [nersc](#))
  - Facility tools and services
  - Demos/tutorials?
  - Answer questions at 2am local time...
- Support (dedicated people)
  - Have to span physics and facility
  - Support forums
- User Monitoring
  - Why is my job taking so long to start?
- UX Feedback
  - You can't allocate a GPU right now because...





# Flexibility

“Hey – I liked Ben’s talk.  
How can I try that out?”

Analysis is the wild west of tools, computing, etc.

