

# SSL Year 3 Plans

IRIS-HEP Retreat  
May 29, 2020



# Three major themes

- Facility R&D
  - *SSL to inform Tier2 evolution & analysis facilities*
  - *Multi-site, hyperconverged infrastructure*
  - *Supporting flexible, reproducible deployments*
- Supporting IRIS-HEP Grand Challenges
  - *Analysis Challenge*
- Accelerated Data Delivery R&D
  - *Explore hardware acceleration at different points in the infrastructure*



# Facility R&D

Multi-site SSL with minimum of federation to expand access and institutional contributions

- User credentials with CILogon
- Federation
- Storage

Volunteer candidate sites - Tier2 sites?

- mwt2, aglt2, unl, uw, ucsd, caltech were mentioned

## **Milestone**

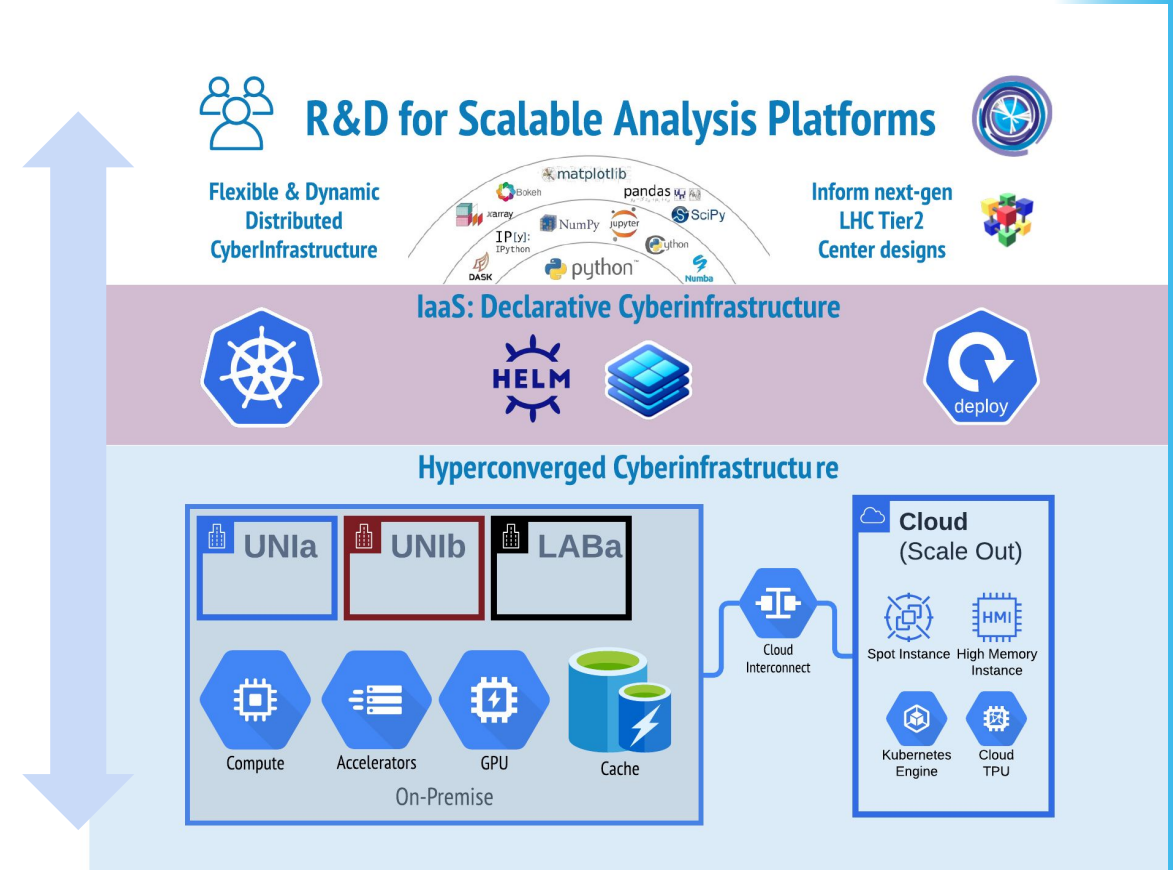
- Four SSL sites providing a k8s resources
  - *A subset providing multi-cluster scheduling & user environments (e.g. JupyterLab) to support analysis challenge*



# Supporting the Grand Analysis Challenge

Outline of goals [here](#) & [here](#)

- Fits with vision of declarative CI
- **Milestone:**
  - *Prepare K8s clusters and other prerequisites including monitoring & analytics*
  - *Scale: 200 TB, 1500 core*
  - *With AS & DOMA, execute challenge*



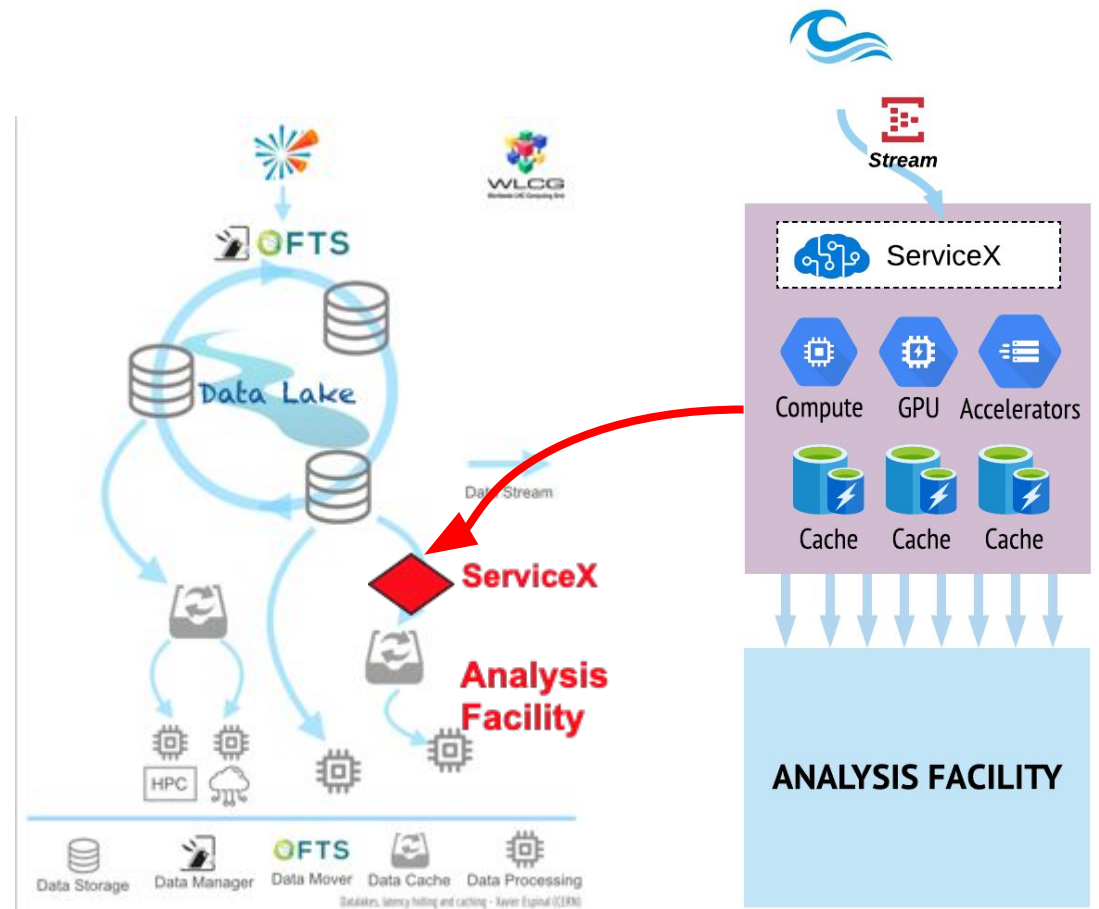


# Accelerated Data Delivery

High level goal is to explore (hardware) accelerating data delivery from ROOT format to columnar formats

Two milestones

- **Opportunity assessment** which profiles baseline transformer performance, evaluates technology options (Dec 2020)
- If cost/benefit relative to ServiceX on [cluster baseline](#) indicates, build **prototype system and benchmark its performance** (July 2021)





# Year 3 Summary

- Four sites offering k8s resources configured with multi-cluster scheduler or equivalent
- Prepare resources, instrument with analytics, contribute facility services for analysis grand challenge
- Assessment study on hardware accelerated data delivery, and if cost/benefit indicates:
  - *Build prototype hardware accelerated ServiceX system and benchmark*

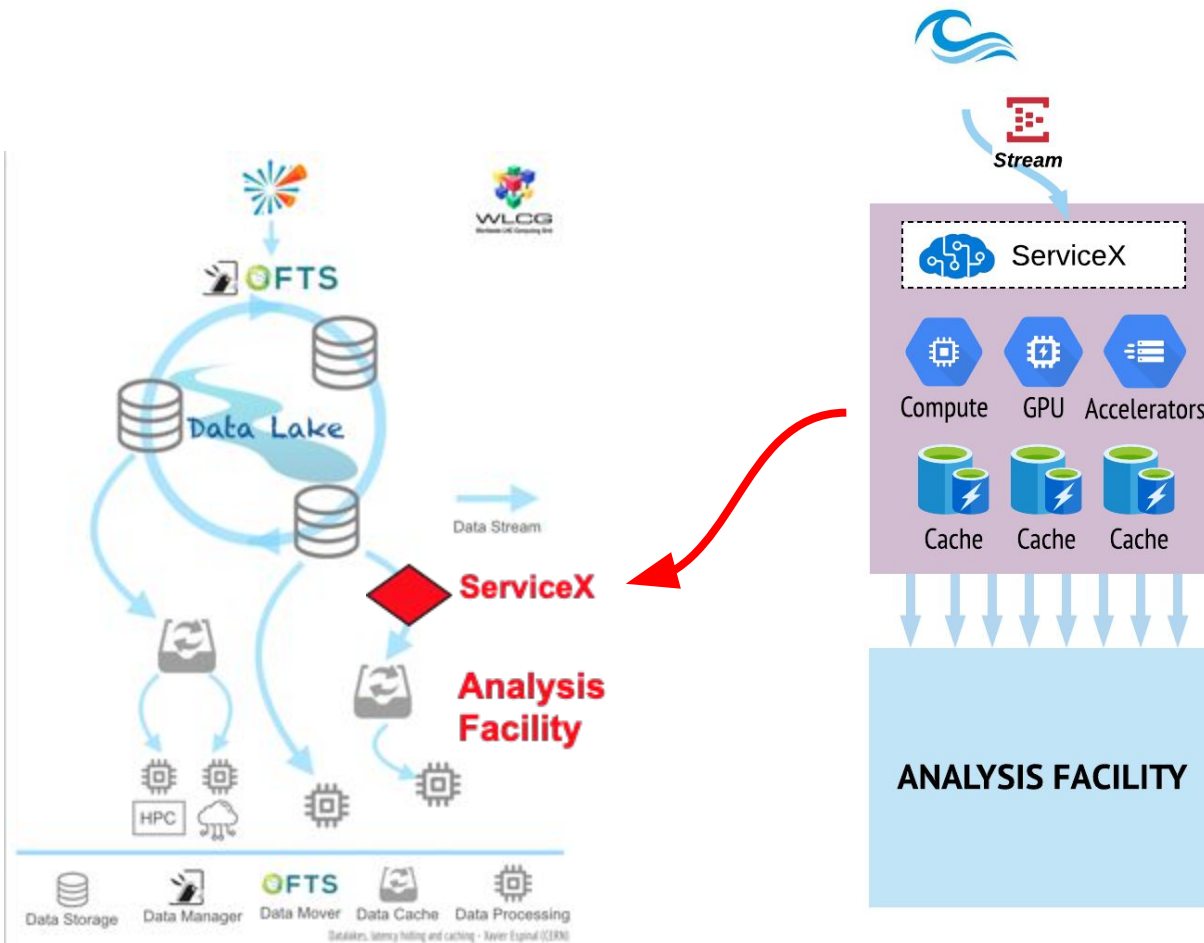
Questions?



extra



# Accelerator Delivery Testbed



- Ingest at 50-100 Gbps
- Uncompress ROOT format
- Reformat to columnar format
- Apply user filters
  - *cuts & columns*
- Write out reformatted, filtered events
- Benchmark





# Accelerated Delivery - details

Demonstrate ingestion via ServiceX on a single host that achieves 50-100Gbps sustained ingest rate while doing the following:

- Uncompress root baskets and reformat into columnar data format to support much higher speed on reads later. Maybe add some indexing for fast retrieval if useful?
- Apply an event filter on keeping only events for which at least one object within a container of objects within a basket exceeds some user defined threshold.
  - *Assume filtering such that  $O(10)\%$  of events survive on output (5-50% should all work well enough).*
- Apply an object list filter
  - *User defined set of baskets are to be kept while all else is dropped.*
  - *This can be done even without uncompressing.*
- Apply a simple algorithm like 4-vector addition or alike, possibly also more complex thing like JEC (apply a scalefactor to all objects in a specific container according to their place in a 2D map of eta-pT, i.e. object characteristic)
- Combine all three of the above.



# Computational Storage with CSSDs

