

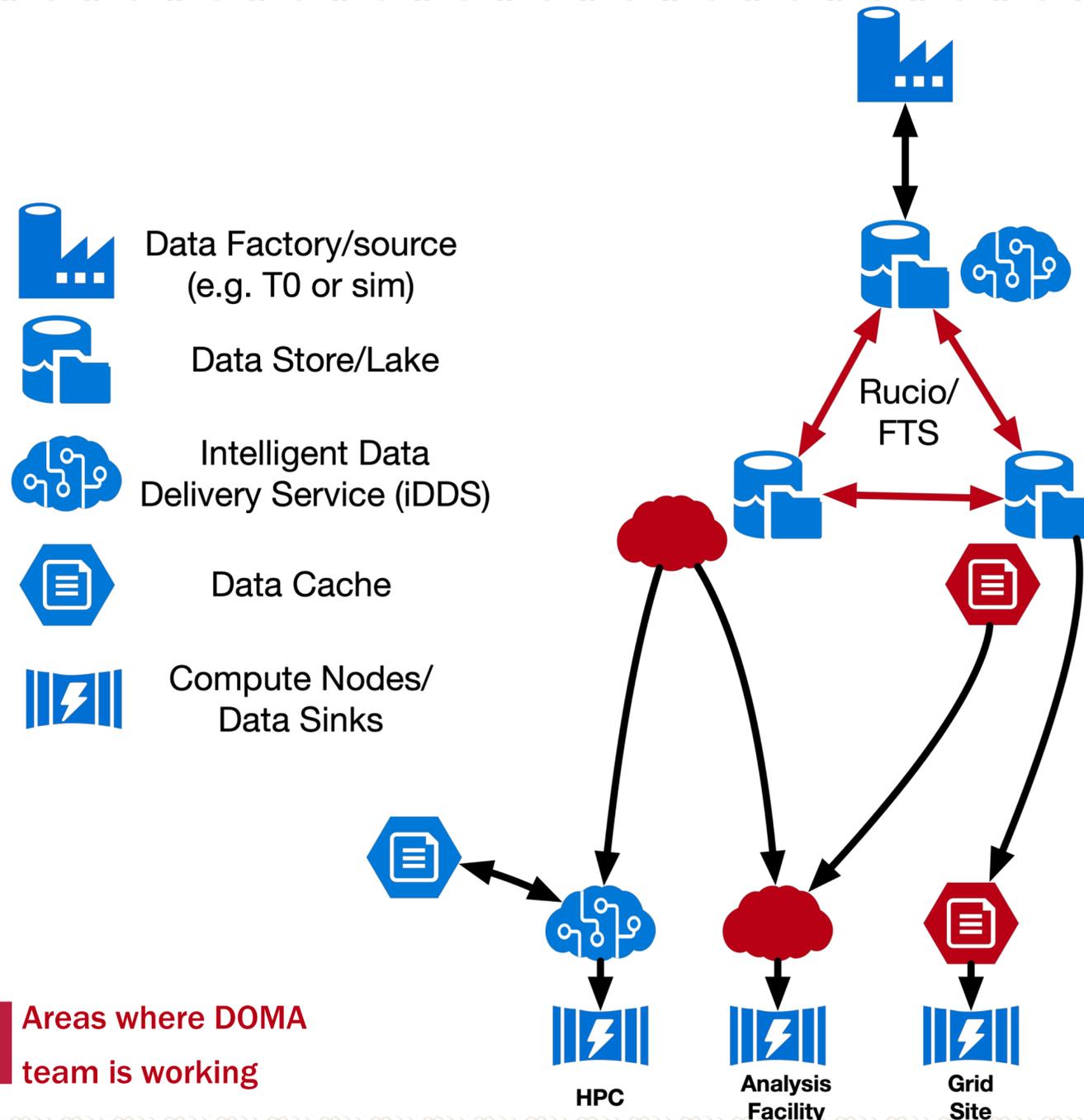
**DOMA –
News, Status, Plans
(Updates)**



MORGRIDGE
INSTITUTE FOR RESEARCH
CORE COMPUTATION

FEARLESS SCIENCE

Going Back to Yesterday



The big picture here didn't change; big-picture milestones and deliverables are unchanged. Spent our breakout sessions talking about proposed data challenges.

- We didn't have the right folks here to talk about the iDDS piece that is “next to” / “inside” the data lake.
- That said, that piece is most relevant to the milestones 12 months ago – we will have a bit of time to integrate that.

Data Lake Prototypes - again

We have distinct deadlines in 3, 6, and 12 months. Rather than treat the milestones independently, I'd rather have them become “data lake challenges”. Something along the lines of:

- **Challenge #1 (3mos):** Use ServiceX to deliver 1M events to SkyHook.
- **Challenge #2 (6mos):** Via Rucio & HTTP, orchestrate movement of 200TB of CMS NanoAOD through a ServiceX instance on the SSL and process for a “real analysis” with Coffea.
- **Challenge #3 (12mos):** Demonstrate processing of 1PB of events via IDDS (maybe on an HPC?), transform them using ServiceX, cache in SkyHook, and deliver to Coffea for analysis.

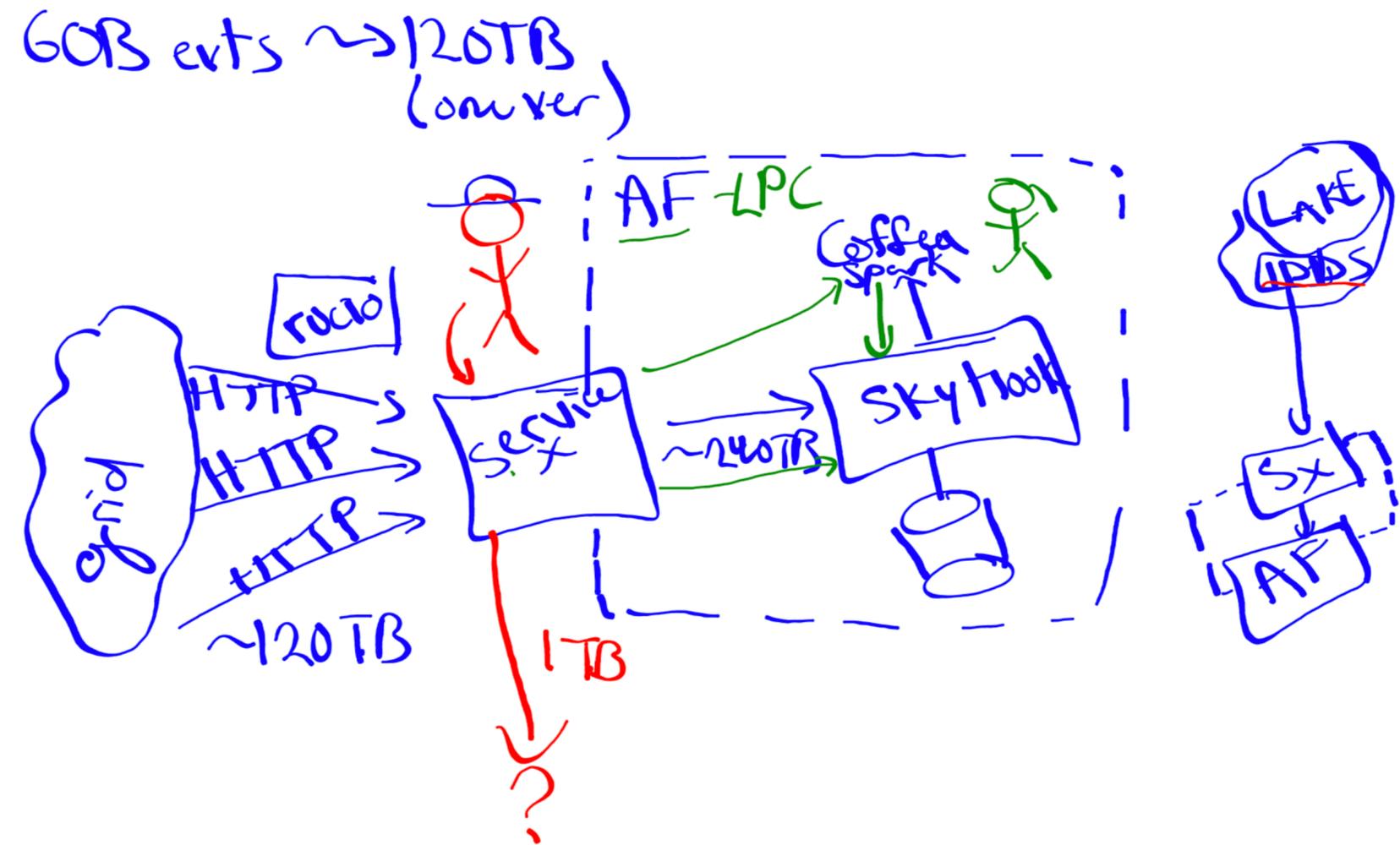
Don't mind the contents of the above three – the important aspect is increasing scale. We will discuss the contents at today's DOMA breakout post-lunch.

- Organizing along these lines helps illustrate the impact of the DOMA area.

Challenge #2

Via Rucio & HTTP, orchestrate movement of 200TB of CMS NanoAOD (~60B of Run2) through a ServiceX instance on the SSL and process for a “real analysis” with Coffea.

- March 1
- Didn't get to discuss it (ran out of time!), but we also decided we want to repeat this demo twice – once for CMS data, once for ATLAS.



Challenge #3

(Did I mention we ran out
of time?)



What's next?

We have enough “input from the crowd” to get the big picture, but not all of the details.

- In the near term, we will be putting two “challenges” documents into the IRIS-HEP shared folder to produce “2 pager” descriptions.
 - Goal is to have these done by the next DOMA group meeting (2 weeks).
- There is a longer discussion around the third challenge we still need to have.
- These documents will feed into milestones & metrics for the challenges themselves.
 - Still need to work on ‘impact metrics’ for the experiments...

In the meantime, December 1 is coming up soon!



MORGRIDGE
INSTITUTE FOR RESEARCH
CORE COMPUTATION

morgridge.org

This material is based upon work supported by the National Science Foundation under Grant No. [1836650](#). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

FEARLESS SCIENCE