DOMA – Year 3 & Beyond

MORGRIDGE
INSTITUTE FOR RESEARCH
CORE COMPUTATION

FEARLESS SCIENCE

The DOMA Team

UCSD

UIUC

U Chicago

Not shown: Chen Zou, graduate student

Morgridge

UNL

Starting ~next month

Wisconsin

Joined late 2019

UCSC

The DOMA team is a distributed team working across UIUC, Morgridge, U Chicago, UCSD, UCSC, UNL, UW.

MORGRIDGE INSTITUTE FOR RESEARCH

Impressions from the Retreat

DOMA serves as a nexus within the institute:

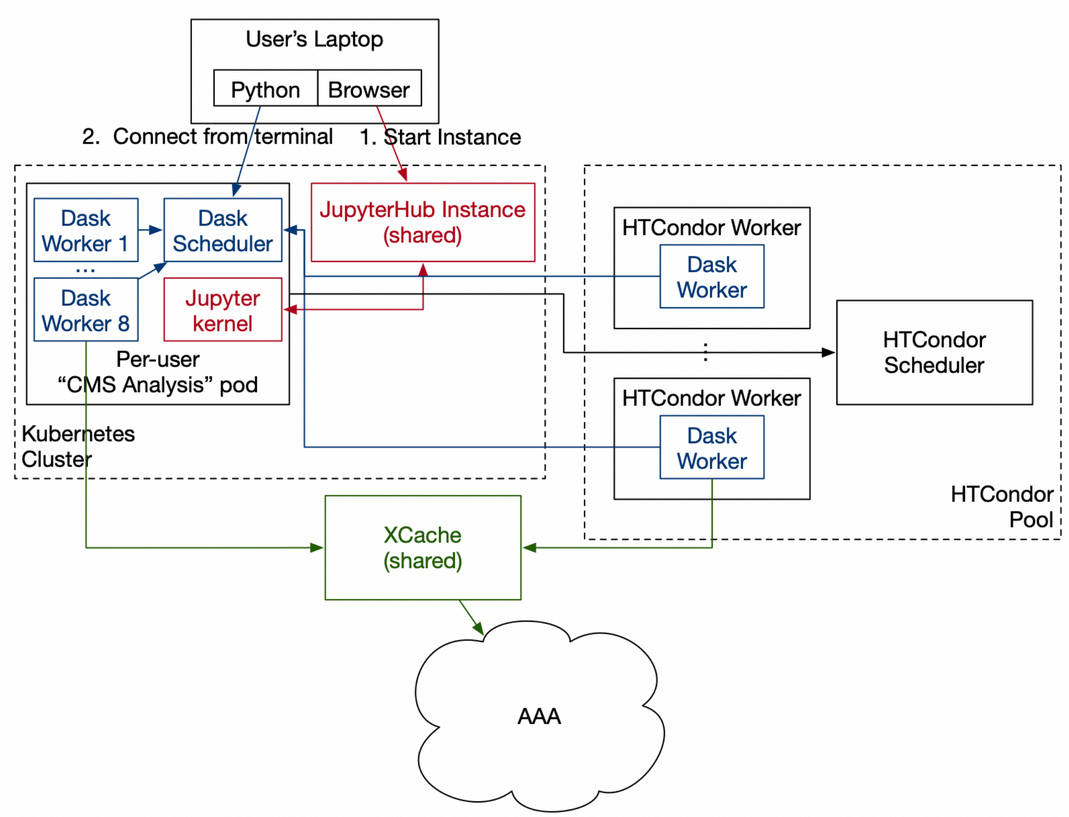
- OSG-LHC / DOMA / SSL – overlap especially on production topics. DOMA does work on several systems (IDDS, XCache, HTTP-TPC) that delivers data to applications in the production infrastructure.
- DOMA / AS – manage a joint project (ServiceX); DOMA is responsible for getting data (particularly, columnar data!) to analysis applications.

Theme: Integrate

Integrate

Integrate

Thinking about Systems...

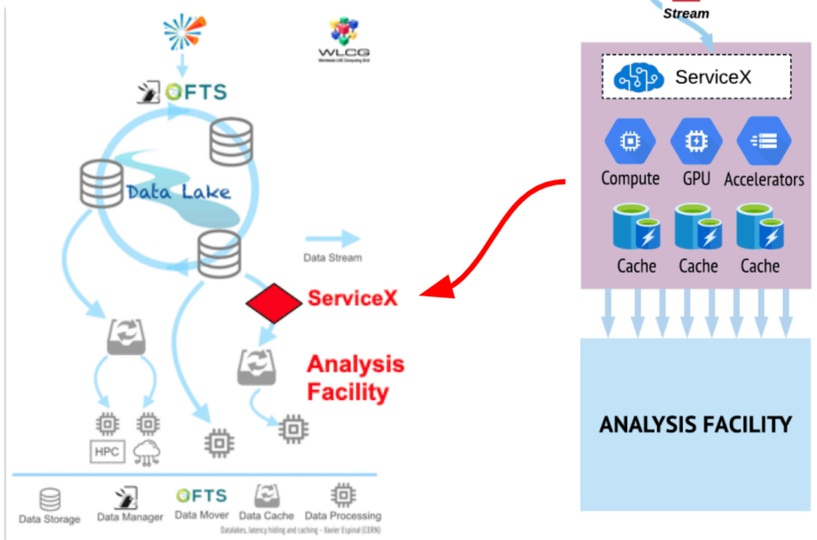


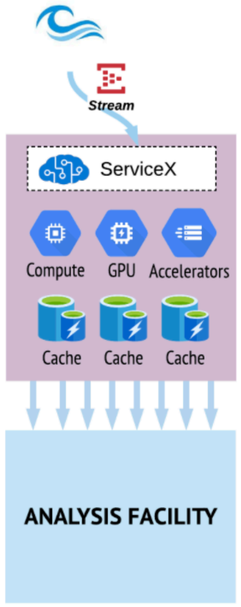
Excited by the demo on Wednesday of the Nebraska Analysis Facility (see recording).

- In IRIS, R&D is critical – but the Institute spans from R&D to operations.
  - We need to consider the entire pipeline.
- An analysis facility is important to DOMA as it provides a place for our work to land.
  - Items targeting production have a significant preexisting architecture to deliver to.

Joint Work with SSL

Accelerator Testbed





Apply hardware acceleration to data delivery

- Ingest at 50-100 Gbps
- Uncompress ROOT format
- Reformat to columnar format
- Optionally apply filter

We are far away from deploying accelerators to our storage systems!

However, by tying it to ServiceX we are trying to connect it to the pipeline.

Next Level of Internal Milestones – Analysis Oriented

Year 3 milestones:

- Production ServiceX for ATLAS and CMS, ready for users: 1 December 2020?
  - Webpages, documentation, tutorials!
  - Ready to deliver data to either S3 or SkyHook.
  - Has a userbase of XYZ teams. Thoughts on XYZ?
- Integrate ServiceX, SkyHook, and Coffea into an analysis facility prototype: 1 Sept 2021.
- SkyHook: User documentation, one at-scale demo analysis.

Needed: Analysis Facility Blueprint?

Can we wait until Year 3 even?

Year 3 Milestones – Production Oriented

- Get at least 1 USATLAS site and 1 USCMS site >30% non-GridFTP in production.
  - Have one site hit >50% non-GridFTP.
  - End of Year?
- Demonstrate use of IDDS in two ATLAS production workflows.
- A total of 4 sites with XCache installs at “production scale” (1PB / month).

Year 3 – Grand Challenges

I'll note both proposed Grand Challenges involve DOMA.

- Includes personnel / approaches (e.g., asking people to deploy DOMA services via k8s) and software products (SkyHook).
- Planning to redirect significant effort toward these activities –will replace the prior modeling work.
- Too soon to set exact milestones – but assume there will be one each year.
  - What's the forum for following up in the next month as we work on the PEP?

What does a successful Year 4 look like?

Using my crystal ball, what does a successful IRIS-HEP Year 4 look like?

- Multiple sites have successfully retired GridFTP endpoints in favor of HTTP-TPC.
- IDDS is integrated into the ATLAS tape carousel operations in preparation for HL-LHC operations.
- ServiceX/IDDS reliably delivers columns to ATLAS and CMS users in an analysis facility.
- At least one LHC site has transitioned to a caching-only mode.

These are all extensions of work done in the design phase – and directly tackle the original motivations for working on DOMA toward HL-LHC.

Slide Unchanged from Tuesday!

MORGRIDGE

INSTITUTE FOR RESEARCH

CORE COMPUTATION

morgridge.org

FEARLESS SCIENCE