(US) CMS Analysis Facilities

Burt Holzman, Fermilab

OSG All-Hands Meeting, Joint USCMS/USATLAS Session

Mar 5, 2021

Analysis Facilities are not new

CAF specs

The intention of the CAF specs from 1997 seem to have been to define central analysis facilities that would provide a level of "user satisfaction" similar to Run I.

The CMS User Analysis Farm (UAF)

- Introduction
- News
- How to get access to the cluster?

Analysis Facilities are not old

SnowMass2021

The Evolution of HEP Analysis Facilities

Paolo Calafiura <pcalafiura@lbl.gov>
V0, 8/31/2020

Analysis Facilities for Late-Stage Analysis

Kevin Lannon (klannon@nd.edu)[1], Paul Brenner[1], and Mike Hildreth[1]

[1]Univeristy of Notre Dame

Snowmass Letter of Interest - Analysis Facilities - CompF5

Burt Holzman <burt@fnal.gov> (Fermilab), Lindsey Gray (Fermilab), Bo Jayatilaka (Fermilab), Nick Smith (Fermilab)

Why evolve Analysis Facilities?

- Traditionally in HEP we have done **row-wise** analysis
  - Process one event at a time serially
  - Scripts and batch jobs (and now also notebooks)
  - We need to continue to support this

- Trending towards **column-wise** (tidy/big data) analyses
  - **Low-latency** access to CPU resources
  - Fast access to a significant amount of **disk**
  - Large investment from industry

- Low latency implies <100% resource utilization
  - **Coscheduling** analysis with other computation is appealing

Core elements for analysis facilities

- As a foundation layer, use a **containerized infrastructure** (Kubernetes-based).
- A **Jupyterhub** deployment as the glass
- Support “typical” (**row-wise**) analysis needs
- Support columnar analysis low latency application (e.g. Apache Spark, **Dask**, …) and frameworks (**Coffea**)
- Deploy **analysis services** (ServiceX, Skyhook, pyhf, …)
- Packaging and redeployment
  - e.g. Demonstrate running FNAL AF at a US CMS Tier 2

Coffea: a columnar analysis framework for HEP

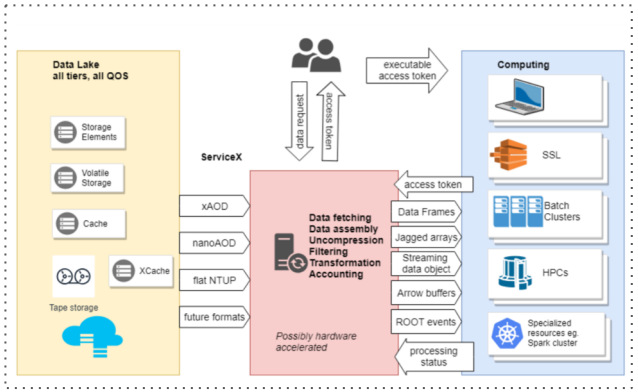
- Developed by the US CMS research program
  - Driven by Lindsey Gray, Nick Smith @ **Fermilab**

- Provides a **user interface** for columnar analysis with missing pieces of the stack filled in get users from ROOT files to histograms; some of this is transitional

- Components include
  - Lookup tools
    - Experiment-specific corrections in array programming context
  - Histogram tools (API will be *hist* package + mplhep)
  - NanoEvents
    - Wrap “flat TTree” as high-level lazy-access awkward array; enables cross-references, self-references, vector semantics, dynamic attributes, …
  - Processors
    - Wrap analysis code (utilizing the above) and take care of map-reduce

Analysis Services – a sampling

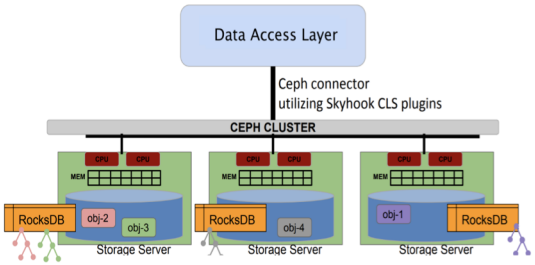
ServiceX



ServiceX provides user level ntuple production

- Converts experiment-specific datasets to columns (e.g. NanoAOD)
- Enable simple cuts or simple derived columns and fields
  - *Heavy-weight analysis will still happen via some separate processing toolchain (like CRAB)*

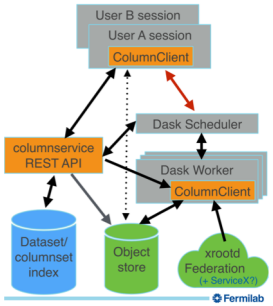
Skyhook DM



SkyhookDM converts ROOT files to columnar format

- *Ceph*-side C++ plugins transition from on-disk format to desired memory format
- Uses Dask workers to distribute data to clients
- Data delivered as Arrow tables or via Arrow Dataset API

Columnservice

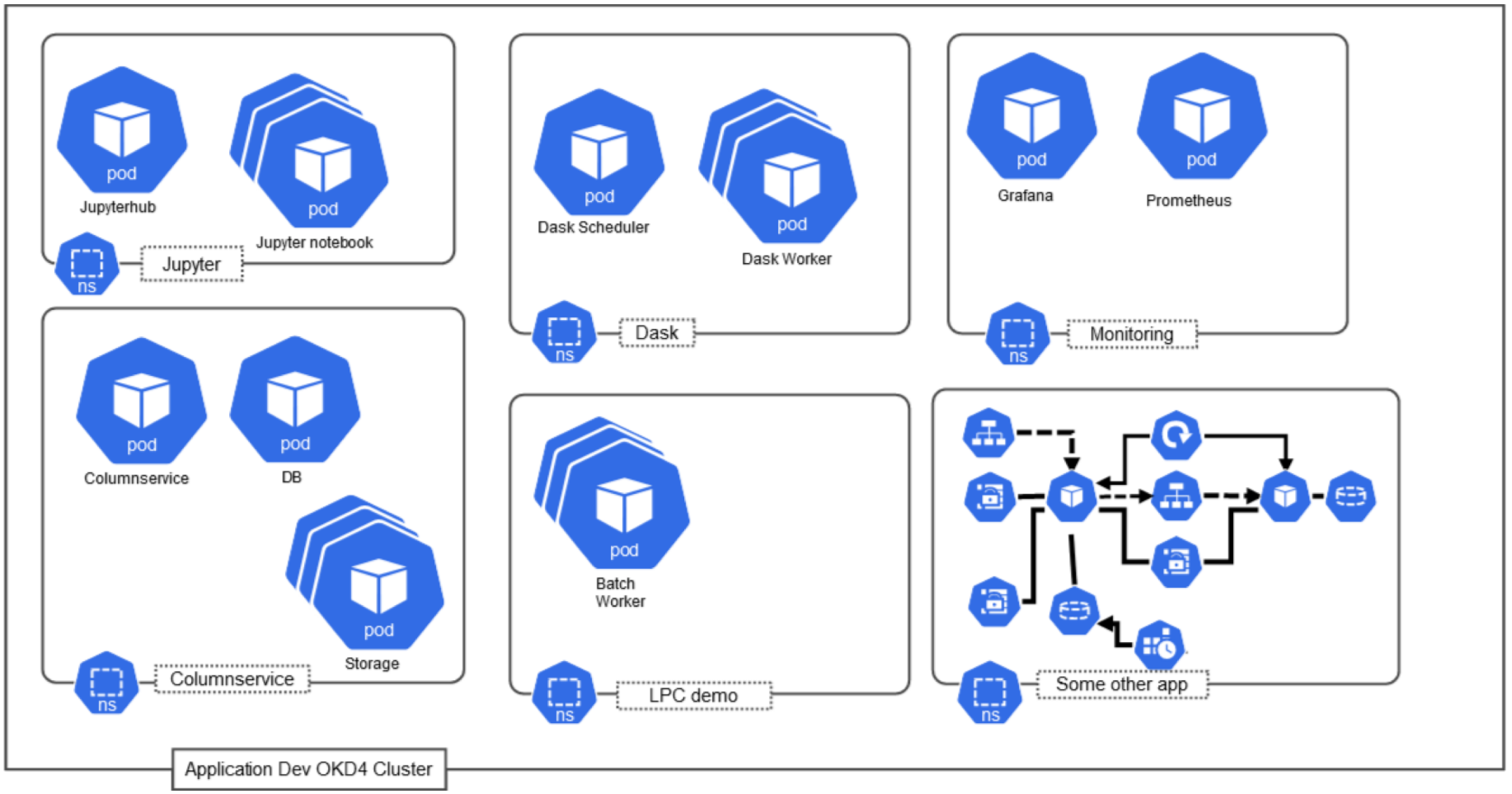


Columnservice caches columnar data for multiple users

- Shared input cache at column granularity
- Derived columns only constructed and cached on access
- Unified metadata and dataset schema database

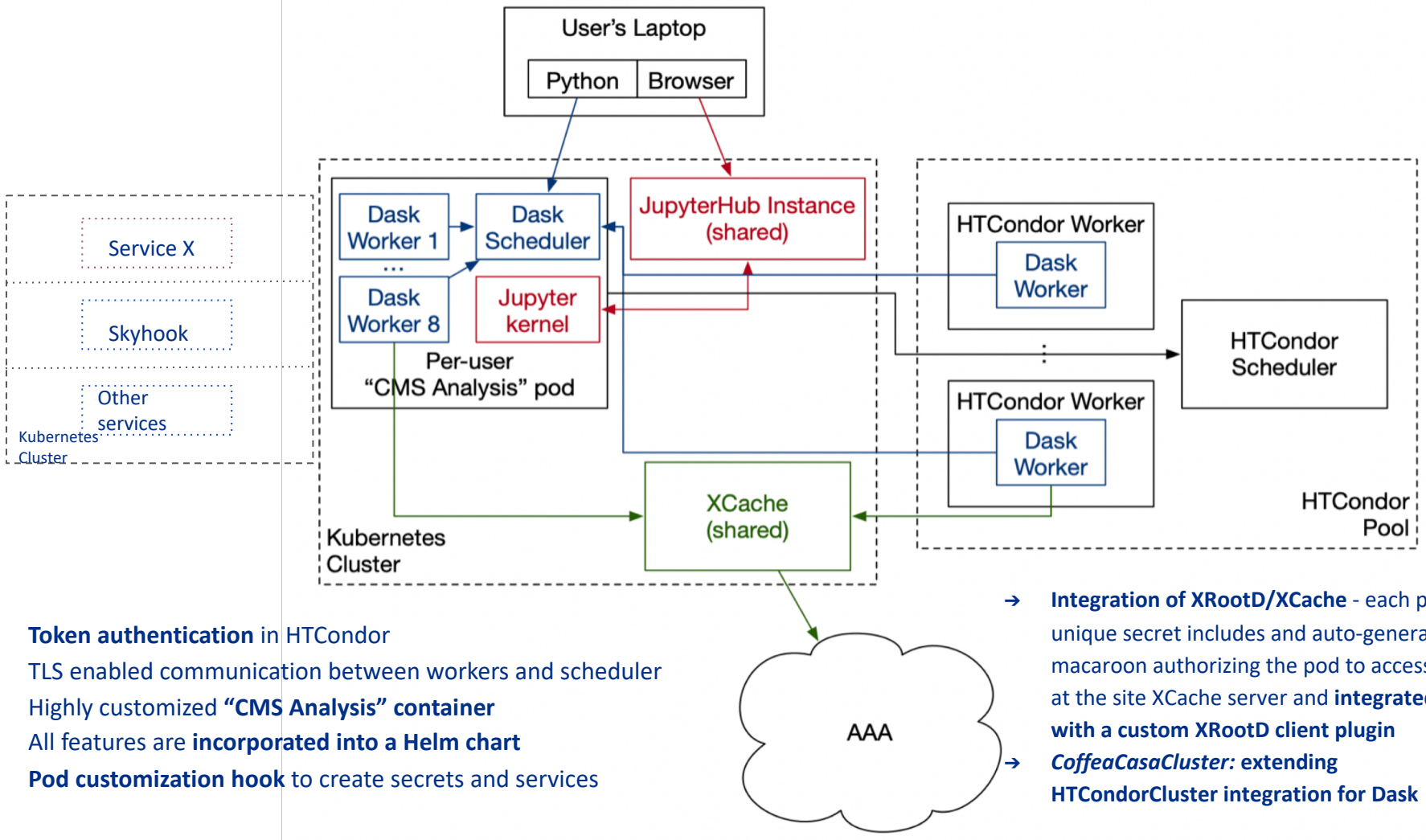
Fermilab

Fermilab Analysis Facility: pre-prototype



Fermilab

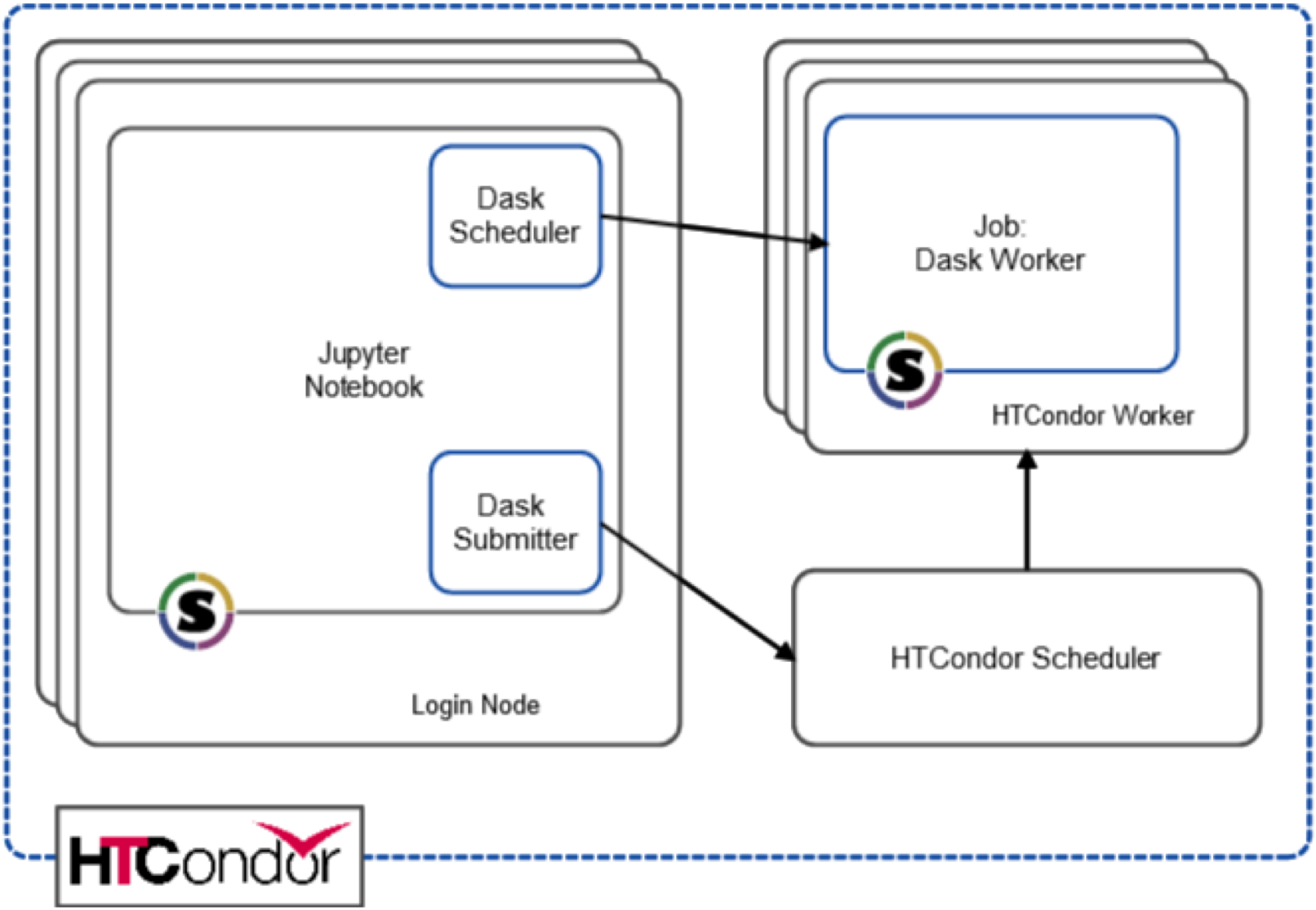
Analysis Facility "Coffea-casa" @ UNL



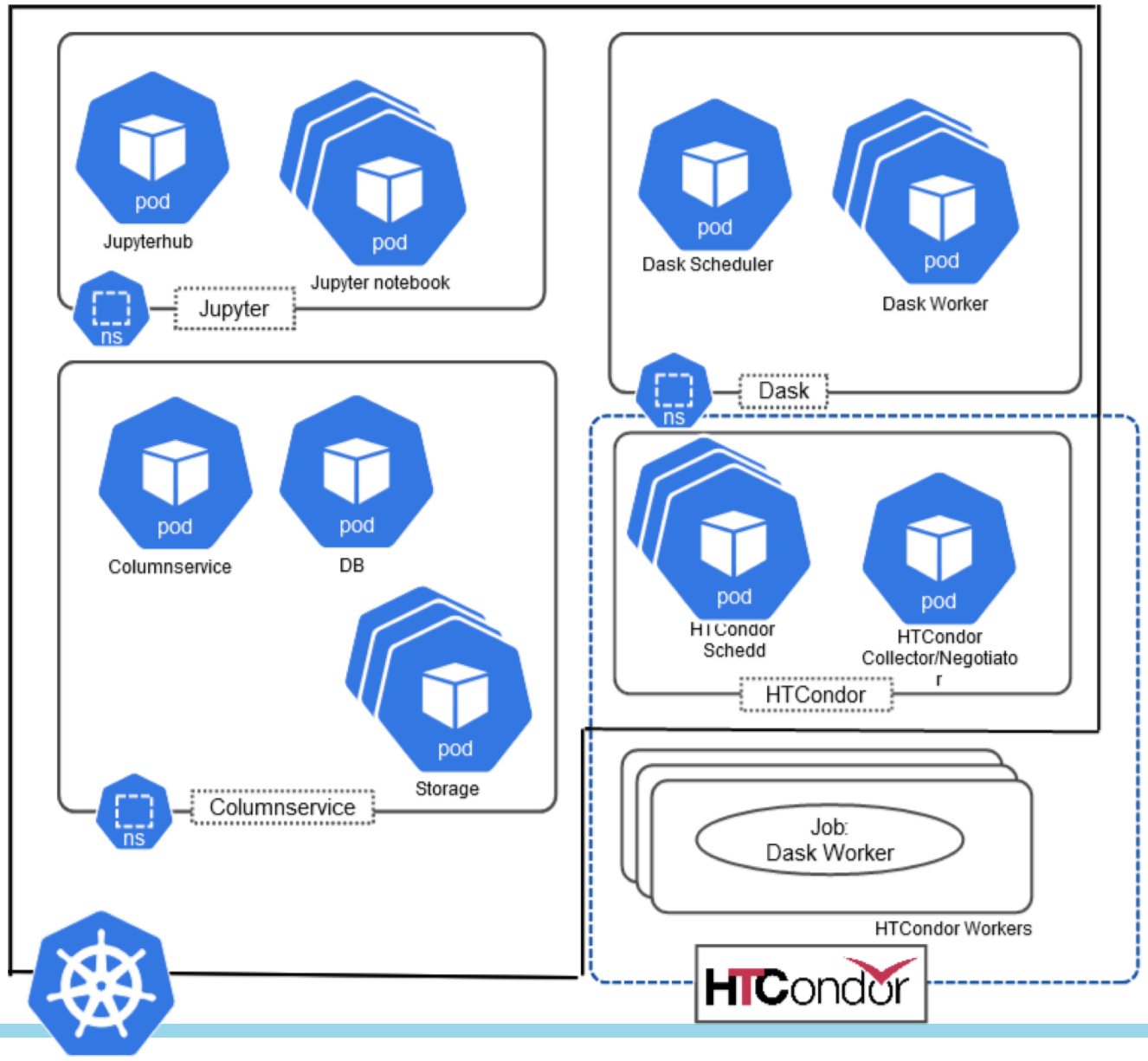
- **Token authentication** in HTCondor
- TLS enabled communication between workers and scheduler
- Highly customized **"CMS Analysis" container**
- All features are **incorporated into a Helm chart**
- **Pod customization hook** to create secrets and services

- **Integration of XRootD/XCache** - each pod's unique secret includes and auto-generated macaroon authorizing the pod to access files at the site XCache server and **integrated with a custom XRootD client plugin**
- ***CoffeaCasaCluster:*** **extending HTCondorCluster integration for Dask**

Fermilab interim solution – Dask on HTCondor



Fermilab Analysis Facility plans – adding elasticity



Fermilab