A Pilot Analysis Facility at CERN, Architecture, Implementation and First Evaluation

A. Sciabà, A. Delgado, B. Jones, D. Castro, E. Tejedor, E. García, G. Guerrieri, M. Schulz
CERN (IT department)

Context and motivation

- A working group was established at CERN to:
  - Design and implement a pilot of an Analysis Facility
  - Obtain feedback from early testers
  - Validate real demand before investing more effort
- Pilot designed to augment the current offering with:
  - Interactive computing on big datasets, with analysis built on frameworks like RDataFrame and coffea
  - Possibility to scale out
- Several other AF initiatives have been setup within the LHC community
- Perception of demand from:
  - Forums: HSF Analysis Facilities WG, WLCG Pre-CHEP Workshop 2023
  - Users: first request for coffea + LxBatch in 2021

CERN's Analysis Facility

- CERN provides infrastructure and services for analysis since a long time
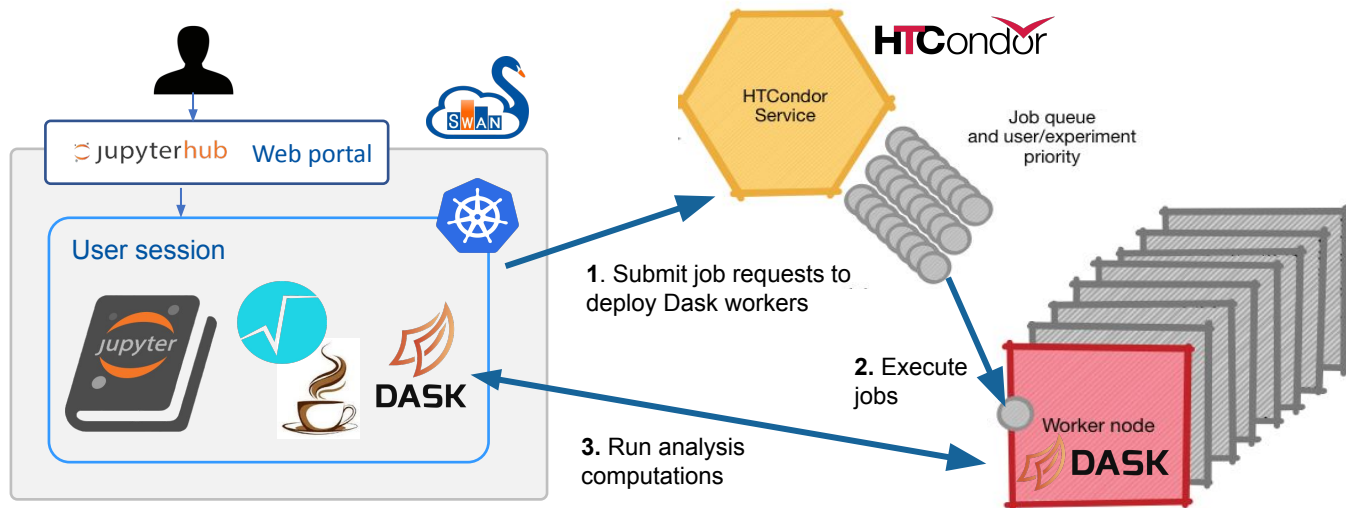
The pilot

- Focus on **scale out of interactive analysis**
  - On **already existing** CERN Batch system resources
  - Via RDataFrame / coffea + Dask

Resource management for pilot's interactive jobs

- Pilot's interactive jobs benefit from dedicated resources in the CERN HTCondor pool
  - Dedicated buffer of ~2k cores for quick allocation of resources (needed to feel interactive!)
  - Extra resources from common pool, job start time **subject to experiment quotas**