



Contribution ID: 10

Type: **not specified**

Combining cloud-native workflows with HTCondor jobs

Thursday, September 24, 2020 3:15 PM (20 minutes)

The majority of physics analysis jobs at CERN are run on high-throughput computing batch systems such as HTCondor. However, not everyone has access to computing farms, e.g. theorist wanting to make use of CMS Open Data, and for reproducible workflows more backend-agnostic approaches are desirable. The industry standard here are containers leveraged with Kubernetes, for which computing resources can easily be acquired on-demand using public cloud offerings. This causes a disconnect between how current HEP physics analysis are performed and how they could be reused: when developing a fully “cloud native” computing approach for physics analysis, one still needs to have access to the ten-thousands of cores available on classical batch system to have sufficient resources for the data processing.

In this presentation, I will demonstrate how complex physics analysis workflows that are written and scheduled using a rather small Kubernetes cluster can make use of CERN’s HTCondor installation. An “operator” is used to submit jobs to HTCondor and—once completed—collect the results and continue the workflow in the cloud. The audience will also learn the important role that software containers and Kubernetes play in the context of open science.

Desired slot length

20

Speaker release

Yes

Primary author: LANGE, Clemens (CERN)

Presenter: LANGE, Clemens (CERN)

Session Classification: Workshop session

Track Classification: HTCondor user presentations