Contribution ID: **315** Type: **Oral**

Design Pattern for Analysis Automation on Interchangeable, Distributed Resources using Luigi Analysis Workflows

Thursday 14 March 2019 15:30 (20 minutes)

In particle physics, workflow management systems are primarily used as tailored solutions in dedicated areas such as Monte Carlo production. However, physicists performing data analyses are usually required to steer their individual workflows manually which is time-consuming and often leads to undocumented relations between particular workloads.
We present the luigi analysis workflow (law) Python package which is based on the open-source pipelining tool luigi, originally developed by Spotify. It establishes a generic design pattern for analyses of arbitrary scale and complexity, and shifts the focus from executing to defining the analysis logic. Law provides the building blocks to seamlessly integrate with interchangeable remote resources without, however, limiting itself to a specific choice of infrastructure. In particular, it introduces the paradigm of complete separation between analysis algorithms on the one hand, and run locations, storage locations, and software environments on the other hand.
To cope with the sophisticated demands of end-to-end HEP analyses, law supports job execution on WLCG infrastructure (ARC, gLite) as well as on local computing clusters (HTCondor, LSF), remote file access via most common protocols through the Grid File Access Library (GFAL2), and an environment sandboxing mechanism with support for Docker and Singularity containers. Moreover, the novel approach ultimately aims for analysis preservation out-of-the-box.
Law is developed open-source and entirely experiment independent. It is successfully used in ttH cross section measurements and searches for di-Higgs boson production with the CMS experiment.

Authors: RIEGER, Marcel (Rheinisch-Westfaelische Tech. Hoch. (DE)); ERDMANN, Martin (Rheinisch-Westfaelische Tech. Hoch. (DE))

Presenter: RIEGER, Marcel (Rheinisch-Westfaelische Tech. Hoch. (DE))

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research