Contribution ID: **410** Type: **Poster**

Law: End-to-End Analysis Automation over Distributed Resources

Thursday 27 October 2022 16:10 (30 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, complex workflows manually, frequently involving job submission in several stages and interaction with distributed storage systems by hand. This process is not only time-consuming and error-prone, but also leads to undocumented relations between particular workloads, rendering the steering of an analysis a serious challenge.
This contribution presents 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 concept 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, Slurm, LSF), remote file access via various protocols using the Grid File Access Library (GFAL2), and an environment sandboxing mechanism with support for sub-shells and virtual environments, as well as Docker and Singularity containers. Moreover, the novel approach ultimately aims for analysis preservation out-of-the-box.
Law is developed open-source and independent of any experiment or the language of executed code. Over the past years, its user-base increased steadily with applications now ranging from (pre-)processing workflows in CMS physics objects groups, to pipelines performing the statistical inference in most CMS di-Higgs searches, and it serves as the underlying core software for large scale physics analyses across various research groups.

Experiment context, if any

I am with CMS. The presented software is experiment-independent.

References

Significance

I've presented this topic in an earlier stage at ACAT 2019, but over the past years the user-base multiplied, making it the most widely used workflow management tool (apart from central production pipelines) at CMS. Law will be used for many Run 3 analyses, and apart from a gentle introduction into the key concepts and main features, I intend to show a demonstrator for a full-blown and not necessarily CMS-related physics analysis with law.

Primary author: RIEGER, Marcel (Hamburg University (DE))

Presenter: RIEGER, Marcel (Hamburg University (DE))

Session Classification: Poster session with coffee break