



Contribution ID: 116

Type: Poster

Parametrizing workflows with ParaO and Luigi

Workflow tools provide the means to codify complex multi-step processes, thus enabling reproducibility, preservation, and reinterpretation efforts. Their powerful bookkeeping also directly supports the research process, especially where intermediate results are produced, inspected, and iterated upon frequently.

In Luigi, such a complex workflow graph is composed of individual tasks that depend on one another, where every part can be customized at runtime through parametrization. However, Luigi falls short with regards to the steering of parameters, accounting for the consequences thereof, and the modification or reuse of task graphs.

This is where the parameter handling of ParaO shines: it has vastly extended key mechanics and value coercion while automatically propagating their effects throughout the task graph. Since the dependencies are described through parameters too, the same principles can be used to freely alter or transplant (parts of) the task graph, thereby empowering reuse. At the same time, ParaO remains largely compatible with Luigi and packages building upon it, such as Law.

Significance

This contribution motivates why and how parametrized workflows are powerful tool to orchestrate entire data analyses i.e. in HEP. It discusses the shortcomings of a current sophisticated implementation (i.e. luigi) and details the principles of how to address them while remaining mostly compatible with existing tooling and extensions (i.e. law).

References

<https://indico.cern.ch/event/1375573/contributions/6089483/>

Experiment context, if any

CMS

Authors: FISCHER, Benjamin (RWTH Aachen University (DE)); ZINN, Felix Philipp (Rheinisch Westfaelische Tech. Hoch. (DE)); ERDMANN, Martin (Rheinisch Westfaelische Tech. Hoch. (DE)); FARKAS, Mate (Rheinisch Westfaelische Tech. Hoch. (DE)); EICH, Niclas (Rheinisch Westfaelische Tech. Hoch. (DE)); DIEKMANN, Svenja (Rheinisch Westfaelische Tech. Hoch. (DE))

Presenter: FISCHER, Benjamin (RWTH Aachen University (DE))

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research