Orchestrating Scientific Workflows with Maestro Workflow Conductor
https://github.com/LLNL/maestrowf
Francesco Di Natale (LLNL)

Workflow Challenges
- Running HPC/software workflows is complex and error prone.
- External solutions usually employ a wide range of technologies that violate security policies.
- Users often dislike fighting with tools and “black boxes”.
- Other tools assume specifics, making it harder for users to develop their own methodologies.
- Using ML in current workflows is extremely difficult.

A Standard Workflow Experience
- Provide the means to facilitate an agile and repeatable approach to scientific workflows.
- Allowing users to focus on exploring the science, not process.
- Workflows should be represented in a manner that is clear, concise, and natural.
- A workflow tool should not assume responsibility for any specific methodology.

Maestro Study Specifications
Maestro makes use of a markup file called a “study specification”. A study contains all the fixed variables, steps, and parameters for running a study. The specification is analogous to the workflow for a physical experiment documented in a lab notebook.

Automation Made Easy
Maestro parses the specification and constructs internal representations for the data contained in the study specification. Maestro is then capable of cleverly expanding the workflow into a DAG, which is then automatically monitored and executed on compute resources.

Future Direction and Vision
- Study concept can be used to generalize workflow definitions for other tools to perform UQ, optimization, ML sampling, etc.
- Additional functionality to add parameters to existing studies, restart failed steps, add steps.

Collaborators & Users
- Maestro's study specification is both machine- and human-readable and is analogous to a page in a laboratory notebook that would be used to document a physical experiment.

Enabling Scientific Research
Maestro provides a general and simple way to define and automate computational science workflows. It’s helped users achieve multiple high-level strategic initiatives at LLNL and is open sourced for wider community use.

Maestro user testimonials:
- “Maestro has made me simulation data-rich for the first time.”
- “Maestro allowed me to focus on the engineering aspects rather than the minutia of managing the ensemble of test cases. Its structured output/workspaces also facilitated automating the post-simulation workflow to get the engineering work done faster and to quickly ask new questions as the dataset was explored.”

Maestro improves productivity by allowing users to focus on science.