

Workflows on Maxwell

Supporting data-centric science often involves the movement of data across file systems, multi-stage analytics and visualization. Workflow technologies can improve the productivity and efficiency of data-centric science by orchestrating and automating these steps (1). DESY-IT provides some support for the Nextflow, Swift and Fireworks tools.

We will add further workflow tools in the near future, so this is strictly work in progress!

Nextflow

As advertised on <https://www.nextflow.io/>: Nextflow enables scalable and reproducible scientific workflows using software containers. It allows the adaptation of pipelines written in the most common scripting languages. Its fluent DSL simplifies the implementation and the deployment of complex parallel and reactive workflows on clouds and clusters.

Swift

The [Swift](#) scripting language provides a simple, compact way to write parallel scripts that run many copies of ordinary programs concurrently in various workflow patterns, reducing the need for complex parallel programming or arcane scripting. Swift is very general, and is in use in domains ranging from earth systems to bioinformatics to molecular modeling.

Fireworks

[FireWorks](#) is a free, open-source code for defining, managing, and executing scientific workflows. It can be used to automate calculations over arbitrary computing resources, including those that have a queueing system. Some features that distinguish FireWorks are dynamic workflows, failure-detection routines, and built-in tools and execution modes for running high-throughput computations at large computing centers. It uses a centralized server model, where the server manages the workflows and workers run the jobs.

Notes

(1) Large parts of the text has been borrowed from NERSCs very comprehensive documentation: <https://docs.nersc.gov/jobs/workflow-tools/>

(2) The workflow tools are available on Maxwell only.