Supporting data-centric science involves the movement of data, multi-stage processing, and visualization at scales where manual control becomes prohibitive and automation is needed. Workflow technologies can improve the productivity and efficiency of data-centric science by orchestrating and automating these steps.

Let us help you find the right tool
Do you have questions about how to choose the right workflow tool for your application? Are you unsure about which tools will work on UA HPC systems? Please open a ticket at explain you would like help choosing a workflow tool, and your ticket will be routed to experts who can help you.

Snakemake

Snakemake is a tool that combines the power of Python with shell scripting. It allows users to define workflows with complex dependencies; users can easily visualize the job dependency graph and track which tasks have been completed and are still pending.

Nextflow

Nextflow is a data-centric workflow management tool written in Groovy, which facilitates complex and reproducible scientific computational workloads.

Makeflow

Makeflow is a workflow engine for large scale distributed computing. It accepts a specification of a large amount of work to be performed, and runs it on remote machines in parallel where possible.