

Nextflow Introduction

Nextflow is a workflow management system designed to streamline the execution of complex data pipelines.

It supports scalable and reproducible workflows, making it ideal for bioinformatics and other data-intensive fields.

Benefits Over Traditional Bash Scripting

- **Scalability:** Automatically manages parallel execution and scales from local to cloud-based environments.
- **Reproducibility:** Uses a declarative language to ensure consistent results and easy sharing.
- **Resource Management:** Integrates with various environments and schedulers for efficient resource allocation.
- **Modularity:** Allows for reusable, modular processes, simplifying pipeline maintenance.
- **Error Handling:** Includes built-in error handling and retry mechanisms.
- **Data Provenance:** Tracks data flow for better debugging and result tracking.
- **Caching and Resume:** Features caching of intermediate results and the ability to resume failed or interrupted workflows, saving time and resources.

Nextflow enhances efficiency, scalability, and maintainability compared to traditional bash scripting.

How do I use Nextflow?

Below is a video going through some of the Nextflow basics on our HPC

<https://www.youtube.com/watch?v=5UkWwRclytw>

There are a number of different docs and training materials that can be found regarding nextflow usage.

<https://www.nextflow.io/docs/latest/index.html>

Official Nextflow training video

<https://www.youtube.com/watch?v=wbTMbJTo1xo>

