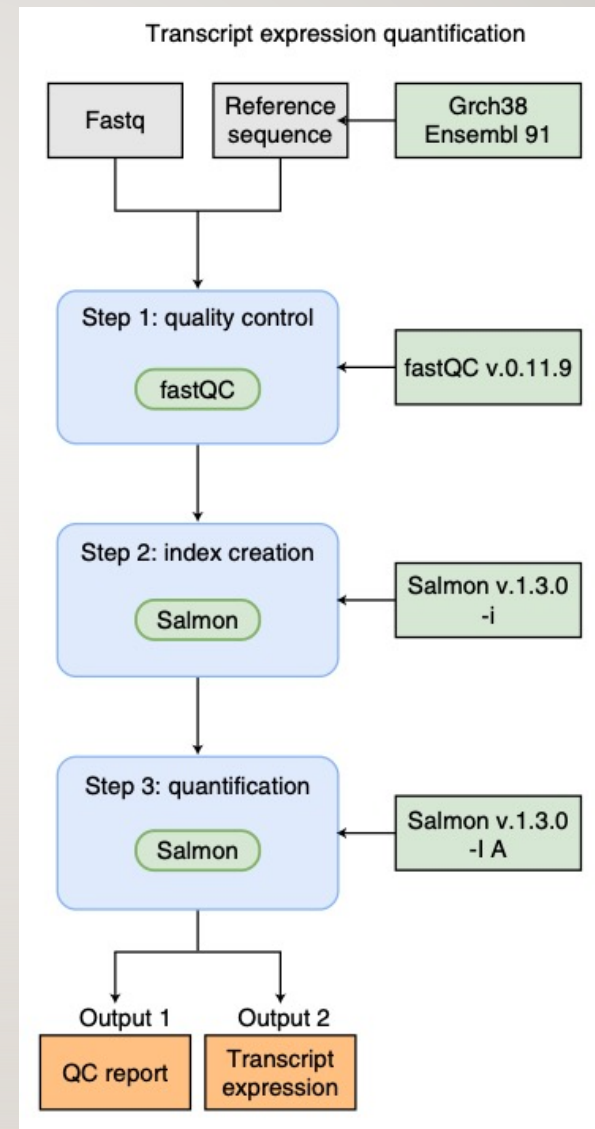


WORKFLOW MANAGEMENT SYSTEMS

A **workflow** is a defined sequence of tasks or activities that need to be completed in a specific order to achieve a desired outcome or goal.



<https://doi.org/10.1038/s41592-021-01254-9>

Historically, developed using custom scripts or **Make** files

Makefile

```
# Specify the alignment file to be generated
alignment: trimming
    command to align

# Specify the trimming required to make the alignment
trimming :
    command to trim
```

Run **make** command in the directory where the Makefile resides




A **Workflow Management System (WMS)** is a software or platform that provides an infrastructure to setup, execute, and monitor workflows.

Why do we need workflow management Systems (WMS)?

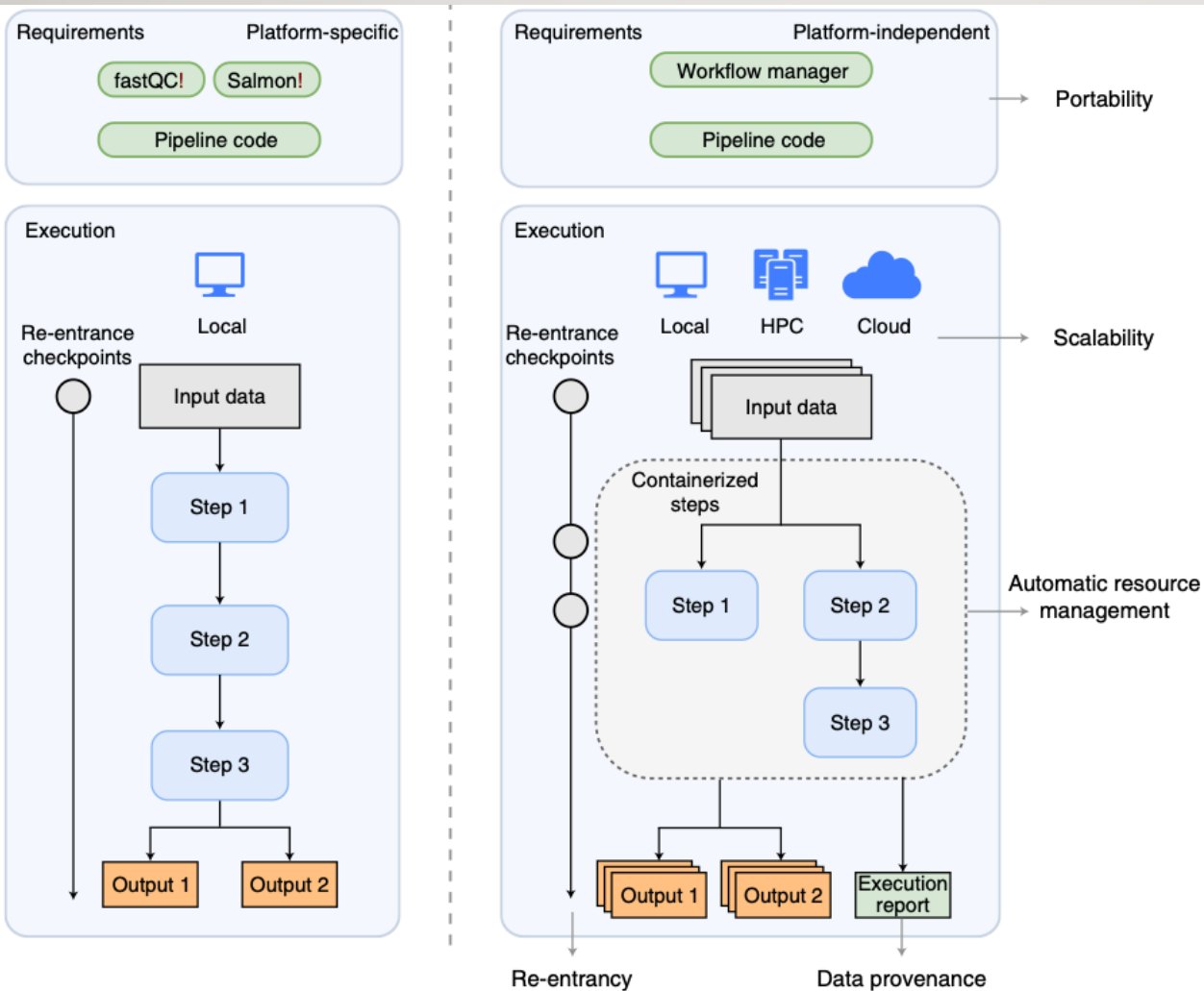


As projects grow or age, it becomes increasingly difficult to keep track of all the parts and how they fit together.

Why Workflow Management Systems (WMS)?

- To automate series of bioinformatic processing steps.
- To standardize analysis for large projects or core facilities (repetitive task)
- Portability: simplify deployment of complex pipelines
- Scalability   
Local HPC Cloud
- Optimise computation resources (parallelization, times, ...)
- Improve reproducibility

Traditional pipeline **VS** Workflow manager



- traditional pipeline implementations are coupled to the local compute environment and are sensitive to changes in software or data versions.
- workflow manager decouples the code from the execution environment

containerized software makes local software installation requirements unnecessary. execution reports can track parameters and versions, providing transparency and data provenance.

<https://doi.org/10.1038/s41592-021-01254-9>

What Workflow Management System

There are over 150 workflow managers currently in use and under development

<https://github.com/pditommaso/awesome-pipeline>

<https://github.com/common-workflow-language/common-workflow-language/wiki/Existing-Workflow-systems>

What Workflow Management System

Overview of workflow managers for bioinformatics

Tool	Class	Ease of use ^a	Expressiveness ^b	Portability ^c	Scalability ^d	Learning resources ^e	Pipeline initiatives ^f
Galaxy	Graphical	●●●	●○○	●●●	●●●	●●●	●●○
KNIME	Graphical	●●●	●○○	○○○	●●○	●●●	●●○
Nextflow	DSL	●●○	●●●	●●●	●●●	●●●	●●●
Snakemake	DSL	●●○	●●●	●●○	●●●	●●○	●●●
GenPipes	DSL	●●○	●●●	●●○	●●○	●●○	●●○
bPipe	DSL	●●○	●●●	●●○	●●○	●●○	●○○
Pachyderm	DSL	●●○	●●●	●○○	●●○	●●●	○○○
SciPipe	Library	●●○	●●●	○○○	○○○	●●○	○○○
Luigi	Library	●●○	●●●	●○○	●●○	●●○	○○○
Cromwell + WDL	Execution + workflow specification	●○○	●●○	●●●	●●○	●●○	●●○
cwltool + CWL	Execution + workflow specification	●○○	●●○	●●○	○○○	●●●	●●○
Toil + CWL/WDL/Python	Execution + workflow specification	●○○	●●●	●○○	●●●	●●○	●●○

Galaxy a powerful choice for users with or without computational expertise who would like to assemble and run custom bioinformatics workflows.

<https://www.nature.com/articles/s41592-021-01254-9>