

Clean and Reproducible

MiCM Workshop 2020-06-17

This workshop is divided into two parts

Me talking

This workshop is divided into two parts

Me talking

Us doing

Us doing

This workshop is divided into two parts



Me talking

Us doing

Us doing

Goals:

Write basic workflows

Understand how to scale to HPC/cloud

nextflow

Snakemake

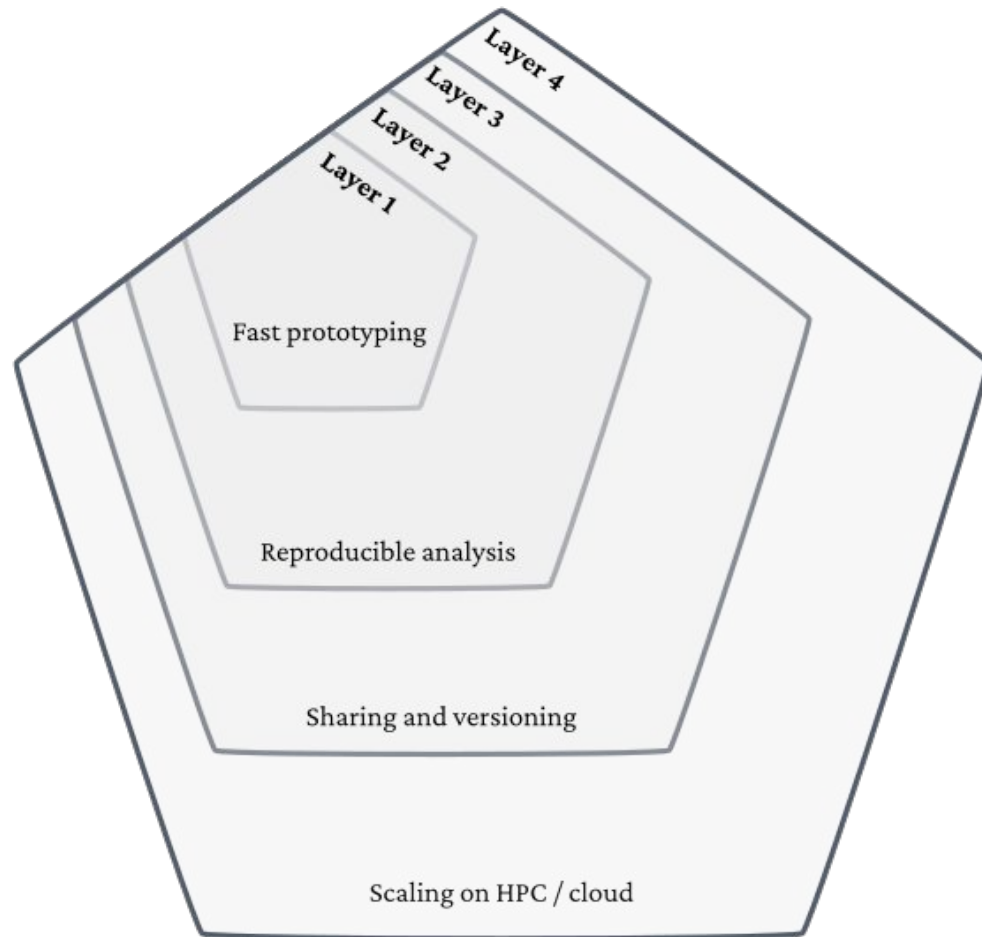
bpipe

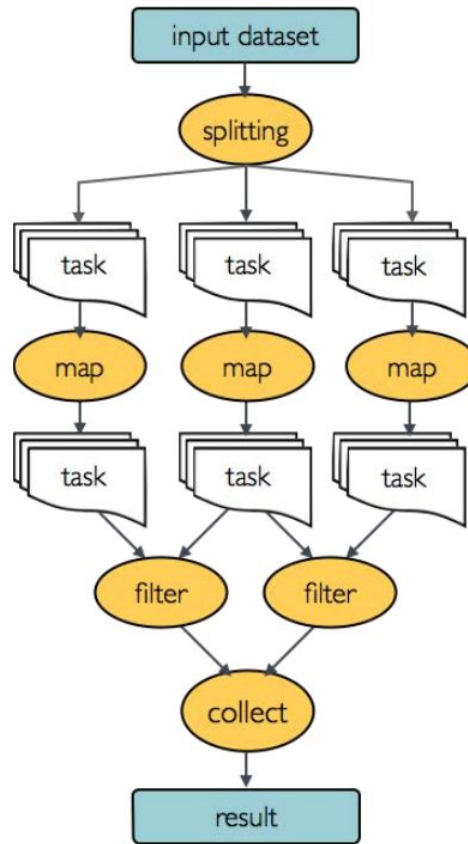
toil

luigi

scipipe

Cromwell/WDL

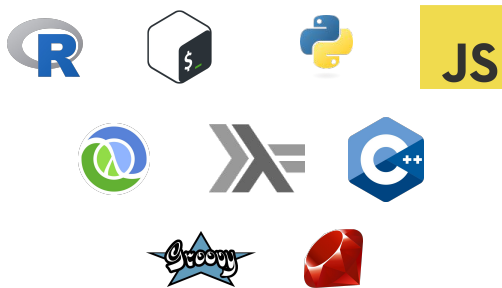




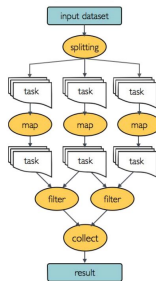
The dataflow programming model defines a directed graph of operations.



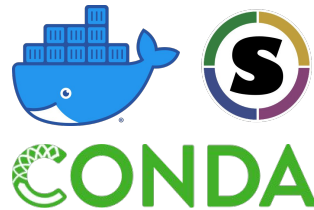
Write in any language



Define process graph



Explicit software dependencies



All under version control

GitHub



GitLab

ATLASSIAN



Bitbucket

Orchestration of tasks onto compute



nextflow
script

nextflow
runtime

Hello, Nextflow!

Version control

Built-in Operators

Adding a second process

Adding a second process

Switching tools

Publishing outputs

Exposing parameters

Moving to HPC

Moving to Cloud

Cloud-specific parameters

Bioinformatics example

Clean and Reproducible

MiCM Workshop 2020-06-17