Snakemake

Making High-Performance Computing a bit easier

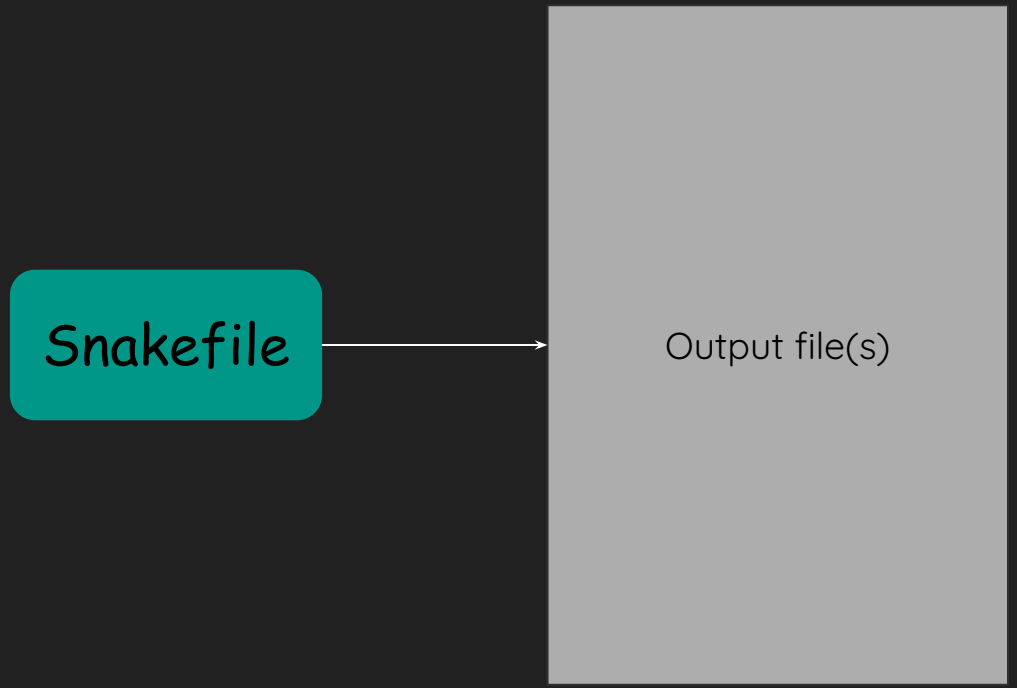
Snakemake

The Snakemake **workflow management system** is a **tool** to create reproducible and scalable data analyses.

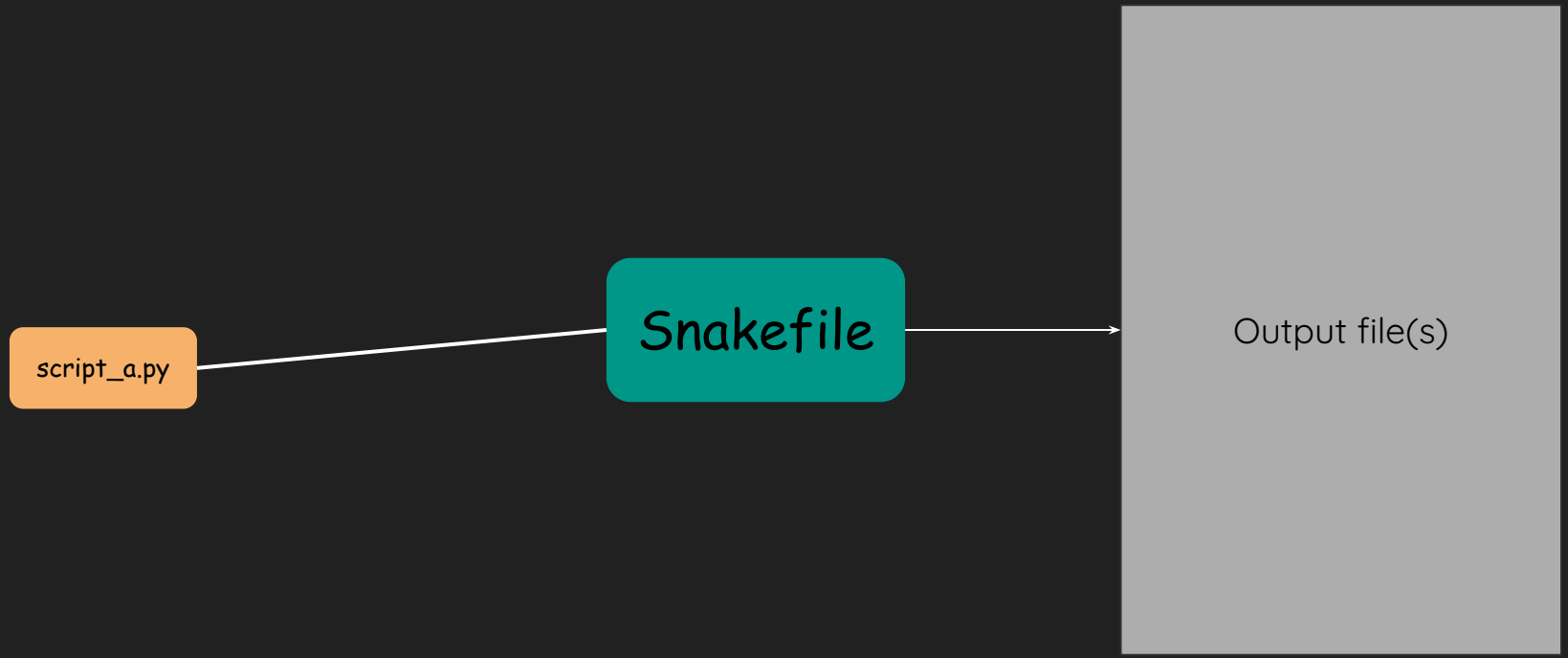
Easy use of multiple threads without the dangers of multithreading.

Snakemake pipeline structure

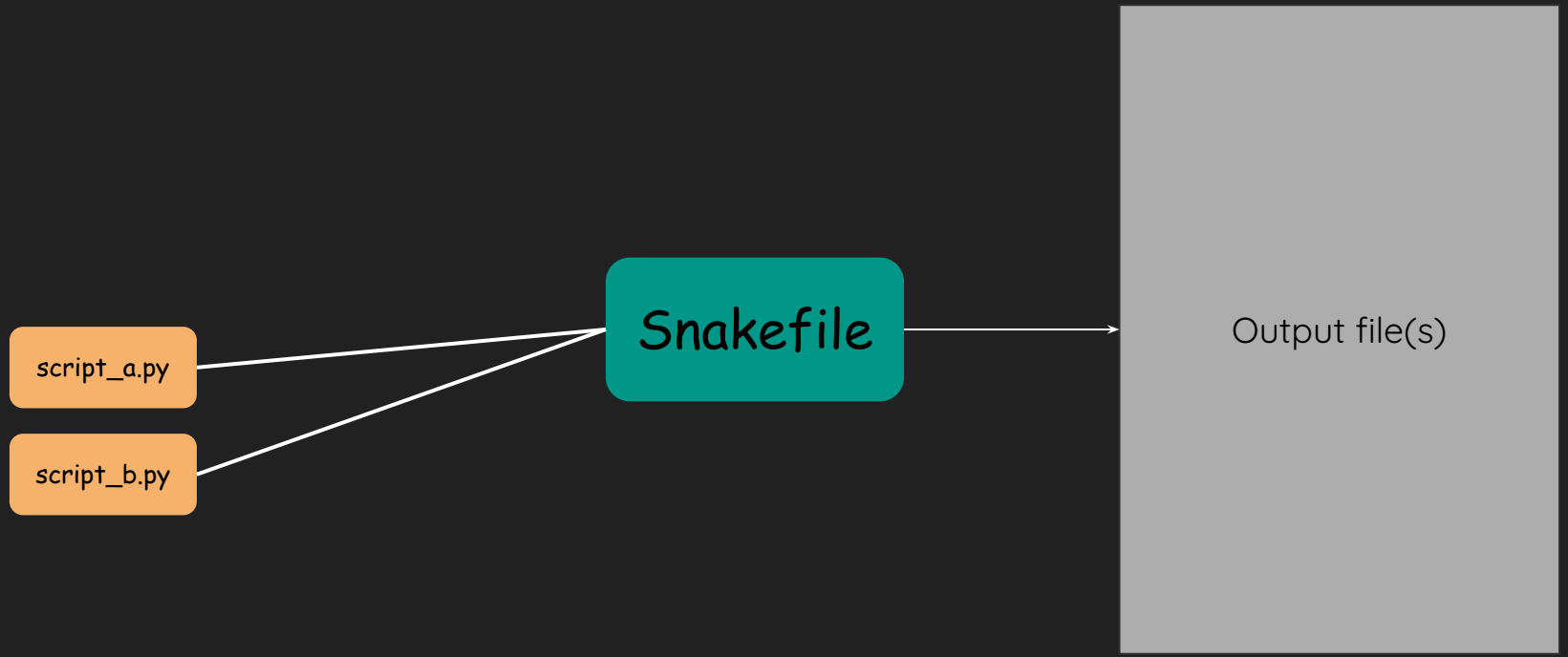
Snakemake pipeline structure



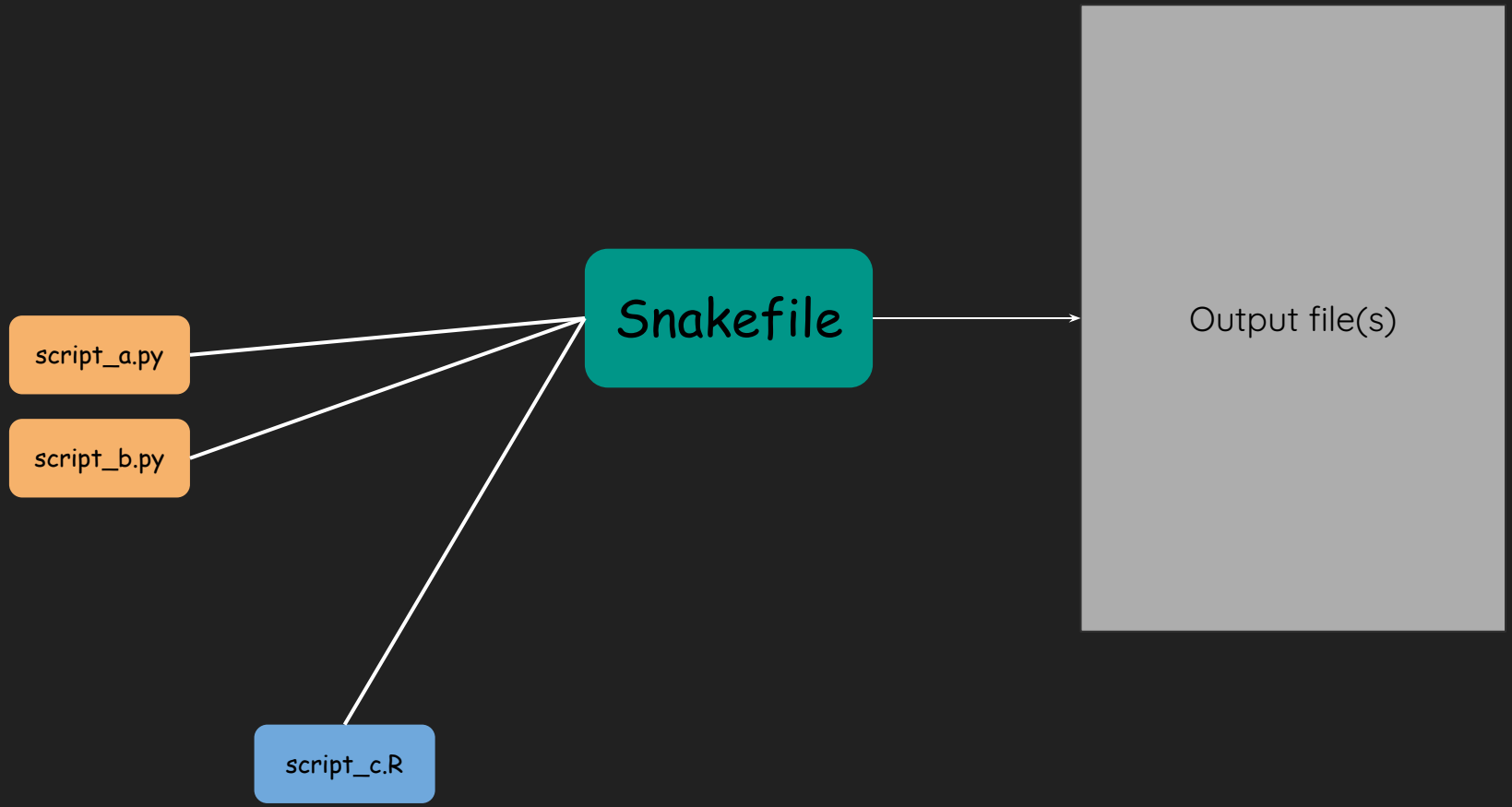
Snakemake pipeline structure



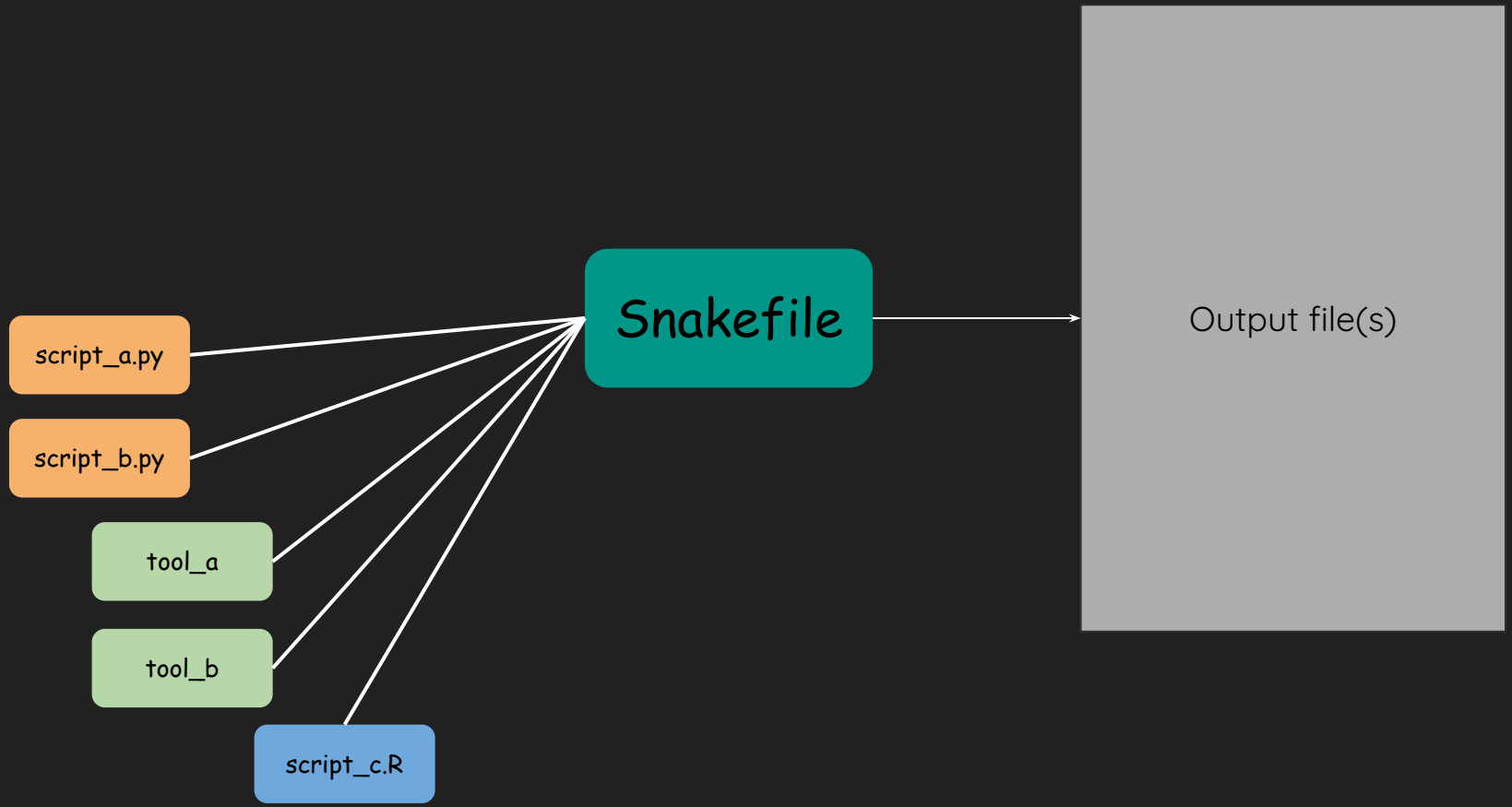
Snakemake pipeline structure



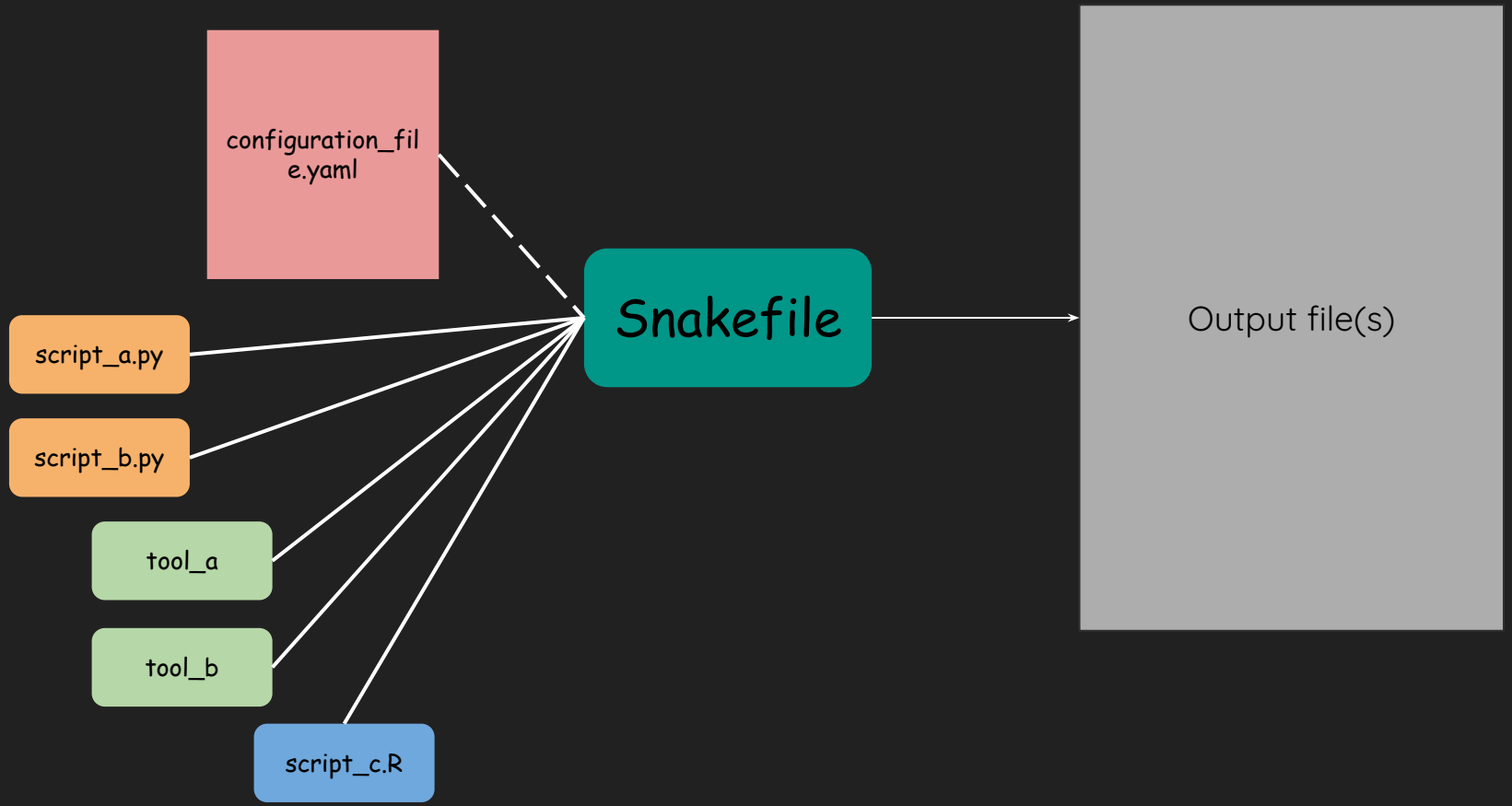
Snakemake pipeline structure



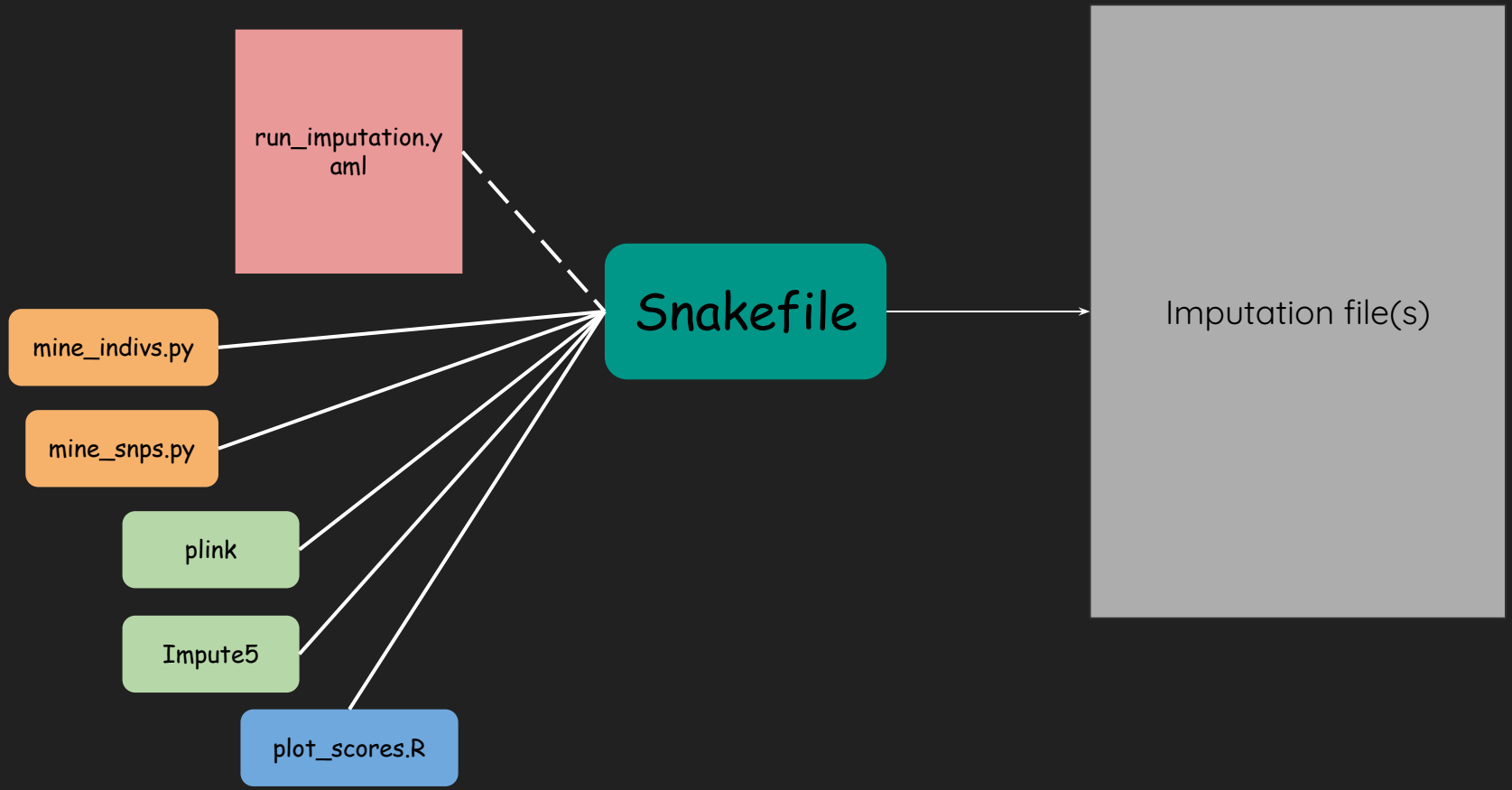
Snakemake pipeline structure



Snakemake pipeline structure

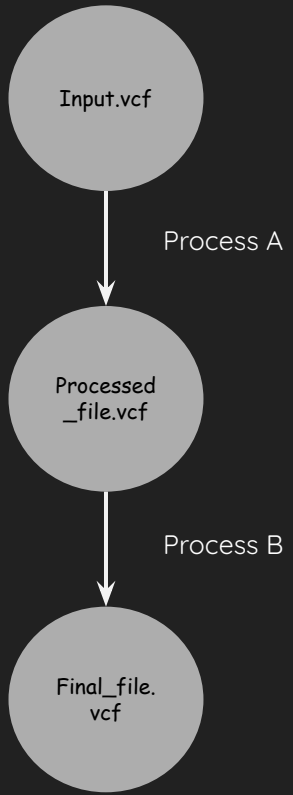


Snakemake pipeline structure

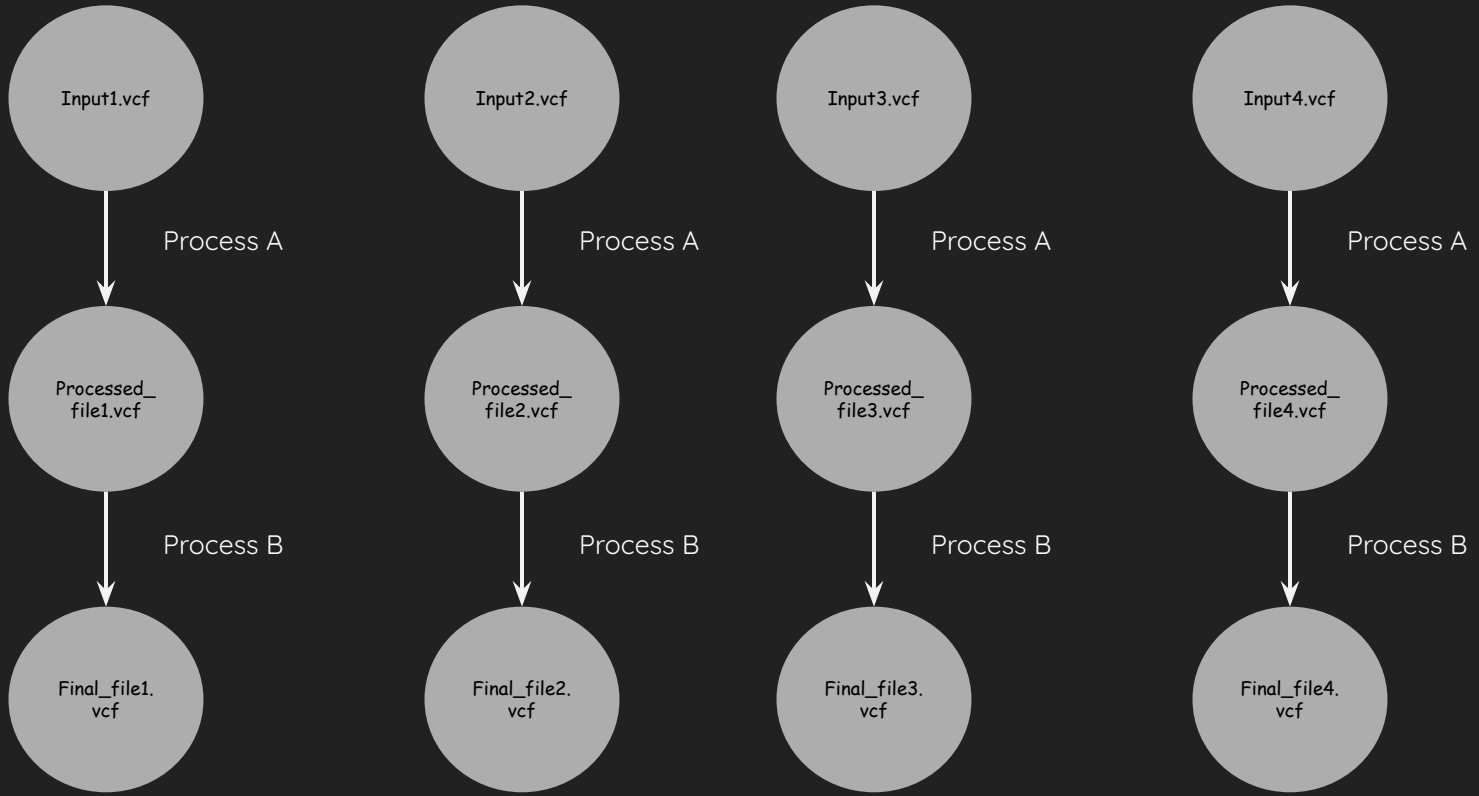


The process

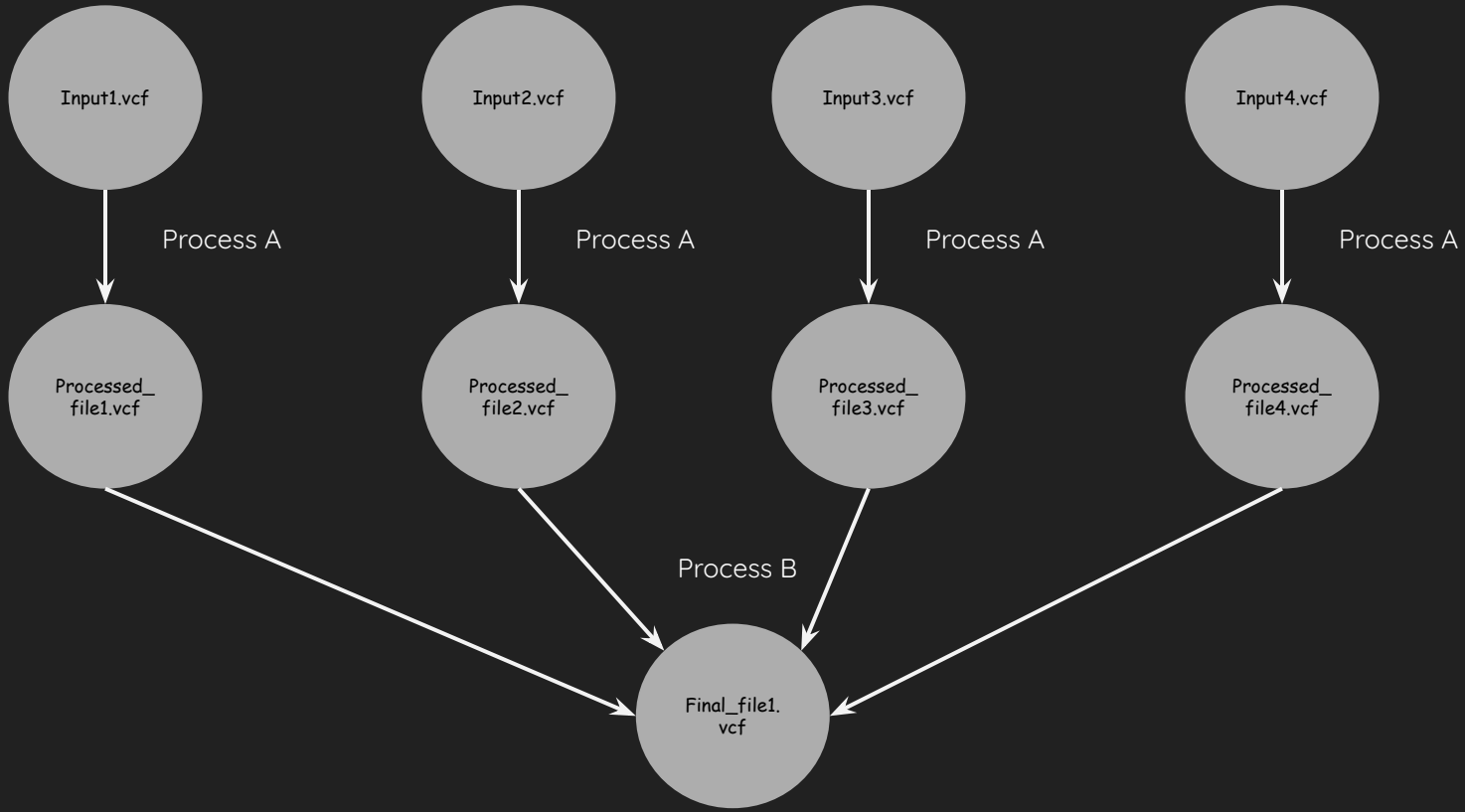
[Directed Acyclic Graph]



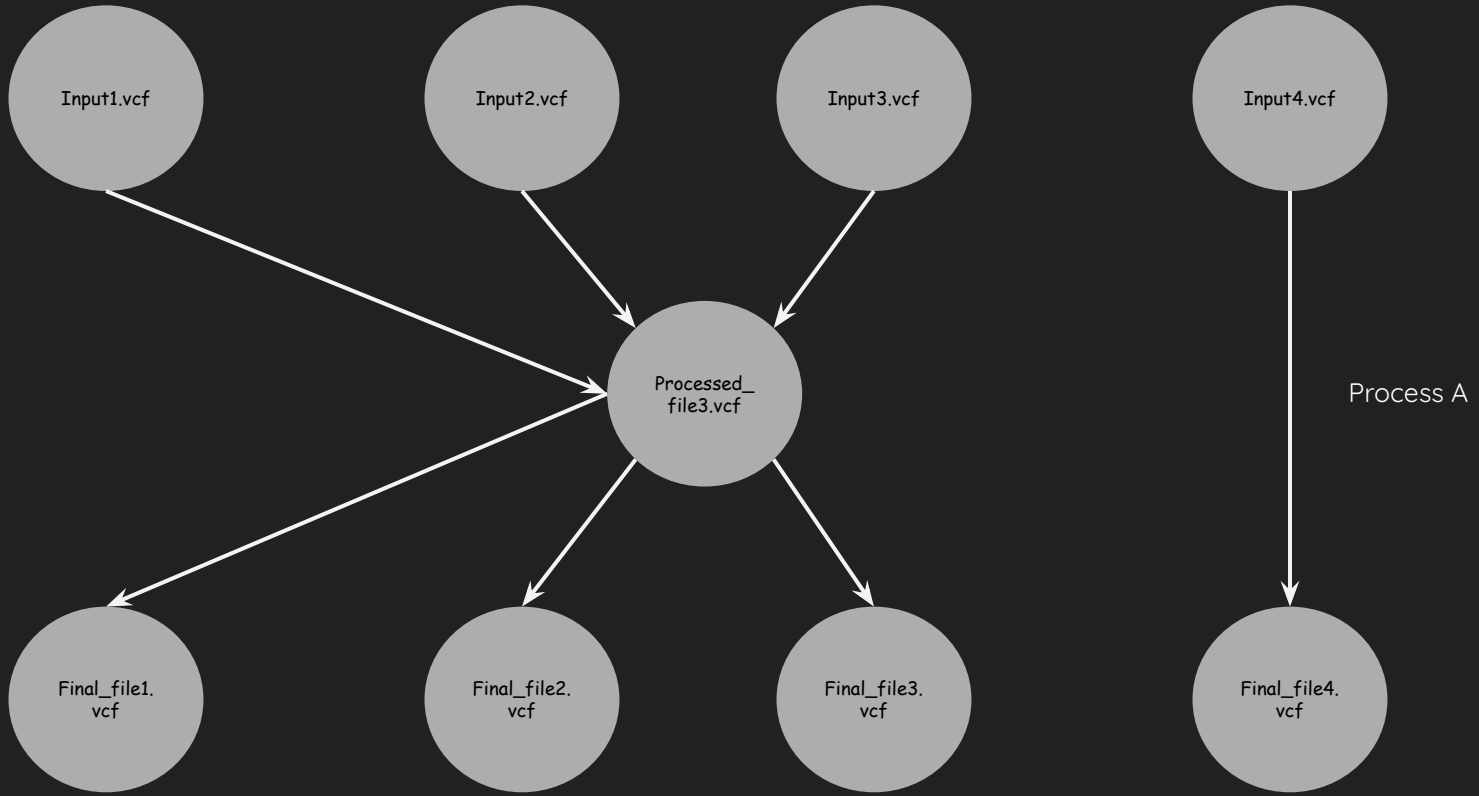
The process [Directed Acyclic Graph]



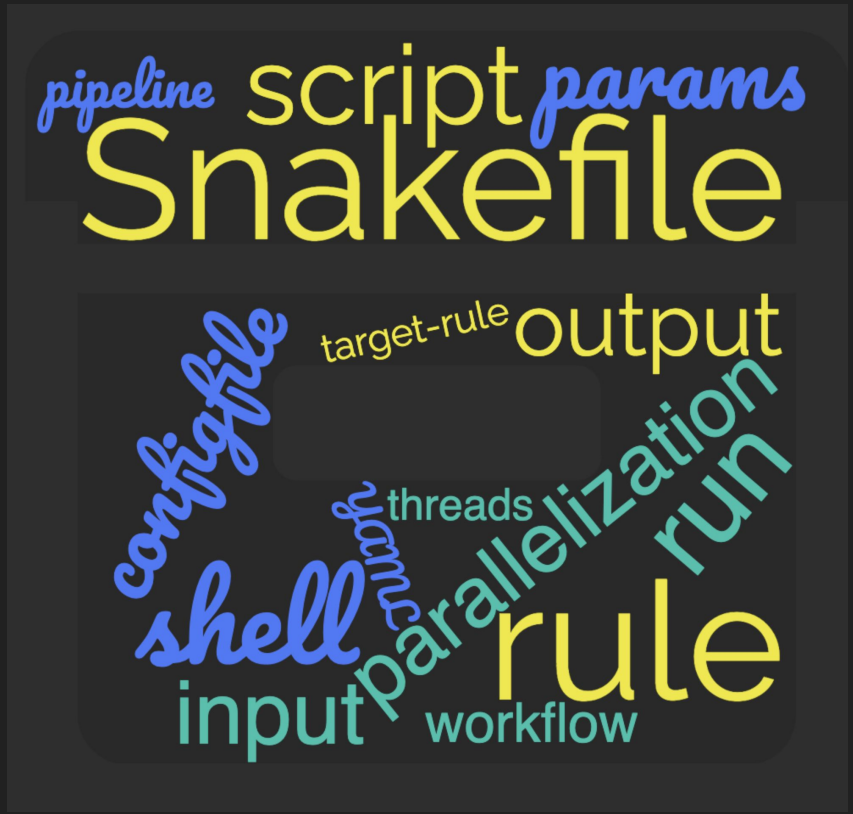
The process [Directed Acyclic Graph]

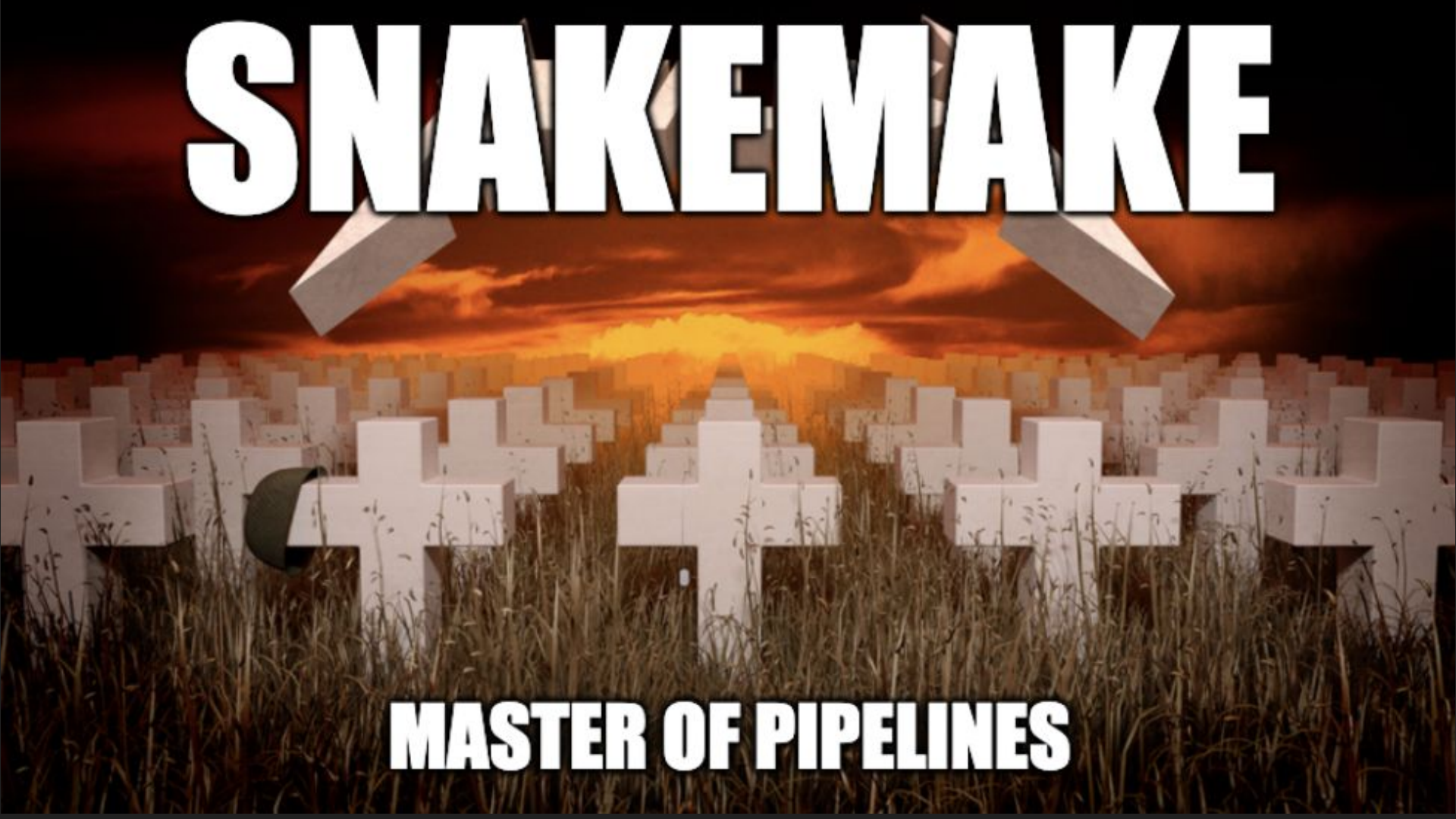


The process [Directed Acyclic Graph]



Important terms/Keywords




SNAKEMAKE
MASTER OF PIPELINES

Exam

Exercise 1

- Create a **schematic** of the imputation process.

- **No explanation** of the process.
- Be a minimalist but show the important things:
  - **Steps** (e.g. filtering)
  - **Logic** (step order)
  - **Files** necessary (use names that reflect the file contents)

- Max 3 slides
- Submit a PDF

Exercise 2

- Provide explanation for the schematic
  - Why this step?
  - **Why** this tool?
  - **What** is in this file?
  - **Which** arguments should be used? (for tools)

- Short answers will be appreciated.

- Submit a PDF (one page)
  - Font: Times New Roman
  - Size: 12

Exercise 3

- Download files.
- Build a pipeline for genotype imputation.

- You should edit the **Snakefile** only.
  - Use arguments when using tools.
  - Target file:
    - All_chromosomes.vcf
  - Make sure to keep the file structure.
  - It is **pseudo-code** no need to run it.

- Submit the Snakefile only!

Thank you