Building a reliable and reproducible scientific workflow using Nextflow

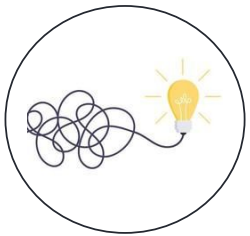
Daniel Valle Millares
(Bioinformatics Platform - CIBERINFEC)

BU-ISCIII
29-03 de octubre de 2025
1ª edición

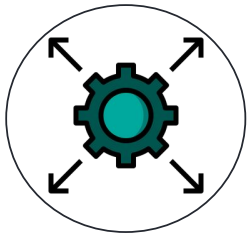
Index

1. Scripting on the cluster — Slurm: sbatch & job arrays
2. Parallelization on our cluster — OpenMP vs MPI (when to use each)
3. **From scripts to workflows — building a reproducible pipeline (Nextflow preview)**
4. Wrap-up & Q&A

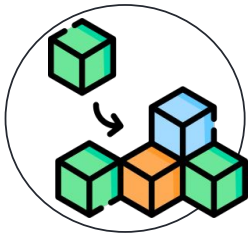
From scripts to workflows — building a reproducible pipeline (Nextflow preview)



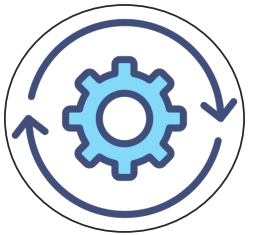
Domain Specific Language



Scalability for Massive Data Analysis



Automation & Modularity



Reproducibility Across Platforms

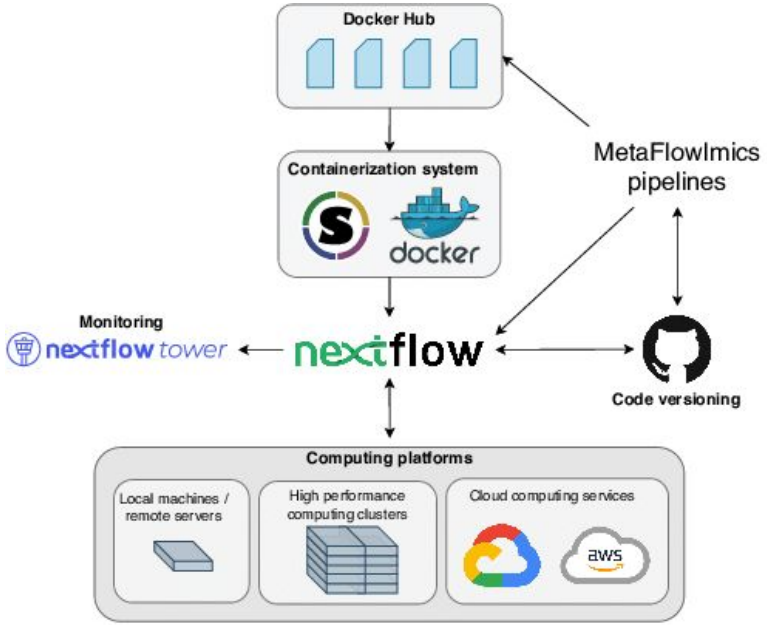
From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Why Nextflow?

- Scripts → OK for one step.
- Workflows → connect many steps

- Nextflow adds a thin abstraction layer:
  - **You**: describe the pipeline and dataflow
  - **Nextflow**: handles SLURM job configuration and submission

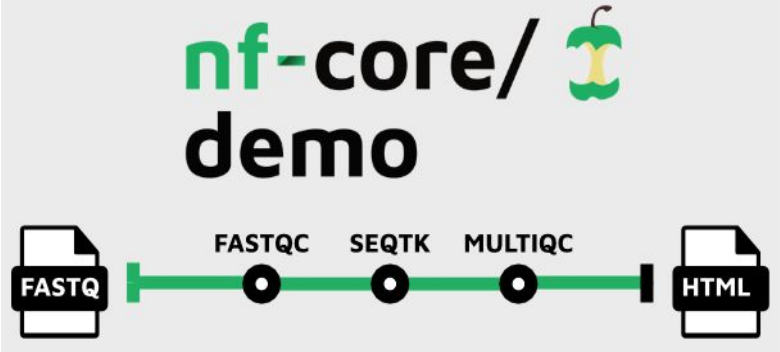
- Benefits:
  - Not need to worry about:
    - Deploy slurm jobs
    - Parallelism
    - Resume tasks
  - Portability: minimum configuration to use your workflow in a laptop, HPC, Cloud…
  - Reproducibility (containers, safe resume, versioning…)



From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Nextflow on Slurm

1) Create/Use a nextflow pipeline



https://nf-co.re/demo/dev/

From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Nextflow on Slurm

1) Create/Use a nextflow pipeline
2) Create a **nextflow.config** file

```
process {
  executor = 'slurm'                          // enviar tareas a Slurm
  queue    = 'middle_idx'                     // partición/cola por defect
  jobName  = { "$task.name-$task.hash" }      // nombre de job legible y
único
}
```

From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Nextflow on Slurm

1) Create/Use a nextflow pipeline
2) Create a nextflow.config file
3) **Create sbatch script**

```
#!/bin/bash
#SBATCH --job-name=nf_demo
#SBATCH --partition=middle_idx
#SBATCH --time=12:00:00
#SBATCH --cpus-per-task=2                # Recursos SOLO para el
controlador de Nextflow
#SBATCH --mem=2G
#SBATCH --output=logs/%x-%j.out
#SBATCH --error=logs/%x-%j.err

# Carga las dependencias para ejecutar Nextflow
module purge
module load Nextflow/24.04.2
module load singularity/3.7.1

mkdir -p 01-nextflow-demo-results
# Ejecuta nf-core/demo (workflow preparado)
# Importante: le indicamos que lea el archivo de configuración
nextflow run nf-core/demo \
  -profile test,singularity \
  -c nextflow.config \
  --outdir 01-nextflow-demo-results \
  -resume
```

From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Nextflow on Slurm

Monitor execution -- stdOUT

```
N E X T F L O W  ~  version 24.02.0
Launching `main.nf` [elegant_burnell] DSL2 - revision: a1b2c3d4f5
executor >  slurm (6)
[3a/5f1c2b] process > DOWNLOAD_TEST_DATA          [100%] 1 of 1 ✓
[b9/904773] process > FASTQC (sample_1)           [100%] 1 of 1 ✓
[6e/12ab34] process > FASTQC (sample_2)           [100%] 1 of 1 ✓
[2d/aa77cc] process > ECHO_HELLO (1)              [100%] 4 of 4 ✓
[7f/33dd44] process > MULTIQC                      [100%] 1 of 1 ✓

Completed at: 2025-09-01 11:42:17
Duration    : 2m 12s
CPU hours   : 0.1
Succeeded   : 8
```

From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Nextflow on Slurm

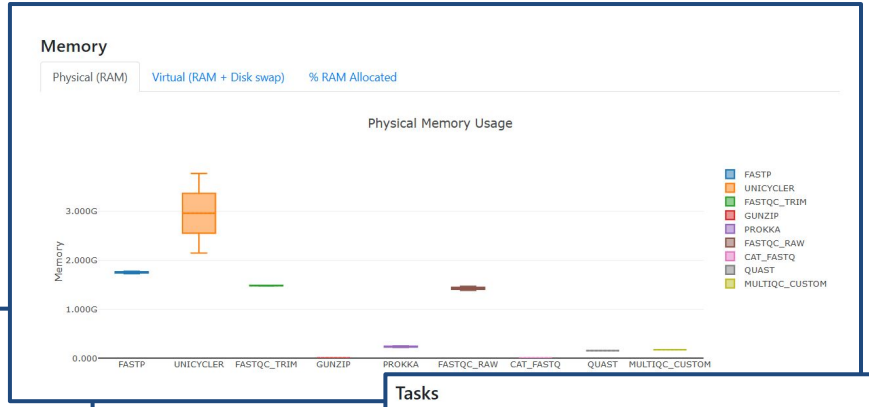
Monitor execution -- squeue -me

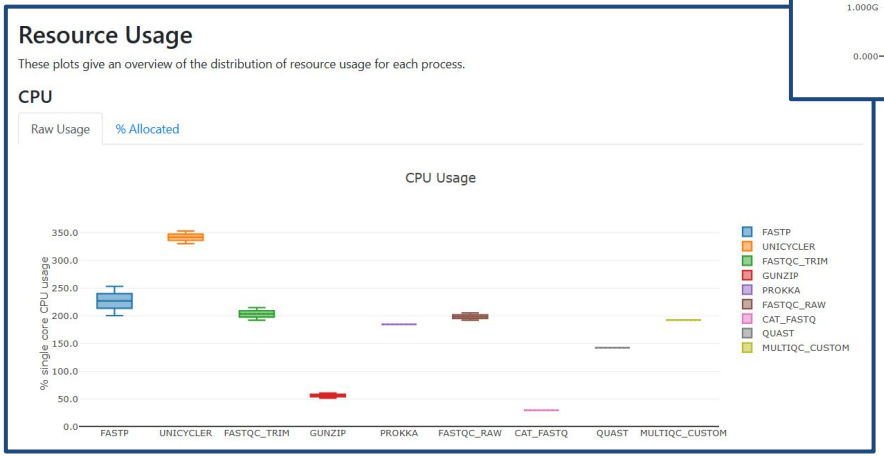
```
JOBID PARTITION  NAME                          USER    ST  TIME  NODES NODELIST(REASON)
     9101201 short_idx  nf_demo                       daniel   R  0:01      1 ideafix03
     9101202 short_idx  FASTQC - a1b2c3d4            daniel   R  0:00      1 ideafix05
     9101203 short_idx  FASTQC - e5f6a7b8            daniel   R  0:00      1 ideafix08
     9101204 short_idx  FASTQC - 9c0d1e2f            daniel  PD  0:00      1 (Resources)
     9101205 short_idx  MULTIQC - 55aa33ff           daniel  PD  0:00      1 (Dependency)
     9101206 short_idx  ECHO_HELLO - deadbeef        daniel  CG  0:02      1 ideafix02
     9101207 short_idx  CLEANUP - f00dbabe           daniel  PD  0:00      1 (Dependency)
     9101208 short_idx  SUMMARY - cafed00d           daniel  PD  0:00      1 (Dependency)
```

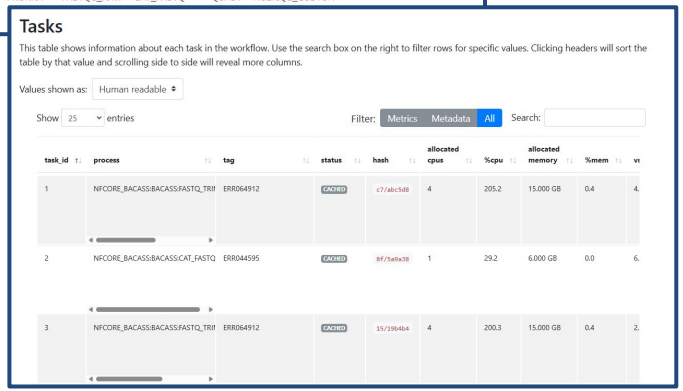
From scripts to workflows — building a reproducible pipeline (Nextflow preview)

Nextflow on Slurm

Monitor execution -- Fancy mode







Thank you for your attention

Questions?

HANDS-ON