wellcome
connecting
science

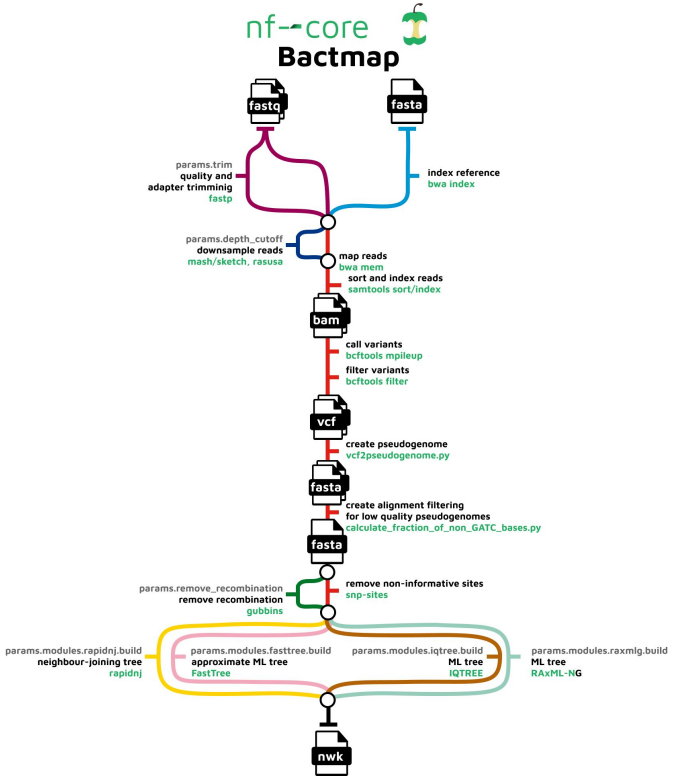
Bioinformatics Essentials

Dr. Jacqui Keane

Overview

- Bioinformatics Software

- Bioinformatics Pipelines



Learning Outcomes

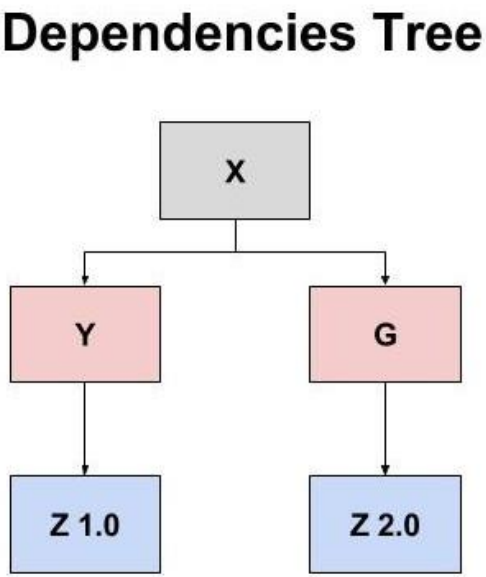
- You can expect to be able to:
  - Describe what a software package manager is and why are useful
  - Install a software package manager (conda/mamba)
  - Install bioinformatics software with the conda package manager
  - Describe what a workflow manager is and why useful
  - Install Nextflow and run a Nextflow pipeline

Overview

- Bioinformatics Software

Bioinformatics Software

- Many bioinformatics software
- Complex to install and manage
  - - e.g. conflicting dependencies
- Use software package manager

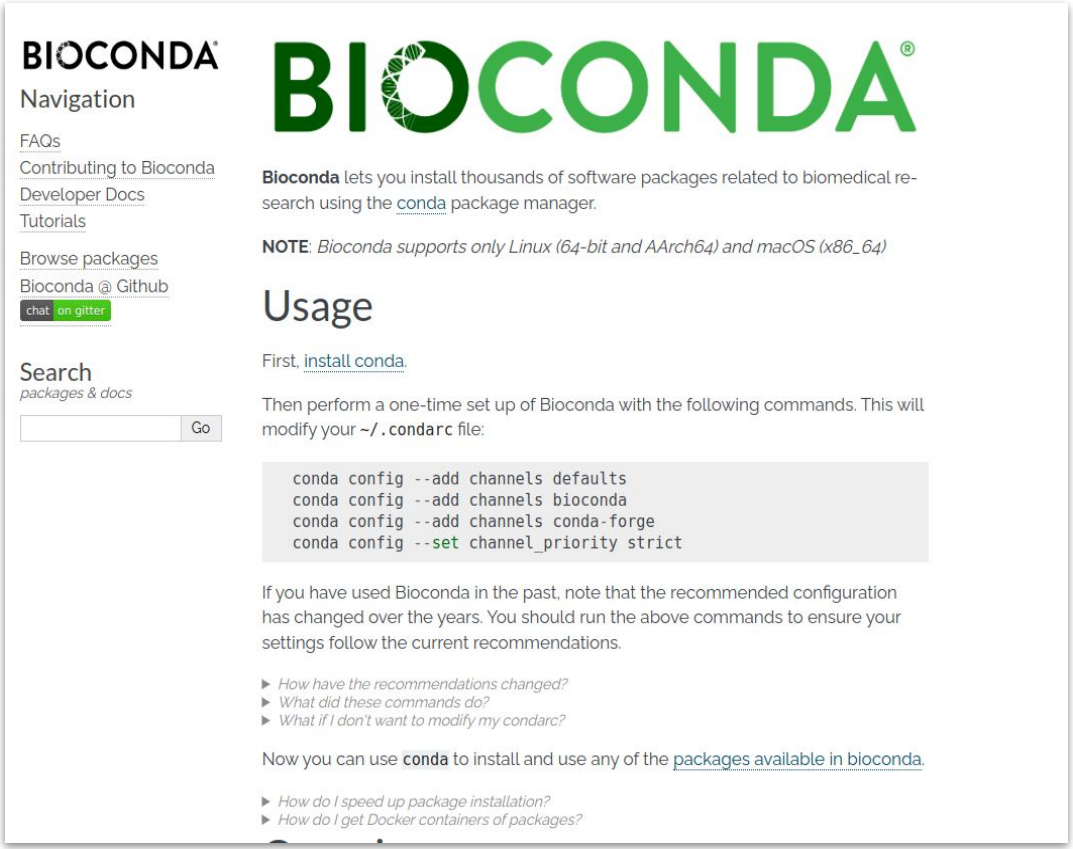


Conda

- Automates the process of installing software and their dependencies
- Software applications available via channels
- A channel online locations of where packages are stored

e.g. conda-forge, r, bioconda

Bioconda channel



BIOCONDA

Navigation

FAQs
Contributing to Bioconda
Developer Docs
Tutorials

Browse packages
Bioconda @ Github
chat on gitter

Search
packages & docs

Go

BIOCONDA®

Bioconda lets you install thousands of software packages related to biomedical research using the conda package manager.

NOTE: *Bioconda supports only Linux (64-bit and AArch64) and macOS (x86_64)*

Usage

First, install conda.

Then perform a one-time set up of Bioconda with the following commands. This will modify your `~/.condarc` file:

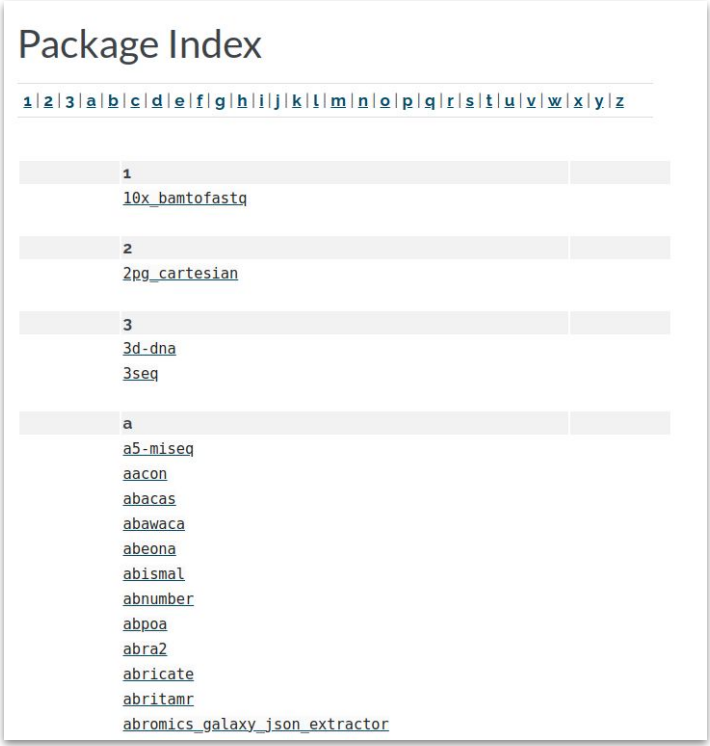
```
conda config --add channels defaults
conda config --add channels bioconda
conda config --add channels conda-forge
conda config --set channel_priority strict
```

If you have used Bioconda in the past, note that the recommended configuration has changed over the years. You should run the above commands to ensure your settings follow the current recommendations.

▸ *How have the recommendations changed?*
▸ *What did these commands do?*
▸ *What if I don't want to modify my condarc?*

Now you can use `conda` to install and use any of the packages available in bioconda.

▸ *How do I speed up package installation?*
▸ *How do I get Docker containers of packages?*



Package Index

1 | 2 | 3 | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z

1
10x_bamtofastq

2
2pg_cartesian

3
3d-dna
3seq

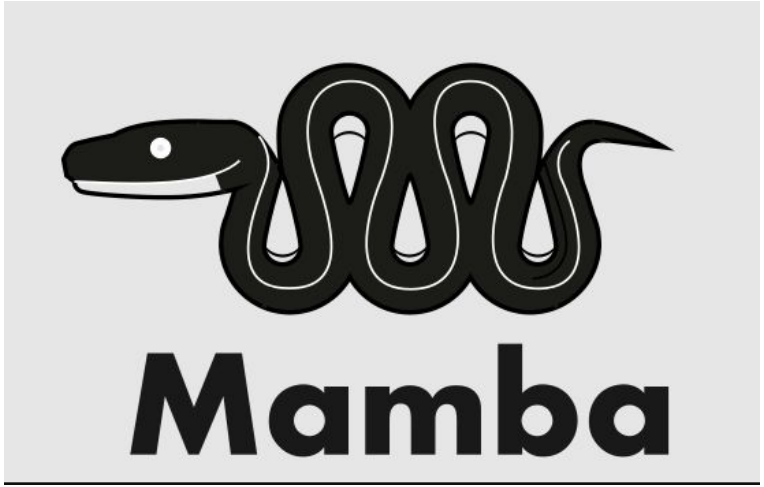
a
a5-miseq
aacon
abacas
abawaca
abeona
abismal
abnumber
abpoa
abra2
abricate
abritamr
abromics_galaxy_json_extractor

Conda environments

- A separate space (box) where you install specific versions of software/set of softwares
- Prevent conflicts between different softwares by allowing you to manage dependencies separately
- Create, activate, and switch between environments easily

Mamba

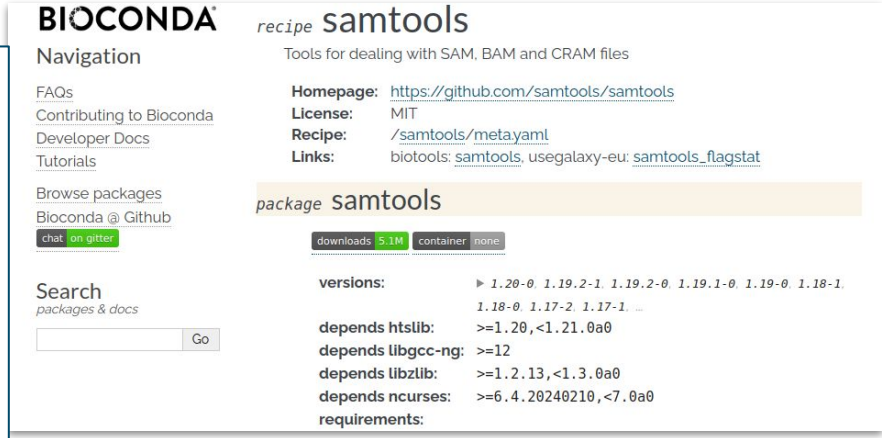
- A software package manager similar to Conda
- Faster and uses less memory, good for installing software faster and using less computer resources
- Fully compatible with conda



Demo

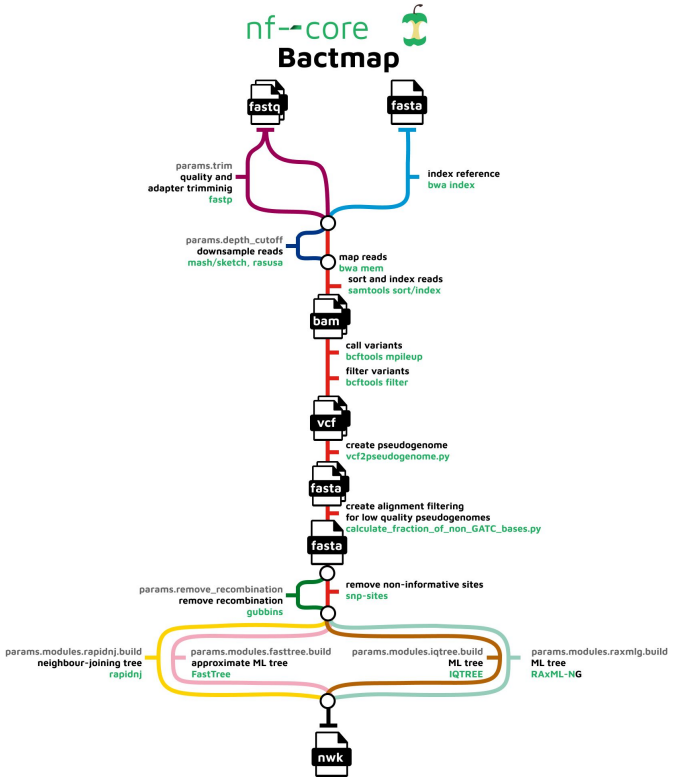
- Bioinformatics Software
  - Install conda from: https://docs.conda.io/projects/conda/en/latest/user-guide/install/linux.html
  - Install mamba from: https://mamba.readthedocs.io/en/latest/installation/mamba-installation.html
  - Use conda to install samtools

```
$ conda info --envs
$ conda list
$ samtools --version
$ conda create -n samtools-1.21 samtools=1.21.0
$ conda activate samtools-1.21
$ conda info -envs
$ conda list
$ samtools --version
$ conda deactivate
```



Overview

- Bioinformatics Pipelines

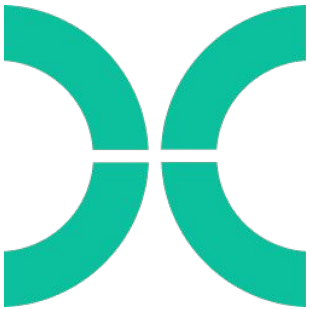


Bioinformatics Pipelines

- Automate analysis using a pipeline
  - e.g. bash script
- Re-write the pipeline for different platforms
- Use workflow/pipeline management system

Nextflow

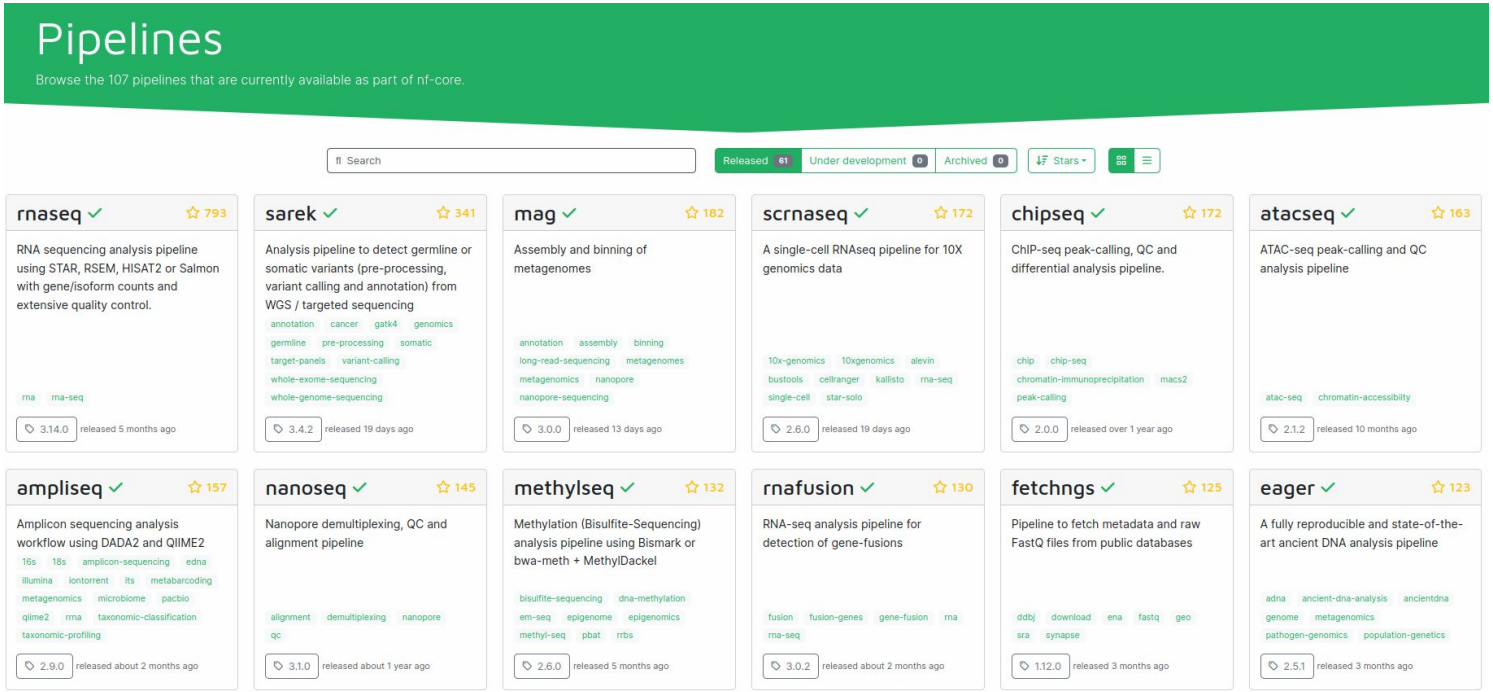
- Use it to write and run data-intensive workflows/pipelines
- Nextflow consists of a
  - ***language*** to write pipelines
  - ***engine*** to run pipelines
- Write a pipeline once and run everywhere!



nextflow

nf-core

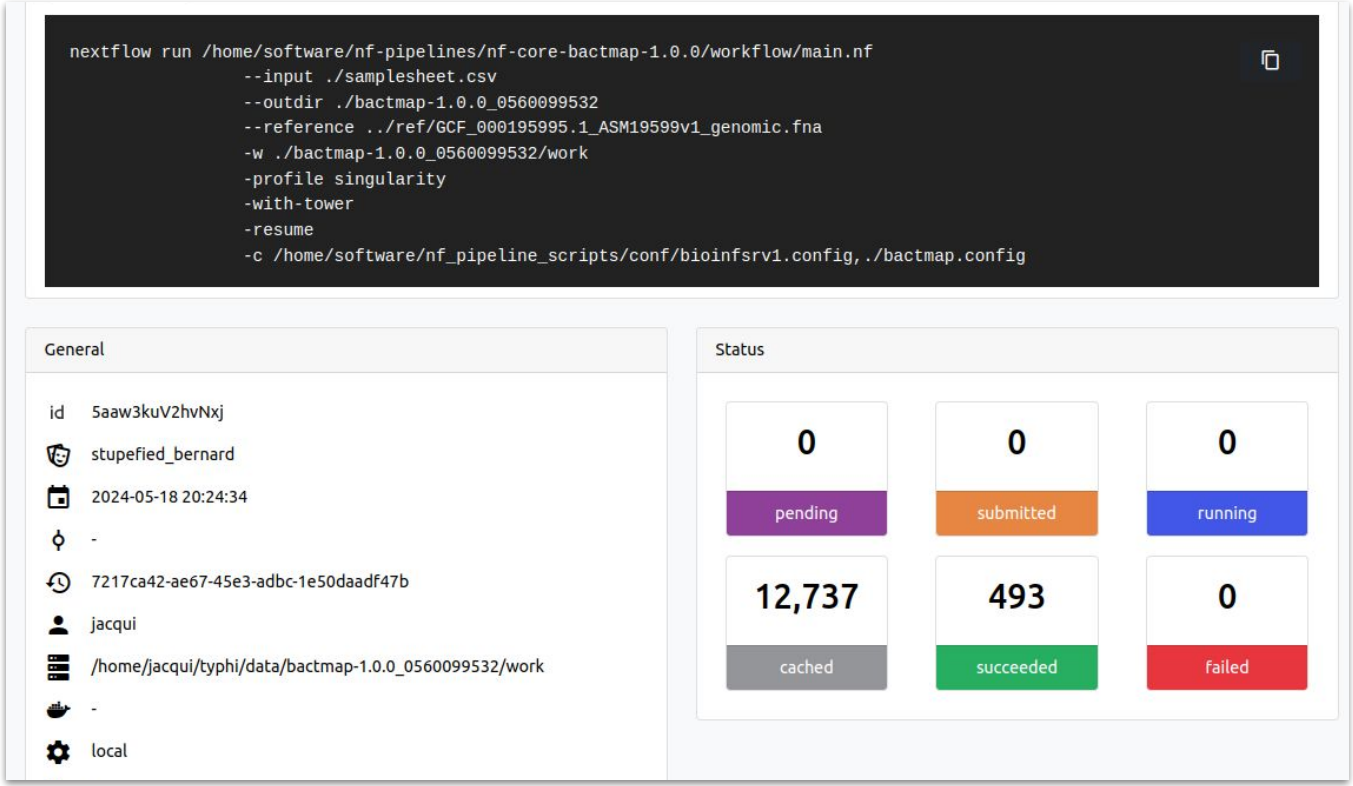
- Community effort to collect a set of best practice analysis pipelines built using Nextflow



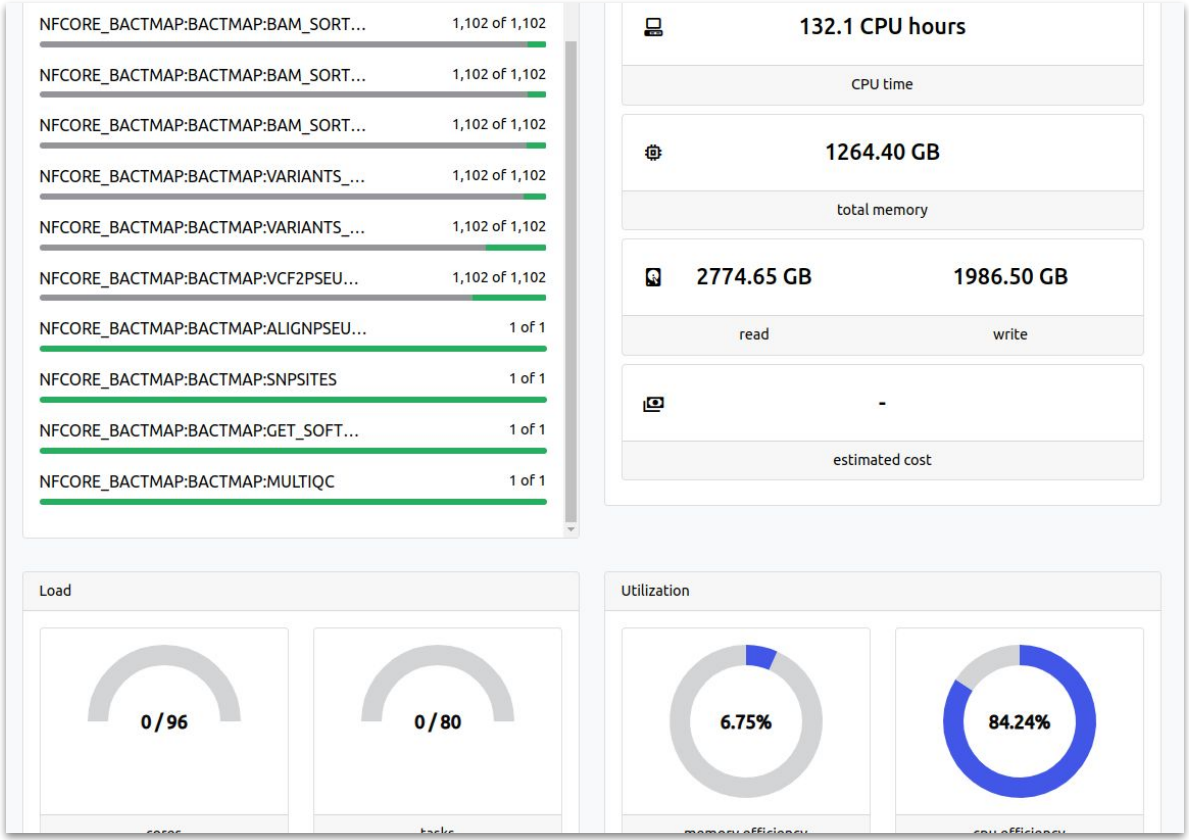
Sequera Platform (Nextflow Tower)

- Manages and tracks the executions of Nextflow pipelines

```
nextflow run /home/software/nf-pipelines/nf-core-bactmap-1.0.0/workflow/main.nf
                --input ./samplesheet.csv
                --outdir ./bactmap-1.0.0_0560099532
                --reference ../ref/GCF_000195995.1_ASM19599v1_genomic.fna
                -w ./bactmap-1.0.0_0560099532/work
                -profile singularity
                -with-tower
                -resume
                -c /home/software/nf_pipeline_scripts/conf/bioinfsrv1.config,./bactmap.config
```



Sequera Platform (Nextflow Tower)

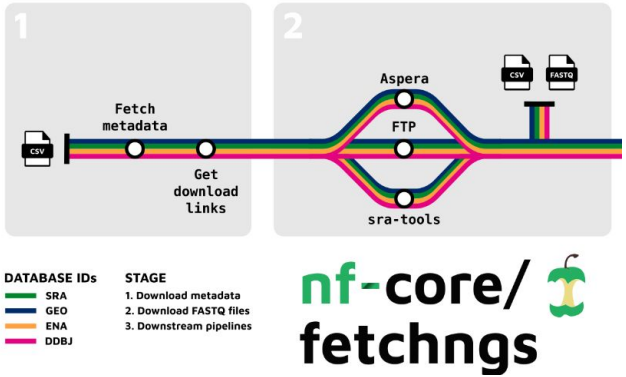


Demo

- Bioinformatics Pipelines
  - Install Nextflow and nf-core
  - Install nf-core pipeline

    - fetchngs
  - Run the fetchngs pipeline

```
$ conda create -n nf-pipelines
$ conda activate nf-pipelines
$ conda install nextflow nf-core
$ nf-core list
$ nf-core list | grep -i rna
```




```
$ wget
https://github.com/sylabs/singularity/releases/download/v4.1.2/singularity-Ce_4.1.2-jammy_amd64.deb
$ sudo dpkg -i singularity-ce_4.1.2-jammy_amd64.deb
$ nextflow run nf-core/fetchngs -profile singularity –input ids.csv –outdir .
```

Questions?