The IFB Core Cluster Infrastructure

FAIR Bioinfo 2020

Gildas Le Corguillé & Julien Seiler
IFB Core Cluster taskforce

TP1 - Snakemake over SLURM

TP1 - Snakemake over SLURM

Exercice 1: connect to the cluster through SSH

- Connect to the IFB Core Cluster using your SSH client
- Find the name of your current working directory

TP1 - Snakemake over SLURM

Exercice 1: connect to the cluster through SSH

- Connect to the IFB Core Cluster using your SSH client

Use your favorite SSH client to connect the cluster:

```
ssh <username>@core.cluster.france-bioinformatique.fr
```

Replace `<username>` with your username.

Beware: at the password prompt, the characters you type are not printed on the screen, for obvious security reasons.

TP1 - Snakemake over SLURM

Exercice 1: connect to the cluster through SSH

- Connect to the IFB Core Cluster using your SSH client
- Find the name of your current working directory

```
pwd
```

The result should look like this :

```
/shared/home/jseiler
```

TP1 - Snakemake over SLURM

Exercice 2: Get your environment ready

- Create a folder for your workflow
- Upload your workflow
- Download your input data
- Load the snakemake module and all required tools

TP1 - Snakemake over SLURM

Exercice 2: Get your environment ready

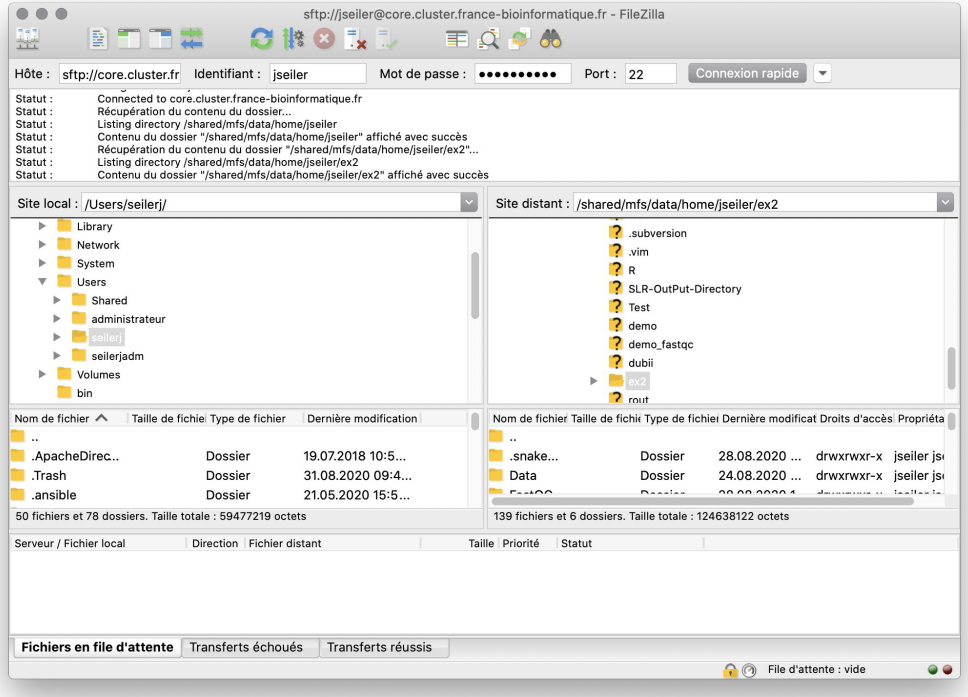
- Create a folder for your workflow

```
$ mkdir fairbio
```

TP1 - Snakemake over SLURM

Exercice 2: Get your environment ready

- Create a folder for your workflow
- Upload your workflow



Use **FileZilla** :
Host: `core.cluster.france-bioformatique.fr`
Login: `<your login>`
Password: `<password>`
Port: 22

Click Quick connection
Move your smk and yml files from your local computer to the fairbio forlder on the cluster

TP1 - Snakemake over SLURM

Exercice 2: Get your environment ready

- Create a folder for your workflow
- Upload your workflow
- Download your input data

The data used for the snakemake tutorial are available on Zenodo :

Go to zenodo.org

Search for 

Copy/paste the download link 

TP1 - Snakemake over SLURM

Exercice 2: Get your environment ready

- Create a folder for your workflow
- Upload your workflow
- Download your input data

Download the data archive using wget :

```
$ wget -O FAIR_Bioinfo_data.tar.gz
https://zenodo.org/record/3997237/files/FAIR_Bioinfo_data.tar.gz?download=1
```

Untar the archive :

```
$ tar -xvzf FAIR_Bioinfo_data.tar.gz
```

TP1 - Snakemake over SLURM

Exercice 2: Get your environment ready

- Create a folder for your workflow
- Upload your workflow
- Download your input data
- Load the snakemake module and all required tools

```
$ module load snakemake fastqc bowtie2 samtools subread
```

TP1 - Snakemake over SLURM

Exercice 3: Run snakemake

- Run your workflow using `srun`
- Run your workflow using --cluster mode
- Run your workflow using --drmaa mode

TP1 - Snakemake over SLURM

Exercice 3: Run snakemake

- Run your workflow using `srun`

```
srun --cpus=12 snakemake --jobs $SLURM_CPUS_PER_TASK -s ex2_o6.smk
--configfile ex2_o1.yml
```

Drawbacks : Wait for 12 cores to be available on the same CPU node

TP1 - Snakemake over SLURM

Exercice 3: Run snakemake

- Run your workflow using `srun`
- Run your workflow using --cluster mode

```
rm -r Result Tmp FastQC Logs

snakemake --cluster "sbatch" --jobs=20 -s ex2_o6.smk --configfile ex2_o1.yml
```

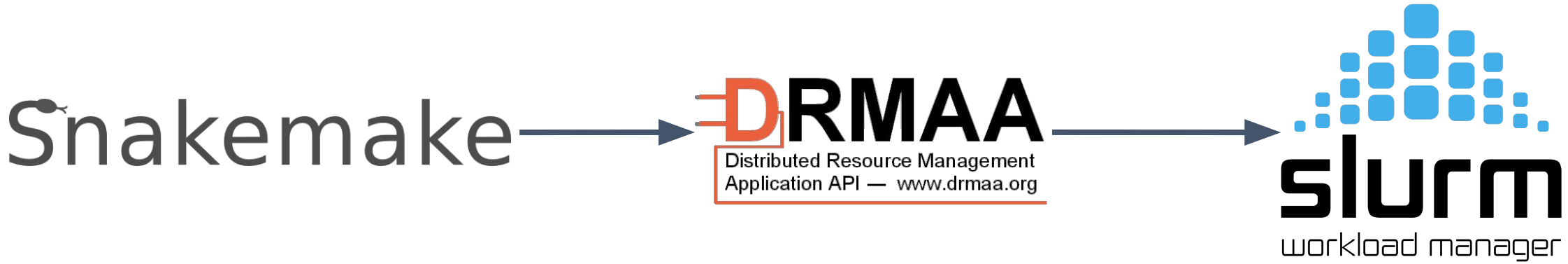
Drawbacks : no control on workflow execution (you can’t stop it)

TP1 - Snakemake over SLURM

Exercice 3: Run snakemake

- Run your workflow using `srun`
- Run your workflow using --cluster mode
- Run your workflow using --drmaa mode

Distributed **R**esource **M**anagement **A**pplication **API**



TP1 - Snakemake over SLURM

Exercice 3: Run snakemake

- Run your workflow using `srun`
- Run your workflow using --cluster mode
- Run your workflow using --drmaa mode

```
rm -r Result Tmp FastQC Logs

snakemake --drmaa --jobs=20 -s ex2_o6.smk --configfile ex2_o1.yml

WorkflowError:
Error loading drmaa support:
Could not find drmaa library.  Please specify its full path using
the environment variable DRMAA_LIBRARY_PATH
```

TP1 - Snakemake over SLURM

Exercice 3: Run snakemake

- Run your workflow using `srun`
- Run your workflow using --cluster mode
- Run your workflow using --drmaa mode

```
module load slurm-drmaa

echo $DRMAA_LIBRARY_PATH

snakemake --drmaa --jobs=20 -s ex2_o6.smk --configfile ex2_o1.yml
```

Useful links

Request an account:
https://my.cluster.france-bioinformatique.fr

Community support:
https://community.france-bioinformatique.fr/

Learn SLURM in 5 minutes:
https://asciinema.org/a/275233

IFB Core Cluster Documentation
https://ifb-elixirfr.gitlab.io/cluster/doc/