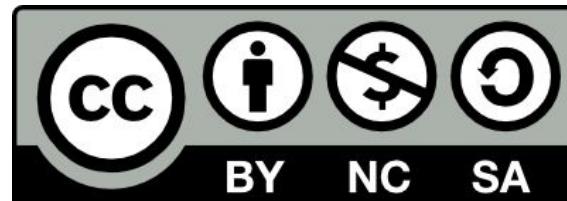




# i-Trop cluster Slurm Initiation

[www.southgreen.fr](http://www.southgreen.fr)

<https://southgreenplatform.github.io/trainings>





Julie ORJUELA-  
BOUNIOL<sup>1</sup>, IE  
Bioinformaticienne  
25%

Ndomassi TANDO, IE  
Administrateur système  
100%  
Animateur plateau

Aurore COMTE, IE  
Bioinformaticienne  
20%

Valérie NOEL, TCS  
Bioinformaticienne  
25%

Bruno GRANOUILLAG<sup>3</sup>, IE  
Systèmes d'information  
20%

Christine TRANCHANT-  
DUBREUIL, IE  
Bioinformaticienne  
20%

Alexis DEREEPER<sup>2</sup>, IE  
Bioinformaticien  
20%

**Provisioning of  
softwares and HPC  
ressources**

**Analysis software  
and IS development**

**Plateau bioinformatique**

**Help and support  
to IRD teams**

**Training to north  
and South**

- Request forms:

<https://itrop-glpi.ird.fr/plugins/formcreator/front/formlist.php>

- Accounts
- Softwares
- Projects

- Incidents: contact [bioinfo@ird.fr](mailto:bioinfo@ird.fr)



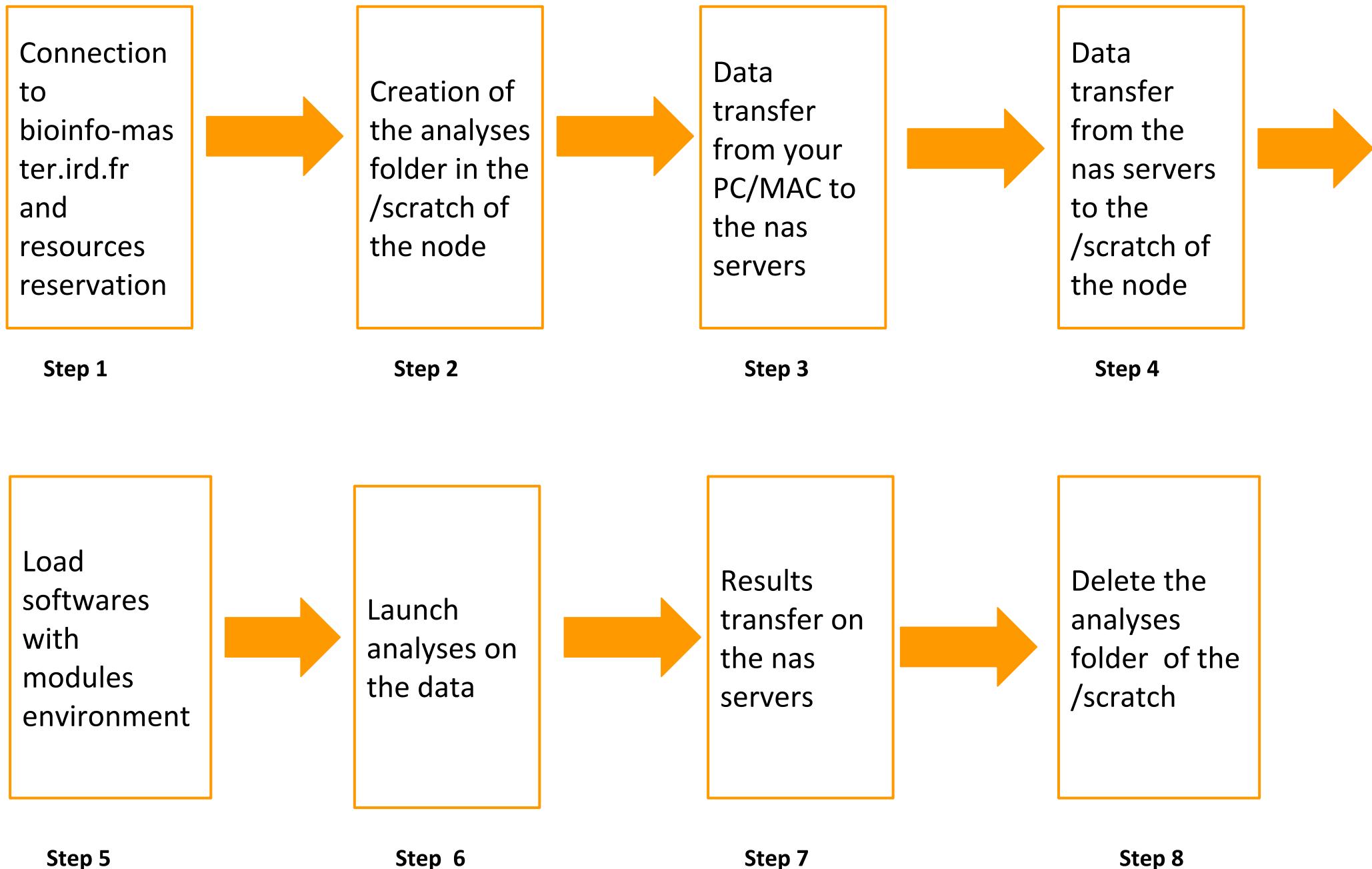
- Howtos:

<https://southgreenplatform.github.io/trainings/hpc/hpcHowto/>

- Slurm Tutorials:

<https://southgreenplatform.github.io/tutorials//cluster-itrop/Slurm/>

# Analyses steps of the cluster





# Practice

Step 1: Connection, sinfo

1

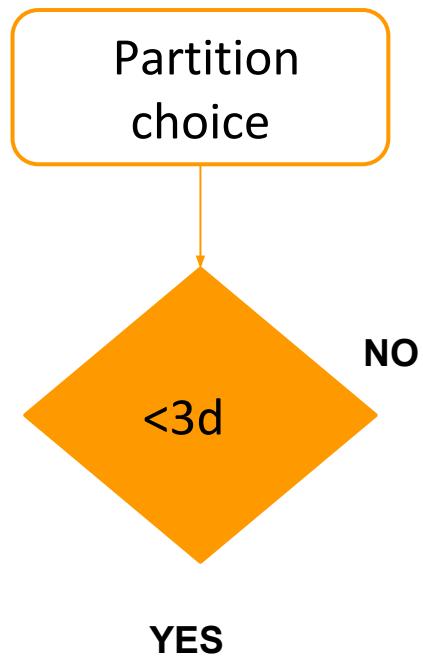
Go to the [Practice 1 and 2](#) of github

# Partitions

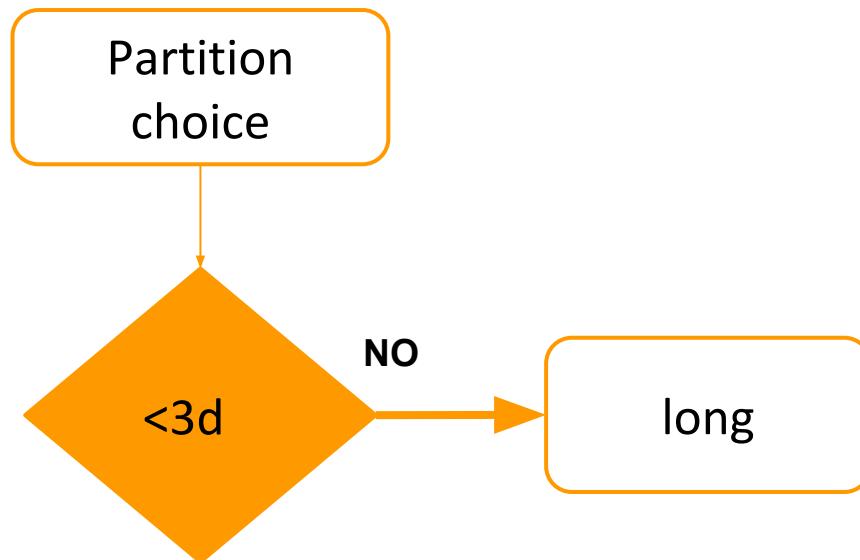
Partitions	Use	RAM on nodes	Core on nodes
short	Short Jobs < 1 day (higher priority, interactive jobs)	48 to 64 GB	12 cores
normal	Short Jobs max 3 days	64 Go to 96 GB	12 to 24 cores
long	45 days >long jobs > 3 days	48 GB	12 to 24 cores
highmem	Jobs with more memory needs	144 GB	12 to 24 cores
supermem	Jobs with much more memory needs	1TB	40 cores
gpu	Need for analyses on GPU cores	192GB	24 cpus and 8 GPUs cores

\*Request to do with arguments

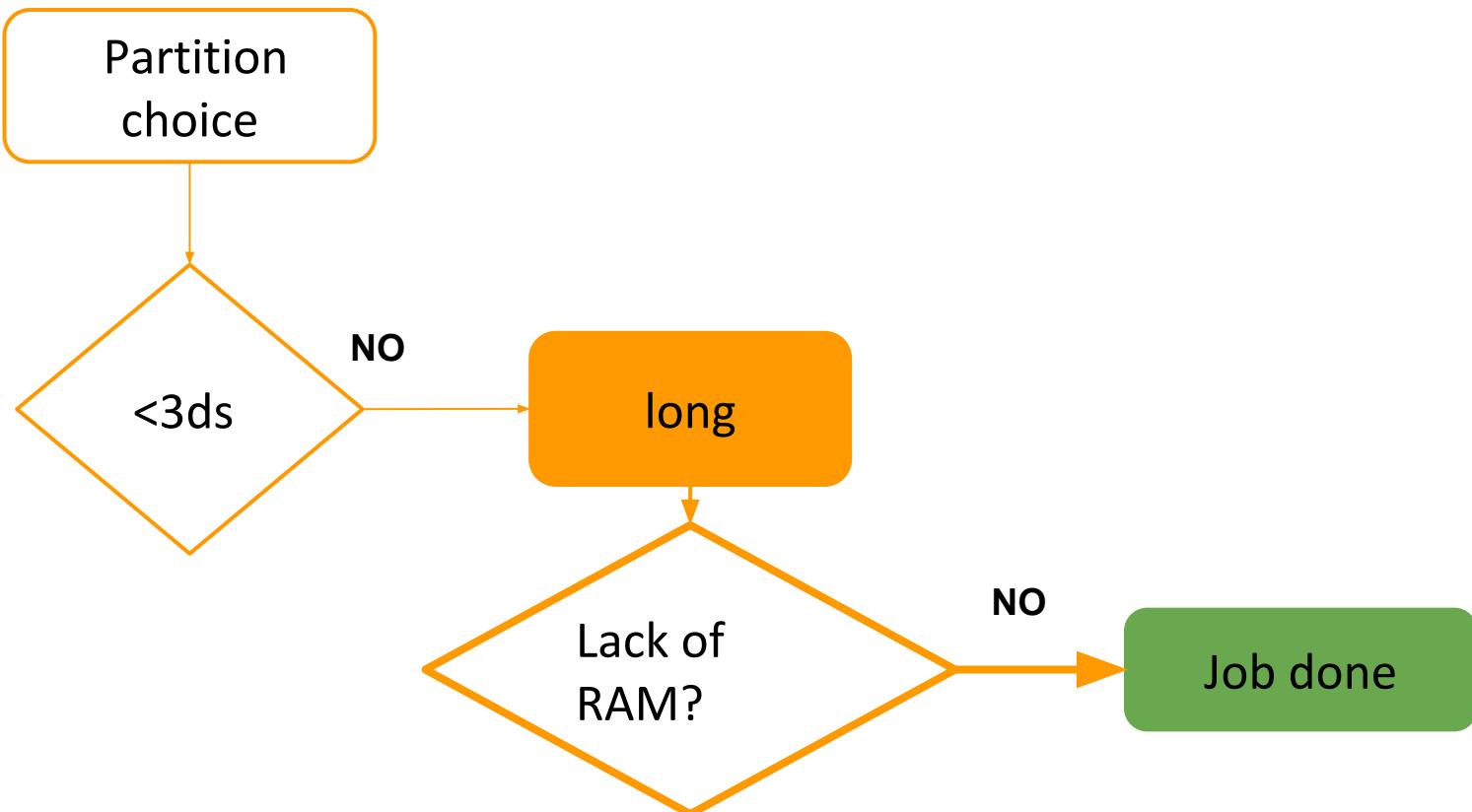
# How do I choose the partition?



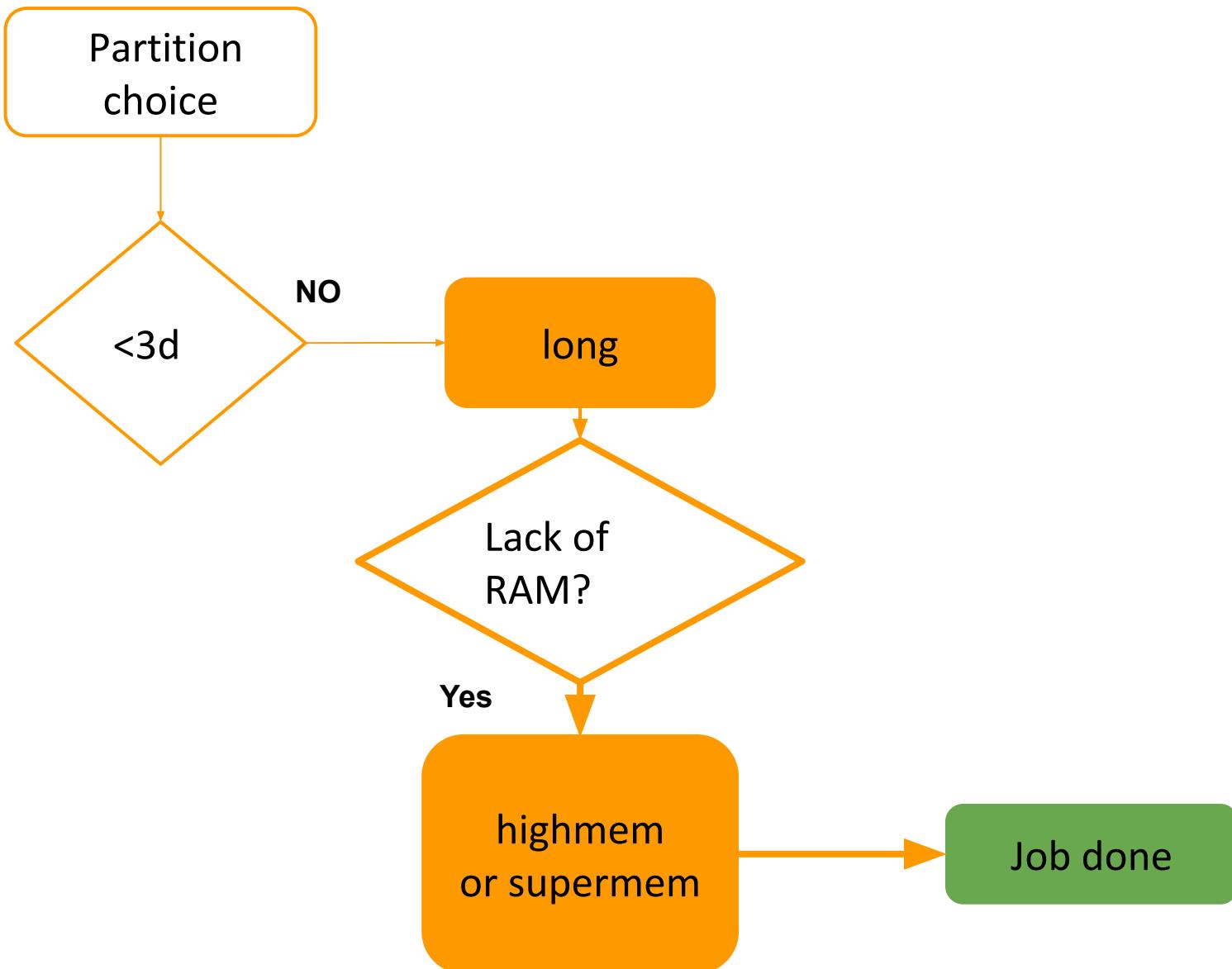
# How do I choose the partition?



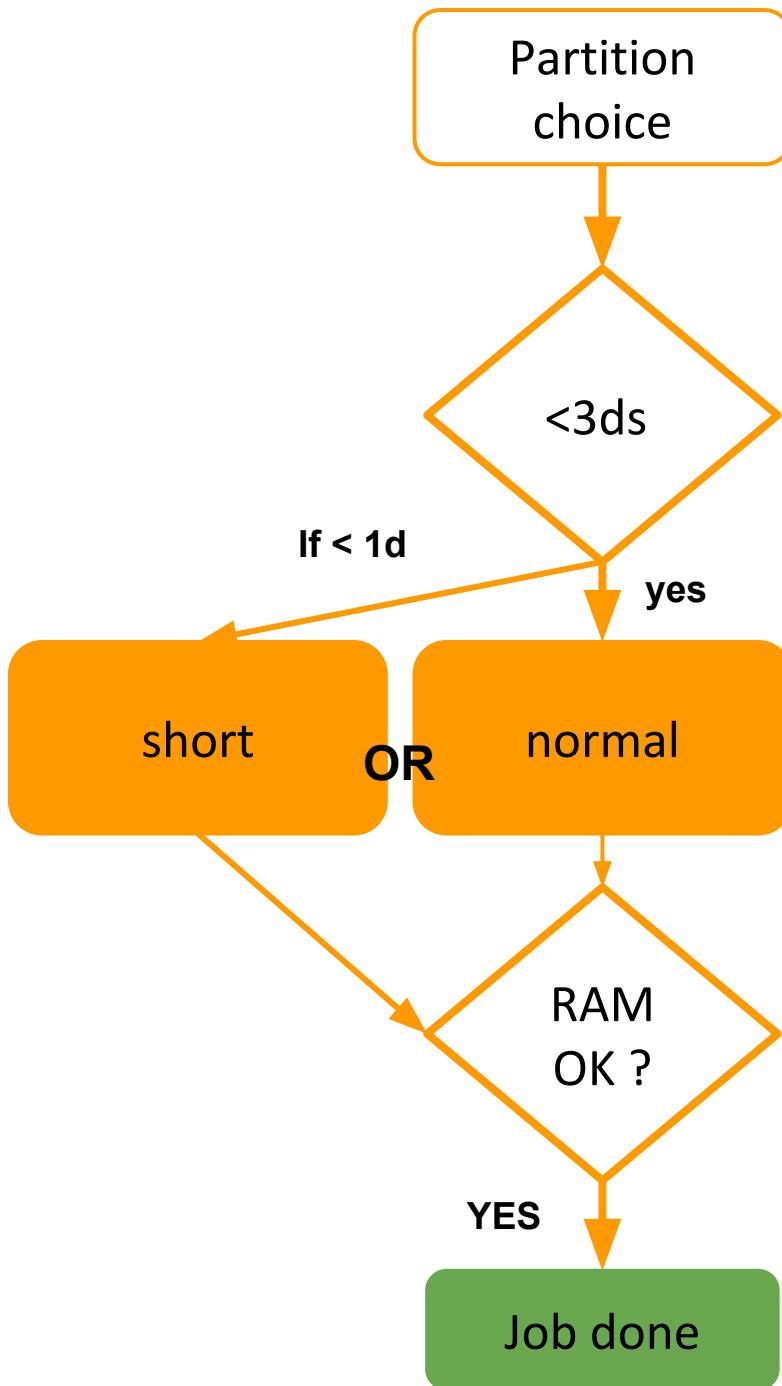
# How do I choose the partition?



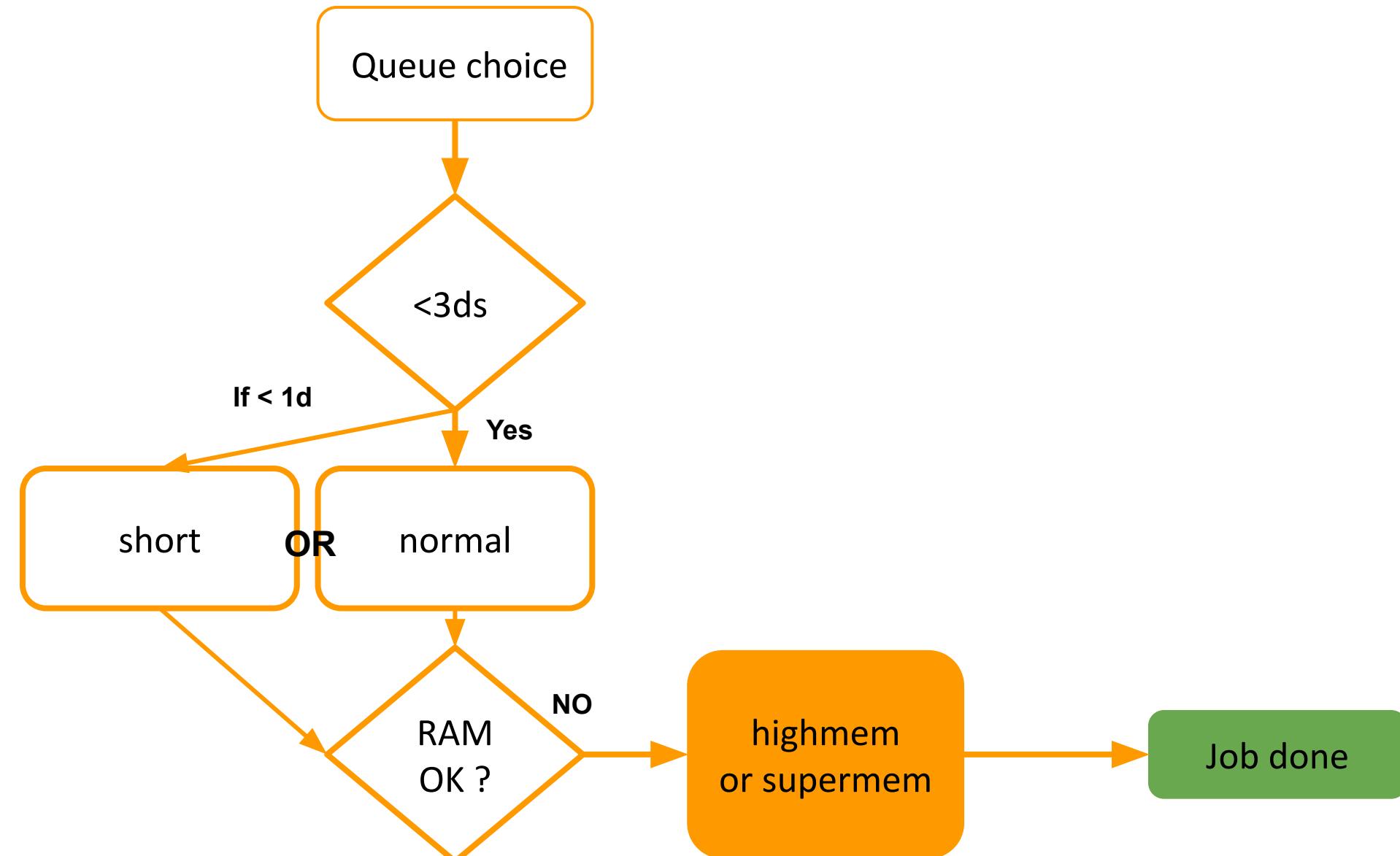
# How do I choose the partition?



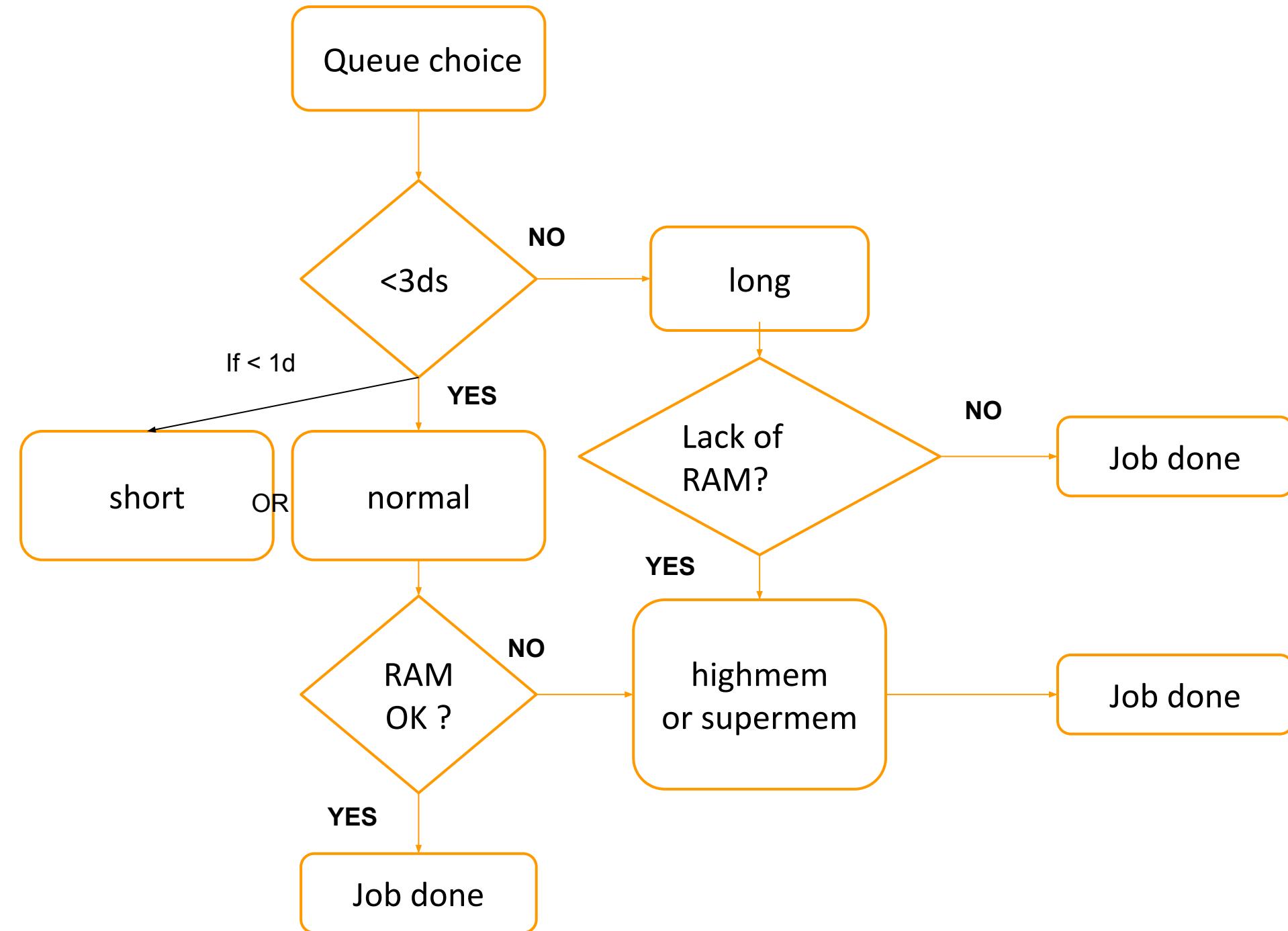
# How do I choose the partition?



# How do I choose the queue?



# How do I choose the queue?

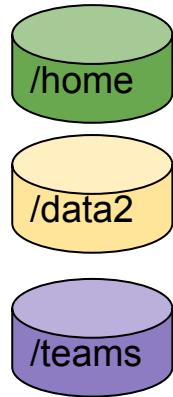


# Particular case : gpu partition

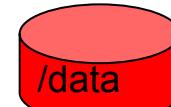
- Partition to work on GPUs processors : basecalling, MiniOn etc..
- Restricted access to gpu\_account group
- Request access with arguments to do here:

<https://itrop-glpi.ird.fr/plugins/formcreator/front/formlist.php>

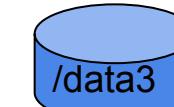
# cluster i-Trop disk partitions



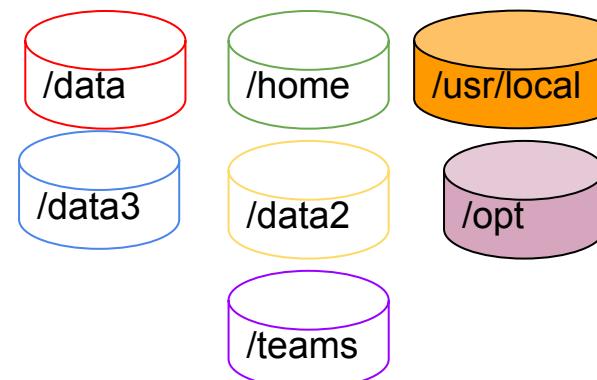
**bioinfo-nas.ird.fr**



**bioinfo-nas2.ird.fr**

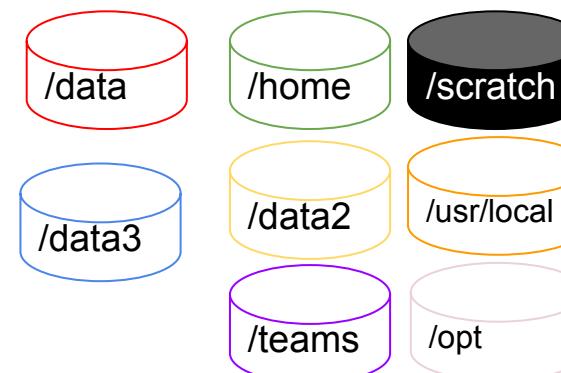


**bioinfo-nas3.ird.fr**



**bioinfo-master.ird.fr**

**Virtual links to the other servers partitions**



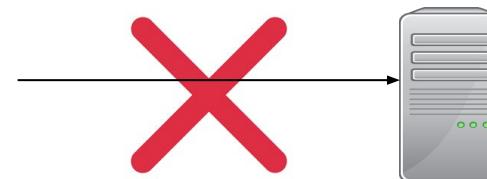
**25 nodes**



# Data transfer on i-Trop cluster



PC/MAC



direct transfer  
via filezilla  
forbidden

**bioinfo-master.ird.fr**  
**91.203.34.148**



# Practice

**Step 3 and 4: scp to nodes**

4

*Go to the [Practice4](#) of the github*

- Allow to choose the version of software you want to use
- 2 types of softwares :
  - bioinfo : includes all the bioinformatics softwares (example BEAST)
  - system : includes all the system softwares(example JAVA)
- Overcome the environment variables

- 5 types of commands :
  - See the available modules :  
`module avail`
  - Obtain infos on a particular module:  
`module whatis + module name`
  - Load a module :  
`module load + modulename`
  - List the loaded module :  
`module list`
  - Unload a module :  
`module unload + modulename`
  - Unload all the modules :  
`Module purge`



# Practice

## Step 5: module environment

5

*Go to the [Practice5](#) of the github*

# Launch a command from the prompt

- Load the software version to launch
- Launch the data analysis

```
$~ command <options> <arguments>
```

With *command*: the command to launch

# Launch a command line job

- Execute a bash command via srun
- Launch the command from a node
- We use:

```
$~ srun <options> <command>
```

With *command*: the command to launch



# Practice

**Step6: launch the analysis**

6

*Go to the [Practice6](#) of the github*

# Main Slurm commands

Commande	Description	Exemple
srun --time=0X:00 --pty bash -i	Interactive way to connect to a node for X minutes	srun --time=02:00:00 --pty bash -i Connection for 2h
salloc --time=0X:00	Allocate one or several nodes for a later use	Salloc -N 2 --p short --time=05:00
sbatch	Launch analyses in background via a script	sbatch script.sh
sinfo	Informations on partitions	sinfo
sinfo -N l	Informations on nodes of the partitions	sinfo -N l
squeue	Infos on all jobs	squeue -u tando
scontrol show job <job_id>	Infos on the active job <job_id>	scontrol show job 1029

# Options des commandes sbatch, srun, salloc

Options	Description	Exemple
--job-name=<name>	Name the job	sbatch --job-name=tando_blast
-p <partition>	Choose a partition	sbatch -p highmem
--nodelist=<nodeX>	Choisir a particular node	Sbatch -p normal --nodelist=node14
-n <nbre_cpus>	Launch with several cores	srun -n 4
--mail-user=<emailaddress>	Send a email	sbatch --mail-user=ndomassi.tando@ird.fr
--mail-type=<event>	Send a email when : END: end of the job FAIL: abortion BEGIN: beginning of job	Sbatch ---mail-type=BEGIN
--workdir=[dir_name]	Precise the working directory	sbatch s--workdir=/scratch/tando script.sh

Plus d'infos sur Slurm ici: <https://southgreenplatform.github.io/tutorials//cluster-itrop/Slurm/#part-2>



# Practice

## Step 7: Retrieve the results

7

*Go to the [Practice7](#) of the github*

# Delete results from scratchs

- Scratch= temporary spaces
- Verify that the copy is OK before
- Use rm command

```
cd /scratch  
rm -rf nom_rep
```



# Practice

## Step8: Data deletion

8

*Go to the [Practice8](#) of the github*

# Scripts to visualize/delete données temporary data

- Scripts location: /opt/scripts/scratch-scripts
- Visualize data on scratchs: scratch\_use.sh

```
sh /opt/scripts/scratch-scripts/scratch_use.sh
```

- Delete data on scratchs: clean\_scratch.sh

```
sh /opt/scripts/scratch-scripts/clean_scratch.sh
```

# LAUNCH A JOB

# Advantages

- Scheduler choose resources automatically
- Use up to 24 cores at the same time
- Possibility to configure this choice
- Jobs launch in background
  - possibility to turn off your PC/MAC
  - automatic results retrieving

# Launch a batch job

- Execute a script via rm
- Use:

```
$~ sbatch script.sh
```

with *script.sh* : the name of the script

# Options des commandes sbatch, srun, salloc

Options	Description	Exemple
--job-name=<name>	Name the job	sbatch --job-name=tando_blast
-p <partition>	Choose a partition	sbatch -p highmem
--nodelist=<nodeX>	Choisir a particular node	Sbatch -p normal --nodelist=node14
-n <nbre_cpus>	Launch with several cores	srun -n 4
--mail-user=<emailaddress>	Send a email	sbatch --mail-user=ndomassi.tando@ird.fr
--mail-type=<event>	Send a email when : END: end of the job FAIL: abortion BEGIN: beginning of job	Sbatch ---mail-type=BEGIN
--workdir=[dir_name]	Precise the working directory	sbatch s--workdir=/scratch/tando script.sh

Plus d'infos sur Slurm ici: <https://southgreenplatform.github.io/tutorials//cluster-itrop/Slurm/#part-2>

# Bash scripts syntax

First part of the script (in green): sge execution options with the key word #SBATCH

```
#!/bin/bash

##### Configuration SLURM#####
## Name of the job:
#SBATCH --job-name=test
## Name of the output file:
#SBATCH --output=res.txt
## Number of tasks
#SBATCH --ntasks=1
## Execution Time Limit
#SBATCH --time=10:00
#####
```

In the 2nd part of the script: the command to execute

```
#####Partie exécution des commandes #####
```

```
nom_variable1="valeur_variable1"  
nom_variable2="valeur_variable2"
```

```
sleep 30  
hostname
```



# Practice

Launch a script with sge

9

Go to the [Practice9](#) of the github

# Satisfaction survey

Thank you to fill up the form at this URL:

<https://itrop-survey.ird.fr/index.php/562934?lang=fr>

## Citations

If you use i-Trop Bioinformatics resources.

Thank you for citing with:

“The authors acknowledge the IRD itrop HPC (South Green Platform) at IRD montpellier

for providing HPC resources that have contributed to the research results reported within this paper.

URL: <https://bioinfo.ird.fr/> - <http://www.southgreen.fr>”

- Include a budget for bioinformatics resources in your answer to projects funding
- A need in hard drives, renewal machinesetc...
- Available quotations
- Contact [bioinfo@ird.fr](mailto:bioinfo@ird.fr) : help, needs definition, quotations...

# Thank you for your attention !



Le matériel pédagogique utilisé pour ces enseignements est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions (BY-NC-SA) 4.0 International:

<http://creativecommons.org/licenses/by-nc-sa/4.0/>