

i-Trop cluster Slurm Initiation

www.southgreen.fr

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





Emmanuelle Beyne
20% ETP

Aurore COMTE
20% ETP

Bruno GRANOUILAC
50% ETP

Valérie NOEL
25% ETP

Julie ORJUELA-BOUNIOU
25% ETP

Ndomassi TANDO
100% ETP

Christine TRANCHANT-DUBREUIL
20% ETP



Cluster certification in process

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

- Howtos:

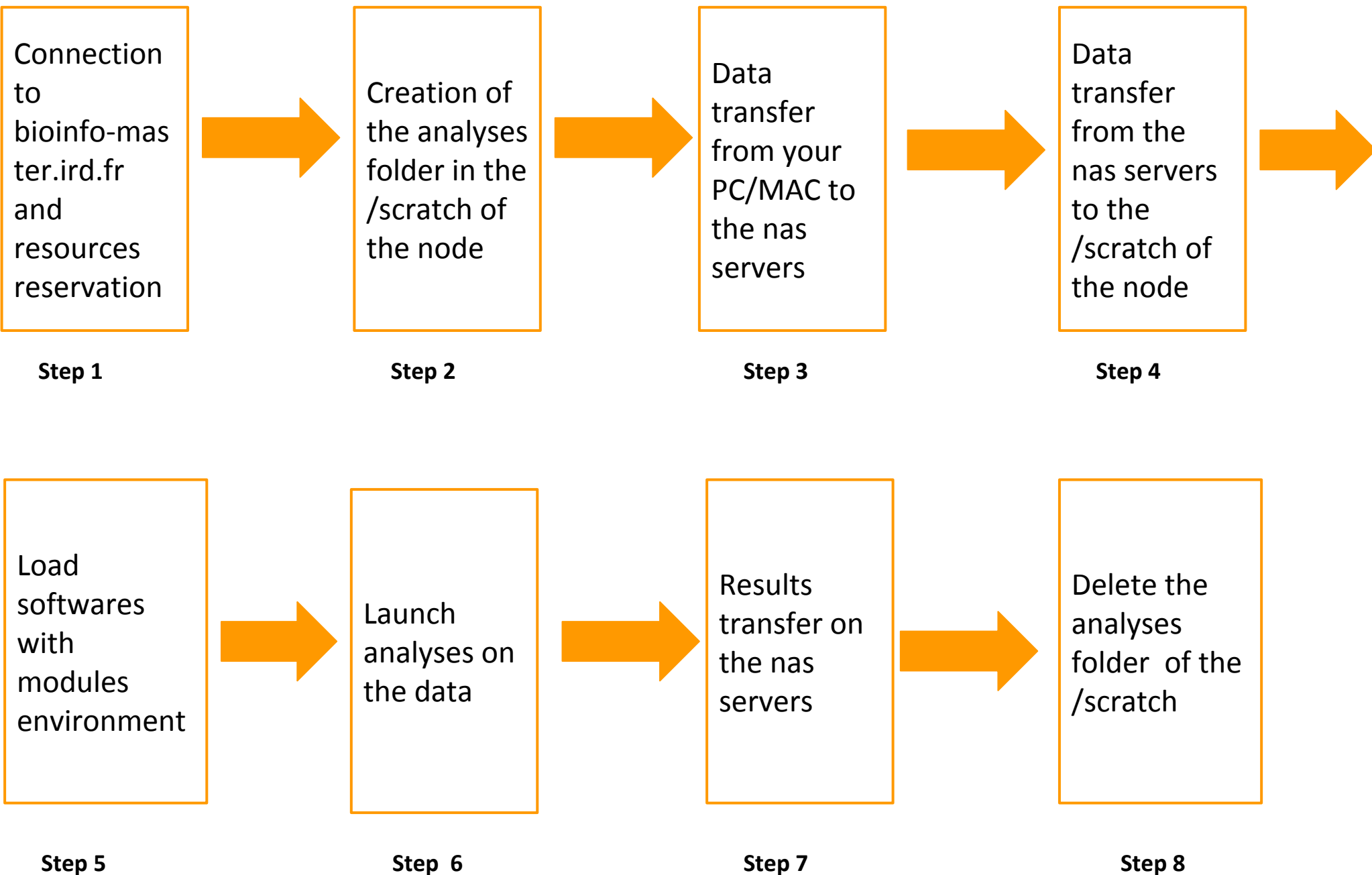
<https://southgreenplatform.github.io/tutorials//cluster-itrop/hpchowto/>

- Slurm Tutorials:

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



Analyses steps of the cluster





Practice

Step 1: Connection, sinfo

1

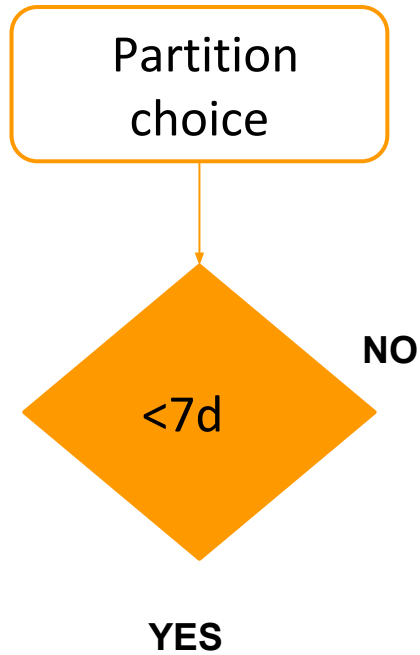
Go to the [Practice 1 and 2](#) of github

Partitions	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 7 days	64 Go to 96 GB	12 to 24 cores
r900	Short Jobs max 7 days	32GB	16 cores /scratch 117GB
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

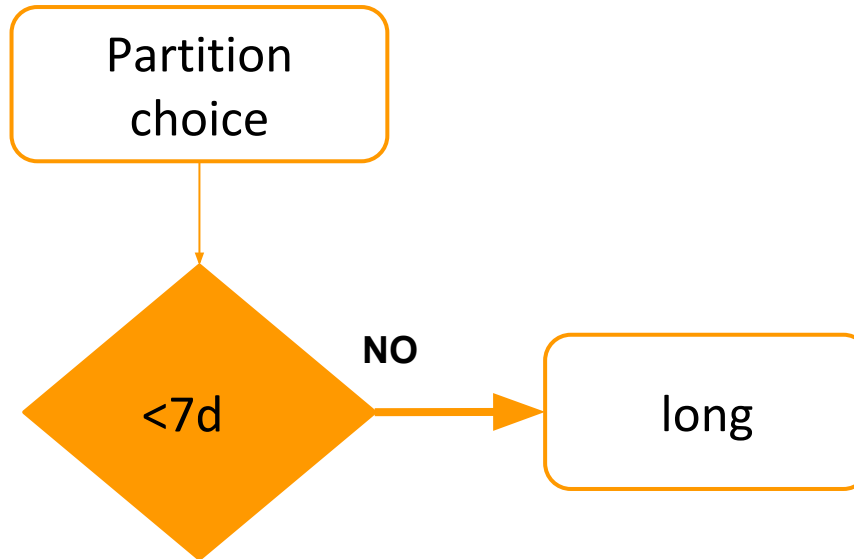
- Partition to work on GPUs processors : basecalling, MiniOn etc..
- Restricted access to gpu_account group
- Request access with arguments to do here:

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

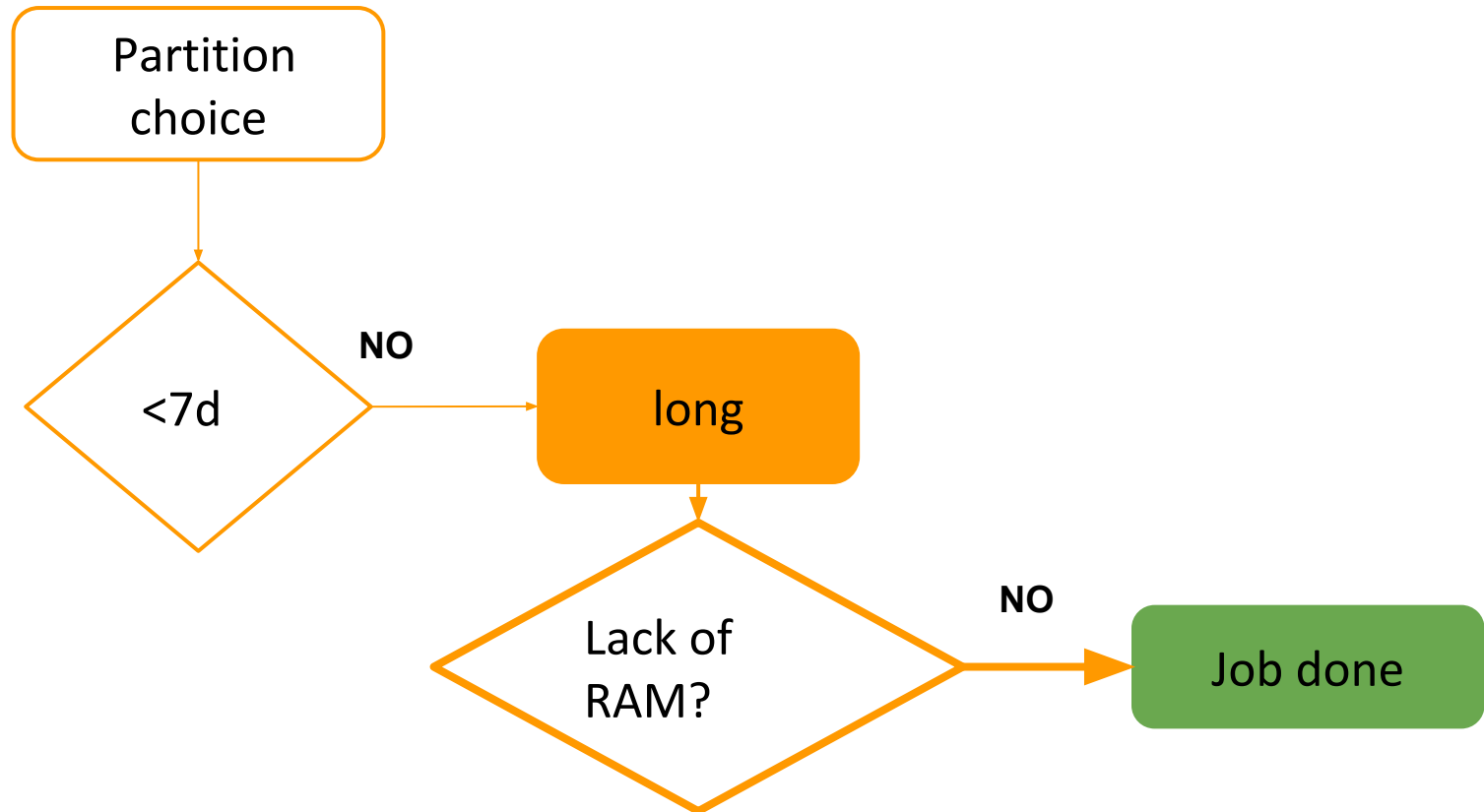
How do I choose the partition?



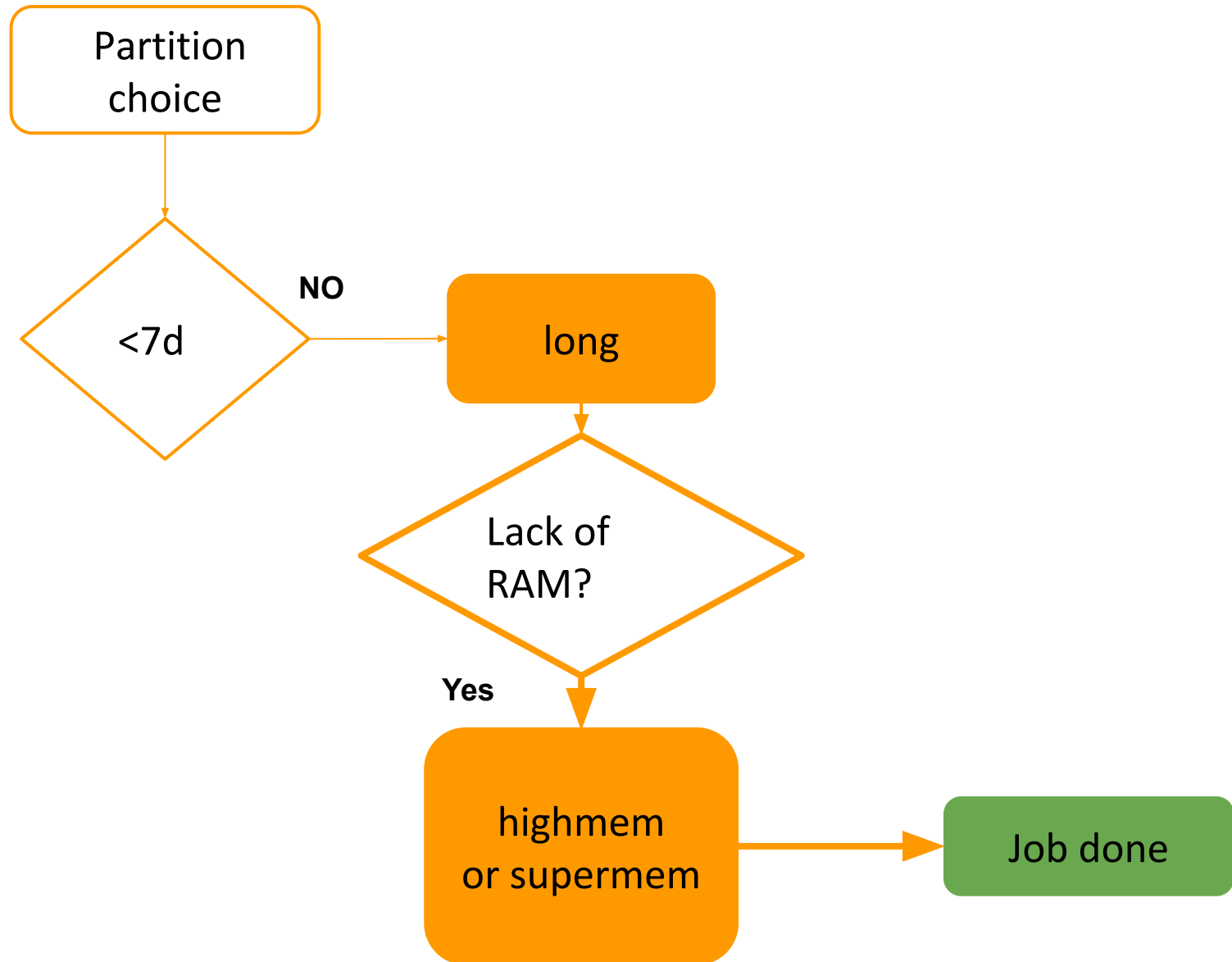
How do I choose the partition?



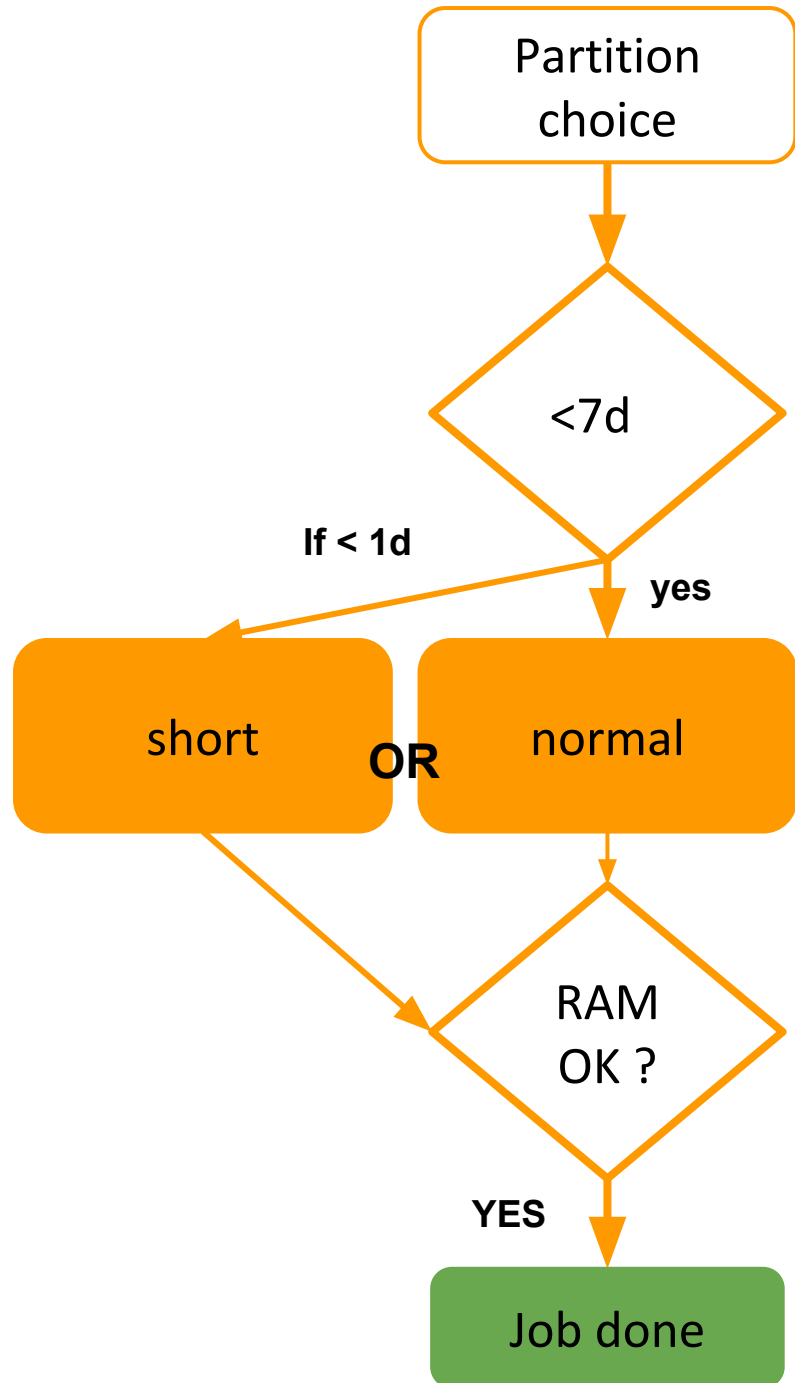
How do I choose the partition?



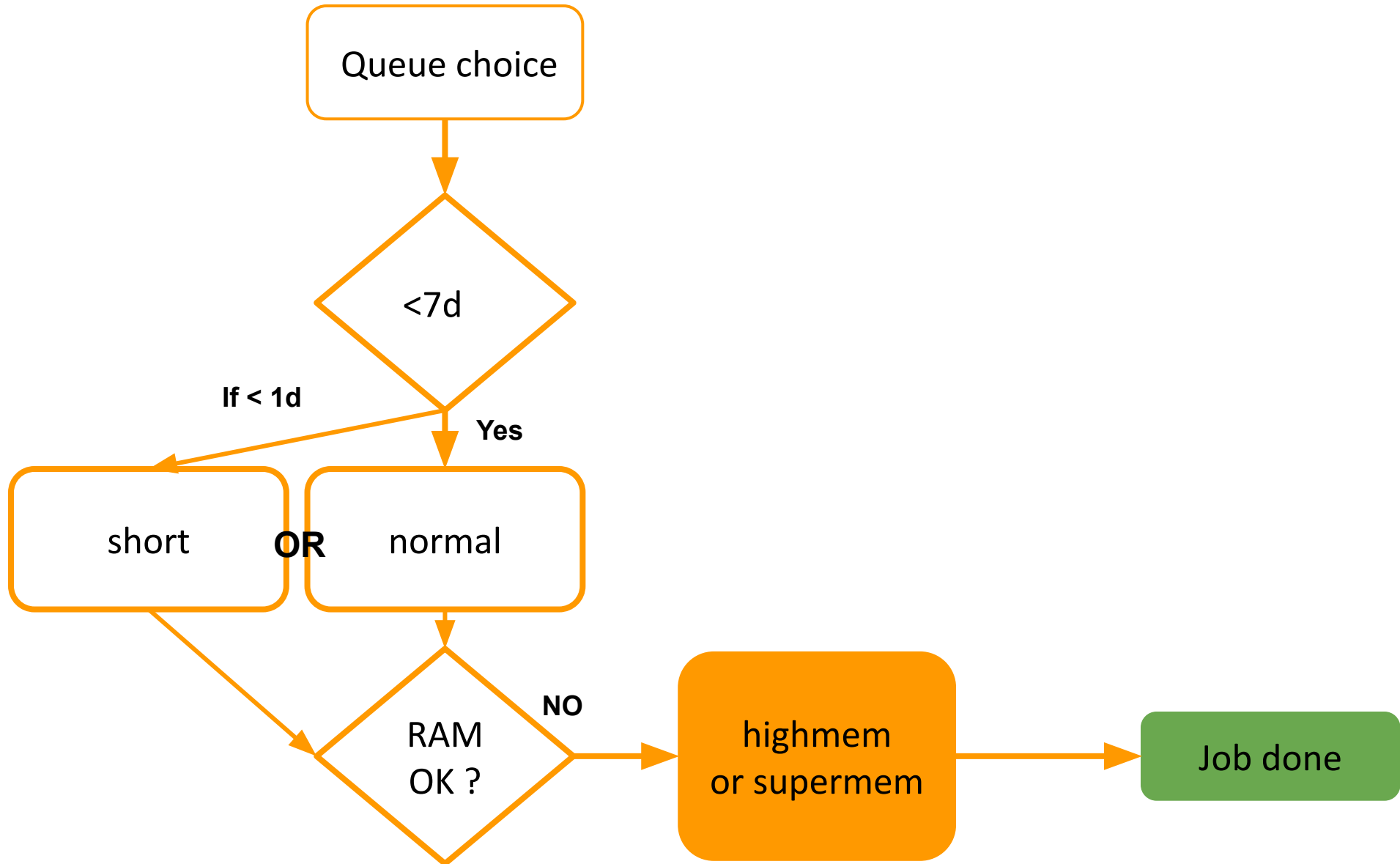
How do I choose the partition?



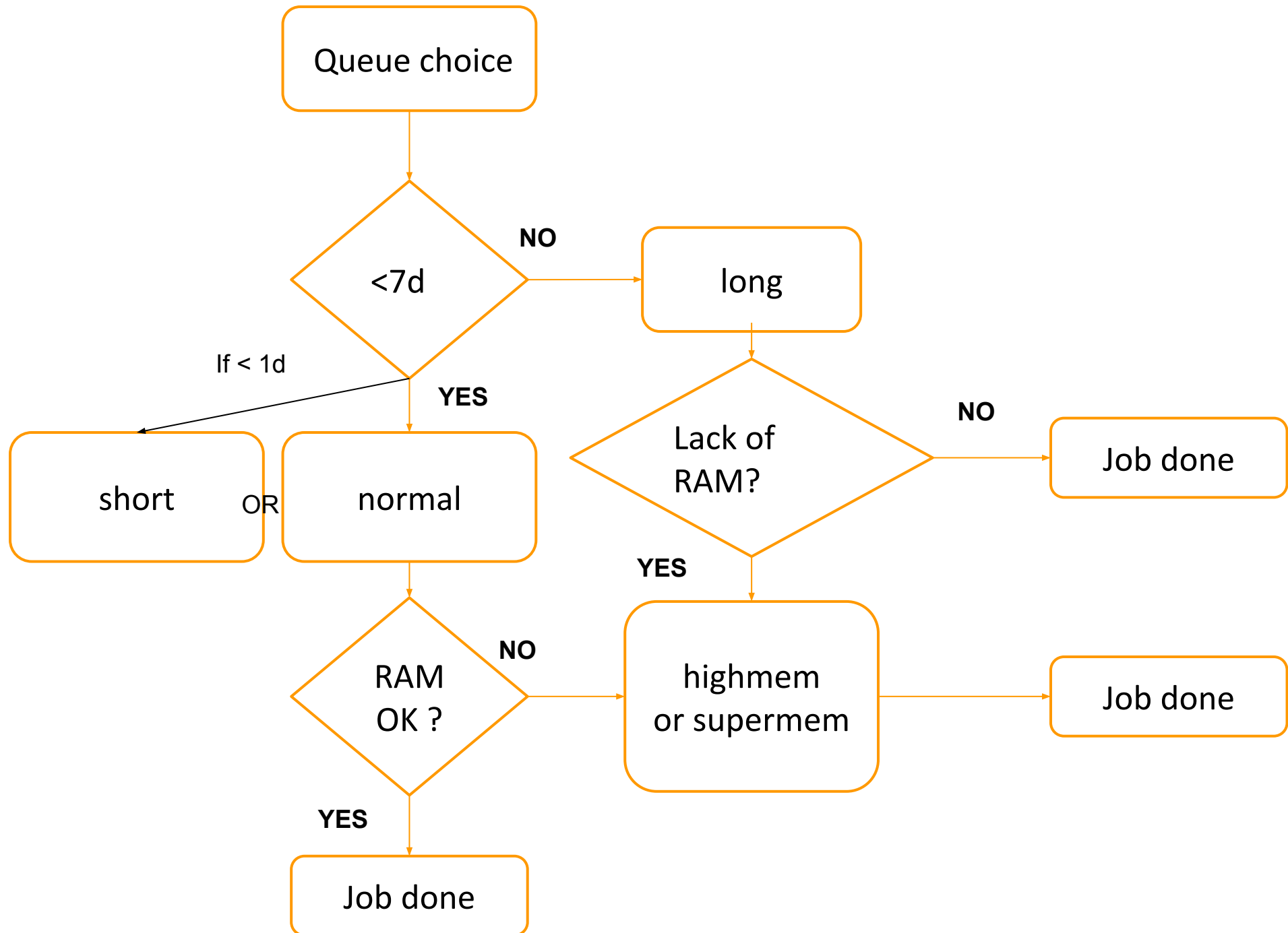
How do I choose the partition?



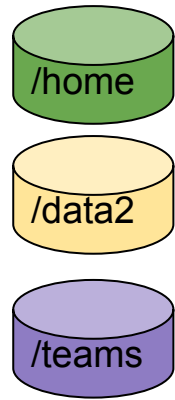
How do I choose the queue?



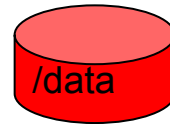
How do I choose the queue?



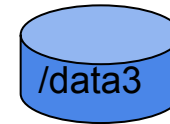
cluster i-Trop disk partitions



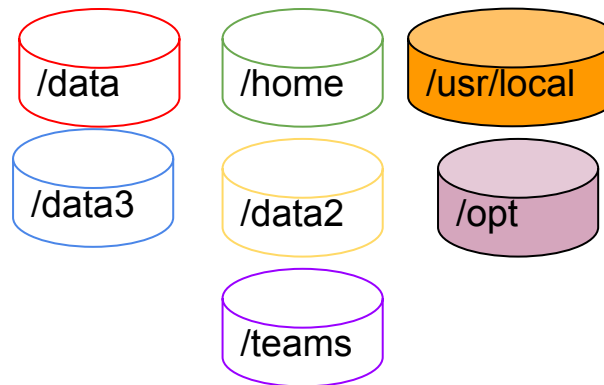
bioinfo-nas.ird.fr



bioinfo-nas2.ird.fr



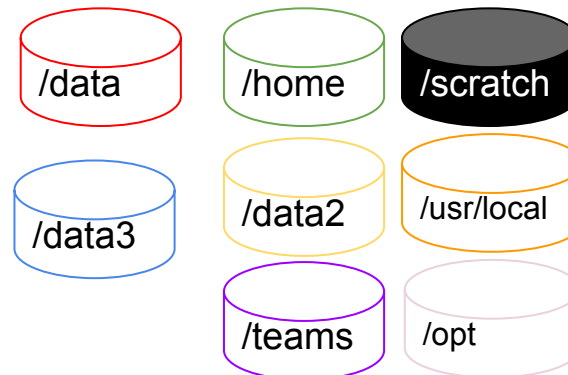
bioinfo-nas3.ird.fr



bioinfo-master.ird.fr



**Virtual links to the
other servers
partitions**

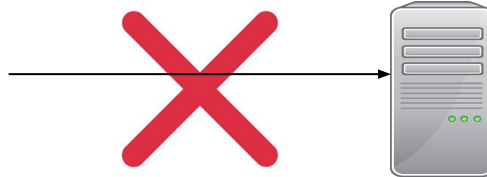


27 nodes





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

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

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

With *command*: the command to launch



Practice

Step6: launch the analysis

6

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



Practice

Step 7: Retrieve the results

7

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

- 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 scratches: scratch_use.sh

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

- Delete data on scratches: clean_scratch.sh

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

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

More infos here : <https://southgreenplatform.github.io/tutorials//cluster-itrop/Slurm/#part-2>

Options of sbatch, srun, salloc commands

Options	Description	Exemple
<code>--job-name=<name></code>	Name the job	<code>sbatch --job-name=tando_blast</code>
<code>-p <partition></code>	Choose a partition	<code>sbatch -p highmem</code>
<code>--odelist=<nodeX></code>	Choose a particular node	<code>sbatch -p normal --odelist=node14</code>
<code>-n <nb_tasks></code>	Launch several instance of a command	<code>srun -n 4</code>
<code>-c <nb_cpu_per_task></code>	Allocate the number of cpus per task	<code>srun -n 4 -c 2 hostname</code>
<code>--mail-user=<emailaddress></code>	Send a email	<code>sbatch --mail-user=ndomassi.tando@ird.fr</code>
<code>--mail-type=<event></code>	Send a email when : END: end of the job FAIL: abortion BEGIN: beginning of job ALL: all events	<code>sbatch ---mail-type=BEGIN</code>
<code>--workdir=[dir_name]</code>	Precise the working directory	<code>sbatch --workdir=/scratch/tando script.sh</code>

LAUNCH A JOB

- 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

- Execute a script via Slurm
- Use:

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

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

Options of sbatch, srun, salloc commands

Options	Description	Exemple
<code>--job-name=<name></code>	Name the job	<code>sbatch --job-name=tando_blast</code>
<code>-p <partition></code>	Choose a partition	<code>sbatch -p highmem</code>
<code>--odelist=<nodeX></code>	Choose a particular node	<code>sbatch -p normal --odelist=node14</code>
<code>-n <nb_tasks></code>	Launch several instance of a command	<code>srun -n 4</code>
<code>-c <nb_cpu_per_task></code>	Allocate the number of cpus per task	<code>srun -n 4 -c 2 hostname</code>
<code>--mail-user=<emailaddress></code>	Send a email	<code>sbatch --mail-user=ndomassi.tando@ird.fr</code>
<code>--mail-type=<event></code>	Send a email when : END: end of the job FAIL: abortion BEGIN: beginning of job ALL: all events	<code>Sbatch ---mail-type=BEGIN</code>
<code>--workdir=[dir_name]</code>	Precise the working directory	<code>sbatch s--workdir=/scratch/tando script.sh</code>

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 Slurm

9

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

Thank you to fill up the form at this URL:

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

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 machines etc...
- Available quotations
- Contact 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/>