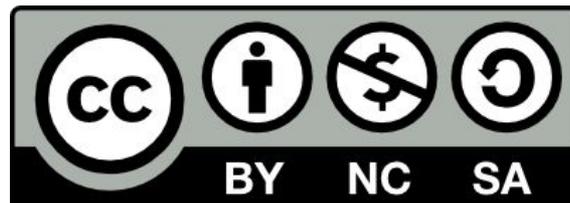


Initiation HPC cluster

www.southgreen.fr

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





Julie ORJUELA-BOUNOL¹, IE
Bioinformaticienne
25%



Ndomassi TANDO, IE
Administrateur systeme
100%
Animateur plateau



Christine TRANCHANT-DUBREUIL, IE
Bioinformaticienne
20%

Aurore COMTE, IE
Bioinformaticienne
20%

Alexis DEREPPER², IE
Bioinformaticien
20%



Valérie NOEL, TCS
Bioinformaticienne
25%



Bruno GRANOUILAC³, IE
Systèmes d'information
20%

Mise à disposition
de ressources
de calcul et
logicielles

Développement de
logiciels d'analyse
et de SI

Plateau bioinformatique

Assistance et
support aux
équipes

Formations au Sud
et au Nord

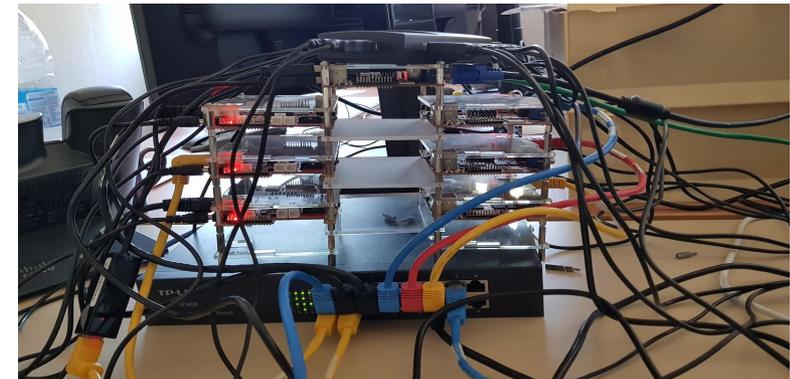
ARCHITECTURE

- une unité logique de plusieurs serveurs
- une unique machine puissante
- une puissance de calcul élevée
- Une plus grande capacité de stockage
- Une fiabilité supérieure
- Une plus grande disponibilité des ressources

- une unité logique de plusieurs serveurs
- une unique machine puissante
- une puissance de calcul élevée
- Une plus grande capacité de stockage
- Une fiabilité supérieure
- Une plus grande disponibilité des ressources



- une unité logique de plusieurs serveurs
- une unique machine puissante
- une puissance de calcul élevée
- Une plus grande capacité de stockage
- Une fiabilité supérieure
- Une plus grande disponibilité des ressources



CALCUL



- **Noeud maître**
Gère les ressources et les priorités des jobs
- **Noeuds de calcul**
Ressources (CPU ou mémoire RAM)

CALCUL



- **Noeud maître**
Gère les ressources et les priorités des jobs
- **Noeuds de calcul**
Ressources (CPU ou mémoire RAM)
- **Serveur(s) NAS**
Stockage

STOCKAGE



- **1 Noeud Maître**



master.univ-ouaga.bf

Rôle :

- Lancer et prioriser les jobs sur les nœuds de calcul
- Accessible depuis Internet
- Connexion :

```
ssh login@master.univ-ouaga.bf
```

- **1 Noeud Maître**



master.univ-ouaga.bf

Rôle :

- Lancer et prioriser les jobs sur les nœuds de calcul
- Accessible depuis Internet
- Connexion :

```
ssh login@master.univ-ouaga.bf
```

- **1 Noeud de Calcul**

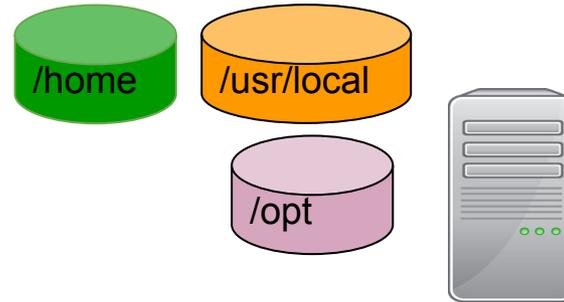


node0

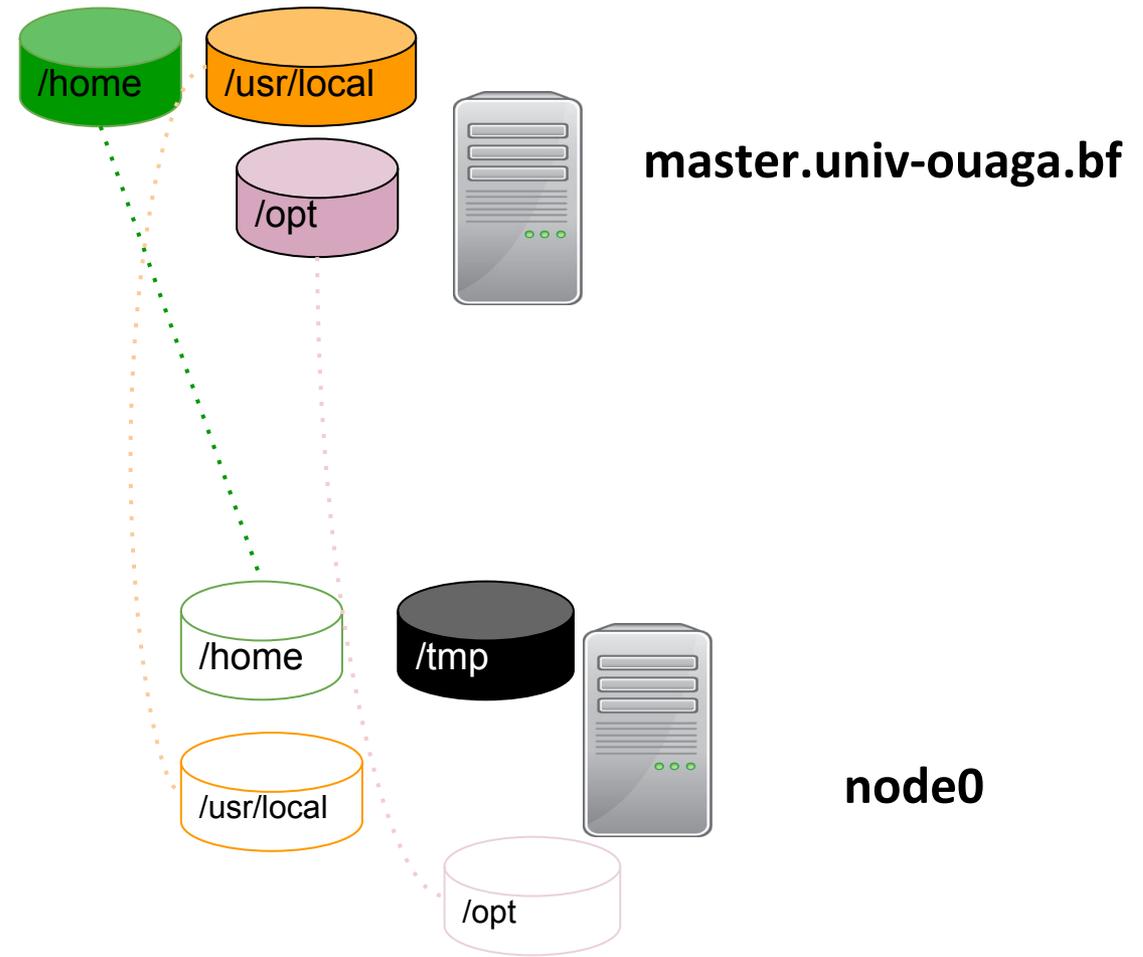
Rôle :

- Utilisé par le maître pour exécuter les jobs/calculs
- Pas accessible depuis Internet
- Nom: node0

**Partitions sur la
machine master**



master.univ-ouaga.bf



Liens virtuels vers les partitions de node0

Connexion à
master.oua
ga-univ.bf
et
réservation
de
ressources



Etape 1
salloc,srun
ou sbatch



Practice

Etape 1: Connexion, info

1

Aller sur le [Exercice 1](#) du github

partition	noeud	Caractéristiques RAM noeuds	Caractéristiques coeurs noeuds	Caractéristique partition
main	node0	256 Go	28 coeurs	Temps infini
short	node0	256 Go	28 coeurs	Limitée à 1 jour

Connexion à
master et
réservation
de
ressources



Création du
répertoire
d'analyse
/tmp sur le
noeud
réservé

Etape 1

Etape 2

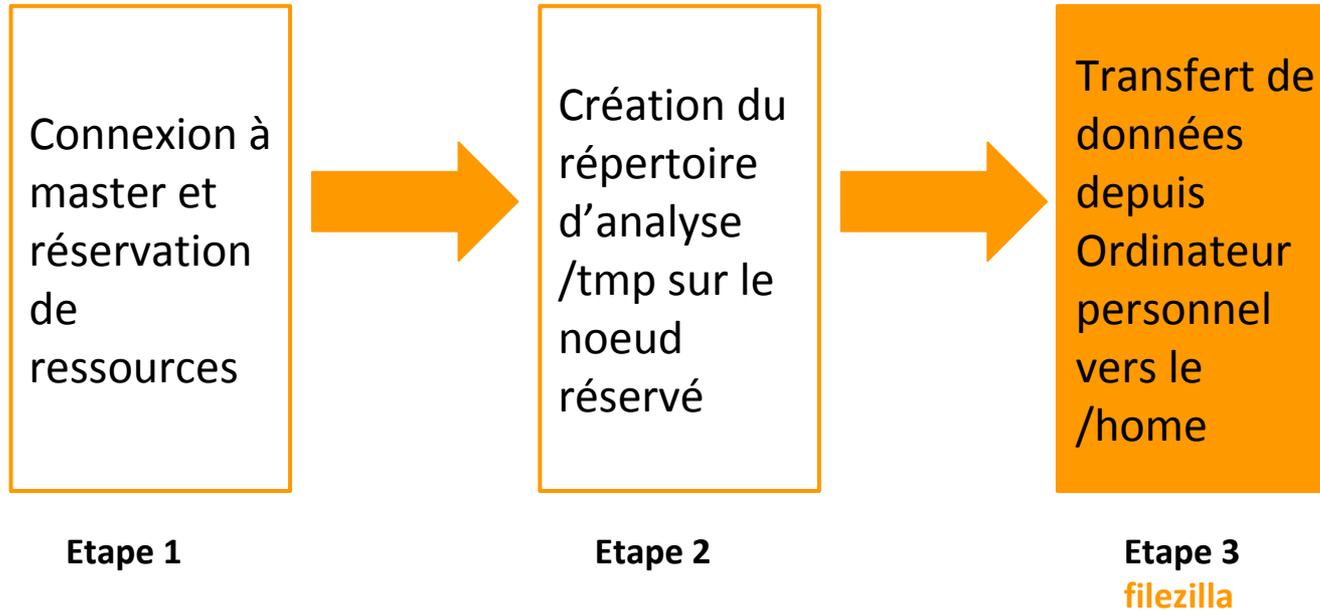


Practice

Etape 2: srun, partition

2

Aller sur l' [Exercice2](#) du github



Copier les données depuis son ordinateur personnel vers le /home si les données à analyser ne sont pas sur le cluster

Transferts de données sur le cluster en temps normal

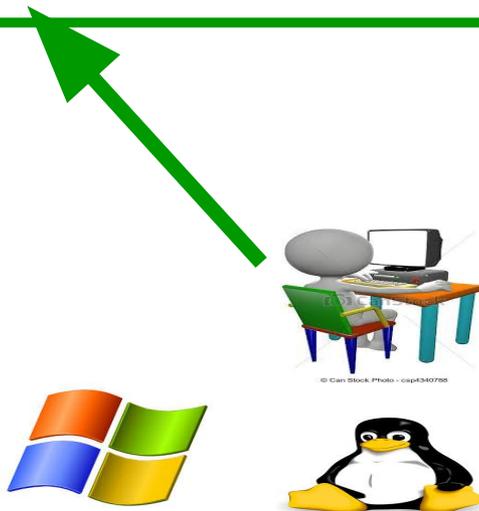
/home ou /opt



master.univ-ouaga.bf

Hostname :
master.univ-ouaga.bf
Login : cluster account

Password : cluster
password
Port : 22



Transferts de données sur le cluster pour le tp

/home ou /opt

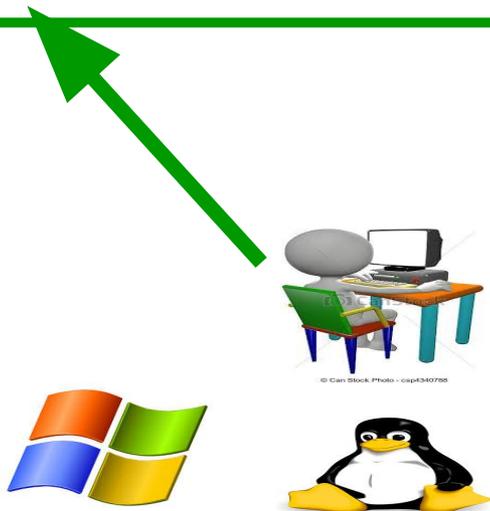


192.168.4.102

Hostname :
192.168.4.102

Login : cluster
account

Password : cluster
password
Port : **22**



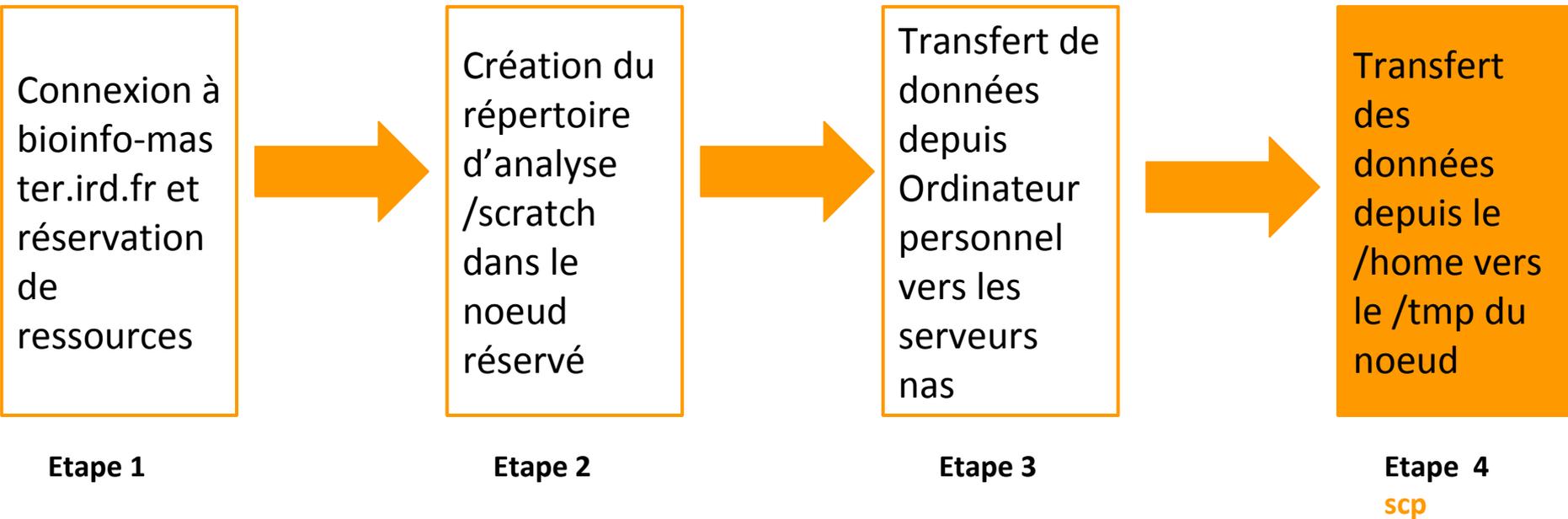


Practice

Etape3: filezilla

3

Aller sur l' [Exercice3](#) du github



- Copie entre 2 serveurs distants :

```
scp -r source destination
```

- Syntaxe si la source est distante :

```
scp -r nom_serveur:/chemin/fichier_a_copier repertoire_local
```

- Syntaxe si la destination est distante :

```
scp -r /chemin/fichier_a_copier nomserveur:/chemin/repertoire_distant
```

Ex: `scp -r master:/home/formationX/repertoire /tmp/formationX/`



Practice

Etape4: scp vers noeuds

4

Aller sur l' [Exercice4](#) du github

- Permet de choisir la version du logiciel que l'on veut utiliser
- Surpassent les variables d'environnement

- 5 types de commandes :
 - Voir les modules disponibles :
`module avail`
 - Obtenir une info sur un module en particulier :
`module whatis + module name`
 - Charger un module :
`module load + modulename`
 - Lister les modules chargés :
`module list`
 - Décharger un module :
`module unload + modulename`
 - Décharger tous les modules :
`Module purge`

Connexion à
bioinfo-mas-
ter.ird.fr et
réservation
de
ressources

Etape 1

Création du
répertoire
d'analyse
/scratch
dans le
noeud
réservé

Etape 2

Transfert de
données
depuis
Ordinateur
personnel
vers les
serveurs
nas

Etape 3

Transfert
des
données
depuis le
/home vers
le /tmp du
noeud

Etape 4

Charger ses
logiciels avec
module
environment

Etape 5
module



Practice

Etape5: module environment

5

Aller sur l' [Exercice5](#) du github

- Charger la version du logiciel à lancer
- Lancer l'analyse des données

```
$~ commande <options> <arguments>
```

Avec *commande*: la commande à lancer

- Exécuter une commande bash via srun
- Lance la commande sur un noeud
- On utilise la commande:

```
$~ srun "commande"
```

Avec *commande*: la commande à lancer

Connexion à bioinfo-master.ird.fr et réservation de ressources

Etape 1



Création du répertoire d'analyse /scratch dans le noeud réservé

Etape 2



Transfert de données depuis Ordinateur personnel vers les serveurs nas

Etape 3



Transfert des données depuis les nas vers le /scratch du noeud

Etape 4



Charger ses logiciels avec modules environment

Etape 5



Lancer les analyses sur les données

Etape 6



Practice

Etape6: lancer l'analyse

6

Aller sur l' [Exercice6](#) du github

- Copie entre 2 serveurs distants :

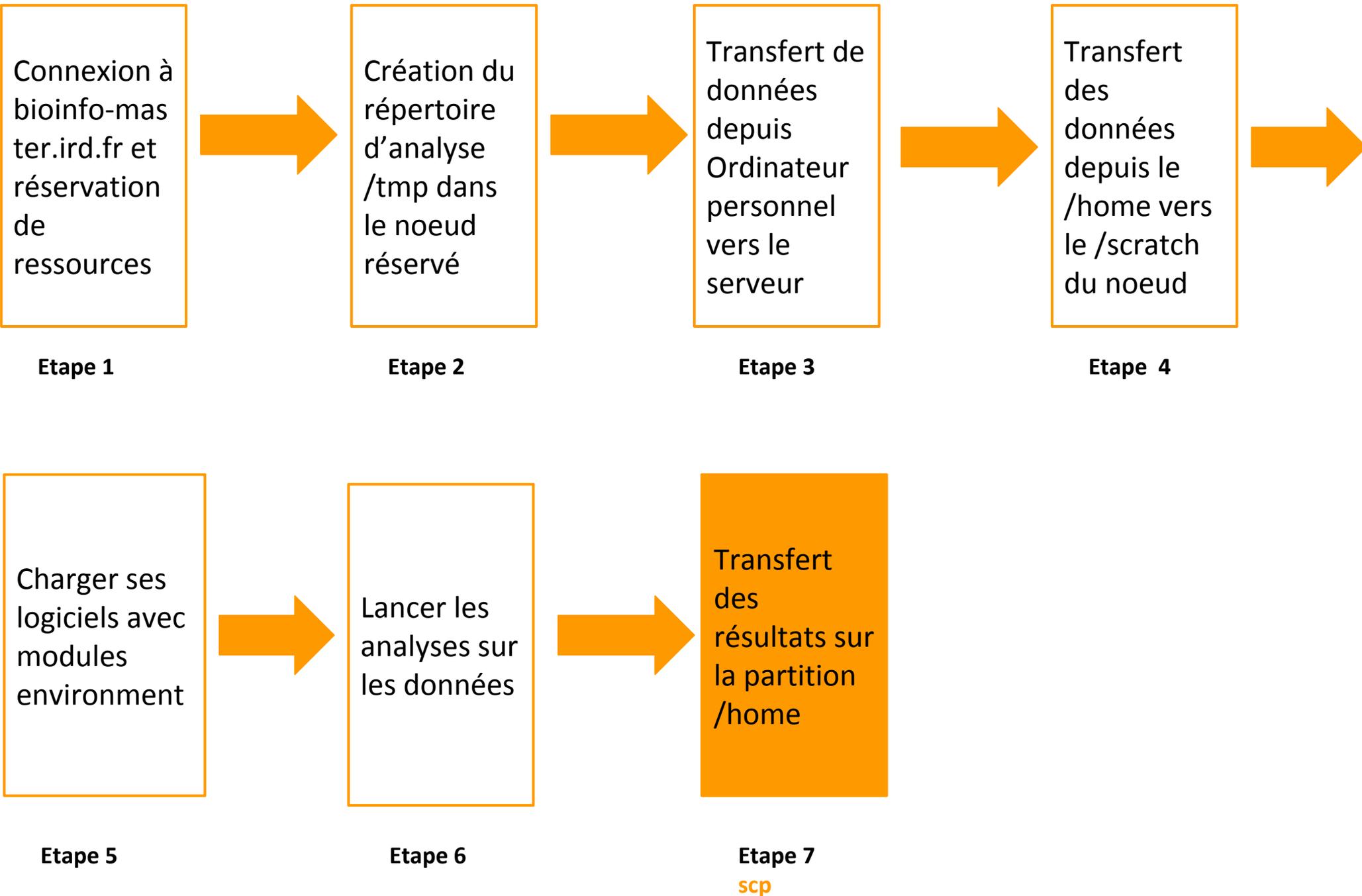
```
scp source destination
```

- Syntaxe si la source est distante :

```
scp nom_serveur:/chemin/fichier_a_copier repertoire_local
```

- Syntaxe si la destination est distante :

```
scp /chemin/fichier_a_copier nomserveur:/chemin/repertoire_distant
```





Practice

Etape7: Récupérer les résultats

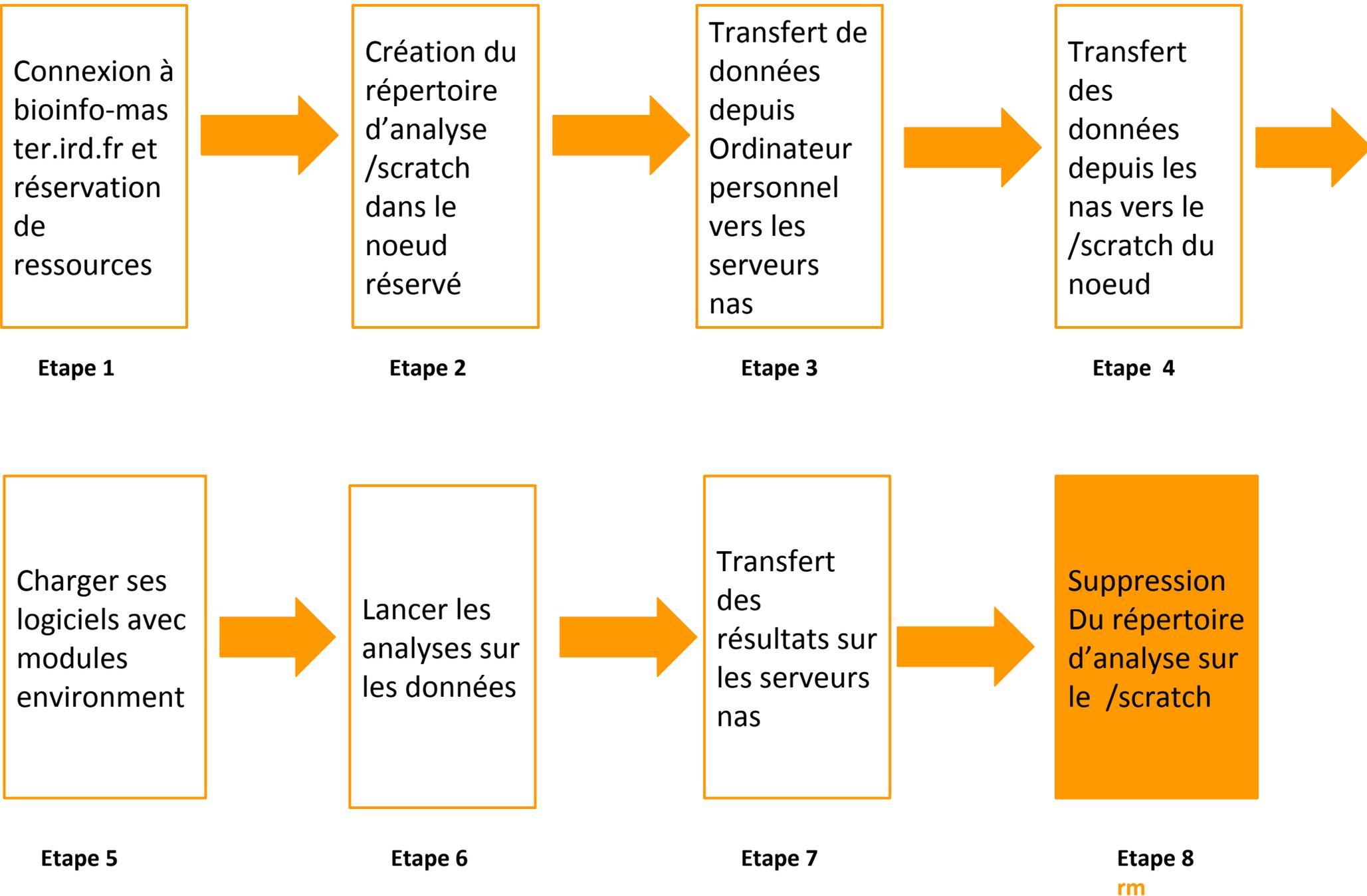
7

Aller sur l'[Exercice7](#) du github

- /tmp= espaces temporaires
- Vérifier la copie des résultats avant
- Utiliser la commande rm

```
cd /tmp  
rm -rf nom_rep
```

Etapes d'une analyse sur le cluster





Practice

Etape8: suppression des données

8

Aller sur l'[Exercice8](#) du github

LANCER UN JOB

- Le scheduler choisit les ressources automatiquement
- Lancer des jobs utilisant jusqu'à 24 coeurs
- Possibilité de paramétrer ce choix
- Jobs lancés en arrière plan
 - possibilité d'éteindre son ordinateur
 - récupération des résultats automatique

- C'est le fait d'exécuter un script bash via sge
- On utilise la commande:

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

Avec `script.sh` : le nom du script

Options	Description	Exemple
<code>--job-name=[name]</code>	Donner un nom au job	<code>sbatch --job-name=tando_blast</code>
<code>-p partition</code>	Choisir une partition en particulier	<code>sbatch -p main job.sh</code>
<code>--odelist=<nodeX></code>	Choisir un noeud en particulier	<code>srun --odelist=node0</code>
<code>-N <nombre de coeurs></code>	Lancer avec plusieurs coeurs	<code>Srun -N 2</code>
<code>--mail-user=<adresse_email></code>	Envoyer un mail	<code>sbatch --mail-user=ndomassi.tando@ird.fr</code>

Voir plus d'options disponibles ici:

[Options de base avec Slurm](#) dans la rubrique Les principales options disponibles pour lancer une analyse sous Slurm:

Dans la première partie du script on renseigne les options d'exécution de slurm avec le mot clé #SBATCH (partie en vert)

```
#!/bin/bash

##### Configuration SLURM#####
## On définit le nom du job
#SBATCH --job-name=test
## On définit le nom du fichier de sortie
#SBATCH --output=res.txt
## On définit le nombre de tâches
#SBATCH --ntasks=1
## On définit le temps limite d'exécution
#SBATCH --time=10:00
#####
```

Dans la 2e partie du script on renseigne les actions à effectuer

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

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

```
sleep 30  
hostname
```



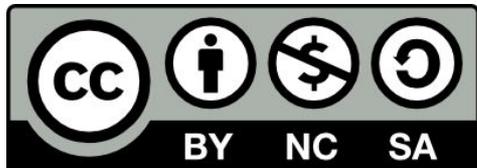
Practice

Lancer un script avec qsub

9

Aller sur l' [Exercice9](#) du github

Merci pour votre 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/>