Introduction à l'utilisation du Mésocentre de Franche-Comté



K.Mazouzi C.Clerget

8 décembre 2016

Plan

1. Accès aux ressources
2. Environnement de travail
3. Soumissions de jobs
4. Contacts

Plan

1 Accès aux ressources

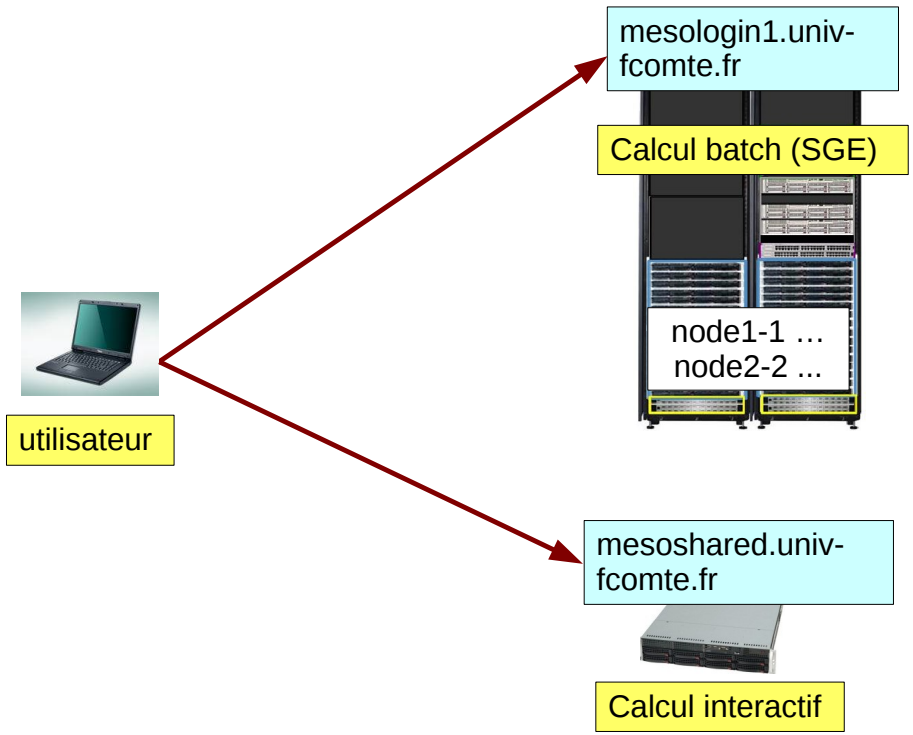
2 Environnement de travail

3 Soumissions de jobs

4 Contacts

Deux points d'accès

- Lumière
  - Calcul en mode batch
  - Connexion sur un noeud spécifique
  - Soumissions de jobs
- Mesoshared
  - Programmes interactifs
  - Programmes graphiques



Sécurité

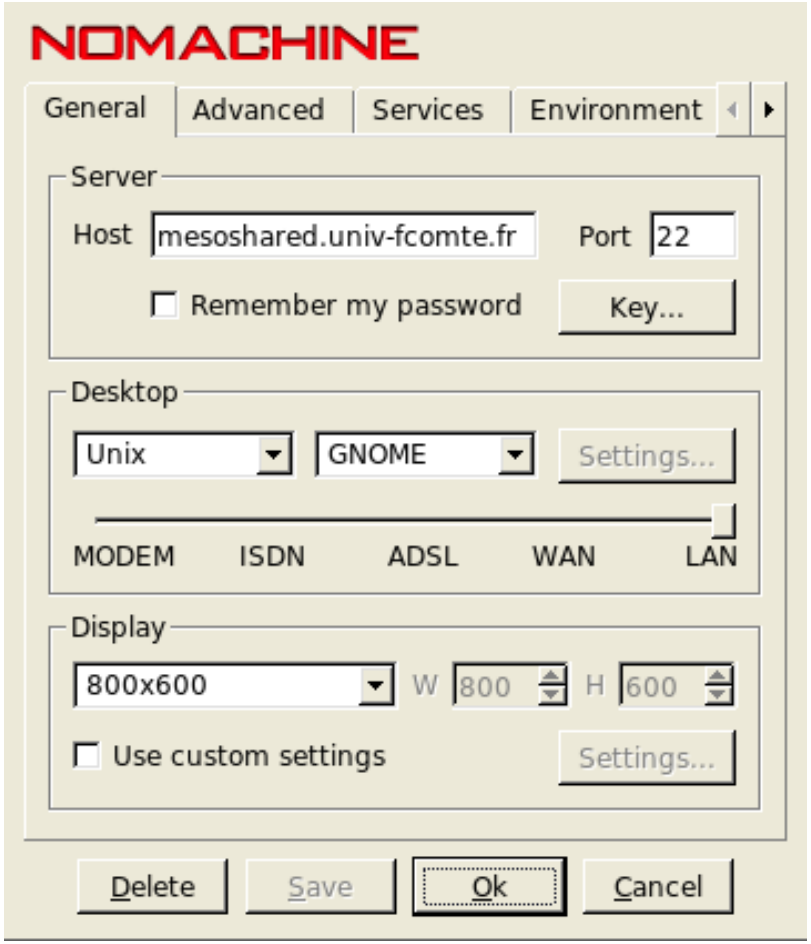
- Toutes les ressources sont sur le réseau privé de l'Université
- Accès depuis l'extérieur requiert à un **accès VPN**
  - Certificats et service gérés par le CRI de l'Université
  - Demande à effectuer en priorité auprès de votre responsable informatique ou directement au CRI le cas échéant
  - Documentation Windows/Linux/Mac OS X : `http://vpn.univ-fcomte.fr`

Outils de connexion

	Windows	Linux et Mac OS X
Ligne de commande	PuTTY	SSH (inclus)
Transferts de fichiers	WinSCP	Client FTP/SCP (Filezilla...)
Session graphique	NXClient	NXClient ou SSH (X11)

- PuTTY : `http://www.putty.org`
- WinSCP : `http://winscp.net`
- NXClient : Voir liens téléchargements wiki

Exemple : NXClient



Exemple : NXClient



Plan

1 Accès aux ressources

2 Environnement de travail

3 Soumissions de jobs

4 Contacts

Environnement : Stockage

Répertoire personnel (variable $HOME)

- Lecture/écriture sur les noeuds de connexion
- Lecture seule sur les noeuds de calcul
- Limité à 10 Go sauvegardés

Répertoire de travail (variable $WORK)

- Lecture/écriture sur tous les noeuds
- Limité à 600 Go par défaut non sauvegardés
- Solution spécialisée (BeeGFS), accès très performant
- Répertoire temporaire également accessible via $TMPDIR

Logiciels

Récupération de la liste des logiciels disponibles :

```
$ module avail
--------------- /Softs/lumiere-modules ---------------
abinit/7.10.5                   bio/sumatra/1.0
altair/HyperWorks/11.0          bio/t-coffee/10.00
ansys/fluent/15.0.7             bio/t-coffee/9.03
bio/bamtools/2.3.0              bio/t-coffee/default
bio/bcftools/1.1                bio/velvet/1.2.09
...
```

Chargement d'un module :

```
$ module load matlab/r2014b
$ matlab
```

Logiciels

Récupération de la liste des modules chargés :

```
$ module list
Currently Loaded Modulefiles:
  1) matlab/r2014b
```

Déchargement d'un module :

```
$ module rm matlab/r2014b
$ matlab
```

Chargement automatique à la connexion :

```
# Ajouter cette ligne au fichier ~/.bashrc
module load matlab/r2014b
```

Plan

1 Accès aux ressources

2 Environnement de travail

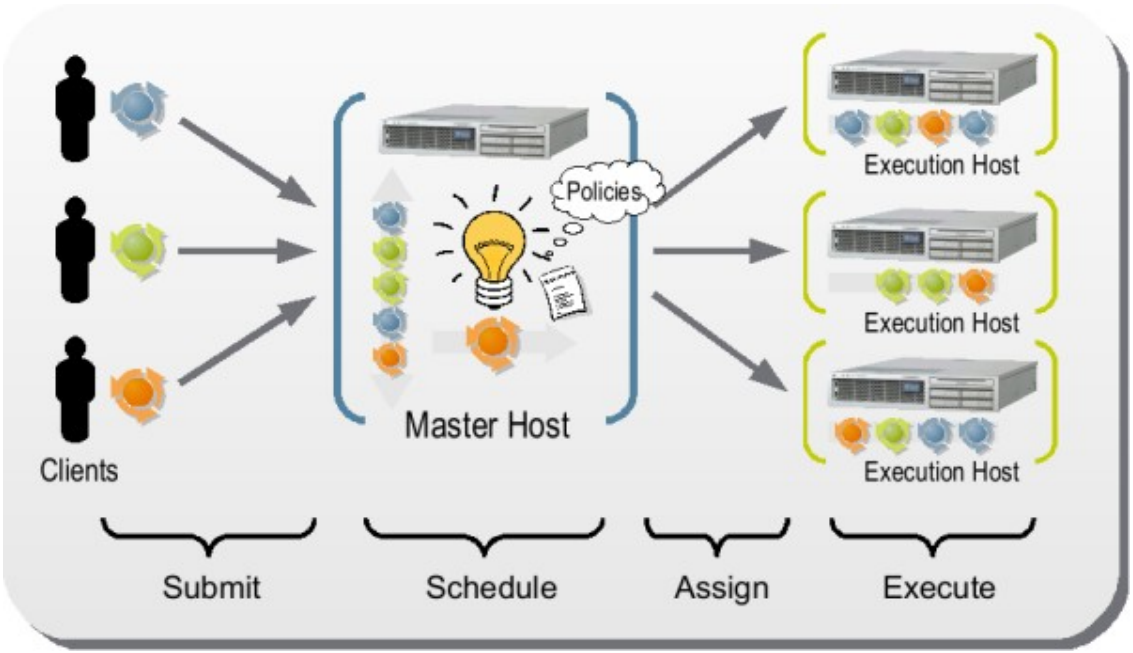
3 Soumissions de jobs

4 Contacts

SGE

SGE : un gestionnaire de ressources

- Partage des ressources
- Gestion de jobs soumis en file d'attente (queues)
- Système de priorité
- Passage au fur et à mesure que les ressources se libèrent



Terminologie

Qu'appelle-t-on une *queue* ?

Une *queue* est une file d'attente dans laquelle s'accumulent les jobs en attente de traitement par le cluster.

Qu'appelle-t-on un *job* ?

Un *job* est un petit programme (shell) contenant la définition de l'environnement dans lequel il doit être exécuté.
$\Rightarrow$ Numéro unique (jobid)

Files d'attentes

Le choix de la bonne file d'attente est importante pour assurer une exécution correcte et rapide des travaux soumis à SGE.

Quelles sont les files d'attente disponibles ?

`Commande : qconf -sql`

```
$ qconf -sql
3dvisu.q
all.q
bigmem.q
formation.q
parallel.q
tesla.q
xphi.q
```

Files d'attente : cluster lumière

Nom	Description
all.q	Travaux séquentiels et parallèles en mémoire partagée (une semaine)
parallel.q	Travaux paralléles en mémoire distribuée (une semaine)
tesla.q	Jobs utilisant les cartes Tesla (16 coeurs, 6 cartes)
xphi.q	Jobs utilisant les cartes Xeon Phi (12 coeurs, 4 cartes)
bigmem.q	Jobs nécessitant une quantité de mémoire importante (8 coeurs, 96Go de mémoire)
3dvisu.q	File d'attente pour la visualisation haute-performance

Attention, ces informations peuvent être amenées à changer, toujours se reporter au wiki utilisateur en priorité !

Soumission de jobs : qsub

qsub [options] [script]

- **options** : ressources (file d'attente, mémoire, nombre de slots...)
- **script** : fichier texte contenant les commandes à lancer
- En pratique, les options sont souvent intégrées dans le script pour éviter de les indiquer à chaque lancement

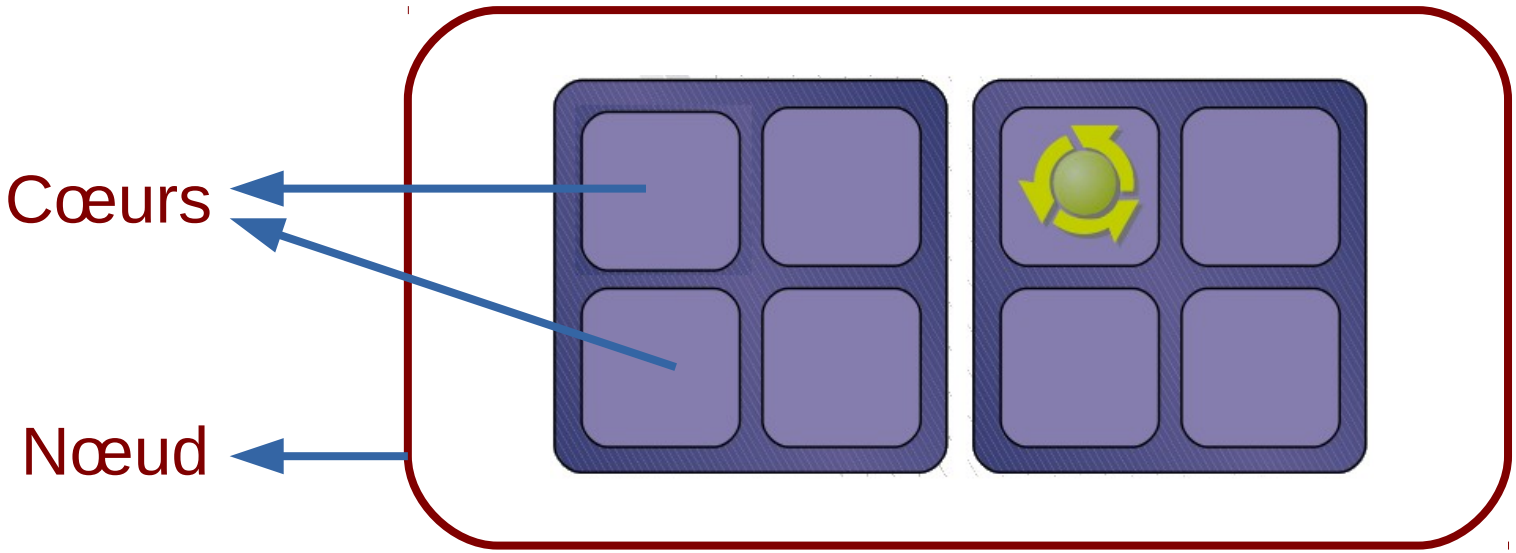
```
$ qsub -q all.q -l h_vmem=4G matlab.sge
```

La commande `man` (`man qsub`) permet d'obtenir de l'aide sur la plupart des commandes Unix.

Job séquentiel

Définition

Un job séquentiel s'exécute sur un et un seul **coeur** d'un noeud particulier



Job séquentiel

Exemple de script SGE :

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -o $JOB_NAME.$JOB_ID.out
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err

appli_seq param1 param2 param3
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err

appli_seq param1 param2 param3
```

```
qsub script.sge
```

Job séquentiel

Exemple de script SGE :

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err

appli_seq param1 param2 param3
```

```
qsub script.sge
```

```
$ qstat
job-ID  prior   name       user    state queue                  slots
-----------------------------------------------------------------------
90024   0.75000 test_sge   user1       r all.q@node1-50             1
```

Job parallèle

Définition

Un job parallèle s'exécute sur plusieurs **coeurs**

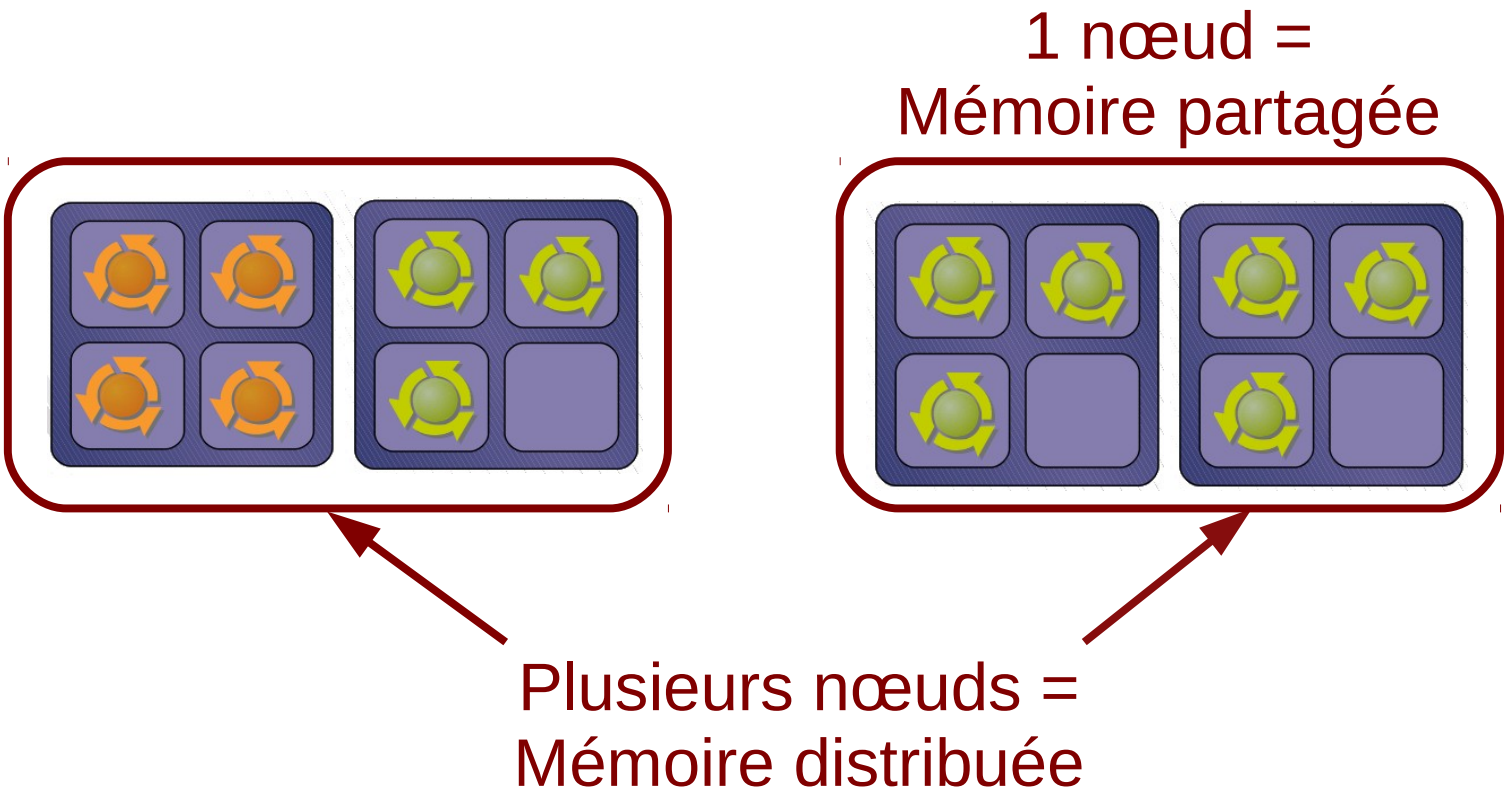


Tableau de tâche

Contexte

Lancer plusieurs instances du même programme en parallèle sur des données différentes.

Exemple

Lancer une application **appli** 100 fois avec des jeux de paramètres différents : `input_1,...,input_100`

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -t1-100
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err

./appli input_$TASK_ID
```

Application parallèle OpenMP

Contexte

L'application peut tirer parti de plusieurs coeurs sur un même noeud

$\Rightarrow$ utilisation de l'environnement parallèle **openmp** et de la variable `OMP_NUM_THREADS`

```
#!/bin/bash
#$ -q all.q
#$ -N test_sge
#$ -pe openmp 8 # environnement parallele + slots
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err

export OMP_NUM_THREAD=$NSLOTS

./appli_openmp input
```

Application parallèle MPI

Contexte

L'application peut s'exécuter sur plusieurs noeuds

$\Rightarrow$ utilisation de l'environnement parallèle **mpi**

```
#!/bin/bash
#$ -q parallel.q
#$ -N test_sge
#$ -pe mpi 32
#$ -o $JOB_NAME.$JOB_ID.out
#$ -e $JOB_NAME.$JOB_ID.err

module load ompi

mpirun -np $NSLOTS ./appli_mpi
```

Suivi des jobs

`qstat`, `qhost` : suivi des jobs en cours d'exécution et des ressources disponibles.

État du job	Signification
`d`	deleting
`t, r`	transfering, running
`s, S, T`	suspending, threshold
`R`	restarted
`w`	waiting
`h`	hold
`E`	error
`q`	queuing

FAQ

Pourquoi mon job est-il en "Eqw" ?

Suite à une erreur dans votre script. Vérifier notamment que `#!/bin/bash` est en première ligne du shell, penser à passer **dos2unix** sur les fichiers windows.

Pourquoi mon job reste-t-il bloqué (en "qw") ?

SGE ne dispose pas des ressources demandées : pour plus d'informations, il est possible d'utiliser `qstat -j jobid`

Comment indiquer le temps d'exécution du job ?

En utilisant l'option `-l h_rt` :

- En ligne de commande : `qsub -l h_rt=02:30:00 job.sge`
- Dans le script : `#$ -l h_rt=02:30:00`

FAQ

Comment demander une certaine quantité de mémoire ?

En utilisant l'option `-l h_vmem` avec la quantité de mémoire demandée
par slot : `#$ -l h_vmem=4g`
Attention : le job sera tué en cas de dépassement !

Plan

1 Accès aux ressources

2 Environnement de travail

3 Soumissions de jobs

4 Contacts

Contacts et informations

- **Site officiel :** `http://meso.univ-fcomte.fr`
- **Wiki utilisateur :** `http://mesowiki.univ-fcomte.fr`
- Liste de diffusion : `meso-utilisateurs@univ-fcomte.fr`
- En cas de problème : `svpmeso@univ-fcomte.fr`
- **Citer l'utilisation / Reporter vos publications au mésocentre**

Questions ?