

05/04/2016

Accueil

Présentation du CINES

Rappel de commande de base Linux

Utilisation de l'OOCIGEN

Pause Café

Code Gaussian

06/04/2016

Accueil

Code Gaussian

TP

Pause Café

TP

Visite de la machine d'OOCIGEN

Connexion et répertoire courant :

\$ ssh **votre_identifiant@server**

Windows : Putty, Cygwin, Xming

\$ pwd : affiche le chemin d'accès du répertoire courant

\$ ls : affiche le contenu du répertoire courant

\$ ls /home/.../

\$ ls *com

\$ cd : change de répertoire

\$ cd ../../, cd ../ , cd ~/

\$ mkdir -p nom_répertoire : créer un répertoire

Rappel de commande de base Linux



Édition de fichiers :

\$ Vi nom_fichier : éditeur de base
deux modes : commande et insertion
\$ Vim, emacs

En général

\$ echo "chaîne caractère"
\$ echo "toto" | tee nom-de-fichier
\$ var=toto
\$ ech \$var
\$ date
\$ who :nom de connexion, terminal,
heure de connexion, nom d'hôte distant,
ou numéro de terminal X
\$ whoami
\$ whereis
\$ top
\$ kill
\$

i	activer le mode insertion
a	Idem i mais après le curseur
:w	sauvegarder le fichier
:wq	sauvegarder le fichier et quitte vi
:q !	forcer vi à se terminer
/toto	rechercher mot en descendant
?toto	rechercher mot en montant
n/N	répéter la recherche dans le même sens / sens opposé
:%s/tot/col/g	changer tot par col dans tout le fichier.
:1,5s/tot/col/g	changer toto par coco dans les lignes de 1 à 5.
r	remplacer un caractère
u	annuler la dernière opération
y	copier une ligne
p/P	colle les lignes après/avant le curseur
dd	supprimer une ligne

Visualisation de fichiers :

\$ cat nom_fichier : affiche le contenu du fichier texte

\$ more nom_fichier

\$ head -n fichier

\$ tail -n fichier

\$ wc -lwc nom_de_fichier

Manipulation de fichiers

\$ cp fich_source fich_dest

\$ cp chem_fich_source chem_fich_dest

\$ mv fich_source fich_dest

\$ mv fich_source rep_dest

\$ rm -i fichier

\$ chmod -option(ugo+/-rwx) nom_fich

\$ diff fich1 fich2

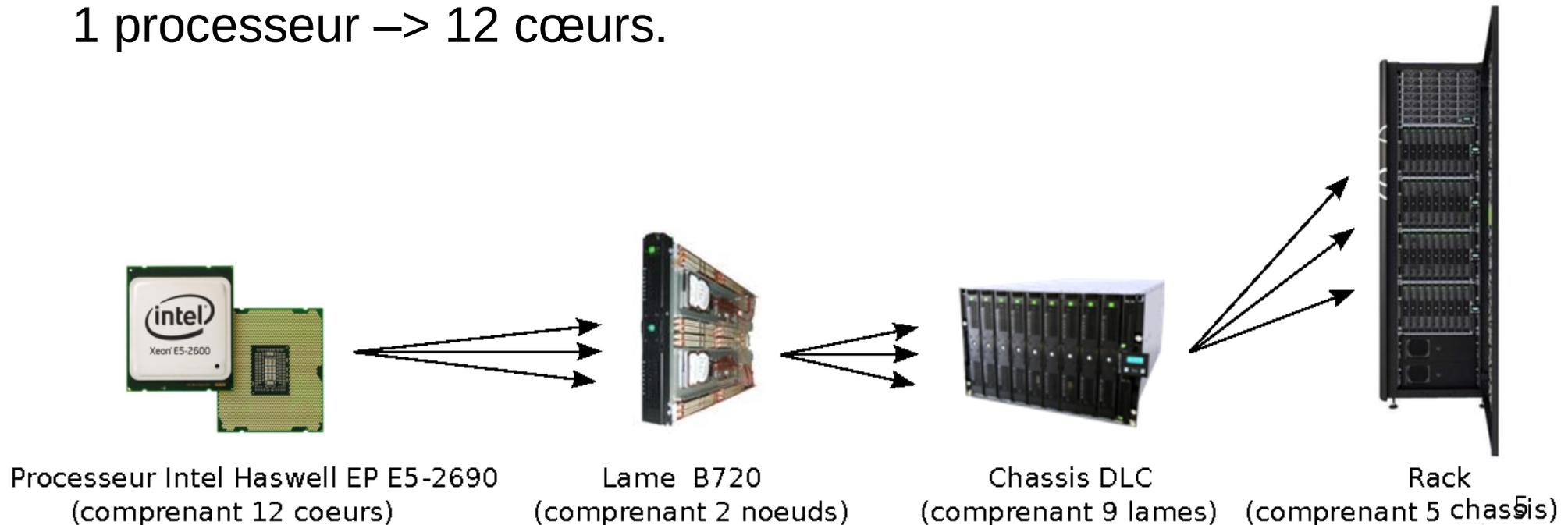
\$ grep 'motif' fichier

\$ find /chemin -critère (name, type, size, group) expression

Utilisation d'OCCIGEN

Occigen : un supercalculateur parallèle, Puissance 2.1Pflo/s.
Acquise par GENCI, fournie par Bull,
maintenue, hébergée, administrée par CINES.

- 5 nœuds de connexion (50, 51, 52, 53, 54).
- 2106 nœuds de calculs
50 544 cœurs (24 racks):
 - 1 rack → 5 châssis ; 1 châssis → 9 lames.
 - 1 lame → 2 nœuds ; 1 nœud → 2 processeurs.
 - 1 processeur → 12 cœurs.



Utilisation d'OCCIGEN



Occigen :

- **CPU :**

Intel Haswell EP v3 E5-2690

12 cœurs, 2.6GHz, 500Gflop/s

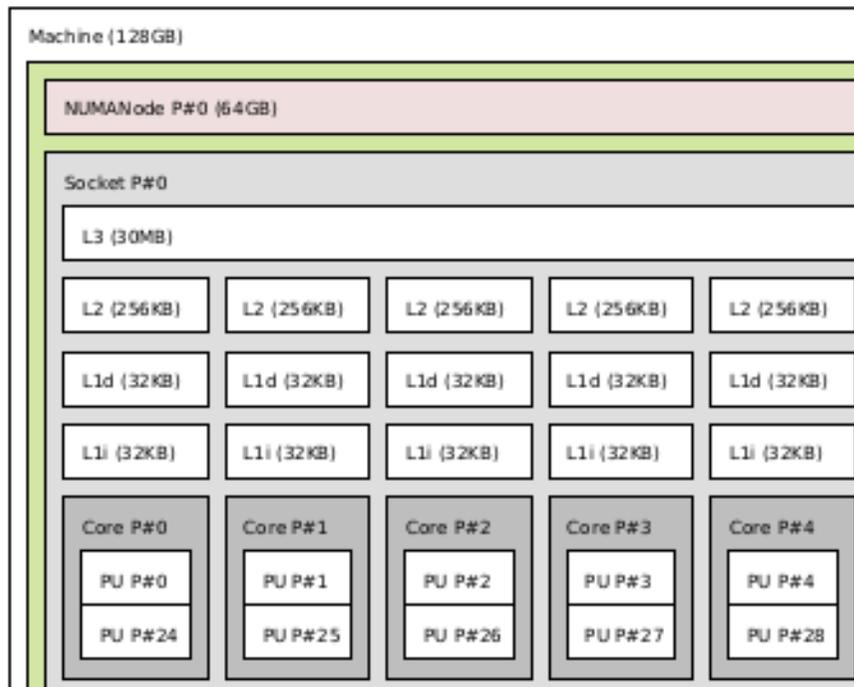
3 niveaux de cache :

2X32Ko/core ;

1X256Ko/core ;

1X 30Mo/sokets.

Bande passante mémoire 68.2Go/s



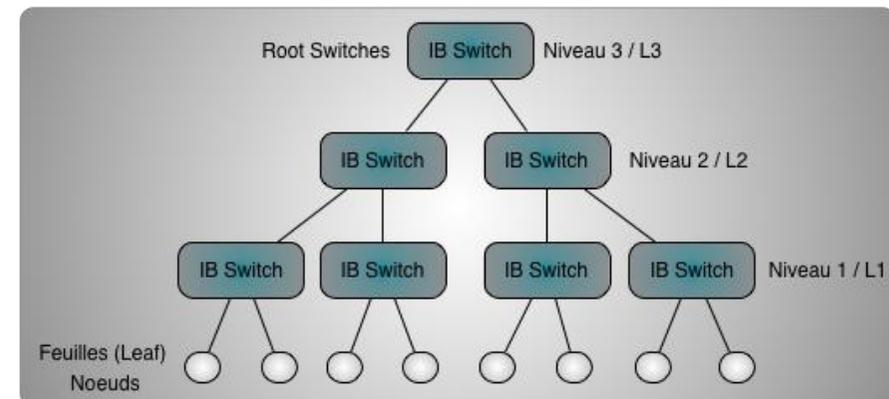
- **Réseau :**

Interconnexion Infiniband FDR :

Débit 56 Go/s

Topologie: Fat Tree

Latence inférieure à 3 μ s



Architecture matérielle:

- **Répertoires de fichiers :**

- /home (montage PanFS – 240To – débit 5Go/s)
 - /scratch (montage Lustre – 5Po – débit 106Go/s)
 - /store (montage Lustre – 2Po – débit 50 Go/s)

- **Espaces de données :**

- 3 espaces personnels : /home, /scratch et /store.

- Accès rapide via des variables :

- cd **\$SCRATCHDIR, \$STOREDIR, \$HOMEDIR.**

- 2 espaces partagés au sein du groupe :

- /sharedhome et /sharedscratch**

- Accès rapide via des variables :

- cd **\$SHAREDSCRATCHDIR, \$SHAREDHOMEDIR**

- Nettoyer régulièrement les fichier inutiles : `rm -ir` ou `rm -i`

- Sauvegarder les données : `tar zcvf votre_archive.tar.gz votre_dossier/`

Architecture matérielle:

- **Espaces de données :**

3 espaces disponibles : /home, /scratch et /store.

Nom	Quota	accessibilité	Perf	Sauvegarde	utilisation
/home	40G	Login+Calcul	+	oui	Compilation, tests, stockages : exécutable, librairies
/scratch	4T	Login+Calcul	+++	non	Lancement des jobs, stockages temporaire,
/store	100000 fichiers	Login+Calcul	++	oui	Stockages des résultats et données

Accès et transfert de données:

- **Accès :**

Windows : Putty, Cygwin, Xming

Linux : ssh login@occigen.cines.fr (option -X si nécessaire)

- **Transfert :**

Du CINES vers l'extérieure : autorisation nécessaire

De l'extérieure vers CINES : possible, via scp, sftp, gftp

Environnement logiciel:

- **module** : permet d'utiliser et afficher les librairies et logiciels installés sous /opt/software :

module av : afficher tous les modules disponibles

module load : charger une librairie ou un code à utiliser

module rm : décharger un module

module purge : décharger tous les modules

module list : afficher les modules chargés dans votre compte

module show : afficher les infos d'une librairie ou un code.

Slurm : gestionnaire de travaux

• Fonctionnement général :

→ remplissage de la machine

→ partage équitable des ressources informatiques via 2 mécanismes :

Fair share : participer au calcul de la priorité du job.

Backfill : optimiser le remplissage de la machine sans impacter la priorité

établie des jobs

→ Temps d'attentes impossible à prévoir

→ Durée maximale par défaut d'un job est 24h

→ Si checkpoint/restart de votre job est impossible :

Demande auprès de svp@cines.fr de passer au job long > 24h :

• Différentes partitions :

all : cas normal

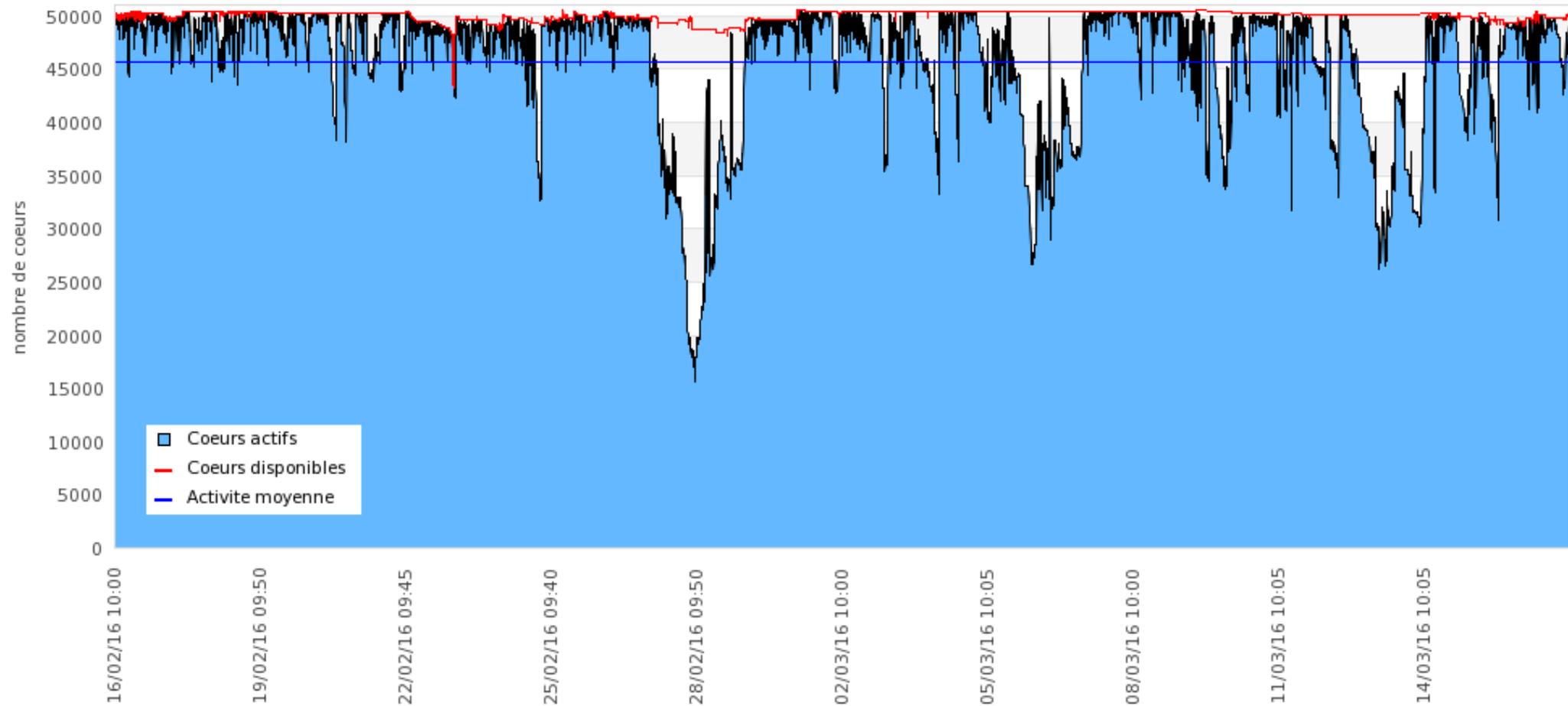
allc : job court

shared : nœuds partagés

BLOCKED_SCRATCH, _STORE, _WORK : dépassement de quota

Slurm : Aperçu de l'activité d'occigen

occigen, Bullx DLC, [50544 coeurs] [du 16/02/16 10:00 au 17/03/16 09:55]
Disponibilite: 99.1%(50080 coeurs) Activite: 90.5% (45741 coeurs)



Slurm : Exécution

- **Plusieurs mode de soumission :**

 - Mode batch : `sbatch script.slurm`

 - Mode interactif : `salloc -N 1 -n 10 -t 10:00 "commande", exit.`

- **Commandes de suivi :**

 - Suivi des jobs : `squeue -u login, watch squeue -u login`

 - Détails d'un jobs : `scontrol show job jobID`

- **Commandes d'actions sur les jobs :**

 - Suspendre un job de la queue : `scontrol hold jobID`

 - Remettre un job dans la queue : `scontrol release jobID`

 - Supprimer un job : `scancel jobID`

- **Décompte des heures consommées :**

 - En mode exclusif : `nombre_nœuds * 24`

 - En mode partagé : `1 * Max[nombre_cœurs, mémoire/5]`

Utilisation d'OCCIGEN



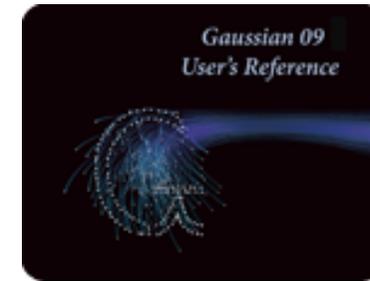
Site web du CINES

<https://www.cines.fr/>

The screenshot shows the homepage of the CINES website. At the top, there is a navigation bar with links for 'Marchés publics', 'Historique', 'FAQ', and 'Français'. A search bar is located on the right. The CINES logo is prominently displayed. Below the logo is a horizontal menu with links for 'Accueil', 'Présentation', 'Calcul', 'Archivage', 'Services', 'Formations', 'Actualités', 'Annonces', and 'Contact'. A 'FLASH INFO' section features a large image of a server room and a headline 'Archivage pérenne' with a brief description. To the right of this section is a list of links: 'Recrutement et stages', 'Documentation technique', 'Ressources', 'Comment calculer au CINES?', 'Comment archiver au CINES?', and 'Espace utilisateur'. Below this is an 'ARTICLES' section with three featured articles: 'Forum des archivistes : 30 mars - 1er avril 2016, Troyes', 'Corriger vos fichiers PDF avec Facile', and 'Deux nouveaux services versants entrent en production sur le site d'archivage pérenne du CINES'. At the bottom, there are four expandable sections: 'Le CINES', 'Le CINES & L'Europe', 'Comité Des Utilisateurs', and 'OCCIGEN : Le Nouveau Supercalculateur'.

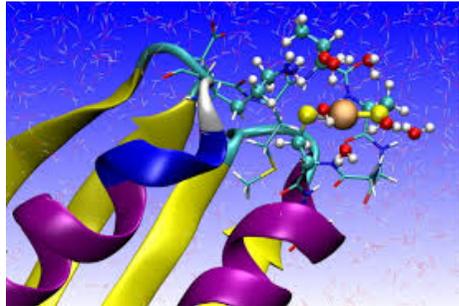
Gaussian

code numérique de chimie théorique

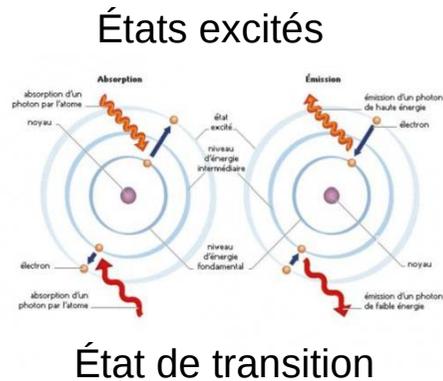


permet :

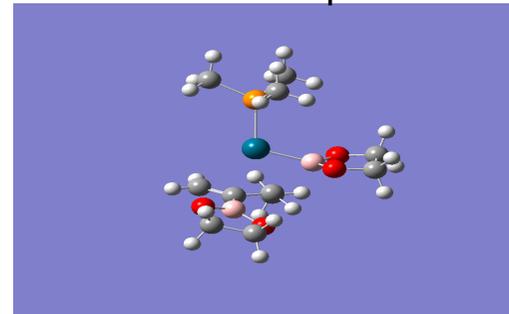
Réactions enzymatiques



Mécanismes réactionnels



Calcul de fréquence



Optimisation de géométrie

en utilisant :

les méthodes : ab initio (MP2, HF,...) , DFT (B3LYP,...), méthodes composites (CBS-QB3),...

les bases : 3-21G, 6-31G, 6-311G,...



Utilisation optimale du code Gaussian sur OCCIGEN

Points problématiques du code Gaussian :

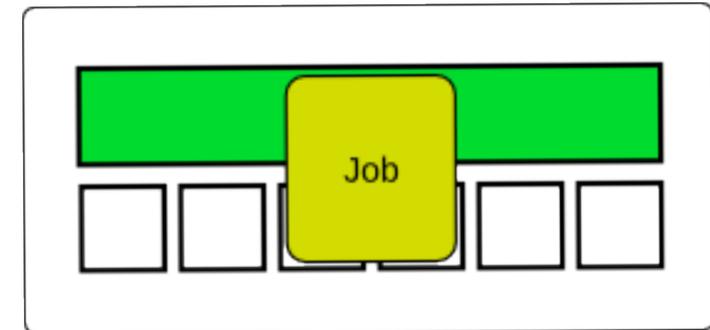
1) Faible scalabilité :



OpenMP ;



gaspillage des ressources informatique.



2) Gestion des fichiers de sortie :

Fichiers inutiles pour les utilisateurs.

3) Très demandés (voir l'article de RFCT)

Quels sont les codes de chimie :

que vous avez cités dans votre demande ?

qui devraient être disponibles sur les centre nationaux, locaux ou régionaux ?

1^{er} Réponse : **Gaussain avec un % important**

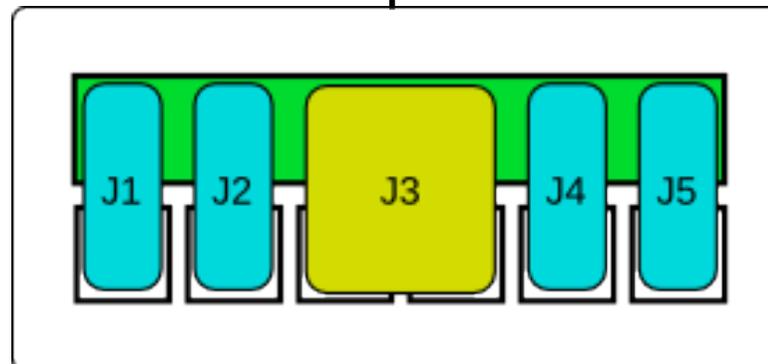
Solution efficace : retravailler le code source Gaussian :
MPI ou hybride MPI-OpenMP ;
Fichiers d'entrée et fichiers de sortie ;
Suppression des fichiers inutiles.

Solutions proposées par CINES :

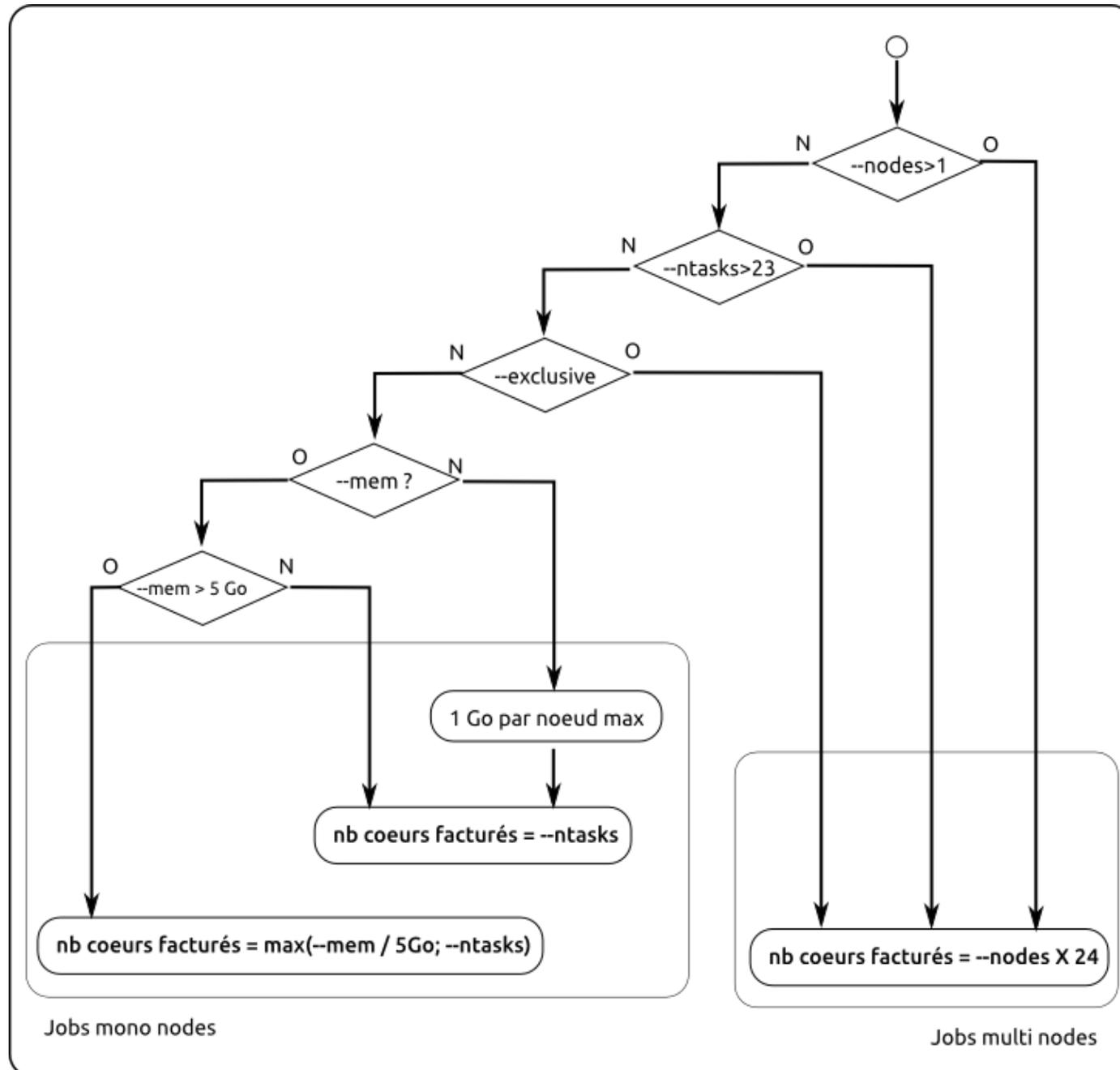
Bonne pratique de l'utilisation Gaussian sur OCCIGEN ;
Nœuds partagés ;
Formations au cours de l'année ;
Équipe Support : compétences en chimie théorique.

Nœuds partagés : plusieurs jobs peuvent tourner de façon simultanée dans un nœud. L'intérêt de nœuds partagés réside dans :

- ➔ la facturation \sim (nbre cœurs, mémoire) ;
- ➔ l'exploitation de toute la capacité de la machine d'OCCIGEN .



Noeuds partagés



Bonne pratique de l'utilisation Gaussian sur OCCIGEN

- Installer la version binaire G09 Intel EM64T
- 5 systèmes et deux types de calcul
- ~ 150 jobs avec un temps variable (~ min à 130h)

Sulfure de dimethyl	Yperite	Sarin	Molécule Y	Molécule Z
3	7	8	15	26
CBS-QB3 Opt + Freq	CBS-QB3 Opt+Freq	CBS-QB3 Opt+Freq	Opt+Freq	Opt+Freq
CH3-S-CH3	S(Cl-CH2-CH3)2	CH3CH(CH3)OP(F) (O)CH3	C, N, O, S, H	C, N, O, Cu, H

- ➔ Script de lancement ;
- ➔ Fichier d'entrée de Gaussain ;
- ➔ Optimisation des calculs en fonction de nbre de proc.

Script de lancement Slurm :

```
#!/bin/bash
#SBATCH -J job_name
#SBATCH --mail-type=ALL
#SBATCH --mail-user=name_user@server
#SBATCH --nodes=1 # ne pas modifier
#SBATCH --ntasks=n # n compris entre 1 et 24 cœurs.
#SBATCH --mem=1000 <MB> # par défaut 1G
#SBATCH --time=12:00:00
#SBATCH --output= nom_de_votre_fichier
```

Charger le module de Gaussian

```
module purge
module load gaussian
```

Initialisation de Gaussian

```
export g09root
export OMP_NUM_THREADS=$SLURM_NPROCS
source $g09root/g09/bsd/g09.profile
```

Script de lancement Slurm :

Création d'un dossier temporaire de calcul :

```
Export GAUSS_SCRDIR=$SCRATCHDIR/GAUSSIAN/$SLURM_JOB_NAME.$SLURM_JOBID  
mkdir -p $GAUSS_SCRDIR
```

Recopier vos fichiers d'entrée vers le dossier temporaire de calcul :

```
cp -f vos_inputs $GAUSS_SCRDIR
```

Penser à recopier aussi vos fichiers (chk, rwf) s'il s'agit d'un redémarrage :

```
cp -f vos_fichiers $GAUSS_SCRDIR
```

```
cd $GAUSS_SCRDIR
```

Lancement de Gaussian :

```
time g09 < input.file.gjf > output.file.log
```

Quelques informations concernant votre calcul

```
echo "job_id      : $SLURM_JOB_ID"  
echo "name_job   : $SLURM_JOB_NAME"  
echo "nbre_nodes  : $SLURM_NNODES nodes"  
echo "nbre_cores  : $SLURM_NPROCS cores"  
echo "OMP_NUM_THREADS : $OMP_NUM_THREADS"
```

Retourner vers la ligne de commande

```
$ sbach nom_de_votre_script_de_lancement
```

Gestion des fichiers de sortie :

À la main :

Recopier le fichier .log vers le répertoire de soumission

```
cp *.log chemin_vers_votre_répertoire
```

Recopier le fichier chk, s'il existe :

```
cp *.chk chemin_vers_votre_répertoire
```

Si Gaussian a planté ou a été arrêté, recopier aussi le fichier.rwf

```
cp *.rwf chemin_vers_votre_répertoire
```

Attention si vous avez d'autres fichiers de sortie.

Vider le répertoire Scratch

```
rm -ri $GAUSS_SCRDIR
```

Gestion des fichiers de sortie :

Avec le script de lancement :

```
export GAUSS_SUBMIT_DIR=$PWD
```

```
cd $GAUSS_SCRDIR
```

```
time g09 <input.com> output.log
```

Recopier le fichier .log vers le répertoire de soumission

```
cp *.log $GAUSS_SUBMIT_DIR
```

Recopier le fichier chk, s'il existe :

```
if [ -f '*.chk' ]
```

```
then
```

```
  cp *.chk $GAUSS_SUBMIT_DIR
```

```
fi
```

Si Gaussian a planté ou a été arrêté, recopier aussi le fichier.rwf

```
if [ -f '*.rwf' ]
```

```
then
```

```
  cp *.rwf $GAUSS_SUBMIT_DIR
```

```
fi
```

Attention si vous avez d'autres fichiers de sortie

Vider le répertoire Scratch

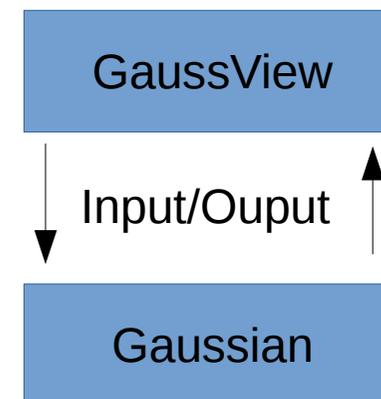
```
rm -rf $GAUSS_SCRDIR
```

Fichier d'entrée de Gaussian :

```
%Mem=2GB          !Link 0 section : Allocation de la mémoire
%NprocShared =12   !Link 0 section : Allocation de la mémoire
%chk=trio.chk      !Link 0 section : Fichier scratch
#P B3LYP/6 – 31 Opt !Route section :

Thiophene         !Titre de la section

0 1               ! Molecule specification
S   0.000  0.000  1.198638
```



Link 0 section :

%Mem : détermine la quantité de mémoire virtuelle que Gaussian va utiliser pour les allocations dynamiques :

- La valeur de **%Mem** doit être identique à la valeur de **#SBATCH –mem**.
- À valoriser avant d'utiliser : GaussView, faire des tests.
- Max ~ 116G

%NprocShared : nombre de cœurs physiques que Gaussian va utiliser pour lancer le calcul :

- La valeur de **%NprocShared** doit être identique à la valeur de **#SBATCH –ntasks**.
- À optimiser avant d'utiliser : faire des tests.
- Max = 24 cœurs.

Fichier d'entrée de Gaussian :

Les fichiers outputs de Gaussian :

Checkpoint file	« name.chk »
Read-Write file	« name.rwf »
Two-ElectronIntegral file	« name.int »
Two-ElectronIntegral file	« name.d2e »
Scratch file	« name.skr »

Par défaut, Gaussian :

Détruit tous les fichiers outputs non-nommés après un calcul terminé normal ;
Garde tous les fichiers outputs après un calcul arrêté ou stoppé anormal

Link 0 section :

%chk : définir le nom et le chemin de fichier.chk

%rwf : définir le nom et le chemin de fichier.rwf

#Save : sauvegarder les fichiers outputs non-nommés même après un calcul terminé normal ;

#NoSave : sauvegarder uniquement les fichiers outputs nommés après un calcul terminé normal. Ex :

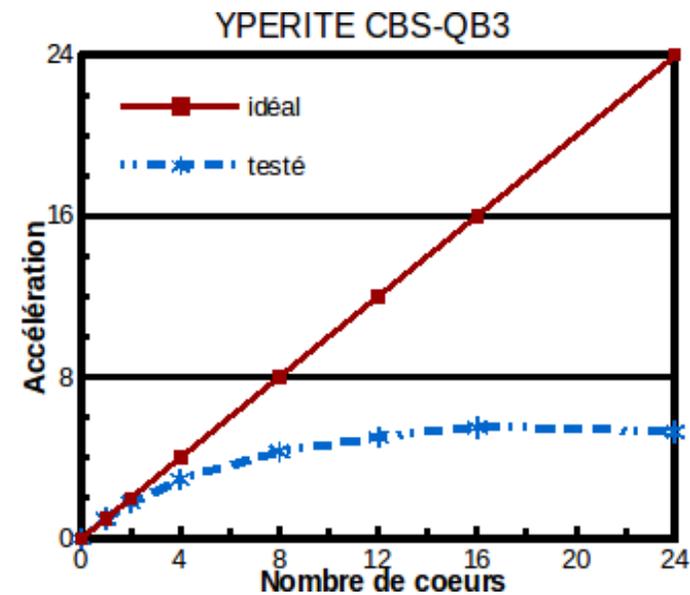
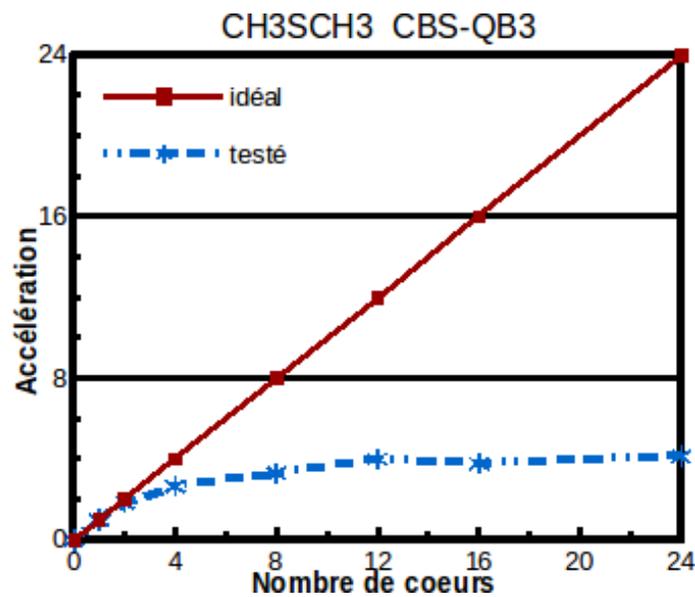
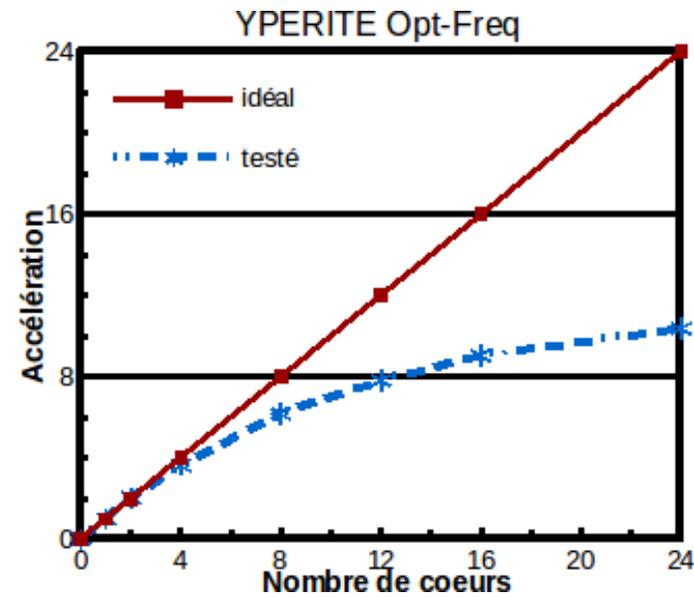
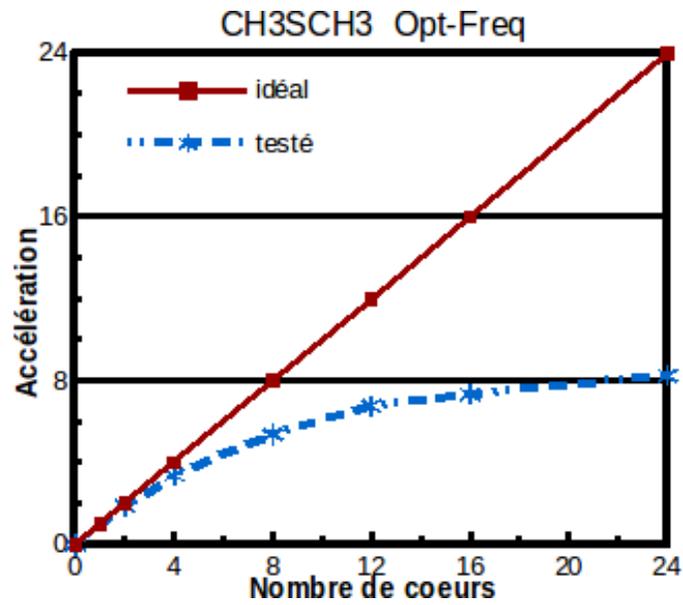
%RWF=/chem/./water fichier sera supprimé.

%NoSave

%Chk=water fichier sera sauvegardé

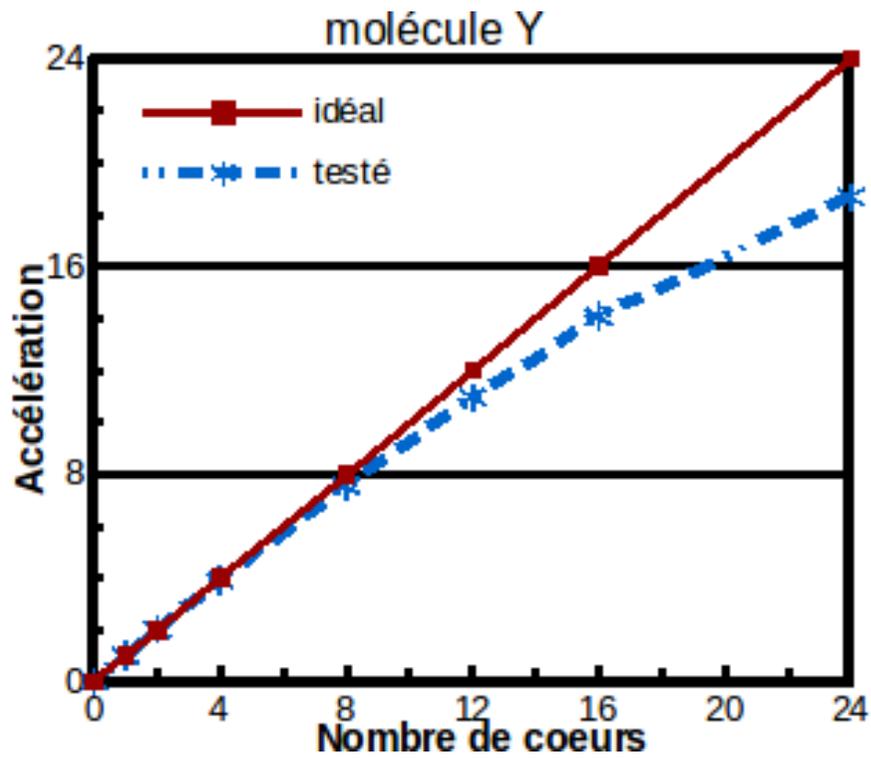
#P : Pour avoir plus d'infos dans le fichier de sortie : infos de la convergence de SCF, temps d'exécution, ensemble de données (exécutable utilisé)

Résultats de l'optimisation



Résultats de l'optimisation

15 atomes, C, N, O, S, H



26 atomes, C, N, O, Cu, H

