

Bonnes pratiques d'utilisation du cluster Agap

Historique des versions

Référence : IN-SCI-Info-004		Gestionnaire : Responsable qualité
Version	date de version	Historique des modifications
Version 01	10/06/2014	Création
Version 02	09/07/2015	Adaptation pour CC2

Sommaire

1. Objectif	1
2. Préambule	1
3. Nœuds de calcul et gestion mémoire	1
4. Espaces de stockage et quota	1
5. Galaxy	2
6. Quelques commandes UNIX	2
7. Fiche de visas.....	4

1. Objectif

L'objectif du document est de décrire les bonnes pratiques de l'utilisateur du cluster de calcul (cc2-login).

2. Préambule

Avant d'utiliser le cluster de calcul, consulter la documentation utilisateur pour la soumission de calculs sur l'infrastructure (cf. doc « Liste des Liens URL »).

Si nécessaire envisager une formation Linux.

3. Nœuds de calcul et gestion mémoire

Limitation mémoire gérée par file d'attente.

Par défaut :

32 Go de mémoire par job sur la queue bigmem.q

4 Go de mémoire par job pour les autres queues (web.q, normal.q)

En cas de dépassement, le job est automatiquement supprimé.

Conseils :

Choisir sa file d'attente en fonction de son besoin mémoire et de la durée présumée du calcul.

Si nécessaire, demander davantage de mémoire pour l'exécution du calcul (paramètre SGE -l mem_free)

4. Espaces de stockage et quota

Le cluster de calcul est organisé autour de deux espaces de stockage :

L'espace NAS (répertoire homedir et répertoires projet) destiné à stocker les données permanentes. Celles-ci sont sauvegardées quotidiennement avec une rétention de 3 semaines.

L'espace work (lien work sous votre homedir, arborescence /work) destiné à stocker les données temporaires de calcul. Cet espace est équipé de disques plus rapides et d'un file-system plus performants pour les calculs.

Limitation de l'espace disque géré par utilisateur ou par projet :

Quota de 150Go par répertoire homedir utilisateur (NAS).

Quotas spécifiques pour chaque répertoire projet (NAS).

Pas de quota en volume pour l'espace work.

Conseils :

Associer l'équipe ID au montage du projet afin de dimensionner le cluster à VOS besoins

Réfléchir dès le début du projet à la politique de gestion et d'archivage des données.

Faire du ménage pour supprimer les données temporaires ou non nécessaires.

Eviter les duplications de données entre utilisateurs et préférer le partage via liens symboliques.

Compresser et archiver les données brutes, si possible sur support externe

Conserver uniquement les FastQ obtenus après les étapes de nettoyage des données brutes

Compresser les données issues d'observations

Extraire uniquement les individus/régions du génome intéressant

Compresser les VCF après analyse

5. Galaxy

Prohiber impérativement l'utilisation d'accents

Ne pas s'affoler à l'affichage de l'erreur « proxy error » : en cas de problème de connexion supprimer les cookies sur son navigateur web

Créer plusieurs historiques et les nettoyer

User et abuser du partage des bibliothèques

Pour être sûr de supprimer une analyse, supprimer l'historique associé

6. Quelques commandes UNIX

1. Compression de données : gzip et bzip2

Gzip

- Algorithme le plus populaire sous GNU/Linux
- Compresser un fichier :
gzip file1
- Décompresser un fichier gzippé :
gunzip file1.gz
gzip -d file1.gz
- Compression optimisée :
gzip -9 file1

- Compresser plusieurs fichiers en un :
`gzip -c file1 file1 > file3.gz`

Bzip2

- plus efficace mais plus lent
- Compresser un fichier :
`bzip2 file1`
- Décompresser un fichier bzipé :
`bunzip2 file1.bz2`
`bzip2 -d file1.bz2`

2. Archivage de données : tar

- Archiver le plus populaire sous GNU/Linux
- Préserve permissions, propriétaires et liens symboliques .
!!! Attention aux liens !!!
- Archiver plusieurs fichiers ou dossiers : option c
`tar cvf archive.tar file1 file1`
`tar cvf archive.tar rep1`
- Lister une archive : option t
`tar tvf archive.tar`
- Extraire une archive : option x
`tar xvf archive.tar`
- 2 options de compression :
z : gunzip
j : bzip2
- Archiver et compresser :
`tar cvzf archive.tar.gz rep1`
`tar cvjf archive.tar.bz2 rep1`
- Décompresser et extraire :
`tar xvzf archive.tar.gz`
`tar xvjf archive.tar.bz2`

3. Suppression de données : rm

- Supprimer un fichier :
`rm file.txt`
- Supprimer un répertoire (et toute son arborescence) :
`rm -r directory`
- Sans demande de confirmation :
`rm -f file.txt`
`rm -rf directory`
- Supprimer une liste de fichiers :
`xargs rm < liste.txt`
- Supprimer les fichiers situés sous le répertoire courant et modifiés il y a plus de 3 jours (cette commande peut également être utile quand rm plante dû à un trop grand nombre de fichiers) :
`find . -type f -mtime +3 -print0 | xargs -0 rm`
- Prudence :
Ne jamais utiliser le compte root



4. Liens symboliques : ln

- Créer un lien symbolique :
ln -s path_to_source my_link
- Supprimer un lien symbolique :
unlink my_link

7. Fiche de visas

	Rédacteur	Testeur	Vérificateur	Approbateur
Nom	B. Pitollat	S. Sidibé-Bocs	P. Turquay	S. Sidibé-Bocs
Fonction	Ingénieur système ID	Responsable équipe ID	Responsable qualité	Responsable équipe ID
Visa				
Date				