

Sommaire

Memento des services GenOuest	2
Création d'un compte	2
En pratique	3
Comment se connecter au cluster	3
Sous linux	3
Sous windows	3
Se repérer sur le cluster	4
Stockage des données	4
Quotas et utilisation	4
Copie de fichiers	4
Installer un nouvel outil	5
Est-ce que cet outil est déjà sur le cluster ?	5
1 à 10 utilisations, comment faire ?	6
Installer votre propre environnement sur le cluster et votre outil	6
Installer votre outil sur le cluster avec Conda	6
Installer votre outil avec un environnement virtuel python	7
Utiliser Singularity pour lancer des containers sur le cluster	8
Autres solutions pour tester des outils	8
Utiliser une VM sur genostack	8
Accès et clés SSH	9
Lancer et se connecter aux instances	9
Utiliser Go-Docker	10
Demande d'installation sur le cluster	10
En résumé comment choisir la ressource à utiliser ?	10
Banques de données disponibles sur le cluster	11
Lancer un calcul / Soumission de job	11
Quelques commandes utiles	12
Comment encapsuler un calcul ?	13
Les services proposés par CeSGO	14
Collaboration	14
Data-access	15
Projects	16
Research sharing	17
Instant	17
FAQ	18

Memento des services GenOuest

Bonjour, vous êtes nouveau-elle sur la plate-forme GenOuest, nouveau-elle dans l'équipe Symbiose, vous avez des besoins en terme d'outils, de données, de stockage. Que peut faire la plate-forme GenOuest pour vous ?

La plate-forme GenOuest a été créée en 2002, elle propose un environnement complet :

- Puissance de calcul
- Capacité de stockage
- Logiciels
- Expertise

Elle est reconnue grâce à différents labels :

- IBiSA depuis 2007
- Plate-forme stratégique de l'INRA (CNOC)
- Membre de l'IFB (Institut Français de Bioinformatique)
- ISO9001 depuis 2008.

Création d'un compte

Rendez-vous sur : <https://my.genouest.org/manager/#/register> (lien également utile en cas de perte de mot de passe). Une fois votre compte créé, n'oubliez pas d'aller faire un tour ici : <https://www.genouest.org/howto/> (ou de continuer votre lecture)

Vous trouverez des réponses à vos questions :

- Comment créer un compte sur la plate-forme ?
<https://www.genouest.org/howto/#account>
- Comment se connecter au cluster ? <https://www.genouest.org/howto/#cluster>
- Quelques informations sur le stockage : <https://www.genouest.org/howto/#storage>
- Comment restaurer des fichiers supprimés ?
<https://www.genouest.org/howto/#snapshots>
- Comment utiliser le cloud (Genostack) : <https://www.genouest.org/howto/#genostack>
- Quelques informations sur les banques de données :
<https://www.genouest.org/howto/#biomaj>
- Quelques informations sur les outils disponibles :
<https://www.genouest.org/howto/#soft>
- Comment utiliser Conda pour installer un nouvel outil ?
<https://www.genouest.org/howto/#conda>
- Comment utiliser Virtualenv pour installer de nouveaux modules pythons ?
<https://www.genouest.org/howto/#virtualenv>
- Comment lancer des jobs sur le cluster SGE : <https://www.genouest.org/howto/#sge>
- Comment utiliser Go-Docker ? <https://www.genouest.org/howto/#godocker>
- Comment contacter le support : support@genouest.org

En pratique:

Pour optimiser son utilisation, notre cluster (genocluster) utilise la soumission de travaux par l'intermédiaire de SGE. SGE fonctionne avec des files d'attente, chacune comprenant un certain nombre de nœuds de calcul. Chaque file a des propriétés spécifiques : temps maximum d'exécution, nombre maximums de nœuds, priorités, etc.

Comment se connecter au cluster

Sous linux

- Création d'une clé ssh, en local dans un terminal :

```
ssh-keygen -t rsa -b 4096
```

Cette commande vous demandera un mot de passe (qui permet de protéger votre clé ssh, également appelé passphrase). Elle va créer deux fichiers :

```
$HOME/.ssh/id_rsa  
$HOME/.ssh/id_rsa.pub
```

- *id_rsa* est votre clé privée et ne doit en aucun cas être divulguée.
- *id_rsa.pub* est votre clé publique. Elle doit être copiée et collée sur <https://my.genouest.org/> (après s'être connecté-e, dans l'onglet "Add public SSH key", ne pas oublier de cliquer sur le bouton Add)

Ensuite vous pouvez vous connecter, à partir d'un terminal en utilisant cette commande :

```
$ssh <your-login>@genossh.genouest.org
```

Sous windows

Plusieurs options sont envisageables et il vous revient de choisir celle qui vous plaît le plus. Le programme Putty est une excellente option :

- Putty : <https://www.chiark.greenend.org.uk/~sgtatham/putty/>
- Tutoriel vidéo fourni par GenOuest: <https://www.youtube.com/watch?v=N3BzP7KiOvg>



Si un mot de passe vous est demandé, c'est que votre clé SSH n'est pas reconnue. (Attention, il ne s'agit pas de la passphrase éventuelle de votre clé)

Assurez-vous que la clé publique soit bien copiée dans l'interface web en une seule ligne et sans espaces dans la partie centrale de la clé (de la forme *ssh-rsa xxxxx nom_de_la_clé*).

Assurez-vous de bien charger votre clé lors de la connexion:

- Linux : le fichier *id_rsa* est chargé automatiquement. Si le nom est différent, indiquez le chemin vers la clé avec l'option *-i* lors du *ssh* (ou avec *ssh-add /chemin/vers/clé*).
- Windows : Assurez-vous de bien indiquer votre clé dans Putty /Pageant

Si le problème persiste, veuillez contacter le support Genouest.

Se repérer sur le cluster

Stockage des données

Une fois connecté-e, vous avez accès à trois répertoires qui sont disponibles sur tous les nœuds de calcul :

- **Votre home** : /home/genouest/<votre-groupe>/<votre-login> chaque utilisateur a un quota de 100Gb. Quelques commandes utiles :
 - Accès direct :

```
cd ~
```

- Utilisation de l'espace disque :

```
quota -s
```

- **Répertoire de groupe** : Vous avez accès à un ou plusieurs répertoires de groupe : /groups/<votre-groupe> qui sont partagés avec vos différentes équipes. Actuellement, il y a un total de 200 To de stockage partagé entre tous les groupes. Chaque groupe possède un quota spécifique.

Un mécanisme de snapshot (sauvegarde à un instant donné) est disponible sur ces volumes, vous pouvez retrouver vos fichiers supprimés par erreur dans le répertoire ~/.snapshots

- **Omaha-beach** : /omaha-beach/<votre-login> vous offre un espace de calcul haute performance associé à un espace de stockage de 120 Go. **Il n'offre pas de stockage de longue durée**, ni de snapshot. Vous pouvez voir votre utilisation avec la commande :

```
pan_quota /omaha-beach/
```

Quotas et utilisation

- En règle générale, les utilisateurs devraient éviter d'écrire les données temporaires des jobs dans les répertoires /home et /groups.
- Ces répertoires sont plus adaptés au stockage qu'à de nombreuses opérations de lecture/écriture. Veuillez utiliser le volume /omaha-beach à la place.
- Nous vous conseillons d'anticiper vos besoins en matière de stockage.
- Si vous prévoyez de générer une grosse quantité de données pour votre projet, veuillez nous contacter au préalable pour vérifier que nous en avons la capacité.
- **Les snapshots ne sont pas des backups. Ils protègent d'une erreur utilisateur, mais pas d'une panne machine.**
- **Veillez utiliser une solution de backup annexe pour assurer la pérennité de vos données.**



Copie de fichiers

- Données distantes

Si vos données sont accessibles depuis un serveur distant, vous pouvez les télécharger en lançant la commande `wget` depuis votre home. (**N'oubliez pas que de nombreuses banques sont déjà disponibles sur le cluster : `ls /db/`**)

- Données locales

Si vous souhaitez transférer des données depuis votre ordinateur local, vous pouvez utiliser la commande `scp`.

Pour récupérer un fichier (du cluster vers votre pc) :

```
scp -p <nom>@<machine>:<fichier_source> <fichier destination>
```

Pour envoyer un fichier (de votre ordinateur vers le cluster):

```
scp -p <fichier_source> <login>@<machine>:fichier_destination
```

Par exemple, pour transférer la donnée locale *data1* dans le dossier *dossier2* de votre home, vous pouvez utiliser la commande suivante:

```
scp data1 login@genossh.genouest.org:/chemin/vers/dossier/2/
```

La commande `scp data1 login@genossh.genouest.org:` copiera la donnée directement dans votre home.

Pour les gros volumes de données, veuillez privilégier un client sftp (par Exemple FileZilla).

- **Interface graphique**
 - Vous pouvez utiliser **FileZilla** avec le protocole sftp pour transférer vos données de la même façon. (En utilisant votre clé ssh comme authentification).
 - Vous pouvez également télécharger des données de petite taille (< 1 Go) dans votre home/omaha-beach/groups via **l'interface web de Cesgo data access** :

<https://data-access.cesgo.org>

Installer un nouvel outil

Quelques questions à se poser avant d'installer un nouvel outil en *qrsh* sur le cluster :

Est-ce que cet outil est déjà sur le cluster ?

- **Liste des outils présents sur le cluster et utilisation**
<https://www.genouest.org/outils/softwaremanager/app.php/>

Les outils installés sont disponibles dans /local. Pour utiliser un outil, vous devez activer son environnement. Cet environnement peut être soit spécifique à l'outil, soit générique à un langage. Par exemple pour charger l'environnement de python 2.7, on utilise la commande :

```
./local/env/envpython-2.7.sh
```

C'est-à-dire que dans chacun de vos scripts, il faudra donc insérer une ligne d'initialisation pour les programmes que vous allez exécuter. Cette commande permet de configurer automatiquement le PATH, les librairies, tout l'environnement nécessaire au bon fonctionnement de votre outil.

Pour lister les environnements disponibles :

```
ls /local/env/env*
```



N'utilisez pas la version de python/perl présente sur les nœuds directement afin

d'éviter des problèmes de version. Veuillez toujours sourcer un environnement.

1 à 10 utilisations, comment faire ?

Installer votre propre environnement sur le cluster et votre outil

Pourquoi créer son propre environnement ? Parce que quand on commence à travailler sur beaucoup d'outils et beaucoup de projets qui n'ont pas tous besoin des mêmes librairies, dépendances, versions de Python... cela peut vite devenir compliqué. Un environnement permet d'avoir une installation propre à un outil et complètement isolée du reste, mais aussi et surtout de ne pas modifier l'environnement des autres utilisateurs. Vous pouvez le créer dans votre home, ou dans les répertoires liés au groupe (ce qui permet de le partager aux personnes qui travaillent avec vous).

Dans tous les cas, commencez par vous connecter sur un nœud de calcul avec la commande :

```
qrsh
```

Vous avez ainsi accès à l'équivalent d'un terminal sur le cluster, ce mode interactif vous permet de tester vos scripts avant de les lancer avec la commande *qsub*, mais aussi de vérifier que toutes les variables d'environnement sont à disposition.

Un *qrsh* simple vous donne uniquement accès à un cœur. Vous n'avez pas non plus de contrôle sur la quantité de RAM disponible.



Il est préférable de vous servir de *qrsh* pour préparer des environnements plutôt que des scripts si ces paramètres sont importants.

Installer votre outil sur le cluster avec Conda

Conda (<http://conda.pydata.org/docs/>) est également disponible pour créer des environnements de travail. Les environnements Conda se comportent de manière similaire aux *virtualenv* python, mais avec des packages Conda au lieu de modules python.

Une liste des packages Conda est disponible ici : <https://anaconda.org/anaconda/repo>

Les paquets Conda sont enregistrés dans des "Canaux". Quand vous effectuez une recherche pour un paquet, vous pouvez voir à quel canal/propriétaire il appartient. Il est nécessaire d'avoir configuré ces canaux pour pouvoir télécharger les paquets associés. Par défaut, les canaux bioconda, conda-forge et R sont configurés sur le cluster.

Le canal Bioconda en particulier est spécialisé dans les outils de Bio-informatique. Vous pouvez ajouter d'autres canaux si besoin.



• Les canaux privés peuvent poser des risques de sécurité. Si possible, utilisez les canaux publics.

<https://conda.io/docs/user-guide/tasks/manage-channels.html>

Pour utiliser Conda, commencez par sourcer le fichier d'environnement (depuis un nœud) :

```
qrsh
. ./local/env/envconda.sh
```

Avec Conda, vous pouvez créer autant d'environnements que vous le souhaitez, chacun contenant les paquets dont vous avez besoin. Il vous suffit d'activer un environnement pour accéder aux outils (un seul environnement peut être actif à un moment donné).

Pour créer un nouvel environnement contenant biopython, deeptools (v2.3.4), bowtie et blast, lancez :

```
qrsh
./local/env/envconda.sh
conda create -p ~/my_env biopython deeptools=2.3.4 bowtie blast
```

Pour l'activer :

```
qrsh
./local/env/envconda.sh
source activate ~/my_env
```

Pour le désactiver :

```
source deactivate
```



- **N'oubliez pas de sourcer Conda avant d'activer votre environnement.**

Installer votre outil avec un environnement virtuel python

Les versions de python disponibles sur le cluster n'ont pas forcément les modules dont vous aurez besoin. Il vous est possible d'utiliser des *virtualenv* pour installer tous les modules que vous souhaitez. Il vous faut d'abord sourcer la version de python que vous voulez (2.7.15 ou 3.6.3) et créer le *virtualenv* (sur un nœud). Dans l'exemple, avec Python 3.6.3, après vous être connecté-e à un nœud grâce à *qrsh* :

```
qrsh
./local/env/envpython-3.6.3.sh
virtualenv ~/my_new_env
```

Ces commandes vont créer un répertoire `~/my_new_env` qui contiendra une version basique de python 3.6.3. Pour utiliser ce *virtualenv*, il vous faut l'activer :

```
qrsh
./~/my_new_env/bin/activate
```

Votre prompt va alors changer, indiquant que vous utilisez le *virtualenv*. Vous pouvez maintenant installer les modules dont vous avez besoin avec la commande *pip* :

```
pip install biopython
pip install py
```

Quand vous avez fini de customiser le *virtualenv*, il vous suffit de le désactiver :

```
deactivate
```



- Il vous suffira juste d'activer le *virtualenv* dans vos scripts pour l'utiliser.
- **Pas besoin de sourcer Python à nouveau.** Vous pouvez directement utiliser votre environnement.
- Vous pouvez créer autant de *virtualenv* que vous voulez et simplement supprimer le répertoire quand vous aurez fini.

Utiliser Singularity pour lancer des containers sur le cluster

[Singularity](#) est une nouvelle technologie permettant l'utilisation de containers sur un cluster de calcul. Comme pour Docker, il est possible de lancer des applications isolées du reste du système. Singularity permet cependant de sécuriser les containers en ne permettant pas l'accès à l'utilisateur root dans le container, permettant ainsi de l'utiliser sur un cluster.

Singularity est pour le moment installé sur les nœuds 31 à 39 du cluster. L'installation sur les autres nœuds de calcul est en cours.



- Singularity ne fonctionnera que sur ces nœuds. Assurez-vous de les sélectionner au moment du lancement du job.

Pour utiliser Singularity, il vous faut sourcer le fichier d'environnement :

```
./local/env/envsingularity-2.4.5.sh
```

Vous pouvez alors lancer un container Singularity, par exemple :

```
singularity run shub://GodloveD/lolcow
```

Vous pouvez également lancer des containers Docker via Singularity :

```
singularity shell docker://biocontainers/bowtie2
```

Il est possible de monter des dossiers dans votre containers (home, ou /db par exemple) :

```
singularity shell -B /db:/db docker://biocontainers/bowtie2
```

Vous pouvez trouver plus de détails sur l'utilisation de Singularity sur le [site officiel](#).

Autres solutions pour tester des outils

Et si finalement vous avez besoin de faire encore plus de tests, et d'installations, et tester de nombreux outils, que vous avez peur de tout casser, vous pouvez aussi (toujours avec votre compte GenOuest) :

Utiliser une VM sur genostack

La connexion se fait avec le domaine Users, votre nom d'utilisateur et mot de passe Genouest :

<https://genostack.genouest.org/dashboard/auth/login/?next=/dashboard/>

Besoin d'aide : <http://www.genouest.org/outils/genostack/getting-started.html>

Vous disposez ainsi d'un environnement avec tous les droits d'installation, avec des quotas d'utilisation par défaut :

- 10 instances
- 50 Go de Ram
- 20 CPUS
- 100 Go de volumes

Chaque instance est montée sur un disque de 20 Go, et expire au bout de deux mois.

Accès et clés SSH :

Veillez vous connecter au frontal openstack.genouest.org au lieu de genossh.genouest.org (en utilisant votre login GenOuest). Pour accéder à vos instances après leur création, vous aurez besoin d'utiliser une clé SSH. Vous pouvez soit copier la clé privée que vous utilisez pour vous connecter au frontal (déconseillé si vous n'utilisez pas de passphrase) ou en créer une nouvelle avec la commande :

```
ssh-keygen -t rsa -C "your_email@example.com"
```

Dans tous les cas, copiez la clé publique associée :

```
cat ~/.ssh/id_rsa.pub
```

Puis connectez-vous à l'interface web de Genostack (<https://genostack.genouest.org>) avec vos identifiants GenOuest. Vous pouvez rajouter votre clé publique dans Project -> Compute -> Key Pairs , en sélectionnant "Import Key pair". Le nom de la clé sert uniquement à la sélectionner dans une liste si vous en utilisez plusieurs.

Lancer et se connecter aux instances :

Pour lancer une instance, dirigez-vous dans Project -> Compute -> Instances et sélectionnez "Launch Instance".

Plusieurs paramètres doivent être remplis avant de pouvoir lancer l'instance :

- Dans le premier onglet, vous pouvez choisir le nom et le nombre d'instances que vous voulez.
- Le second onglet vous permet de choisir la base pour l'instance (OS classique ou customisé)
- Le troisième onglet vous laisse sélectionner la taille en terme de cœurs et RAM de votre instance.

Vous pouvez ensuite lancer l'instance avec le bouton "Launch instance". Après quelques secondes, l'adresse IP de l'instance va apparaître dans l'interface.

Vous pouvez vous connecter en ssh à votre instance depuis la machine openstack.genouest.org avec la commande :

```
ssh root@ip.de.votre.instance
```

(Il est possible que l'instance soit en cours de démarrage, si la connexion SSH est refusée, veuillez attendre quelques secondes avant de recommencer)



Si un mot de passe (autre que la passphrase) vous est demandé, c'est que votre clé SSH n'est pas reconnue.

- Veuillez vérifier que vous avez bien configuré votre clé publique dans l'interface
- Vérifiez aussi que votre clé publique correspond bien à votre clé privée.
- Si votre clé privée ne s'appelle pas `id_rsa`, vous aurez besoin de la préciser avec l'option `-i` lors de la connexion ssh.

Utiliser Go-Docker:

Go-Docker est un outil qui permet de soumettre des scripts shell dans des containers Docker disponibles en interface web ou en ligne de commande

Comment y accéder ? <https://godocker.genouest.org/>

Comment ça marche ? <https://bitbucket.org/osallou/go-docker/wiki/Usertutorial>

Un petit exemple : <https://www.cesgo.org/collaboration/docs/go-docker-et-blast/>

Comment soumettre un job ?

<https://godocker.atlassian.net/wiki/spaces/GOD/pages/3309576/NewJobSubmit>

Demande d'installation sur le cluster :

Contact du support GenOuest pour la demande d'installation : support@genouest.org

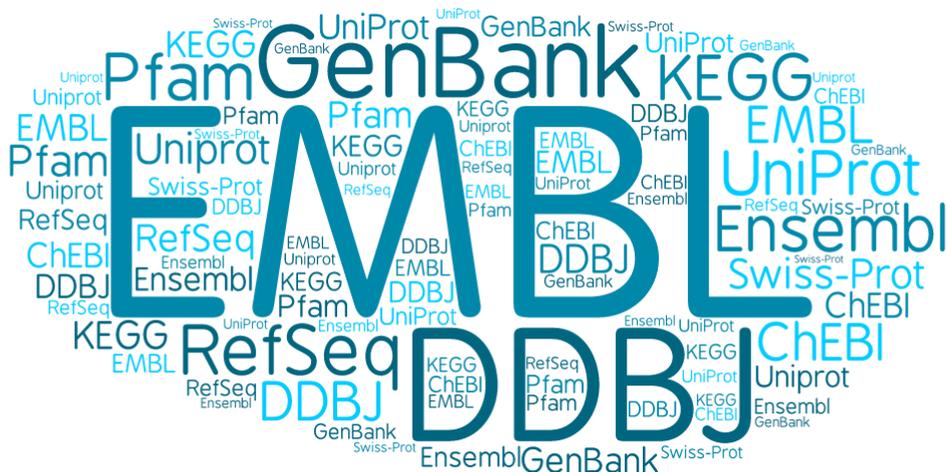
En résumé comment choisir la ressource à utiliser ?

- **Cluster**
 - Cores 800 CPUs / Storage 1200 Tb
 - Ligne de commande / utilisation de 'q*' pour les jobs
 - Beaucoup de données
 - Maîtrise de la ligne de commande et d'un langage de programmation
- **Galaxy**
 - Cores 62 CPUs / RAM 1000Gb
 - Jobs créés via l'interface web / Nombreux outils disponibles
 - Interface web
 - Echange de workflows
- **Cloud**
 - Cores 384 CPUs / RAM 2Tb / Storage 3Tb
 - Basé sur OpenStack / root sur les machines virtuelles (VM)
 - Besoin d'outils difficiles à installer/Tests d'outils
 - Tests et développements sans avoir peur de tout casser
- **GO-Docker**
 - Cores 112 CPUs / RAM 200Gb / Storage 900Gb
 - Outil pour le lancement de job docker (web ou ligne de commande)
 - Besoin d'outils difficiles à installer/Tests d'outils
 - Tests et développements sans avoir peur de tout casser

Banques de données disponibles sur le cluster

Les banques sont disponibles dans le dossier /db du cluster. Ce répertoire est disponible sur tous les nœuds de calcul, sur la plupart des images (Debian, Ubuntu, Centos) de Genostack et sur Go-Docker. Les mises à jour se font automatiquement, et les banques sont directement indexées grâce à [BioMAJ](#).

La liste des banques présentes sur le cluster est disponible ici (ainsi que des informations complémentaires sur la version et le chemin direct) : <https://banks.genouest.org/app/#/bank>



Lancer un calcul / Soumission de job :



Il est interdit de lancer des calculs directement sur les frontaux (genocluster2.irisa.fr or genossh.genouest.org, les outils n'étant pas accessibles depuis les frontaux).

Pour effectuer des calculs, il est possible de soumettre un job au cluster (*qsub*) ou de se connecter directement à un nœud (*qrsh*). Avec *qsub*, vous devez soumettre un script (incluant de sourcer les outils et environnement nécessaires) et le soumettre avec les options voulues.

De la documentation est disponible ici : <https://wikis.utexas.edu/display/CCBB/sge-tutorial>

Lors de la soumission d'un job, il sera automatiquement dispatché sur un des nœuds de calcul lorsque les ressources seront disponibles.

Les nœuds ont des caractéristiques différentes en terme de CPU (8 à 40 cœurs) et RAM (48 à 512 Go).

Vous pouvez utiliser la commande suivante pour afficher les nœuds disponibles, avec leurs caractéristiques et charge actuelle :

```
qhost
```

La soumission de jobs se fait avec la commande *qsub*. Vous pouvez ensuite monitorer vos jobs avec *qstat*.

Voici un exemple de job :

```
#!/bin/bash
## request Bourne shell as shell for job
#$ -S /bin/bash
## print date and time
date
## Sleep for 20 seconds
sleep 20
## print date and time again
date
```

Écrivez le script dans un fichier *simple.sh* et soumettez-le au cluster avec la commande :

```
qsub simple.sh
```

Le job sera visible avec la commande *qstat* et avec *qstat -j xxx* (xxx = votre id de job).

Une fois que le job est terminé, vous devriez voir deux nouveaux fichiers dans votre home : *simple.sh.oxxxxxxx* et *simple.sh.exxxxxxxxxx*. Ces fichiers contiennent les sorties *stdout* et *stderr* (ils peuvent être vides).

Pour des jobs plus complexes, vous aurez besoin de demander une quantité de ressources pour le job (Par défaut, 1 seul CPU vous est attribué).

- Vous pouvez demander plus de CPUs avec l'option *-pe make*. Exemple :

```
qsub -pe make 8 job_script.sh
```

Cette commande demande que le job soit placé sur un nœud avec 8 cœurs disponibles.

- Vous pouvez également faire une demande en quantité de RAM. Si votre job dépasse cette quantité, il sera arrêté. Cette commande permet de vous assurer que vous serez placé sur un nœud avec au moins cette quantité de RAM disponible. Sans préciser cette limite, vous n'aurez pas de contrôle sur la quantité de RAM disponible pour votre job. Si votre job nécessite 15 Go de RAM, il vous faut utiliser l'option *h_vmem* comme suit:

```
qsub -l h_vmem=15G job_script.sh
```

- Vous pouvez combiner une demande en cœurs et RAM. Gardez en mémoire que votre job pourra rester en attente plus longtemps selon vos demandes.



Les ressources sont communautaires. Si vous utilisez beaucoup de ressources pour un temps très long, merci de nous contacter avant.

Quelques commandes utiles

De nombreuses commandes sont disponibles pour monitorer vos job sur le cluster :

- Suppression d'un job :

```
qdel job_id
```

- Suppression de tous les jobs d'un utilisateur :

```
qdel -u user_id
```

- Connaître les nœuds et les ressources disponibles :

qhost

Plusieurs options sont ainsi disponibles pour *qsub* et vous pouvez également les inclure dans le script lui-même. Par exemple :

```
#$ -S /bin/bash
##Force bash l'utilisation de bash.
#$ -l mem=1G
##Demande 1 Go de RAM
#$ -N MadScience_3
##Choisit le nom du job.
#$ -cwd
##Utilise le répertoire de travail actuel
#$ -m be
##Envoie un email au début (b) et à la fin (e) du job
#$ -t 1-1000
##Dit à SGE de lancer 1000 fois le job, numérotés de 1 à 1000.
```

Le numéro du job est accessible depuis le script avec `$SGE_TASK_ID`.

Une liste exhaustive des options est disponible ici :

<http://gridscheduler.sourceforge.net/htmlman/htmlman1/qsub.html>.

La plupart des options liées aux ressources et environnements de travail fonctionneront avec *qrsh*.



- **Ne lancez pas de calculs directement sur les frontaux. Ces machines possèdent des ressources limitées et ne servent qu'aux accès au cluster.**
- **Vous pouvez soumettre vos jobs avec un *qsub*, ou vous connecter à un nœud directement avec *qrsh*.**

Comment encapsuler un calcul ?

Chaque calcul doit donc être encapsulé dans un script pour pouvoir être lancé sur le cluster. Un exemple de script est donné ci-dessous:

```
#!/bin/bash
#$ -S /bin/bash
#$ -M <votre-adresse>@mail.fr
#$ -N nom_du_script
#$ -m bea

# initialisation environnement
./local/env/envncbi.sh

# le calcul proprement dit
blastall -p blastx -d genbank -i seq.fasta
```

Les services proposés par CeSGO :

CeSGO propose un environnement virtuel de travail complet pour les chercheurs.

Les services CeSGO incluent :



Créer des groupes



Partager des idées



Echanger



Rédiger et partager des documents



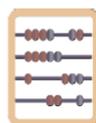
Mettre en place un site web



Gérer et partager des fichiers



Gérer les données scientifiques



Lancer des calculs



Gérer des projets

Collaboration

Collaboration (<https://www.cesgo.org/collaboration>) gère la partie échange de documents et d'informations entre membres d'un même projet (groupe).

Vous pouvez créer votre propre groupe, et inviter d'autres utilisateurs à le rejoindre.

Dans ce groupe, vous pouvez partager et éditer des documents, créer des événements, et échanger des informations via un forum de discussion.

Vous pouvez recevoir sur demande des notifications basées sur l'activité du groupe.

Chaque membre du groupe peut également créer de manière simple et rapide un site WordPress, qui sera disponible sous l'url <https://www.cesgo.org/mysite>.

Des urls personnalisées (mysite.genouest.org) sont également disponibles sur demande.

LE PROJET CESGO

Pour utiliser les services CeSGO, vous devez avoir un compte sur la plate-forme GenOuest.

CeSGO
collaboration

Echanger avec votre communauté

Plus de détails
Utiliser

CeSGO
data-access

Synchroniser et partager vos données

Plus de détails
Utiliser

CeSGO
research sharing

Stocker et publier vos données de recherche

Plus de détails
Utiliser

CeSGO
projects

Gérer vos projets efficacement avec un kanban

Plus de détails
Utiliser

CeSGO
instant

Discussion directe avec vos partenaires

Plus de détails
Utiliser

 UNION EUROPÉENNE
UNANIEZH EUROPA

 L'Europe s'engage
en Bretagne / Avec le Fonds européen de développement régional

Ici, l'Union européenne cofinance un projet pour mettre en place une structure e-science proposant un environnement virtuel de recherche dédié aux communautés des sciences de la vie du réseau Biogenouest.

ARTICLES RÉCENTS

Bonne année de l'équipe CeSGO 4 janvier 2018

Collaboration simultanée dans data-access 18 décembre 2017

L'accès au données GenOuest maintenant accessible via le service data-access 30 octobre 2017

Formation CeSGO 11 septembre 2017

Nouvelle section 'vidéos' 27 juillet 2017

EVENEMENTS

<< Fév 2018 >>						
l	m	m	j	v	s	d
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	1	2	3	4

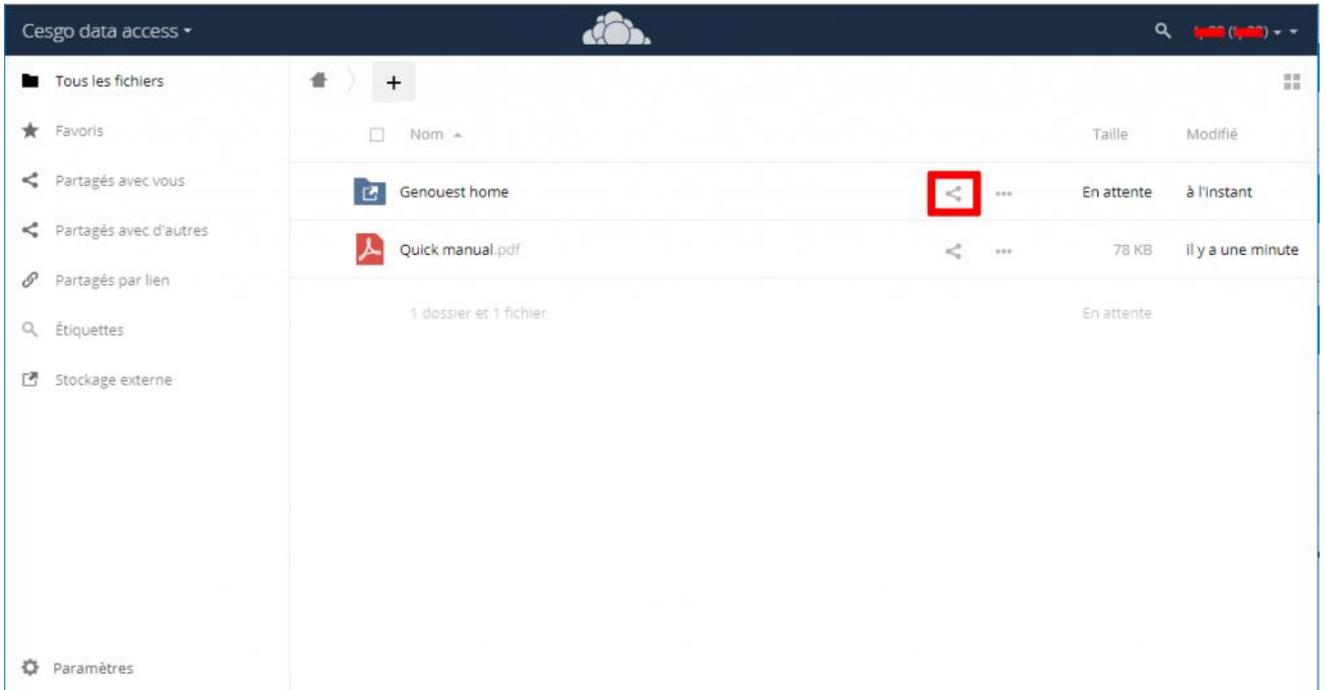
Data-access

Data-access (<https://data-access.cesgo.org>) est une interface web à laquelle vous pouvez lier des stockages externes. Par défaut, vos répertoires home/omaha-beach/groups sont disponibles.

Vous pouvez ajouter vos propres stockages externes, comme un google drive.

Vous pouvez ensuite créer des liens directs (adaptable en terme de sécurité et durée) vers vos données, pour les partager avec d'autres utilisateurs via un lien de téléchargement.

Vous pouvez aussi transférer des données entre deux points de montage.

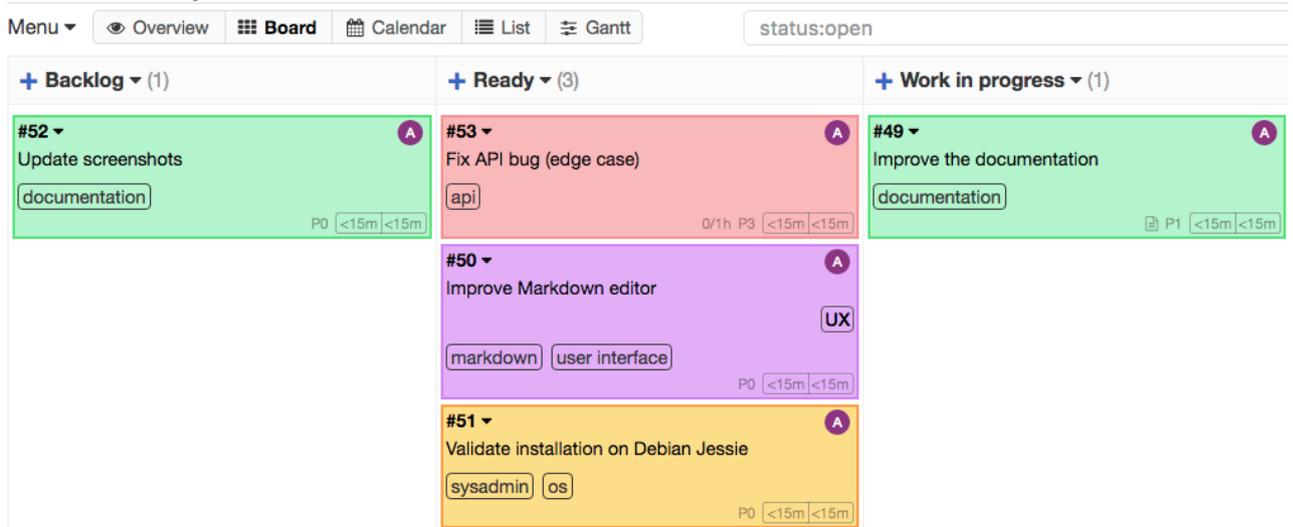


Projects

Basé sur Kanban, Projects (<https://projects.cesgo.org>) offre un interface de type Trello, permettant de créer et gérer les tâches de vos projets.

Vous pourrez ensuite partager/attribuer ces tâches à d'autres utilisateurs.

KB Demo Project



The screenshot shows a Kanban board for 'KB Demo Project' with a menu at the top including 'Overview', 'Board', 'Calendar', 'List', and 'Gantt'. The board is divided into three columns: 'Backlog (1)', 'Ready (3)', and 'Work in progress (1)'. The 'Ready' column contains three tasks: #53 'Fix API bug (edge case)' with 'api' tag, #50 'Improve Markdown editor' with 'markdown' and 'user interface' tags, and #51 'Validate installation on Debian Jessie' with 'sysadmin' and 'os' tags. The 'Work in progress' column contains task #49 'Improve the documentation' with 'documentation' tag. Each task card shows priority (P0, P1, P3) and time estimates (<15m).

Research sharing

Basé sur l'outil Seek, Research sharing (<https://seek.cesgo.org/>) est un service permettant de faciliter la gestion de données scientifiques.

Lors de votre première connexion, vous pouvez demander à créer ou rejoindre un projet de recherche. Vous pouvez alors ajouter des éléments à ce projet : (données, modèles, publications..)

Vous pouvez également ajouter des expérimentations biologiques (tests, études..)

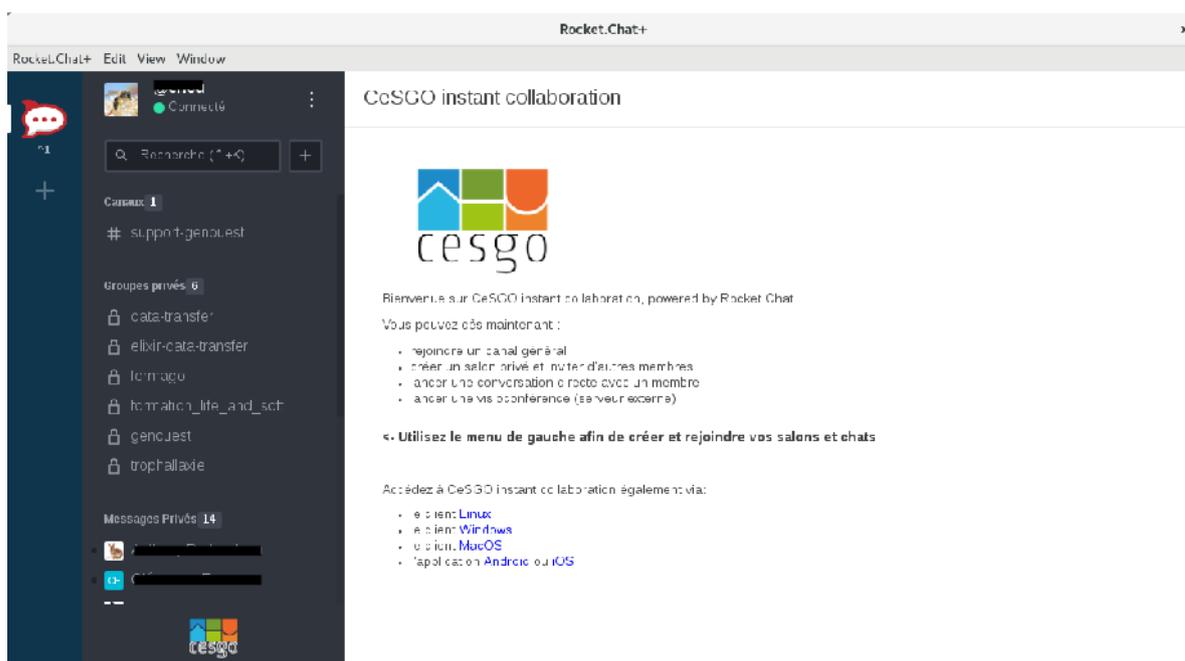
Ces données peuvent être partagées de manière contrôlée avec d'autres utilisateurs.

En utilisant les liens de partage Data-access, vous pouvez simplement ajouter des données de votre home vers Seek.

Seek vous permet d'avoir tout votre contenu scientifique à un seul endroit.

Instant

Instant (<https://instant.cesgo.org/>) est une application de messagerie instantanée, vous permettant d'échanger rapidement avec d'autres personnes.



Vous pouvez trouver plus d'information sur CeSGO à l'adresse suivante :

<https://www.cesgo.org>

FAQ

Comment récupérer toutes les variables d'environnement dans un script à soumettre en *qsub* ?

```
# Export de toutes les variables d'environnement  
#$ -V
```

Ou directement dans la commande *qsub*.

Comment récupérer un fichier supprimé ?

Un mécanisme de snapshot est en place sur les volumes `/home/genouest/`, si vous effacez un fichier par erreur vous pouvez le récupérer dans le répertoire `~/snapshots`. (en le copiant via la commande *cp*)

Comment déployer 10 fois le même container sur Go-Docker ?

Utilisation de l'API (avec *godjob*)

[Tutorial](#)

Comment supprimer tous ses jobs en attente ?

```
for id in `qstat -u <user> | grep qw | cut -d" " -f1`; do qdel $id; done
```

Comment copier de gros volumes de données sur le cluster ?

Il est préférable d'utiliser un client sftp (par exemple FileZilla).