

Lettre d'information

Cette lettre d'information est destinée aux membres des équipes utilisant la plate-forme Genotoul-Bioinfo. Elle a pour but de répondre aux questions et commentaires que vous nous avez fait remonter via le questionnaire de satisfaction annuel et de vous parler des évolutions de l'équipe, les nouveaux outils, services, conditions d'utilisation, projets et formations mis en place.



Logiciels

Vous cherchez à lancer un outil ? Pour vous aider, nous avons mis en place des scripts pour trouver les modules à charger :

search_module NomDuLogiciel (ou une partie du nom du logiciel)

De la même façon vous pouvez trouver les packages python :

search_Python_package
NomDuPackage

Et pour les packages R :

search_R_package NomDuPackage

Attention dans ce dernier cas, la recherche est sensible à la casse.



Bonnes pratiques

Nous avons commencé un guide de bonnes pratiques permettant de réduire l'empreinte carbone des calculs. Nous allons l'enrichir progressivement. Pour le moment sur cette page disponible sur [ce lien](#), nous donnons quelques conseils et nous relayons deux papiers récents sur le sujet. L'un met à disposition un outil pour estimer l'empreinte carbone de vos jobs selon les ressources réservées (Lanelongue *et al*, 2021), et l'autre a comparé l'empreinte carbone de différents logiciels de bioinformatique (Grealey *et al*, 2022). Pour ajuster au mieux vos paramètres de réservation : utilisez la commande `seff <jobid>` sur 1 job test en état COMPLETED.

Retour sur l'enquête de satisfaction

Vous avez été nombreux (119 soit plus de 11%) à répondre à notre questionnaire de satisfaction annuel et nous vous en remercions chaleureusement. Suite à vos retours nous avons pris plusieurs décisions telles que : remettre à jour la formation à l'utilisation des workflows nextflow, proposer de nouvelles visioconférences d'informations courtes (1 heure) sur des points précis... Nous avons aussi remis le prompt en couleur sur le cluster.

Vous nous avez également demandé la possibilité d'avoir un Jupyter notebook et un Rstudio en interactif sur le cluster. Nous avons récemment mis ce service en place. Pour le moment il est nécessaire de faire une [demande](#) via notre site web pour y accéder. Ensuite, suivez les tutoriels sur notre FAQ rubrique [jobs submissions](#) à la question « How to use OOD ? » pour savoir comment les utiliser.

Nous avons aussi acheté et mis à votre disposition 4 cartes GPU A100, certaines sont partitionnées. Pour pouvoir y accéder, il faut, au préalable, avoir fait une demande via notre [formulaire](#). La procédure de réservation est explicitée dans la FAQ rubrique [jobs submissions](#) à la question « How tu use GPU node ? ».

Nous vous rappelons également qu'il est possible de louer de l'espace disque work supplémentaire via une [demande de ressources](#) sur notre site web.

Vous nous avez également parlé de lenteur sur notre site web, il y avait effectivement un problème de montage NFS désormais résolu, cela devrait améliorer les choses. Plusieurs utilisateurs nous ont aussi mentionné que la recherche sur notre FAQ fonctionnait mal. Nous en sommes désolés. Nous avons décidé d'y travailler pour chercher une solution à ce problème.

Vous êtes de plus en plus nombreux à avoir déclaré lire notre newsletter et nous vous en remercions.

Enfin, autre point relatif à la communication, suite au piratage qu'à subi INRAE fin 2023, nos mails étaient filtrés par de nombreux EPST, il nous était alors impossible de communiquer de façon fiable.



Quelques rappels sur le cluster

Un cluster est un ensemble de serveurs de calcul appelés nœuds qui sont interconnectés. Les nœuds de calcul du nouveau cluster sont munis d'un processeur plus rapide que sur l'ancien et la RAM par nœud a été multipliée par 4 (128 cœurs et 2 To de RAM par nœud). Notre cluster est décrit sur notre site web sur cette [page](#) (pour avoir plus de détails, cliquer sur chaque composant de l'image). Le cluster est une ressource partagée entre plus d'un millier d'utilisateurs.

Nous avons remarqué que le cluster n'était pas toujours utilisé à 100 % de ses capacités, car des utilisateurs réservent plus de RAM que nécessaire. Il est donc important de réserver les ressources dont votre job aura besoin au plus juste. Vous trouverez de la documentation utile sur notre FAQ [jobs submissions](#) et en utilisant la commande `man sbatch`.

Lorsque vous lancez vos jobs, ils seront tout d'abord placés en file d'attente et leur priorité de passage sera calculée par l'ordonnanceur (slurm) selon de nombreux critères dont les ressources demandées et leur lancement effectif dépendra également des ressources disponibles.

Les nœuds du cluster peuvent écrire sur l'espace work (performant mais pas répliqué) mais pas sur le save car ce dernier n'est pas adapté. Le save sert à stocker les résultats finalisés après vérification, il est répliqué sur site distant.

Suite à plusieurs incidents récents qui ont nécessité un redémarrage des serveurs frontaux, nous vous rappelons que conformément à notre charte, **aucun traitement (R, test, vscode, awk, etc...) ne doit être lancé sur les serveurs frontaux.**

Cycles d'apprentissage

Où se former en info / bioinfo / biostat ?

En partenariat avec Sigenae et SaAB (MIAT), nous vous proposons des cycles d'apprentissage récurrents comme Linux et Cluster (2 fois dans l'année) ou l'utilisation de sed et awk. Il reste actuellement quelques places pour la session suivante :

Une session « RNASeq, analyse bioinformatique et statistique » de 3 jours et demi du 14 au 17 mai prochain. Les pré-requis pour pouvoir suivre cette session sont Linux, Cluster et avoir une connaissance de base de R.

D'autres sessions seront programmées cet automne via la page « [training](#) » de notre site web.

Les [tarifs](#) et [le formulaire d'inscription](#) sont disponibles sur notre site internet.

La plateforme [GenoToul Biostat](#) met aussi en place différentes formations. Ces cycles d'apprentissage vont de l'initiation à la statistique (avec R ou Asterics) aux réseaux de neurones et deep learning en passant par les modèles linéaires et linéaires généralisés et l'intégration de données. Son catalogue est disponible sur [ce lien](#).

Si vous ne trouvez pas la formation que vous souhaitez, [l'IFB \(l'Institut Français de Bioinformatique\)](#) référence une grande diversité de cycles d'apprentissage en bioinformatique, comme des sessions sur les principes FAIR, la phylogénie ou l'assemblage, le scaffolding et l'annotation en particulier les formations [EBAIL](#). Pensez également aux formations permanentes de vos différents instituts. Elles proposent de nombreuses formations en informatique telles que Python ou Git. Le [CSIESR](#) propose de nombreuses formations en ligne ou en présentiel sur des thématiques ayant trait à l'informatique : Linux, Docker, OpenAI, Python, devops, Git etc. Votre établissement de tutelle doit être adhérent pour que vous puissiez vous y inscrire, mais de nombreux établissements le sont.

De nombreuses ressources existent sur internet telles que des tutoriaux pour apprendre à se servir des notebooks jupyter.

De notre côté, suite à vos demandes, nous prévoyons de remanier et de proposer le cycle d'apprentissage sur l'utilisation des workflows nextflow et nous réfléchissons à mettre en place un nouveau cycle d'apprentissage à propos de Python pour la bioinfo. Nous vous tiendrons au courant via nos lettres d'informations, nos mails et notre compte X (ex-twittr).

Enfin, si vous souhaitez que nous vous accompagnions dans vos analyses (par exemple via un suivi régulier), contactez-nous via le formulaire [Ask for a project](#) sur notre site web pour collaborer avec nous.



Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

@BioinfoGenotoul
<http://bioinfo.genotoul.fr>