

Le groupe ingé de la Plateforme bilille organise un séminaire
mardi 20 mars, de 10h à 12h
à l'**Institut de Biologie de Lille** 1, rue du Professeur Calmette 59000 Lille.



UTILISATION DU FRAMEWORK NEXTFLOW

pour le soin et la recherche en bioinformatique

Fabrice BONTE (CHU de Lille), Frédéric LEMOINE (Pasteur Paris), Christophe DEMAY (CHU de Lille)

- 10h00 : Introduction aux problèmes de reproductibilité/portabilité, à la parallélisation et à la containerisation par Fabrice Bonte
- 10h40 : Utilisation de Nextflow pour l'analyse de données phylogénétiques. Descriptions des avantages de Nextflow et de la courbe d'apprentissage par Frédéric Lemoine
- 11h20 : Utilisation de Nextflow dans l'infrastructure HPC du CHU de Lille : annotation de données de séquençage haut-débit, communication avec une base de données et centralisation de logs par Christophe Demay



Centre Hospitalier Régional
Universitaire de Lille



Institut Pasteur

Abstract :

Face à la croissance actuelle des masses de données biologiques et à la complexité des processus d'analyse, l'utilisation d'outils du type gestionnaires de workflows est de plus en plus nécessaire à l'exécution et au partage des analyses bioinformatiques.

La parallélisation et l'informatique distribuée sont devenus indispensables pour s'attaquer à ce type de problème, mais les outils existants sont encore peu utilisés, soit par manque de support, soit à cause d'un modèle qui s'adapte mal aux particularités des analyses, soit à cause d'un apprentissage ardu.

Nextflow est l'un des outils permettant de répondre à ces besoins. Il est basé sur le modèle de programmation par flux de données : il simplifie l'écriture de pipelines parallèles et distribués, et permet ainsi à l'utilisateur de se concentrer sur la logique fonctionnelle du processus d'analyse et d'exécuter simplement des workflows complexes d'analyse de données.

Un script Nextflow est composé d'un ensemble de processus pouvant être écrits dans n'importe quel langage exécutable par la plate-forme Linux (BASH, Perl, Ruby, Python, R, etc). Les entrées et les sorties transitent entre ces processus dans des « channels ».

Nextflow fournit ainsi une couche d'abstraction entre la logique du pipeline et la couche d'exécution, afin qu'il puisse être exécuté sur plusieurs plateformes sans changement.

De plus, il fournit des connecteurs pour les gestionnaires de ressources SGE, LSF, SLURM, PBS et HTCondor ainsi que pour les plates-formes cloud Kubernetes et Amazon AWS.

Compatible avec les technologies de containers Docker et Singularity, Nextflow permet l'écriture de pipelines autonomes, facilement reproductibles pour lesquels la gestion de versions est grandement simplifiée.

Enfin, la plateforme GitHub offre la possibilité de partager des pipelines et de les exécuter rapidement.