

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Présents

Membres de la plateforme :

Jean-François Gibrat (JFG)

Valentin Loux (VL)

Éric Montaudon

Véronique Martin

Valérie Vidal

Sandra Dèrozier

Franck Samson

Membres du CSU :

- Cécile Neuveglise (MICALIS)
- Stéphane Chaillou (MICALIS)
- Sylvain Marthey (GABI)
- Jordi Estellé (GABI)
- Luc Jouneau (BDR & VIM)
- Pierre Nicolas (MaIAGE)
- Olivier Collin (Plateforme Genouest, Rennes)
- Marie-Laure Martin-Magniette (URGV, APT)
- Raphaël Leblois (INRA Montpellier)

Excusés

Éric Guédon (MICALIS), Andréa Rau (GABI)

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Ordre du jour de la réunion

- Présentation générale de la plateforme
- Point sur les ressources fournies par la plateforme
- Points divers
- Questions des membres du CSU

Fichier joint : 20150127-CSU_Migale.pdf

Accueil

Après un tour de table, JFG rappelle les missions du Conseil Scientifique des Utilisateurs :

- Représenter les utilisateurs
- Rôle d'intermédiaire (descendant et montant) entre les utilisateurs et la plateforme.

Le CSU a été composé en demandant aux unités principales utilisatrices de la plateforme de désigner des représentants. Ceux-ci sont majoritairement (cf statistiques d'utilisation de la plateforme) des représentants d'unité INRA de Jouy en Josas, mais il y a également quelques représentants « extérieurs » : Marie-Laure Martin-Magniette pour l'Agro Paris Tech (et l'URGV/IPS2), Raphaël Leblois pour l'INRA de Montpellier. Olivier Collin est invité en tant que responsable opérationnel d'une plateforme proche de Migale, la plateforme GenOuest de Rennes.

Composition du CSU 2015 :

Représentants de **MICALIS** :

- Cécile Neuveglise
- Stéphane Chaillou
- Éric Guédon

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Représentants de **GABI** :

- Andréa Rau
- Sylvain Marthey
- Jordi Estellé

Représentant de **BDR** et de la **VIM** :

- Luc Jouneau

Représentant de **MaIAGE**:

- Pierre Nicolas

Membres **externes** :

- Olivier Collin (Plateforme Genouest, Rennes)
- Marie-Laure Martin-Magniette (URGV, APT)
- Raphaël Leblois (INRA Montpellier)

Présentation générale de la plateforme

VL informe le CSU du départ effectif depuis le 1^{er} janvier de Franck Samson de la plateforme MIGALE. VL prend la responsabilité opérationnelle par intérim de la plateforme.

VL rappelle les missions générales de la plateforme. Il présente la composition de l'équipe en 2014 (~5,25 ETP). La plateforme est labélisée RIO, certifiée ISO 9001 depuis avril 2011. La plateforme est labélisée « Plateforme Stratégique INRA » depuis 2011. Elle appartient au CATI MIAGO et fait partie du pôle APLIBIO de l'Institut Français de Bioinformatique. Au 27 janvier 2015 elle possède 589 utilisateurs identifiés, dont 80% INRA (45 % du centre de Jouy en Josas). Ce chiffre rassemble tous les comptes, mêmes temporaires ou n'ayant pas été accédés depuis longtemps.

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Mise à disposition de ressources :

La plateforme met à disposition une architecture complète de ressources pour la bioinformatique. Pour cela elle dispose de 13 serveurs physiques, d'un cluster de calcul d'environ 580 processeurs et de plus de 150 To de stockage (dont 50 sont sauvegardés). 163 utilisateurs « actifs » ont utilisé la plateforme ces 3 derniers mois (sur les 589 comptes). En moyenne 250 000 jobs sont envoyés sur le cluster chaque mois, pour un total de 5 millions de jobs en 2014. Le cluster est utilisé à environ 45% de ses capacités, ce qui pour un centre de calcul de notre taille est raisonnable. Chercher à améliorer significativement ce chiffre se ferait sans doute au prix de la souplesse d'utilisation. Le top 10 des utilisateurs représente environ 20% de l'utilisation totale du cluster. La file la plus utilisée est la file « short » (calculs de durée inférieure à 12h), ce qui est cohérent avec le type d'utilisation majoritaire (calculs parallélisables par les données). La politique d'arbitrage des priorités sur le cluster est revue régulièrement au sein de la plateforme. Pour l'instant tous les utilisateurs ont la même priorité de départ et le même quota (illimité) d'heures de calcul. L'utilisation des ressources n'est pas homogène sur l'année et on remarque des pics d'utilisation par période (différentes selon les années). La priorisation des calculs suivant les ressources consommées (fair-share) permet de lisser ces pics et de minimiser leurs impact sur les utilisateurs.

En 2014, un gros effort d'investissement (~80 k€) a été porté sur une solution de stockage centralisée (NetApp). Cette solution est en cours de déploiement et permettra une meilleure extensibilité du stockage de la plateforme. De plus, l'administration des espaces de stockage est grandement simplifiée. La migration des « anciens » espaces de stockage vers le nouveau sera faite progressivement, en prévenant les utilisateurs à l'avance.

De nouveaux nœuds ont été installés sur le cluster, en remplacement des nœuds n6x les plus anciens (3 blades M620 de 2*10 cœurs et 96 Go de mémoire). Le rapport nombre de cœur / mémoire est légèrement supérieur à ce qui était proposé auparavant (4 Go).

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Mis à disposition des outils et des banques :

Plus de 140 outils bioinformatiques et 70 packages R sont mis à disposition et maintenus sur la plateforme. 70 banques de données publiques sont mise à jour et mise à disposition dans le répertoire /db accessible depuis l'ensemble de l'infrastructure. Plus de 70 bases de données relationnelles sont hébergées sur les serveurs de la plateforme.

La demande d'ajout ou de mise à jour de banque se fait sur le site de la plateforme, par le biais d'un formulaire.

La mise à disposition d'outils accessibles en ligne de commande s'accompagne ces dernières années d'un effort conséquent pour mettre à disposition une partie de ces outils sous l'interface de gestion de workflows Galaxy. Galaxy est maintenant le cadre privilégié pour la mise à disposition d'outils bioinformatiques à destination d'un public non-expert. Une trentaine d'outils ou de suites logicielles (QIIME, Mothur) sont mis à disposition. Cette offre logicielle a été accompagnée d'une offre soutenue de formation (une cinquantaine d'utilisateurs formés en 2014) et de support applicatif. Plus de 150 personnes ont un compte sur l'instance Galaxy de la plateforme Migale.

Formations :

La plateforme Migale continue son investissement conséquent dans la formation, avec 13 modules présentés en 2014 dans le cadre du cycle « Formation à la bioinformatique par la pratique ». Plus de 150 personnes ont été formées par 18 formateurs différents (en majorité de la plateforme ou de MaIAGE). La plateforme a mis en place avec la Formation Permanente Nationale de l'INRA et la Cellule bioinformatique une formation autour des données RNASeq proposant sur trois jours une formation à Galaxy, l'analyse bioinformatique sous cet environnement et l'analyse statistique sous RStudio. Cette formation, bien évaluée par les participants, sera rejouée à l'automne 2015. Migale continue de proposer un contenu équivalent à cette formation dans son cycle, mais pour un public plus expert.

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Projets :

La plateforme participe aux projets « Investissement d'Avenir » France Génomique, Institut Français de Bioinformatique et BioDataCloud.

Conceptions et développements :

Trois projets sont présentés :

- IGO est un portail web intégratif des bases de données développé à MaIAGE.
- Le projet « truite » a permis la mise en place d'un Genome Browser sur des données privées (et publiques) de l'unité GABI
- Le projet « rat » a eu pour but la mise en place d'un outil de stockage et de manipulation de données complexes et hétérogènes.

Ces trois projets ont demandé un investissement de la plateforme en terme de développement ou de (co-)encadrement de stagiaires de master. Le nombre limité de mini-projets s'explique par les efforts conséquents faits autour du déploiement et d'intégration d'outils dans Galaxy.

Animation et communication :

La plateforme est représentée au bureau des réseaux PEPI Ibis (bioinformatique et statistique) et gestion de données.

La plateforme anime le réseau des bioinformaticiens de Jouy en Josas.

Le lien avec les utilisateurs de la plateforme se fait de deux manières :

- Une lettre d'information (migazette) semestrielle
- La liste user-migale@jouy.inra.fr à laquelle tous les utilisateurs sont dorénavant inscrits.

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

Points particuliers d'information et discussion avec le CSU :

L'unité MIG a fusionné avec l'unité MIAJ au 1^{er} janvier 2015. Migale est donc maintenant une équipe de l'unité MaIAGE (Mathématiques et Informatique Appliquées des Génomes aux Organismes). Cela n'a pas d'impact direct sur les contours et missions de la plateforme.

Dans le cadre de la certification ISO 9001, la plateforme sera auditée mi-avril sur son processus « Conception et développement ».

Migale est volontaire pour être assez rapidement hébergée dans le futur Data Centre INRA Ile de France, la salle machine du bâtiment 233 étant saturée. Cependant, nous manquons d'information et de visibilité pour pouvoir établir un planning prévisionnel.

Le CSU a confirmé l'intérêt qu'il portait à Galaxy et noté son importance pour mettre à disposition de façon conviviale et uniformisée les outils. Cette activité se fait cependant aux dépens de l'activité « mini-projets », l'investissement en temps (développement, maintenance, formation...) étant conséquent.

Les coûts de la plateforme sont en augmentation. Le prix des fluides a augmenté, le financement récurrent de l'INRA par la CNOC (~20 k€) ne suffit pas à les payer. Le différentiel était pris auparavant en charge par l'unité ou par la plateforme par ses financements propres (projets collaboratifs) mais la diminution de ces derniers nous oblige, comme d'autres plateformes à réfléchir à d'autres modèles de financement. VL rappelle que la plateforme GenoBioToul facture ses utilisateurs à partir de 100 000 heures de calcul et fait payer le stockage. Si le premier utilisateur des ressources CPU de la plateforme avait acheté le même temps de calcul chez un fournisseur Cloud Privé (Amazon), celui-ci lui aurait été facturé entre 5k€ et 25k€. L'idée est d'inciter les gros utilisateurs à participer au financement de la plateforme. Celle-ci restera gratuite et ouverte à tous les académiques pour une utilisation raisonnable, mais au-delà d'un seuil qui reste à définir, les utilisateurs seront incités à monter un partenariat ou facturés sous forme de prestation de service. Le CSU accueille favorablement cette proposition. Les utilisateurs seront

COMPTE-RENDU Conseil Scientifique des Utilisateurs de la plateforme MIGALE

Rédigé par : V. Loux

Date : 27/01/2015

informés de cette réflexion qui se déroule également à un niveau nation dans le cadre de l'IFB.

L'année 2015 est une année charnière pour la plateforme avec le départ de deux personnes : Franck Samson en mobilité et Annie Gendrault à la retraite.

Un poste d'IR responsable opérationnel de plateforme est ouvert au concours 2015. Valérie Vidal remplace Annie Gendrault pour la gestion et la mise à jour des bases de données. Cependant Valérie n'est affectée à la plateforme que sur $\frac{1}{4}$ de son temps de travail et toutes les missions d'Annie (conseil, animation,...) ne sont pas reprises.

L'arrivée de VL ainsi que l'ouverture d'un poste MEM IE « analyse de données omiques » au concours 2015 permet l'élargissement des missions de la plateforme vers l'analyse primaire de données NGS. C'est un travail en cours, qui doit se faire en partenariat avec le département MICA afin que l'offre de service à la communauté MEM soit claire et cadrée. Le CSU sera informé des évolutions des missions de la plateforme.