Lancement d'E-Biothon, une plateforme en réseau pour accélérer et faire progresser la recherche en bio-informatique

Paris, France - 19 nov. 2013: Le CNRS, IBM, Inria, l'Institut français de Bioinformatique et la start-up innovante SysFera déploient E-Biothon, une plateforme Cloud expérimentale (1) pour accélérer et faire progresser la recherche en biologie, santé et environnement.

Disposant de 200 téraoctets (10¹² octets) de stockage et d'une puissance de calcul de 28 téraflops (10¹² flops), elle fournira aux chercheurs et à l'ensemble de la communauté scientifique un portail applicatif et une puissance de calcul importante. Ceci permettra d'aborder le traitement des données complexes de la biologie d'aujourd'hui afin de mettre au point les logiciels applicatifs de demain. La plateforme est présentée du 18 au 21 novembre 2013 lors de l'événement majeur du calcul haute performance : le salon Supercomputing (SC 13) à Denver.

La France a toujours été à la pointe dans la recherche médicale, notamment en ce qui concerne les grandes "épidémies" et pathologies de notre temps (SIDA, cancer ou encore diabète). L'analyse génétique et protéomique des virus ou des patients apparaît de plus en plus importante pour aider à découvrir de nouveaux traitements. Les avancées technologiques récentes, tel que les séquenceurs haut-débit, permettent aux chercheurs en biologie d'avoir accès à des quantités gigantesques d'informations brutes (des péta-octets de données sont générées par an) sur la composition des virus, des bactéries ainsi que sur l'espèce humaine. Analyser ces données pour en extraire du sens, est une tâche ardue qui nécessite d'énormes quantités de traitements informatiques.

C'est pour accélérer ces traitements que le CNRS, IBM, Inria, l'Institut français de Bioinformatique et SysFera se sont associés (2) pour mettre à la disposition des chercheurs cette plateforme de Cloud, hébergée à l'Idris (3), le centre du CNRS pour le calcul numérique intensif de très haute performance, situé à Orsay. Associant un portail applicatif et une puissance de calcul importante, elle permettra de mettre au point les logiciels et les applications qui permettront d'accélérer la recherche en biologie et en santé, en particulier en génomique et en protéomique. L'objectif est de faire progresser plus rapidement la compréhension des maladies génétiques, notamment les maladies neuromusculaires et d'accélérer drastiquement la découverte de nouveaux traitements de rupture. Elle vise aussi à accélérer la recherche en écologie-biodiversité afin de mieux comprendre notre environnement.

La plateforme est constituée de systèmes haute performance IBM Bluegene/P représentant une puissance de 28 téraflops associée à 200 téraoctets de stockage, et de la solution SysFera-DS qui offre aux utilisateurs un portail web d'accès aux ressources de calcul. À travers ce portail, les chercheurs ont accès à tout un environnement de travail leur permettant d'exécuter simplement les traitements informatiques en lien avec les analyses dans les domaines de la génomique, protéomique et métabolomique, puis de gérer les données générées, tout cela à partir d'un simple navigateur web.

Dans un premier temps, trois applications pilotes ont été déployées, notamment dans les domaines de l'épidémiologie et de la bio-diversité. Après cette phase initiale de déploiement, l'objectif est maintenant d'ouvrir cette plateforme soutenue par France Grilles et l'Institut français de Bioinformatique, à l'ensemble de la

communauté scientifique.

Notes

(1) Le Cloud computing est l'accès via un réseau de télécommunications (souvent internet) à la demande et en libre-service, à des ressources informatiques (un parc de machines, d'équipement de réseau et de logiciels...) partagées configurables.

(2) IBM assure la mise en configuration opérationnelle et le support des systèmes haute performance IBM Bluegene/P, le CNRS héberge et administre ces calculateurs à l'IDRIS et est en charge, avec Inria du support utilisateurs. Le portail d'accès aux ressources déployé est la solution SysFera-DS développée par SysFera.

(3) L'Idris (Institut du développement et des ressources en informatique scientifique), fondé en novembre 1993, est un centre d'excellence en calcul numérique intensif au service des unités de recherche tributaires de l'informatique extrême, tant sur les aspects applicatifs (simulations aux grandes échelles) que sur ceux liés aux recherches inhérentes au calcul de haute performance (infrastructures de calcul, méthodes de résolution et algorithmiques associées, traitement des grands volumes de données, etc.).