

Stage en informatique scientifique (H/F)

Niveau d'études : Master

Durée du stage : 6 mois, démarrage souhaité en janvier 2021.

Sujet : Workflow vs. Dataflow : Concepts, défis et simulation pour le calcul haute performance

Localisation : CC-IN2P3 – Campus de la Doua, Villeurbanne

Candidatures : par email à frederic.suter@cc.in2p3.fr et matthieu.moy@univ-lyon1.fr (joindre CV, relevé de notes et lettre de motivation).

Gratification : 591,51 euros/mois

Contexte

Ce stage d'une durée de 6 mois s'effectuera dans le cadre du projet ANR DARK-ERA (2021-2024) et pourra déboucher sur une poursuite en thèse. Il fera l'objet d'un co-encadrement entre le Centre de Calcul de l'IN2P3 (Campus LyonTech la Doua) et le Laboratoire de l'Informatique du Parallélisme (Campus de Gerland)

Le ou la candidate recherché(e) est en Master 2 en informatique ou en école d'ingénieur avec un profil "Système et Réseaux" et de bonnes compétences en C/C++, Linux et programmation concurrente. La rigueur, l'autonomie et la curiosité sont également des compétences recherchées.

Sujet

Le projet international de radiotélescope *Square Kilometer Array* (SKA) générera des flux de données à des débits sans précédent (plusieurs Terabits par seconde) sur lesquels une chaîne d'algorithmes complexes devra être exécutée en temps réel et ce sans aucune capacité de stockage et pour un budget énergétique de seulement 1 MWatt pour 250 pétaflops.

De telles exigences nécessitent de concevoir une architecture innovante orientée *dataflow* et fortement hétérogène (combinaison de CPU, GPU, FPGA, MPPA, ...)

L'un des objectifs du projet ANR DARK-ERA est de développer un outil de prototypage rapide pour les applications *dataflow* afin de fournir des analyses précoces en termes d'utilisation de la mémoire, de latence, de débit et de consommation d'énergie. Cet outil comblera les forces de deux logiciels de simulation déjà existants : PREESM [1] et SimGrid [2]. PREESM évalue avec précision les performances d'un seul nœud hétérogène tandis que SimGrid simule avec précision les communications entre les nœuds.

L'un des obstacles à ce couplage de PREESM et SimGrid est que si PREESM est adapté aux applications orientées *dataflow* [3], SimGrid est quant à lui historiquement adapté aux applications orientées *workflow* [4]. Dataflows et workflows sont deux manières conceptuellement proches de modéliser des applications exprimant du parallélisme de tâches et du parallélisme de données, mais présentent néanmoins des différences fondamentales.



L'objectif de ce stage sera d'identifier les concepts communs et les différences majeures entre ces deux modèles d'applications, puis de déterminer dans quelle mesure l'approche orientée workflow suivie par SimGrid peut-être adaptée/modifiée/étendue pour permettre la simulation d'applications orientées dataflow. En fonction de l'avancée du stage, la proposition d'un simulateur complet et l'étude d'algorithmes d'ordonnancement pourront également être envisagées.

Environnement de travail

Le candidat effectuera son stage au Centre de Calcul de l'Institut national de physique nucléaire et de physique des particules (CC-IN2P3), situé sur le campus de la Doua à Villeurbanne.

Le CC-IN2P3 est l'un des quatre centres de calcul nationaux français. Il met à disposition de ses 2 500 utilisateurs regroupés dans 80 collaborations scientifiques près de 30 000 cœurs et 340 pétaoctets de stockage. Cette infrastructure est principalement dédiée au stockage et au traitement de données issus de grands instruments scientifiques dans le domaine de la physique, tels que des accélérateurs de particules, des satellites ou des télescopes.

Le CC-IN2P3 a contribué à des découvertes scientifiques majeures lors de la dernière décennie telles que le boson de Higgs par les détecteurs ATLAS et CMS du grand collisionneur d'hadrons (LHC) au CERN, ou la première observation des ondes gravitationnelles par les expériences LIGO et VIRGO.

Références bibliographiques

- [1] M. Pelcat, K. Desnos, J. Heulot, C. Guy, J.-F. Nezan, and S. Aridhi. PREESM: A Dataflow- based Rapid Prototyping Framework for Simplifying Multicore DSP Programming. In Proceedings of the 6th European Embedded Design in Education and Research Conference (EDERC), pages 36–40, 2014.
- [2] H. Casanova, A. Giersch, A. Legrand, M. Quinson, and F. Suter. Versatile, scalable, and accurate simulation of distributed applications and platforms. *Journal of Parallel and Distributed Computing*, 74(10):2899–2917, 2014.
- [3] E. A. Lee and D. G. Messerschmitt. Synchronous dataflow. In Proceedings of the IEEE 75(9):1235–1245, 1987)
- [4] E. Deelman, T. Peterka, I. Altintas, et al. The Future of Scientific Workflows. *The International Journal of High Performance Computing Applications*. 32(1) : 159-175, 2018.

