Étude de l'impact de la topologie réseau et de la co-allocation sur les performances d'applications MPI

Encadrant: Frédéric Suter, Centre de Calcul de l'IN2P3. (frederic.suter@cc.in2p3.fr)
Durée: de 4 à 6 mois (possibilité de rémunération pour 1 à 2 stagiaires)
Lieu de stage:
Centre de Calcul de l'IN2P3
21 avenue Pierre de Coubertin
CS70202
69627 Villeurbanne cedex

Contexte et motivation

Dans de nombreuses disciplines, les résultats scientifiques sont désormais obtenus grâce à l'utilisation intensive de l'informatique qui permet l'analyse de grands volumes de données et l'exécution de simulations numériques de plus en plus précises. Cette augmentation des besoins en calcul haute performance des scientifiques se traduit par un accroissement permanent de la taille et des capacités des grands supercalculateurs. Les machines parallèles les plus puissantes au monde sont ainsi composées de plusieurs millions de cœurs interconnectés par des topologies réseaux de plus en plus complexes (fat tree, tores multidimensionnels, dragonfly, slimfly, megafly, ...). La taille et la complexité des applications s'exécutant sur ces machines sont également en perpétuelle augmentation.

Il devient alors d'une part de plus en plus difficile d'évaluer les performances de ces grands systèmes informatiques au delà des quelques benchmarks utilisés pour réaliser les classements mondiaux mais qui ne sont que peu représentatifs de l'utilisation réelle de ses machines. La diversité et la complexité des applications scientifiques compliquent d'autre part énormément le dimensionnement des futurs supercalculateurs, autrement la détermination de la configuration matérielle la plus efficace et la plus rentable.

Afin de s'affranchir de la complexité des applications parallèles tout en conservant la capacité d'étudier leurs performances, de nombreux efforts ont été entrepris afin de développer des *proxy-applications* qui reflètent les caractéristiques des applications réelles, le schéma de communication ou l'algorithme par exemple, sans s'encombrer de la partie métier de ces applications.

La simulation est une approche permettant de palier les difficultés liées à l'évaluation des performances d'applications distribuées s'exécutant sur des supercalculateurs. Parmi les outils disponibles, SimGrid est une boite à outils dont les principales qualités sont son noyau de simulation efficace et ses modèles réseaux dont la validité a été démontrée. Ce logiciel existe depuis près de 20 ans, dispose d'une base solide d'une dizaine de développeurs et a été utilisé dans près de 500 publications scientifiques. Depuis quelques années, SimGrid dispose d'une interface utilisateur appelée SMPI qui permet de simuler des applications parallèles MPI écrites en C, C++ et Fortran sans modification grace à plusieurs mécanismes complexes au niveau système. Le principe de SMPI est de replier complètement une application parallèle sur un seul processeur tout en gardant une bonne qualité prédictive des performances de cette application.

Sujet

L'objectif principal de ce stage est de combiner plusieurs développements récents autour de SMPI afin d'étudier l'impact de différentes topologies réseaux sur les performances de différentes applications MPI notamment lorsque celles ci sont doivent partager cette ressource réseau. Pour atteindre cet objectif, différents points d'étude seront considérés qui offrent chacun de nombreuses possibilités de recherche et de développements.

- Si certains modèles de topologies réseaux ont déjà été intégrées dans SimGrid, les propositions les plus récentes (dragonfly-plus, slimfly, megafly, ...) restent encore à modéliser, développer, valider et intégrer
- Des efforts récents ont permis de pouvoir exécuter de nombreuses *proxy-applications* avec SMPI. Ces travaux se limitent pour l'instant à la faisabilité de l'exécution simulée. L'analyse, voire l'amélioration, de la qualité des simulations de ces applications reste néanmoins à effectuer.
- Le large éventail de comportements offert par ces proxy-applications peut alors permettre de mener une campagne d'évaluation des performances de ces applications en fonction de la topologie de la machine et donc de déterminer quelle topologie correspond le mieux à quelle classe d'applications.
- Enfin, il est désormais possible d'exécuter plusieurs applications MPI au sein d'une même simulation. Cela ouvre la voie à l'étude de l'impact du partage des resources sur des applications réelles. L'une des perspectives à plus long terme dans cette voie est de pouvoir grace à la simulation guider le placement des applications au sein d'un grand supercalculateur afin de minimiser l'impact de ces applications entre elles.

Ce sujet est vaste et complexe mais à l'avantage d'offrir différents angles d'approche en fonction des goûts du ou des candidats (modélisation, développement, expérimentation, analyse de performance, ...). En revanche, l'un des points communs à toutes les facettes de ce sujet sera la méthodologie à mettre en oeuvre pour mener le stage. La tenue de cahiers de laboratoire électroniques sera ainsi fortement encouragée. L'un des avantages principaux de cette démarche est de favoriser la reproductibilité (et par conséquent la confiance que l'on peut avoir) et le partage des résultats obtenus.

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 petaoctets 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.