

# Support exécutif scalable pour les architectures hybrides distribuées

Master 2 – Sujet de Mémoire de recherche

2011 – 2012

## 1 Contexte

- Thème : Supports et AlgoriThmes pour les Applications Numériques hAutes performanceS (Satanas)
- Équipe : Runtime (INRIA / LaBRI)
- Encadrants :
  - Olivier Aumage, 05 24 57 41 19, [olivier.aumage@inria.fr](mailto:olivier.aumage@inria.fr)
  - Samuel Thibault, 05 24 57 41 43, [samuel.thibault@labri.fr](mailto:samuel.thibault@labri.fr)
  - Emmanuel Agullo, 05 24 57 41 50, [emmanuel.agullo@inria.fr](mailto:emmanuel.agullo@inria.fr)
- Lieu du stage : INRIA Bordeaux – Sud-Ouest, 351 cours de la Libération, bât. A29 bis, 33405 Talence Cedex
- Mode de financement : ANR (projet ProHMPT)
- Mots-clés : parallélisme, gpus, calcul distribué, scalabilité, équilibrage de charge

## Contexte

L'équipe Runtime (INRIA/LABRI) travaille à la conception de supports d'exécution d'applications scientifiques sur des machines de calcul parallèles et distribuées, dans le domaine que l'on nomme *calcul intensif*. Le but fondamental de ces supports d'exécution est d'assurer la *portabilité des performances* des applications en permettant à celles-ci de *maximiser* l'exploitation du potentiel de performances des machines tout en *minimisant* ou en éliminant l'effort d'adaptation nécessaire à l'utilisation de telles applications sur une machine donnée.

Le support exécutif StarPU conçu et développé au sein de Runtime dans le cadre de la thèse de Cédric Augonnet est destiné plus particulièrement à l'exploitation des architectures *hétérogènes*, c'est-à-dire les architectures comportant des unités de calcul généralistes (processeurs multicœurs habituels) associées à des unités de calcul accélératrices (processeurs graphiques/GPU, notamment). StarPU offre un modèle de programmation à base de tâches élémentaires potentiellement parallèles et de relations de dépendances explicites ou implicites entre ces tâches. Le rôle de StarPU est ensuite principalement d'assurer l'ordonnement de ces tâches élémentaires de calcul en prenant les décisions d'affecter telle ou telle tâche sur telle ou telle unité de calcul, et ceci en fonction de divers paramètres tels que les dépendances, les coûts estimés des différentes solutions, etc. StarPU s'occupe également des travaux afférents tels que les transferts de données vers ou depuis les mémoires embarquées des unités accélératrices. Cependant, les travaux sur StarPU ont pour l'instant principalement porté sur l'aspect "intra-nœud". La scalabilité au delà d'une dizaine de nœuds n'est pas garantie.

## Description du sujet

Le cadre de cette étude concerne un noyau de calcul en algèbre linéaire dense qui constitue une partie importante de certains codes de calcul utilisés en production au CEA/DAM/CESTA. Il s'agit plus précisément des factorisations LLt, appliquées à des matrices denses symétriques définies positives. Jusqu'à récemment, les implémentations performantes de ces noyaux de calcul concernaient uniquement les grappes de machines multiprocesseur homogènes. ScaLAPACK ([www.netlib.org/scalapack/](http://www.netlib.org/scalapack/)) est probablement la bibliothèque la plus répandue offrant ce type de service, et est utilisée au CEA/DAM/CESTA. Des travaux récents en collaboration avec l'Université du Tennessee (J. Dongarra), associant des développements au niveau de la bibliothèque MAGMA et du support exécutif StarPU ont permis d'obtenir des performances de tout premier plan sur des algorithmes de factorisation de LLt sur machines multicoeurs hybrides dotées d'accélérateurs NVIDIA.

L'objectif de ce stage est d'étudier les performances de ce type d'approche lorsqu'un grand nombre de nœuds sont utilisés. Le but à terme de cette étude est donc d'élaborer une version parallèle (prototype) pour machines distribuées de la factorisation LLt en s'appuyant conjointement sur StarPU et sur MPI. Il faudra donc, pour cela, travailler à l'optimisation du prototype au niveau distribué et notamment de s'assurer que le prototype puisse fonctionner efficacement sur des grappes classiques et grappes hétérogènes de taille significative (passage à l'échelle), telles que les grappes communément mises en œuvre dans les centres de calcul nationaux. En particulier, dans le mode de fonctionnement actuel, chaque nœud déroule l'intégralité du graphe de tâches. L'enjeu de ce volet sera donc de s'affranchir de cette contrainte tout en préservant la cohérence de fonctionnement de chacun des nœuds vis-à-vis de la session de calcul dans son ensemble.

## Liens et références

- Équipe Runtime : <http://runtime.bordeaux.inria.fr/>
- Support exécutif StarPU : <http://runtime.bordeaux.inria.fr/starpu>
- Bibliothèque MAGMA : <http://icl.cs.utk.edu/magma/>
- Équipe associée Morse avec l'UTK : <http://hiepac.bordeaux.inria.fr/eagullo/morse/>