POLYTECH LILLE

Project proposal for exchange students

Université de Lille 1 SCIENCES ET TECHNOLOGIES

Topic: High Performance Programming and Computing for Big Data Applications

Keywords: Parallel programming and computing, distributed computing, *pagerank* algorithm, linear algebra

Host laboratory : CRISTAL/CNRS (in relation with the « *maison de la simulation de Lille* »)

Supervisor : Prof. Serge G. PETITON

E- Mail : serge.petiton@univ-lille1.fr

http://www.cristal.univ-lille.fr/?rubrique27&eid=26

Laboratory : CNRS/CRISTAL

Description

Exascale *hypercomputers* are expected to have highly hierarchical architectures with nodes composed by *lot-of-core* processors and accelerators. These supercomputers combine processors of different types, namely multi-core processors and massively parallel processors, such as graphics cards (GPU) or INTEL MIC, that combine dozen of cores with high compute performance but with restricted capabilities. Methods have to be redesigned and new ones introduced or rehabilitated in terms of communication optimizations and data distribution.

Data science and sparse linear algebra are two important classes of applications which generate very large scale graphs, or sparse matrices, well-adapted for two-levels parallel and distributed programing paradigms. The *pagerank* algorithm, important for many big data applications, is a good example as we manipulate both large very large unstructured graphs and very large sparse matrices.

The main goal of this master thesis is to use the two-level programming associating YML, developed in France to describe graph of components, and PGAS languages, such as XMP developed in Japan, for data science applications based on *pagerank* methods. Applications for marketing applications will be study as the main example, based on a Hiden Markov chain model.

Level of studies :

Bachelor's level (undergraduate) ☐

Master's level (postgraduate) ■

Duration :

3 months ☐
4 months ☐
5 months ■

Type of evaluation
Written report and oral defense

No funding available from the laboratory.