Sujet :

Responsables : George Bosilca, Thomas Herault at the Innovative Computing Laboratory, The University of Tennessee, Knoxville (http://icl.utk.edu) and Mathieu Faverge, IPB, Bordeaux
Téléphones : +1 (865) 974-6321
Courriels : bosilca@icl.utk.edu, herault@icl.utk.edu, mathieu.faverge@inria.fr

Présentation du sujet :

With the advent of manycore and accelerated computing, machines have become more complex than ever to program and understand. As a consequence, computationally intensive operations, among others, have switched to task-based dataflow and dynamic scheduling approaches, to extract the maximal coarse grain parallelism, and exploit the full computational power of the heterogeneous hardware. While this change is progressing, tools to analyze, debug, and tune the performance of applications must be adapted to the new programming paradigm.

PaRSEC [1,2] is a generic framework that provides task-based dataflow computation on distributed memory machines with accelerators for high performance computing. It embeds tracing capability, allowing the final user to gather arbitrary information either by sampling or per task, and critical timings of scheduling and communication. Once gathered, this information must be synthesized, filtered, and visualized to present a comprehensive, but clear, picture of the application efficiency and bottlenecks. PaRSEC uses a Pandas [3] flexible database representation to allow easy scripting in Python of such filters, aggregators, and conversion tools.

On the visualization front, state of the art tools, like Vite [4], leverage on existing performance analysis tools approach to project the execution trace on a classical Gantt-like diagram, showing the resource usage and the timely distribution of computational task execution. Although this visualization provides a first insight for the end user, it hides critical information, like the data dependency between tasks, which allow a more precise understanding of scheduling decisions. Recent tools, coming from different communities, propose to expose this information through a graph visualization that can be coupled with the Gantt-chart representation. However, these tools are capable of presenting 10s to 100s of tasks, where a typical PaRSEC application will expose hundred of thousands to millions of tasks.

The goal of the internship will be to evaluate filtering and aggregation techniques, and scalable visual representations, to expose this information in a concise and accessible manner to the end user.

Mot-clés : DAG, performance visualization, ViTE

Commentaires : This internship requires good skills in C++, Qt, OpenGL and Python. Basic applied statistical concepts are welcomed. This internship is a joint effort between Inria Bordeaux – Sud-Ouest and The University of Tennessee, Knoxville. The internship location might be shared between the two institutions.

Références :

[1] George Bosilca, Aurelien Bouteiller, Anthony Danalis, Thomas Herault, Pierre Lemarinier, Jack Dongarra: DAGuE: A generic distributed DAG engine for High Performance Computing. Parallel Computing 38(1-2): 37-51 (2012)

[2] http://icl.cs.utk.edu/parsec/

[3] http://pandas.pydata.org

[4] http://vite.gforge.inria.fr