DASH: Data-Aware Scheduling at Higher scale

Spring/Summer 2018

Guillaume Aupy, INRIA Bordeaux
guillaume.aupy@inria.fr

Context In the past, scheduling algorithms have mostly been developed around the constraints linked to computing. While computing power of supercomputers keeps on increasing at an exponential rate, their capacity to manage data movement experiences some limits. It is expected that this imbalance will be one of the key limitation to the development of future High-Performance Computing (HPC) applications. We propose to rethink how the data created by the applications and stored during the computation (also known as I/O for Input/Output) is managed in supercomputers.

As an example, in 2013, Argonne (a research lab in the US) upgraded its house supercomputer: moving from Intrepid (Peak performance: 0.56 PFlop/s; peak I/O throughput: 88 GB/s) to Mira (Peak performance: 10 PFlop/s; peak I/O throughput: 240 GB/s). While both criteria seem to have improved considerably, the reality behind is that for a given application, its I/O throughput scales linearly (or worse) with its performance, and hence, what should be noticed is a downgrade from 160 GB/PFlop to 24 GB/PFlop!

In future large-scale platforms, the way I/O movements are scheduled is more and more critical to optimize performances. In this project we propose to add a layer of data-movements scheduling to the usual job scheduling in super-computers. More specifically, the novelty of this project is to account for known HPC application behaviors (periodicity, limited number of concurrent applications) to define data scheduling strategies.

Internship Program and objectives During the internship, the student will work on:

- Modeling HPC applications and HPC platforms, looking for structural arguments on the shape of I/O movement and computations
- Developing and analysing new scheduling algorithms that take those structural arguments into account. As an example of natural extensions of our preliminary work[1]:
  1. The shape of a compute vs I/O period
  2. The multiplicity of entry-point into the job scheduler (multiple I/O nodes).

More information about the context and subject is available at `http://gaupy.org/ressources/files/dash_anr_proposal.pdf`

Prerequesite: Good knowledge in algorithmics and analysis of algorithms; Basic knowledge of probability theory; Basic knowledge in programming[2]

Scientific context Co-advised by Emmanuel Jeannot. Research will be performed in the Tadaam team (`https://team.inria.fr/tadaam/`) in Bordeaux.
We are looking for an intern that wants to continue as a PhD student. Funding is secured both for the internship and a PhD following the internship via the DASH ANR project. This also includes generous travel money during the PhD thesis.

[1] `http://gaupy.org/ressources/pub/reports/RR-9037.pdf`

[2] Depending on the liking/expertise of the student, there will be a possibility of implementing and experimenting those algorithms on top tier HPC platform.