CS554 Project Ideas

GeMTC:Hadoop – Supporting Hadoop Applications on Accelerators

Overview

This project aims to provide support for MapReduce applications on accelerators such as NVIDIA GPUs, AMD GPUs, or Intel Xeon Phi MIC. To program GPUs you typically use CUDA (NVIDIA) or OpenCL (AMD). The Intel Xeon Phi contains traditional x86 cores, which are programmed using traditional programming languages(C, OpenMP, Pthreads) or new emerging languages (e.g. OpenCL and OpenACC).

For this project you will write a framework (in any language such as CUDA, C/PThreads/SCIF, OpenCL, OpenMP, or OpenACC) to enable MapReduce applications such as those written in Hadoop to be executed on accelerators. If appropriate, the GeMTC framework could be leveraged. If other frameworks already exist to allow Hadoop programs to run on accelerators, these frameworks should be evaluated and compared to the native Hadoop stack running on commodity processors.

Relevant Systems and Reading Material

- GeMTC – http://datasys.cs.iit.edu/projects/GeMTC
- Xeon Phi - http://software.intel.com/en-us/mic-developer
- Swift – http://swift-lang.org
- Hadoop - http://hadoop.apache.org/
- Scott J. Krieder, Justin M. Wozniak, Timothy Armstrong, Michael Wilde, Daniel S. Katz, Benjamin Grimmer, Ian T. Foster, Ioan Raicu. "Design and Evaluation of the GeMTC Framework for GPU-enabled Many-Task Computing", ACM HPDC 2014; http://datasys.cs.iit.edu/publications/2014_HPDC14_GeMTC.pdf

Preferred/Required Skills

- Required: C or C++
- Preferred: CUDA, OpenMP, OpenACC, OpenCL, Threaded programming

Evaluation

You will be responsible for writing several test applications such as WordCount, Grep, and Sort, that utilize this framework. Experiments can be done on a single node, on a single Xeon Phi or an NVIDIA GPU on the Jarvis cluster.

Project Mentor

Ioan Raicu, iraicu@cs.iit.edu