**ABSTRACT**

Harp-DAAL stands for a new framework based on the HPC-ABDS concept that aims to bridge Big Data Applications with HPC platforms and methods. By interfacing our previous Harp [1][2][3] framework with Intel’s Data Analytics Accelerator Library (DAAL), we demonstrate that traditional HPC kernels written in native C/C++ could well serve in accelerating Java written Big data and machine learning applications on emerging many-core architectures such as Intel’s Knights Landing Xeon Phi processors.

**MATERIALS AND METHODS**

Harp[4] is a Hadoop plug-in made to abstract communication by transforming map-reduce programming models into map-collective models.

1. Harp has MPI-like collective communication operations that are highly optimized for big data problems.
2. Harp has efficient and innovative computation models for different machine learning problems.

DAAL[5] is Intel’s open-source highly optimized library that helps speed big data analytics by providing algorithmic building blocks for all data analysis stages and for offline, streaming, and distributed analytics usages.

**PERFORMANCE COMPARISON BETWEEN HARP-SGD AND HARP-DAAL-SGD**

Harp-DAAL-SGD accelerates the original Harp-SGD by around 20% thanks to faster computation kernels.

- Test for MovieLens with Dim 128 on 2 nodes of Juliet

**EXPERIMENTS ON KNL’S ARCHITECTURE**

Harp-DAAL-SGD benefits from KNL’s AVX-512 VPU with around 3x speedup. There is an overall 1.5x - 4x speedup over Haswell.

**CONCLUSIONS**

Harp-DAAL aims to provide fast machine learning solutions for Big Data applications. We have shown a better performance than the original Harp framework on both Haswell CPU and emerging many-core KNL Xeon Phi processors. Through extensive tests on KNL, highly optimized DAAL kernels can exploit the threads and memory resources of the new Xeon and Xeon Phi architectures. It shows 4x speedup over Haswell, an over 90% threads utilization, as well as a stat-of-art convergence speed.

**REFERENCE**


