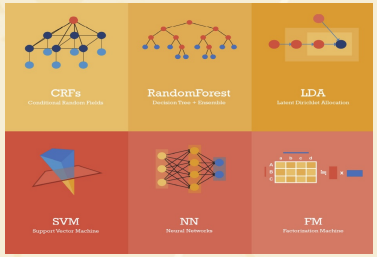
Harp-DAAL: A new Fast Machine Learning Framework on Emerging Architectures

Langshi Chen, Bingjing Zhang, Peng Bo, Judy Qiu IPCC@IU

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 Acceleration 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.

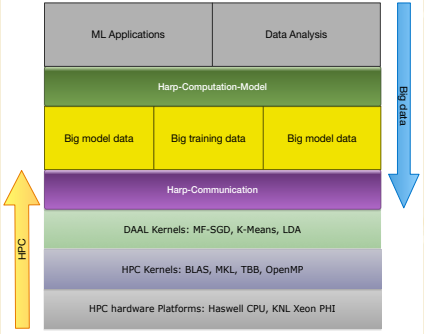


MATERIALS AND METHODS

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.

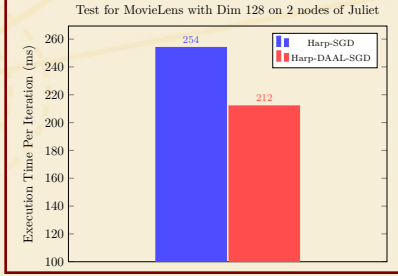
DAAL[5] is Intel's open-source highly optimized library that helps speed big data analytics by providing algorithmic building blocks for all data analysis stages and for offline, streaming, anddistributed analytics usages.

Harp-DAAL Framework Interfacing Big data and HPC



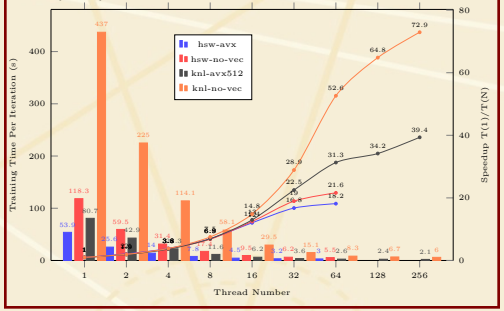
Performance Comparison between Harp-SGD and Harp-DAAL-SGD

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



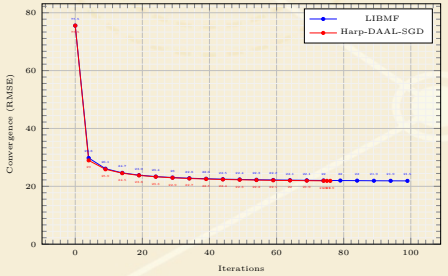
Experiments on KNL's Architecture

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

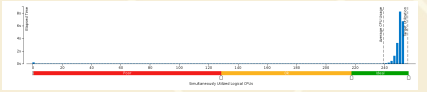


Fast Convergence

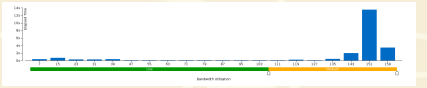
Harp-DAAL-SGD has a better convergence than the state-of-art LIBMF library, using 25% less iterations



It uses 93.5% of KNL's all 256 threads.



It benefits from KNL's High Bandwidth Memory



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

[1] B. Zhang, Y. Ruan, J. Qiu, Harp: Collective Communication on Hadoop, in the proceedings of IEEE International Conference on Cloud Engineering (IC2E2015), March 9-13, 2015.
[2] B. Zhang, B. Peng and J. Qiu, "Parallelizing Big Data Machine Learning Applications with Model Rotation". Submitted to Book Chapter in "Advances in Parallel Computing", 2016.
[3] B. Zhang, B. Peng and J. Qiu. "Model-Centric Computation Abstractions in Machine Learning Applications". BeyondMR, 2016.
[4] Harp project https://github.iu.edu/IU-Big-Data-Lab/Harp
[5] DAAL project https://github.com/01org/daal

INDIANA UNIVERSITY BLOOMINGTON
SCHOOL OF INFORMATICS AND COMPUTING