

High Performance Data Analytics



HPDA

*A Multifaceted Program to Accelerate Big Data Streaming Analytics
using High Performance Computing*

Contact

John R. Johnson, HPDA Program Director
john.johnson@pnnl.gov | (509) 375-2651

Mark Raugas
mark.raugas@pnnl.gov | (206) 528-3114

<http://www.pnnl.gov/computing/HPDA/>

PNNL-SA-130618

In collaboration with:



UNIVERSITY of
WASHINGTON



Proudly Operated by **Battelle** Since 1965

Emerging Architectures

System Software for Data-Vortex-based Environments: Developing a high-level system software stack for Data Vortex systems to improve performance and programmability of both data analytics and traditional HPC applications.

Partitioned Global Address Space Models for Data-Vortex-based Environments: Implementing PGAS models for Data Vortex systems to enable a wide class of codes to be evaluated on the architecture.

RISC-V for Graph Processing: Customizing the RISC-V Open System on a Chip (OpenSOC) for irregular, graph-centric workloads.

Systems

Abstract Hypergraph Machine: Developing an Abstract Hypergraph Machine runtime for both multi-edge and vertex-centric computation. Fundamental Hypergraph Analytics and Benchmarking of Large Scale Algorithms.

SHAD: Flexible and scalable data structures for irregular workloads.

Large-Scale Graphs

Performance Modeling on Property Graphs: Developing new modeling capabilities for property graphs that include graph-generation algorithms and providing a quantitative understanding of graph queries at extreme scales.

Topological Data Modeling: Analyzing big data with high complexity using TDM techniques, including Abstract Simplicial Complexes and Sheaf Theory.

Machine Learning

Graph-centric Deep Learning: Studying clustering, analysis, and prediction tasks on graphs using Deep Learning.

Data

Simulation of Large-scale Graphs: Developing HPC-based simulation tools that can generate large-scale data sets containing labeled activity to test graph analytics algorithms.

SCALABLE LIBRARIES FOR HPC PLATFORMS

MaTE_x

Extreme Scale Machine Learning Toolkit using MPI
<https://github.com/matex-org/matex/wiki>

SHAD

Flexible and scalable data structures for irregular workloads
<https://github.com/pnnl/SHAD>

Grappolo

Multithreaded C++ and OpenMP library for graph clustering
<http://hpc.pnl.gov/people/hala/grappolo.html>

HPDA