



Pacific Northwest  
NATIONAL LABORATORY

Proudly Operated by **Battelle** Since 1965

# High Performance Data Analytics

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



## Graph Analytics

Graph modeling, visualization, and evaluation for understanding large, complex networks

## Compute Intensive Analytics

Novel and innovative techniques to solve computationally intensive problems

## Streaming Analytics

New algorithms and approaches to rapidly analyze high-bandwidth, high-throughput streaming data

## Exploratory Data Analysis

Mechanisms to explore and analyze massive streaming data sources to gain new insights and inform decisions



## Emerging Architectures Analysis and Distributed Heterogeneous Testbed

Evaluation, characterization, and modeling of next-generation analytics



UNIVERSITY of  
WASHINGTON



## GRAPH ANALYTICS

- » **Semantic Data Analysis:** Creating high-performance platforms—scaling to hundreds of terabytes of data and offering fault resilience—for integrated analysis across graph-based and relational semantic data structures along with new query languages for this blended data repository
- » **Big Data Visualization:** Developing strategies to learn and visualize relationships in graph data that are several orders of magnitude larger than those currently possible to process with available visualization tools
- » **Data Sets for Graph Analytics Research:** Developing a multi-discipline data generator, ontology, and queries for evaluating the features, performance, and scalability of data analytics tools
- » **Activity-based Analytics:** Building a scalable, model-based, information-theoretic, and dynamic graph analytics infrastructure that extends beyond topological and static approaches
- » **Performance Analysis of Big Graph Data Tools:** Building metrics and strategies to assess the performance of these tools as the data and platforms increase in scale
- » **Activity-based Analytics:** Building a scalable, model-based, information-theoretic, and dynamic graph analytics infrastructure that extends beyond topological and static approaches
- » **Anti-Evasive Anomaly Detection:** Building a system for monitoring high-performance computing systems, looking for anomalies in an environment where an adversary may try to evade detection by reacting to our own monitoring activities

## STREAMING ANALYTICS

- » **Emerging Graph Patterns:** Developing scalable graph algorithms and a network analysis framework to detect precursor events or patterns and mitigate threats
- » **Data Fusion and Compression:** Exploring novel methods for dimensionality reduction and fusion of heterogeneous big data sources
- » **Massive-scale Network Analysis:** Establishing methods for conducting real-time analysis of massive scale streaming data

- » **High-volume, Variable-veracity Subgraph Search:** Exploring computation platforms and algorithms to find patterns in streaming graphs meeting confidence requirements

## COMPUTE INTENSIVE ANALYTICS

- » **Malware Detection:** Developing dynamic and bio-inspired static analysis techniques for malware classification, characterization, and detection.
- » **Resilient MPI for Graph Engine for Multithreaded Systems (GEMS):** Developing fault tolerant system software to ensure continued execution in the presence of faults.
- » **Typography:** Developing techniques to explore multi-million document repositories of unstructured text through a combination of visualization and a suite of computationally demanding class of algorithms to better address analysts' research questions.
- » **MPI-Based Machine Learning:** Supporting the analyst by automatically selecting the correct machine learning algorithm given an assigned problem set on high performance computing (HPC) implementations.
- » **Recurrent Pattern Discovery:** Developing event analysis and sequencing using co-occurrence detection and recurrent temporal pattern discovery using a combination of signal processing, statistical analysis and machine learning algorithms.
- » **Myria Middleware:** Transitioning and deploying an integrated Polystore Data System. Functionality includes query translation/optimization, mixed relational algebra algorithms, graph traversal and workspace abstraction.

## EXPLORATORY DATA ANALYSIS

- » **Exploratory Graph Analysis:** Creating a set of intuitive tools and query languages to facilitate the interactive exploration of semantic databases
- » **Geo-inspired Parallel Simulation:** Developing a parallelized simulation that capitalizes on structural similarities between complex networks that are geographically defined
- » **Cyber Analytic Data Sets:** Generating large-scale, synthetic data for cyber analytics research and development

### Contact

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



U.S. DEPARTMENT OF  
**ENERGY**