MACHINE LEARNING
StreamSmart: Streaming hypothesis generation with human-in-the-loop to provide explanations for decisions.
Event Analysis and Recurrent Pattern Discovery: Perform scalable, near-real-time analysis to detect and identify even co-occurrences and assemble sequences of co-occurrences into recurrent temporal patterns.
Message Passing Interface-based Machine Learning: Combining algorithmic and systems techniques for scaling out deep learning algorithms on very-large-scale systems.
Deep Learning on Multilingual Social Media: Enabling a language-independent view of global events and trends in real time.

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 high-performance computing applications.
Automata Processor: Exploring methods to identify known interaction patterns in rich semantic data.

SYSTEMS
Semantic Data Analysis: Creating high-performance platforms for computing over large sizes of graph and tabular (SQL-like) data.
Resilient Message Passing Interface for Fault-tolerant Runtimes: Designing fault-tolerant MPI runtimes that provide high-fidelity fault detection and fault tolerance for computing-intensive and graph algorithms.
Myria Middleware—Unified Services for Hybrid Big Data Systems: Providing monitoring and analysis services across multiple backends, including plan inspection, performance profiling, per-query, and cluster utilization.
Graph Query Language: Building a toolkit to enable interactive, analyst-driven exploration of very large data represented in a property graph format.

LARGE SCALE GRAPH
Performance Modeling on Property Graphs: Developing new modeling capabilities for property graphs that include graph-generation algorithms and quantitative understanding of graph queries at extreme scales.
Topological Data Modeling: Analyzing big data with high complexity using TDM techniques.
Scalable Approximate Graph Clustering on Streaming Data: Developing clustering techniques for two fundamentally different formulations of the dynamic clustering problem to expose and exploit approximation strategies and achieve effective reduction in time-to-solution.
High-performance Algorithms and Software for Clustering Based on Constrained Low-rank Approximations: Designing scalable algorithms for large-scale problems and affording the ability to produce more accurate solutions faster in noisy real-life applications.
Web-scale Graph Visual Analytics: Exploring a web-scale graph that can overcome scalability challenges in size, cognition, visualization, and computation.

DATA
Big Data Benchmarking Suite for Cybersecurity Analytics: Establishing a benchmark for evaluating performance and bottlenecks of big data systems with a focus on cybersecurity-related workloads and data sets.
Simulation of Large-scale NetFlow Data with Botnet Activity: Developing a high-performance computing-based simulation tool that can generate large-scale NetFlow data sets containing labeled malicious activity to test graph analytics algorithms.

In collaboration with:
SCALABLE LIBRARIES FOR HPC PLATFORMS

**Machine Learning**
- Predict cybersecurity threats in near real time
- Rapidly analyze social media to discover trending events

**Emerging Architectures**
- Evaluate, characterize, and model next-generation architectures
- Develop and deploy systems software to enable novel hardware

**Graph Analytics**
- Discover structure of communities in massive streaming data
- Interactively query and visualize big data at scale

**MaTEx**
Machine Learning Toolkit for Extreme Scale using MPI
http://hpc.pnl.gov/matex/

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

HIGH PERFORMANCE DATA ANALYTICS
Basic and Applied Research, Development, and Technology Transfer