Computational Challenges for Nanophase Materials Sciences

Scientific insight is increasingly dependent on the ability to efficiently integrate and analyze complex multi-modal data.

Analysis of high-dimensional data from experimental techniques at the Center for Nanophase Materials Sciences (CNMS) are designed to probe complementary ranges of time, space, and energy.

**MULTI-MODAL EXPERIMENTS**
- Scanning Transmission Electron Microscopy (STEM)
- Scanning Probe Microscopy (SPM)
- Scanning Tunnelling Microscopy (STM)
- Molecular dynamics (MD) Simulations

Near Real-time Data Analysis

**SCALABLE DATA ANALYSIS**
- Remote execution of advanced algorithms
- Dynamically generate & execute HPC workflows
- Interactive 3D plotting

**ATOMIC ANALYSIS**
- Execution of image processing filters
- Find atomic positions through clustering
- Navigate analysis with the Image Processing Stack UI component

Workflow System Requirements

**ROBUST ANALYSIS IN NEAR REAL-TIME**
Fixed time-window to conduct the experiment at national facilities requires feedback in near-real time to drive the next step in the experiment.

**ROBUST ANALYSIS AT THE EXTREME SCALE**
- Image processing (support from ACUMEN) and machine learning of high dimensionality of experiment data
- Computational complexity of simulation (MD) required to analyze physical dynamics
- Need for quick turnaround of analysis tasks

Robust Data Management

Easily transmit data files to your private BEAM storage area and manipulate remote directories and files as if they were stored locally.

**UNIQUELY IDENTIFIED DATA FILES**
- Each DB record contains provenance attributes enabling full context search and filtering capabilities
- Search by collection date & upload date, instrument name, image resolution, facility user
- User-defined metadata
- Data search and filtering software tool in progress

**DATA MANAGEMENT SOLUTION**
- All steps from raw data to analysis to finalized results can be stored
- Unbroken chain of custody maintains data integrity
- Complete provenance stored in data file
- HDFS Data Format

Multi-tier Architecture Includes OLCF & CADES

**LEVERAGING THE POWER OF CADES**

**HOW DOES BEAM USE CADES?**
- Cluster computing capability
- Infrastructure as a Service (IaaS) provides scalable computing, software support, and database support
- High performance storage that can expand and grow with our needs
- MNMS is installing a new compute cluster in CADES (ready Oct 1, 2015)
- Cray CS400 with 32 compute nodes with 24 cores per node (768 cores total)
- 128 GB of RAM per node (4 TB of RAM total)

**FUTURE ROADMAP**
- Scale to larger data and real-time streaming data flows
- Development of a suite of machine learning methods tailored to CNMS and IFIM problems provided as a toolkit to users
- Multi-modal data integration and analysis across multiple microscopes and experimental modalities
- Continue development of intuitive user-friendly workflows, custom dashboards, and software tools for large-scale integrated data analytics, computational modeling, and collaboration

BEAM is funded by the LDRD Program at ORNL

Powered by Bellerophon