Scalable In-Memory Data Indexing and Querying for Scientific Simulation Workflows

**Target Application**
- S3D combustion simulation (Jacqueline Chen and Hemanth Kolla, Sandia National Laboratory)
- Goal: Identify flamefronts (a transient phenomena) at runtime in combustion simulations

**Data Management Challenges**
- Develop capabilities for interactive queries with low latency, operating on live simulation results
- Efficient data indexing and querying

**Approach**
- Parallel in-memory indexing and querying on staging nodes
- Support SQL-like query syntax and simple querying APIs
- Flexible framework that can integrates different index techniques (currently using FastBit compressed bitmap index)

**Impact**
- Efficiently support online runtime query-driven data analysis for extreme scale scientific simulations
- In-memory approach improves both indexing time (up to 7x) and querying time (up to 35x)

Q. Sun, F. Zhang, T. Jin, H. Bui, M. Parashar, K. Wu, A. Shoshani, H. Kolla, S. Klasky, J. Chen
“Scalable In-Memory Data Indexing and Querying for Scientific Simulation Workflows”, Submitted to CCGrid 2014.