**SIRIUS: Science-driven Data Management for Multi-tiered Storage**

Abstract

This project explores the use of application level knowledge to optimize the times to insight across a workload of multiple applications in a multi-user environment with shared storage and network resources. The basis for this work is the notion of selectable data quality to explore the tradeoffs between accuracy of results, resource requirements, and time to insight on systems with shared, oversubscribed computational and storage resources.

Our thesis is that by adding application level knowledge about data to guide storage system behaviors, we will obtain substantial benefits in the organization, storage, and access to extreme scale data resulting in improved productivity for computational science. We will demonstrate novel techniques to facilitate efficient data placement onto multiple storage tiers, and enable application-guided data reductions to address capacity, bandwidth, and latency. Our goal is to address the associated data management challenges in the context of current and emerging storage landscapes and expedite insights into mission critical scientific processes. We will test our prototypes on current and future Department of Energy (DOE) system with many of today's cutting edge applications to ensure our techniques can be used within our framework on current and next generation systems.