**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. In the last several years’ extreme scale computational science was conducted using dedicated computational resources, with shared resources, such as network and storage, being overprovisioned. This resulted in little to no impact from resource contention. The increased scale of recent, and incoming future systems, has placed a greater emphasis on mitigating contention, which is a focus area of our research.

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 and effective data placement onto multiple storage tiers and enable application-guided data reductions and transformations to address capacity and bandwidth bottlenecks. 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.