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

**Abstract:**

Scientific discovery at the exascale will not be possible without significant new research in the management, storage and retrieval over the long lifespan of the extreme amounts of data that will be produced. Our thesis in this project is that adding application level knowledge about data to guide the actions of the storage system provides substantial benefits to the organization, storage, and access to extreme scale data, resulting in improved productivity for computational science. In this project we will demonstrate novel techniques to facilitate efficient mapping of data objects, even partitioning individual variables, from the user space onto multiple storage tiers, and enable application-guided data reductions and transformations to address capacity and bandwidth bottlenecks. Our goal here is to address the associated Input/Output (I/O) and storage challenges in the context of current and emerging storage landscapes, and expedite insights into mission critical scientific processes.

We aim to reduce the time to insight, not just for a single application, but for the entire workload in a multi-user environment, where the storage is shared among users. We achieve this goal by allowing selectable data quality, by trading its accuracy and error in order to meet the time or resource constraint. We will explore beyond checkpoint/restart I/O, and will address the challenges posed by key data access patterns in the knowledge gathering process. Ultimately, we will take the knowledge from the storage system to provide vital feedback to the middleware so that the best possible decisions can be autonomically made between the user intentions and the available system resources. We will test our prototypes on current and future Department of Energy (DOE) system with many of today's applications to ensure our techniques can be used within our framework on current and next generation systems.