

## **Fast Data Analysis for Scientific Big Data Applications**

Many scientific computing and high-performance computing applications have become increasingly data intensive. Recent studies have started to utilize indexing, subsetting, and data reorganization to manage the increasingly large datasets. In this talk, I will present our study of Fast Analysis with Statistical Metadata intending to boost the data analytics performance via data subsetting and integrating a small amount of statistics into the original datasets. The added statistical metadata illustrates the data shape and provides knowledge of the data distribution; therefore the original scientific libraries can utilize these statistical metadata to perform fast queries and analyses. We will also discuss segmented analysis and pre-analysis concepts and ideas for reducing data movement and to speed up data analysis for scientific big data applications. These concepts and ideas can potentially lead to new data analytics methodologies and can have an impact on scientific discovery productivity.