

Faculty Candidate Seminar

Building HPC Storage Infrastructure for Data-Intensive Science

Dong Dai

Texas Tech University

Abstract: Data-intensive science has emerged as a new, fourth paradigm for scientific exploration. Its core activities, including data capture, data analysis, and data curation, all heavily rely on high-performance storage infrastructure. In this talk, I will illustrate how we build the storage infrastructure in high-performance computing (HPC) platforms to help scientists 1) access their data quickly, and 2) manage their data efficiently. First, I will focus on the I/O interference issue raised in a highly concurrent environment. Such interference may create stragglers in the system and significantly slow down the data access performance. In our work, we propose a two-choice randomized I/O scheduler that can dynamically identify and avoid the stragglers to improve the data access performance. In the second part, I will focus on the concept of provenance metadata, which records the history of a piece of data, and is critical for many scientific data management tasks like understanding data origins, verifying data quality, and reproducing important results. I will describe how we build new storage infrastructure to collect, store, and query provenance metadata in HPC platforms with extreme performance and scalability.

Bio: Dong Dai is currently a research assistant professor at Computer Science Department, Texas Tech University. Before that, he was a joint postdoctoral researcher at Argonne National Laboratory and Texas Tech University. His research lies in the general area of high-performance computing and data-intensive systems. More specifically, he builds high-performance storage infrastructure to help scientists and their applications access and manage data with extremely high performance and scalability. His research has been published in top-tier HPC conferences (ACM/IEEE SC, ACM HPDC, and PACT) as well as other major venues and journals such as IEEE Cluster and ParCo. He also served as PIs and Co-PI for multiple research grants from National Science Foundation.

Date: March 9, 2018

Time: :00 am

Computer Science Building

