

Information Science and Technology Seminar Speaker Series



Gagan Agrawal
Ohio State University

Tools for Managing, Querying, and Processing Large-Scale Scientific Datasets

Wednesday, August 14, 2013

3:00 - 4:00 PM

TA-03, Bldg. 1698, Room A103 (MSL Auditorium)

Abstract: This talk will describe research on a number of 'big-data' topics at Ohio State. Initially, we will describe a framework for efficient 'MapReduce'-style processing developed in our lab. This system supports a variant of the popular MapReduce API, and has the following key features: 1) it can be up to a factor of 10 or more efficient than Hadoop for many scientific data analysis and data mining tasks, 2) it can directly work on top of scientific data formats like NetCDF and HDF5, and 3) it can be used not only on multi-core CPU clusters, but also accelerators and CPU-GPU clusters.

Next, we will describe a data management framework, where 'database'-like functionality is provided as an 'add-on' on top of massive scientific datasets (including those stored in formats like NetCDF and HDF5). Without requiring any reformatting or reloading of data, we can support selection and projection queries, a number of array operations, sampling services, and other analysis often desired in a scientific environment. This work involves not only a number of optimizations on top of the existing work on bitmap indices, but also novel applications of these methods, for sampling and dealing with noisy data.

Biography: Gagan Agrawal is a professor of computer science at Ohio State University. He received a BS degree from IIT Kanpur, and MS and PhD degrees from University of Maryland, College Park. He has worked in a number of research areas, including parallel compilation and runtime support, data mining, and grid and cloud computing. His recent research is focused on two areas: tools and programming models for accelerator-based computing, and managing and processing large-scale datasets.

For more information contact the technical host Curt Canada, cvc@lanl.gov, 665-7453 or James Ahrens, ahrens@lanl.gov, 667-5797.