UNCLASSIFIED

Information Science and Technology Seminar Speaker Series

Bill Howe
University of Washington

Myria/Kamook: Federated Big Data Programming across HPC and Enterprise Environments

Wednesday, August 12, 2015

3:00 - 4:00 PM

TA-3, Bldg. 1690, Room 102 (CNLS Conference Room)

Abstract: There has been a "Cambrian explosion" of big data systems proposed and evaluated in the last eight years, but relatively little understanding of how these systems or the ideas they represent compare and complement one another. In enterprise and science situations, “one size is unlikely to fit all”: we see analytics teams operating multiple systems simultaneously in a common ecosystem. However, the highest level of abstraction for interoperability achieved in practice is basically at the file system (e.g., lustre in HPC environments and HDFS in enterprise environments). But we see a convergence around data models (relations, arrays, graphs) and collection-at-at-time programming models (relational algebra, parallel dataflow, linear algebra).

The design space therefore seems narrower than the implementation space, suggesting an opportunity to build a suite of common interfaces to all these seemingly disparate systems to make them easier to compare, easier to use together, and perhaps to improve overall performance by affording cross-platform, federated optimization. In aprticular, we see a role for federation in bridging the gap between high-performance computing in computational science and the high-throughput computing being advanced by industry, allowing different kinds of platforms and design alternatives to co-exist in a single ecosystem.

I'll describe the Myria system (and the Kamook optimizer) we are building at the University of Washington to achieve thse goals. We provide a common programming algebra and compile this intermediate representation to multiple back-end systems, including a parallel global address space (PGAS) model designed for use on high-performance clusters.

I'll also describe a family of algorithms for scalable community detection we are exploring for use with Myria as a challenge problem.

Biography: Bill Howe is the Associate Director of the UW eScience Institute and holds an Affiliate Assistant Professor appointment in Computer Science & Engineering, where he studies data management, analytics, and visualization systems for science applications. Howe has received two Jim Gray Seed Grant awards from Microsoft Research for work on managing environmental data, has had two papers elected to VLDB Journal's "Best of Conference" issues (2004 and 2010), and co-authored what are currently the most-cited papers from both VLDB 2010 and SIGMOD 2012. Howe serves on the program and organizing committees for a number of conferences in the area of databases and scientific data management, and serves on the Science Advisory Board of the SciDB project. He has a Ph.D. in Computer Science from Portland State University and a Bachelor's degree in Industrial & Systems Engineering from Georgia Tech.

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

Hosted by the Information Science and Technology Institute (ISTI)

Los Alamos NATIONAL LABORATORY EST. 1943

ISTI INFORMATION SCIENCE & TECHNOLOGY INSTITUTE