SPEAKER: Michael Franklin, Computer Science, University of Chicago

TOPIC: A Retrospective on the AMPLab and the Berkeley Data Analytics Stack

DATE: Friday, December 9

TIME: 3:30 pm

PLACE: OLS 011 (note change in venue)

HOST: Jane Qi (yq2h)

Abstract: The Algorithms, Machines and People Laboratory (AMPLab) was launched by a group of systems and machine learning faculty at UC Berkeley in early 2011 and was awarded an NSF CISE Expeditions in Computing grant in 2012. The goal of the lab is to develop a new approach to large-scale data analytics (i.e., Big Data processing) that seamlessly integrates the three main resources available for making sense of data at scale: Algorithms (machine learning, statistical and query processing techniques), Machines (scalable clusters and elastic cloud computing), and People (as individual analysts and as crowds). The lab has had significant impact on the Big Data software landscape through the development of a freely-available Open Source software stack called BDAS: the Berkeley Data Analytics Stack. BDAS is a comprehensive analytics platform that has been the incubator for a number of influential systems including the Mesos cluster resource manager (now Apache Mesos), the Spark in-memory computation framework (now Apache Spark), and the Tachyon distributed storage system (now called Alluxio). It contains interfaces for streaming analytics, distributed machine learning, high-performance SQL processing and graph analytics, among others. While serving as a unifying artifact for dozens of Ph.D. and Postdoctoral researchers, BDAS software features prominently in many industry discussions of the future of the Big Data analytics ecosystem - a rare degree of impact for an academic project.

The AMPLab is in the final year of its planned six-year existence. In this talk I will provide an overview of AMPLab and BDAS with a focus on identifying the overarching themes of this large software research project. I will highlight some risks we took and a few that we didn’t and will then provide some thoughts on the future of data analytics systems and the best ways for academic researchers to influence that future.

Bio: Michael Franklin is the Liew Family Chair of Computer Science and a Sr. Advisor to the Provost for Computation and Data at the University of Chicago. He joined U. Chicago in summer 2016 to help implement a major expansion of the Computer Science Department and to initiate a cross-campus effort in Data Science. He is also an Adjunct faculty member at UC Berkeley, where he was the Thomas M. Siebel Professor of Computer Science and former Chair of the Computer Science Division. Prof. Franklin is currently serving as PI of the Algorithms, Machines, and People Laboratory (AMPLab) expedition. In addition to its status as a NSF CISE Expeditions project, AMPLab works with dozens of industrial sponsors including foundering sponsors Amazon Web Services, Google, IBM, and SAP. Prof. Franklin is a PI of the NSF Western Region Big Data Innovation Hub and was formerly a co-PI and Executive Committee member for the Berkeley Institute for Data Science, part of the Moore and Sloan Foundations' initiative to advance Data Science Environments. He is an ACM Fellow, a two-time winner of the ACM SIGMOD "Test of Time" award, has several recent "Best Paper" awards and two CACM Research Highlights selections, and is a recipient of the Outstanding Advisor Award from the Computer Science Graduate Student Association at Berkeley.