Making Sense of Big Data With the Berkeley Data Analytics Stack

Michael Franklin

Thursday, Oct. 9
Lecture, 4 p.m.

McMurtry Auditorium, Duncan Hall
Reception, 5 p.m.

Abstract
In this talk, I’ll first outline the motivation and insights behind our research approach and describe how we have organized to address the cross-disciplinary nature of Big Data challenges. I will then describe the current state of a freely available Open Source software stack called BDAS: the Berkeley Data Analytics Stack with an emphasis on our newest efforts, including the GraphX graph processing system, the MLBase machine learning platform, and the SampleClean framework for combining sampling and hybrid human/computer data cleaning. Finally, I will present our current views of how all the pieces will fit together to form a system that can adaptively bring the right resources to bear on a given data-driven question to meet time, cost and quality requirements throughout the analytics life cycle.

Bio
Michael Franklin is the Thomas M. Siebel Professor of Computer Science and chair of the computer science division at the University of California at Berkeley. He has more than 30 years of experience in the database, data analytics and data management fields as a researcher, lab director, faculty member, entrepreneur and software developer. Franklin is also the director of the Algorithms, Machines and People Laboratory (AMPLab) at UC Berkeley. The AMPLab currently works with 23 industrial sponsors, including founding sponsors Amazon Web Services, Google and SAP, and received a National Science Foundation CISE Expeditions in Computing award in March 2012. Franklin is a co-PI and executive committee member for the Berkeley Institute for Data Science, part of a multicampus initiative to advance data science environments. He is an ACM Fellow, a two-time winner of the ACM SIGMOD Test of Time award and received the outstanding Adviser Award from the Computer Science Graduate Student Association at Berkeley.