GraphLab: Machine Learning for Big Data in the Cloud

Carlos Guestrin
CSE, University of Washington

Abstract
Today, machine learning (ML) methods play a central role in industry and science. The growth of the Web and improvements in sensor data collection technology have been rapidly increasing the magnitude and complexity of the ML tasks we must solve. This growth is driving the need for scalable, parallel ML algorithms that can handle "Big Data."

In this talk, I will also describe the GraphLab framework, which naturally expresses asynchronous, dynamic graph computations that are key for state-of-the-art ML algorithms. When these algorithms are expressed in our higher-level abstraction, GraphLab will effectively address many of the underlying parallelism challenges, including data distribution, optimized communication, and guaranteeing sequential consistency, a property that is surprisingly important for many ML algorithms. On a variety of large-scale tasks, GraphLab provides 20-100x performance improvements over Hadoop. In recent months, GraphLab has received thousands of downloads, and is being actively used by a number of startups, companies, research labs and universities.

Biography
Carlos Guestrin is the Amazon Professor of Machine Learning at the Computer Science & Engineering Department of the University of Washington. His previous positions include Associate Professor at Carnegie Mellon University and senior researcher at the Intel Research Lab in Berkeley. He is also the co-founder of Flashgroup, a start up focused on addressing information and social overload on the web. Carlos received his PhD and Master from Stanford University, and a Mechatronics Engineer degree from the University of Sao Paulo, Brazil.