

Distinguished Speaker Series
April 22 at 11:30am, NTDP F223

Cumulon: Simplifying Matrix-Based Data Analytics in the Cloud

Dr. Jun Yang



Abstract

Cumulon is a system aimed at simplifying the development and deployment of statistical analysis of big data in public clouds. Cumulon allows users to program in their familiar language of matrices and linear algebra, without worrying about how to map data and computation to specific hardware and cloud software platforms. Given requirements in terms of time, monetary cost, and risk tolerance, Cumulon automatically makes intelligent provisioning, configuration, and execution decisions---from the type and number of machines to acquire, to the choice of blocking factors for matrix multiply. For clouds with auction-based markets, where the cost and availability of computing resources vary according to market conditions, Cumulon helps users decide how to bid for such resources and how to cope with market volatility. In this talk, I will share our experience in building Cumulon, including the alternatives explored and the lessons learned.

Bio

Jun Yang is a Professor of Computer Science at Duke University, where he has been teaching since receiving his Ph.D. from Stanford University in 2001. He is broadly interested in databases and data-intensive systems. He is a recipient of the NSF CAREER Award, IBM Faculty Award, HP Labs Innovation Research Award, and Google Faculty Research Award. He also received the David and Janet Vaughan Brooks Teaching Award at Duke. His current research interests lie in making data analysis easier and more scalable for scientists, statisticians, and journalists.