CENTER FOR NEURO-ENGINEERING & COGNITIVE SCIENCE & THE DEPARTMENT OF INDUSTRIAL ENGINEERING PRESENTS...

Declarative Big Data Analytics: Applications and Tools

Friday, October 24th, 4:00 p.m.
Location: CBB Room 106

Shivakumar Vaithyanathan
Senior Manager, Cognitive Computing
IBM

In February of 2011 a computer program defeated the two best human players on one of the most challenging question/answer game shows of all time, Jeopardy! Named, Watson, after IBM's founder, this event was largely recognized as a significant milestone in the advancement of Big Data Analytics. In this talk I will give a brief introduction to several applications of this Big Data analytics technology to tasks ranging from retail applications of social media analytics such as lead generation to applications in investment and equity research. I will briefly describe the major analytic phases: text analytics, semi-structured data processing (joins, group-by, aggregation), and statistical/predictive modeling core to these applications. At IBM we are building tools and technologies to support each of these analytic phases and in particular we are building declarative languages for these phases. While the declarative nature of the language abstracts away the need for programmer-optimization, the syntax of these languages is designed to appeal to the corresponding communities. In the second part of the talk I will describe SystemML in which ML algorithms are expressed in a higher-level language and are compiled and executed in a MapReduce environment. This higher-level language exposes several constructs including linear algebra primitives that constitute key building blocks for a broad class of supervised and unsupervised ML algorithms. The algorithms expressed in SystemML are compiled and optimized into a set of MapReduce jobs that can run efficiently on a cluster of machines. I will end the talk with a discussion on speeds, feeds and comparisons.

Speaker Biosketch

Shivakumar Vaithyanathan is an IBM Fellow and Chief Scientist for Big Data Analytics and manages the Machine Learning Systems group at IBM Research. His research is at the cusp of three areas namely natural language processing, machine learning and databases. Multiple technologies developed under his direction ship with several IBM products including IBM's Big Data products. Prior to IBM, Shivakumar was one of the founding members of the Altavista Group at Digital. He has co-authored more than 40 papers in major conferences including, ACL, EMNLP, SIGMOD, VLDB, ICML, NIPS and UAI and was an invited keynote speaker at the 2011 German Database Conference and 2011 ACM SIGIR Industrial Track. He was an associated editor for Journal of Statistical Analysis and Data Mining from inception to 2012.

For additional information, please contact Bhavin Sheth at brsheth@uh.edu