

Workshop on Big Data Analytics

Date: 20 February 2013, Wednesday

Time: 9.00am - 5.00pm

Venue: National University of Singapore, LT18

Location map and driving instructions: <http://www.comp.nus.edu.sg/maps/location.html>

Register at: <https://register.comp.nus.edu.sg/talk/>

Event website at: <http://www.comp.nus.edu.sg/events/BigDataWorkshop.html>



H V Jagadish
Bernard A Galler Collegiate Professor
University of Michigan



Cyrus Shahabi
Professor of Computer Science
& Electrical Engineering
University of Southern
California



David Maier
Maseeh Professor of
Emerging Technologies
Portland University



Ke Yi
Associate Professor
Hong Kong University of
Science and Technology



Anindya Datta
Associate Professor
National University
of Singapore



Jeffery Ullman
Stanford W Ascherman
Professor
Stanford University



Roland Yap
Associate Professor
National University
of Singapore

Big Data: Challenges and Opportunities outside the Box

Speaker: H V Jagadish, Bernard A Galler Collegiate Professor of Electrical Engineering and Computer Science, University of Michigan

Synopsis: The promise of data-driven decision-making is now being recognized broadly, and there is growing enthusiasm for the notion of "Big Data." In fact, it is hard to avoid mention of Big Data anywhere we turn today. This talk will present an overview of the challenges and then delve into some Big Data issues that may not be the first ones that come to your mind when you hear the term.

TransDec: A "Big Data" Framework for Decision Making in Transportation Systems

Speaker: Cyrus Shahabi, Professor of Computer Science & Electrical Engineering, University of Southern California

Synopsis: The vast amounts of transportation datasets (traffic flow, incidents, etc.) collected by various federal and state agencies are extremely valuable in 1) real-time decision-making, planning, and management of the transportation systems, and 2) conducting research to develop new policies to enhance the efficacy of the transportation systems. This talk will present a data-driven framework, dubbed TransDec (short for Transportation Decision-Making), which enables real-time integration, visualization, querying, and analysis of dynamic and archived transportation data.

Analysing High-Volume Data Streams

Speaker: David Maier, Maseeh Professor of Emerging Technologies, Department of Computer Science, Portland University

Synopsis: Velocity is one of the "Three V's" of Big Data (along with Volume and Variety). This talk discusses recent tools and technologies for analyzing high-volume data streams, including out-of-order processing, data-stream warehouses and modifications of both relational and map-reduce technologies to support data streams.

Tracking Distributed Data

Speaker: Ke Yi, Associate Professor of Computer Science, Hong Kong University of Science and Technology

Synopsis: Consider a model where k players each receive a stream of elements over time, and they communicate with the goal of tracking some function of all the elements received so far continuously in real time. This talk will present a brief overview of the results in this model, with a few selected problems explained in more detail: count-tracking, frequent items, and random sampling.

Research Problems at the Intersection of Big Data and Mobile Applications

Speaker: Anindya Datta, Associate Professor, National University of Singapore

Synopsis: The tremendous rise of mobile computing devices in general, and mobile applications (aka apps) in particular, have given rise to the fastest growing consumer segment in the history of human merchandising. Apps generate massive amounts of data, which in turn can be mined to uncover intelligence of great use to players in the app ecosystem, such as publishers and advertisers. In this talk, we shall provide an overview of our research in this area, including collection systems that we have set up to acquire data and processing systems we have set up to uncover intelligence. We will describe, at a high level, a variety of problems we are investigating, and drill down into a specific important research question: how to decipher the true popularity ranks of apps.

Algorithm Design for MapReduce

Speaker: Jeffery Ullman, Stanford W Ascherman Professor of Computer Science, Stanford University

Synopsis: After a brief review of how MapReduce works, we shall look at the trade-off that needs to be made when designing MapReduce algorithms for problems that are not embarrassingly parallel. In particular, the less data that one reducer is able to handle, the greater the total amount of data that must be communicated from mappers to reducers. We can view this trade-off as a function that gives the "replication rate" (average number of copies of an input communicated from mappers to reducers) in terms of the "reducer size" (number of inputs that can be accommodated at a reducer).

Computing with MapReduce on Large Graphs

Speaker: Roland Yap, Associate Professor, National University of Singapore

Synopsis: The growth of the Internet has led to the creation of many important large graphs such as social networks and the graph of the web itself. The size of these graphs lead to new challenges in graph algorithms since such graphs are far larger than what can be feasibly processed with conventional memory-resident algorithms. We demonstrate how to use MapReduce for a classic graph algorithm, namely, maximum-flow. By designing an algorithm with a high-degree of parallelism and exploiting properties of the graphs, we are able to scale to very large graphs.



School of
Computing

Leading the World with Asia's Best

Register At

