

NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

IS Teaching SEMINAR

Title: Hadoop: A Distributed Infrastructure for Big Data Analytics

Speaker: Mr. Xu Xiaoying, Ph.D. Candidate

Date/Time: 5 September 2014, Friday, 02:00 PM to 03:00 PM

Venue: MR1, COM1-03-19

Chaired by: Mr. Chen Qing, Ph.D. Candidate
(chenqing@comp.nus.edu.sg)

ABSTRACT:

Big data is a hot topic nowadays. This seminar is going to introduce the basic concepts of Hadoop which is a distributed infrastructure to support big data storage and analytics. First, a brief introduction of big data will be given, followed by the Hadoop file system and MapReduce programming framework. We will then discuss other components like Mahout and Hive in the Hadoop ecosystem. Simple demos including running MapReduce Job in Amazon cloud computing platform and using Hortonworks Sandbox to test Hive script will also be presented.

BIODATA:

Mr. Xu Xiaoying is currently a Ph.D. candidate and teaching assistant in Department of Information Systems, School of Computing, National University of Singapore. He received his bachelor degree from School of Software Engineering, South China University of Technology. His research interests include personalized recommender system, text mining and real-time information retrieval. His work has been published in leading IS conferences like ICIS and IFIP.

FACULTY DISCUSSANT: Associate Professor Kaushik Dutta