Tutorial Title: Big Data Management and Mining

Abstract

To handle emerging complex data at massive scale from web, social network, and sensor network etc., "big data analytics” and “big data management” areas are emerging. Many traditional assumptions are not working, instead, new query and programming interfaces are required, and new computing models are emerging. The tutorial will focus on data mining and machine learning algorithms for analyzing very large amounts of data or Big data. Map Reduce and No SQL system will be used as tools/standards for creating parallel algorithms that can process very large amounts of data. The following concepts will be covered: Hadoop, Mapreduce, NoSQL systems (Cassandra, Pig, Hive, BigTable, HBASE), Storm, Spark, Large scale supervised machine learning, Data streams, Clustering, and Applications including recommendation systems, Web and security.

Tutors’ short bio:

Latifur R. Khan is currently a full professor (tenured) in the Computer Science department at the University of Texas at Dallas (UTD), where he has been teaching and conducting research since September 2000. He received his Ph.D. degree in Computer Science from the University of Southern California (USC), USA in August of 2000. Dr. Khan's research areas cover big data management and analytics, data mining, multimedia information management, and semantic web. He has published more than 180 papers including more than 45 journal papers. He is an ACM Distinguished Scientist. He has chaired several conferences and serves (or has served) as associate editor on multiple editorial boards including IEEE Transactions on Knowledge and Data Engineering (TKDE) journal. He has conducted tutorial sessions in prominent conferences such as ACM WWW 2005, IEEE CloudCom 2013, MIS2005, and DASFAA 2007.

More information about her research can be found at: http://www.utdallas.edu/~lkhan.