



Course CS 6350.002 – Big Data Management and Analytics
Professor Dr. Anurag Nagar
Term Spring 2016
Meetings Fri 4:00 PM – 6:45 PM, Location: ECSS 2.412

Professor's Contact Information

Office Phone TBA
Office Location ECSS 4.403
Email Address axn112530@utdallas.edu
Office Hours MW 2:30 – 3:45 PM
Other Information Walk-in during office hours is encouraged. At other times, email is the best method of contact.
MidTerm Friday March 4, 2015
Final Exam TBA

General Course Information

Pre-requisites, Co-requisites, & other restrictions Database Management Systems, JAVA (intermediate/advanced), Linux OS, Machine Learning/AI (co-requisite)

Course Description

Popular relational database systems like IBM DB2, Microsoft SQLServer, Oracle, and Sybase are struggling to handle massive scale of data introduced by the Web, Social network and cyber physical systems. Organizations have to deal with extremely large datasets. To handle emerging data at massive scale, "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 course 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 course material will be drawn from textbooks as well as recent research literature. The following topics will be covered this year: Hadoop, Mapreduce, NoSQL systems (Cassandra, Pig, Hive, BigTable, HBASE, SPARK), Storm, Large scale supervised machine learning, Data streams, Clustering, and Applications including recommendation systems, Web and security.

Learning Outcomes

Ability to understand of conceptual, logical and physical organization of big data
Ability to understand of large data processing using Map-Reduce
Ability to understand of NoSQL models, theory and practices
Ability to understand of data modeling, indexing, query processing for big data
Ability to understand of recommendation systems for big data
Ability to understand of unsupervised learning for big data
Ability to Understand of supervised learning for big data
Ability to communicate and work on team software project

- Texts & Materials**
- B1: Jimmy Lin and Chris Dyer, Data-Intensive Text Processing with MapReduce, Morgan & Claypool Publishers, 2010.
<http://lintool.github.com/MapReduceAlgorithms/>
 - B2: Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Addison-Wesley April 2005.
 - B3: Anand Rajaraman and Jeff Ullman, Mining of Massive Datasets, Cambridge Press,
<http://infolab.stanford.edu/~ullman/mmds/book.pdf>
 - B4: Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000. 550 pages. ISBN 1-55860-489-8. [Optional]

Assignments & Academic Calendar

Assignments will be through eLearning.

For academic calendar see:

<https://www.utdallas.edu/academiccalendar/files/AcademicCalendarSpring2016.pdf>

Course Policies

Grading (credit) Criteria	20% Midterm, 20% Final, 40% Assignments, 15% Project, and 5% Class Participation and Quizzes. In order to obtain an “A” or “A-” grade a student must perform well in the examinations, as well as in the assignments, and project. This is the minimum requirement, and satisfying this requirements does not guarantee an A or A- grade.
Late Work	5% penalty for each late day
Class Attendance	Voluntary, but strongly recommended.
Make-up Exams	Make-up examinations will be offered only if the student has a valid medical reason and produces a doctor’s letter. If a student misses an examination or quiz without prior notice, the student will forfeit the right to take a make-up examination or quiz at a later date. Exceptions will be made in rare situations, and entirely at the discretion of the instructor. Any request for such make-up examination should be accompanied by supporting document, e.g. a doctor’s note.
Classroom Citizenship	Please be considerate of fellow students and the instructor. Please turn off all electronic devices during class hours. Participate actively in classroom and online (through eLearning) discussion.
Comet Creed	<i>This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:</i> <i>“As a Comet, I pledge honesty, integrity, and service in all that I do.”</i>
UT Dallas Syllabus Policies and Procedures	<i>The information contained in the following link constitutes the University’s policies and procedures segment of the course syllabus.</i> <i>Please go to http://go.utdallas.edu/syllabus-policies for these policies.</i>

The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.