



Course CS 6350.002 – Big Data Management and Analytics

Professor Anurag Nagar, Ph.D.

Term Spring 2020

Meetings Through eLearning Collaborate videos

Professor's Contact Information

Office Phone 972-883-6345

Office Location ECSS 4.610

Email Address Anurag.Nagar@utdallas.edu

Office Hours Virtual office hours: Monday 12 noon to 5 PM or by appointment

Other Information Email is the best method of contact.

General Course Information

Pre-requisites, Co-requisites, & other restrictions Database Management Systems, JAVA (intermediate/advanced), Linux OS

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.

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

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://lntool.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.
	• B5: Tom White, Hadoop: The Definitive Guide, O'Reilly Media 4 th edition, April 2015
	• B6: Matei Zaharia, Spark: The Definitive Guide, O'Reilly Media 1 st edition, Feb 2018

Assignments & Academic Calendar

Exam 1 Friday Mar 13 at the testing center

Exam 2 Friday May 1 online through eLearning

Assignments will be through eLearning.

The academic calendar for the class will be posted on eLearning

Course Policies

Grading (credit) Criteria	25% Midterm, 25% Final, 25% Assignments, 10% Project, and 15% 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	You are expected to turn in your homework on time. To take care of extenuating circumstances, you will be allowed a total of 4 free late days for the entire semester. After those free days have been used, there will be a 10% penalty for each late day. You will only be able to turn the homework late up to 48 hours (2 days) after the due date. After that, the submission will be closed.

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

CS Department Attendance Policy	No compulsory attendance after spring break because of virtual classes. The CS department takes attendance and class participation seriously. Three consecutive absences lead to one letter grade drop. Four consecutive absences lead to an F. Official policy is stated here: http://cs.utdallas.edu/education/graduate/attendance-policy/
Extra Credit	None
Make-up Exams	No makeup exams at a later/earlier date will be scheduled for any student unless a written medical note is provided.
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.
Field Trip Policies	None
Student Conduct and Discipline	<p>The University of Texas System and The University of Texas at Dallas have rules and regulations for the orderly and efficient conduct of their business. It is the responsibility of each student and each student organization to be knowledgeable about the rules and regulations that govern student conduct and activities. General information on student conduct and discipline is contained in the UTD publication, A to Z Guide, which is provided to all registered students each academic year.</p> <p>The University of Texas at Dallas administers student discipline within the procedures of recognized and established due process. Procedures are defined and described in the <i>Rules and Regulations, Board of Regents, The University of Texas System, Part I, Chapter VI, Section 3</i>, and in Title V, Rules on Student Services and Activities of the university's <i>Handbook of Operating Procedures</i>. Copies of these rules and regulations are available to students in the Office of the Dean of Students, where staff members are available to assist students in interpreting the rules and regulations (SU 1.602, 972/883-6391).</p> <p>A student at the university neither loses the rights nor escapes the responsibilities of citizenship. He or she is expected to obey federal, state, and local laws as well as the Regents' Rules, university regulations, and administrative rules. Students are subject to discipline for violating the standards of conduct whether such conduct takes place on or off campus, or whether civil or criminal penalties are also imposed for such conduct.</p>

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

<p>Academic Integrity</p>	<p>The faculty expects from its students a high level of responsibility and academic honesty. Because the value of an academic degree depends upon the absolute integrity of the work done by the student for that degree, it is imperative that a student demonstrate a high standard of individual honor in his or her scholastic work.</p> <p>Scholastic dishonesty includes, but is not limited to, statements, acts or omissions related to applications for enrollment or the award of a degree, and/or the submission as one's own work or material that is not one's own. As a general rule, scholastic dishonesty involves one of the following acts: cheating, plagiarism, collusion and/or falsifying academic records. Students suspected of academic dishonesty are subject to disciplinary proceedings.</p> <p>Plagiarism, especially from the web, from portions of papers for other classes, and from any other source is unacceptable and will be dealt with under the university's policy on plagiarism (see general catalog for details). This course will use the resources of turnitin.com, which searches the web for possible plagiarism and is over 90% effective.</p>
<p>Email Use</p>	<p>The University of Texas at Dallas recognizes the value and efficiency of communication between faculty/staff and students through electronic mail. At the same time, email raises some issues concerning security and the identity of each individual in an email exchange. The university encourages all official student email correspondence be sent only to a student's U.T. Dallas email address and that faculty and staff consider email from students official only if it originates from a UTD student account. This allows the university to maintain a high degree of confidence in the identity of all individual corresponding and the security of the transmitted information. UTD furnishes each student with a free email account that is to be used in all communication with university personnel. The Department of Information Resources at U.T. Dallas provides a method for students to have their U.T. Dallas mail forwarded to other accounts.</p>
<p>Student Grievance Procedures</p>	<p>Procedures for student grievances are found in Title V, Rules on Student Services and Activities, of the university's <i>Handbook of Operating Procedures</i>.</p> <p>In attempting to resolve any student grievance regarding grades, evaluations, or other fulfillments of academic responsibility, it is the obligation of the student first to make a serious effort to resolve the matter with the instructor, supervisor, administrator, or committee with whom the grievance originates (hereafter called "the respondent"). Individual faculty members retain primary responsibility for assigning grades and evaluations. If the matter cannot be resolved at that level, the grievance must be submitted in writing to the respondent with a copy of the respondent's School Dean. If the matter is not resolved by the written response provided by the respondent, the student may submit a written appeal to the School Dean. If the grievance is not resolved by the School Dean's decision, the student may make a written appeal to the Dean of Graduate or Undergraduate Education, and the dean will appoint and convene an Academic Appeals Panel. The decision of the Academic Appeals Panel is final. The results of the academic appeals process will be distributed to all involved parties.</p> <p>Copies of these rules and regulations are available to students in the Office of the Dean of Students, where staff members are available to assist students in interpreting the rules and regulations.</p>

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

Incomplete Grades	<p>As per university policy, incomplete grades will be granted only for work unavoidably missed at the semester's end and only if 70% of the course work has been completed. An incomplete grade must be resolved within eight (8) weeks from the first day of the subsequent long semester. If the required work to complete the course and to remove the incomplete grade is not submitted by the specified deadline, the incomplete grade is changed automatically to a grade of F.</p>
Disability Services	<p>The goal of Disability Services is to provide students with disabilities educational opportunities equal to those of their non-disabled peers. Disability Services is located in room 1.610 in the Student Union. Office hours are Monday and Thursday, 8:30 a.m. to 6:30 p.m.; Tuesday and Wednesday, 8:30 a.m. to 7:30 p.m.; and Friday, 8:30 a.m. to 5:30 p.m.</p> <p>The contact information for the Office of Disability Services is: The University of Texas at Dallas, SU 22 PO Box 830688 Richardson, Texas 75083-0688 (972) 883-2098 (voice or TTY)</p> <p>Essentially, the law requires that colleges and universities make those reasonable adjustments necessary to eliminate discrimination on the basis of disability. For example, it may be necessary to remove classroom prohibitions against tape recorders or animals (in the case of dog guides) for students who are blind. Occasionally an assignment requirement may be substituted (for example, a research paper versus an oral presentation for a student who is hearing impaired). Classes enrolled students with mobility impairments may have to be rescheduled in accessible facilities. The college or university may need to provide special services such as registration, note taking, or mobility assistance.</p> <p>It is the student's responsibility to notify his or her professors of the need for such an accommodation. Disability Services provides students with letters to present to faculty members to verify that the student has a disability and needs accommodations. Individuals requiring special accommodation should contact the professor after class or during office hours.</p>

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

Religious Holy Days	<p>The University of Texas at Dallas will excuse a student from class or other required activities for the travel to and observance of a religious holy day for a religion whose places of worship are exempt from property tax under Section 11.20, Tax Code, Texas Code Annotated.</p> <p>The student is encouraged to notify the instructor or activity sponsor as soon as possible regarding the absence, preferably in advance of the assignment. The student, so excused, will be allowed to take the exam or complete the assignment within a reasonable time after the absence: a period equal to the length of the absence, up to a maximum of one week. A student who notifies the instructor and completes any missed exam or assignment may not be penalized for the absence. A student who fails to complete the exam or assignment within the prescribed period may receive a failing grade for that exam or assignment.</p> <p>If a student or an instructor disagrees about the nature of the absence [i.e., for the purpose of observing a religious holy day] or if there is similar disagreement about whether the student has been given a reasonable time to complete any missed assignments or examinations, either the student or the instructor may request a ruling from the chief executive officer of the institution, or his or her designee. The chief executive officer or designee must take into account the legislative intent of TEC 51.911(b), and the student and instructor will abide by the decision of the chief executive officer or designee</p>
Withdrawal from Class	<p>The administration of this institution has set deadlines for withdrawal of any college level courses. These dates and times are published in that semester's course catalog. Administration procedures must be followed. It is the student's responsibility to handle withdrawal requirements from any class. In other words, I cannot drop or withdraw any student. You must do the proper paperwork to ensure that you will not receive a final grade of "F" in a course if you choose not to attend the class once you are enrolled.</p>
Comet Creed	<p><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></p> <p><i>"As a Comet, I pledge honesty, integrity, and service in all that I do."</i></p>
UT Dallas Syllabus Policies and Procedures	<p><i>The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus.</i></p> <p><i>Please go to http://go.utdallas.edu/syllabus-policies for these policies.</i></p>

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