Remote / Online Course Syllabus

Course Information

Course Number/Section	CS 6350.002
Course Title	Big Data Management and Analytics
Term	Fall 2020
Meeting Times	M, W 2:30 – 3:45 PM

Professor Contact Information

Professor	Anurag Nagar, Ph.D.
Office Phone	972-883-6345
Email Address	anurag.nagar@utdallas.edu
Office Location	ECSS 4.610
Online Office Hours	M, W 12:30 - 2:30 PM through MS Teams
Exams	Exam 1: Monday October 5
(Tentative)	Exam 2: Wednesday December 2

Course Modality and Expectations

Instructional Mode	Remote For details, see https://www.utdallas.edu/fall-2020/fall-2020-registration-information/
Course Platform	This course will be delivered live using MS Teams. Each student will receive a meeting invite through their UTD email.
Expectations	You are encouraged to attend live sessions, which will also be recorded and posted on MS Stream after each class. This course will use Honorlock – an online exam proctoring tool. To successfully take an exam, you must have a web camera with microphone, a laptop or desktop computer (no tablets/phones), Chrome browser, a reliable internet connection and your photo ID. You will be prompted to install the Honorlock Chrome Extension (which you can remove after you finish the test). You will then access the exam within your eLearning course and go through the authentication process. The web camera will monitor you throughout your test. Please see the Testing Guidelines and Support Information for additional information.
Asynchronous Learning Guidelines	If you choose the asynchronous option, you can view the posted lectures, and submit quizzes and assignments by the mentioned deadlines. You will be required to take exams during the same time block as the rest of the class. For more details, see https://www.utdallas.edu/fall-2020/asynchronous-access-for-fall-2020/

COVID-19 Guidelines and Resources

The information contained in the following link lists the University's COVID-19 resources for students and instructors of record.

Please see http://go.utdallas.edu/syllabus-policies.

Class Participation

Regular class participation is expected regardless of course modality. Students who fail to participate in class regularly are inviting scholastic difficulty. A portion of the grade for this course is directly tied to your participation in this class. It also includes engaging in group or other activities during class that solicit your feedback on homework assignments, readings, or materials covered in the lectures (and/or labs). Class participation is documented by faculty. Successful participation is defined as consistently adhering to University requirements, as presented in this syllabus. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

Class Recordings

The instructor may record meetings of this course. Any recordings will be available to all students registered for this class as they are intended to supplement the classroom experience. Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, students are expressly prohibited from recording any part of this course. Recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. If the instructor or a UTD school/department/office plans any other uses for the recordings, consent of the students identifiable in the recordings is required prior to such use unless an exception is allowed by law. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

Class Materials

The Instructor may provide class materials that will be made available to all students registered for this class as they are intended to supplement the classroom experience. These materials may be downloaded during the course, however, these materials are for registered students' use only. Classroom materials may not be reproduced or shared with those not in class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

Course Pre-requisites, Co-requisites, and/or 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.

Student Learning Objectives/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

Required Textbooks and Materials

Required Texts
None

Suggested Course Materials

Suggested Readings/Texts

• 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.
• B5: Tom White, Hadoop: The Definitive Guide, O'Reilly Media 4th edition, April 2015
• B6: Bill Chambers, Matei Zaharia, Spark: The Definitive Guide, O'Reilly Media, February 2018

Textbooks and some other bookstore materials can be ordered online or purchased at the UT Dallas Bookstore.

Technical Requirements

In addition to a confident level of computer and Internet literacy, certain minimum technical requirements must be met to enable a successful learning experience. Please review the important technical requirements on the Getting Started with eLearning webpage.

Course Access and Navigation

This course can be accessed using your UT Dallas NetID account on the eLearning website.

Please see the course access and navigation section of the Getting Started with eLearning webpage for more information.

To become familiar with the eLearning tool, please see the Student eLearning Tutorials webpage.

UT Dallas provides eLearning technical support 24 hours a day, 7 days a week. The eLearning Support Center includes a toll-free telephone number for immediate assistance (1-866-588-3192), email request service, and an online chat service.

Communication
This course utilizes online tools for interaction and communication. Some external communication tools such as regular email and a web conferencing tool may also be used during the semester. For more details, please visit the Student eLearning Tutorials webpage for video demonstrations on eLearning tools.

Student emails and discussion board messages will be answered within 3 working days under normal circumstances.

Distance Learning Student Resources
Online students have access to resources including the McDermott Library, Academic Advising, The Office of Student AccessAbility, and many others. Please see the eLearning Current Students webpage for more information.

Server Unavailability or Other Technical Difficulties
The University is committed to providing a reliable learning management system to all users. However, in the event of any unexpected server outage or any unusual technical difficulty which prevents students from completing a time sensitive assessment activity, the instructor will provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor and also contact the online eLearning Help Desk. The instructor and the eLearning Help Desk will work with the student to resolve any issues at the earliest possible time.

Academic Calendar
Posted on eLearning

Grading Policy
Grading will be done using the breakdown shown below:

Homework Assignments: 25% of the course grade
Quizzes and Class Participation: 15% of the course grade
Term Project: 10% of the course grade
Exam 1: 25% of the course grade
Exam 2: 25% of the course grade

There will be no extra credit assignments or exams. Grading will be done on a relative scale.

Course Policies
Make-up exams
Make-up exams will only be allowed for verified medical emergencies.

Extra Credit
None

Late Work
Barring extenuating circumstances, all problem sets must be turned in on the date specified. Assignments turned in within 24 hours of the due date will receive 90% of its score. Assignments turned in within 48 hours of the due date will receive 80% of its score. Assignments more than 48 hours late will not be accepted.

Class Participation
Class participation is encouraged from all students.

Classroom Citizenship
Please be considerate of fellow students and the instructor. Please turn off all electronic devices during class hour. Participate constructively in classroom discussion.

Comet Creed
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:

"As a Comet, I pledge honesty, integrity, and service in all that I do."

Academic Support Resources
The information contained in the following link lists the University's academic support resources for all students.

Please go to Academic Support Resources webpage for these policies.

UT Dallas Syllabus Policies and Procedures
The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus.

Please go to UT Dallas Syllabus Policies webpage for these policies.

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