Big Data Syllabus

Course Information

Course Number/Section	MIS 6346.001
Course Title	Big Data
Term	Fall 2020

Professor Contact Information

Professor	Dr. Judd D. Bradbury
Office Phone	972-883-4873
Mail Contact	e-Learning Course Messages (first priority)
Office Location	JSOM 3.220
Office Hours	Tuesday 4:00 – 5:00 PM

Course Modality and Expectations

Instructional Mode	Remote/Virtual
Course Platform	Course content will be delivered in a fully digital manner posted on eLearning. Collaborate will be used for synchronous lectures, virtual course meetings, and optional labs. Connection links will be provided for all meetings using an eLearning announcement/email. Office hours will be conducted using MS Teams for students that have requested them.
Expectations	Students should engage lectures and course materials in a timely manner as designated in the syllabus schedule. Students should follow the course policies.
Asynchronous Learning Guidelines	Students enrolled in the course may engage asynchronous learning. All lecture content will be recorded and posted in eLearning. Asynchronous students can easily use the course schedule in the syllabus as a guide. Students that choose to participate asynchronously are required to watch the recordings of course meetings and synchronous lectures. All students are required to engage live sessions or recordings along with course announcements to ensure they understand course direction and requirements.

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

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. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

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 Description

The course covers Big Data concepts, architecture, and hands-on use of several tools in the Hadoop Ecosystem. The course will cover the theoretical as well as hands-on. The tools covered include Linux, Hadoop, Sqoop, Flume, Hive, Pig, and Spark.

The course is technically demanding requiring programming and quick assimilation of a large number of tools. Please understand that in order to effectively learn the concepts and approaches in this course, you will need to work very hard for long hours every week. If you like to wait until the day before your assignments are due to begin your work, you should drop this course. If you are up for this personal responsibility, please join us!

Student Learning Objectives
- Learn about Linux, Hadoop, HDFS, MapReduce, Hive, Pig, Spark, Sqoop, and Flume
- Be able to successfully ingest and manipulate data in Hive
- Perform analysis in Spark

Required Textbooks
The following textbooks are available electronically for free in the UT Dallas Library. Click on the link below and search for the title.

http://www.utdallas.edu/library/

Linux The Linux Command-line: A Complete Introduction, 1st Edition (January 2012) *by William E. Shotts, Jr.*
No Starch Press
https://www.nostarch.com/tlcl
https://www.amazon.com/Linux-Command-Line-Complete-Introduction/dp/1593273894

Hadoop Hadoop: The Definitive Guide, 4th Edition (March 2015) *by Tom White*
O'Reilly Publishing

Pig Programming Pig, 2nd Edition (December 2016) *by Alan Gates and Daniel Dai*
O'Reilly Publishing

Spark Spark: The Definitive Guide, 1st Edition (February 2018) *by Matei Zaharia and Bill Chambers*
O'Reilly Publishing

Required Materials

Registration with Polleverywhere.com

Laptop with a minimum of 16GB of RAM to run a virtual machine sandbox. You need to be able to troubleshoot your own connectivity issues. We cannot be responsible for every hardware/firewall/security configuration.

We will be running a Cloudera virtual machine sandbox.

Course Policies
A solemn duty will be upheld by your professor to maintain a level playing field for all students.

- **Assignments are due at midnight on the due dates defined in the syllabus**
- **Assignments must be submitted through e-learning**
- **Do not use the Microsoft Edge browser for submissions as it will damage documents**

- **Students are provided only one submission attempt per assignment**
- **Late assignments receive a 25% discount more than 7 calendar days late is a 0/100**
- **Completing assignments is 100% a student responsibility, we help, we do not tutor**
- **Interim quizzes will be provided to ensure students study the material each week**
- **Students are required to attend every class, speak, and monitor e-learning daily**
- **Missed exams will be provided with a 0/100**
- **We do not provide extra credit assignments**
- **A signed note from a medical doctor will be required for any grading impacted policy**
- **Colds, headaches, upset stomach and/or flu are not acceptable excuses for missing class or deliverables**
- **Students will receive the grades they earn, requests to change grades will be deleted**
- **We will only discuss grade calculation errors, we will not discuss your grade preference**

Communication
This course utilizes online tools for interaction and communication. For more details, please visit the eLearning Tutorials webpage http://www.utdallas.edu/elearning/students/eLearningTutorialsStudents.html for video demonstrations on eLearning tools.

- **Instructor will communicate with announcements and discussion board postings**
- **Communication sequence is:**
    1. **Discussion board posting**
    2. **Attend optional lab**
    3. **Course message to professor (If no response to discussion board in 24 hours)**
- **Individual concerns or questions should be sent in a course message**
- **Professor and teaching assistants are off duty on weekends**
- **Discussion board postings should always include:**
    1. **Exercise and step number**
    2. **Description of the last step that worked**
    3. **Screen shot of the error**
- **Discussion board postings are not markers that can be placed on the professor**
- **Blanket e-mails or lobbying your classmates is inappropriate student conduct**
- **Discussion board posts regarding course policies and/or answers will be deleted**

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 http://www.utdallas.edu/elearning/students/getting-started.html#techreqs on the Getting Started with eLearning webpage http://www.utdallas.edu/elearning/students/getting-started.html.

Technical Specifications (Recommended)

RAM - 16GB
Processor - Intel i5 2.4 Ghz (minimum)
Graphics Processor - 512 MB (Dedicated)

Course Access and Navigation
The course can be accessed using the UT Dallas NetID account at: https://elearning.utdallas.edu. Please see the course access and navigation http://www.utdallas.edu/elearning/students/getting-started.html#courseaccessandnav section of the site for more information.

To become familiar with the eLearning tool, please see the Student eLearning Tutorials http://www.utdallas.edu/elearning/students/eLearningTutorialsStudents.html.
UT Dallas provides eLearning technical support 24 hours a day/7 days a week. The eLearning Support Center http://www.utdallas.edu/elearninghelp services include a toll-free telephone number for immediate assistance (1-866-588-3192), email request service, and an online chat service.

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 page http://www.utdallas.edu/elearning/students/cstudents.htm for details.

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 http://www.utdallas.edu/elearninghelp. The instructor and the eLearning Help Desk will work with the student to resolve any issues at the earliest possible time.

Assignments & Academic Calendar

WEEK BEGIN	TOPIC/LECTURE	READING	ASSESSMENT / ACTIVITY	EXERCISE DUE DATE
1 8/17	Big Data Overview	**The Linux Command Line, Chapters 1-4.**	**Exercise 1 – Setup Virtual Sandbox**	**8/21**
2 8/24	Linux	**The Linux Command Line, Chapters 5,6, and 12.**	**Exercise 2 – Linux**	**8/28**
3 8/31	Hadoop HDFS	**Hadoop: The Definitive Guide, Chapter 1 and Chapter 3.**	**Exercise 3 – Hadoop HDFS**	**9/4**

4 9/7	Map Reduce	**Hadoop: The Definitive Guide, Chapter 2, Chapters 4, and Chapter 7.**	**Exercise 4 – Map Reduce**	**9/11**
5 9/14	Flume & Sqoop	**Hadoop: The Definitive Guide, Chapters 14.** **Hadoop: The Definitive Guide, Chapters 15.**	**Exercise 5 – Flume** **Exercise 6 – Sqoop**	**9/18**
6 9/21	Hive 1	**Hadoop: The Definitive Guide, Chapter 17.**	**Exercise 7 – Hive 1**	**9/25**
7 9/28	Hive 2		**Exercise 8 – Hive 2**	**10/2**
8 10/5	Midterm Exam Review		**Midterm Exam**	**10/9**
9 10/12	Pig 1	**Hadoop: The Definitive Guide, Chapter 16.**	**Exercise 9 – Pig 1**	**10/16**
10 10/19	Spring Break		Spring Break	**10/23**
11 10/26	Pig 2	**Programming Pig, Chapters 4-6.**	**Exercise 10 – Pig 2**	**10/30**
12 11/2	Spark Overview	**Spark The Definitive Guide, Chapter 2-3.**	**Case Study 1 - Pig & Hive**	**11/6**

13 11/9	Spark Operations	**Spark The Definitive Guide, Chapter 4-6.**	**Exercise 11 – Data Frames**	**11/13**
14 11/16	Advanced Spark	**Spark The Definitive Guide, Chapter 7-8 & Chapter 10.**	**Case Study 2 - Spark Machine Learning**	**11/20**
15 11/23	Machine Learning in Spark Exam Review	**Spark The Definitive Guide, Chapter 24 & 29.**	**Case Study 2 - Spark Machine Learning**	**11/27**
16 11/30			**Final Exam**	**12/4**

Exam Procedures

Exams will be administered electronically using a remote eLearning online exam.

Grading Policy

Weights

Assignments(best 9 of 11)		38%
Case Studies		10%
Midterm Exam		25%
Final Exam		25%
Quiz, Class Policy Test & Course Evaluation		2%
Total		100%

Grading Scale

Scaled Score	Letter Equivalent
>= 93	A
>= 90	A-
>= 86	B+
>= 83	B

>= 80	B-
>= 76	C+
>= 70	C
< 70	F

Comet Creed

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

Scholastic Dishonesty Referrals

Cheating in course sections is pervasive and consistent. We will report all instances of scholastic dishonesty. If you attempot to answer quiz questions presented live in a classroom, from a remote location, as if you are present, you are cheating. Students are required to use their own personal computer for any and all assignments for this class. Some students make poor decisions to copy another student's work instead of doing their own. When they get caught they suggest someone just borrowed by computer. You will be responsible for any copying performed from your computer. Usage of another student's system, or provision of your system to another student without providing prior notification to the professor is scholoastic dishonesty.

Students cheating on an item in this course will receive a scholastic dishonesty referral including a requested sanction of a zero for the entire category of their final course grade. (Example: Cheating on Assignment 7 will earn a requested sanction of a zero for all homework assignments.)

Students should review the link on student policies to ensure they understand all of the actions that are considered cheating.

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 http://go.utdallas.edu/syllabus-policies for these policies.

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