Syllabus

BUAN 6346.505/MIS 6346.505 Big Data Analytics

Dr. Waseem Shadid

Spring -2024 / Wednesday 7:00 pm – 9:45 pm.

Traditional Classroom / JSOM 2.722

Professor Information

Email Address: Waseem.Shadid@utdallas.edu
Office Location: Microsoft Office Teams
TA Information Rajarshi Boggarapu (rajarshi.boggarapu@utdallas.edu)
Office Hours: TBD

Course Modality and Expectations

Traditional learning mode Face-to-face	The course will be taught face-to-face. Instructor and students meet according to the schedule.
Course Platform	Blackboard
Expectations	Class Attendance is mandatory. Lectures will NOT be recorded.
Asynchronous Learning	Asynchronous learning is not available in this class. Students aquire Asysnchronous learning mode need to drop and re-enroll a different class

Pre-requisites

MIS 6326. MIS 6320, or BUAN 6320

Course description

The big Data landscape is continuously evolving as new technologies emerge and existing technologies mature. This is a comprehensive course covering Spark, and key elements of the Hadoop Ecosystem used in developing end-to-end applications for processing Big Data, efficiently. Students who complete this course will understand key concepts of Spark and Hadoop, and they will learn to apply Spark and Hadoop tools in developing applications for solving the types of problems faced by enterprises and research institutions today. The tools covered in this course include Sqoop, Hive, Impala, Pig, Flume, and Spark.

Learning Outcomes

1) Students will be able to describe architecture and methods for storage and provision in Hadoop
2) Students will develop competency in storing, querying, and processing data in HDFS
3) Students will demonstrate competency in importing different types of data into Hadoop. In addition, students will learn Spark - a framework for processing data
4) Students will learn steps involved in processing data in Hadoop environment from end-to-end perspective

Optional Text and Materials (Professor will provide PDF copy for the following)

- O'Reilly Sqoop Cookbook by Ting and Cecho
- O'Reilly Programming Hive by Rutherglen, Wampler, and Capriolo
- O'Reilly Programming Pig by Alan Gates
- O'Reilly Learning Spark by Karau and Zahaia

Hardware Requirements

- You must have at-least **8GB RAM** on your computer
- You must bring your laptop to every class due to the hands-on nature of the class
  (Please send me an email if you won't be able to meet the minimum Hardware requirements)

Software Used

Cloudera VM – will be made available by the instructor

Tentative Assignments & Academic Calendar*

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

1	01/17/2024	• Introductions and course details • Syllabus Overview and Expectations	Lab#1
2	01/24/2024	• Hadoop Architecture • Data Storage in Hadoop • Hands-on: Basic Linux Commands	Lab#2
3	01/31/2024	• YARN • Hadoop Ecosystem • Hands-on: HDFS Commands	
4	02/07/2024	**Sqoop** • Sqoop Architecture • Sqoop commands • Importing data using Sqoop • Sqoop Hands-on	Lab#3 Lab#4
5	02/14/2024	**Hive and Impala Lecture 1** • Impala and Hive Architecture • Hive commands	Lab#5
6	02/21/2024	**Impala and Hive** • Impala and Hive Hands-on • Handling Complex data structures in Hive • Handling JSON format in Hive • Data Partitioning in Hive	Lab#6 Lab#7
7	02/28/2024	**Review**	
8	**03/06/2024**	**Midterm Exam**	**Testing Center**
9	**03/13/2024**	**Spring Break**	
10	03/20/2024	**Introduction to Streaming Systems in Hadoop** • Flume architecture • Flume Hands-on • Pig Hands-on	Lab#8 Lab#9
11	03/27/2024	**Introduction to Spark** • Spark Shell • Spark Context	Lab#10
12	04/03/2024	Resilient Distributed Datasets RDDs	Lab#11
13	04/10/2024	**Working with RDDs** • Generic and Pair RDDs Spark Hands-on • Loading data into Spark from different sources • Spark SQL and Data Frames • More Spark SQ Transforming and Querying Data Frames	Lab#12 Lab#13
14	04/17/2024	**More advanced topics in Spark RDDs** • Spark SQL and Data Frames	Lab#14

		• More Spark SQL • Transforming and Querying Data Frames	
15	04/24/2024	**Project presentation**	Lab#15 Lab#16
16	05/01/2024	**Review**	
17	**05/08/2024**	**Final Exam**	**Testing Center**

Exam Registration

Registration Deadlines: All students must reserve a time slot no later than **48 hours prior to exam appointment time** at https://ets.utdallas.edu/testing-center/students/ **WALK-IN APPOINTMENTS ARE NOT ALLOWED - NO EXCEPTIONS!**

Course Policies

- The Labs are to be completed in class and are due by the end of class, unless otherwise stated in eLearning.
- Makeup Exam: There is no makeup exams. In case of medical emergency, a medical report is required including physician information.
- Missing exam: Any missing exam without medical report will be graded as Zero.
- Assignments must be submitted through eLearning. Emailed submissions are not accepted.
- Late Assignments: Subject to 10% penalty, 30% penalty after the third day.
- Class Attendance: Students who fail to attend class regularly are inviting scholastic difficulty. Absences may lower a student's grade where class attendance and class participation are deemed essential by the instructor.
- UTD Syllabus Policies and Procedures: Please visit https://go.utdallas.edu/syllabus-policies
- Cheating will not be tolerated. When I find evidence of cheating, the documentation is turned over to the Office of Community Standards and Conduct. (https://www.utdallas.edu/conduct/dishonesty/)

Academic Integrity:

In general, academic dishonesty involves the abuse and misuse of information or people to gain an undeserved academic advantage or evaluation. The common forms of academic dishonesty include:

- Cheating – using deception in the taking of tests or the preparation of written work, using unauthorized materials, copying another person's work with or without consent, or assisting another in such activities.
- Lying – falsifying, fabricating, or forging information in either written, spoken, or video presentations.
- Plagiarism—using the published writings, data, interpretations, or ideas of another without proper documentation

  Plagiarism includes copying and pasting material from the internet into assignments without properly citing the source of the material. Episodes of academic dishonesty are reported to the Vice President for Academic Affairs. The potential penalty for academic dishonesty includes a failing grade on a particular assignment, a failing grade for the entire course, or charges against the student with the appropriate disciplinary body.

Grading scale

Top 25% Students or greater than or equal to 90	**A**
Next 25% Students or greater than or equal to 85 but less than 90	**A-**

Next 25% Students or greater than or equal to 80 but less than 85	**B+**
Remaining	**B and below**

Calculated Grade Weights**

- Labs / Assignments (25%)
- Exam I (Midterm) – (15%)
- Exam 2 (Final) – (20%)
- Project – (15%)
- Quizzes (10%)
- Class attendance (10%)
- Class participations and engagements (5% - Hidden)

***The calculated grade weights are subject to change at the discretion of the Professor.*

Classroom citizenship

- Cell phone use is not allowed during class or exam.
- eLearning will be used for class content.
- Slides and other class materials will be posted after class is held.
- Class announcements (e.g., change in assignment dates) will be posted in the eLearning announcements. It is the students' responsibility to regularly check the announcements (typically by having the announcement automatically forwarded to their email accounts).

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 https://go.utdallas.edu/syllabus-policies for these policies.
- The University of Texas at Dallas is committed to providing reasonable accommodations for all persons with disabilities. The syllabus is available in alternate formats upon request. If you are seeking classroom accommodations under the Americans with Disabilities Act (2008), you are required to register with the Office of Student AccessAbility, located in the Administration Building, Suite 2.224. Their phone number is 972-883-2098, email: studentaccess@utdallas.edu and website is https://studentaccess.utdallas.edu . To receive academic accommodations for this class, please obtain the proper Office of Student AccessAbility letter of accommodation and meet with me at the beginning of the semester.

Academic Support Resources

- The information contained in the following link lists the University's academic support resources for all students.
- Please see http://go.utdallas.edu/academic-support-resources.

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