UPDATED Aug 19, 2022

Big Data - Course Syllabus
BUAN 6346 / MIS 6346 - Fall 2022

Course Information

Course Number/Section BUAN 6346.001/MIS 6346.001
Course Title: Big Data
Term: Fall 2022 from Aug 22, 2022 to Dec 16, 2022
Class Hours & Location Thursday 4PM to 6:45PM @JSOM 12.210

Professor Contact Information

Professor	Antonio Paes
Email Address	antonio.paes@utdallas.edu *(preferred contact method)*
Office Hours	Please schedule appointment by email

Course Description

The course covers:

- Theoretical and Practical aspects of Big Data
- Business driver for Big Data
- Distributed systems
- Hadoop framework and related tools
    - Installation
    - Hadoop Architecture
    - HDFS
    - Yarn
    - MapReduce
    - Sqoop
    - Flume
    - Hive
    - Spark
    - Current and future concepts of big data

UPDATED Aug 19, 2022

Student Learning Objectives/Outcomes

- Understanding Big Data concepts, architectures and analysis
- Understanding Hadoop Ecosystem and its tools
- Learning fundamentals of Spark framework for processing data
- Developing hands-on experience with the tools to move, process and query data
- Reach business like conclusions

Text Books

Main Books

- Hadoop: The definitive guide by Tom White
- Practical Data Science with Hadoop and Spark by Ofer Mendelevitch
- Learning Spark by Jules S. Damji
- Big Data - Principles and Best practices of scalable real-time data systems by Nathan Marz

Addition Books

- Hadoop Application Architectures by Mark Grover
- Hadoop in 24 Hours by Jeffrey Aven
- Practical Hive by Scott Shaw
- Hadoop in Practice by Alex Holmes
- Apache Spark in 24 Hours by Jeffrey Aven
- Programming Hive by Edward Capriolo

Required Materials

Laptop: with at least 8GB RAM.
Virtualization Environment: Virtual Box must be installed
We will go through the installation process of all the required software.
Download Ubuntu 20.04 LTS (https://ubuntu.com/download/desktop)
All software can be used/downloaded at no cost

UPDATED Aug 19, 2022

Course Materials

All notes will be posted on eLearning (https://elearning.utdallas.edu/)

Tentative Course Schedule*

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

Week	Topic	Assignment & Others
1	First Day - Introduction - Syllabus - History: From files to DBs to Big Data - Big Data Architecture	
2	- Hadoop general Architecture - Installing VM, Linux, Hadoop and its dependencies - Big Data - Ethics and Social impact	
3	- HDFS - Career Paths and certification - Description of the group project	
4	- Project - Data Sources - Yarn - MapReducer	
5	- Flume - Sqoop	
6	- Hive 1 - Basic and Data Analysis	
7	- Hive 2 - Data Analysis	
8	**Mid Term Exam**	
9	- Spark - Basics	

	- Installation	
10	- Spark Hands On - Spark SQL - Spark RDD - Jupiter Notebooks on Spark - Cloud Solutions - Dataframes - Data Mesh - Future of Big Data	
11	- Projects Presentations	
12	**Final Exam**	

About Classes

- Classes content will be posted weekly

About Assignments

Individual Homework Assignments:

- There will be multiple individual homework assignments during the semester.
- Assignments must be submitted through eLearning on time (please do not wait till last minute, put at least one hour buffer between the deadline and your planned submissions time)
- Submissions **emailed** to the Instructor and/or TA will **not count**.
- Delayed assignments **will not be graded**

Group Project:

- Groups of will be formed, randomly selected by the instructor

Course Policies

UPDATED Aug 19, 2022

- Makeup Exam: There are no makeup exams. In case of a medical emergency, a medical report is required including physician information.
- Missing exam: Any missing exam without a medical report will be graded as Zero.
- Assignments must be submitted through eLearning. Emailed submissions are **not accepted and will be graded as zero.**
- 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

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.

UPDATED Aug 19, 2022

Grading Scale

Grade	Min	Max
A+	96	100
A	93	96
A-	89	92
B+	85	88
B	81	84
B-	77	80
C+	73	76
C	69	72

Calculated Grade Weights**

- Assignments (20%)
- Mid Term Exam (20%)
- Final Exam (30%)
- Group Project (30%)

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

Classroom citizenship

- 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).

UPDATED Aug 19, 2022

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.

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.