

## ***BUAN/MIS 6346 Big Data Course Syllabus Fall 2019***

---

### **Course Information**

<i>Course Number/Section</i>	MIS 6346/ BUAN 6346 Section 501
<i>Course Title</i>	Big Data
<i>Term</i>	Fall 2019
<i>Class Hours &amp; Location</i>	Tuesdays 7:00 - 9:45pm. JSOM 1.102
<i>TA:</i>	Jayanth Jayaraman <jxj180006@utdallas.edu>
<i>TA Office Hours</i>	TBD

### **Professor Contact Information**

<i>Professor</i>	Dr. Hakki C. Cankaya
<i>Email Address</i>	<a href="mailto:hakki.cankaya@utdallas.edu">hakki.cankaya@utdallas.edu</a> (email is preferred method to communicate with the instructor)
<i>Office Hours</i>	Tuesdays 6:00 - 6:50pm in JSOM 3.604 or by appointment

### **Course Description**

The course covers theoretical and practical aspects of Big Data, Hadoop and Spark Ecosystem including implementation, concepts, architecture, and tools. The course plans to cover the following tools and systems: Hadoop, HDFS, Sqoop, Hive, Impala, Flume, and Spark.

### **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

### **Required and Optional Textbooks and Materials**

#### *Required Text Books*

- 1) Big Data Science and Analytics: A Hands-On Approach, by Arshdeep Bagha and Vijay Madiseti. ISBN- 9780996025539
  
- 2) Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney, ISBN-13: 978-1491957660, ISBN-10: 1491957662

#### *Optional Book(s)*

- 3) Spark: The Definitive Guide, by Bill Chambers and Matei Zaharia. ISBN: 978-1-491-91221-8
- 4) Hadoop: The definitive guide by Tom White
- 5) Linux Command Line by William Shotts

### *Required Materials*

Laptop: with at least 8GB RAM.

Virtualization Environment: You are required to set up a virtualization environment (VirtualBox or VMware) on your Laptop and install Cloudera Distribution Student VM. You need to have at least 8GB RAM on your Laptop if you plan to run the VM on your computer. Another alternative is to buy service on a public cloud service like Microsoft Azure or others to host the Cloudera VM (this option will not be supported by the course instructor or TA). You will be responsible for your Laptop virtualization environment and/or your cloud service; the instructor will not have time and/or resources to solve problems with students' computing environment and connectivity issues.

Python IDE: You need to set up Python on your computer and bring it to every class (software is available at no cost).

### **Course Materials**

#### *Class Notes*

Class notes and Cloudera slide decks will be posted on eLearning.

### **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 Instructor and/or TA will not count.

**Group Project:** Students will be formed into groups for term project.

### **Tentative \* Course Schedule**

<b>WEEK/ DATES</b>	<b>TOPIC/LECTURE</b>	<b>ASSIGNMENTS</b>
1 Aug.20	Syllabus, Introduction to Databases. Intro to Big Data and Big Data stack	Big Data Text book Ch1, Ch2 HW-0
2 Aug. 27	Big Data Platforms and Patterns, Distributed systems.	Big Data Text book Ch3, Ch4 HW-1

3 Sep.3	Linux review, Hadoop Eco system and Architecture, NoSQL	HW-2
4 Sep.10	Python intro. Data Acquisition. Importing relational data with Sqoop	HW-3
5 Sep.17	Hive&Impala, Data formats, Python data structures, functions, and files.	HW-4
6 Sep.24	Data formats, Data file partitioning.	HW-5
7 Oct.1	Flume, Spark Basics.	HW-6
8 Oct.8	Exam 1	
9 Oct.15	Working with Spark RDD, Aggregating data Map-reduce	HW-7
10 Oct.22	Midterm Project reviews, Spark Apps	
11 Oct.29	Spark Data Processing	HW-8
12 Nov.5	Spark Data Processing II	
13	Project Presentations I	

Nov.12		
14 Nov.19	Project Presentations II	
15 Nov.26	Fall Break. No Class	
16 Dec.3	Exam 2	

\* Timeline and content of the course schedule are subject to change by the instructor.

### Grading Policy

- 30% Exam 1
- 30% Exam 2
- 20% Homeworks
- 20% Project

### Grading Scale:

93-100 A
90-92 A-
85-89 B+
80-84 B
75-79 B-
65- 74 C+
60-64 C
Below 60 F

### Course Policies

#### *Make-up exams*

There will be no make-up exams, except for medical emergency (written statement justifying the situation from a physician required). The written statement should include the physician's address and phone number for the verification purpose. Any missed exam will earn zero grade.

#### *Extra Credit*

There is no extra credit offered for the course.

### *Late Work*

Any assignment that is not submitted by the closing time of the eLearning system for that particular assignment is a late assignment. Late assignments are subject to penalties: 1 day (24 hour window following the submission time) late brings 20% penalty. The work will not be accepted, if it is late more than a day. As the best practice, pls institute at least one hour buffer between the submission deadline for the assignment and your planned submission time.

### *Classroom Citizenship*

Use of cell phone is not allowed in the class and during the exams. Cheating during the exams and/or assignments will not be tolerated and will be turned over to the responsible office in the school.

### **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.”

### **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.***