MATH 6397 – Big Data Analytics

Class Number: 18225
Semester: Summer 2019 [June 03 – July 24]
Time: Friday 3:00 – 5:00 PM
Class Room: AH 301

Instructor: Dr. Dvijesh Shastri
Office: PGH 677
E-mail: shastriD@uhd.edu
Office Hours: 2:00 – 3:00 PM [Friday] or by appointment

Catalog Description: The course introduces concepts and techniques in managing and analyzing large data sets for data discovery and modeling. Topics include big data storage systems, parallel processing platforms, and scalable machine learning algorithms.

Course Prerequisites: With consent of the instructor.

Learning Objectives: After taking this course, students should be able to

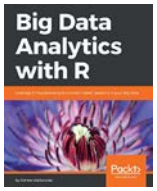
LO1. Explain how big-data has become a new norm of life and what challenges are associated with it in building models and ultimately, discovering knowledge.
LO2. Apply tools and algorithms required to store and analyze big data. In particular, the students will obtain knowledge on big-data processing workflow using the Hadoop Distributed File System (HDFS) and the MapReduce framework.
LO3. Explain and utilize big data storages including relational databases such as MySQL, PostgreSQL, and MariaDB, and non-relational databases such as document-based NoSQL, MongoDB, and a distributed Apache Hadoop-complaint HBase database.
LO4. Design highly scalable systems for analyzing big data though the Apache Spark ecosystem.

Mode of Instruction: The mode of the instruction is hybrid, meaning the course features both face-to-face and online learning. The face-to-face learning will occur during the class time. For online learning, the course material (e.g., online assignments, PowerPoint slides, etc.) will be made available on blackboard periodically, typically, right before and/or right after face-to-face meetings. The online component will require students to assume responsibility of an active learner, completing tasks necessary to understand the material.

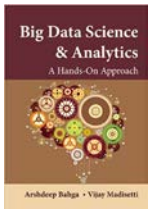
Online Course Support: The Blackboard system (https://elearning.uh.edu) will be used for online course material. As the semester progresses, various materials will be posted there including lecture notes, projects, and course announcements.

Reading Material

- **Textbook:** Big Data Analytics with R: Leverage R Programming to uncover hidden patterns in your Big Data, by Simon Walkowiak, 2016, 1st Edition, Packt Publishing. **(TB)**
  ISBN 13: 978-1786466457



- **Recommended book:** Big Data Science & Analytics: A Hands-On Approach by Arshdeep Bahga, Vijay Madisetti, 2016, 1st Edition, VPT. **(RB)**
  ISBN 13: 978-0996025539



Course Topics: Selected topic from the textbook's chapters 1 to 8 will be covered in the course. See the tentative course outline for details.

Topic Prerequisites: Students are expected to have background in the following:

- Knowledge of computer programming (preferably in R and Python) at a level sufficient to write a reasonably non-trivial computer program.
- Familiarity with basic statistics and linear algebra.

Workload: 10-12 hours/week

Course Grade: Course grades will be determined as follows:

Assignment	Weight
Final Exam (Comprehensive)	50 %
Homework Assignments (5) [Group mode]	30 %
Quizzes (5) [Individual mode]	20 %

Your final course grade will be determined by the standard college formula based on your course average:

	95.00-100.00 → A	90.00-94.99 → A-
87.00-89.99 → B+	83.00-86.99 → B	80.00-82.99 → B-
77.00-79.99 → C+,	73.00-76.99 → C	70.00-72.99 → C-
67.00-69.99 → D+,	63.00-66.99 → D	60.00-62.99 → D-
00.00-59.99 → F		

Late Submission and Make-up Policies

- **Homework assignments** are to be completed and turned in *by the due date*. **For each late day, 30% of the total possible points will be deducted** (a day ends at the due time). No work will be accepted more than 3 days late.
- **Quizzes** are to be completed and turned in *by the due date*. No labs will be accepted after the due date.
- **Exam:** A make-up exam will *only* be given in cases of documented emergencies. It is your responsibility to contact me with documentation of your emergency.
- All missed grades will be recorded as zeros.

CLASS POLICIES

Class Attendance Policies: Regular class attendance of all class meetings is expected of every student enrolled in this class. "Your failure to attend class, or make contact with your instructor to adequately explain your absence by the 10th class day of the semester will result in your being administratively dropped from this course. Being dropped from this course may affect your enrollment status and/or your financial aid eligibility."

Student Conduct In Class Policy

Any acts of classroom disruption that go beyond the normal rights of students to question and discuss with instructors the educational process relative to subject content will not be tolerated, in accordance with the Academic Code of Conduct described in the Student Handbook.

Electronic Devices In Class Policy

Cellular phones, pagers, CD players, radios, and similar devices are prohibited in the classroom and laboratory facilities. Calculators and computers are prohibited during examinations and quizzes, unless specified. Reasonable laptop-size computers may be used in lecture for the purpose of taking notes.

Academic Dishonesty: You are encouraged to generally discuss assignments with fellow students, but may not copy their solution or code. **Doing so constitutes academic dishonesty which will be sanctioned with a grade of F in the course.** See https://www.uh.edu/provost/policies/honesty/ for more information on UH's policy on academic dishonesty.

Campus Carry Law

Beginning August 1, 2016, the new campus carry law that was signed by Governor Abbott on June 13, 2015 allows persons with a state mandated concealed handgun license (CHL) to carry a concealed handgun in certain areas on campus so long as the area has not been designated by the University as an exclusion zone. The University's campus carry policy can be found here:
http://www.uh.edu/af/universityservices/policies/mapp/07/070105.pdf .

Tentative Course Outline

Week	Date	Topic	Quiz	Homework
1	6/07 Fri	**Introduction to Big Data** • TB: Chapter 1 • RB: Chapter 1	Quiz-0	HW-0
2	6/14 Fri	**R for Big Data** • TB: Chapter 3	Quiz-1	HW-1
3	6/21 Fri	**Hadoop and MapReduce Framework** • TB: Chapter 4 • RB: Chapters 2 and 3	Quiz-2	HW-2
4	6/28 Fri	**Relational Databases** • TB: Chapter 5	Quiz-3	HW-3
5	7/05 Fri	**Non-Relational Databases** • TB: Chapter 6	Quiz-4	HW-4
6	7/12 Fri	**Spark Ecosystem** • TB: Chapters 7 and 8	Quiz-5	HW-5
7	7/19 Fri	**Final Exam (1 hr. 30 min)**		