Georgia State University

J. Mack Robinson
COLLEGE of BUSINESS

MSA 8040
Data Management for Analytics

Syllabus

Instructor: Dr. Houping Xiao
Email: hxiao@gsu.edu
Office: Buckhead 329
Class Meetings: Fall 2019
Buckhead 306
Thursday, 2:00pm – 4:30pm **OR** 7:15pm – 9:45pm (please check your CRN number)

Course Description:

The course is intended to introduce concepts in both Database and Data Mining areas for unstructured data analytics. Topics related to Database include 1) Database and database design concepts, including relational database model, entity relationship modeling and Normalization; 2) Structured Query Language, focusing on MySQL; 3) NO-SQL, focusing on MongoDB. For unstructured data mining, we will cover the following topics: 1) unstructured data generation using Web Scraping in python and data store and query via both SQL and No-SQL; 2) Sentiment Analysis; and 3) Topic Modeling, focusing on LDA. More details of each topic, please refer to the tentative course schedule.

Course Objectives:

By the end of the semester, students will be able to:
- Explain in own words concepts in both SQL and NO-SQL
- Differentiate SQL and NO-SQL
- Build and implement query in both SQL and NO-SQL databases
- Differentiate structured and unstructured data
- Generate unstructured data using Web Scraping
- Apply unstructured data analytics techniques to solve real problems

Textbooks:
- [1] Carlos Coronel and Steven Morris. Database systems: Design, Implementation, & Management. 13th Edition. Cengage Learning. ISBN-13: 978-1337627900

Recommended References:
- https://docs.mongodb.com/manual/
- https://dev.mysql.com/doc/refman/8.0/en/tutorial.html
- http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf
- [2] Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques. 3rd Edition. Elsevier. ISBN-10: 0123814790. ISBN-13: 9780123814791

IN-CLASS Course Structure

The course is designed as a combination of in class lecture hands-on work. Please refer to the tentative course schedule below for the lecture time. The goal of this pedagogical approach is to introduce the theoretical concepts and reinforce practical implementation in MySQL, MongoDB and Python such that students can indeed master skills on their own after class. *Effort outside of reviewing and practicing the class materials are highly recommended.*

OUT-OF-CLASS Assistance

Office Hours: Thursday 4:30pm—6:00pm

Location: Buckhead 329

If you have questions about course concepts, please come to the Office Hours. If you could not come to the office hour, please email me to set up a meeting. To maximize the effectiveness of the meeting to help you better understand content you are struggling with, **in your email please include the specific questions you have and would like to discuss about the course content.**

Coursework and Grading

Coursework
- Class participation
- Mid-term Exam
- Projects

Grade Calculation
- 15% Class Participation

  The attendance of this class is required and regularly, please sign the attendance sheet when entering the class.
- 25% In-Class Mid-term Exam

  There will be one in-class mid-term exam. The mid-term exam will cover the material discussed in the lectures and exercises before the mid-term. Exam missed due to an unexcused absence may not be made up (please refer to the Make-up policy in page 4).

  *You are not allowed to collaborate on the exam. Both you and your collaborator will receive a score of **ZERO** if any infraction is noticed and established. In addition, other actions may also be taken.*
- 60% Projects (TWO 15% individual projects and ONE 30% final group project)
  1. Individual projects (30%)

     We will have **2** individual projects, including two individual projects (15% for each)

     Code of your solutions should be submitted through iCollege **on the day that they are due**!
  2. 30% Project (20% Project Chapter + 40% Report + 40% Presentation)

     A group-based final project (**at most 3 people per group**) will be released on Week 10 or 11 to evaluate the students' understanding on the topics covered in the course. Each project includes a project chapter, a final report of their findings and a presentation to present their work. The grade will be based on the evaluation on their reports and demos from the instructor.

     Please check Guidelines on each project carefully!
     1. **Late submission policy:** deliverables submitted after due date will only lose 10% of your total scores for each day after the due. No work is accepted after one week after due day.
     2. **Peer-Evaluation (this is for the final group project)**:

        To better achieve fairness in the class, at the end of the course you will be asked to evaluate yourself and the other members of your group on completing the project. These ratings are used for gauging team members' contributions. The grade you and your group members receive will depend in part on these peer evaluations. Rate each member based on the following criteria: (1) participation in group activities, (2) quality of work, (3) quantity of work, (4) finishing assigned work on time, and (5) ability to work as a team member. Please use the following scale to assign scores:

5	Exceptional effort, above and beyond the call of duty
4	Above average effort
3	Normal effort (this is the expected score!)
2	Below average effort
1	Unacceptable effort

Then, submit the following note to the instructor:

Your Name: _____ Score: _____

Team Member #2: _____ Score: _____

Team Member #3: _____ Score: _____

Note: Please include a brief reason for any group member scoring either a "1" or a "5." I expect everyone to be thoughtful and diligent in completing this evaluation. ***You may get ZERO for the project if you receive "1"s from all other group members.***

Grading

Class participation	15%
Mid-term Exam	25%
Individual Project 1	15%
Individual Project 2	15%
Final Group Project	30%

Grading Policy

The following grading scale will be used to translate number grades to letter grades:

96.5 – 100	A+
92.5 – 96.4	A
89.5 – 92.4	A-
85.5 – 89.4	B+
82.5 – 85.4	B
79.5 – 82.4	B-
69.5 – 79.4	C+
59.5 – 69.5	C
55.0 – 59.4	C-
50.0 – 54.9	D
0 – 49.9	F

Responsibility for Learning and The Honor Code

Being responsible for your own learning does not mean that you must always work in isolation. However, when working in groups we encourage you to be mindful of how much effort and Learning you are experiencing. Below, we outline our expectations for work in this course.

For projects, I encourage students to work together to solve and understand the problems. Nevertheless, each student is responsible for demonstrating he or she has good grasp of the material. Ultimately, each student's project solution should reflect his or her own learning and be written in the students' own words. While students may work together to figure out how to solve the problems, each student must run his or her own analyses and turn in their own output. For the Exam, each should work independently, no discussion is allowed. Under no circumstance should a student email his or her project solutions, project reports, and codes to a classmate. Working together (for the project) is for the purpose of collaborating, not copying.

"As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity." As listed on https://deanofstudents.gsu.edu/student-conductpolicy-on-academic-honesty/.

Make-up Exam Policy

There is one mid-term exam in this course. Date for the exam is already set on the Tentative Course Schedule below. If there is an excusable reason for being unable to be present during the exam dates, please let me know as soon as possible to schedule a make-up exam. The make-up exam if at all possible will take place before the scheduled exam date. Students with unexcused absences for an exam will earn a 0 on the exam.

Special Needs

Students who wish to request accommodation for a disability may do so by registering with the Office of Disability Services. Students may only be accommodated upon issuance by the Office of Disability Services of a signed Accommodation Plan and are responsible for providing a copy of that plan to instructors of all classes in which accommodations are sought.

Tentative Course Schedule (Topics)

The course syllabus provides a general plan for the course; deviations may be necessary.

Week	Date	Topic	Corresponding Readings	Project
1	08/29/2019	Introduction to Database: Database and design concepts, Relational database model	Chapter 1, 2, & 9	
2	09/05/2019	SQL: ER Model, ER diagrams, normalization, and SQL	Chapter 3, 4, & 6	
3	09/12/2019	MySQL: Introduction to MySQL Server and Workbench, Data types	Chapter 7, 8, & Online materials	
4	09/19/2019	MySQL: Functions, Operators and Statement Syntax	Chapter 7, 8, & Online materials	**P1 Out**
5	09/26/2019	MySQL: SQL	Chapter 7, 8, & Online materials	
6	10/03/2019	MySQL: Advanced SQL	Chapter 7, 8, & Online materials	
7	10/10/2019	MySQL: Procedure & Trigger	Online materials	
8	10/17/2019	Mid-term Exam		**P1 Due P2 Out**
9	10/24/2019	Feedback on Mid-term Exam	Online materials	
10	10/31/2019	Introduction to Unstructured data analysis And MongoDB	Online materials	
11	11/07/2019	MongoDB: CRUD & Aggregation	Online materials	**P2 Due P3 Out**
12	11/14/2019	Web Scraping	references	
13	11/21/2019	Sentiment analysis and Topic Modeling	references	
14	11/28/2019	Thanksgiving Break		
15	12/05/2019	**Final Presentation and Report Due (P3 Due)**		

Your constructive assessment of this course plays an indispensable role in shaping education at Georgia State. Upon completing the course, please take the time to fill out the online course evaluation