Department Computational and Data Sciences

College of Science

4400 University Drive, MS 6A12, Fairfax, Virginia 22030

Phone: 703-993-5836; Fax: 703-993-9300; https://cos.gmu.edu/cds/

CSI 695
Scientific Databases

1. **General Information**

Instructor:	Dr. Olga Gkountouna
Where:	Blackboard Collaborate (synchronous online mode)
When:	Wednesday, 4:30-7:10pm. Aug 24 - Dec 16, 2020
Course website:	Blackboard
Credits:	3
Prerequisites:	recommended CDS 302/502 or equivalent, or permission of instructor.
Office Hours:	By appointment (*please email me at ogkounto@gmu.edu*).

2. **Course Description**

The main focus of this course is the design and implementation of *Database Management Systems* (DBMS). The topics to be covered include the E-R and Relational models, the use of the SQL query language, file storage, indexing, transactions and concurrency control, query processing and optimization. Furthermore, the course will cover more advanced topics on NoSQL databases with an emphasis on MongoDB. Finally, the course includes an overview of Big Data analytics, as well as an introduction to machine learning and its applications.

The course will include not only the theoretical foundations, but also practical applications, using real datasets, both during in-class examples and homework assignments. Students will develop class projects using real data of their preference. The final deliverables of this class are: a final presentation in class, and a project report written in the format of a scientific article.

3. **Learning Outcomes**

By the end of the course each student will

- have a broad knowledge on fundamentals, theory and applications of Database Management Systems;
- be able to articulate and effectively communicate concepts and ideas related to Data Management and Databases to experts, non-experts, and other professionals in a work environment;
- have the ability to appropriately apply the knowledge acquired in the course for various hypothetical and real-world database design and management tasks;
- have experience with indexing, accessing, querying, mining and visualizing data from scientific databases;
- have experience with free and open-source programming packages of python/R, in order to apply data mining solutions to big data sets, and
- be able to properly visualize and interpret the results.

4. **Format**

The course will be taught as a combination of lectures and discussions, in a *synchronous* online *mode*. The lectures will be broadcast through **Blackboard Collaborate**. The students do a class project instead of exams. They will present their project proposal in class, as well as give a presentation of the final project in the end.

5. **Textbooks**

No required textbook. Course slides and reading material will be provided via Blackboard.
Recommended book: Silberschatz, Korth, Sudarshan, "*Database System Concepts*", McGraw-Hill.

6. **Technology Requirements**

Software. Students may choose the programming language that they prefer to write their applications code. The course will be using the SQL query language, the MongoDB query language, and Python as a high-level programming language. For practice in class you will need *SQLite*, and *Anaconda Python 3.x*. It is recommended that you practice at home with another DBMS, such as PostreSQL or MySQL. Python is recommended for data mining/machine learning applications. For the final project report, the use of *Overleaf* online LaTeX editor is strongly recommended.

A basic understanding of computer programming principles and knowledge of any programming language or pseudo-code is desirable.

Hardware. You will need access to a Windows or Linux or Macintosh computer with at least 2 GB of RAM and a fast, reliable broadband Internet connection (e.g., cable, DSL). To use the computers in a classroom, you may download SQLite on a memory stick and run your SQL code in that directory. For the amount of computer hard disk space required to take an online course, consider and allow for the space needed to: 1) install the required and recommended software and, 2) save your course assignments.

7. **Grades**

Each assignment and written exam will be given a numerical grade on a 0-100 scale. Some assignments may include bonus tasks. At the end of the term all the marks will be totaled as a <u>weighted average</u> according to the following weights:

Assignments	**20%**
Project Proposal	**10%**
Final Project	**70%**

Please note that in general all assignments may not have the same weight. The weight of each individual assignment will be indicated on the assignment form.

8. **Class Project**

Projects should include implementation of a database application or data mining/machine learning analysis on real data sets, based on the course materials and the discussions in class. The final deliverables are the source code, slides of the final class presentation, and a final project report that describes the data, the system/methods used, the findings (if any), and a discussion on the shortcomings and directions for potential future work. Projects should be submitted through the **Blackboard course website**.

9. **Assignments:**

The course will include few written assignments on selected topics from the material covered in class and in the assigned reading. All assignments are mandatory. Typically, 1 week will be allocated for every assignment.

Assignments should be done **through the Blackboard course website**.

> **Please note**: Assignments should be submitted only through the Assignment submission section of the Blackboard system - DO NOT email assignments directly to the instructor.

10. Course outline (tentative)

In this course we will cover the following topics. Please note that the topics and their order are subjected to change at the discretion of the instructor. This class will not follow a rigid format, but will instead adjust to the talent levels and interests of the students.

Date	Lect. #	Topic	Assignment due dates
08/26	1	Introduction: Why Databases? Course overview. The Relational Model, Relational Algebra	
09/02	2	Introduction to SQL DDL and DML basic queries	
09/09	3	SQL – intermediate and advanced queries SQL practice in class (interactive)	
09/16	4	Database Design: The Entity-Relationship Model, Reduction of E-R to Relational Schema	Assignment 1
09/23	5	Application Design: Complex Data Types, Application Development	
09/30	6	Big Data Analytics Intro to Data Mining and Machine Learning	
10/07	7	**Project Proposals Presentations / Discussion**	Project Proposal
10/14	8	Supervised learning: classification, regression Unsupervised learning: clustering, ARM	
10/21	9	Storage management: Physical Storage, File structures Indexing	
10/28	10	Query Processing Query Optimization	Assignment 2
11/04	11	Transactions Concurrency Control	
11/11	12	NoSQL Databases: Key-value, Graph stores, Object-Based Document-oriented Databases; Intro to MongoDB	
11/18	13	MongoDB query language practice in class Blockchain Databases	
11/25	14	**Thanksgiving Recess (No classes)**	
12/02	15	**Project Presentations**	
12/09	16	**Project Presentations**	
12/16	17	**FINAL PROJECT DEADLINE**	Final Project: Report, Slides, and Code

11. **Late paper submission:**
    The late submission policy is the following:
- Submissions within the first day (24 hours) after the deadline: -**5**% credit.
- Submissions within the second day after the deadline: -**10**% credit.
- Submissions within the third day after the deadline: -**30**% credit.
- No assignments after the third day (i.e., 72 hours after the deadline) will be accepted.

    Exceptions to this policy may be granted, given *serious circumstances* and *at the discretion of the Instructor*.

12. **General guidelines for ASSIGNMENT preparation and submission**
    a. Grades of assignments will be based on:
       - **Academic merit** of your answers.
       - **Conciseness** and **completeness** of your answers.
       - **Organization** and **presentation**. Please organize your report in a logical fashion so that your answers could be easily identified.
    b. Please remember that your assignment is a professional document, and should therefore be formatted and constructed accordingly. All assignments are to be typed. Hand-written assignments will not be accepted.
    c. Submission of your answers will be made exclusively through Blackboard. Please do **not** email your assignments.
    d. The electronic submission of your assignment report has to be in **PDF format**, unless stated otherwise.
    e. Each assignment submission should include a cover page with the following information: assignment title, assignment number, student name, and submission date.
    f. Please make sure you have a backup of all the materials you submit.

13. **Course website:**
    The course has a Blackboard website. This website will provide you a single portal through which you may obtain lecture notes, retrieve assignment data and, review links to additional materials, and receive special announcements. You are required to visit the course website once per day. Please notify ITU (and, if necessary, the instructor) if you encounter any problems accessing this website.

14. **Electronic communication:**
    All course related correspondence, should be made through GMU email accounts. Please always use only your GMU email to contact the Instructor. You may not receive a reply immediately, but typically emails will be answered within 2 business days.

15. **Student Expectations:**
    - **Academic Integrity:** Students must be responsible for their own work, and students and faculty must take on the responsibility of dealing explicitly with violations. The tenet must be a foundation of our university culture. [See http://academicintegrity.gmu.edu/distance].
    - **Honor Code:** Students must adhere to the guidelines of the George Mason University Honor Code [See http://oai.gmu.edu/the-mason-honor-code/].
    - **MasonLive/Email (GMU Email):** Students are responsible for the content of university communications sent to their George Mason University email account and are required to activate their account and check it regularly. All communication from the university, college, school, and program will be sent to students solely through their Mason email account.

[See https://masonlivelogin.gmu.edu].

- **Patriot Pass:** Once you sign up for your Patriot Pass, your passwords will be synchronized, and you will use your Patriot Pass username and password to log in to the following systems: Blackboard, University Libraries, MasonLive, myMason, Patriot Web, Virtual Computing Lab, and WEMS. [See https://password.gmu.edu/index.jsp].
- **University Policies:** Students must follow the university policies. [See http://universitypolicy.gmu.edu]. Responsible Use of Computing - Students must follow the university policy for Responsible Use of Computing. [See http://universitypolicy.gmu.edu/policies/responsible-use-of-computing].
- **University Calendar:** Details regarding the current Academic Calendar. [See http://registrar.gmu.edu/calendars/index.html].
- **Students with Disabilities:** Students with disabilities who seek accommodations in a course must be registered with the George Mason University Office of Disability Services (ODS) and inform their instructor, in writing, at the beginning of the semester [See http://ods.gmu.edu].
- Students are expected to follow courteous Internet etiquette at all times; see http://www.albion.com/netiquette/corerules.html for more information regarding these expectations.

16. **Student Services:**
    - **University Libraries:** University Libraries provides resources for distance students. [See http://library.gmu.edu/distance and http://infoguides.gmu.edu/distance_students].
    - **Writing Center:** The George Mason University Writing Center staff provides a variety of resources and services (e.g., tutoring, workshops, writing guides, handbooks) intended to support students as they work to construct and share knowledge through writing. [See http://writingcenter.gmu.edu]. You can now sign up for an Online Writing Lab (OWL) session just like you sign up for a face-to-face session in the Writing Center, which means that you set the date and time of the appointment! Learn more about the Online Writing Lab (OWL).
    - Counseling and Psychological Services: The George Mason University Counseling and Psychological Services (CAPS) staff consists of professional counseling and clinical psychologists, social workers, and counselors who offer a wide range of services (e.g., individual and group counseling, workshops and outreach programs) to enhance students' personal experience and academic performance [See http://caps.gmu.edu].
    - **Family Educational Rights and Privacy Act (FERPA):** The Family Educational Rights and Privacy Act of 1974 (FERPA), also known as the "Buckley Amendment," is a federal law that gives protection to student educational records and provides students with certain rights. [See http://registrar.gmu.edu/privacy].

17. **Notice of mandatory reporting of sexual assault, interpersonal violence, and stalking:**
    As a faculty member, I am designated as a "Responsible Employee," and must report all disclosures of sexual assault, interpersonal violence, and stalking to Mason's Title IX Coordinator per University Policy 1202. If you wish to speak with someone confidentially, please contact one of Mason's confidential resources, such as Student Support and Advocacy Center (SSAC) at 703-380-1434 or Counseling and Psychological Services (CAPS) at 703-993-2380. You may also seek assistance from Mason's Title IX Coordinator by calling 703-993-8730, or emailing titleix@gmu.edu.

Disclaimer: Any typographical errors in this Course Outline are subject to change and will be announced in class. The date of the final examination is set by the Registrar and takes precedence over the final examination date reported by the instructor.

Note: Recording is permitted *only with the prior written consent of the professor* or if recording is part of an approved accommodation plan.