# CDS 302 Syllabus
## Scientific Data and Databases
Fall 2021 - Section: 002

**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/

## CDS 302
### Scientific Data and Databases

## General Information

- **Classroom:** ENGR 1110, Online meetings (if needed) will use Blackboard Collaborate Ultra
- **Meeting Time:** Monday, 4:30pm - 7:10pm
- **Course Content:** Lecture Slides, reading material, example code, data files, etc. will be made available weekly on Blackboard. I may post a video of a lecture on a few weeks (Note: I may miss the September 13th class due to a schedule conflict. I will probably cover this week using a video lecture.)
- **Course Website:** Blackboard - students should check the website frequently for announcements.
- **Credit Hours:** 3
- **Specialized Designation:** Writing Intensive in Major
- **Prerequisites:** CDS-101 or CDS-130 or equivalent, or permission of the instructor

## Instructor

- **Instructor:** Joe Boone (Graduate Lecturer)
- **Email:** jboone@gmu.edu
- **Office Hours:** Online (Blackboard Collaborate), please arrange directly with me via e-mail

## Graduate Teaching Assistant

- TA: Mailun (Alan) Zhang
- Email:mzhang23@gmu.edu
- **Office Hours:** To be announced

## Course Description

The main focus of this course is the application of various Database Management Systems (DBMS) and related tools in various contexts. The topics to be covered include using SQL for queries and data manipulation in a relational DBMS, structured data modeling and design, DBMS topics such as indexing, hashing, file storage, query processing. The course will include not only theoretical foundations, but also practical applications, using real data sets. There will be in-class examples and homework assignments for storing, managing and querying databases. Finally, the course will include discussions on more advanced topics, beyond traditional Relational DBMS including scientific data formats and semi-structured and unstructured data. Note that this is a writing intensive course, we will visit the basics of scientific writing, including how to write articles using LaTeX as the preferred typesetting system.

## 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 others in a professional environment
- Have the ability to appropriately apply the knowledge acquired in the course for hypothetical and real-world database design and management tasks
- Have experience with indexing, querying, mining and visualizing data from scientific databases using SQL
- Have the necessary writing skills for creating scientific manuscripts, articles and reports, using the appropriate math notations, tables, references, citations, etc.
- Have the ability to use a high level language (we will use Python) for manipulation of data in a Database Management System.

## Format

This section of the course is an in-person section. We will meet each week at the indicated time and location. There will be a Midterm and a Final exam, as well as several assignments - some of which will involved coding assignments. Classes will typically include part lecture and part instructor led demonstration. Students will get the most out of the class if they bring a laptop and follow along with the demonstrations in class. There will frequently be student exercises during class involving the tools demonstrated.

## Textbook

There is no required textbook. All reading material will be provided via Blackboard. A recommended text is Silberschatz, Korth, Sudarshan, "Database System Concepts", McGraw-Hill. There is generally a reading assignment each week.

## Technology Requirements

**Software:** This course will be using LaTeX as the main text editor for writing assignments. The use of Overleaf online LaTeX editor is strongly recommended. Alternatively, you may use another LaTeX editor of your choosing. The course also will also be using SQL as a query language, and Python as a programming language. For assignments you will need SQLite, DBbrowser for Sqlite, and Anaconda Python 3.x (or other Python distribution). This semester we will also take a hands-on look at NoSQL databases, the specific tools to be announced later. A basic understanding of computer programming principles and knowledge of any programming language or pseudo-code is assumed.

**Hardware:** You will need access to a Windows, Linux or Macintosh computer with at least 2 GB of RAM and to a fast, reliable broadband Internet connection (e.g., cable, DSL). 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.

## 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 weighted average according to the following weights:

- Assignments: 50%
- Midterm Exam: 25%
- Final Exam: 25%

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

The course grading scheme will be as follows: A+ (97+%), A (93%), A- (90%), B+ (87%), B (83%), B- (80%), C+ (77%), C (73%), C- (70%), D+ (67%), D (60%).

## Exams

The course includes mandatory midterm and final exams. The exams will be taken in person in class. If for any reason an exam must be taken online, then we will use the RESPONDUS Lockdown Browser. Please install it. The material covered in the exams will be announced in class. A student who cannot complete a course examination or complete a course assignment because of an illness or other compelling reason can apply for extension of time to complete the assignment. If RESPONDUS is required, see these links:

**Watch** this video to get a basic understanding of LockDown Browser: https://www.respondus.com/products/lockdown-browser/student-movie.shtml

**Download** and install the browser from: https://download.respondus.com/lockdown/download.php?id=133435885

**Note:** The date of the final examination is set by the Registrar.

# Assignments

The course will include several written assignments on selected topics from the material covered in class and in the assigned reading. All assignments are mandatory. Typically, one 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. **<u>DO NOT</u> email assignments directly to the instructor or TA.**

# Course Outline

We will cover the following topics in the following order (subject to instructor revision):

| Date | Week | Topic | Assignment |
|------|------|-------|------------|
| 08/23 | 1 | Introduction: Why Databases? Course Overview. <br> Scientific Writing: Introduction to LaTeX and OverLeaf | |
| 08/30 | 2 | LaTeX: Math Notations, Equations, Tables, References <br> The Relational Model: schemas, keys, diagrams and algebra | A1: Writing |
| 09/06 | - | No Class - Labor Day | |
| 09/13 | 3 | SQL Part 1: Introduction to basic queries <br> Basic SQL DDL and SQL querying practice (SQLite) | A2: SQL queries |
| 09/20 | 4 | SQL Part 2: String operators, set operations, <br> aggregate functions, ordering, grouping, having-clause | |
| 09/27 | 5 | SQL Part 3: nested queries, set membership, joins, self-joins, <br> with-clause | A3: Advanced Queries |
| 10/04 | 6 | SQL Part 4: More SQL practice <br> Exam Review/Preparation | |
| 10/12 | 7 | MIDTERM EXAM <br> (Note: We meet on Tuesday this week) | |
| 10/18 | 8 | Data Modeling Part 1: The ER Model | A4: E-R Model |
| 10/25 | 9 | Data Modeling Part 2: Reduction of E-R to Relational Schema | |
| 11/01 | 10 | RDBMS Internals Part 1: Relational Storage and File Structure <br> Indexing and Hashing: Basics, Ordered Indices | |
| 11/08 | 11 | RDBMS Internals Part 2: Indexing: Dense vs. Sparse Index <br> Multilevel Index, B*-Tree Structure; B+-Tree Operations | A5: E-R to Relational Model |
| 11/15 | 12 | RDBMS Internals Part 3: Indexing: B-Tree; Bitmap Index; <br> Static Hashing, Dynamic Hashing; Hash file org., Query Processing | |
| 11/22 | 13 | Beyond Structured Data Part 1: Scientific Data and Formats <br> Semi-Strctured Data, NoSQL Databases | A6: Indexing |
| 11/29 | 14 | Beyond Structured Data Part 2: Key Value Databases <br> Document Databases, Graph Databases | |
| 12/06 | 15 | Reading Day - No Class | |
| 12/13 | 16 | FINAL EXAM | |

## Late Assignments

The late assignment policy is as follows:

- 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.

## General Guidelines for Assignment Preparation and Submission

- 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.
- 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.
- Submission of your answers will be made exclusively through Blackboard. Please do not email your assignments.
- The electronic submission of your assignment report has to be in **PDF format**, unless stated otherwise.
- Each assignment submission should include a cover page with the following information: assignment title, assignment number, student name, and submission date.
- Please make sure you have a backup of all the materials you submit.

## Course Website

The course uses Blackboard as its 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.

## Electronic Communication:

All course related correspondence should be made through GMU email accounts. Please use only your GMU email to contact the Instructor. You may not receive a reply immediately but I will do my best to reply within 24 hours.

## 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].
- **Internet Etiquette:** 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.

## 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].

### 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

The instructor reserves the right to modify this syllabus at any time during the course to improve the learning experience and classroom environment. The pacing of the course and the list of covered topics may also be altered in response to student progress. Any change to the syllabus will be announced via the course website.

The course objectives reflect what a student is expected to understand by the end of the course after putting in the necessary time and effort both inside and outside the classroom and completing all assignments. These outcomes are not guaranteed. Students will get more out of the course the more they put into it. Any acquired skills and knowledge will fade over time if not reviewed or practiced after the course concludes.

# A/V Recording

There is to be no video or audio recording during online sessions, with the group or individually, unless approved by the instructor.

---