CS 504 Principles of Data Management and Mining — Spring 2021

Online. See https://cs.gmu.edu/~hrolenok/teaching/cs-504-spr2021/index.html for current syllabus. Be sure to check your @gmu.edu email and blackboard regularly for updates.

Instructor: Brian Hrolenok
@gmu.edu email: hrolenok
Office Hours: TBD, ONLINE. See Blackboard for more information.

TA: Archange Destine
@gmu.edu email: adestine
Office Hours: TBD, ONLINE. See Blackboard for more information.

Note about online instruction

Due to the size of the class, and the limitations that places on the ability to maintain safe distance in the classroom, this course is being offered in an **online**, **asynchronous** format this semester. There are no scheduled lecture times, and all the material will be made available via Blackboard. Office hours will be held using videoconferencing software (tentatively Blackboard collaborate), and the instructor and TA will be monitoring email and Piazza for questions that arise outside of scheduled office hours. Responses to emails are usually within the next business day, but may be delayed due to unforseen circumstances, so **do not wait until the last minute**. Since there are no scheduled class meetings, it is crucial that you frequently monitor your school email and Blackboard to make sure you do not miss any announcements, due dates, policies, or other information.

Course description

CS 504 is a 3-credit introductory course that combines elements from a diverse field of topics, but with a slightly different focus than a course on databases or machine learning. In this course, the central object of study will be **data**: how to store and query efficiently, how to mine for hidden structure, and particularly the properties which make these tasks more or less difficult. Topics to include intro to database management, ER and relational data models, SQL and NoSQL, classification, and clustering. This course will include several individual programming and report based assignments, several quizzes, and a final exam. All assessments will be turned in online.

IMPORTANT NOTE: This course cannot be taken for credit by students of the MS CS, MS IS, MS ISA, MS SWE, CS PhD or IT PhD programs. Please check with the registrar if you are unsure about whether this course will count towards your degree program.

Learning objectives:

- To develop problem solving skills and analytical thinking from a data science perspective.
- To introduce a broad survey of modern approaches to data management and analysis.
- To develop the design and programming skills that will help you to interface with data at scale.

Prerequisites. The only prerequisite for this course is graduate standing, but please note that certain majors cannot take this course for degree credit (see catalog).

Textbook: There is no required textbook for this class. You may find the following texts useful as a supplemental references:

- Silberschatz, Korth, and Sudarshan. Database System Concepts. McGraw-Hill., 2019. ISBN: 9780078022159
- Provost, Foster, and Tom Fawcett. Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc., 2013. ISBN:9781449361327

Homeworks

All assignment submissions will be handled through blackboard, and are due by the date and time listed there. Unless otherwise explicitly stated in the assignment description, all homeworks are **individual effort**. Submissions by email will not be accepted.

Late Policy

You have three free late days to be used at your discretion thoughout the semester. That means you might turn in one assignment two days late or two different assignments one day late, etc. A free late day is "used" one minute after an assignment due date. A second free late day is "used" 24 hours and one minute after the due date. A third free late day is used 48 hours and one minute after the due date. After the free late days are exhausted, you will receive a 20% penalty per day. Late days can only be applied to homeworks, they cannot be applied to quizzes or exams.

Quizzes & Exams

Throughout the semester, there will be several *participation quizzes* given via blackboard. These will be short, multiple-choice, and you will receive <u>full credit</u> as long as you attempt the quiz by the given due date. Feedback (which will be useful when studying for the final exam) can only be made available if you attempt the quiz, and will remain available until the final exam period begins.

The final exam will be given during the final exam period (May 3-May 7), and you will be able to pick a time to begin within a window. The length of the exam will depend on the number of questions, but will not exceed 2 hours and 45 minutes. There will be no make up for this exam unless previously arranged *(well in advance)*.

Grading policies

Your TA and I will strive to provide you reasonably detailed and timely feedback on every assessment. If you have any questions about any of your grades please reach out to us, either by coming to scheduled office hours or via your "@gmu.edu" email address. If there is an error with your grade, please contact us within a week of when feedback is returned, otherwise we might not be able to change it.

Point breakdown:

- Homeworks: 75% total (split evenly between 4-5 homeworks)
- Participation quizzes: 10%
- Final exam: 15%

This course is graded according to the standard graduate grading policies, with the following cutoffs:

- A: ≥ 90
- A-: ≥ 88, < 90
- B+: ≥ 85, < 88
- B: ≥ 80, < 85
- B-: ≥ 78, < 80
- C: ≥ 70, < 78
- F: < 70

Academic Integrity

Please familiarize yourself with both the University wide [honor code](), as well as the one specific to the department of [Computer Science](). Violations of academic integrity will be reported to the Honor Committee. Course failure is a common recommended outcome for students found in violation.

Accommodations

If you need academic accommodations, please make sure you contact the instructor at the beginning of the semester or as soon as possible. Also make sure to contact GMU's Disability Services, available online ([https://ds.gmu.edu](), [ods@gmu.edu]()) and by phone (703.993.2474), which coordinates all academic accommodations. After you have contacted ODS, you still need to contact the instructor so that appropriate arrangements can be made.

Topic outline

- Introduction to database management
- ER & EER model
- Relational data model and ER & EER to relational mapping
- SQL
- NoSQL
- Introduction to data mining concepts
- Classification
- Clustering