

COMP 7/8116: Advanced Database Systems

Xiaofei Zhang

Spring, 2024

E-mail: xiaofei.zhang@memphis.edu
Office Hours: Upon Request
TA: Manish Kasu (mjkasu@memphis.edu)

Web: Canvas
Class Hours: 13:00–14:25 TR
Class Room: Dunn Hall 124

Course Description & Objectives

This project-based seminar-like course will be covering a number of advanced topics in modern database systems. The course will be organized as a mix of lectures and paper presentations on the recent development of distributed database systems. The purpose of this course is to prepare graduate students with the philosophy, methods, and thinking of modern database systems for their career in data science or data management research.

Required Materials

- Course notes and the reading list will be available on Canvas.

Prerequisites

Students are expected to understand the fundamentals of databases (e.g., COMP3115), algorithms (e.g., COMP4030), and operating systems (e.g., COMP4270) each at least at the level of an introductory course.

Workload & Mark Breakdown

Project (40%)

The students are encouraged to come up with their own projects under the three topic cohorts covered in class: 1) distributed database; 2) big data processing framework; 3) graph computing. The instructor may suggest some projects as well. Projects are intended to be mini-research projects on a topic of the course. Example formats include (but are not limited to):

- Build a new application on top of one of the systems.

- Build a new feature/extension to an existing system.
- Implement a new algorithm that replaces a core part of a system (e.g., the scheduling algorithm of a system).

The most important thing is to pick a project on a topic you really like! You will meet with the instructor several times during the term to discuss your progress. You have about 3 weeks to decide on the project.

Project Deliverables: project proposal, project presentation (15%), project report (15%), and source code (10%)

Group Work Policy: Given the comprehensive nature of this course and valuable collaborative skills it fosters, a group project is adopted as the primary assessment approach. Recognizing the importance of equitable contribution and to ensure that all group members actively participate and contribute, a stringent evaluation framework is implemented. Each group member will be required to submit a statement outlining each group member's specific role and contributions to the project. These statements will serve as a means to assess the dedication and involvement of each participant. The instructor reserves the right to review these statements and allocate grades to individual members based on their demonstrated efforts and contributions.

Paper Presentation (20%)

Each student will present **two** papers from the given reading list (which will be available on Canvas). Please consult the instructor first if you plan to present a paper out of the list.

- Please use slides for the presentation. The content in your slides should be your own but you can use others' materials, e.g., figures from the paper we are reading, when necessary, and by crediting your source on your slide.
- Your presentation should answer the following questions:
 - What is the problem?
 - Why is it important?
 - Why is it hard? Why don't naive/previous methods work?
 - What is the solution to the problem the authors propose?
 - What interesting research questions does the paper raise?
 - How does the paper relate to the existing work?
 - In your opinion, what is the major contribution and impact of the work?
- If you are presenting a system, it would be good to show a little demo of how to configure and run that system.

Attendance (10%)

Attendance is mandatory for this course. Students are encouraged to actively participate in the Q&A session in class.

Assignment (30%)

There will be 3 assignments (10% each) with a mixture of written questions and coding tasks. Students can work with any programming language that they feel comfortable with. Note that it is the student's responsibility to make sure the submission is valid and readable.

Course Schedule (Tentative)

Week	Topic
1	Overview & Introduction
2, 3	Distributed query & transaction processing
4	P2P & Multidatabases
5, 6	1st round paper presentation
7	HDFS & MapReduce
8	<i>Spring break</i>
9	Spark & Streaming
10, 11	2nd round paper presentation
12	Graph processing and NoSQL systems
13, 14	Project presentation

Assessments & Grading

7000-level v.s. 8000-level

Students enrolled in 7000 and 8000 sessions will have the same amount of workload, but 8000-level students are expected to work on more challenging problems for the course project.

Grading Scale

We will calculate final letter grades in two different ways; then each student will receive the higher of the two letter grades. One way is a fixed grading scale, with the following cutoffs:

$$A \geq 92 \quad B \geq 80 \quad C \geq 68 \quad D \geq 56 \quad F < 56$$

The other way is a curve, with the following percentages of students receiving each grade:

$$A : 18\% \quad B : 28\% \quad C : 28\% \quad D : 20\% \quad F : 6\%$$

We will feel free to give an F to any student who clearly did not put effort into the course (or an A+ to any student with truly exceptional performance).

Course Policies

Plagiarism or cheating behavior in any form is unethical and detrimental to proper education and will not be tolerated. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. Plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students

are allowed and encouraged to discuss with each other and look up resources in the literature, but appropriate references must be included for the materials consulted, and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor's discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the Office of Student Accountability for further disciplinary action. For further information on the UofM code of student conduct and academic discipline procedures, please refer to <http://www.memphis.edu/studentconduct/misconduct.htm>

Within this class, you are welcome to use foundation models (ChatGPT, GPT, DALL-E, Stable Diffusion, Midjourney, GitHub Copilot, and anything after) in a unrestricted fashion. However, you should note that all large language models still have a tendency to make up incorrect facts and fake citations, code generation models have a tendency to produce inaccurate outputs, and image generation models can occasionally come up with highly offensive products. You will be responsible for any inaccurate, biased, offensive, or otherwise unethical content you submit regardless of whether it originally comes from you or a foundation model. If you use a foundation model, its contribution must be acknowledged in the hand-in; you will be penalized for using a foundation model without acknowledgment. Having said all these disclaimers, the use of foundation models is encouraged, as it may make it possible for you to submit assignments with higher quality, in less time.