Fall 2024

CS 631-003: Data Management System Design

Shantanu Sharma

CS 631: Database Management System Design

Fall 2024

Credit: 3

Instructor: Dr. Shantanu Sharma

Location: **CKB 204**

Class sessions: Thursday 2:30 – 5:20 PM

Instructor Office: GITC 4201B

Instructor email: shantanu.sharma@njit.edu

Office Hours: Monday (10:20 AM – 11:15 AM) or by appointment over email. No office hours on 9/9.

Office Hours Location: GITC 4201B

Prerequisite:

Knowledge of C and data structures. Acquaintance with fundamental notions of relational database technology. Mathematical properties and usage of database programming languages. Methods of database design and conceptual modeling. Methods of physical storage for database information. Fundamental notions of concurrency control and recovery in database systems.

Course Overview:

This course will give a broad overview of database management systems. Fundamental concepts of databases will be explained. Topics include database system architecture, data modeling, storage of databases, query languages, indexing, concurrency control, transaction management, and big data processing systems (Hadoop and Spark), etc.

The outcome of the course:

Students will:

1. Gain knowledge of the requirements and applications of the database management system.
2. Understand the fundamental components and operations supported by a database management system.
3. Gain experience with a DBMS.

Primary Textbook:

1. Lecture slides will cover major topics.
2. Database management systems (3rd edition) by Raghu Ramakrishnan and Johannes Gehrke.
3. Database System Concepts (7th edition) by Avi Silberschatz, Henry F. Korth, and S. Sudarshan.

Lecture slide availability: *Slides will be online before the lecture. If there is any change in the slides during or after the lecture, they will be updated after the lecture.*

Other reference books: Students may refer to the following book to gain a deeper knowledge of databases. However, reading all such books is not mandatory.

1. Fundamentals of Database Systems (7th edition) by Ramez Elmasri and Shamkant Navathe
2. Database Systems: The Complete Book by Héctor García-Molina, Jeffrey Ullman, and Jennifer Widom
3. Architecture of a Database System by Joseph M. Hellerstein, Michael Stonebraker, and James Hamilton
4. Principles of Database and Knowledge-Base Systems, Volume I, by Jeffrey D. Ullman.

Grading Policies for Master's Students:

Quiz	10
Assignments	30
Mid-term exam	24
End-term final exam	30
Attendance	6
Total	**100**
Bonus point (see details below)	***10***

I reserve the right to make small adjustments to grade weights or to add/remove assignments, project, or quizzes as the need arises.

Grade Letter for Master's Students:

Grade letters will be allocated as follows:

Grade letter	**Points**
A	80 and above
B+	70 – 79
B	60 – 69
C+	50 – 59
C	40 – 49
F	Below 40

Quizzes: (10 points) for **both** Master's and PhD Students

1. Each quiz will contain ***10 to 20 questions*** (multiple-choice, fill-in-the-blank, or small questions).
2. Each quiz will be in the classroom.
3. ***15-20 minutes*** will be given to return the sheet.
4. Each student must submit their own quiz. No groups will be allowed.
5. The quizzes will be on the following dates:
   a) 09/19
   b) 10/10
   c) 10/31
   d) 11/21
   e) 12/05

6. **Grading of quizzes:** An average of the best of the four quizzes will be taken in the final grade.
   For example, if one student receives the following points in the above quizzes: 10, 9, 8, 10, 10, respectively, then only the following points will be considered 10, 9, 10, 10 (points of the 3rd quiz will not be counted). To include the quiz points in the final grade, we do the following: (10+9+10+10)/4
7. The grade of each quiz will be published before the following quiz.

Assignments: (30 points) for Master's students

- Three assignments will be given.
- Each student must submit their own assignments. No groups will be allowed.
- Each assignment of 10 points.
- The assignment will be published on the following date and will be due as mentioned in the following table

Assignment number	Published on web	Due date and time
1	10/01	10/15 11:59PM
2	11/01	11/15 11:59PM
3	11/15	12/05 11:59PM

- All assignments must be submitted on Canvas on/before the due date.
- *Assignments should* ***NOT be handwritten*****.** Students may use Microsoft Word, Google Docs, or Overleaf.
- The grade for each assignment will be published before publishing the next assignment, except for the third assignment.
- ***Late submission*** of the assignment will be allowed with the given reasons. Each day will cause a deduction of 1 point. However, assignments will not be accepted after the 20th of each month and result in **zero points**. Without reason, late submissions are ***not*** allowed.

Mid-term exam (24 points) and End-term (30 points): for Master's students

- Students can carry **books, printed notes, and printed slides**. However, Internet access is not allowed.
- Mid-term date: 10/31 or 11/07 (depending on the progress of the course)
- End-term date: will be announced.

Attendance (6 points): for Master's students

1. There will be 3-4 guest lectures. The tentative date and time are as follows: 9/30, 10/16, 10/23, 11/20 – 2:30 to 3:30 PM in GITC.
2. Each lecture will carry 1.5 points.
3. Attending and participating in each lecture will result in some points.
4. *If a guest lecture will not happen, then points will be added to the final exam.*

Bonus points:

1. There is a provision for earning ten bonus points.
2. However, if a student obtains 100 points in the final grade without a bonus, then their final grade will be 100. Students who earn a final grade between 80 and 100 will secure an A, and the bonus points will not impact them.
3. To earn the 10 bonus points, a student needs to meet the instructor before 03/20.

Academic Integrity

Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at: http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf.

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. ***Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university.*** *If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu*

Note for Students with Disabilities

We will follow the university rules in this regard.

Weekly Schedule of the Lectures

(Tentative -- *the schedule may change according to the progress in the lectures*)

Week number	Topic to be covered
1	Introduction
2	Relational Model and Algebra
3	
4	SQL
5	
6	
7	Data storage
8	Review class – tentative
9	Indexes
10	
11	Query execution and optimization
12	
13	Database design
14	Big-data, MapReduce, Hadoop, Spark
15	Discussion and others (ACID vs BASE)

Grading Policies for PhD Students:

Quiz (Biweekly in the class)	10
Assignments	20
Mid-term exam	15
End-term final exam	25
Research paper presentation	10
Research-based project	20
Total	**100**

I reserve the right to make minor adjustments to grade weights or to add/remove assignments, project, or quizzes as the need arises.

Grade Letter for PhD Students:

Grade letters will be allocated as follows:

Grade letter	**Points**
A	85 and above
B+	75 – 84
B	60 – 74
C+	50 – 59
C	40 – 49
F	Below 40

Assignments: (20 points) for PhD students

- Two assignments will be given.
- Each student must submit their own assignments. No groups will be allowed.
- Each assignment of 10 points.
- The assignment will be published on the following date and will be due as mentioned in the following table

Assignment number	Published on web	Due date and time
1	11/01	11/15 11:59PM
2	11/15	12/05 11:59PM

- All assignments must be submitted on Canvas on the due date.
- *Assignments should **NOT be handwritten**.* **PhD Students must use Overleaf**.
- ***Late submission*** of the assignment will ***not*** be allowed.

Mid-term exam (15 points) and End-term (25 points): for PhD students

- Students can carry **books, printed notes, and printed slides**. However, Internet access is not allowed.
- End-term date: will be announced.

Research paper discussion and presentation (10 points)

1. A list of papers on advanced database systems will be published on Canvas in September.
2. All students need to use **an identical template** for slides that will **be given by the instructor**.
3. Presentation time will be at least 20 mins and at most 30 mins and Q/A time will be 10 mins.
4. Students may use the slide content from the published paper. No point will be deduced if a student does not prepare their content in slides**.**
5. **Slides must be emailed and discussed with the instructor at least 7 days before the presentation.**
6. Additional rules/aids will be published on Canvas.
7. Evaluation criteria:

Criteria	**Points**
Introduction and **precise problem statement** of the paper, i.e., what problem the paper is trying to solve and why such a problem is important	1
Clear presentation on the proposed method and show how the proposed solution can overcome the problem you stated in the first part. Note: Often, a paper presents many techniques; please focus on one or two techniques and explain them well, instead of making slides of the entire paper and not presenting the entire paper well.	5
Experimental results	1
Quality of slides (Animation, less text, no fancy slides, prefer the style of the slides that I use in the class)	1.5
Q/A by all in the class	1.5
Total	**10**

Research-based Project (20 points)

1. **Purpose**

   The purposes of the project are: (1) to allow you to explore a particular topic in data processing and (2) to understand how DBMSs work.
2. Group: You may form a group of size 2.
3. **Due date**: 12/05 (as mentioned in the first class).
4. **Late submission**: Not allowed.
5. The project must have **user interfaces**. For example, users enter their queries and see the intermediate and final results after decryption.
6. **DBMS**: You need to use MySQL DBMS. Do not use any other DBMS.
7. **Dataset**: will be provided.
8. **Programming Language**: Java.
9. **Project Report**
   a. **Must be written on overleaf.**
   b. It must be detailed, and your view on the project must be included. The report must contain each bit-by-bit information that you want to provide and go over during project development.
   c. Must include the following:
      i. An informal overview of the algorithm in your own words.
      ii. Cover the part on how you execute code
   d. **The number of pages**: The report has no page limit. I suggest about 8-10 pages. The length should be appropriate for what you have to say. The report must include an appendix, and the number of pages of the appendix is not counted as the 8-10 pages of the report.
   e. **Appendix**:
      i. Must include the screenshots of the project.
      ii. Must include the following numbers/graphs. You may put graphs/numbers in the form of a table in your primary report. Or you can use any innovative method to show the following:
         1. The system setting on which you execute your project, like RAM, CPU, disk/SSD.
         2. The time to execute the algorithm and different settings.

f. **Font**: Use 10 font size and Times.

g. **Margin**: 1 inch from each side.

h. **Abstract**: An informative abstract must be included.

10. **Project slides:** Each project must have 7-10 slides, describing your project, objective, approach selected, and the outcome.
11. **Code and related files:** Submit the entire code, datasets, and other files you used.
12. **Well-documented code: *The code must be well-documented***.
13. Each group member needs to submit the **evaluation form** on CANVAS (given at the end of this instruction file) for their group member.
14. **Deliverables: must be uploaded to a Google folder that I will share with you in the first week of December.**
    a. Code + data file
    b. Report
    c. Slide
15. **Grading**: The project will be evaluated based on scientific merit (correctness and completeness), effective presentation (clarity, organization, English usage, document completeness), and appropriateness of the written technical report to the assignment.
    a. ***Once you submit your project, I will meet each group in my office and check the project code on their system.***
    b. **The project will include 20 points out of 100.** These 20 points will be partitioned as follows:

Topic	**Points**
Presentation of the project work	2
Report in tex– covering all details	3
Project (code and file submission), execution of the project at my end, and matching against the reported results in the report	12
Well-documented code	3
Total	**20**

Academic Integrity and Note for Students with Disabilities

We will follow these rules as mentioned on page 6.