CSCI 4707: Practice of Database Systems

Fall 2023

Instructor:

Chang Ge (cge@umn.edu)
Welcome to CSCI 4707

Instructor:
Chang Ge
Office: 4-213 Keller Hall
Email: cge@umn.edu  (Prefix subject with CSCI4707)

Office Hours:
- Monday: 1:45 – 2:30 PM
- Wednesday: 3:50 PM – 4:30 PM
- Alternative OHs may be scheduled by email

CS&E offers two sessions of the same course
- This is the 002 session
- Differences in exams
Welcome to CSCI 4707

- **TA:**
  - Yuchuan Huang
    Email: huan1531@umn.edu
    Lead TA on all labs
  - Arig Mostafa
    Email: mosta041@umn.edu
    Lead TA on all HWs
  - Shengya Zhang
    Email: zhan9051@umn.edu
    Lead TA on all exams

- Office hour will be posted shortly
- Prefix subject with CSCI4707-002 when emailing TAs
Grading Policy

- Four Homeworks: 20% (5% each) (Groups of 2 or less)
- Three Labs: 30% (Groups of 3 or less)
  - Lab # 1 (6%) → SQL
  - Lab # 2 (12%) → Inside the engine
  - Lab # 3 (12%) → Inside the engine
- Two Midterm Exams: 25% (12.5% each)
- Final Exam: 25%
- Late submissions will not be graded
Welcome to CSCI 4707

- **Text Book:**
  Raghu Ramakrishnan and Johannes Gehrke
  • Chapters: 1-5, 8, 9, 12-19

- **Two useful references:**
  - The complete book
  - Encyclopedia of Database Systems
Welcome to CSCI 4707

- Class web page:
  - https://www-users.cse.umn.edu/~cge/teaching/csci4707/
  - Course schedule, slides

- Other platforms we will use:
  - Canvas, for announcements
  - Piazza, for discussions, Q&A
  - Gradescope, for submitting HWs and labs
What we will study in the Course

- **PART 1: Outside the Database Engine as a User**
  - How to do conceptual database design (Chapter 2)
  - How to do a logical database design (Chapter 3)
  - How to query the database (Chapters 4, 5)
  - How to enhance the logical design (Chapter 19)

- **PART 2: Inside the Database Engine**
  - Storage and indexing modules (Chapters 8-10)
  - Query processing & optimization (Chapters 12-15)
  - Transaction Processing (Chapters 16-18)
What Is a DBMS?

- How do you store your phone contacts:

<table>
<thead>
<tr>
<th>Name</th>
<th>Home</th>
<th>Cell</th>
<th>Address</th>
<th>Email</th>
</tr>
</thead>
</table>

- Designing, storing, and managing such “simple” tables is the core of DBMS
What Is a DBMS?

- A very large, integrated collection of data.
- Models real-world enterprise.
  - Entities (e.g., students, courses)
  - Relationships (e.g., Madonna is taking CSCI 1234)
- A Database Management System (DBMS) is a software package designed to store and manage databases.
History of DBMS

- Early 60’s
  - First general-purpose DBMS by Charles Bachman at General Electric (First recipient of ACM Turing Award in databases in 1973)

- Late 60’s
  - Hierarchical data model developed at IBM
  - The SABRE system for airline reservation is jointly developed by American Airlines and IBM where several people can access the same data through a network
History of DBMS

- **70’s**
  - The relational data model is proposed by Edgar Codd at IBM (ACM Turing Award at 1981)
  - Two main prototypes for relational database management systems are developed. Ingres at UCB and System R at IBM (Mike Stonebraker received the ACM Turing Award at 2014)
  - Peter Chen (MIT) proposed the entity-relationship model

- **80’s**
  - SQL query language is developed (part of System R)
  - The concept of read/write transactions is developed to allow concurrent execution of database operations (Jim Gray received the ACM Turing Award at 1998)
  - Commercial databases are in the market (DB2, Oracle, Informix)
History of DBMS

- 90’s – 10’s
  - DBMSs are well-established in industry and academia
  - More applications
  - Big data: volume, velocity, and variety

- 10’s – present
  - New DBMS architectures
  - Cloud native
  - DBMS as a service
Files vs. DBMS

- Special code for different queries
- Must protect data from inconsistency due to multiple concurrent users
- Crash recovery
- Security and access control