Term: Fall 2025   **Subject:** Computer Science & Engineering (CSE)   **Number:** 512

Course Title: **D**istributed **D**atabase **S**ystems (CSE 512)

Homework 01 (20 Points)
Due on Monday, August 2nd at 11:59 pm

Motivation: Multinational corporations operate in diverse locations and manage multiple projects simultaneously. Tracking employees, responsibilities, project budgets, and compensation across regions requires a robust relational database. A well-structured design ensures accurate reporting, reduces redundancy, and supports efficient decision-making at a global scale.

Problem Statement: You are hired as a database designer for a multinational company that manages projects across several countries. Each project may be located in a different city or country and requires employees with specific job titles and responsibilities. Employees can be assigned to multiple projects at once, sometimes in different regions, with varying durations. Salaries are standardized by job title but may need to be compared across projects and locations for analysis. The company wants to,

- Maintain clear relationships among employees, projects, assignments, and salaries.
- Avoid redundancy and anomalies by using proper normalization.
- Support complex queries for reporting, including identifying global staffing patterns, comparing budgets, and analyzing labor costs across multiple locations.

The schema includes the following four relations:

- **EMP**(ENO, ENAME, TITLE)
- **ASG**(ENO, PNO, RESP, DUR)
- **PROJ**(PNO, PNAME, BUDGET, LOC)
- **PAY**(TITLE, SAL)

Intuition: Employees (EMP) work on projects (PROJ). Their assignments are recorded in ASG with a responsibility (RESP) and a duration in months (DUR). Titles map to salaries through the PAY relation.

Q1. Entity–Relationship Diagram (ERD) - Draw an ERD illustrating the relational database schema. Show   and attributes, primary key, foreign key, relationships, and cardinalities. (5 points)

Q2. Relational Schema Definition – Define and explain the terms - relation, attributes and tuples, primary key, foreign key, and integrity constraints in the context of the given schema. (5 points)

Q3. Normalization - Provide two normalization cases to avoid redundancy and anomalies in the context of the given schema. (5 Points)

For each,

- State the functional dependency or anomaly
- Explain the normalization step (1NF/2NF/3NF)
- Describe how this improves the design.

Q4. Write at least five queries for the given schema. For each query, provide a natural language description of what the query should return, and its equivalent SQL expression. You may choose your queries from the list of operations below, but each query must use a different operation (no repeats): (5 Points)

- Relational Algebra operations: Selection, Union, Set Difference, Cartesian Product, Insertion, Theta (θ)–Join, Semi-join, Division

Submission Instruction:

- Submit a single PDF file containing the complete solutions for Q1–Q4.
- All solutions must be handwritten (digital handwriting, e.g., stylus, iPad, OneNote, is also acceptable).
- Typed answers will not be accepted.
- You may scan or take clear photos of your work. *(Unclear photos used will be rejected.)*
- The ER Diagram may be either drawn by hand or created digitally using any appropriate diagramming tool.