# CSE 412/598 – Database Management – Midterm Exam

March 12, 1:45 - 2:55pm (1hr 10mins)

**Name:**

**ASU ID:**

**Circle One:** **412** **598**

**Instructions:**

- **Write your Name and ASU ID.**
- The exam is open notes, open books and closed mouth.
- **Duration**: 70 mins **Pages**: 5 pages
- Use only space provided on each page to solve the problems. We will not read solutions that spill over to the next page. Do not use this booklet as draft – use scrap paper.
- Write neatly. If we cannot decipher your handwriting, we will not grade it.
- **Total Points**: 110, (with 10 points bonus). Good Luck!

Problem 1- **(25 points)** ER Design

Perform the ER design for the following simple enterprise. List all entities and relationships with all their attributes/roles and constraints. Especially indicate key constraints and participation constraints.

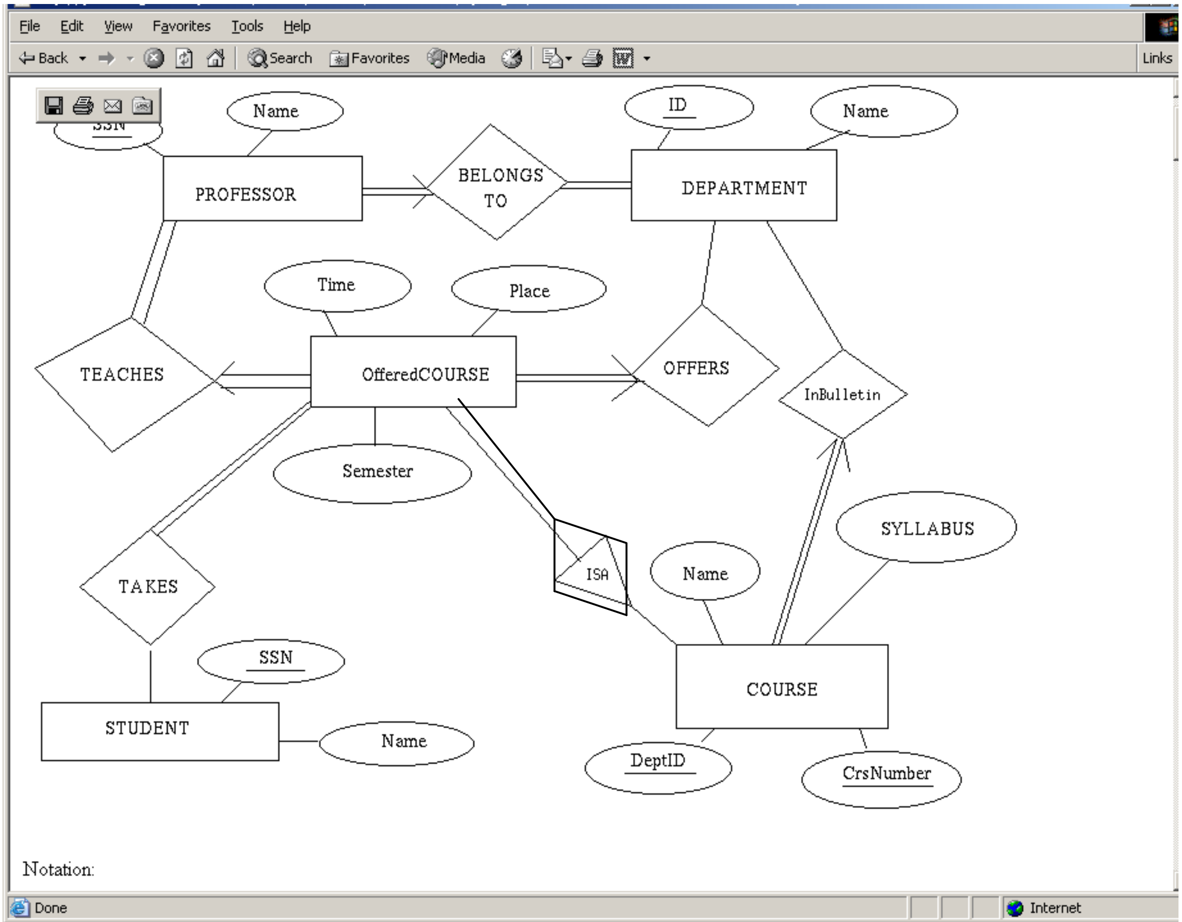
**No need to** convert the ER design into the relational model.

The Enterprise: A small part of a university enrolment system. There are students who take courses, and there are professors who teach them. Your design should allow for the fact that not all courses in the curriculum are offered in the current semester. For current courses, you should provide meeting place and time but only general information is needed for courses that are not offered this semester.

Professors teach courses and they work for university departments. Assume that *every* professor teaches at least one course but no more than two courses and *every* professor belongs to exactly one department. Students may or may not take courses.

If necessary, you can make additional *reasonable* assumptions regarding attributes and constraints.

Use this page for your ER Diagram.




Problem 2 – **(30 points)** Relational Algebra

Consider the following relational schema:
**employee** ( ssn, name, gender, address, salary, supervisorSSN, dnumber)
**department** ( dnumber, dname, managerSSN)
**deptLocation** ( dnumber, dlocation)
**project** (pnumber, pname, plocation)
**worksOn** ( emplSSN, pnumber, hours)

Answer these queries using the relational algebra: **(each 10 points)**

1) Find each employee (ssn) who does not work on any project.

$\pi_{ssn}$ (employee) - $\pi_{emplSSN}$ (worksOn)

2) Find each department (dname) who has employees working on projects at locations other than a department's location (i.e. plocation =/= dlocation)

$\pi_{dname}$ ($\sigma_{plocation <> dlocation}$ (
deptLocation $\bowtie_{dnumber}$ department $\bowtie_{dnumber}$ employee $\bowtie_{ssn=emplSSN}$ worksOn $\bowtie_{pnumber}$ project ))

3) Find all projects (pnumber) that are being worked on in all departments. (A project is being worked on in a department iff at least one employee in that department is working on that project.)

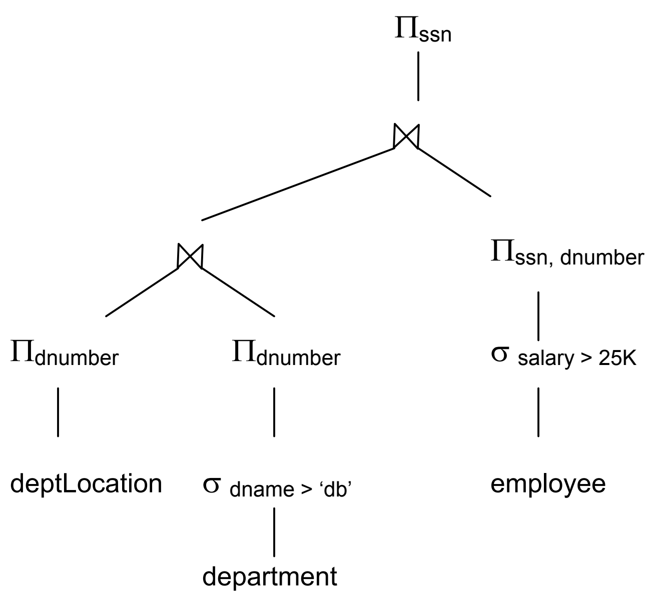
$\pi_{pnumber,\ dnumber}$ (worksOn $\bowtie_{ssn=emplSSN}$ employee) / $\pi_{dnumber}$ (department)

Problem 3 – **(25 points)** Query Optimization

Consider the following relational algebra query for the above schema in Problem 2.

$$\Pi_{ssn}\,(\sigma_{\,salary > 25K\ \wedge\ dname = \text{'db'}}\,((\text{employee} \bowtie \text{department}) \bowtie \text{deptLocation}))$$

Assuming |department| = |deptLocation| < |employee|, write an efficient optimized relational algebra expression that is equivalent to the above query.




Problem 4 – **(30 points)** Relational Design

Answer the following questions: **(each 10 points)**

1) Prove or disprove the soundness of the following inference rule for the FDs: **XY → Z and Z → W implies X → W**. Use a tuple based argument applied to an arbitrary relation. (Hint: use relation R with attributes X, Y, W, Z)

| X | Y | Z | W |
|---|---|---|---|
| 1 | 1 | 1 | a |
| 1 | 0 | 2 | b |

The above table disproves the soundness.

2) Consider the schema R = ABCDE with the set of FDs F={AB→CD, C→ B}. Is the decomposition {AE, BC, ACD} of R lossless with respect to F?

| | A | B | C | D | E |
|---|---|---|---|---|---|
| AE | a | | | | e |
| BC | | b | c | | |
| ACD | a | | c | d | |

After applying C→ B we have

| | A | B | C | D | E |
|---|---|---|---|---|---|
| AE | a | | | | e |
| BC | | b | c | | |
| ACD | a | b | c | d | |

After applying AB→CD the table does not change. Since we have no rows complete with a,b,c,d,e the decomposition is **lossy**.

3) Consider the schema R = ABCDEFGH with the following set F of FDs:

F = {ABH → C, A→ DE, BH → FG, F→ G}

Find a lossless dependency preserving 2NF and 3NF decomposition. Is your 3NF decomposition in BCNF?

**ABH** is a candidate key.

2NF: <ABHC>, <ADE>, <BHFG>

3NF: <ABHC>, <ADE>, <BHF>, <FG>

The above 3NF is in BCNF since whenever X→ Y on R, X is a super key of R.