CS348: INTRODUCTION TO DATABASE SYSTEMS
(Sample Midterm)

TIME: 80 minutes

NAME AND ID HERE: ________________________________

NOTE 1: This is a closed book examination. For example, class text, copies of overhead slides and printed notes may not be used. There are 12 pages. The last page, only, may be separated and used as an easy reference for the second question. Answer all questions in the space provided.

NOTE 2: Some of the questions in this examination are open ended; however, they can be answered to the level discussed in class by short organized answers. It is recommended that you spend part of your time organizing your answer, rather than writing down ideas in the order they occur to you. The conciseness and organization of your answers will be taken into consideration in the grading.

NOTE 3: There are 100 marks in total, 5 of which are bonus marks. As a guide to managing your time, the marks awarded for each question are indicated in parenthesis at the start of the question.

NOTE 4: You are also being tested on your ability to understand the questions. In the case of a perceived ambiguity, state a clear assumption and proceed to answer the question.

NOTE 5: Cheating is an academic offence. Your signature below indicates that you understand and agree to the University's policies regarding cheating on exams.

PLEASE SIGN HERE: ________________________________
(27 marks; continued on next two pages) General questions about database systems, the relational model and about the SQL standard. Answer each of the following using no more than a few sentences in each case.

(a) Explain why it might be desirable to use an ability to have more than one external schema over an enterprise’s conceptual schema. What kind of data independence is achieved by using this ability?

(b) Explain each of the following terms.
   
   1. physical schema

   2. isolation

   3. durability
(c) Does it make sense for a relation to have more than one primary key? Does it make sense for a relation to have more than one possible choice for a primary key? Justify your answers.

(d) Explain how null values in SQL should be understood.
(e) Assume relation $R$ has attributes $\{A, B, C\}$, with $A$ as the primary key. Can the following SQL query be expressed in the relational calculus satisfying the range restricted syntax? Justify your answer.

```sql
select B from R
group by B
having count(*) > 2 and min(C) < 10
```
II. (45 marks; continued on next three pages) Consider the following SQL data definition for maintaining information about employees at a hypothetical company.

CREATE TABLE emp
    (    num INTEGER NOT NULL,
         name VARCHAR(20) NOT NULL,
         dept VARCHAR(20) NOT NULL,
         sal INTEGER NOT NULL,
         boss INTEGER NOT NULL,
    PRIMARY KEY (num),
    FOREIGN KEY (boss) REFERENCES emp (num) );

You can assume the following:

1. There is one person, the president, that has herself/himself as the boss;

2. All other employees have a boss they report to that is someone else;
   and

3. There are no cycles in the boss hierarchy for anyone other than the president. (A cycle would exist if, for example, Fred was the boss of Mary and Mary was in turn the boss of Fred.)

Also, an employee is referred to as a *manager* if the employee has at least one other distinct employee with the employee as boss.
Write each of the following queries in both the relational calculus and in SQL.

(a) The number and name of each employee, excluding the president, together with the number and name of the employee’s boss.
(b) The number and name of each employee with a salary among the lowest in his/her department, and that has a boss in a different department that has a lower salary.
Write the following query in SQL only.

(e) The number and name of each manager together with the total salary for all employees for whom he or she is the boss or for whom he or she is the boss of the boss.
III. (18 marks; continued on next page) Questions on the SQL standard and application development. Answer each of the following using no more than a few sentences in each case.

(a) In what way can a SQL view behave differently from a SQL table? Justify your answer.

(b) Give one reason why it can be desirable to adopt the CLI protocol when developing applications.
(c) Give two reasons why it can be desirable to adhere to the static embedded SQL protocol when developing applications.

(d) Explain the purpose served by each of the following.

1. an SQL trigger

2. an indicator variable

3. a cursor
IV. (10 marks) Questions on ER modelling. Answer each of the following using no more than a few sentences in each case.

(a) Explain how agents communicate references to a subordinate entity in a weak entity set.

(b) (Bonus) Is it possible to “over-constrain” an entity set E in an ER diagram, that is, for the rest of an ER diagram to imply that E must not have any entities? Justify your answer.
The following is reproduced from Question II. You may detach this page to assist composing your answers to Question II. NOTE: do not write on this page.

CREATE TABLE emp
    (  num INTEGER NOT NULL,
        name VARCHAR(20) NOT NULL,
        dept VARCHAR(20) NOT NULL,
        sal INTEGER NOT NULL,
        boss INTEGER NOT NULL,
        PRIMARY KEY (num),
        FOREIGN KEY (boss) REFERENCES emp (num) );