UNIVERSITY OF WATERLOO
Computer Science CS 338 Final Examination
Makeup Examination

EXAMINOR: G. E. Weddell TIME: 120 minutes

This is an open book examination. In particular, class text, copies of overhead slides and printed notes may be used. Do all questions.

NOTE 1: 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 2: There are 100 marks in total. As a guide to managing your time, the marks awarded for each question are indicated in parenthesis at the start of each question.

Question 1. (28 marks) Answer each of the following.

(a) Explain what it means for a functional dependency $X \rightarrow Y$ to be logically implied by a set of functional dependencies F; that is, for $X \rightarrow Y$ to be a member of F^+. Can this be determined efficiently?

(b) Again, let R denote a set of attributes and F a set of functional dependencies over R. Outline an algorithm for finding a candidate key for R.

(c) In two or three sentences, explain why at most one clustering index can be declared on a relation. Why is a clustering index preferable to a nonclustering index? Justify your answer by suggesting a simple SQL query for which a clustering index on an attribute will enable the query to be evaluated more efficiently than would be possible with a nonclustering index on the same attribute.

(d) In what way is more indexing desirable? In what way is more indexing undesirable?

(e) Indicate whether each of the following statements is true or false. In each case, justify your answer using no more than three sentences.

1. The SQL standard excludes any consideration for the internal schema of a relational database.
2. In a SQL database system that conforms to the SQL standard, it is not possible for a transaction to retrieve (read) a tuple that has been updated by another active transaction.
3. Once defined, a SQL view can be used exactly like a SQL base table.
4. If a database system does not maintain statistics, it will be unable to process queries.

Question 2. (25 marks) Assume you are developing an information system for budget management that will use a relational database system to store budget data. An initial analysis phase of the project has resulted in the following (informal) description of the relevant data for the system.

- A department has a number of employees and projects.
- One of the employees in the department is the manager.
- An employee has a name.
- Each employee who is not the manager reports to another employee in the department.
- A project has a project name, a weekly budget and a number of assigned employees.
- A given employee can be assigned to any number of projects.
- An employee is assigned to a project for a given number of hours per week.
- An employee is either permanent or temporary.
- Permanent employees have a yearly salary.
- Temporary employees are paid at an hourly rate.

Create an E-R diagram that represents the budget data. Include additional comments in your design that explain any unusual representation decisions.

Question 3. (25 marks) Consider the following SQL data definition for maintaining information about employees at a hypothetical company.

```
CREATE TABLE Employee
   (  Num              INTEGER NOT NULL,
      Name             VARCHAR(20) NOT NULL,
      Dept             VARCHAR(20) NOT NULL,
      Salary           INTEGER NOT NULL,
      BossNum          INTEGER NOT NULL,
   PRIMARY KEY (Num),
   FOREIGN KEY (BossNum) REFERENCES Employee (Num) );
```

It is assumed that the president has herself/himself as boss, that all other employees report to someone else and that there are no cycles in the reporting hierarchy for anyone other than the president. (A cycle would exist if, for example, Fred's boss was Mary and Mary's boss was Fred.) Write SQL queries on this schema to answer each of the following.

(a) The count of the number of departments in the company.

(b) The name of each employee, excluding the president, together with the name and salary of the employee's boss.

(c) A sorted list of department names, together with the total salary of all employees for each department.

(d) The average number of employees in a department.

(e) A sorted list of the names of all employees not managing any other employees or that have a manager in a different department.

Question 4. (22 marks) Consider again the Employee table in Question 3 and answer each part.

(a) For the following SQL query on this table: (1) perform an initial translation of the query into the relational algebra, and (2) derive an equivalent query in the relational algebra in which you perform selections as early as possible, and in which more selective joins appear earlier in any join order. (Note that projections need not be considered when modifying the query in part (2).)

```
SELECT DISTINCT E.Name
FROM Employee E, Employee BB, Employee B
WHERE E.BossNum = B.Num
AND B.BossNum = BB.Num
AND BB.BossNum = BB.Num
AND B.Dept = 'Toy'
```

Explain the result returned by this query in English.

(b) Consider the query:

```
SELECT DISTINCT E1.Name as FirstEmp, E2.Name as SecondEmp
FROM Employee E1, Employee E2
WHERE E1.Num <> E2.Num
AND EXISTS( SELECT * FROM Employee B
            WHERE E1.BossNum = B.Num
            AND E2.BossNum = B.Num )
```

How many tuples can the query compute in the worst case if there are 100 tuples in the Employee table itself? How many tuples would be computed in the best case if there are again 100 tuples in the Employee table? Briefly justify your answer in both cases.