UNIVERSITY OF TORONTO AT SCARBOROUGH
Spring Final 2001

Database Management Systems – CSCD34

Duration — 2 hours

Examination Aids: One $8\frac{1}{2}$" $\times$ 11" sheet of paper.

- **Make sure your examination booklet has 12 pages (including this one).**

Family Name: ____________________ Given Names: ____________________

Student #: ____________________

1.	________	/ 15
2.	________	/ 15
3.	________	/ 25
4.	________	/ 10
5.	________	/ 15
Total	________	/ 80

Question 1 True or False [15 points in total]

Write 'T' or 'F' next to each question. 1 for each correct answer, -1 incorrect answer, 0 null answer

____ 1. Every serializable schedule has an acyclic serializability graph.

____ 2. In the simple nested loops join algorithm, it is better to put the smaller relation in the inner loop.

____ 3. The I in "ACID" stands for "integrity."

____ 4. Strict two-phase locking guarantees that schedules are serializable.

____ 5. In processing a project-select query, projection should be performed before selection whenever possible.

____ 6. Every dependency-preserving decomposition is lossless.

____ 7. If a relation scheme has no non-trivial functional dependencies, then it is in Boyce-Codd Normal Form.

____ 8. A nested SQL query consists of more than one query block.

____ 9. The natural join in relational algebra is commutative.

____ 10. For a selection condition of the form $\sigma_{R.A=c}$, the cost of using an unclustered B^+ tree to find all the records that satisfy the condition is proportional to the height of the tree.

____ 11. The number of disk accesses required to find a data entry with a given key value in a B^+ tree grows proportionally to the height of the tree.

____ 12. If V_1 and V_2 are two SQL variables, the value of the conditional expression `(V1 = V2) AND (NOT V1 = V2)` is always FALSE.

____ 13. A typical relational query optimizer always considers all possible query plans and chooses the one with the lowest cost.

____ 14. Two actions on a data item conflict whenever at least one of them is a read.

____ 15. Under the 2PL protocol, a value written by a transaction T is not read or overwritten by any other transaction until T either commits or aborts.

Question 2 [15 marks in total]

Consider two relation schemes describing car sales at all stores of a chain of automobile dealerships: a tuple in `R[InvoiceNum,Store,Model,DateID,Price]` describes a sale of model `Model` at price `Price` at store `Store`, on date `DateID`. Dates are represented by a numeric identifier; a tuple in `D[DateID,Day,Month,Year,DayOfWeek]` describes each date identifier. The primary key of `R` is `InvoiceNum`, the invoice number, and the primary key of `D` is `DateID`.

1. (3 marks) Define in SQL a **view** `DOWSales[InvoiceNum,Store,DOW,Price]` where each tuple represents a sale of a car at a particular store on a particular day of the week, and the sale price.

2. (6 marks) Express in SQL the query that gives, for each store and each day of the week, the average number of cars sold at that store on that day of the week, over the life of the database. (You may use the view defined in part 1 if you wish).

3. (6 marks) Define the *performance* on a day of the week as the total dollar amount sold on that day of the week on all stores over the life of the database. Express in SQL the query that finds the days of the week with the best performance. (You may use the view defined in part 1 if you wish).

Question 3 [25 marks in total]

Consider a database with two relations: `Person(SIN, firstname, lastname, birthdate)` and `Spouse(SIN1,SIN2)`. (SIN = Social Insurance Number) `SIN` is the primary key of `Person`, and `Spouse` has two candidate keys, `SIN1` and `SIN2`.

(a) [3 marks]
Write SQL CREATE TABLE statements to create these two relations, with all apropriate integrity constraints. Assume that if person with SIN s1 is married to person with SIN s2, then only the tuple (min(s1,s2),max(s1,s2)) appears in the relation *Spouse*. Assume also that every SIN that appears in `Spouse` also appears in `Person` (but not conversely).

(a) [3 marks]
Give a relational algebra expression that lists the SIN's of all persons who have the same birth date as their spouse.

(b) [4 marks]
Give a relational algebra expression that lists the names of all the single (i.e. unmarried) people in the database.

(c) [4 marks]
Give a relational algebra expression that lists the SIN's of each pair of persons whose spouses have the same last name. That is, pairs (s_1, s_2) such that s_1 and s_2 are different and the spouse of the person with SIN s_1 has the same last name as the spouse of the person with SIN s_2. If pair (s_1, s_2) appears in the list, then pair (s_2, s_1) should not.

(d) [6 marks]
Give a relational algebra expression that lists those last names such that every person with that last name is married.

(e) [5 marks]
Write a program in a programming language of your choice, using Embedded SQL, that given a last name supplied by the user, prints out the first and last names of all persons who are married to someone with that last name. Do not worry about elaborate user interactions or formatting of the output.

Question 4 [10 points in total]

For the following schedule of three transactions T_1, T_2 and T_3:

$$R_1(x), W_2(x), W_1(x), W_3(x).$$

(a) [2 marks]
Construct the serialization (precedence) graph.

(b) [4 marks]
If the schedule is conflict serializable, give one equivalent serial schedule; if the schedule is not conflict serializable, state the reason.

(c) [4 marks]
If the schedule is serializable, give one equivalent serial schedule; if the schedule is not serializable, state the reason.

Question 5 [15 points in total]

Consider the following relational schema where FK stands for foreign key and all primary keys are underlined.

`Employee [` **Eid**`, EName, Age, Dept ]` Employee[Dept] FK of Dept
`Dept [` **Did**`, DName, Manager ]` Dept[Manager] FK of Employee, Unique(Manager)

Consider the following query: "return the names of all departments that are managed by an employee under age 40".

Employee has 100,000 tuples stored on 100 pages. Dept has 1,000 tuples stored on 10 pages. There are only 3 buffer pages available.

1. (3 marks) Write the query above in relational algebra.

2. (3 marks) Draw the query tree for the query above.

3. (4 marks) Draw a query evaluation plan for the query above.

4. (5 marks) Suppose the join in the query is implemented as a block nested join, with `Dept` as the outer relation. How many buffer pages are needed to reduce the cost of this join to the minimum possible number of I/Os, and what is this number? (Use the cost model used in the textbook to count the number of I/Os).