UNIVERSITY OF TORONTO AT SCARBOROUGH
Spring Final 2000

DATABASE MANAGEMENT SYSTEMS – CSCD34

Duration — 2 hours

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

- **Make sure your examination booklet has 11 pages (including this one).**

Family Name: ______________________ Given Names: ______________________

Student #: ______________________

| | | |
|---|---|---|
| 1. | ________ | / 15 |
| 2. | ________ | / 20 |
| 3. | ________ | / 20 |
| 4. | ________ | / 15 |
| 5. | ________ | / 15 |
| Total | ________ | / 85 |

**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 relational algebra expression can be expressed by an equivalent SQL SELECT statement.

____ 2. An E/R diagram containing weak entity sets cannot be translated to a relational schema.

____ 3. The C in "ACID" stands for "concurrency."

____ 4. Two-phase locking guarantees that cascading aborts do not occur.

____ 5. Every serializable schedule is conflict-serializable.

____ 6. Using an unclustered index to implement a $\sigma$ operation may in some cases be slower than doing a sequential scan.

____ 7. If a relation scheme is in Boyce-Codd Normal Form, then it is also in Third Normal Form.

____ 8. An SQL view cannot be updated.

____ 9. The division operator in relational algebra is associative.

____ 10. CHECK constraints in SQL make foreign key constraints unnecessary.

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

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

____ 13. To join four relations $R_1$, $R_2$, $R_3$, $R_4$, one possible plan that the System R optimizer would consider consists of first joining $R_1$ and $R_2$, then $R_3$ and $R_4$, and then joining the two partial results.

____ 14. Strict two-phase locking guarantees deadlocks cannot occur.

____ 15. A file cannot have clustered indexes on two different fields.

**Question 2** [20 marks in total]

Suppose we have a relation *Sales* describing the sales of a chain of stores, with attributes *Store, Item, Colour, Day, Month, Year, Price*. A tuple *(s,i,c,d,m,y,p)* in this relation means that item $i$ of colour $c$ was sold at store $s$ on the date given by $d$, $m$ and $y$, for a price $p$.

Express the following queries in SQL:

1. (3 marks)Total sales (dollar value) for each store

2. (3 marks)Total sales of red items for each store

3. (4 marks) Total sales sold by each store each year

4. (4 marks) Number of different items sold by each store (e.g., if a store sells one kind of chair, one kind of couch, and one kind of coffee table, the number for that store is 3).

5. (6 marks) Stores that have sold the largest number of different items.

**Question 3** [20 points in total]

Consider a system of three transactions $T_1$, $T_2$, and $T_3$. Transaction $T_1$ adds a \$10 bonus to a customer's savings account and then a \$5 bonus to a customer's chequing account. Transaction $T_2$ adds a 5% bonus to the same customer's chequing account. Transaction $T_3$ sets both balances, chequing and savings, to \$0, for the same customer. Call the savings account data item for this customer $s$ and the chequing account data item $c$.

1. (4 marks) Write down a serial schedule for $T_1$, $T_2$ and $T_3$.

2. (6 marks) Write down a non-serial schedule for this same set of transactions that is conflict serializable. Justify its conflict serializability.

3. (5 marks) Consider the schedule

$$S : R_1(s), W_1(s), R_2(s), R_2(c), W_2(s), W_2(c), R_1(c), W_1(c), W_3(s), W_3(c).$$

Show that this schedule is not conflict serializable by exhibiting the precedence graph.

4. (5 marks) Show that the schedule $S$ above is serializable.

**Question 4** [15 points in total]

The following schema describes information about web pages.

`Page[URL, Author, Title , Keyword ]`

A tuple `(u,a,t,k)` of the `Page` relation gives a description of a web page with URL `u`, author `a`, title `t`, and containing keyword `k`. Each page has exactly one author, exactly one title, and is uniquely identified by a URL. A page may contain many keywords. Assume all domains are strings of length 50.

1. (3 marks) Give the SQL table creation statements for the relation above, including keys, foreign keys, and null constraints.

2. (3 marks) Give a set of functional dependencies for relation `Page`. Show that this relation scheme is not in Boyce-Codd Normal Form.

3. (3 marks) Find a BCNF decomposition of relation scheme `Page`.

4. (6 marks) Using the decomposed schema you obtained in the last part, write a program in a language of your choice, using Embedded SQL, that implements a simple *web search engine*: given a keyword supplied by the user, the program prints out the URL's of all the pages that contain this keyword. Do not worry about elaborate user interactions or formatting of the output.

**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 Employee, Unique(Manager)

Consider the following query: “return the names of all employees under age 40 who manage some department”.

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 plan for the query above.

4. (5 marks) Suppose the join in the query is implemented as a block nested join, with `Employee` as the outer relation. How many buffer pages are needed to reduce the cost of this join to 500 I/O operations?