Prof. Li-Yan Yuan

CMPUT 391: Database Management Systems

Final Examination

April 19, 2001

It is a close-book examination and the time for the test is 120 minutes. There are ten (10) questions with 10 marks each. Good luck to all of you.

1. One of the new features in the object-relational database systems is the support of nested relations. (Note that a relation is nested if its table cells may also be a table.) Give an example to demonstrate that nested relations can lead to efficiency gains, in terms of space.

   You need to present a real-world-like scenario, two sample tables, (one is nested and one is not), such that the scenario can be represented by both tables but the nested table is more efficient.

2. Consider a relational database about a university with the following three relations

   teach(Prof, Course)
   take(Student, Course, Grade)
   advise (Prof, Student)

   The first relation indicates the courses a prof teaches; the second tells what courses each student takes and the corresponding grades; and the last indicates advisors of a student.

   Write an SQL query to list all the professors who never teach any course taken by more than five students.

3. Consider a database consisting of the following tables.

```
employee( e_id, e_name, position, salary )
works( e_id, d_id, pct_time )
dept( d_id, budget, manager_id )
```

   The first table indicates the job title and salary for each employee, the second shows the percentage of time that a given employee works for a given department, and the last one indicates the budget and boss of each department.

   Write SQL triggers to ensure that (1) each manager must be an employee, (2) if the manager of a department is deleted (i.e., either a tuple in `dept` is deleted or a tuple in `employee` with e_id matching manager_id of a tuple in `dept`), all the employees in the department must also be deleted.

   Note that the company will certainly NOT fire all the employees in a department when the manager of the department is changed (updated).

4. List **ALL** non-trivial multivalued dependencies satisfied by the following relation. (Note that $X \twoheadrightarrow W$ is non-trivial if $X \cap W = \emptyset$ and $XW \subset R$.)

A	B	C
a2	b2	c1
a1	b1	c2
a1	b1	c3
a2	b3	c1
a2	b2	c3

5. Consider the universal relation $R = ABCDEGHK$ and the following set F of functional dependencies

$$A \rightarrow BC$$
$$AC \rightarrow BD$$
$$E \rightarrow H$$
$$BG \rightarrow H$$
$$H \rightarrow E$$

Find a join loss-less, dependency preserving and 3NF decomposition of R. Is your database schema BCNF ?

6. The following is cited from a question and its solution in Assignment.

> In designing a relational database, why might we choose a non-BCNF design?
> **Answer:** For some applications, there exists no database schema that is both BCNF and dependency preserving. Therefore, one may choose to design a database schema that is in 3NF and dependency preserving, instead of BCNF.

Give an example to further demonstrate the answer.

You need only to present a set R of attributes and a set F of FDs such that there exists no decomposition D of R that is in BCNF and dependency preserving with respect to F.

7. Draw the precedence graph of the following schedule and determine whether the schedule is serializable. (Note that all instructions, except lock and unlock, are omitted. We assume the transaction that requests a read-lock on Q will read from Q and one that requests the write-lock on Q will read from and write into Q.)

T_1	T_2	T_3	T_4
	read-lock(A)		
	write-lock(B)		
	unlock(A)		
		write-lock(A)	
	unlock(B)		
read-lock(B)			
		unlock(A)	
			read-lock(B)
read-lock(E)			
			unlock(B)
write-lock(C)			
unlock(E)			
unlock(B)			
		write-lock(B)	
		unlock(B)	
unlock(C)			
			read-lock(C)
			unlock(C)
	write-lock(C)		
	unlock(C)		

8. The strict two phase locking protocol has two rules:

   (a) If a transaction T wants to read from (or write into) an item it must request and hold a read (or write) lock on the item.

(b) All locks held by a transaction are released when the transaction is committed.

Is a schedule observing the strict two phase lock protocol is always recoverable ? Justify your answer.

9. Consider the relations $r_1(A, B, C)$ and $r_2(A, D)$. Assume that there are no primary keys, except the entire schema, for both tables; r_1 and r_2 have 10,000 and 20,000 tuples respective; and the selection cardinality of two tables are $V(A, r_1) = 1000$ and $V(A, r_2) = 2000$. Further assume that the memory can hold only 1000 tuples from either table.

   (a) Estimate the size of $r_1 \bowtie r_2$.

   (b) Design an efficient algorithm for computing this join, and justify your design. Specify any reasonable assumptions you think fit, and use whatever knowledge you have regarding the database and file systems.

   (You need to present your algorithm in a pseudo language, and justify your design. Recall that some possible choices are the nested loop approach, using an access structure, merge-sort join, hash-join, etc.)

10. Consider the following two tables containing information about the marketing campaign of a magazine. The first three columns show the age, salary, and residency state/province of a potential customer and the subscription column shows whether the person subscribed to the magazine. Please use the first table to construct a decision tree that helps to predict whether a person is going to subscribe to the magazine. The accuracy of your decision can be checked using the second table. The more accurate and simpler, the higher the mark.

age	salary	state	subscription
45	56k	TN	Yes
45	75k	FL	No
32	59k	CA	Yes
62	92k	FL	No
56	42k	TN	Yes
51	74k	TX	Yes
64	38k	AB	No
61	79k	TX	Yes
28	120k	AB	No
19	28k	CA	Yes
21	65k	TX	Yes
28	44k	BC	Yes

age	salary	state	subscription
27	73k	FL	No
66	25k	TN	No
47	69k	AB	Yes
52	28k	TN	Yes
63	59k	TX	No
48	65k	TX	Yes