CS4604 Final Exam

Dec 17, 2002

Please enter the following information:

- **Name:**

- **ID:**

GOOD LUCK and Happy Holidays!
Do not write below this line

Problem	Max Score	Score
1	30	
2	10	
3	10	
4	10	
5	20	
6	10	
7	10	
8	15 (XC)	
Total	100	

1. (30 points) Short answer questions:

   (a) Carefully state the condition to check if a given Datalog query is safe.

   (b) Assume that there is a referential integrity constraint from a column of table A to a column of table B. What types of operations on A can violate the constraint?

   (c) Assume that there is a referential integrity constraint from a column of table A to a column of table B. What types of operations on B can violate the constraint?

   (d) What is the SQL query to do a natural join of two relations A and B?

   (e) Mention one property of decompositions that is *not* satisfied by 4NF.

   (f) How many non-trivial MDs are possible in a two-column relation?

   (g) Three main approaches to database tuning were mentioned in class. Mention them.

   (h) State whether true or false: $\pi_L(R \cup S) = \pi_L(R) \cup \pi_L(S)$. Assume that R and S are sets.

   (i) State whether true or false: $\pi_L(R \cup S) = \pi_L(R) \cup \pi_L(S)$. Assume that R and S are bags.

   (j) What is 'as bad as stems yet' an anagram for?

2. (10 points) Design an E/R diagram for the following situation about books, authors, book publishers, book sellers, and book stores. Books have an ISBN number (the key), a name, and type of printing (e.g., hardcover or softcover). A book can be written by many authors and one author can write many books. Authors have a social security number (key) and a name. Publishers are the companies that manage the printing and production of books (e.g., 'Prentice Hall'). Publishers have a name (key), year of inception, and address. A given book has a unique publisher. Book sellers are the organizations that help deliver books to consumers (e.g., 'Barnes and Noble'). Book sellers maintain book stores to sell books; a given book seller can have many stores (and must have at least one store, to qualify as a book seller). A book store is one of two types - 'brick,' or 'web-based' (and cannot be both).

   For instance, the book seller 'Borders' maintains many brick stores (three in Washington DC, one in Raleigh, and so on) as well as many web-based stores (e.g., `borders.com`, `borders.com.uk`). The book seller 'Amazon' on the other hand has only web-based stores. The bookseller 'Littletown books' has only one single brick store (in Blacksburg), and so on. Brick book stores have a name (key) and address. Web-based book stores have a name (key) and a URL. Some book stores sell books made by many publishers but some book stores are dedicated to selling books made by only one publisher. Try to model as much as you can, and write *Notes* if necessary. Failing to write *Notes* implies that your diagram is exact.

3. (10 points) Consider the relation R(A,B,C,D) with FD $B \rightarrow D$ and MD $AB \rightarrow\rightarrow C$. Decompose the relation into a collection of relation schemas in 4NF. For full credit, show clearly all the steps needed to obtain the final decomposition.

4. (10 points) Consider the relation `Students(id,name,address,gpa)`. Write an SQL query to find the name that appears most often in the `Students` relation.

5. (20 points) Consider the relation `GasStations(id,address,company,revenue)` with tuples such as (1,'323 Main Street, Blacksburg', 'Exxon', $200K), and so on. There are many gas stations in the `GasStations` relation, but they are all franchisees of a small number of companies such as 'Shell', 'Chevron', 'Exxon' etc. Write a query in Datalog to find the company of the gas station that has the **second highest** revenue. If there are many such gas stations, your query should identify the companies of each of these gas stations.

6. (10 points) Let R be a relation with schema:

$$(A_1, A_2, \cdots, A_n, B_1, B_2, \cdots, B_m)$$

and let S be a relation with schema $(B_1, B_2, \cdots, B_m)$; that is, the attributes of S are a subset of the attributes of R. The *quotient* of R and S, denoted $R \div S$, is the set of tuples t over attributes $A_1, A_2, \cdots, A_n$ (i.e., the attributes of R that are not attributes of S) such that for every tuple s in S, the tuple ts, consisting of the components of t for $A_1, A_2, \cdots, A_n$ and the components of s for $B_1, B_2, \cdots, B_m$, is a member of R.

Give an expression of relational algebra that is equivalent to $R \div S$.

7. (10 points) We have seen how to use relational algebra to write queries but it can also be used as a language to express constraints on a relational schema. Consider the relation `Students(id,name,address,gender,gpa)`. Assume that the FD `id` $\rightarrow$ `address` holds in `Students`. This FD can be expressed as:

$$\sigma_{\text{S1.id=S2.id AND S1.address<>S2.address}}(S1 \times S2) = \{\}$$

where $S1$ and $S2$ are copies of `Students`. In other words, the constraint is saying that there can be no combination of tuples that satisfy the condition where the `id` fields are the same but the `address` fields are different.

Use relational algebra to express the following constraints:

- The only legal values of `gender` are 'M' or 'F'.

- There is a referential integrity constraint from the `id` field of `Students` to the `ssn` field of `Individuals(ssn,birthdate,typeofemployment)`.

8. (Extra credit question: 15 points) Consider the relation `GraduatingStudents(id,name,age, starting_salary)`. Write an SQL query to find the median starting salary of graduating students.