CS4604 Final Exam

Please enter the following information:

- **Name:**

- **ID:**

GOOD LUCK!
Have a nice summer!
Do not write below this line

Problem	Max Score	Score
1	30	
2	10	
3	10	
4	10	
5	10	
6	10	
7	6	
8	4	
9	10	
10	15 (XC)	
Total	100	

1. (30 points) Short answer questions:

   (a) Of the three query languages — relational algebra, Datalog and SQL — which one of them is *not declarative*?

   (b) SQL obeys the *closure* property. What does this mean?

   (c) The SABRE online database (`www.sabre.com`) is a huge airline reservation facility that handles bookings and seat assignments for nearly a third of the world's air traveling population. Consider two scenarios: (i) Two travel agents Alice and Bob are selling seats for the same flight at the same time, and (ii) Alice and Bob are selling seats for the same flight at different times. Which of the four ACID properties ensures that nothing bad happens in the first scenario that wouldn't have happened in the second? Just identify the letter.

   (d) (True/False) 3NF removes redundancies but cannot be guaranteed to preserve dependencies.

   (e) A database system can sometimes reorder and rearrange queries so that they can be executed more efficiently. In other words, it can perform query optimization. What data structure present inside the database system assists in this activity?

   (f) Give a real-life example of weak entity sets.

   (g) An example was given in class of an interesting inference made by a data mining program on analyzing supermarket transactional data. What was this example?

   (h) If relation R has m tuples and relation S has n tuples, what is the minimum number of tuples that $R \cup S$ can have (assuming set-theoretic semantics)?

   (i) Give a real-life example of a multivalued dependency (MD).

   (j) (True/False) CS4604 is an interesting course.

2. (7 points) Design an E/R diagram for a university database with the following requirements. The database models instructors (identified by their SSN), courses (identified by a string, such as 'CS4604'), and semesters (such as Fall '99). Instructors offer courses in certain semesters. Any particular instructor might offer the same course in different semesters. An instructor can offer more than one course in a given semester. A given course in a particular semester can be taught by more than one instructor. In addition, the university requires a list of *potential instructors* for each course; i.e., these are instructors who *can teach* that particular course. This information is used for obtaining initial class assignments of courses to instructors.

   (3 points) After drawing your diagram, if there are any aspects of the above english description that cannot be modeled (or perhaps should be enforced by integrity constraints or other domain-specific constraints), point them out.

3. Suppose we have a relation schema `R(A,B,C)` with functional dependency $A \rightarrow B$. Suppose also that we decide to decompose this schema into `S(A,B)` and `T(B,C)`.

   - (5 points) Give an example of an instance of relation R for which the above decomposition is lossless. Also explain which normal form is violated in R.
   - (5 points) Give an example of an instance of relation R for which the above decomposition is lossy. Show how the projection of your instance on S and T, when subsequently rejoined do not yield the original relation instance.

4. (10 points) Consider the relation `Student(ID,name,address,gpa)`. Write a query to find the name of the student with the second highest gpa (If there are many such students, your query should output all of them). Your query should be written in relational algebra.

5. (a) (4 points) Consider the two relations `Student(ID,car)` and `Sportscar(car,maker)`. Write a Datalog query to find all the 'non-sports cars' owned by students.

(b) (3 points) Assume that you are given an SQL query as follows:

```
SELECT a,b,c,...
FROM R1,R2,R3,...
```

i.e., there is no `WHERE` part; only `SELECT` and `FROM` are present in the query. Then, this query is a (pick one of the below:)

- Cartesian Product
- Natural Join
- Theta-Join

(c) (3 points) What is the difference between $R \bowtie S$ and $R \bowtie_C S$ where the condition C is that $R.A = S.A$ for each attribute A appearing in the schemas of both R and S?

6. (10 points) Consider the following mutually recursive Datalog query which attempts to identify the odd and even natural numbers.

```
Odd(x) <- Number(x), NOT Even(x).
Even(x) <- Number(x), NOT Odd(x).
```

It was shown in class that this is not a *safe* query. It is unsafe not because of mutual recursion, but due to the fact that there is no 'unique interpretation' to the query. Example: Under the above two rules, the number 6 could be either in `Odd` or in `Even` — both are acceptable!

Explain how you would rewrite the query so that you can identify the odd and even natural numbers with a unique interpretation (*Hint: This problem was solved in class on Monday, April 12, 1999*).

7. (6 points) Consider the two relations

   ```
   HasTaken(student-ssn,course-number)
   RequiredForGraduation(course-number)
   ```

   with their obvious interpretations *i.e.,* `HasTaken` identifies the courses taken by each student and `RequiredForGraduation` lists all the courses required to graduate. For example, if `RequiredForGraduation` contains 14 courses, this means that all those 14 courses are required to graduate (you may assume that all the `course-numbers` in `RequiredForGraduation` will be different). Write a query in relational algebra to find all the students who have satisfied all the requirements for graduation. If you use a symbol other than $\{\cup, \cap, -, \pi, \sigma, \bowtie, \bowtie_C, \times\}$, you must first define it.

8. (4 points) Consider the following schema:

   `Movie(title,length,studio,producer)`

   Write an SQL query to find those movies that are longer than *Schindler's List*. You may assume that there is only one movie called *Schindler's List*.

9. (10 points) Mr. Big, the president of 'Relations-R-Us' is throwing a party for the entire company. The organization has a hierarchical structure; i.e., it is a tree with Mr. Big at the root. The entire tree has a lot of levels (unknown). Now, in order that the party is interesting for all attendees, Mr. Big does not want both an employee and his or her immediate supervisor to attend. Of course, Mr. Big has to be present because the party is at his residence. Your task is to figure out how to make up the guest list.

    - (5 points) First explain, how you would model this situation by relations. In other words, what are the relations that you will design, what are the columns, how many are there etc.? What kind of data will they contain? Explain with some examples.
    - (5 points) How would you write a query that uses your relations to make up the guest list? This query can be written in any of the three notations - relational algebra, Datalog or SQL.

10. (Extra credit question) Consider a DBMS for the Bank of Littletown, USA, that is being granted outsider access for auditing purposes. The trick is to ensure the privacy of individuals while allowing the use of data for statistical purposes. Assume that we execute two queries on this DBMS:

    - a query that requests the sum of the bank account balances for all customers in Littletown.
    - a query that returns the number of customers living in Littletown.

    If the answer to the second query is 1, we have inadvertently compromised the details of that one customer who happens to live in Littletown! A simple way to overcome this security breach is to place a restriction such as:

    Restriction: *Disallow* all queries that involve fewer than, say, 10 individuals.

    There is still a security breach with this mechanism.

    - (5 points) Describe how a malicious hacker can obtain the account balance of any specific individual in spite of the above restriction.
    - (10 points) Now that you have described the critical flaw, describe a mechanism by which you can overcome such conspiracies. In other words, describe more restrictions that will prevent this problem. Of course, this can go on and on, as hackers get more and more clever but this question just asks for two initial ideas.