CSE 344 Final Examination

December 14, 2011, 8:30am - 10:20am

Name: _____

Question	Points	Score
1	20	
2	15	
3	10	
4	20	
5	20	
6	10	
7	5	
Total:	100	

- This exam is an open book exam.
- You have 1h:50 minutes; budget time carefully.
- Please read all questions carefully before answering them.
- Some questions are easier, others harder; if a question sounds hard, skip it and return later.
- Good luck!

1 E/R Diagrams and Constraints

1. (20 points)

   (a) (10 points) Design an E/R diagram describing the following domain:

   - A **Person** has attributes **pid** (key) and **name**.
   - A **Skier** is a type of **Person** with attribute **aptitude**.
   - A **Snowboarder** is a type of **Skier**.
   - A **PairOfSkis** has attribute **sid** (key) and **model**.
   - A **Snowboard** has attribute **sid** (key) and **model**.
   - A Skier **owns** zero or more PairOfSkis. The ownership relation has a **purchase_price**. A PairOfSkis is owned by at most one Skier.
   - A Snowboarder **owns** zero or more Snowboards. The ownership relation has a **purchase_price**. A Snowboard is owned by at most one Snowboarder.
   - a Person can **rent** a PairOfSkis or a Snowboard. A person cannot rent more than one PairOfSki or one Snowboard at the same time. A person cannot rent a PairOfSki and a Snowboard at the same time either. A piece of equipment can be rented by at most one person at a time. The rental comes with a **start_date** and an **end_date**.

   Answer (Draw an E/R Diagram):

(b) (10 points) Write the SQL `CREATE TABLE` statement for the **owns** relation between Skier and PairOfSkis. Make sure that your statement specifies the PRIMARY KEY and any FOREIGN KEYS. Additionally, we would like to enforce the constraint that `purchase_price` be greater than zero.

Answer (Write a CREATE TABLE statement):

2 Conceptual Design

2. (15 points)

   (a) (5 points) Consider the following relational schema and set of functional dependencies. List all superkey(s) for this relation. Which of these superkeys form a key (i.e., a minimal superkey) for this relation? Justify your answer in terms of functional dependencies and closures.

   R(A,B,C,D,E) with functional dependencies AB $\rightarrow$ E and D $\rightarrow$ C.

   Answer (Find all the superkeys and keys):

(b) (10 points) Decompose R into BCNF. Show your work for partial credit. Your answer should consist of a list of table names and attributes and an indication of the keys in each table (underlined attributes).

Answer (Decompose R into BCNF):

3 SQL and Views

3. (10 points)

   (a) (5 points) Discuss **two** scenarios where a view can be helpful in a database management system.

   Answer (Discuss two benefits of views):

(b) (5 points) Explain why it is not always possible to perform SQL UPDATE/DELETE/INSERT statements on top of a view.

Answer (Explain why updating through a view is not always possible):

4 Transactions

4. (20 points)

Consider a database consisting of a single relation R:

R:

A	B
1	10
2	0

(a) (10 points) The following two transactions run concurrently on this database:

Line	T1	T2
1	begin transaction;	begin transaction;
2	update R set B = B-10 where A=1;	select sum(B) from R;
3	update R set B = B+10 where A=2;	commit;
4	commit;	

Is it ever possible for T2 to see a value of zero in its output? Explain why or why not.

Answer (Discuss if a schedule where T2 sees a zero value is possible):

(b) (10 points) The following two transactions run concurrently on this database:

Line	T1	T2
1	begin transaction;	begin transaction;
2	insert into R values (3,150)	select sum(B) from R;
3	commit	select sum(B) from R;
4		commit;

Is it ever possible for T2 to see a different value as the output of the `select sum(B) from R` statements?

Answer (Discuss if a schedule where T2 sees different values is possible):

5 Parallel Data Processing

5. (20 points)

   (a) (10 points) We have two very large tables $R(A, B)$ and $S(A, C)$ where all attributes are integers. The data in each table is *randomly* distributed across three servers N_1, N_2, and N_3. Explain how a parallel DBMS can compute $R \bowtie_{R.A=S.A} S$ in parallel using all three servers.

   Answer (Explain parallel join):

(b) (10 points) List **two features** that are different between a parallel relational DBMS and the Pig system that you used in HW6.

6 DBMS as a Service

6. (10 points)

   (a) (5 points) Today, instead or purchasing physical machines, an alternate approach to running a database management system is to run it in the cloud. One of the key promises of the cloud is the illusion of infinite resources and the ability for users to elastically grow and shrink the resource consumption of their DBMS.

   Why can it be difficult to scale a relational DBMS?

   Answer (Discuss the challenge of scaling a relational DBMS):

(b) (5 points) Indicate one approach that the so-called "NoSQL" systems take to overcome the above challenge.

Answer (Indicate a feature of NoSQL systems designed to facilitate scalability):

7 Data Integration and Data Cleaning

7. (5 points)

   (a) (5 points) Company A recently bought company B. The two companies want to integrate their databases. Indicate **three** challenges that the integration will face. For each challenge give a concrete example.

   Answer (Indicate three challenges of data integration *with concrete examples*:)

That's it! Happy Holidays!