CSCI585 Spring '18 Midterm Exam

March 9th, 2018

CLOSED book and notes. No electronic devices. DO YOUR OWN WORK. Duration: 1 hour. If you are discovered to have cheated in any manner, you will get a 0 and be reported to SJACS. If you continue working on the exam after time is up you will get a 0.

Signature: _____

Problem Set	Number of Points
Q1	5
Q2	5
Q3	5
Q4	5
Q5	5
Q6	5
Q7	5
Total	**35**

Q1. (5 points total) ER MODELING

Design ERD using Crow's foot notation for the following problem:

Computer Science department needs to design a database to manage computer labs using the following information:

- Each lab has one unique identifier, name, administrator name, and many computers.
- Each computer has a unique identifier, configuration information (CPU, RAM, hard drive) and location (in one of the labs).
- Each class has a unique identifier, class name, and instructor's name.
- Each class can have lab hours in multiple labs and one lab can be registered for multiple classes. A timestamp is stored to indicate a class is registered for a lab session.

Q2. (5 points total) SQL
After the Oscars award ceremony last Sunday, you have been contacted by the organizers to write some queries. Their database consists of the following tables:
MEMBERS (MEMBER_ID, NAME).
MOVIES (MOVIE_ID, RELEASE_YEAR, TITLE, DIRECTOR).
REVIEWS (REVIEW_ID, *MEMBER_ID*, *MOVIE_ID*, TEXT, REVIEW_DATE, RATE).
ACTORS (NAME, *MOVIE_ID*).
Primary keys of every table are underlined while foreign keys are italic. The RELEASE_YEAR attribute of a movie is a number, such as 2018.

A (2 points) Display unique member IDs of all the members who reviewed at least one of the
movies reviewed by user with member ID "M1". The list of member IDs must exclude "M1".

B (1 point) Delete all reviews that have the term "horrible" in their text. If the text contains "XhorribleX" where X refers to any character(s), its review must be deleted as well.

C (2 points) Display the actors' names and average rating for the movies with the highest average rating.

Q3. (5 points total) NORMALIZATION

Convert the following table into:

a. The 1NF. (1 point)
b. The 2NF. (2 points)
c. The 3NF. (2 points)

Show the dependency diagram for each form and identify the primary key for each table.

Parent_ID	Parent_Name	Home_Address	Children_Names	Enrollment	Start_Hour	End_Hour	Daycare_ID	Daycare_Location
1	Alice	627 Green St., LA	Mike, Sara	Full	7am	5pm	324	1214 Hover St., LA
2	Brad	93 27th St., LA	Liam	Morning	7am	12pm	324	1214 Hover St., LA
2	Brad	93 27th St., LA	Nina	Full	7am	5pm	324	1214 Hover St., LA
3	Claire	45 Pico Blvd., LA	Luke	Full	7am	5pm	324	1214 Hover St., LA
4	Tom	1308 55th Pl., SD	Sara	Afternoon	1pm	5pm	564	453 5th Ave., SD
5	Alice	433 Maple St., SD	Tony, Yara	Full	7am	5pm	564	453 5th Ave., SD

Q4. (5 points) TRANSACTION MANAGEMENT

A (3 points) What does ACID in ACID properties stand for? Give an example of a scenario where atomicity is violated.

B (2 points) What is two-phase locking (2PL)? Give an example to illustrate how deadlock may happen with two phase locking.

Q5. (5 points) QUERY OPTIMIZATION

Consider the three following tables for an online-sale database and all attributes are neither indexed nor sorted.

1. CUSTOMER (cid, name, age), cid is the primary key.
2. PRODUCT(pid, seller), pid is the primary key.
3. TRANSACTION(tid, cid, pid), tid is the primary key.

And we want to execute the following SQL query:

```
SELECT T.tid, C.name
FROM TRANSACTION T, CUSTOMER C, PRODUCT P
WHERE C.cid = T.cid
AND P.pid = T.pid
AND seller = 'Olivera'
AND C.age >= 25
AND C.age <= 34
```

Assuming:

- There are 100 rows in CUSTOMER, 5,000 rows in PRODUCT and 10,000 rows in TRANSACTION.
- There are 100 different sellers equally distributed in PRODUCT.
- Customers's ages range from 20 to 44 (both inclusive) equally distributed in CUSTOMER.
- cid and pid are independently equally distributed in TRANSACTION.

Now our task is to optimize the query with a Cost-based optimizer. **Suppose the cost of running a SELECT operation is the number of rows in the source table** and **the cost of running a JOIN operation is the total rows of the two source tables**. If we execute the query with following access plan, the cost will be 5,050,015,100.

STEP	OPERATION	COST	ESTIMATED RESULT ROWS
A1	Join T and C	15,000	50 milliion
A2	Join A1 and P	50,000,100	5 billion
A3	Select rows in A2 with all conditions	5 billion	40 (Explained below)

The possibility of C.cid = T.cid is 1/100 for there are 100 different cid. The possibility of P.pid = T.pid is 1/5000 for there are 5000 different pid. The possiblity of seller = 'Olivera' is 1/100 for there are 100 different sellers. The posibility of C.age >=25 and C.age <= 34 is 10/25. Since all conditions are independent, the number of result rows in A3 is about 5 billion/100/5000/100*(10/25)=40.

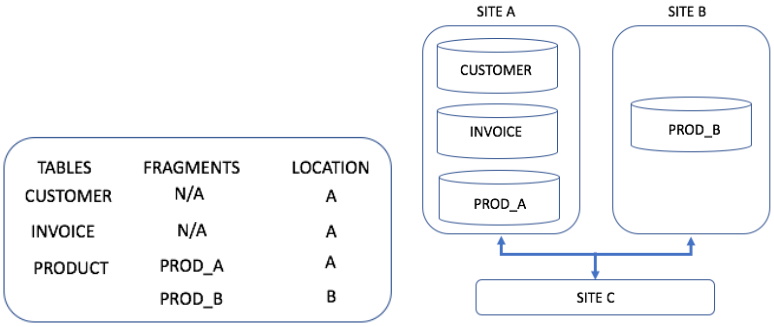
T, C and P are abbreviations for TRANSACTION, CUSTOMER and PRODUCT, respectively.

Do you have a better access plan to execute the query with a lower total cost? Please fill the following form (on the next page!) about your access plan with STEP 1 given.

- You don't have to fill all rows depending on how many steps in your access plan.
- Try not to ruin this form. There should be enough room in each cell for you to answer and make corrections.

STEP	OPERATION	COST	ESTIMATED RESULT ROWS
B1	Select rows in C with ages between 25 and 34	100	40
B2			
B3			
B4			
B5			
B6			
B7			

Q6. (5 points) DISTRIBUTED DATABASES



TABLES	FRAGMENTS	LOCATION
CUSTOMER	N/A	A
INVOICE	N/A	A
PRODUCT	PROD_A	A
	PROD_B	B

For the DDBMS above, specify the type of operation the database must support (remote request, remote transaction, distributed transaction or distributed request) to perform each of the following operations at SITE C:

a.
```
SELECT   *
FROM     PRODUCT
WHERE    PROD_QOH > 20;
```

b.
```
SELECT   CUS_NAME, INV_TOTAL
FROM     CUSTOMER, INVOICE
WHERE    CUSTOMER.CUS_NUM = INVOICE.CUS_NUM;
```

c.
```
BEGIN WORK;
UPDATE   PRODUCT
SET              PROD_QOH = PROD_QOH + 5
WHERE    PROD_NUM = '123';
INSERT   INTO CUSTOMER(CUS_NUM, CUS_NAME, CUS_STATE)
         VALUES('111', 'Tommy Trojan', 'CA' );
COMMIT WORK;
```

Q7. (5 points) DB SECURITY, WEB TECHNOLOGIES, BUSINESS INTELLIGENCE

A (1 point) Contrasting between activities of a "database administrator" (DBA) and a "data administrator" (DA), who sets policies and standards?

B Which Web technology has a class named DataSet?

C (1 point) Name the components of Star schema.

D. (1 point) Is snowflake schema normalized or denormalized?

E. (1 point) Name the two extensions SQL offers for OLAP.