CSE 530 – Midterm Exam

Name:

(Print ***CLEARLY***)

Question	Points Possible	Points Earned
1	22	
2	20	
3	20	
4	23	
Total	85	

Question 1 – Transactions and Locking

You are given the following schedule:

R2(A)
R2(Y)
R3(Z)
W2(Y)
R1(A)
W3(Z)
W2(A)
W1(Z)
W1(A)

(5 points) Prove whether or not this schedule is conflict serializable.

(5 points) If the above schedule is serializable, determine an equivalent serial schedule and prove that they are equal. If the above schedule is not serializable, use the same transactions to create a non-serial schedule that is serializable schedule and prove that it is serializable.

(5 points) Will basic two phase locking lead to better or worse performance than strict two phase locking (in general)? Explain how you know.

(5 points) There are two categories of deadlock resolution: deadlock prevention, and deadlock detection. Under what circumstances is it preferred to use deadlock prevention over deadlock detection? Explain how you know.

(2 points) We use transactions to solve certain types of problems that arise with concurrent operations. Name the problem that exists in the following schedule:

R1(X), R1(Y), R2(X), W2(X), W1(Y), R2(Y), R1(X)

Question 2 – Data Warehousing

We are interested in building a data warehouse to track survey results. For the sake of this problem, you should assume that all survey questions will have numerical answers. You should assume that we want to track many surveys, and that each survey will have multiple questions.

Some questions that we might want to answer using this data warehouse include:

How many people took each survey?

What was the average response for each question?

On which date did we receive the most responses?

Using this information, answer the following questions:

(5 points) What dimensions should exist in this data warehouse?

(2 points) What measures should exist in this data warehouse?

(3 points) In class we spent time identifying hierarchies in our dimensions. Some examples of such hierarchies include time (Date -> Month -> Year) and location (Zip code -> City -> State -> Country). Give an example of a hierarchy that exists in this database – you cannot use date or location as an answer here, those have already been given.

(5 points) List all of the columns that would exist in the fact table.

(5 points) List two differences between OLTP and OLAP database design. You should answer this question generally, it is not based on the data warehouse from the previous problems.

Question 3 – Distributed Databases

A company has an existing database that it uses to track product shipments. We wish to distribute the database in the following way:

- There are two shipping centers. Shipping center one is responsible for all video game shipments, while shipping center two is responsible for all computer shipments. We wish to place the data for these types of products (and only these types of products) at each site. We wish to keep the information for all products at a third site, the company headquarters.
- Each product has information about how much our company paid for it as well as how much we wish to sell it for. This pricing information is needed at the company headquarters, but is not needed at the shipping centers.
- We also must keep track of what vendor each product was purchased from. This vendor information is kept only at the shipping centers, not at company headquarters.

(5 points) What type(s) of fragmentation are being used? Explain how you know.

(5 points) An alternative form of distribution could simply replicate the entire database at each of the three sites. List one advantage and one disadvantage of using this implementation over the fragmented database.

(10 points) Company headquarters suspects that a vendor is overcharging them, so they wish to review all shipments from that vendor. This vendor sells video games as well as computers.

The following query is submitted to the fragmented database at company headquarters. We wish to compute the results of this query and then make sure that the results end up back at company headquarters.

$$\pi_{VendorName,ProductName,Price}(VENDOR \bowtie_{vendor.vendorid=products.vendorid} PRODUCTS)$$

You are given the following column sizes:

VendorName (vendors table): 10 bytes

ProductName (products table): 10 bytes

Price (products table): 4 bytes

VendorID (vendors table): 4 bytes

VendorID (products table): 4 bytes

The Products table at shipping center one contains 100 rows, the Products table at shipping center 2 contains 150 rows, and the Vendors table contains 50 rows.

What is the minimum number of bytes that must be sent to complete this query?

Question 5 – MongoDB

We wish to construct a MongoDB database that tracks what quizzes have been given. A quiz has a course number and a quiz number associated with it. A quiz can contain multiple choice, fill in the blank, and true false questions.

(10 points) Construct an example of a quiz document. This document should contain all of the information about one quiz, including the questions. Make sure to include one question of each type in your example. You should not use any references in your document. You do not need to include any actual data in your document, just the structure will suffice.

(5 points) Write a mongo DB query that will retrieve all quizzes for CSE530. You can assume that the quiz documents are in a collection called "quizzes" in a database called "db."

(5 points) In the previous problems you were asked not to use any references in your document design, however it is possible and sometimes desirable to use references in a MongoDB document. Explain when using references might be a better implementation choice.

(3 points) In class we discussed how NoSQL databases are often schemaless, however there is often an implied schema. Explain what is meant by an implied schema.