

# CSE 530 – Midterm Exam

---

Name:

(Print CLEARLY)

---

Question	Points Possible	Points Earned
1	25	
2	15	
3	20	
4	25	
Total	85	

## Question 1 – Transactions and Locking

(6 points) Transactions are said to abide by four properties called the ACID properties. Some of these properties can be manipulated by programmers/administrators while other properties are tightly controlled by the system. Name at least one ACID property that can be modified by a database programmer/administrator and explain how it can be modified.

Atomicity

(6 points) In class we used a precedence graph to determine whether schedules were serializable or not. Give at least two reasons why this technique is not used in typical database implementations.

1. if the number of transactions grows up to very large, it is very impractical considering efficiency.
2. our goal is not to check whether schedules are serializable or not but to generate ones, thus we usually use 2PL to create.

(7 points) Does strict two-phase locking prevent deadlock? If yes, explain how. If no, provide an example of a schedule that obeys two-phase locking but also creates deadlock.

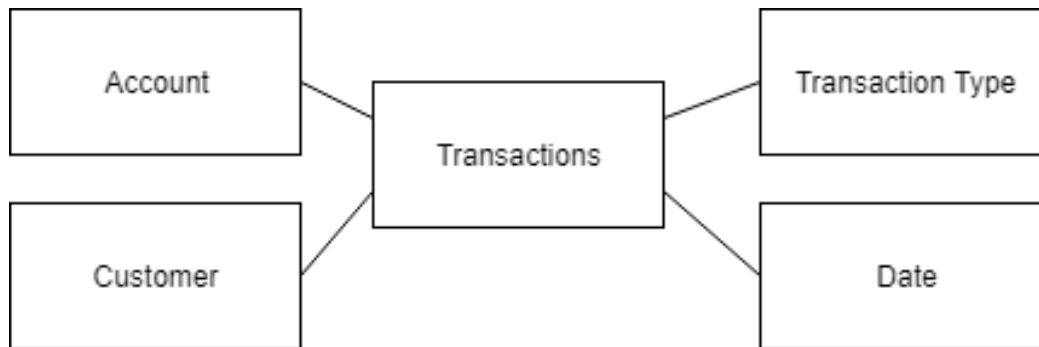
No.

(6 points) What is an intention lock? Name at least one way that intention locks can lead to increased performance.

In terms of multiple granularity locking, intention locks are table-level locks that indicate which type of lock (share or exclusive) a transaction requires later for a row in table. There are three types of intention locks: intention shared lock, intention exclusive locks, shared intention exclusive. Intention lock is useful when one wants to update all records in a file (file), the other wants to update a single record (row) with only two types of locks (read/write), we need to check all further down locks, which is painful. with intention locks, we don't need to check all and we might just check whether there is an intention lock. Clearly, this will lead to a better performance

## Question 2 – Data Warehousing

You are given the following star schema:



(3 points) What is the granularity of the above data warehouse?

granularity is decided by dimension, thus the answer is  
Account, Transaction Type, Date and Customer

(2 points) Name a measure that you would expect to find in this data warehouse.

Transaction Amount

(10 points) Starting with a base cube that contains all of the dimensions (Account, Transaction Type, Customer, Date) state which cube operations would be necessary to examine all deposit transactions that took place in Q3 of 2017 for customers that have two or more cars in the state of Missouri.

### Question 3 – Distributed Databases

(7 points) Explain at least two differences between how transactions are handled in a distributed database vs. a centralized database.

1. for distributed database, we don't consider ACID properties usually, which is checked for centralized database
2. for distributed database, global transaction manager is in charge of coordinator

(7 points) What is the difference between reliability and availability? How does distribution affect the reliability and availability of a system?

reliability: probability that a system is running  
, measured by mean time between failures, it's expected  
and more speculative than availability; e.g expect at least  
once failure every five years

availability: probability that the system is continuously  
available, measured as percentage and it's predicted;  
replication increase the availability; e.g for 10 years, how  
often it crashes

Distribution can increase availability by increasing the  
number of replications

(6 points) Under what circumstances would it be preferable to use a distributed catalog as opposed to a centralized catalog?

1. If the centralized catalog goes down, my database is unavailable (if I want to maximize my availability)
2. If nodes of network are very close to each other
3. If we are not changing the catalog very often

## Question 4 – MongoDB

(25 points) In class we discussed how MongoDB allows us to have much more flexibility with our database design. We also discussed how the design choices we make involve tradeoffs.

For this problem, you need to design a MongoDB database for a simple social network. This social network allows users to create posts and also allows users to like posts that have been made. Your designs should include user information (username, email address), post information (title, content) as well as a way of tracking what user made each post and what users have liked each post.

You must come up with two different designs. Once you have created your database designs, discuss the tradeoffs that exist between them.

Design 1:

Design 2:

What tradeoffs exist between the two designs that you have created? Name at least two advantages and disadvantages of each design.