UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Engineering, Built Environment and IT
Department of Computer Science

Database Systems
COS 221

Examination Opportunity 3(EO3)

23 June 2022 (12:30 to 14:30)

Examiners

Internal: Dr Linda Marshall and Mr Andrew Broekman
External: Avinash Singh

Instructions

1. Read the question paper carefully and answer all the questions.
2. The assessment opportunity comprises of **4** questions on **4** pages. Refer to the table below for the marks per question.
3. **2** hours have been allocated for you to complete this paper. You will download the paper and can write it offline. You need to photograph (make sure your photographs are clear) or scan all your pages and create a **single PDF** (with .pdf extension) of your answers.
4. An upload slot will be open on the module ClickUP page under the **Examination Opportunities** menu option for the duration of the examination opportunity (12:30 to 14:30) and then for an additional 30 min to give enough time to download this paper, create the PDF containing your answers and then upload your PDF. **No late submissions will be accepted. Papers uploaded in the incorrect format will not be accepted either.**
5. Make sure you include your student number, name and a picture of your student card on the first page of your submission. Leave some space for your marks to be written on the first page. Each subsequent page of your submission must include at least your student number and name at the top of the page.
6. This paper is **take home** and is subject to the University of Pretoria Integrity statement provided below.
    - You are allowed to consult any literature.
    - You are not allowed to discuss the questions with anyone.
    - You may not copy from online resources. All answers must be in your own words.
7. If you have any queries when writing the paper, post them in good time to the COS 221 WhatsApp group or use the chat functionality on the Blackboard Collaborate session for EO3. Make sure your post in general enough as not to give away any answers. These two channels will be used to deliver any general information during the Examination Opportunity.

Integrity statement:
The University of Pretoria commits itself to produce academic work of integrity. I affirm that I am aware of and have read the Rules and Policies of the University, more specifically the Disciplinary Procedure and the Tests and Examinations Rules, which prohibit any unethical, dishonest or improper conduct during tests, assignments, examinations and/or any other forms of assessment. I am aware that no student or any other person may assist or attempt to assist another student, or obtain help, or attempt to obtain help from another student or any other person during tests, assessments, assignments, examinations and/or any other forms of assessment.

Question:	1	2	3	4	Total
Marks:	12	10	36	36	94

Full marks is: 85

Short Questions

1. For each of the statements given below, state whether the statement is *True* or *False*.
   (a) If T_1: `read_item(A)` and T_2: `read_item(A)` then the order of T_1 and T_2 does not matter. (1)
   (b) Two-phase locking protocol ensures conflict serialisability and safety from deadlocks. (1)
   (c) Using checkpoints reduces overhead. (1)
   (d) Active database applications augment Relational database systems with events, conditions and actions. (1)
   (e) Valid and Transaction time are two time dimensions in temporal database systems. (1)
   (f) Relational database systems can deal with all aspects required for spatial information storage and retrieval. (1)
   (g) Content identification in multimedia databases is a manual process. (1)
   (h) Search engines are an example of information retrieval systems and make use of crawlers to glean information from the web. (1)
   (i) When indexing information that has been retrieved from documents, it is necessary to invert the index. (1)
   (j) A Read-Write lock combination is not compatible. (1)
   (k) "No", is the correct entry in the table for a Write lock with a Write lock combination. (1)
   (l) The more strict the concurrency control, the less concurrent the transactions are. (1)
2. Complete the following sentences by filling in the missing words.
   (a) A strict schedule is also ___(i)___ which in turn is ___(ii)___. (2)
   (b) The ___(iii)___ data model of which the Relational model is an example of, is too restrictive. (1)
   (c) Another term for horizontal partitioning is ___(iv)___. (1)
   (d) Provide three (3) categories of NO SQL databases, ___(v)___, ___(vi)___ and ___(vii)___. (3)
   (e) The three desirable properties of distributed systems with replicated data are: ___(viii)___, ___(ix)___ and ___(x)___. (3)

Long Questions

3. Consider the relational schema comprising of the relations `Guitars`, `Players` and `LastPlayed` below and answer the questions that follow.

```
Guitar(gId, Brand, Price)
Player(pId, Name, Age)
LastPlayed(gId, pId, Date)
```

   (a) Identify the functional dependencies for each of the relations in the given schema. (3)
   (b) For each relation in the schema, identify the primary key(s). (3)
   (c) What are the relational integrity constraints for the given schema? (3)
   (d) Draw an ER diagram for the schema. Assume `Date` can be further broken down into `Year`, `Month` and `Day`. (4)
   (e) Write an SQL query to retrieve dates and names of players below the age of 25 who last played Gibson brand guitars that cost more than 3000. (3)
   (f) Translate your SQL query to a Relational Algebra (RA) query. Use Natural joins when joining the tables. (3)
   (g) Draw the initial query tree for the RA query given as an answer in Question 3(f) above. (3)
   (h) i. What is a *join selection factor*? (2)
   ii. Using the idea behind *join selection factor*, optimise your initial query tree. (3)
   iii. Translate your optimised query tree to an optimised RA query. (3)
   (i) i. How does *nested-loop join* differ from *sort-merge join*? (2)

ii. Write a simple algorithm for a *nested-loop join*. (4)

4. Consider the three transactions T_1, T_2, and T_3 given below and answer the questions that follow.

T_1: $r_1(X)$; $r_1(Z)$; $w_1(X)$;
T_2: $r_2(Z)$; $r_2(Y)$; $w_2(Z)$; $w_2(Y)$;
T_3: $r_3(X)$; $r_3(Y)$; $w_3(Y)$;

(a) For each of the following schedules, draw the precedence graph and determine whether the schedule is conflict serialisable. If the schedule is conflict serialisable, write the conflict equivalent serial schedule of transactions (e.g., T_3; T_2; T_1). If the schedule is not conflict serialisable, identify a set of conflicting actions that prevents it from being conflict serialisable.

i. S_1: $r_1(X)$; $r_2(Z)$; $r_1(Z)$; $r_3(X)$; $r_3(Y)$; $w_1(X)$; $w_3(Y)$; $r_2(Y)$; $w_2(Z)$; $w_2(Y)$; (3)

ii. S_2: $r_1(X)$; $r_2(Z)$; $r_3(X)$; $r_1(Z)$; $r_2(Y)$; $r_3(Y)$; $w_1(X)$; $w_2(Z)$; $w_3(Y)$; $w_2(Y)$; (3)

(b) Given schedule S_3:

S_3: $r_3(X)$; $r_1(X)$; $r_2(Z)$; $r_1(Z)$; $r_2(Y)$; $w_2(Z)$; $r_3(Y)$; $w_2(Y)$; $w_1(X)$; $w_3(Y)$;

Assume locks have been applied to S_3 as follows:

T_1	T_2	T_3
		read_lock(X)
read_lock(X)		read_item(X)
read_item(X)		
	write_lock(Z)	
	read_itemi(Z)	
read_lock(Z)		
read_item(Z)		
unlock(Z)	write_lock(Y)	
	read_item(Y)	
	write_item(Z)	read_lock(Y)
		read_item(Y)
	unlock(Z)	
	write_item(Y)	
	unlock(Y)	
write_lock(X)		
write_item(X)		unlock(X)
unlock(X)		write_lock(Y)
		write_item(Y)
		unlock(Y)

i. Which transaction follows the Two Phase Locking (2PL) protocol? (2)

ii. Apply the 2PL protocol to one of the other transactions in the schedule. (2)

iii. Which transaction makes use of lock conversion? (1)

iv. Provide an example, using the given transactions, where the transactions can result in deadlock. Explain why you say the transactions are in deadlock. (3)

v. What benefit does strict 2PL provide? Explain the drawback to this benefit. (2)

vi. How does 2PL differ from *multiple granularity locking*? Use relevant examples. (6)

(c) i. How does the recovery manager ensure atomicity and durability of transactions? (2)

ii. What are the roles of the Analysis, Redo, and Undo phases in ARIES recovery algorithm? (3)

(d) Consider the execution, with a crash, given in the table below.

LSN	ARIES Record
10	BEGIN CHECKPOINT
20	END CHECKPOINT
30	update T_1, P5
40	update, T_2, P3
50	COMMIT T_2
60	update, T_3, P3
70	ABORT, T_1
80	CRASH, RESTART

The update entries are of the form `update, TransactionID, PageID`. You may assume the entries in the Transaction Table are:
(T_1, 8, in progress) and (T_4, 3, commit)
The Dirty Page Table entries are:
(P_3, 9) and (P_5, 8).
Answer the questions that follow. Be precise about the points at which every phase begins and ends.

i. Provide the contents of the tables during the Analysis phase of the ARIES recovery algorithm. (3)

ii. What is done during Redo? (3)

iii. What is done during Undo? (3)