

UNIVERSITY OF BOLTON
CREATIVE TECHNOLOGIES
COMPUTING GROUP

SEMESTER TWO EXAMINATION 2018/2019

ADVANCED DATABASE SYSTEMS

MODULE NO: CPU6007

Date: Tuesday 21st May 2019

Time: 14:00 – 16:00

INSTRUCTIONS TO CANDIDATES:

There are **SIX** questions.

Answer **FOUR** questions.

All questions carry equal marks.

Marks for parts of questions are shown in brackets.

All rough work must be written in your answer book.

Creative Technologies
 Computing Group
 Semester Two Examination 2018/2019
 Advanced Database Systems
 Module No. CPU6007

Question 1

This question deals with distributed database management systems

- a) Discuss three advantages of a distributed database management system (DDBMS) **(3 marks)**
- b) Compare and contrast homogeneous and heterogeneous distributed database management systems. **(4 marks)**
- c) Using the concepts of Selection and Projection explain the difference between horizontal and vertical fragmentation. **(4 marks)**
- d) Discuss four alternative strategies regarding placement of data relating to distributed databases. **(4 marks)**
- e) A company uses a distributed relational database to store product information, customer information and sales records. Specifically the following three relations are used:

Customer (customerId, fname, lname, city) 4000 records on server X

Product (productId, name, type, price) 750 records on server Y

Sale (customerId, productId) 2800 records on server X

To list the customers from Hamburg who bought televisions there are several possible strategies to execute the query.

Calculate the response times for the following strategies:

- i. Move the product relation to server X and process the query there. **(2 marks)**
- ii. Join the Customer and Sale relations on server X, select tuples where the city is Hamburg, and move the result to server Y for matching with customers who bought televisions. **(3 marks)**
- iii. Join the Customer and Sale relations on server X, select tuples where the city is Hamburg, and for each one of these in turn check at server Y to determine whether they bought televisions. **(5 marks)**

ASSUMPTIONS

- Each tuple in each relation is 40 characters long
- Each tuple in a join result is 120 characters long
- There are 30 types of product available and
- There are 500 customers from Hamburg
- Computation time is negligible compared to communication time
- Data is transmitted at 1500 characters per second
- The message delay is 5 seconds

PLEASE TURN THE PAGE....

Creative Technologies
 Computing Group
 Semester Two Examination 2018/2019
 Advanced Database Systems
 Module No. CPU6007

Question 2

This question deals with transaction management and recovery.

- a) With reference to two of the four ACID properties explain why recoverability is required for database transactions. **(4 marks)**
- b) Explain the function of checkpoints in terms of buffers and secondary storage **(4 marks)**
- c) The following is a transaction log. For each of the transactions explain what recovery option is needed following the crash using Immediate Update and Deferred Update: **(10 Marks)**

```
[start_transaction, T1]
[write_item, T1, D, 0, 20]
[start_transaction, T3]
[commit, T1]
[write_item, T3, C, 14, 30]
[checkpoint]
[start_transaction, T2]
[write_item, T2, B, 0, 15]
[commit, T3]
[write_item, T2, A, 0, 20]
[commit, T2]
[start_transaction, T5]
[write_item, T5, B, 15, 12]
[start_transaction, T4]
[write_item, T4, A, 20, 30]
[write_item, T4, D, 20, 25]
→ → CRASH ← ←
```

- d) How does shadow paging differ from log-based recovery techniques? **(7 marks)**

PLEASE TURN THE PAGE....

Creative Technologies
Computing Group
Semester Two Examination 2018/2019
Advanced Database Systems
Module No. CPU6007

Question 3

This question deals with managing concurrency

- a) The **Lost Update** problem is one of the issues that needs to be addressed in controlling concurrency. Illustrate the Lost Update problem using two transactions as an example. **(6 marks)**
- b) Explain how 2-Phase locking ensures that transactions are serializable. **(6 marks)**
- c) With the aid of an example, explain the issue of deadlock as it affects database transactions. How this problem is normally resolved? **(4 marks)**
- d) When using timestamps, what would happen in each of the following scenarios?
- A read on an item which has been updated by a younger transaction
 - A write on an item which has been updated by a younger transaction
 - A write on an item which has been read by a younger transaction
- (3 marks)**
- e) Consider the timestamp protocol and three transactions T1, T2 and T3. Given the rules you have outlined in your answer to the previous question, what happens to each of these transactions:
- R1(Y) W2(X) W3(Y) R3(Z) W1(X) W2(Z)
- Note: R=Read, W=Write, 1, 2 or 3 represents the transaction and X, Y and Z are Locks
- (6 marks)**

PLEASE TURN THE PAGE...

Question 4

This question relates to Data Warehousing and Data Mining

- a) Design a data warehouse structure for a national car dealer chain to provide business decision makers with the important data they need. The company records information about its dealerships, customers and cars sold in its database. For each dealership it records location and manufacturer. For each car it records make, model and total cost. For customer it records name and address. For each sale the dealership, date, salesman and car are recorded. Using the four-step dimensional modelling process design a star schema for the dealership **(9 marks)**
- b) In data mining terms, explain how Classification and Value Prediction are used as Predictive Modelling methods. **(8 marks)**
- c) Identify two methods used for Descriptive Modelling and explain how that can be used to derive useful information from a data set. **(8 marks)**

Question 5

This question relates to Online Analytical Processing (OLAP)

- a) What features of OLAP would support a national supermarkets chain business Analytics **(6 marks)**
- b) Analyse the statement, "OLAP is just an extended set of grouping functions". **(6 marks)**
- c) Describe the architecture, characteristics, and issues associated with each of the following categories of OLAP tools:
- i MOLAP
 - ii ROLAP
 - iii HOLAP
 - iv DOLAP **(8 marks)**
- d) Discuss the relationship between data warehousing and OLAP. **(5 marks)**

PLEASE TURN THE PAGE...

Creative Technologies
Computing Group
Semester Two Examination 2018/2019
Advanced Database Systems
Module No. CPU6007

Question 6

This question relates to Object Oriented Database Management Systems

- a) Identify the operational reasons for implementing a OODBMS as opposed to a RDBMS. **(6 marks)**
- b) Explain the concept of Pointer Swizzling and its roll in achieving acceptable performance in object DDMS's **(6 marks)**
- c) Discuss three disadvantages and three advantages of an object oriented database system (OODBMS) **(6 marks)**
- d) The Object-Oriented Database Manifesto documents thirteen mandatory features for an OODMS. Identify seven of these features and analyse their impact on Database Design **(7 marks)**

END OF QUESTIONS