KOLEJ UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

ACADEMIC YEAR 2020/2021

APRIL/MAY EXAMINATION

COMPUTER SCIENCE BACS3183
ADVANCED DATABASE MANAGEMENT

FRIDAY, 7 MAY 2021 TIME: 9.00 AM – 12.00 NOON (3 HOURS)

BACHELOR OF COMPUTER SCIENCE (HONOURS) IN DATA SCIENCE
BACHELOR OF COMPUTER SCIENCE (HONOURS) IN SOFTWARE ENGINEERING

BACHELOR OF INFORMATION TECHNOLOGY (HONOURS) IN INTERNET TECHNOLOGY
BACHELOR OF INFORMATION TECHNOLOGY (HONOURS) IN SOFTWARE SYSTEMS DEVELOPMENT

Instructions to Candidates:

Answer **ALL** questions in the requested format or template provided.

- This is an open book final online assessment. You MUST answer the assessment questions on your own without any assistance from other persons.
- You must submit your answers within the following time frame allowed for this online assessment:
  - The deadline for the submission of your answers is **half an hour** from the end time of this online assessment.
- Penalty as below **WILL BE IMPOSED** on students who submit their answers late as follows:
  - The final marks of this online assessment will be reduced by 10 marks for answer scripts that are submitted within 30 minutes after the deadline for the submission of answers for this online assessment.
  - The final marks of this online assessment will be downgraded to zero (0) mark for any answer scripts that are submitted after one hour from the end time of this online assessment.
- Extenuation Mitigating Circumstance (EMC) encountered, if any, must be submitted to the Faculty/Branch/Centre within 48 hours after the date of this online assessment. All EMC applications must be supported with valid reasons and evidence. The UC EMC Guidelines apply.

FOCS Additional Instructions to Candidates:

- Include your **FULL NAME**, **STUDENT ID** and **PROGRAMME OF STUDY** in your submission of answer.
- Read all the questions carefully and understand what you are being asked to answer.
- Marks are awarded for your own (original) analysis. Therefore, use the time and information to build well-constructed answers.

BACS3183 ADVANCED DATABASE MANAGEMENT

STUDENT'S DECLARATION OF ORIGINALITY

By submitting this online assessment, I declare that this submitted work is free from all forms of plagiarism and for all intents and purposes is my own properly derived work. I understand that I have to bear the consequences if I fail to do so.

Final Online Assessment Submission
Course Code:
Course Title:
Signature:
Name of Student:
Student ID:
Date:

BACS3183 ADVANCED DATABASE MANAGEMENT

Question 1

a) Consider the simplified relational database shown in Figure 1, which is used to store a company's IT staff details and project assignments.

StaffDetail

StaffID	SName	DOB	DivisionNo	DivName	DivSupervisorID
S101	Kim Tan	16-05-1992	1	Software Testing	S114
S102	Jan Ang	01-11-1994	2	Analysis & Design	S223
S114	John Smith	28-12-1988	1	Software Testing	S114
S200	Mohan	14-08-1996	2	Analysis & Design	S223
S223	Peter	12-06-1987	2	Analysis & Design	S223
S227	Sam Toh	22-03-1989	3	Support	S227

StaffProject

StaffID	ProjectNo	ProjectName	HoursAssigned	SName
S101	P8	Sales Analysis	10	Kim Tan
S114	P8	Sales Analysis	6	John Smith
S101	P10	Work Automation	12	Kim Tan
S102	P11	Payroll Generation	25	Jan Ang
S200	P11	Payroll Generation	18	Mohan

Figure 1: Staffs' Project Assignment

(i) In what normal form is each of the two tables shown in Figure 1? Support your answer by explaining the functional dependency(ies) that exist(s) in each table. Any assumptions, if necessary, should be stated clearly. (8 marks)

(ii) Discuss the data anomaly problem(s) if any, that can arise with each of the tables in Figure 1. (12 marks)

(iii) Derive the Third Normal Form tables from those shown in Figure 1. The primary and foreign keys (if any) should be clearly indicated. (6 marks)

b) Consider the tables in Figure 2 used by Neo Enterprise which sells electrical products to retailers.

Customer (CustCode, CustName, Address, Postcode, CustPhone, CreditLimit)
Product (ProdCode, ProdName, QuantityOnHand, CostPrice)
Invoice (InvoiceNo, CustCode*, InvoiceDate)
InvoiceLine (InvoiceNo*, LineNo, ProdCode*, Quantity, SellingPrice)

Figure 2: Product Sales

(i) Create a view that contains for each product its code, name, total units sold as well as total sales value during the first three months of 2021. The result should be sorted in descending order of the total units sold. Discuss how this view can enhance database security for the company. (10 marks)

(ii) Assume that the following is a frequently processed query:

```
Select CustCode, CustName, CustPhone
From Customer
Where CreditLimit between 5000 and 9000;
```

Question 1 b) (ii) (Continued)

Propose the kind of index which should be defined that can speed up the query. Your answer should include the proposed index key and whether it is a hash or B^+-tree index. Justify your answer.
Discuss the disadvantage(s), if any, of creating this index. (10 marks)

(iii) Between clustering and non-clustering index, discuss which is suitable for range searches. (4 marks)

[Total: 50 marks]

Question 2

a) Consider a non-serial schedule of two concurrent transactions T1 and T2, as shown in Figure 3. Assume that A and B have the initial values of 180 and 60 respectively.

Time	T1	T2
t_0	begin-trans	
t_1	read (A)	
t_2	read (B)	begin-trans
t_3		read (B)
t_4	B = A - B	
t_5		B= B * 2;
t_6	write (B)	
t_7	commit	write (B)
t_8		read (A)
t_9		A = A + 100
t_{10}		write (A)
t_{11}		commit

Figure 3: Concurrent Transactions

(i) Discuss the lost-update problem that occurs in the schedule shown in Figure 3. Explain, with the aid of a diagram how this problem can be avoided using the versioning approach. (14 marks)

(ii) The two-phase locking protocol (2PL) is another solution to the lost-update problem but it may cause deadlock of transactions. With the aid of a diagram, explain how deadlock can happen during the execution of T1 and T2 with 2PL protocol. (8 marks)

b) Differentiate clearly between the rollforward recovery and rollback recovery in conjunction with a transaction log file. Explain how the recovery procedure is different if the checkpoint facility is provided by the DBMS. (10 marks)

Question 2 (Continued)

c) QExpress is a company providing courier services throughout Malaysia. It has many offices throughout the peninsular Malaysia, Sabah and Sarawak. As the company has experienced tremendous increase of business activities especially in the past two years, the top management has decided to distribute its operations and staff details according to the states. They also have decided that staff payroll details will be processed by the headquarter office based in Petaling Jaya.

The centralized database has the following relational schema:

Branch (BranchNo, BAddress, BTelNo, BEmail, State)
Staff (StaffNo, SName, SAddress, STelNo, Gender, DOB, Position, Salary, BranchNo*)
Customer (CustNo, CName, CAddress, CTelNo, CEmail, BranchNo*)
Vehicle (VehRegNo, MaxLoad, BranchNo*)
CustOrder (OrderNo, CustNo*, OrderDate, DeliveryDate, DeliveryAddress, LoadWeight, LoadDesc)
VehicleAllocated (OrderNo*, VehRegNo*)

(i) It was suggested that snapshot replication be adopted for the Branch table. Give your opinion on this suggestion. (5 marks)

(ii) Discuss the data distribution strategy that is suitable for the Staff table. Justify your answer. (7 marks)

(iii) Distinguish between location transparency and replication transparency with appropriate scenarios based on the relational schema given above. (6 marks)

[Total: 50 marks]