GX1307

OLLSCOIL Na hÉIREANN
THE NATIONAL UNIVERSITY OF IRELAND

NATIONAL UNIVERSITY OF IRELAND, GALWAY.

SEMESTER II EXAMINATIONS 1998/1999

Third University Examination in Information Technology

CT325 DATABASE SYSTEMS

Prof. D Bell
Dr. G. Lyons
Mr. C. O' Riordan

Time allowed: **Three hours**

Answer 4 questions
Question 1 is compulsory
All questions carry equal marks

Q.1 a) Describe the main components of the relational model. Your answer should include a description of data structures, integrity contraints and operators.

b) Given the following relational schema and interpretation:

SLSREP:	Rep_No, Rep_Name, Addr, Status
CUSTOMER:	Cust_No, Cust_Name, Addr, Balance, Rep_No
ORDER_INX:	Ord_No, Cust_No, Date
ORDER_LINE:	Ord_No, Item_No, Ord_Price
ITEM:	Item_No, Item_Desc, Item_Price, Colour

(Keys are underlined)
The SLSREP relation stores information regarding the sales representatives - unique number, name, address and status. The CUSTOMER relation is used to maintain information on customers. The attribute Rep_No in CUSTOMER acts as a foreign key to the SLSREP relation. ORDER_INX stores information on orders placed by customers. The ORDER_LINE relation is used to maintain information on orders - the items and the price. The final relation, ITEM, stores information on items - number, description, price and colour.

Develop SQL queries to satisfy the following information needs:

i) List all customers (with balance greater than 100) who have ordered an item with a green colour.
ii) List the name of all sales representatives who represent the customer(s) who have the lowest balance.
iii) List all customers who placed an order on the same date as a customer with customer number 147.
iv) List the customer number and customer name of all customers who have placed an order with the highest price (Ord_price) containing less than 3 items.

c) Develop an operator tree representing a query to satisfy the information need:

"List all customers (with balance greater than 100) who have ordered an item with a green colour".

Outline briefly an algorithm for heuristic optimisation and apply this algorithm to the operator tree.

d) Explain, with examples, what is meant by the view update problem.

Q.2 a) Extendible, dynamic, and linear hashing are three techniques to allow hashing to a dynamically increasing file. Choose one technique and show how the hashing would proceed using the following key field values:

69, 60, 92, 71, 59, 21, 74, 15, 20, 28, 43
(Note: Assume bucket size of 2).

b) Describe, briefly, the structure of a B-Tree. Outline an insertion algorithm and show how the tree grows when inserting records with the following key values:

6, 4, 2, 8, 10, 12, 13, 15, 19

Outline the differences in structure and insertion algorithms for B trees and B+ trees.

c) Calculate the size of a B Tree (in blocks) and the number of block accesses required to access a specified record given:
40,000 records, average fill factor of 69%, Block size = 512 bytes,
Search field = 9 bytes, Record pointer = 7 bytes, Block pointer = 6 bytes

Would a B+ Tree be more or less efficient?

Q.3 a) With respect to transactions, the properties of atomicity, consistency, preservation, isolation, durability and serializability are usually desired. Explain the underlined terms.

Explain the term *conflict-serializability* and outline an algorithm to test a schedule for conflict serializability.

b) With respect to recovery, explain the main entries used in a system log. Outline recovery algorithms for a system operating under:
   i) immediate update protocol.
   ii) deferred update protocol.

c) Two phase-locking and timestamping are common approaches to ensuring concurrency. Outline *either* approach and present pseudo-code for the primitives used.

Apply either approach to the following schedule:
(If using 2 phase locking, assume shared and exclusive locks)

T1	**T2**	**T3**
	read(X)	
	read(Y)	
	write(X)	
		read(Y)
		read(Z)
read (Y)		
read(Z)		
	write(Y)	
		write(Y)
write(Z)		
		read(X)
		write(X)

Q.4 a) Describe the relational database design process. Your answer should include an overview of conceptual modelling, mapping to relational schema, choice of keys and indexing strategies.

b) Define first, second, third and Boyce Codd normal forms.

Given **R** = {A, B, C, D, E, F, G, H, I, J}together with functional dependencies:
{A, B} -> {C, D, E, F},
{B} -> {G}
{A} -> {H},
{G} -> {A},
{H} -> {I},
{I} -> {J}.

Normalise the relation **R** such that it satisfies BCNF.

c) Explain, and illustrate with examples, when denormalistaion is appropriate.

d) Outline the main factors involved when designing fragmentation, replication and allocation schemas for a distributed database.

Q.5 a) Write a short note comparing the relational, object-relational and object-oriented models.

b)

i) Outline a recovery mechanism that may used in a distributed database.

ii) The semi-join operator is often used to increase efficiency of queries in distributed systems. Explain the operation of the semi-join operator.

Given two sites S1 and S2 containing relations SLSREP and CUSTOMER respectively. Assume a query:
"List all customers (name and number) together with the name of their sales representative" is issued at another site, S3.

Given that SLSREP has 1000 tuples and CUSTOMER has 25,000 tuples, calculate the number of bytes transferred using the semi-join operator.

Assume the following attribute sizes (in bytes):

Rep_No:	8,	Rep_Name	20
Addr	100,	Status:	2
Cust_No	8,	Cust_Name	20
Balance	4,	Addr (cust)	100

c)

i) With respect to deductive databases, explain *backward-chaining*, *forward-chaining* and *rule-safety.*

ii) Traditional relational operators may be mapped into the form of Datalog rules. For example, given fact sets **R1**(A, B, C) and **R2**(D, E, F), **R1** intersect **R2** may be defined by:
intersect_R1_R2 (X, Y, Z) :- **R1**(X, Y, Z), **R2** (X, Y, Z).

Develop Datalog rules to implement:

i) Union
ii) Cartesian Product
iii) Equijoin (assume third attribute in R2 acts as foreign key to first attribute of R1).