

OLLSCOIL na hEIREANN
THE NATIONAL UNIVERSITY of IRELAND

NATIONAL UNIVERSITY OF IRELAND, GALWAY

SPRING EXAMINATIONS 2000

Third University Examination in Information Technology

CT332 DATABASE SYSTEMS II

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

Time allowed: **Three hours**

Answer 4 questions including **Question 1**
All questions carry equal marks

Q.1. i) Given the following company database schema:

EMPLOYEE:	<u>RSI</u> , Fname, Lname, Salary, Address, Age, Dno
DEPARTMENT:	<u>Dno</u> , Dname, Description
DEPT_LOCN:	<u>Dno</u> , Location
PROJECT:	<u>Pno</u> , Pname, Budget, Proj_Desc
WORKS_ON:	<u>Ssn</u> , <u>Pno</u> , Hours

provide SQL queries for the following:

- Given the name and salary of all employees working in a department named "Testing".
- List all employees (Fname, Lname) who earn more than 35,000 and have worked more than 30 hours on any project.
- Using the join method, list all employees who work in the same department as "Tom Smith" and who earn more than "Tom Smith".

- d) List the average number of hours worked by all employees who have worked on more than 4 projects.
- e) Of all employees who have worked on less than 3 projects, list the name(s) of the employee with the lowest salary.
- ii) Develop SQL code to provide a view *high_earners* that includes those employees with a salary over 35,000. Explain what is meant by the view update problem and explain how it may arise.
- iii) Integrity constraints are used to ensure the integrity of the data in a database. Describe the integrity constraints usually enforced by a DBMS. Discuss, briefly, how users may apply their own integrity constraints and discuss the associated benefits and disadvantages.

- Q.2.**
- i) Explain briefly the types of problems that may arise if concurrency control is not enforced.
 - ii) Two phase locking is a common approach to ensuring concurrency. Describe this protocol. Outline pseudo-code for the primitives involved: *read_lock*, *write_lock* and *unlock*. How might the *upgrade* and *downgrade* primitives be implemented?
 - iii) Given the following schedule, show how it would proceed under two phase locking (assuming shared and exclusive locks and the presence of upgrade and downgrade facilities).

T _a	T _b	T _c
read(Y)		
write(Y)		
	read(X)	
	write(X)	
		read(Z)
		write(Z)
read(X)		
write(X)		
	read(Z)	
		read(Y)
		write(Y)

- iv) Discuss how the two phase locking protocol could be extended to work for distributed databases.

Q.3 i) Explain the terms primary, clustering and secondary indexes

ii) Extendible, dynamic and linear hashing are three approaches to allow hashing to an dynamically expanding and contracting file.

Choose any of the three techniques and outline its operation. Illustrate how the file may contract and expand given deletions and insertions.

iii) Describe the operation of a B+Tree. Your answer should include a description of its structure and the operation of an insertion algorithm.

Given records with the following key values:

16, 18, 21, 23, 6, 8, 35, 40, 44, 48

show how a B+Tree would expand.

iv) Discuss the advantages of a B+Tree over a B Tree.

Q.4 i) Discuss the ACID properties of a transaction.

ii) Describe the main entries in a system log.

iii) Assuming the immediate update protocol, show how the recovery process would proceed given the following log:

```
[start_transaction, T1]
[read, A, T1]
[write, A, 10, 12, T1]
[start_transaction, T2]
[read, A, T2]
[write, A, 12, 13, T2]
[commit, T1]
[read, B, T2]
[write, 50, 60, T2]
[checkpoint]
[start_transaction, T3]
[read, C, T3]
[write, C, 45, 60, T3]
[read, D, T3]
[read, C, T2]
```

[read, D, T2]
[write, C, 60, 61, T2]
[write, D, 100, 0, T2]
[commit, T2]
<crash>

How would the log differ under a deferred update protocol ?

- iv) Describe how the two phase commit may be used in a distributed database to ensure the atomic nature of transactions.

Show that this approach works by considering the different types of failures that may occur, for example, failure of a participating site, failure of the coordinator, network failure, failure at commit point, during execution etc.

- Q.5** i) Parallel databases have been used to improve the efficiency of query processing. Explain, with examples, the following techniques:

- a) Interquery parallelism
- b) Intraquery parallelism
- c) Interoperation parallelism
- d) Intraoperation parallelism

- ii) Discuss how heuristic approaches are used in query optimisation. Outline an algorithm for this approach. Illustrate its operation on a query to satisfy the following need (based on schema in Q.1)

“List all employees (Fname, Lname) who earn more than 35,000, who work for the Testing department and have worked more than 30 hours on any project.”

- iii) In distributed databases, the amount of network traffic is often taken as a metric of the cost of the execution of a query. Describe, with an example, a technique to reduce the amount of data transferred in multi-site joins.