

OLLSCOIL NA hÉIREANN GAILLIMH
NATIONAL UNIVERSITY OF IRELAND GALWAY

SPRING/SUMMER EXAMINATIONS 2003

Third University Examination in Information Technology (3IF121)

Database Systems II (CT332)

Professor P. Nixon
Professor G. Lyons
Mr. C. O'Riordan

Time allowed: *three* hours.

Attempt 3 questions including Question 1

1. (a) The following is a fragment of a relational database schema representing a section of a company database. Keys are underlined.

EMPLOYEE (PPS, Fname, Lname, Address, Salary, Age, Dno)
DEPARTMENT (Dno, Dname, Location, Description)
PROJECT (Pno, Pname, Budget, Plocation)
WORKS_ON (PPS, Pno, Hours).

The EMPLOYEE relation is used to maintain information regarding employees in the company. The *PPS* acts as a primary key. The *Fname* and *Lname* attributes store information regarding a person's name. The *Address* attribute is used to store the employee's address. The *salary* and *age* attributes maintain the salary and age of the employee. The *Dno* attribute acts as a foreign key, related to the primary key *Dno* in DEPARTMENT.

The DEPARTMENT relation is used to store information on departments; the name, location and description are maintained.

The PROJECT relation is used to store information on projects: a project number which acts as a key, the name, the budget and the location.

The WORKS_ON relation represents a many-many relationship between the EMPLOYEE and PROJECT entities. A composite key {*PPS*, *Pno*} is used. *PPS* acts as a foreign key to the EMPLOYEE relation and the *Pno* acts as a foreign key to the PROJECT relation. The *hours* attribute is used to maintain a record of the number of hours an employee has worked on a particular project.

Develop SQL queries to satisfy the following information needs:

- (i) List all employees (first and last names) who work for a department named "Research".
- (ii) List all employees (first and last names) who earn more than 35000 and have worked more than 30 hours on a project located in "Galway".

- (iii) Using the join method, list all employees (first and last names) who work in the same department as "Tom Smith".
- (iv) Of all the employees who have worked on less than 3 projects, list the employee with the highest salary.

(10)

- (b) Outline the process of heuristic optimisation. For the information need:

List all employees who earn more than 35000 and have worked more than 30 hours on a project located in "Galway",

develop an operator tree that represents an efficient strategy for the above query. (12)

- (c) Describe algorithms, and their associated advantages or disadvantages, that may be used to evaluate a *join* (\bowtie) operator. (11)

- 2. (a) Explain, briefly, the importance of concurrency control in multi-user databases. (6)

- (b) Answer (i) or (ii)

- (i) Time-stamping has been used to enforce conflict serializability in schedules. Outline the time-stamping approach. Your answer should include pseudo-code for any primitives involved. Outline the advantages and disadvantages of the time-stamping approach.

- (ii) Two-phase locking is an approach to guarantee conflict serializability. Outline the two-phase locking approach. Your answer should include pseudo-code for any primitives involved. Outline the advantages and disadvantages of the two-phase locking protocol.

(11)

- (c) Show how the following schedule would proceed under either two-phase locking or time-stamping:

T1	T2	T3
read_item(X)		
read_item(Y)		
	read_item(Z)	
	read_item(Y)	
		read_item(X)
		read_item(Z)
write_item(X)		
write_item(Y)		
	write_item(Z)	
		write_item(X)

(8)

- (d) Explain the meaning of a commit point of a transaction. Discuss the recovery process for a system operating under the immediate update protocol. (8)

3. (a) Discuss the properties of a well designed relational schema. (6)

(b) Describe a procedure you might adopt to develop a relational schema from an EER diagram.

Your answer should include a description of how you would deal with entities (strong and weak), attributes (atomic, composite, multi-valued), relationships (binary (1:1, 1:N, N:M), n-ary), specialisations and categorisations.

Include examples, where appropriate, to illustrate your examples. (13)

(c) Explain how the process would differ if you were developing a schema suitable for an object-oriented database. (8)

(d) Given $R = \{A, B, C, D, E, F, G, H, I, J, K\}$ and the following functional dependencies:

- $\{A, B, C\} \rightarrow \{F, G, H\}$
- $\{D\} \rightarrow \{A\}$
- $\{A, B\} \rightarrow \{D\}$
- $\{C\} \rightarrow \{E\}$
- $\{H\} \rightarrow \{I\}$
- $\{I\} \rightarrow \{J, K\}$

Decompose R to a set of relations such that all relations satisfy BCNF. (6)

4. (a) Given relations:

EMPLOYEE (PPS, Fname, Lname, Address, Salary, Age, Dno)
DEPARTMENT (Dno, Dname, Location, Description)

stored at sites s_1 and s_2 respectively, and a query that involves returning the last name of all employees working for a department named "Testing", show how the semi-join operator may be used to evaluate the query. Specify which attributes and tuples are shipped at each stage. (9)

(b) With respect to distributed databases, explain, with examples, how techniques used in centralised databases, can be extended to handle concurrency control in a distributed database. (8)

(c) With respect to parallel databases, explain, with examples, the following terms:

- (i) Inter-query parallelism
- (ii) Intra-query parallelism
- (iii) Inter-operation parallelism
- (iv) Intra-operation parallelism

(8)

(d) Describe, briefly, the main difference between the relational, object-relational and object-oriented models. (8)