Victoria University of Wellington

DEGREE EXAMINATIONS — 2000 COMP302

MID-YEAR

COMP 302

Database Systems

Time Allowed: 3Hours

Instructions: Answer all questions.
Make sure your answers are clear and to the point.
Calculators and foreign language dictionaries are allowed.
No reference material is allowed.
There are 180 possible marks on the exam.

Question 1. SQL and Relational Algebra (30 marks)

Consider the relational database EMPLOYEE_PROJECTS defined in the SQL DDL in the following manner

```
CREATE TABLE Employee(
EmpId CHR(8) PRIMARY KEY,
Fname CHAR(10) NOT NULL,
Lname CHAR(10) NOT NULL,
Salary INT(4) NOT NULL,
Manager CHR(8),
CONSTRAINT eefk FOREIGN KEY (Manager) REFERENCES Employee (EmpId));

CREATE TABLE Project(
ProjId CHR(6) PRIMARY KEY,
Pname CHAR(26) NOT NULL,
Pmgr CHR(8),
CONSTRAINT pefk FOREIGN KEY (Pmgr) REFERENCES Employee (EmpId));
```

```
CREATE TABLE Works(
EmpId CHR(8) NOT NULL,
ProjId CHR(6) NOT NULL,
NoOfHours INT(2),
CONSTRAINT wpk PRIMARY KEY (EmpId, ProjId),
CONSTRAINT wefk1 FOREIGN KEY (EmpId) REFERENCES Employee (EmpId),
CONSTRAINT wpfk2 FOREIGN KEY (ProjId) REFERENCES Project (ProjId)
);
```

a) Write the following SQL queries on the EMPLOYEE_PROJECT database:

- Retrieve the number of employees that have a manager. (5 marks)
- Retrieve EmpId, Fname and Lname of employees whose Lname starts with 'M', and who have salary greater than 2,000 and less than 3,000. (5 marks)
- Retrieve EmpId, Fname, Salary and NoOfHours of all employees that work on project with Pname = 'Vic'. (2 marks)
- For every employee, retrieve EmpId, number of projects he/she works on, and the total sum of hours he/she works on these projects. (5 marks)
- Retrieve EmpId and Fname of all employees that are not managers of any project. (5 marks)
- Retrieve EmpId and Salary of all employees that earn more than employee with EmpId = '30'. Order by Salary descending. (5 marks)

b) Use relational algebra to write the query "Retrieve EmpId, Fname, Salary and NoOfHours of all employees that work on project with Pname = 'Vic' " (3 marks)

Question 2. Enhanced Entity Relationship Data Model

(30 marks)

a) Map the relational database schema EMPLOYEE_PROJECT defined in question 1 of this examination to the entity relationship data model by drawing an ER diagram. State any assumption you made, and use only one of the available conventions to represent relationship cardinalities.

(10 marks)

b) Draw an EER diagram of the conceptual schema for the part of a University database, described in the following manner: (20 marks)

- Lecturers, general staff and students are persons.
- Persons are uniquely identified by values of PerId (person's ID), and have Fname (first name), and Lname (last name).
- Additionally:
    - every student has at least one Major and one NoOfPts (number of points) for each major,
    - lecturers have Position and AcQual (academic qualification),
    - general staff have GenPos (general position),
- A lecturer teaches at most one course, whereas a student takes at least one course,
- Courses are uniquely identified by values of CouId (course ID), and have Cname (course name),
- Every course is taught by at least one lecturer, and can be taken by many students, but there can exist courses that are not taken by any students,
- Each course uses at least one textbook,
- A textbook is uniquely identified by the course, which uses that book, and by OrdNo. Attribute OrdNo is the ordinal number of the book in the list of the textbooks of a particular course. The book also has Title and Publisher attributes,
- Every book has at least one author,
- Authors are uniquely identified by AuthId (author ID), and have Afname (author first name) and Alname (author last name),
- Each author has written at least one book,
- The property of the association between authors and books is AuthSeqNo (author's sequence number).

Question 3. Functional Dependencies and Normalization

(40 marks)

a) Which constraint is crucial to understand update anomalies? (1 mark)

b) Describe insertion, deletion and modification anomalies. (3 marks)

c) What is the name of the relation we have to start with to show that two of its projections have an additive (lossy) join? Give a short example. (4 marks)

d) Which condition has to be satisfied when joining two relations to be sure we are not going to get spurious (wrong) tuples in the resulting table? (2 marks)

e) Functional dependencies represent important database constraints. A DBMS is capable of checking these constraints only if one condition is met. Describe that condition in one sentence. Is that condition in any relationship with normal forms? (Explain your answer) (2 marks)

f) State whether each of the following inference rules is true or false. If claimed false, give a relation that satisfies functional dependencies in the left-hand side of the inference rule but does not satisfy the dependencies in the right-hand side.

- $\{W \rightarrow Y, X \rightarrow Z\} \models \{WX \rightarrow Y\}$ (2 marks)
- $\{XY \rightarrow Z, Y \rightarrow W\} \models \{WX \rightarrow Z\}$ (2 marks)
- $\{X \rightarrow Z, Y \rightarrow Z\} \models \{X \rightarrow Y\}$ (2 marks)
- $\{X \rightarrow Y, XY \rightarrow Z\} \models \{X \rightarrow Z\}$ (2 marks)
- $\{XY \rightarrow Z, Z \rightarrow Y\} \models \{Z \rightarrow X\}$ (2 marks)

(* $\models$ means "implies" *)

g) A database design expert defined the following set of attributes

$$U = \{A, B, C, D, E, F\},$$

and the following set of functional dependencies on U

$$F = \{A \to B, B \to A, B \to C, D \to E, E \to D, D \to F\}.$$

He proved that F represents a minimal cover.

Then he started to design a relational database schema, and came to the following two sets of attributes

$$R_1 = \{A, B, C\} \text{ and } R_2 = \{D, E, F\}.$$

You are invited to finish his design, but you may not change the attributes of R_1 and R_2. To finish the design, you are instructed:

- To find the set of functional dependencies F_1 on R_1 and set of functional dependencies F_2 on R_2, by starting from the set of functional dependencies F and attribute sets R_1 and R_2 (in other words to find projections $F_1 = F|_{R_1}$ and $F_2 = F|_{R_2}$) (3 marks)
- To find all the candidate keys of R_1 and R_2 with regard to F_1 and F_2, respectively, and to designate primary keys, (3 marks)
- To prove the highest normal form satisfied by R_1 and R_2, (3 marks)
- To prove whether R_1 and R_2 represents an attribute and constraint preserving decomposition of (U, F), (2 marks)
- To prove whether R_1 and R_2 represents a non additive join decomposition of (U, F), and if not to transform the set of relation schemes $\{R_1, R_2\}$ to a non additive (lossless) join decomposition of (U, F), (5 marks)
- To define referential integrity constraints for the obtained set of relation schemas. (2 marks)

Question 4. Object - Oriented Data Model and Databases

(30 marks)

a) Name at least two application areas where you expect object - oriented database will suit better than relational. (2 marks)

b) Name at least two application areas where you expect relational database will suit better than object - oriented. (2 marks)

c) Relational and object - oriented approach differ in many aspects. Comment and briefly explain how they differ with regard to:

- domain types, (2 marks)
- object and tuple structure, (2 marks)
- object and tuple identification and association, (2 marks)
- mechanisms to achieve persistence, (3 marks)
- type of data manipulation language in current use. (2 marks)

d) Map the part of the University database schema, described in question 2 b) of this examination, to Objectivity DDL. (Minor errors in syntax will not be penalized.) To do the mapping, use either the description given or the EER diagram that you developed. (15 marks)

Question 5. Query Processing and Optimization (30 marks)

Consider the SQL DDL database description given in question 1 of this examination, and the following SQL query

SELECT e.EmpId, p.ProjId, w.NoOfHours

FROM Employee e, Project p, Works w

WHERE e.EmpId = p.Pmgr AND e.EmpId = w.EmpId AND

w.NoOfHours => 50;

a) Draw the heuristic optimization tree that corresponds to this query.

(10 marks)

b) Suppose that:

- NoOfHours is monthly amount of hours assigned to an employee to work on a project, and the total number of hours assigned to any employee can't exceed 150 hours per month,
- values of NoOfHours are evenly distributed between 0 and 99, and each value appears with nearly equal frequency in the Work table,
- there are 2,000 employee records with bfr_e = 10 (bfr means blocking factor),
- there are 100 project records with bfr_p = 10,
- there are 5,000 work records with bfr_w = 25,
- there are 5 buffers of equal capacity provided in main memory for storing the file blocks during the execution of above mentioned query,
- intermediate and final result are materialized (stored as temporary tables - files on disk),
- when a join is performed, it is an equijoin (there are two columns for join condition attributes in the result),
- whenever an operation that alters the record's size is performed, the result is written in the intermediate file so that disk blocks correspond to the capacity of the main memory buffers (the blocking factor is appropriately increased or decreased, so to fit best to buffer capacity, and the number of blocks at the output of the operation can be less or greater than the number of blocks at the input, respectively).

Calculate the execution cost of the SQL query, defined at the beginning of this question. State any assumption you make. (20 marks)

Question 6. Transaction Processing and Concurrency Control

(20 marks)

Consider the following two transactions:

T_1	T_2
read_item(X)	read_item(X)
set X:= X - 1	set X:= X + 1
write_item(X)	write_item(X)

a) Concurrent execution of these two transactions leads to a well-known concurrency anomaly. What is the name of this anomaly? (1 mark)

b) Show that every serial execution of these two transactions preserves the consistency of the database. (3 marks)

c) Show a concurrent execution of T_1 and T_2 (where the model of concurrency is interleaving) that produces a schedule that is not conflict serializable. Explain why it is not serializable. (4 marks)

d) Is there a concurrent execution of T_1 and T_2, which produces a conflict serializable schedule? If so, explain what it is and show that it is conflict serializable. If not, explain why there is no conflict serializable schedule of T_1 and T_2. (4 marks)

e) Show that two - phase locking protocol ensures conflict serializability of T_1 and T_2, and briefly explain why. (4 marks)

f) If two - phase locking is applied, and we consider only T_1 and T_2, is it possible either dead lock or starvation can occur? (Explain your answer.) (4 marks)

****

COMP 302 **end**