Victoria University of Wellington

DEGREE EXAMINATIONS — 1997

(MID-YEAR)

COMP 302

Database Systems

Time Allowed: 3 hours.

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

1. **Relational model**

   (a) Consider the relation instance R.

R

X	Y	Z	W
3	4	11	1
1	2	4	3
1	2	12	1
2	1	5	9
1	2	14	4
3	4	3	9
1	2	15	2

i. Using the relational algebra aggregate operator, give an expression that finds the average of attribute Z for each value of attribute X.

[4 marks]

ii. Show the result of applying the expression in part (i) to the relation instance R.

[10 marks]

iii. Using the relational algebra aggregate operator, give an expression that applies the function *count* to the attribute W and *minimum* to the attribute Z, while grouping on the attributes X and Y.

[4 marks]

iv. Show the result of applying the expression in part (iii) to the relation instance R.

[10 marks]

(b) A recognised problem in the database field is called the *parts explosion problem*. The problem can be explained as follows. A given database contains a parts inventory relation $P(P\_Number, P\_Name, \ldots)$ and another relation $PP(Major\_Part\_Number, Minor\_Part\_Number)$ showing which parts contain other parts as immediate components (or subparts). An example of a possible instance of PP is:

PP

Major_Part_Number	Minor_Part_Number
P1	P2
P1	P3
P1	P4
P2	P3
P2	P5
P3	P5
P3	P6
P4	P3
P4	P6
P7	P8
P7	P9

The problem can be stated "Find all components for a given part (say $P1$) to all levels (ie. all subparts of $P1$, all the subparts of the subparts of $P1$ etc.)".

i. Can you give a relational algebra expression to find this result in general? If so, give the expression. If not, explain why you cannot.

[5 marks]

ii. If the problem was to find all components of a given part to 2 levels, could the query be expressed in relational algebra? If so, give the expression. If not, explain why you cannot.

[5 marks]

(c) Consider the following relation instances:

R

A	B	C	D
2	4	6	8
1	3	5	7
1	2	5	6
2	4	3	4
1	3	8	9
2	4	5	6
1	2	3	4

S

A	B
2	4
1	2

i. Show what the following relational algebra query produces.
$R \div S$

[10 marks]

ii. Is the result a relation?

[2 marks]

iii. Rewrite this relational algebra expression using only operators from the following set $\{\sigma, \pi, \cup, -, \times\}$. Show that this expression produces the same result.

[10 marks]

2. **Normalisation**

   (a) i. In one sentence, define the term "functional dependency".
   [5 marks]

   ii. In a typical organisation, who specifies the functional dependencies that hold among the attributes of a relation schema?
   [2 marks]

   iii. Why can we not deduce a functional dependency from a particular instance of a database?
   [3 marks]

   iv. Should a relation schema always be normalized to 3NF? Why or why not?
   [4 marks]

   (b) Consider the relation R, which has attributes that hold schedules of courses at a university;

   $R = \{course\_no,\ offering\_dept,\ credits,\ course\_level,\ instructor\_name,$
   $semester,\ year,\ days\_hours,\ room\_no,\ no\_of\_students\}.$

   The attribute *days_hours* is an integer whose range is from $1 - 50$, where 1 represents Monday at 8am, 10 represents Monday at 5pm, 11 represents Tuesday at 8am, and 50 represents Friday at 5pm.

   Suppose that the following functional dependencies hold on R:

   $\{course\_no,\ semester,\ year,\ room\_no,\ day\_hrs\} \rightarrow$
   $\{offering\_dept, credits, course\_level,\ no\_of\_students,\ instructor\_ssn\}$

   $\{course\_no,\ semester,\ year\} \rightarrow \{no\_of\_students\}$

   $\{course\_no\} \rightarrow \{offering\_dept,\ credit\_hours,\ course\_level\}$

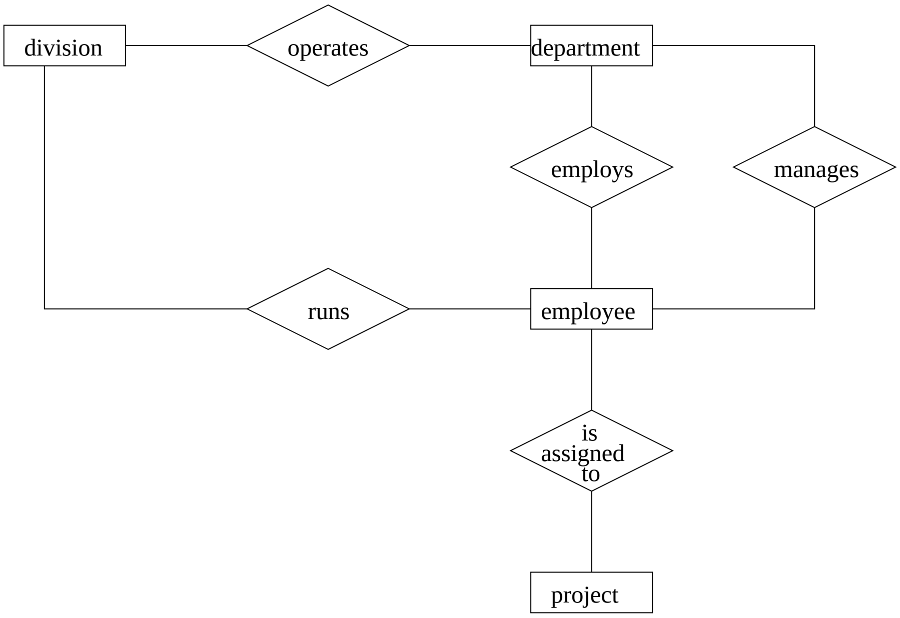
   i. Which set of attributes forms a key of R?
   [4 marks]

   ii. A relation schema R is in Boyce-Codd Normal Form (BCNF) if whenever a functional dependency $X \rightarrow A$ holds in R, then X is a superkey of R. Decompose R into BCNF relations. List any assumptions you make.
   [12 marks]

   iii. Are the resulting relations in part (ii) also in 3NF? If they are, prove that they are. If they aren't, prove that they are not.
   [5 marks]

3. **Enhanced entity relationship diagrams**

(a) Consider the following entity relationship diagram.



Redraw the above diagram to include the following structural constraints.

i. A department employs many employees, but each employee is employed by one department.

ii. Some employees, known as "rovers", are not assigned to any department.

iii. A division operates many departments, but each department is operated by only one division. Each division must have at least one department and each department must belong to a division.

iv. An employee may be assigned to many projects and a project may have many employees assigned to it.

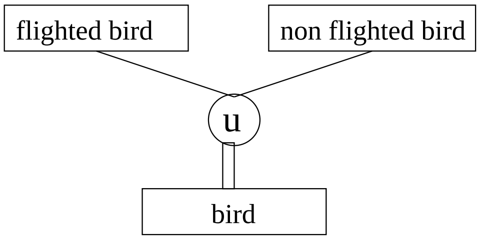
v. A project must have at least one employee assigned to it.

vi. One of the employees manages each department. Each department must have a manager.

vii. One of the employees runs each division. Each division must have someone running it.

[10 marks]

(b) Consider the following enhanced entity relationship diagram.



i. Describe the relationships and constraints that are modelled in the diagram above.

[2 marks]

ii. Draw another enhanced entity relationship diagram, using the specialisation/generalisation construct, that shows the same information.

[3 marks]

(c) Draw an enhanced entity relationship diagram that models the following classification.

i. Animals can be mammals or oviparous animals.

ii. Mammals do not lay eggs. Oviparous animals lay eggs.

iii. Mammals feed their young milk. Oviparous animals feed their young predigested food.

iv. An animal is called ornithorhyncus if it is both a mammal and an oviparous animal, i.e. it lays eggs and feeds its young with milk.

[5 marks]

4. **Query processing and query optimisation**

(a) Consider relations *employee* and *department*. There is a primary index on *dnumber* in relation *department*.

employee

fname	lname	address	bdate	dno
Jim	Bulge	7 Willis St	19 Dec 1955	1
Jenny	Skipper	19 Victoria St	2 Feb 1960	2
Locky	Schmitt	55 Lambton Quay	1 Jan 1966	2
Nicola	Smith	43 Dixon St	10 July 1966	1
Wyatt	Creature	1 Ghuznee St	31 May 1969	3

department

dnumber	dname
1	Research
2	Administration
3	Head Office

Consider the following SQL query.

SELECT fname, lname
FROM employee, department
WHERE dname = "Research" AND dnumber = dno

i. Draw the initial query tree for the query above.

[5 marks]

ii. Draw the series of trees that would follow from the query tree in part (i) when heuristic algebraic optimisation is applied to this query. Describe the heuristic that is applied at each step.

[20 marks]

(b) Consider a **project** operation $\pi_{\langle attribute\ list \rangle}(R)$. If $\langle attribute\ list \rangle$ does not include a key of relation R, then duplicate tuples must be eliminated. This can be done by sorting the result of the operation and then eliminating duplicate tuples, which appear consecutively after sorting.

i. Develop a cost function for this **project** operation. (Remember to include the cost of reading and writing records to and from disk.)

[8 marks]

ii. Explain how the cost function developed in part (i) changes if $\langle attribute\ list \rangle$ does include a key of relation R.

[2 marks]

5. **Concurrency control**

(a) Consider the following two transactions:

T_0:	read(A);	T_1:	read(B);
	read(B);		read(A);
	if A=0 then B:=B+1;		if B=0 then A:=A+1;
	write(B)		write(A);

Let the consistency requirement be $A = 0 \vee B = 0$, with $A = B = 0$ the initial values.

i. Show that every serial execution of these two transactions preserves the consistency of the database.

[6 marks]

ii. Show a concurrent execution of T_0 and T_1 (where the model of concurrency is interleaving) which produces a schedule that is not conflict serializable. Show it is not conflict serializable.

[6 marks]

iii. Is there a concurrent execution of T_0 and T_1 which produces a conflict serializable schedule? If so, explain what it is and show that it is conflict serializable. If not, explain why there isn't.

[6 marks]

(b) Show that the two-phase locking protocol ensures conflict serializability. (A formal proof is not required.)

[6 marks]

(c) If deadlock is avoided, is starvation still possible? Explain your answer.

[6 marks]

****