M358/X

Mathematics and Computing: A Third Level Course
M358 Relational Databases

Friday 22 October 1999 2.30 pm – 5.30 pm

Time allowed: 3 hours

There are **TWO** parts to this paper and you should attempt **BOTH PARTS**. You should attempt **ALL** questions in Part I and **TWO** questions from Part II. Your answers to Part I questions should be written *in the spaces provided* inside this examination paper. Your answers to Part II questions should be written in the answer book(s) provided. Please begin each answer to a Part II question on a new page.

Part I carries 60% of the total examination marks; each question is worth 5 marks.

Part II carries 40% of the total examination marks; each question is worth 20 marks.

The marks for each part of each question are shown thus: [4].

At the end of the examination

Check that you have written your personal identifier and examination number on this examination paper and on each answer book used. **Failure to do so will mean that your work cannot be identified.** Attach this examination paper to the front of the answer book(s) in which you have answered questions from Part II with the paper fastener provided.

Examination Number								
Personal Identifier								

PART I

Answer ALL questions from this part.

Question 1

A company arranges short holiday breaks for its customers. Data about each customer's costs for hotel, catering and extras can be recorded in a table, or tables, in the three ways described below with example data as (a), (b) and (c). Suppose that, additionally, data about the cost of travel should be recorded. Following each description, explain how travel data may be recorded in the same way.

(a) In the first way, each kind of cost is in a separate table with a row for each customer incurring that cost.

Hotel

Customer	*Cost*
C12	562
C23	895

Catering

Customer	*Cost*
C12	827
C41	512

Extras

Customer	*Cost*
C12	121
C41	75

[2]

(b) In the second way, all costs are in one table with a row for each customer and each cost incurred.

Billing

Client	*Class*	*Cost*
C12	Hotel	562
C23	Hotel	895
C12	Catering	827
C12	Extra	121
C41	Catering	512
C41	Extras	75

[1]

(c) In the third way, all costs are in one table with a row for each customer and a column for each kind of cost.

Cost

Client	*Hotel*	*Catering*	*Extras*
C12	562	827	121
C23	895		
C41		512	75

[2]

Question 2

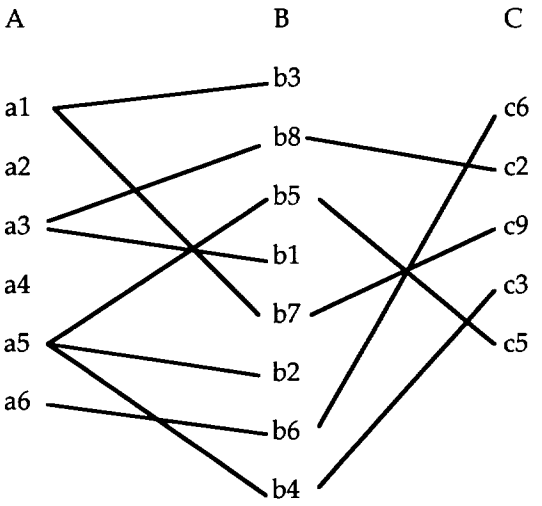


Figure 1 Example relationship occurrences

Figure 1 shows connections between occurrences of the entity types A, B and C. You should assume that all other occurrences of these entity types have relationships with the same properties as shown in this figure. Draw an E-R diagram that gives the degree and participation conditions of these relationships.

[5]

Question 3

(a) Consider the following extension of a relation, *R*.

R

A	*B*	*C*	*D*	*E*
a	1	12	5	red
b	2	12	3	red
a	2	11	4	blue
b	1	13	2	unknown
b	3	13	0	blue
a	3	13	1	blue

(i) What is the cardinality of this relation?

[1]

(ii) What is the degree of this relation?

[1]

(iii) Give the candidate key(s) that may be inferred from these occurrences.

[1]

(b) Three relations, *R*, *S* and *T*, represent three entity types, R, S and T, respectively. Relations *R*, *S* and *T* have the following headings.

R(*A*, *B*, *C*)

S(*D*, *E*, *F*)

T(*G*, *H*, *I*)

(A, B), *D*, *F* and *G* are candidate keys of their respective relations.

Attributes *A* and *D* are defined on the same domain and represent a relationship between *R* and *S*.

Attributes *F* and *G* are defined on the same domain, which is different from that of *A* and *D* and represent a relationship between *S* and *T*.

Give an E-R diagram showing the relationships between the entity types that result from the given attributes. Do *NOT* include participation conditions.

[2]

Question 4

(a) The following relations contain data about the departments in an organisation and the employees that work in each department.

Employee

EmpNo	*Name*	*InDept*	*Salary*
e1	Smith	d3	25000
e2	Newton	d2	22000
e3	Jones	d2	18000
e4	Kay	d1	15000
e5	Carling	d3	15000
e6	Sharpe	d1	19000
e7	Ismail	d1	20000

Department

DeptNo	*Name*	*Head*
d1	engineering	e7
d2	marketing	e3
d3	administration	e1

Give the result of executing the following relational algebra expression.

```
(select
  (project
        (join Employee and Department where EmpNo = Head)
  over Employee.Name, Salary)
where Salary > 18000
```

[2]

(b) The following five relational algebra expressions relate to the two relations *Student*, having attributes *StudentId, Name, Registered, CounsellorNo* and *Region*, and *Enrolment*, having attributes *StudentId, CourseCode* and *TutorNo*, as defined in the course text.

Circle two numbers associated with the two expressions that, when evaluated, give identical relations.

```
1 (project Enrolment over StudentId)
  difference
  (project (select Student where Region <> 4) over StudentId)

2 join Enrolment and (select Student where Region = 4)
        where Enrolment.StudentId  = Student.StudentId

3 select (Enrolment times (select Student where Region = 4))
        where Enrolment.StudentId  = Student.StudentId

4 divide (project Enrolment over StudentId, CourseCode)
  by (project (select Student where Region  <> 4) over StudentId)
  over StudentId

5 (project Enrolment over StudentId)
  intersection
  (project (select Student where Region = 4) over StudentId)
```

[3]

Question 5

The following tables represent details of doctors and their patients, based on the Hospital example in the course text. Note that the example data is only provided to help your understanding and is not required for answering the questions.

doctor

staff_no	*doctor_name*	*position*	*specialism*
110	Liversage	Consultant	Cardiac
131	Kalsi	House Officer	Orthopaedic
156	Hollis	Registrar	Geriatric
200	Tamblin	Consultant	Paediatric
231	Ansell	Consultant	Orthopaedic
232	Shine	Consultant	Paediatric

patient

patient_id	*patient_name*	*consultant_no*	*ward_no*
p04	Anarkali	232	w3
p12	Hook	231	w2
p22	Parsi	232	w2
p26	Colbert	110	w1
p31	Rubinstein	231	w2
p37	Seaward	232	w3
p38	Ming	110	w2
p67	Jarvis	232	w3
p73	Purdy	200	w3
p91	Fellows	232	w2

For these tables, write an SQL query to give:

(a) the staff number and name of each doctor whose specialism is Paediatric;

[1]

(b) the patient identifier and name of each patient in ward w2 who is treated by a consultant having a staff number of either 200, 231 or 232;

[2]

(c) the patient identifier and name of each patient and the specialism of the consultant who is responsible for them.

[2]

Question 6

The following table represents details of doctors, based on the Hospital example in the course text.

doctor

staff_no	*doctor_name*	*position*	*specialism*
110	Liversage	Consultant	Cardiac
131	Kalsi	House Officer	Orthopaedic
156	Hollis	Registrar	Geriatric
200	Tamblin	Consultant	Paediatric
231	Ansell	Consultant	Orthopaedic
232	Shine	Consultant	Paediatric

The following statement is intended to specify a function named **doctor_salary**.

```
CREATE FUNCTION doctor_salary (f_position **********)
RETURNS **********
BEGIN
  IF f_position = 'Consultant' THEN RETURN 60000;
  ELSEIF f_position = 'Registrar' THEN RETURN 30000;
  ELSEIF f_position = 'House Officer' THEN RETURN 18000;
  ELSE RETURN NULL;
  END IF;
END
```

(a) The function specification is incomplete in the two places shown by **********. Give below a specification for each one to complete the statement appropriately.

[2]

(b) Using the **doctor** table, write an SQL query that demonstrates the use of the completed function **doctor_salary** by giving each doctor's name and salary.

[1]

(c) Give the table produced by executing the query you gave in part (b).

[1]

(d) Suppose the following INSERT statement is executed.

```
INSERT INTO doctor (staff_no, doctor_name)
VALUES ('250','Matthews')
```

Give the additional row that would now appear in the result of executing the query you gave in part (b).

[1]

Question 7

Complete each of the sentences about SQL given in parts (a) to (e) by writing in each place shown by _____ the appropriate term from the following list:

host variable, independent, at, built-in function, next, outer, row, concurrent, aggregate function, after, expression, serialized, value, routine, column, before, inner, isolated, procedure, natural, first, function, last

(a) A _____ that is not included in the GROUP BY clause of a query can only appear in the SELECT clause of the query if it is an argument of some _____

(b) The _____ join is a special kind of _____ join, in which the columns to be matched have the same name.

(c) _____ execution of transactions X1 and X2 occurs when X2 may start only when X1 has finished; in this case they are not _____ transactions because only one is executed at a time.

(d) After execution of an OPEN statement, the cursor is positioned _____ the _____ row of the cursor table.

(e) Execution of a _____ results in the return of a single value, whilst execution of a _____ can result in several actions but does not return a value.

[5]

Question 8

A company distributes magazines to many customers, and each customer may subscribe to many magazines. An E-R data model for the company initially includes the entity types Customer, with identifier CustomerCode, and Magazine, with identifier MagazineNo. Both entity types have optional participation in the many-to-many relationship, Subscribes, between them.

A further data requirement is that the company needs to record the date and amount of each customer's payment for each magazine subscription. This data requirement can be represented by resolving the Subscribes relationship using a new entity type, Subscription. Show the effect of this resolution by giving:

(a) an E-R diagram of the three entity types and their relationships, with degree and participation conditions;

[3]

(b) the identifier of the entity type Subscription.

[2]

Question 9

For each of the normal forms in (a), (b) and (c), give the two terms from the following list that appear in its definition:

determinant, functionally dependent, fully functionally dependent, functional dependency, transitively dependent, transitive dependency, primary key, foreign key, alternate key, candidate key, single-valued fact

(a) first normal form

[1]

(b) third normal form

[2]

(c) Boyce-Codd normal form

[2]

Question 10

(a) Suppose a column of a table contains character string values and is not a primary or foreign key of that table. List three properties of the data in this column that should be taken into account in defining the column in an SQL CREATE TABLE statement (just give a word or phrase identifying each property; you do not have to show how it is expressed in SQL).

[3]

(b) Complete each of the following sentences by inserting the appropriate word.

(i) The main reason why a column that contains integer values would be defined as a CHARACTER data type in a table is that numerical _____ are not applicable to the data. [1]

(ii) The scaling factor for a column of numeric data is an example of _____, which is necessary for understanding the meaning of data stored in a database. [1]

Question 11

The course text describes three approaches to distributing data:

1 a client/multi-server system

2 a distributed database system

3 a replication system.

Complete each of the following sentences by giving the number of the approach for which it is valid.

(a) ____ provides location independence. [1]

(b) ____ avoids distributed transactions. [1]

(c) ____ does not provide support for replication. [1]

(d) ____ requires a user process to control distributed processing. [1]

(e) ____ enables an SQL statement to refer to data in more than one database. [1]

Question 12

A supermarket chain needs to devise a marketing strategy to promote a new product line: vitamin enriched fruit juices. To determine whether there is a customer demand for this product line, the supermarket chain wishes to run a pilot scheme that introduces the new product only at a small number of branches, preferably those within a region of the country supplied by a single distribution depot. The supermarket chain plans to mail discount vouchers to households that it believes would be interested in the promotion, and is using Knowledge Discovery in Databases (KDD, as described in the course) to identify those households. Figure 2 shows the steps of the general model for the KDD process, without names for the processes and their results, and a description of how it relates to the supermarket example. (Note that this example is provided to help your understanding of the model and is not required in the answer).

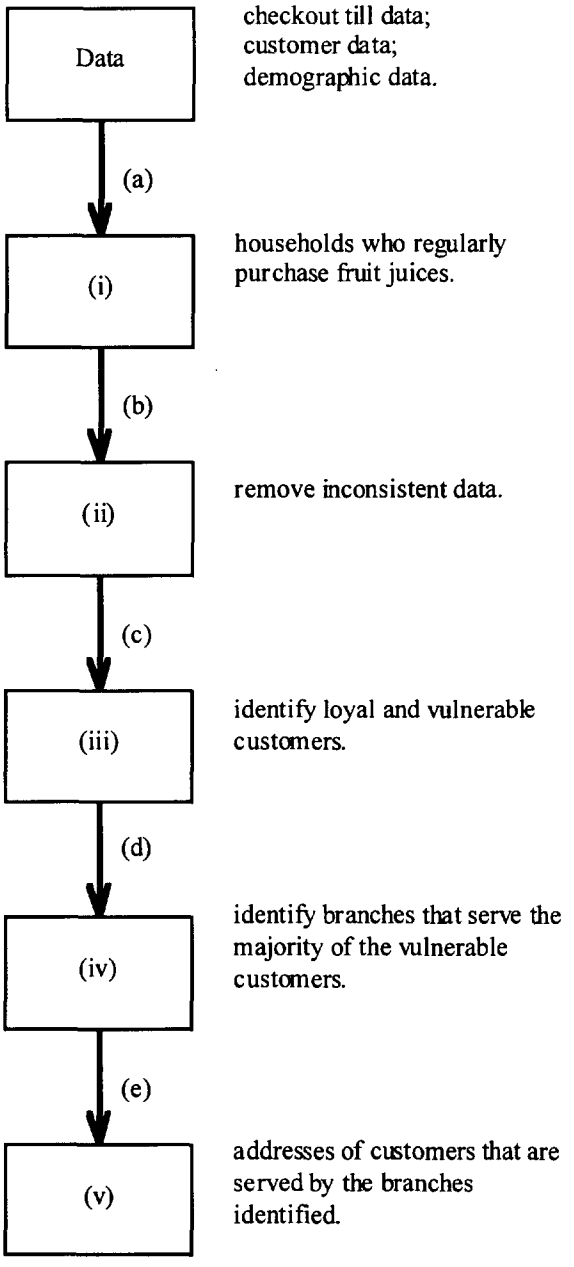


Figure 2 KDD process

In the table below, give the name of each of the processes (a)-(e) and each of the results (i)-(v) for the general model shown in Figure 2.

(a)	(i)
(b)	(ii)
(c)	(iii)
(d)	(iv)
(e)	(v)

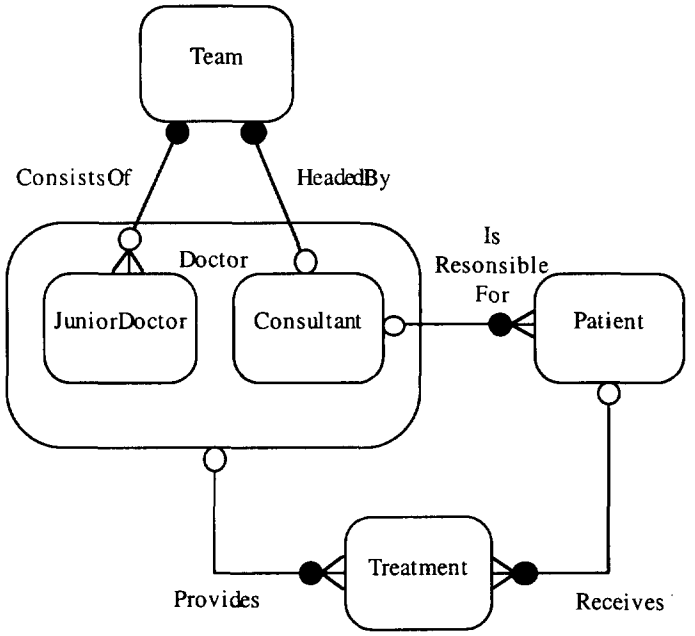
[5]

PART II

Answer **TWO** questions from this part.

Question 13

In an E-R model, entity subtypes can be used as a convenient way of modelling certain complex data requirements. Figure 3 gives a fragment of an E–R model in which entity subtypes are used to represent the requirements of the Hospital example as given in the course.



Entity types

Doctor (StaffNo, DoctorName)
  JuniorDoctor (Position)
  Consultant (Specialism)
Patient (PatientId, PatientName)
Team (TeamCode, TelephoneNo)
Treatment (StaffNo, PatientId, StartDate, Reason)

Constraints

The attribute Position (of JuniorDoctor) may only have a value of Registrar or House Officer.

Assumptions

None.

Figure 3 Hospital example E–R model

The subtype concept has no direct counterpart in relational theory and, for a relational representation of an E-R model containing subtypes, it is possible to choose between *either* representing all subtypes by a single relation (such as the *Doctor* relation as given in the Hospital relational model in the course text) *or* by a relation for each subtype. In this question you are required to produce a relational model representing the E-R model in Figure 3 where the entity subtypes JuniorDoctor and Consultant are each represented by a relation, *JuniorDoctor* and *Consultant* respectively.

(a) Provide a list of relation headings that represent the entity types and relationships of the E-R model for the Hospital example given in Figure 3. [5]

(b) Assume the following domains are defined:

domains

IdentifiersOfPatients = p01 .. p99
PersonNames = string
StaffNumbers = 100 .. 999
PositionsOfJuniorDoctors = (Registrar, House Officer)
SpecialismsOfConsultants = string
Reasons = string
CodesOfTeams = t1 .. t8
Dates = calendardates,

Using these domains, produce a relational model corresponding to the E–R model given in Figure 3. Make sure that entity and referential integrity is maintained in your relational model, and that mandatory participation conditions are enforced (but you do not need to include other constraints). Annotate your model with comments to show how and where you have represented these conditions. [15]

Question 14

The Study database used in the course contains the table **country**, having the columns relevant to this question of **name** and **capital**, and the table **city**, having the columns **name**, **country** and **population**.

An SQL query, based on these tables, is as follows:

```
SELECT country.name, capital, count(*)
FROM country, city
WHERE country.name = city.country
GROUP BY country.name, capital
HAVING MAX(city.population) > 5000.0
```

(a) Describe the logical processing model for this query. [6]

(b) Give, as a sentence, the request that the query answers. [2]

(c) Give one reason why this query cannot be reformulated using a subquery — that is, without using a join. [2]

(d) Give the definition of a view that provides the same data as that resulting from the query, explaining any choices you made. [3]

(e) Explain why this view is not updatable, identifying each aspect of its definition that prevents it being updatable. [4]

(f) Give three general reasons why the use of a view might be beneficial. [3]

Question 15

(a) Derived data may be expected in many situations, such as in the Hospital example for the course where a value for the number of patients on a ward can be derived by counting the patients currently recorded as occupying that ward. Give one general reason why derived data should *not* normally be included in a conceptual data model. [2]

(b) For any derived data required by users, a database administrator has to decide during database design whether to calculate its value each time the data is required — but not store it — or to define that data as a column in a base table so that it is stored in the database.

(i) Describe how a database administrator can enable users to access derived data as a column in a table without it being stored in a base table. [2]

(ii) Describe one possible benefit of storing derived data as a column in a base table. [2]

(c) One potential problem of storing derived data as a column in a base table is maintaining consistency. However, triggers can be used to overcome this problem. Suppose the number of patients on a ward is stored as an additional column **no_of_patients** in the **ward** table. Describe, without giving any SQL statements, the three triggers required to maintain consistency of the column **no_of_patients** in the **ward** table when there are changes to the data in the **patient** table, and explain the action of each trigger. [6]

(d) Suppose that the column **no_of_patients** has the same meaning as in part (c), but that triggers are not used to maintain consistency of this column.

(i) Describe, without giving any SQL statement, a constraint that enforces consistency between the column **no_of_patients** and the **patient** table, and explain how it may cause problems when changes are made to the **patient** table. [4]

(ii) One possible approach to maintaining consistency is to require each application process that updates the **patient** table to update the column **no_of_patients** as well. Describe the issues that have to be taken into account in this approach. [4]

Question 16

(a) Give three reasons why a data warehouse would be implemented as a database separate from the operational systems that support the daily business of an organization. [3]

(b) A model showing various components involved in the building and use of a data warehouse includes a DBMS and four further components.

(i) Give a diagram showing the role of a DBMS and these four components in a data warehouse system. [4]

(ii) Describe the main tasks of each of the four components. [8]

(iii) State which components require certain metadata and explain how it is used. [5]

[END OF QUESTION PAPER]