THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **CO2209 ZB**

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING and COMBINED DEGREE SCHEME

Database Systems

Wednesday 9 May 2018: 10.00 – 13.00

Time allowed: 3 hours

There are **FIVE** questions on this paper. Candidates should answer **FOUR** questions. All questions carry equal marks and full marks can be obtained for complete answers to **FOUR** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

Only your first **FOUR** answers, in the order that they appear in your answer book, will be marked.

There are 100 marks available on this paper.

A handheld calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

Question 1

The relations below hold information about Therapists and Clients, and also Sessions held between Therapists and Clients for treatment.

THERAPIST
Primary Key: TherapistID

TherapistID	Surname	Therapy Centre	Speciality
T3141	Kapoor	Third Eye	Physical Therapy
T5926	Locatelli	Alternatives	Meditation
T5358	Rani	Alternatives	Colour Therapy
T9793	Sen	Eastern Wisdom	Physical Therapy
T2384	Dutoit	Alternatives	Massage
T6264	Senna	.NULL.	Iridology
T3383	Santelli	Third Eye	Colour Therapy
T2795	Hoch	Eastern Wisdom	Acupuncture
T0288	Noshima	Good Vibes	Acupuncture
T4197	Pillai	Third Eye	Massage

CLIENT
Primary Key: ClientID

ClientID	Surname	FirstName
HG90322	Farook	Mohammed
YT74221	Trilling	Lionel
HR03045	Chou	Fan
BB39842	Feltenham	Carol
KR49832	Shoemaker	Aria
PR49933	Hoblein	Johan
TG49272	Zucker	Gunter
YT74334	Henderson	Sheila

SESSION
Primary Key: THERAPISTID + CLIENTID + SESSIONDATE
Foreign Keys: TherapistId references TherapistId in THERAPIST;
ClientID references ClientID in CLIENT

TherapistID	ClientID	SessionDate	Time	Duration
T3141	YT74334	2017-10-23	09.00	0.5
T3141	HG90322	2017-10-23	10.00	1.5
T5358	YT74221	2017-12-06	11.00	1.0
T2384	HR03045	2017-12-11	10.00	0.5
T2384	BB39842	2017-08-07	13.00	1.5
T3383	TG49272	2017-06-19	11.00	2.0
T3141	YT74334	2017-06-19	09.00	1.0
T2795	HR03045	2017-11-13	13.30	1.5
T0288	KR49832	2017-11-24	15.00	3.0
T4197	YT74221	2017-11-28	11.00	0.5
T3141	HG90322	2017-10-19	10.00	2
T5926	HR03045	2017-11-22	09.30	0.5
T5358	HG90322	2017-11-22	10.30	0.5

A. Construct queries in SQL to carry out the following tasks:

1. Get TherapistIDs of Therapists whose Speciality is 'Colour Therapy' and whose Therapy Centre is 'Third Eye'.

[1 mark]

2. List the Surnames of Therapists at Third Eye who have Physical Therapy or Colour Therapy as their Speciality.

[1 mark]

3. Get the surnames of Clients who have Sessions with the Therapist whose TherapistID is T3141.

[1 mark]

4. Find the names of Clients who have had Sessions with Therapists whose Speciality is 'Meditation'.

[1 mark]

5. Get the Surnames of all Therapists whose Therapy Centres we do not know.

[1 mark]

6. Find the total hours worked by the Therapist whose TherapistID is T5926.

[2 marks]

7. Get the ClientIDs of all Clients who have not yet had any sessions with Therapists.

[2 marks]

8. Get the TherapistIDs and total hours worked by each of the Therapists.

[2 marks]

9. Get the TherapistIDs and total hours worked by each of the Therapists from September 2017 onwards.

[2 marks]

10. Get the TherapistIDs and total hours worked, for August 2017, for each of the Therapists whose consultation sessions total fewer than twenty hours during that month.

[3 marks]

11. Create a View that will show only the Surname and Speciality of all Therapists whose Therapy Centre is 'Third Eye'.

[2 marks]

B. A database administrator (DBA) found that queries on a large database of online customer purchases were running more and more slowly as the database tables grew in size. An analysis of the problem revealed that the main table held information on everyone who had ever placed an order, although only about 20% of customers became repeat customers. What could the DBA do to improve performance, given this fact about customers?

[3 marks]

C. What are 'NoSQL' database systems? What kind of data are they designed to deal with?

[2 marks]

D. In a distributed database system, why might we want to replicate some of the data?

[2 marks]

Total = 25 marks

Question 2

The following relation holds information about an Animal Rescue Centre. It was designed by someone who did not understand the relational model, and you have been hired as a consultant to improve it.

The Happy Homes Animal Rescue Centre takes in stray animals, nurses them back to health if necessary, and then re-homes them.

The animal will be given a unique ID number, on a chip inserted under its skin. A single member of staff (recorded in 'StNum') will be assigned to look after the animal, and will typically have several animals to look after. This member of staff has his or her first name recorded.

While in the Centre, the rescued animal will receive professional veterinary care, which will consist of immunization injections, given once only for each type of disease. The rabies injection is given immediately when the animal reaches the Centre; others may be given later. The dates of these injections are recorded for each animal. The Centre buys its vaccination drugs in batches, just one batch at a time for each kind of vaccination. It records the 'Use-By' date for each type of vaccination drug, so that it can re-order the particular drug in good time.

A partial 'snapshot' of this relation might look like the following:

Primary Key: ID + Vaccination

ID	Type	VDate	Vaccination	Use-By	StNum	StaffName
5988	Cat	2017-07-23	rabies	2019-01-31	665	Jill
5988	Cat	2017-08-02	parvovirus	2018-03-30	665	Jill
5988	Cat	2017-08-02	distemper	2018-06-30	665	Jill
5997	Dog	2017-07-30	distemper	2018-06-30	754	Sirhan
5997	Dog	2017-07-30	rabies	2019-01-31	754	Sirhan
5997	Dog	2017-08-11	leptospirosis	2020-02-15	754	Sirhan
2887	Dog	2018-01-04	rabies	2019-01-31	665	Jill
2887	Dog	2018-01-13	leptospirosis	2020-02-15	665	Jill
6502	Cat	2017-08-02	rabies	2019-01-31	754	Sirhan
6502	Cat	2017-08-08	distemper	2018-06-30	754	Sirhan

A. Identify the Functional Dependencies in this table, using the following example: If B is functionally dependent on A, show it this way: A $\rightarrow$ B

[5 marks]

B. This relation is susceptible to insertion, deletion, and update anomalies. Give an example, based on the relation above, of each kind.

[6 marks]

C. Assuming that the Primary Key of this relation is ID + Vaccination, identify the partial and transitive dependencies in the original relation.

[2 marks]

D. Change the schema so that the data in this table is in Boyce-Codd Normal Form, specifying the Primary Key of each new table. **Insert the data** from the original table into your new tables.

[8 marks]

E. Suppose a new procedure was introduced at the Centre, whereby every animal's distinguishing features, and the foods it was averse to, were recorded. Someone has proposed the following relation, which would hold both sets of information. They argue that since there are no determinants, that it does not violate the prescription 'Let every determinant be a candidate key', and is therefore in Boyce-Codd Normal Form. Comment briefly on the problems associated with this design and propose an alternative relational schema that avoids these problems.

[4 marks]

ID	Features	FoodsToAvoid
5988	Scar on abdomen	Fish
5988	White left paw	Fish-flavoured pellets
5988	Missing left canine tooth	Fish-flavoured pellets
5997	None	Fatty foods
6502	Docked tail	Kellogg's pellets
6502	Docked tail	Fish

TOTAL = 25 marks

Question 3

A bakery which bakes 'Deluxe Cakes' buys in many different types of ingredients from independent suppliers – such as flour, chocolate, sugar, *etc.* With these ingredients, it bakes several different types of 'Deluxe Cakes'. These are then sold in quantity to independent pastry retailers.

The baker wants to create a database to record information about its suppliers, cakes, ingredients, and retail pastry distributors.

A particular sort of cake may use many different ingredients. Each ingredient can be used in the baking of several different types of 'Deluxe Cakes'. A kind of 'Deluxe Cake' cannot exist in the database without a set of ingredients that make it up, but an ingredient can be recorded even if it is not currently used in the baking of any existing 'Deluxe Cakes' type, and even if it does not currently have a supplier.

The bakery may have several suppliers for each type of ingredient. A supplier is only recorded if it supplies at least one type of ingredient.

A given Pastry retailer can receive and sell many different types of 'Deluxe Cakes'. No Pastry retailer has a monopoly on re-selling any given type of 'Deluxe Cakes'. A type of 'Deluxe Cake' may exist without yet having any Pastry Retailers which sell it, whereas no Pastry Retailer will be recorded in the database unless it is selling at least one type of the baker's 'Deluxe Cakes'.

A. Draw an Entity/Relationship diagram to represent this situation. Include all entity types and their relationships but *not* their attributes. Indicate cardinality and participation constraints. Use the following conventions:

This illustrates a situation where an instance of 'Entity-type-1' can have a relationship to zero, or one (but no more than one) instance of 'Entity-type-2', and where an instance of 'Entity-type-2' must have that relationship with at least one instance of 'Entity-type-1', and can have it with an unlimited number of them.

[10 Marks]

B. Design a relational schema in Third Normal Form which could hold the information represented by the E/R diagram you drew up for **Part A.** Be sure to indicate both Primary and Foreign Keys.

Suppliers are uniquely identified by *PNums*, ingredients by *IngredientNames*, types of 'Deluxe Cake' by *CakeNames*, and Pastry Retailers by *RetailerCodes*. We also want to record the *ContactPhone* for each supplier, the *calorie-count* of each ingredient, and the *features* (there can be more than one for a single type) of each 'Deluxe Cake' type, and the *address* of each Pastry Retailer.

[10 marks]

C. Suppose our arrangements with our suppliers changed, so that each type of ingredient was supplied by one supplier only. (For example, cloves and ginger were supplied by Supplier K3308 only, and caster sugar by Supplier M9383 only.) What changes, if any, would you have to make to the schema you designed in Part **B**?

[5 marks]

TOTAL = 25 marks

Question 4

A. A catering company prepares dinners for large gatherings, and keeps records of their engagements. It can keep its prices low because it only serves one kind of meal at a given event. No venue can have more than one engagement per night. The table below shows some of their engagements, where they occurred, on what date, what was served, and how many meals were prepared. This information was stored on paper, but now the company wants to put this information in a computerized database.

Venue	Date	Food	Count
Brewer's Hall	2016-11-26	Rack of Lamb	45
Kompass Klub	2016-11-26	Fish Pie	70
Erewhon	2017-01-23	Chicken Kiev	85
S.Side Ctty Centre	2016-12-23	Rack of Lamb	140
Polonium	2016-12-23	Rack of Lamb	65
Kompass Klub	2016-04-23	Chicken Kiev	65
Polonium	2016-04-23	Fish Pie	80
Erewhon	2017-01-05	Lamb Agnon	40
Kompass Klub	2017-01-15	Chicken Kiev	75

(i) The employees charged with implementing this database could not agree on what attribute or attributes should be designated as the Primary Key. One employee proposed that Venue should be the Primary Key. A second argued that Date was the proper choice as Primary Key. A third argued that Venue and Date and Meal together should be the Primary Key. And a fourth said that all four attributes together should be the Primary Key.

For this relation, what bad consequences would ensue (in terms of the requirements stated above) if during the design phase the Primary Key were designated as:

(1) Venue
(2) Date
(3) Venue + Date + Meal
(4) Venue + Date + Meal + Count

[2 marks]

(ii) The employees couldn't agree, but one of them had an idea: why not create a fifth attribute, from the domain of integers, which would automatically be incremented each time a new row was added. Then the first row would be 1, the second 2, the third 3, and so on. These numbers would be generated automatically by the system's auto-increment feature. The values in this column would be guaranteed to be unique, and would thus make the perfect primary key.

Comment briefly on this idea.

[3 marks]

(iii) What *should* the Primary Key for this relation be, and why?

[2 marks]

B. Is it the case that if one relation has a larger degree than another, it must necessarily take up more space on secondary storage? Is it the case that if one relation has a greater cardinality than another, that it must necessarily take up more space on secondary storage? If a relation has both a greater degree and a greater cardinality than another, must it necessarily take up more space on secondary storage?

[2 marks]

C. What is a 'foreign key'? What is the effect of the ON DELETE RESTRICT and ON DELETE CASCADE statements?

[3 marks]

D. What is a data dictionary (also called a system catalogue)? Describe some of the information that it might contain, and how it might be used. In a relational database, how is the 'data dictionary' held?

[5 marks]

E. Briefly discuss the basic differences between the 'internal schema', the 'conceptual schema' and the 'external schema' of a database.

[6 marks]

F. What is meant by 'physical data independence'?

[2 marks]

TOTAL = 25 marks

Question 5

Consider a city-wide library system where books are checked out and returned. The library records these events on a database. It also records when new books are added to the stock. A sample of its database might look like this:

Books
Primary Key: BookCode

BookCode	Title	InStock
56573329104	'The Darwin Award - How to Win'	20
63876863101	'Gone with the Wine'	10
66873429195	'Slower Than Light'	40

Borrowed
Primary Key: Member + BookCode

Member	BookCode	DateOut	DateIn
ABC3009	66873429195	2017-09-21	NULL
ABC3009	63876863101	2017-10-23	2017-11-08
PDQ9977	63876863101	2017-11-16	2017-12-01
TAS9933	76654915433	2017-12-30	NULL

The library keeps its database up to date by recording changes. For example, when a member of the library whose library membership number is ABC3009 borrows a book whose BookCode is 56573329104, the Books table is updated:

```
--1
UPDATE Books
SET InStock = InStock -1
WHERE   BookCode = 56573329104;
```

and the Borrowed table is extended:

```
--2
INSERT INTO Borrowed (Member, BookCode, DateOut)
VALUES 'ABC3009', 56573329104, '2017-10-23';
```

When a new book is added to the stock, a statement like the following might be executed:

```
--3
INSERT INTO Books (BookCode, Title, InStock)
VALUES 63876863101, 'Gone With the Wine', 1;
```

When another copy of a book which is already held by the library is added to the stock, a statement like the following might be executed:

```
--4
UPDATE Books
SET InStock = InStock +1
WHERE   BookCode = 56573329104;
```

Sometimes books which have become damaged are removed from stock. A statement to do that would look like this:

```
--5
UPDATE Books
SET InStock = InStock -1
WHERE   BookCode = 75636119110;
```

When a borrower returns a book, the Books table is updated:

```
--6
UPDATE Books
SET InStock = InStock +1
WHERE   BookCode = 56573329104;
```

and the borrowed table is also updated:

```
--7
UPDATE Borrowed (Member, BookCode, DateIN)
VALUES 'ABC3009', 56573329104, '2017-10-29';
```

Answer the following questions, illustrating your answer with reference to the library and its database.

A. In the context of databases, what is meant by the word 'transaction'?

[2 marks]

B. For the SQL statements above, labelled '--1' to '--7', which, if any, make up transactions? (There may be none, one, or more than one.) Briefly explain your answer.

[8 marks]

C. Although the database is located in the central Main Library, books are also borrowed from branch libraries. (The InStock field includes the total book stock for each book across the central Main Library and the branch libraries.) At the moment, branch librarians email the Main Library with details of borrowed and returned books, and the database is updated centrally at the Main Library. Thus, all borrowings are recorded centrally in the Borrowed table.

However, the proposal has been made that local librarians be allowed to update the database directly themselves via the internet. Using as your example the SQL statements listed above, discuss the problem that could arise if this were done by simply giving them access to the central database.

[5 marks]

D. There are solutions to the problem posed in **C.** Discuss one of them, showing how it would prevent the problem.

[5 marks]

E. Solutions to the 'concurrent access' problem which involve the 'locking' of resources can themselves give rise to another problem. Describe this problem, and approaches to overcoming it.

[5 marks]

TOTAL = 25 marks

END OF PAPER