THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALL

UNIVERSITY OF LONDON

CO2209 ZA

BSc EXAMINATION

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING and COMBINED DEGREE SCHEME

Database Systems

Thursday 9 May 2019: 10.00 – 13.00

Time allowed: 3 hours

DO NOT TURN OVER UNTIL TOLD TO BEGIN

There are **FIVE** questions in 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 following table, FirstTest, records the outcomes of Initial Tests for Patients who are undergoing a preliminary screening test for a particular illness. PatientNum uniquely identifies a patient. If a Patient tests negative in the preliminary screening they will not be re-tested for that illness again. In both cases their results will be stored in the FirstTest Table. If they test Positive, they will be re-tested later along with others who tested positive in the preliminary screening, and that data will be put in a separate table, called ReTests.

FirstTest

PatientNum	**Test**	**TestDate**	**Outcome**
P347769	Diphtheria	2017-11-28	Positive
P347769	Enteritis	2017-12-03	Positive
P398876	Diphtheria	2017-11-28	Negative
P289877	HNV1	2017-11-28	Negative
P289877	Diphtheria	2017-11-28	Positive

The Primary Key of this table is PatientNum + Test.

A. What bad consequences could follow if, in creating the table – before we added the data – we defined the Primary Key of the table as

(1) PatientNum alone?
(2) Test alone?
(3) TestDate alone?
(4) PatientNum + Outcome?
(5) PatientNum + TestDate?

[5 marks]

B. Suppose it is decided to allow Patients to re-test for illnesses for which they initially tested negative, at a later date. Would we need to change the definition of the Primary Key? If we did not change it, what problem might arise? If we changed it, what would the new key be? Explain your answer.

[5 marks]

C. This table records which patient is being treated for which disease(s), with which drug(s). Notice that a patient may (or may not) have more than one disease, and may (or may not) be treated for a specific disease with more than one drug.

Treatment

PatientNum	**Disease**	**Drug**
P347769	Diphtheria	Erythromycin
P347769	Diphtheria	Antitoxin
P347769	Enteritis	Erythromycin
P289877	Diphtheria	Penicillin
P488213	Diphtheria	Erythromycin

(1) What is the Primary Key of this table?

(2) A colleague suggests that this table should be "normalized" by splitting it into two tables, one with the attributes PatientNum + Disease, and the other with the attributes Disease + Drug. He notes that these tables would occupy less storage space than the original table (for instance, the fact that Patient P347769 has Diphtheria would only occupy one row), and that if we wanted to make a query which referenced all three columns, we could just rejoin them.

Comment on the advantages (if any) and/or disadvantages (if any) of following his suggestion.

[5 marks]

D. Relational theory uses the terms "Primary Key" and "Candidate Key". How are these two terms related? How many Candidate Keys are there in the relation FirstTest in Part **A**? In relational theory, can there be two tuples with identical Primary Keys in the same relation? Can there be two tuples with identical Candidate Keys in the same relation?

[5 marks]

E. An organisation stores information about individuals (such as names, addresses, phone/email numbers, *etc.*) in a table with SocialSecurityNumber as Primary Key. Occasionally, it receives full information about individuals, but without the SocialSecurityNumber. It was proposed simply to enter the information into the database when received, and to add the SocialSecurityNumber some time later when known. Comment fully on this approach. Suggest an alternative approach.

[5 marks]

TOTAL = 25 marks

Question 2

A. Sometimes when transactions are being processed, a system failure occurs. When this happens, we can identify five types of transactions, from the viewpoint of the relationship of their start and completion times with respect to the last checkpoint and the point at which the system failed.

(1) What is a "checkpoint"? What happens at a checkpoint?

(2) Describe each of these types of transaction, stating, in each case, the corresponding recovery action that a DBMS must take to maintain a consistent database (a diagram will help your explanation).

[12 marks]

B. Briefly explain the ACID properties which transactions should possess.

[8 marks]

C. What is meant by saying that a set of transactions whose execution operations are interleaved with each other, is "Serializable"? How can this be guaranteed for a set of transactions? What is the "two-phase locking protocol"?

[5 marks]

TOTAL = 25 marks

Question 3

A. What do we mean by the "cardinality" of a relationship between two entity-types in an Entity-Relationship diagram? Illustrate your answer with example diagrams.

[3 marks]

B. A data modeller wants to draw an Entity-Relationship Diagram to show the relationships among a County, its Districts, and the Schools in the County. A County is divided up into Districts. Each District administers several Schools on behalf of the County.

(1) Draw an E/R diagram based on this situation which contains a "Fan Trap".
(2) Draw a second diagram which resolves the fan trap.

[4 marks]

C. Briefly define any **five** of the following terms.

(1) Super Key
(2) Composite Key
(3) Query Optimization
(4) Functional Dependency
(5) Determinant
(6) Null Value

[10 marks]

D. In the context of a distributed database, briefly define the terms "data replication" and "replication independence". What are the advantages of replicating data in a distributed database?

[4 marks]

E. In the last few years, database management systems have arisen which do not exclusively use the relational model. They are designed to deal with collections of data which do not easily fit into the relational "square data" model. Describe a situation where we might wish to consider an alternative to the relational model.

[4 marks]

TOTAL = 25 marks

Question 4

A. A retail sales company holds records on its customers, their memberships of the company's various "customer clubs", and their purchase history. A simplified example of some of the relations in their database is presented here.

Relation: CUSTOMER
Attributes:

PK: CustNum
Surname
AgeGroup — 6 levels … from Young to Elderly.
CrdImt — This Customer's Credit limit
Flag — Run double-check on card purchase if = T

CustNum	Surname	AgeGroup	CrdImt	Flag
C298731	Giap	3	2000	F
C765399	Locatelli	3	3000	F
C790008	Prohorov	5	3000	T
C7543871	Gadar	4	1000	F

Relation: MEMBER – Customers can join various "Customer Clubs" to get special privileges, such as early purchase rights; product testing; offers to buy slightly damaged products

Attributes:

PK:CustomerClub
PK:CustNum
RewardPts – how many "Reward Points" have been earned via that Club membership
FK: CustNum references CustNum in CUSTOMER

CustomerClub	CustNum	RewardPts
EARLYBIRDS	C298731	250
EARLYBIRDS	C765399	275
EARLYBIRDS	C790008	300
TESTERS	C765399	250
SHOPSOILED	C298731	210
SHOPSOILED	C765399	0
SHOPSOILED	C790008	180
BIRTHDAY	C790008	320

Relation: PURCHASE
Attributes:

PK:CustNum
PK: Product
PK: PurchaseDate
Cost

Note that each purchase line is for 1 unit only of the Product.
FK: CustNum references CustNum in CUSTOMER

CustNum	Product	PurchaseDate	Cost
C298731	JPHONE-8	2018-03-21	899
C298731	USB-SD ADAPTER	2018-03-21	12
C765399	USB-SD ADAPTER	2018-04-11	12
C765399	WIFI EXTENDER	2018-05-22	24
C790008	NETBOOK-3	2017-10-08	695

(1) What is the query that will list the Surnames, without repetition, of the Customers who have a Credit Limit greater than 1500?

[1 mark]

(2) What is the query that will list the Surnames of the Customers who are members of the EARLYBIRDS club, who have earned fewer than 250 Reward Points in that club?

[1 mark]

(3) What is the query that will list the Surnames of the Customers who are members both of the EARLYBIRDS Club and the SHOPSOILED Club?

[2 marks]

(4) What is the query that will list the Surnames of the Customers who are members of the EARLYBIRDS Club but *not* the SHOPSOILED Club?

[3 marks]

(5) What is the query that will list the average Cost of all purchases in 2017?

[2 marks]

(6) What is the query that will list the Surnames of the Customers who have the lowest Credit limit of all Customers?

[2 marks]

(7) What is the query that will list all the Products purchased by any Customer, and for each Product, the total number of times it was purchased?

[2 marks]

(8) What is the query that will list the Products purchased by at least 25 Customers, and for each one, the total number of times it was purchased?

[2 marks]

B. Modern database management systems contain a module called an "optimiser". What is the purpose of this module? Give two examples of how its purpose may be achieved.

[5 marks]

C. **(1)** What do we mean by a database "view"?
What kind of alterations to data can take place as the result of creating a view?

(2) What is the SQL for creating a view?

(3) What are two reasons for creating views?

[5 marks]

TOTAL = 25 marks

Question 5

The following table holds information about patients, the tests for diseases administered to them, and the nurses who administer the tests. It is the only table which holds this information.

A patient is only tested for a particular disease once, but may be tested for two or more different diseases on the same date. On any given date, the test for a particular disease (for all patients tested) is administered by just one nurse – this person may administer tests for more than one disease on the same date. PatientIDs and NurseIDs are unique – surnames may not be. Nurses have two days off a week, recorded in DaysOff.

Primary Key: **Patient + Test**

PatientID	Surname	Test	TestDate	Result	NurseID	DaysOff
P347769	Watanabe	Diptheria	2017-11-28	Neg	NA0338	Sun, Mon
P347769	Watanabe	Enteritis	2017-11-28	Neg	NA9883	Fri, Sat
P398876	O'Brien	Diptheria	2017-11-28	Pos	NA0338	Sun, Mon
P289877	Aguado	HNV1	2017-11-28	Neg	NA0338	Sun, Mon
P289877	Aguado	Diptheria	2017-11-28	Pos	NA0338	Sun, Mon
P966333	Imbari	Enteritis	2017-12-03	Neg	NA7221	Sun, Mon
P830732	Gonzales	Diptheria	2017-12-03	Neg	NA9883	Fri, Sat
P480887	Xu	Diptheria	2017-12-04	Neg	NA9883	Fri, Sat
P480887	Xu	HNV1	2017-12-04	Neg	NA7221	Sun, Mon

A. Identify the Functional Dependencies in this table.

[10 marks]

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

[6 marks]

C. When is a relation in "Boyce-Codd Normal Form" (BCNF)?

[1 mark]

D. Split the original table into tables in BCNF, specifying the Primary Key of each table.

[8 marks]

TOTAL = 25 marks

END OF PAPER