THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **CO2209 ZA**

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING and COMBINED DEGREE SCHEME

Database Systems

Date and Time: Monday 9 May 2016: 10.00 – 13.00

Duration: 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 hand held 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

A sewing-machine manufacturer receives sewing-machine components from suppliers, and then uses these components to assemble a number of different models of domestic sewing-machine. These sewing-machines are then sold to domestic goods shops.

In order not to become dependent on a single supplier, the manufacturer always makes sure that any given component-type is supplied by at least two different suppliers. A supplier may exist without yet supplying components. A supplier typically supplies several different types of component.

Each component-type can be used in the assembly of several different models of sewing-machine. A component only exists in the system if it is used in at least one model of sewing-machine.

A sewing-machine model is made up of many different types of component, but always at least one. A model may exist without yet being stocked by any shops.

A given domestic goods shop can receive and re-sell many different models of sewing-machine. No domestic goods shop has a monopoly on re-selling any given sewing-machine model.

A. Draw an Entity-Relationship Diagram to illustrate the relationships among sewing-machine-models, component-types, component-suppliers, and sewing-machine-shops. Include cardinality and participation constraints.

[5 marks]

B. Prepare a normalised relational schema which can record the relationships illustrated in **A.** Assume that sewing-machine-models are identified by Model-numbers, suppliers by Supnums, shops by Shop-names, and component-types by Compcodes. Indicate the primary keys, and foreign keys if appropriate, of each relation. Assume that there already exist 'master relations' for suppliers, components, models, and shops whose primary keys are Supnum, Compcode, Model-Number, and Shop-name, respectively. You may assume these relations are named Master-Supplier, Master-Component, Master-Model, and Master-Shop.

[5 marks]

C. Extend the relational schema in **B.** by showing a schema for a relation that could record how many of which kind of components were shipped to us, and by which supplier, and on what date, for

each delivery we received. For example, the relation should be able to record that Supplier S001 shipped 150 of Component C001 and 200 of Component C002 on 22 November 2015, and that Supplier S002 shipped 25 of Component C002 on 12 January 2016. Be sure to indicate the primary key of each relation.

[4 marks]

D. Suppose our arrangements with our suppliers changed, so that each type of component was supplied by *one* supplier only. (For example, C001 and C005 were supplied by Supplier S003 only, and C002 by Supplier S001 only.) Show the new schema which could record the same information as required by **B.** and **C.** Indicate the primary keys, and foreign keys if appropriate, of each relation.

[5 marks]

E. Suppose that our arrangements with suppliers changed once again, and we sourced each component type from more than one supplier, but we wanted to keep track of which supplier had supplied the particular component types that were used in each model. (We might want to do this if, as sometimes happens, one particular supplier makes certain components to a very high precision, say to 0.001 of a centimetre, while another makes the same component to a precision of only 0.01 of a centimetre. We might have 'luxury' models and 'standard' models, and want to use only high-precision components in the luxury models.)

For example, we might want to record that supplier S001 supplies component C001 to Models M1 and M3, and component C002 to Model M4, while supplier S2 supplies C001 to models M4 and M5, and supplier S003 supplies C001 to Models M1 and M3, and also supplies C002 to Model M4. (In this part of the question we only want to record the relationships among suppliers, components and models, without considering quantities or delivery dates.)

Could we use the relational schema you proposed in **B**, perhaps with new relations derived via SQL expressions, to record this data? If so, show what SQL expression(s) could do this. If not, show a new relational schema that could record this information. In all cases, be sure to note the Primary Key and any Foreign Keys if appropriate, remembering that there already exist 'master relations' for Models, Components, and Suppliers.

[6 marks]

Total = 25 marks

Question 2

A company specializing in the maintenance of computers and related devices stores, in a single table, information about the repairs that its technicians have carried out on its computers and other electronic devices.

It records the serial number of the equipment repaired, a one-word description of the equipment, the employee number and name of the technician who carried out (or supervised) the repair (no more than one name will be recorded), the date the repair was begun, brief notes on the nature of the problem, and the technician's pager number.

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

SerialNo	RepairDate	Description	Technician	Name	Notes	Pager
R299822	11-10-2015	Router	P32014	Sakhel	Dropped packets	#932
PS993301	23-05-2015	Power supply	P32014	Sakhel	Fuse replacement	#932
NB39393	23-12-2008	Netbook	P88317	Alice	Keyboard faults	#290
NB99019	01-09-2015	Netbook	P88317	Alice	Cracked screen	#290
HD30022	13-12-2015	BackupHD	P93858	Ksenia	Fuse replaced	#998
DT83298	23-10-2015	Desktop	P88317	Alice	Boot-up failure	#290
DT83298	23-11-2015	Desktop	P32014	Sakhel	Replaced HD	#932
DT40332	03-12-2015	Desktop	P32014	Sakhel	Sticky Keys	#932

For instance, the table records that on the 11th of October in 2015, a technician named Sakhel, whose employee number is P32014, and whose pager number is #932, repaired a Router, serial number R299822, which was having a problem with dropped packets.

A. Assuming that no device has more than one failure per day, what is the primary key of this table?

[1 mark]

B. Identify the functional dependencies in this table.

[5 marks]

C. This table is susceptible to insertion, deletion and modification anomalies. Give an example, based on the table, of each kind. Assume that no other tables recording information on equipment or technicians exist. Assume that the data you see in the table is all the data the table holds.

[6 marks]

D. Bring the data in this table to BCNF, specifying the specifying the primary and foreign keys where appropriate, of each table.

[9 marks]

E. The company's database has a table, with hundreds of thousands of entries, which records details of its customers. Three of the columns of this table are as follows:

Street	City	PostCode
234 Elm St	Smallville	SM456
249 Elm St	Smallville	SM456
276 Elm St	Smallville	SM457

The company accesses this database frequently, to send out mailings to its customers. A data analyst has pointed out that the table is not fully normalized, since there is a functional dependency from PostCode to City. (In other words, PostCode functionally determines City). However, he recommends that the table not be split into two BCNF tables. What argument might he have made to justify his recommendation?

[4 marks]

Total = 25 marks

Question 3

A. Briefly describe horizontal fragmentation.

When might a database designer want to implement horizontal fragmentation?

Use the following relation, of which a small sample is given, as an example, and show how it might be horizontally fragmented, if we found that almost all urgent queries were of the sort, ***SELECT*** *PartNum* ***WHERE*** *Warehouse = Main-1 , or* ***SELECT*** *PartNum* ***WHERE*** *Warehouse = MAIN-2 or* ***SELECT*** *PartNum* ***WHERE*** *Warehouse = MAIN-3* – that is, almost all queries to return particular PartNums were for tuples whose Warehouse value was MAIN-1, MAIN-2 or MAIN-3 .

Assume that the relation has 2,500,000 tuples, of which about 50,000 have Warehouse equal to Main-1, Main-2, or Main-3. The Primary Key is PartNum.

PartNum	Name	Warehouse	Unit Price
P397303-8	hammer	Main-1	14.50
P476870-3	saw	Main-3	12.00
P973066-4	Grommet inserter	Aux-24	22.00
P398907-2	Flutter valve	Main-4	45.00

[5 marks]

B. Describe the problem of "deadlock" (also known as "the deadly embrace") and discuss ways in which the problem might be solved.

[5 marks]

C. In the context of a database, what is a Data Dictionary (or System Catalogue)? What sort of information is it likely to contain? How would it be used in querying and updating the database?

[5 marks]

D. Briefly define the term “data replication” and “replication independence” in the context of a distributed database. Why do we sometimes replicate data in a distributed database?

[5 marks]

E. Show with an example, including a drawing, how the ‘data access protocol’ can be applied to provide a solution to a concurrent access problem (either the lost update, uncommitted dependency or inconsistent analysis problem).

[5 marks]

Total = 25 marks

Question 4

A. How does SQL differ from standard programming languages, such as Java, PHP or Python?

[4 marks]

B. "The database designer will always be able to find at least one set of attributes which make up any relation's Primary Key."

Is this statement correct? Can there be a relation which has no Primary Key? (Note: this does not refer to whether or not a set of tuples is – correctly or incorrectly – *designated* as the Primary Key by the person designing the relation, but rather whether or not a Primary Key always exists in principle.)

[3 marks]

C. The concepts of Primary Keys, and of Foreign Keys, are often linked to the idea of logical database 'integrity'. In what way do they constrain database integrity? Can the database designer enforce any other kinds of constraint on the database at the logical level, and if so, how? (Note that this question does not refer to 'physical integrity' maintenance via such procedures as backups, concurrent access mechanisms, encryption, or access controls.)

[8 marks]

D. Write brief definitions of any **five** of the following terms to show how they are used, as they apply to relational databases. Only the first five answers will be marked

(1) Relation
(2) Tuple
(3) Intension
(4) Extension
(5) Candidate Key
(6) Compound (or composite) Key
(7) NULL value
(8) Functional Dependency
(9) Determinant

[10 marks]

Total = 25 marks

Question 5

A medicinal herbs company has implemented a database describing its products. Each distinct herb it sells has a unique identifier, HRB-ID, a NAME, a single MEDCODE, and should have -- but doesn't always -- a particular employee who possesses specialist knowledge of that herb who has been designated the 'go to' person for answering customer inquiries about it.

It also wants to record which customer has ordered how much of which product, and the date of the order. A customer never places more than one order for the same product on the same date.

Relation Name: **PRODUCTS**

Attributes:

HRB-ID	Primary Key. A whole number ranging from 0 to 9999.
NAME	A string of characters, from 4 to 36 characters long. Note that more than one product may have the same NAME.
MEDCODE	A whole number, ranging from 1 to 3. This denotes the level of medical warning that accompanies the herb: 1 means completely harmless even if taken in excess, 2 means harmful in excess, 3 means extremely hazardous - not to be taken without a doctor's approval.
EMPNUM	A number, ranging from 0 to 500 000. This field may be empty, in which case we put the keyword NULL there.)

HRB-ID	NAME	MEDCODE	EMPNUM
3458	Beetroot ext.	1	556698
8762	Powdered algae	1	342554
8971	Ginger root	2	342554
9230	Gingko Biloba	1	354211
9231	Gingko Biloba	2	354211
9377	Goji berries	1	NULL
9498	Blueberry ext.	1	354211
9599	Devil's Claw	3	198887

Relation Name:	**ORDERS**	Primary Key: CUSTNUM + HRB-ID + DATE
Attributes:	**CUSTNUM**	A whole number, ranging from 0 to 999999
	HRB-ID	A whole number, ranging from 0 to 9999; this is a Foreign Key which references HRB-ID in PRODUCT
	DATE	The date the order was processed
	QTY	A whole number, must be greater than 0

CUSTNUM	HRB-ID	DATE	QTY
458876	3458	2016-03-21	5
458876	3458	2015-10-12	5
458876	9377	2015-10-12	8
469855	9377	2015-10-12	12
434590	9498	2016-02-09	15
687744	3458	2016-03-21	2
458876	9230	2016-03-21	5
687744	3458	2016-05-12	10

A. Write the SQL expressions that would create these two tables. You do not have to write the statements that would populate them with data.

[6 marks]

B. Write the SQL expressions that would answer the following queries:

1. List the customer numbers of customers who placed orders for the herb whose HRB-ID is 9377.

[1 mark]

2. List the names of all herbs with a MEDCODE greater than 1, sorted on HRB-ID.

[1 mark]

3. How many different products are in the PRODUCTS table?

[1 mark]

4. List the HRB-IDs of all herbs where the employee expert is unknown or missing or has not been assigned yet.

[1 mark]

5. List all HRB-IDs of herbs with 'root' in their name.

[1 mark]

6. List the customer numbers of all customers who have ordered any herb with a MEDCODE greater than 2.

[2 marks]

7. List the total quantity of all herbs ordered after January 1st, 2016.

[2 marks]

8. What is/are the NAME(s) of the herb(s) we stock with the most hazardous MEDCODE. (Note that it is possible from time to time that no herb is stocked with the MEDCODE equal to 3.)

[2 marks]

9. List the customer numbers and total quantities of all herbs with MEDCODE >=2 ordered by each customer.

[2 marks]

10. List the customer numbers and total quantities of all herbs with MEDCODE >=2 ordered by each customer in 2016.

[2 marks]

11. List the CUSTNUMs and total quantities of all herbs ordered by each customer with MEDCODE >= 2 where this total is greater than 5.

[2 marks]

12. List the HRB-IDs of herbs that no one has ever ordered.

[2 marks]

Total = 25 marks

END OF PAPER