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

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 company wants to set up a database to coordinate its business, which is managing absentee-owned flats, houses and commercial property. These are all called 'properties'. Its business involves collecting the rents from the tenants, and overseeing routine maintenance and repairs of the properties. In return for doing this, the company gets a percentage of the rent collected.

The company is divided into several branch offices, each of which is responsible for managing properties within the geographical area where the office is located.

Employees of the company oversee properties. A given property will have only one employee overseeing it, but an employee may oversee many properties. Some employees may not oversee any properties.

All properties for which the company is responsible must be registered with the branch office in their area, but may not yet be assigned to a branch office employee to oversee them.

Properties are owned by owners. A property will have only one owner (and must have one), but an owner may own more than one property.

Properties are rented by a tenant who is legally responsible for paying the rent and for paying for any tenant-caused damage to the property. There will be only one tenant for a given property, but a given tenant may rent more than one property. The company wants to record information about properties, their owners, their tenants, branch offices, and employees.

A. Draw up an Entity/Relationship Diagram, which shows the entity types described above, and the relationships among them. Include cardinality (1:1, 1:N, N:M) information and participation (optional or required) constraints. Do not include attributes.

[13 marks]

B. Draw up a relational schema, which can record the information presented above. Be sure to indicate both primary keys and foreign keys. Your schema should include as few relations as possible.

You may assume that employees are identified by their Employee Numbers. In addition to the branch office they work for and the properties they manage, we need to record their Surnames, FirstNames, and DateHired.

Tenants are identified by Tenant ID Numbers. In addition to their tenancy of properties, we need to record their Surnames and Firstnames. For each tenancy, we need to record the date that it began.

Properties are identified by Postal Addresses. A property has a (unique) Postal Address and a description, and a rental amount.

Owners are identified by Owner Codes. We need to record the Owner's Surname and First Name.

Branch offices are identified by Branch Office Names. We need to record the address of each branch office.

[12 marks]

Total = 25 marks

Question 2

The following table records information about the percentage of ground cover found in certain uniquely numbered plots of land, which are being observed over time to track the growth of vegetation on them.

Each plot is found in a given district, and no plot overlaps districts. Each plot's size in hectares is recorded.

Roughly every two years, an on-site inspection is done to record the amount of ground cover growing on the site. The date of the inspection, and the identification number of the inspector, is recorded, along with the inspector's telephone number.

(Note that newly-hired Inspector R409 has, so far, only inspected one plot.)

PlotNo	District	Hectares	Inspection Date	Inspector	Ground Cover (%)	Phone#
2455	West Lake	32.5	12 Dec 2006	M330	34	7633088852
2455	West Lake	32.5	23 Oct 2008	P528	32	7609865463
2455	West Lake	32.5	7 Nov 2010	M330	33	7633088852
2455	West Lake	32.5	8 Jan 2012	P301	35	7682907965
2455	West Lake	32.5	29 Nov 2014	P301	36	7682907965
2455	West Lake	32.5	02 Jan 2016	R409	36	8097733027
3891	West Lake	12.3	12 Dec 2009	P301	38	7682907965
3891	West Lake	12.3	10 Dec 2010	P528	36	7609865463
3891	West Lake	12.3	23 Oct 2012	P301	36	7682907965
3891	West Lake	12.3	30 Jan 2014	P301	38	7682907965
3891	West Lake	12.3	5 Jan 2016	M330	37	7633088852
2349	Bluestone	46.0	22 Sep 2008	P301	45	7682907965
2349	Bluestone	46.0	19 Jan 2010	S300	46	7629920821
2349	Bluestone	46.0	24 Sep 2011	P301	46	7682907965
2349	Bluestone	46.0	8 Nov 2013	P528	45	7609865463
2349	Bluestone	46.0	24 Nov 2015	P528	48	7609865463
4892	Bluestone	32.5	22 Sep 2012	S300	36	7629920821
4892	Bluestone	32.5	30 Feb 2014	S300	36	7629920821
4892	Bluestone	32.5	18 Dec 2015	P301	37	7682907965

A. What is the primary key of this table?

[2 marks]

B. Identify the functional dependencies in this table.

[3 marks]

C. This table is susceptible to insertion, deletion, and insertion anomalies. Give an example, based on the table, of each kind.

[6 marks]

D. Recast this table into BCNF, specifying the primary keys, and foreign keys if appropriate, of each table. You need only show the attribute names.

[6 marks]

E. Consider the following table, which records the examination results for students who have taken a particular subject. Students who fail a subject are not allowed to re-sit it.

STUDENT-RESULTS

SNUM	SUBJECT	EXAMDATE	MARK
454	History I	23-05-08	65
454	French I	17-06-08	74
538	History I	23-05-08	52
654	Psychology II	12-06-07	57

The primary key of this table is SNUM + SUBJECT.

1. What bad consequences could follow if we misidentified the primary key of the table as
   a) SNUM alone?
   b) SNUM + SUBJECT+ EXAMDATE?

[4 marks]

2. Suppose it is decided to allow students to re-sit exams in a subject which they have failed, in a following year. Would we need to change the definition of the primary key? Explain your answer.

[1 mark]

F. Consider the following table, which records the books and articles written by lecturers. Lecturers are identified by Employee numbers, books by ISBNs, and articles by Serial Item and Contribution Identifiers (SICI). Note that the lecturer whose employee number is P22234 has written one book and two articles, and the lecturer whose employee number is K39423 has written two books and one article.

LECTURER-DETAILS

Lecturer	Book	Article
P22234	1 84195 525 6	1046-8188(199501)13:1<69:FTTHBI>2.0.TX;2-4
P22234	.NULL.	0002-8231(199601)47:1<23:TDOMII>2.0.TX;2-2
K39423	0 86104 068 6	0095-4403(199502)21:3<12:WATIIB>2.0.TX;2-J
K39423	978 1 84489 416 1	.null.

Although this relation does not violate the rule of thumb for normalization "let every determinant be a candidate key" – yet it is a very poor design. Explain why, and suggest an alternative way to represent the information held by it.

[3 marks]

Total = 25 marks

Question 3

A. Briefly describe vertical fragmentation.

When might a database designer want to implement vertical fragmentation?

Use the following relation, of which a small sample is given, as an example, and show how it might be vertically fragmented, if we found that almost all urgent queries were of the sort, *SELECT PartNum WHERE PackingType =[some value], or SELECT PartNum WHERE Location =[some value]* – that is, almost all queries to return particular PartNums were on the attributes Location and/or PackingType.

Assume that the relation has 500,000 tuples. Note that Name and Description are of type STRING, and that values in the Description attribute average 170 bytes in size.

PartNum*	Name	Description	Location	PackingType
65641090	"Saw Blade"	"Carbon steel blade replacement for Tomahawk saws."	WH05	P3
93193648	"Marine anchoring device"	"Double-helical fastening device with copper-plated head"	WH04	P2
55901637	"Deck covering, weather-proof"	"Rubberized canvas tarpaulin with grommet-holes for attaching to deck lugs."	WH93	P3
94134175	"Propeller shaft collar."	"Aluminium collar for propeller shaft, with self-guided machine screws for attachment to motor body."	WH82	P5

[6 marks]

B. 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?

[6 marks]

C. In the context of a distributed database, briefly define the term "data replication," and "replication independence". Why is data sometimes replicated in a distributed database?

[5 marks]

D. Using an example of your own choice, show how a query optimiser can increase the efficiency of query execution in a distributed database where a query involving a selection and a join must unite tuples from two different relations, with very different cardinalities, located at different geographical locations.

[8 marks]

Total = 25 marks

Question 4

A. A database designer is designing a database for a hospital, and wants to design a simple relation, which can store notes taken by a doctor who examines a patient on multiple dates. A patient is never examined more than once on a given date.

The relation must contain the patient's (unique) PatientNo, the date of the examination, and notes made by the examining doctor.

Typical queries on this relation might be 'find all patients who were examined on a given date'.

The designer was told that the average number of examinations per patient is three and therefore proposed the following design, where PatientNo would be the Primary Key.

PatientNo	Exam1 Date	Exam1 Notes	Exam2 Date	Exam2 Notes	Exam3 Date	Exam3 Notes	Exam4 Date	Exam4 Notes	Exam5 Date	Exam5 Notes

1. Comment upon this design, and in particular what problems it might pose.
2. Propose an alternative design, which would not have these problems, being sure to specify the PrimaryKey.

[7 marks]

B. Write brief definitions to show what we mean by any **five** of the following terms, as they apply to relational databases. (Note: if you use the verb 'determines' in your answer, be sure to explain what you mean by this.) Only the first five answers will be marked.

(1) NULL value
(2) Attribute integrity
(3) Compound (or composite) Key
(4) Referential integrity
(5) Candidate Key
(6) Entity integrity
(7) Foreign Key
(8) Functional Dependency
(9) Determinant

[10 marks]

C. Although the relational model of data, and database systems which implement it, remain the industry standard, there are alternative approaches to database design. What sort of situations might lead designers to consider an alternative to the use of a relational database? Briefly describe one or more of these alternatives.

[8 marks]

Total = 25 marks

Question 5

A company, which makes replacement parts for cameras, has a database holding information on them. Each distinct camera-part it sells has a unique PARTNO, and as well has a PARTNAME, a WEIGHT, and a STOCK-COUNT.

The company also records which customer has ordered how much of which part, and the date of the order. A customer never places more than one order for the same part on the same date.

It holds this information in the following two relations, which are shown with some example data.

Relation: PARTS

Attributes:	**PARTNO**	Primary Key. A number, ranging from 1 to 999.
	PARTNAME	A string of characters, from 4 to 36 characters long. PARTNAMEs are not necessarily unique.
	WEIGHT	A number, ranging from 1.0 to 9999.9. This denotes the weight in grams of the part, rounded to one decimal place.
	STOCK-COUNT	A number, ranging from 0 to 10 000.

PARTNO	PARTNAME	WEIGHT	STOCK-COUNT
998	Shutter-sleeve	5	275
943	Battery mount	6	222
883	Lens adjuster	6	208
745	Viewfinder	12	345
511	Viewfinder	15	136
884	Lens replacement	15	NULL

Relation: **ORDERS** Primary Key:CUSTNUM + PARTNO + DATE

Attributes:

CUSTNUM	A whole number, ranging from 0 to 999999
PARTNO	A whole number, ranging from 1 to 999; this is a Foreign Key which references PARTNO in PARTS
DATE	The date the order was processed.
QTY	A whole number.

CUSTNUM	PARTNO	DATE	QTY
458876	998	2016-03-21	1
458876	998	2015-10-15	1
458876	486	2015-10-15	4
469855	486	2015-10-15	2
434590	505	2016-02-09	1
687744	998	2016-03-21	3
458876	745	2016-03-21	1
687744	998	2016-03-23	5
469855	P486	2015-12-10	2

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. Create SQL queries to answer the following questions:

1. List the customer numbers of customers who placed orders on the 12th of October 2015.

[1 mark]

2. How many different parts weigh less than 10 grams?

[1 mark]

3. List the part numbers of all parts where the STOCK-COUNT is unknown or missing.

[1 mark]

4. List the names of all parts, sorted by part number. Do not list any names twice.

[1 mark]

5. List the customer numbers of all customers who have ordered any part with a weight greater than 4, sorted by customer number.

[2 marks]

6. List the total quantity of all parts ordered on or after the 1st of January, 2016.

[2 marks]

7. Find the Part number and name (only) of the heaviest part we stock.

[2 marks]

8. List the customer numbers and total quantities of all parts weighing more than 20 grams, which were ordered by each customer.

[2 marks]

9. List the customer numbers and total quantities of all part weighing more than 20 grams, which were ordered by each customer in 2016, sorted to show the order with the greatest total quantity first, down to the order with the smallest quantity.

[2 marks]

10. List the customer number and total quantity of all parts weighing more than 20 grams ordered by each customer where the total quantity for that customer is greater than 3.

[2 marks]

C. Suppose you had a complex query to put to a very large database, where you were not sure that your query was correct. When you try it, your query returns a result, in the form of a relation with thousands of tuples, but you are not sure that it is the correct result.

What could you do to test your query and then correct it if it was wrong?

[3 marks]

Total = 25 marks

END OF PAPER