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 8 May 2017: 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 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

A. The following database records information about certain deliveries of prepared meals from fast-food suppliers to construction projects. It records the name of each Supplier, its customer rating, the date it was approved, and its delivery fee. For each Project the database records the kind of construction that is underway, the postal code where it is located, its starting date, and its manager. Each Delivery is recorded, showing the delivery date and quantity of meals supplied. The tables are shown below, with some sample data.

SUPPLIER
Primary Key: **SUPCODE**

SUPCODE	SNAME	RATING	DATEAPPROVED	FEE
S001	Meals2U	Average	2014-02-21	7.50
S002	GourmetGrub	Extra	2015-11-06	9.00
S003	LunchToYou	Extra	2014-10-12	7.25
S004	QuickMeals	Average	2014-09-02	8.25
S005	SwiftSoups	Extra	2016-02-12	9.00

PROJECT
Primary Key: **PROJNUM**

PROJNUM	PNAME	LOCATION	STARTDATE	MGR
P01	Hotel	GU3	2017-01-23	B. Jones
P02	Restaurant	GU5	2017-02-03	M. Iqbal
P04	Office Block	GU3	2016-12-21	B. Murphy
P05	Leisure Plaza	GU6	2016-11-09	NULL
P07	Hotel	GU4	2017-02-13	B. Jones
P09	Office Block	GU5	2016-11-21	J. Locke

DELIVERY
Primary Key: **PROJNUM + SUPCODE**

PROJNUM	SUPCODE	DELIVERYDATE	QUANTITY
P04	S003	2017-01-11	12
P07	S004	2017-01-04	15
P04	S003	2016-11-21	25
P01	S001	2016-12-18	20
P07	S001	2017-02-21	35
P09	S004	2017-02-11	18

Construct queries in SQL to answer the following queries:

1) How many projects do not have a manager recorded in the database?

[1 mark]

2) Get the average delivery fee.

[1 mark]

3) Get the project names of projects managed by B. Murphy which have had deliveries from supplier S004.

[1 mark]

4) Find the project names of projects which have had deliveries from suppliers whose fee is more than £8.50.

[1 mark]

5) What is the name (or names) of the supplier(s) with the lowest delivery fee?

[2 marks]

6) Get the supplier codes and total quantities of deliveries from each of the suppliers.

[2 marks]

7) Get the supplier codes and total quantities delivered by each of the suppliers where the total delivery quantity is greater than 100.

[2 marks]

8) Get the supplier codes and total quantities delivered by each of the suppliers that have made deliveries in any year after 2015, where the total delivery quantity is greater than 100.

[2 marks]

9) What are the project numbers of projects which have had deliveries, but have not had a delivery in 2017?

[3 marks]

B. Most databases are indexed. What is an index? What is its purpose? What sorts of attributes are most often indexed? What are the disadvantages of indexing? (You need not describe any particular kind of index.)

[5 marks]

C. Discuss briefly the problems involved in implementing a 'family tree' database, consisting of several pairs of ancestors, each with one or more descendants, who in turn have descendants, and so on.

[5 marks]

Total = 25 marks

Question 2

Consider the following relation, which contains information on publicity tours by authors, organized by their publishers, to publicize the books they have written, via interviews on local radio stations.

BOOKTOURS PRIMARY KEY: ISBN+INTDATE

ISBN	AUTHOR	TITLE	PRICE	PUBLISHER	MAINOFFICE	INTTOWN	INTDATE
9254062218	Cindy Airy	Fast Lane	$24.99	Vanguard	Bourne	San Remo	08.11.2017
9254062218	Cindy Airy	Fast Lane	$24.99	Vanguard	Bourne	Fitzroy	02.07.2017
9254062218	Cindy Airy	Fast Lane	$24.99	Vanguard	Bourne	York	03.07.2017
7738144309	Bill Posters	Trumpets of Dumb	$19.99	Forward	Canberra	San Remo	08.11.2017
0224062530	Bill Posters	Start Procrasti nating Now!	$14.99	Vanguard	Bourne	Leeds	04.07.2017
3149986280	Geri Attick	How to Quit Tweeting	$15.95	McBane	Bourne	Bridger	04.07.2017
9254062218	Cindy Airy	Fast Lane	$24.99	Vanguard	Bourne	Darwin	06.07.2017
0224062530	Bill Posters	Start Procrasti nating Now!	$14.99	Vanguard	Bourne	Glasgow	09.07.2017
7738144309	Bill Posters	Trumpets of Dumb	$19.99	Forward	Canberra	Glasgow	10.07.2017
9254062218	Cindy Airy	Fast Lane	$24.99	Vanguard	Bourne	Chester	09.07.2017

An author gives, at most, only one publicity interview per day. An author may give more than one interview in a given town (identified by the attribute INTTOWN in the relation above) about the same book, on different days.

An interview is about only one book (even if the author is on tour to publicize more than one). A book has only one author. A particular book has only one ISBN. (The ISBN uniquely identifies the book.)

A publisher has a Main Office in only one town. An author can write more than one book. Titles should not be assumed to be unique (that is, two different books may have the same title.)

A. Identify the Functional Dependencies in this table.

[7 marks]

B. This table is susceptible to insertion, deletion, and update anomalies. Give an example, based on publishers and their main offices, of each kind.

[6 marks]

C. 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 BCNF, specifying the Primary Keys of each new table, and showing the extension of the resulting relations.

[6 marks]

E. Consider the following relation, which records an author's 'genres' and the foreign languages which they speak well enough to be interviewed in. Cindy Airy writes only Mysteries, and can be interviewed in Spanish and French. Bill Posters writes in three genres, but can only be interviewed in Spanish. Geri Attick writes in two genres but cannot be interviewed in any foreign language.

AUTHOR	GENRE	LANGUAGES
Cindy Airy	Mysteries	Spanish
Cindy Airy	Mysteries	French
Bill Posters	Horror	Spanish
Bill Posters	Thrillers	Spanish
Bill Posters	Romance	Spanish
Geri Attick	Romance	
Geri Attick	Thrillers	

This relation does not violate Boyce-Codd Normal Form, since it has no determinants. Yet it is a poorly-designed table.

Examine this table and state any problems you see with it. Propose a better relational schema (which may involve new tables) that can record the same information without the problems this table has.

[4 marks]

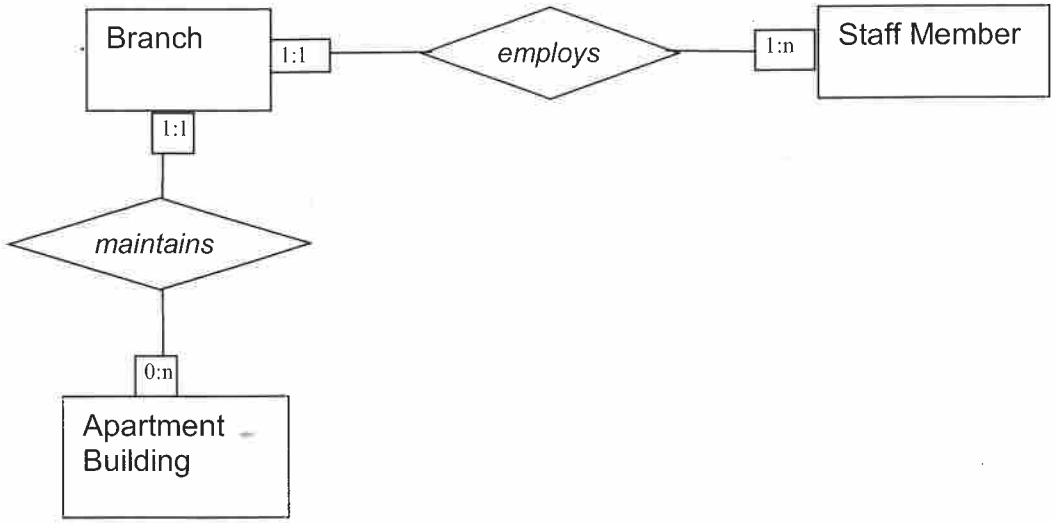
Total = 25 marks

Question 3

A. Consider a situation where a company is responsible for the maintenance and repair of Apartment Buildings. The company has a number of Branches, each of which is responsible for maintaining several Apartment Buildings in its area. Each Branch employs several maintenance Staff, each of whom is responsible for overseeing one or more Apartment Buildings.

The relationships have been modelled this way. A Branch, represented by a BranchNum, employs one or more Staff Members, represented by EmpNums, and a Staff Member must be employed by a single Branch. A Staff Member can oversee several Apartment Buildings, represented by PropNums, but may not yet have been assigned any. An Apartment Building is overseen by only one Staff Member, and will always have one assigned to oversee it.

An analyst has modelled this situation with the following Entity-Relationship Diagram.



The analyst has proposed the following relational schema to implement this Entity-Relationship model.

EMPLOYED-BY
Primary Key: EmpNum
Other Attributes: BranchNum

MAINTAINED-BY
Primary Key: PropNum
Other Attributes: BranchNum

The analyst asserts that this model will always allow us to see which Staff Members work for which Branches, which Apartment Buildings are maintained by which Branches, and, via a simple join, which Apartment Buildings are overseen by which Staff Members. (All of these links have been requested by the client for whom the database is being built.)

However, there is a flaw in this Entity-Relationship diagram and the proposed Relational Schema. What is it, and how can we repair the Entity-Relationship diagram and the Relational Schema to fix it? Draw a better Entity-Relationship Diagram and show a correct Relational Schema.

[6 marks]

B. What is "Brewster's Conjecture" (later proved, to become the "CAP Theorem")?

[5 marks]

C. Some large collections of data consist of 'documents', which do not easily fit the traditional relational 'square' model of domains, attributes, and tuples. How can these documents be structured? In what other way does a collection of documents not resemble the sort of data for which the relational model was designed? (You may wish to illustrate your answer with a short example.)

[5 marks]

D. In a distributed database, a table may be 'horizontally fragmented'. What does this mean, and why would it be done? Illustrate your answer with an example taken from the case of a global company with branches in many countries, which has a database with one logical table holding data on its employees.

[5 marks]

E. Consider a website which allows users to type phrases into a search box and then searches a database for the phrase. One threat to such a site is known as 'SQL injection'. Briefly describe this method of attack and outline one method of dealing with it.

[4 marks]

Total = 25 marks

Question 4

A newly-formed ferry company will run ferries from several ports in England, to several ports on the European continent. Its database needs to record information about its Ferries, Connections, Voyages and Customers.

Connections are regularly-scheduled sailings. They are identified by Connection Numbers. Each Connection has a Departure Port and a Destination Port, a Departure Time and an Arrival Time, a Person-Ticket-Price and a Personal-Ticket-with-Automobile-Berth Price and the days of the week on which the connection will operate. (Some people travel by foot, others bring their cars.)

A Customer is given a unique Customer-Number, has a Family-Name and First-Name, and may have a Passport Number. A Customer can book one or more connections, each for a specific date. A given connection on a particular date is called a *voyage*.

Each Voyage will have many Customers, and occur on one particular ferry which will have a single Captain. A Customer's Booking of a Voyage will be for either a personal seat alone, or a personal seat plus an automobile berth. A given Connection is in one direction between two cities. (Return journeys have different connection numbers.)

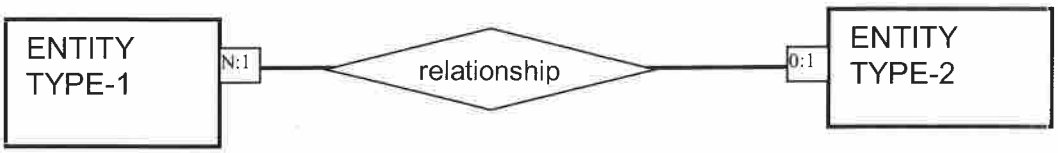
For example, Connection DC05 takes place between Dover and Calais, sails at 9.30 am on Monday through Saturday, arriving in Calais at 15.00, and is always made by a SeaMaster-B type ferry. The ferry company owns several SeaMaster-B ferries, and any one of them may be used on this Connection on a particular date. A Captain may switch between ferries, commanding one vessel on Monday and a different one on Tuesday.)

Ferries will be a certain Model (such as a SeaMaster-B), be identified by a unique Vessel-Identifier, and will have a certain number of seats and a certain number of Automobile berths. (These may differ slightly from Ferry to Ferry, even among Ferries of the same model.)

Ferries, Captains and Connections can exist without having yet been involved in a Voyage, but Customers must have made at least one Booking for a Voyage. A Voyage can exist before any Bookings have been made for it.

An example of a typical transaction might be that Customer K3358, Chris Yan, Passport Number 98077374 booked a place for one person on Connection DC05 for the 17th of March 2017. On this Voyage Chris Yan was allocated Seat 5 in Row K (Recorded as 'K5'). The Ferry used for that particular Voyage had the Vessel-Identifier UKF6622, and was a SeaMaster-B, commanded by Captain Andrew Veil, whose Employee Number is U998A. This particular vessel had room for 75 automobiles and 450 passengers.

A. Draw an Entity/Relationship diagram to represent this situation. Include only the entity types and relationships, but not the attributes. Use the following conventions to show Entity-types, relationships, and cardinality/participation constraints.



This illustrates a situation where 'Entity-type-1' can have a relationship to zero, or one (but no more than one) instance of 'Entity-type-2', and where '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 normalized relational schema which could hold the information represented by the E/R diagram you drew up for **Part A.** (The previous description will give information about the attributes you will need, which you did not have to include on the E/R diagram.) Populate it with the example data given in **Part A**. Be sure to indicate both Primary and Foreign Keys.

[15 marks]

Total = 25 marks

Question 5

A. A 'transaction' in the database sense can be defined as a sequence of database operations (select/insert/update/delete) that must form a single, 'logical' unit of work. (The 'logical' means that from the human point of view, the whole sequence must be completed for the intended operation to be valid.)

The classic example of a transaction in this sense is the sequence of operations involved when someone moves an amount of money from their 'current account' (upon which they can write cheques) into their 'savings account' (upon which cheques cannot be written, but which earns interest).

This involves reading the current account's balance, deducting an amount from it, and then increasing the savings account by the same amount.
The aim of the database system in successfully completing a transaction can be summed up in the acronym 'A.C.I.D.'.

Briefly describe what each of these letters stands for, and illustrate each description by referring to the bank transaction mentioned above.

[8 marks]

B. Most database operations take place on data which has been loaded into a 'buffer'.

1) Briefly define the following terms, in the context of databases:

a) Buffer
b) Log
c) Checkpoint

2) Transactions held in a buffer are vulnerable to a system crash.

Describe how the method of logs and checkpoints deals with this problem, allowing the system to recover from a crash that occurs after a checkpoint. (In this context, a 'crash' means a 'soft crash', that is, one that does not damage non-volatile storage.)

[8 marks]

C. Briefly define each of the following terms, in the context of their use when referring to databases:

1) Null value
2) View
3) Determinant
4) Foreign Key
5) Boyce-Codd Normal Form

[5 marks]

D. What is meant by a "distributed database system"?

From the point of view of data duplication, what are the possible architectures of such a system? What are the advantages and disadvantages of duplicating data at each site of a distributed database system?

[4 marks]

Total = 25 marks

END OF PAPER