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

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

A company has a large central warehouse near a port city, where bulk shipments of textiles are brought by ship. From time to time shipments of textiles are sent out to local centres in different parts of the country, where they are used to make garments.

The company's database has tables which record all the various kinds of textile which are in the central warehouse, and also the textiles which are in its local centres around the country.

In response to the arrival of a shipment of textiles at the central warehouse, the database is updated to reflect the increase in the kind of textile which has come in.

When textiles are shipped out of the warehouse to local centres, the warehouse table is updated to reflect the decrease in the amount of textile held there. The textiles are listed as 'in transit'. When the textiles arrive at a local centre, the tables which record how much textile each local centre is stocking are updated to reflect a corresponding increase. Thus, the company always knows how many textiles of each type it has, and where they are located.

(1) A new shipment of several kinds of textile arrives. The central warehouse table is updated: each tuple corresponding to a particular kind of textile has its 'On-Hand' field increased by the amount of that type of textile which has arrived.

(2) When a local centre requests a shipment of a certain quantity of a certain kind of textile from the central warehouse, that textile's tuple is accessed, the 'On-Hand' field is decreased by the shipment amount, and the 'In-Transit' field is increased by the same amount. When the receiving local centre confirms that the shipment has arrived, the value of the 'In-Transit' field is decreased and the 'On-Hand' field in that kind of textile's tuple in the 'LocalCentre' table is updated to show the increase.

A sample of the relevant attributes of the relations might look like this:

CentralWarehouse

Primary Key: TexCode

TexCode	Description	On-Hand	In-Transit
BO332	Bombazine	230	0
BA054	Batik	0	300
BA068	Batik	250	750
DR888	Drugget	150	125
GA612	Gazar	235	0
HM091	Himroo	550	100
FD482	Foulard	50	0

LocalCentre

Primary Key: CentreNum + TexCode

CentreNum	TexCode	On-Hand
453	BO332	100
453	BA054	0
453	BA068	25
564	BO332	15
564	BA054	235
564	HM091	55
765	BA068	50

Here is a typical set of SQL statements, labelled --1 to --7, that the company's Database Administrator might enter during the course of a day, to keep the database up to date.

/* A telephone call from the Unloading Dock tells us that 500 metres of DR888 has just been unloaded and inspected and is being stored in the Central Warehouse Update the 'OnHand' field at the Central warehouse to show this */

```
--1
UPDATE CentralWarehouse
SET On-Hand = On-Hand + 500
WHERE TexCode = 'DR888';
```

/* We receive a note from the Shipping Department informing us that 100 metres of GA612 has been loaded onto trucks for delivery to our local centres in other parts of the country. Update the 'OnHand' and 'InTransit' fields of the Central Warehouse to reflect this change. */

```
--2
UPDATE CentralWarehouse
SET OnHand = On-Hand - 100, In-Transit = In-Transit+100
WHERE TexCode = 'GA612';
```

/*An email tells us that 100 metres of BA068 has arrived at one of the local centres, 453, to which it was shipped. Adjust the database accordingly. Update the 'In-Transit' field of the Central Warehouse, to indicate that 100 metres of this fabric is no longer in transit ...*/

```
--3
UPDATE CentralWarehouse
SET In-Transit = In-Transit -100
WHERE TexCode = 'BA068';
```

/* ... and Update the 'On-Hand' field of the local centre whose code is 453, to indicate that 100 metres of BA068 has arrived*/

```
--4
UPDATE LocalCentre
SET On-Hand = On-Hand + 100
WHERE   CentreNum = 453  AND TexCode='BA068';
```

/*We get an inspection report on warehouse damage. Update the 'OnHand' field at the Central warehouse to indicate that 25 metres of HM091 has been written off as spoiled.*/

```
--5
UPDATE CentralWarehouse
SET On-Hand = On-Hand - 25
WHERE TexCode = 'HM091';
```

/*An email tells us that 50 metres of the popular fabric BA068 has arrived at local centre 765. Update the On-Hand field of this local factory, to indicate this... */

```
--6
UPDATE LocalCentre
SET On-Hand = On-Hand + 50
WHERE   CentreNum = 765  AND TexCode= 'BA068';
```

/* ...and Update this fabric's In-Transit field for the Central Warehouse. */

```
--7
UPDATE CentralWarehouse
SET In-Transit = In-Transit - 50
WHERE TexCode = 'BA068';
```

Using, where appropriate, examples from the situation just described, answer the following questions.

A. In the database context, what do we mean by a '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. At the moment, when a shipment of textiles arrives at a local centre, the manager there rings up the Central Warehouse and informs them of its arrival. The DBA at the Central Warehouse then updates the database so that it reflects the new situation. It has been proposed that local managers be allowed to update the database directly themselves when an 'In-Transit' shipment arrives at their centre, via the web. Using as your example some of the SQL statements listed above, discuss the problem that could arise if this

were done by simply giving local centre managers 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

Question 2

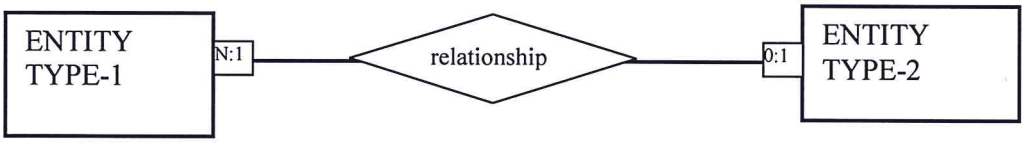
A corporation which manufactures Games Consoles buys in many different types of electronic and other components from independent suppliers, and from them assembles several different models of Games Console. Independent Computer Games shops then buy these Games Consoles to sell retail.

To assure continuity of supply, the manufacturing corporation typically has at least two distinct suppliers for each type of component. A supplier is only recorded if it supplies at least one type of component, but a component type may exist without yet having any suppliers.

A particular model of Games Console is made up of many different types of component. Each component-type can be used in the assembly of several different models of Games Console. A Games Console model cannot exist without a set of components that make it up, but a component can be recorded even if it is not currently used in the assembly of any existing Games Console model.

A given Computer Games shop can receive and sell many different models of Games Console. No Computer Games shop has a monopoly on re-selling any given Games Console model. No shop will be recorded unless it is selling at least one model of the manufacturer's Games Consoles, but a Games Console model may exist without yet having any shops which stock it.

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 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.) Be sure to indicate the Primary Keys.

Suppliers are uniquely identified by *SupCodes*, components by *ComponentNumbers*, Games Console models by *ModelNames*, and Shops by

ShopNames. We also want to record the single main *ContactPhone* for each supplier, the *weight* of each component, and the *features* (there can be more than one for a single model) of each Games Console model, and the *address* of each shop. Be sure to indicate the Primary Keys.

[10 Marks]

C. With the current arrangements, any particular component is delivered by one of several different suppliers. Thus, when a component is used to assemble a Games Console, we don't know which supplier it came from.

Suppose we wanted to change that, so that we knew just which supplier had supplied which individual instance of a component. (As we might want to if a percentage of some components began to fail, and we wanted to see if a particular manufacturer was responsible.)

Would we need to alter our relational schema to record this, or is the information already present in the current schema, just requiring a JOIN? If we needed a new schema, what would it look like? If a new relational schema is required, would the Entity-Relationship Diagram corresponding to it need to be redrawn, and if so, how?

[5 Marks]

TOTAL = 25 marks

Question 3

A. A chemical processing complex is subject to periodic inspections to ensure its safe running. These are recorded in the following relation. No location has the same type of inspection more than once on a given date.

Location	Inspection-Type	Date	Result
Cracking tower 6	Wiring insul.	2017-04-23	Passed
Cracking tower 6	Valve seals	2017-04-23	Passed
Pres.Chamber-B	Valve seals	2017-04-24	NH_3 detected -- replacement scheduled
Mix. Vat 2C	Smoke alarm	2017-04-23	Passed
Mix. Vat 2D	Smoke alarm	2017-04-23	Passed
Pres.Chamber-B	Valve seals	2017-05-03	Passed
Cracking tower 6	Wiring insul.	2016-04-12	Passed

Choosing the attribute or attributes which will be a relation's Primary Key during the design process is a critical part of database design.

In the relation above, what would be the undesirable consequences, when designing the relation, and before trying to load data into it, of designating:

(1) Location as the Primary Key?
(2) Inspection-Type as the Primary Key?
(3) Location + Inspection-Type as the Primary Key?
(4) Inspection-Type + Date as the Primary Key?
(5) Inspection-Type + Date + Result as the Primary Key?
(6) Location + Inspection-Type + Date + Result as the Primary Key?

What *should* the Primary Key for this relation be?

[5 marks]

B. Consider the following table, which is used exclusively for generating address labels for mass postal advertisement mailings, sent out monthly, all within the same country. The full database has several million addresses, and a label-printing run may take several hours. A sample of the data is shown here.

Name	Address-1	Address-2	PostCode	State
Eugenia Cheng	No. 6 Peregrine Heights	Cicero	K3M-3P9	IL
Mariam Mirzakhani	3459 Sierra Way	Palo Alto	G9A-6T3	CA
Emily Noether	Flat 3, Trump Towers	Cicero	K3M-3P9	IL
Sonya Kovalevskaya	55 Carollton Drive	Cicero	K3M-3P9	IL
Mary Cartwright	17 The Hills	Austin	P3Y-8T3	TX

This relation is not fully normalized. There is a functional dependency between Postcode and State. A State consists of many Postcodes. No Postcode straddles more than one State. (In other words, for a given Postcode, there is just one State. PostCode functionally determines State.) This relation could be normalized to 3NF, by splitting the State attribute out to its own separate relation, consisting of PostCode and State, with PostCode as the Primary Key.

Normalizing it like this would avoid the problem of update, addition and deletion anomalies. But would there be any disadvantages to system functioning as a consequence of doing this?

[2 marks]

C. One way to ensure that a relation has a Primary Key that is guaranteed to be unique is to design the relation with an additional column, which is automatically incremented when a new tuple is added. This attribute is essentially a 'record number', and the system guarantees that every tuple has a unique identifier through an 'auto-increment' feature that adds a new number when a new tuple is added.

Are there any disadvantages to doing this? If so, briefly explain them. If not, explain why not.

[2 marks]

D. Almost all databases which are large enough so that most of their content normally resides on disk have one or more of their attributes indexed. Write a brief essay discussing indexing. **Your answer should cover:**

- what is an index?
- why do we index databases?
- what criteria govern which attributes are indexed?
- what are the disadvantages of indexing and what causes the disadvantages?

[8 marks]

E. **(i)** Suppose a database were read-only. Would some form of concurrency control be necessary? Why, or why not?

[1 mark]

(ii) What do we mean by 'physical data independence'? What is independent of what? Why is it considered important?

[2 marks]

F. Write brief definitions of each of the following terms, in the context of the relational model.

(1) Attribute [or Column] Integrity
(2) Referential Integrity
(3) Entity Integrity
(4) Composite Key
(5) View

[5 marks]

TOTAL = 25 marks

Question 4

A small business which does general maintenance work on rented flats has hired you as a database consultant.

They have a database management system, and have already designed a relation to hold information about the jobs they have sent their employees out on. A single 'job' will consist of various tasks carried out at a particular property, by the employees of the business.

Every employee has an Employee Number, and First Name. Employees may be sent to a particular flat to carry out a certain task, on a certain date. Every job has a Job Number, Owner's Phone, and Job Address. Note that an owner may own more than one property. The particular Task that the employee was to carry out on that job is recorded, along with the start date for that particular Task. Employees just carry out one task on a particular job.

As an example, this table can record that the Employee Number 3459, whose first name is 'Mike', was sent out on Job Number 4994, which is at #3, Myrtle Gardens, and whose owner's phone is 898 774, to carry out the Task of installing a sink. Other employees have been sent out on that same job, and Mike was also sent out to that flat, as part of an earlier job, once before.

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

Primary Key: EmpNum + JobNum

EmpNum	Name	JobNum	JobAddress	OwnerPhone	Task	StartDate
3459	Mike	4994	#3, Myrtle Gardens	898 774	Install sink	2017-04-12
0672	Amanda	4994	#3, Myrtle Gardens	898 774	Hang curtains	2017-04-17
3301	Salem	4994	#3, Myrtle Gardens	898 774	Paint interior	2017-04-20
3459	Mike	5633	21 Djugashvilli Villas	954 009	Install sink	2017-03-21
3459	Mike	5701	Flat 6, Bunberry Bldg	898 774	Remove boiler	2017-04-23
3301	Salem	5701	Flat 6, Bunberry Bldg	898 774	Paint interior	2017-04-30
3301	Salem	5633	21 Djugashvilli Villas	954 009	Paint interior	2017-04-13
3459	Mike	2688	#3, Myrtle Gardens	898 774	Install bath	2016-09-20

A. Identify the Functional Dependencies in this table, using the following example: If B is functionally dependent on A, show it this way: A → B. You need not show multi-valued dependencies.

[5 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. 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 found in the original table into your new tables.

[8 marks]

E. It has been proposed to add the following relation to the database, in order to record further information about each property, namely, the gas or electrical appliances each has, and also each flat's accessories. Some of the tuples might look like this.

JobAddress	Appliances	Accessories
#3, Myrtle Gardens	Microwave	Fish tank
#3, Myrtle Gardens	Stove	Throw rug
#3, Myrtle Gardens	Air Conditioner	Throw rug
Flat 6, Bunberry Bldg	Stove	Wall Mirror
Flat 6, Bunberry Bldg	Oven	Large vase
Flat 6, Bunberry Bldg	Oven	Storage chest
Flat 6, Bunberry Bldg	Oven	Bookcase
21 Djugashvilli Villas	Washing machine	Bookcase

This relation has no determinants, thus it does not violate the prescription 'Let every determinant be a candidate key'. Therefore, it is in Boyce-Codd Normal Form. However, there may be problems associated with it. Briefly comment on the problems associated with this design and propose an alternative schema that avoids these problems, while holding the same information. Populate your new relations with the correct data.

[4 marks]

TOTAL = 25 marks

Question 5

The following local government database records information about certain deliveries of building materials from building material suppliers to construction sites, where construction projects are taking place.

The database records each supplier's suppliercode, name, size, the year they were approved as a supplier, and their environmental friendliness rating.

For each construction site, a site number is recorded, along with the type of construction project going on there, a map reference, the year it began, and the name of the supervisor in charge of arranging deliveries. (A given person may be in charge of more than one site.)

Details of deliveries of material to the site by suppliers are also recorded – the site being delivered to, the supplier, the date of delivery, the material delivered, and its quantity in tons.

An extract from the database is given below.

SUPPLIER
PRIMARY KEY: SUPPLIERCODE

SUPPLIERCODE	NAME	SIZE	YEARAPPROVED	RATING
S001	Carillion	Medium	2015	8
S002	Builder Bob's	Large	2013	8
S003	Konstrukt	Large	2013	6
S007	Reliable	Medium	2014	9
S009	Anders Bros	Large	2012	7

SITE
PRIMARY KEY: SITENUM

SITENUM	PROJECT_TYPE	MAPREF	STARTYEAR	SUPERVISOR
T01	School	K34B06	2014	M. Azideh
T02	Roundabout	H93C87	2014	P. Chanders
T04	Hospital	H76B33	2015	NULL
T05	Hotel	A82K77	2015	M. Aziz
T07	Resurfacing	K34B06	2017	M. Azideh
T09	Hotel	Y64G88	2016	K. Elfridi

DELIVERY
PRIMARY KEY: SITENUM + SUPPLIERCODE + DELIVERYDATE + MATERIAL

SITENUM	SUPPLIERCODE	DELIVERYDATE	MATERIAL	QTY_TONS
T04	S003	2017-04-21	Asphalt	10
T02	S007	2017-10-15	Asphalt	20
T04	S003	2017-06-08	Hardcore	15
T01	S001	2017-04-09	Sand	20
T07	S001	2017-02-11	Cement	10
T09	S007	2016-04-30	Asphalt	5
T04	S007	2016-09-24	Sand	25
T04	S007	2016-10-21	Cement	15
T01	S007	2016-04-17	Gravel	10
T04	S002	2017-04-21	Hardcore	35
T01	S007	2017-11-15	Asphalt	15
T09	S002	2017-04-12	Sand	40
T04	S001	2017-10-11	Sand	10

A. Construct queries in SQL to carry out the following tasks. You need not show the results.

1) Get the supplier code and name of building material suppliers whose size is 'Medium' and whose environmental friendliness rating is greater than 7.

[1 mark]

2) Get the site numbers of projects where we do not have a name for the supervisor.

[1 mark]

3) Find the total quantity in tons of building material (QTY_TONS) delivered by supplier S007 in 2016.

[1 mark]

4) Get the construction site numbers of construction sites managed by M. Aziz which have had deliveries from supplier S007:

[2 marks]

5) Find the project types of construction sites which have had deliveries from building material suppliers whose environmental friendliness rating is less than 8.

[2 marks]

6) Get the supplier codes of suppliers who have never delivered asphalt.

[2 marks]

7) Get the supplier codes and total quantities of deliveries made by that supplier for each of the building material suppliers.

[2 marks]

8) Get the supplier codes and total quantities delivered by that supplier for each of the building material suppliers that have made deliveries from 2015 onwards.

[2 marks]

9) Get the supplier codes and total quantities delivered, for 2017, for each of the suppliers whose deliveries totaled one hundred tons or more during that year.

[3 marks]

B. If the database administrator finds that some regularly-used queries are taking a very long time to complete, what steps might be taken to first identify the exact nature of the problem, and then to possibly ameliorate it? List at least **THREE** things – your answers need not be lengthy.

[3 marks]

C. Running a complex query against a very large database risks retrieving results that *seem* valid, but are not. How can this be guarded against?

[3 marks]

D. Describe a situation in which clients might want to implement a database management system, but where the relational model would *not* be an obvious fit for the data and the relationships among the data in that situation.

[3 marks]

Total = 25 marks

END OF PAPER