Past paper Sample SOLUTION

Note:

- The sample solution here provides only “key” part of the solution, it may not be the full solution.
- Also important to note, the sample solution here is not the “only” solution. You should answer questions to demonstrate your understanding, rather than repeating these solutions or lecture notes “word for word”.
- Always write full sentences in exam.

Past paper Sample SOLUTION

Question 1

a) Answer the following questions.

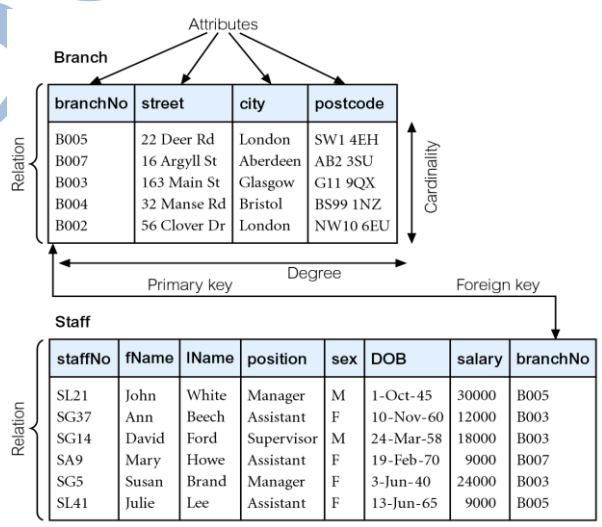
[10 marks]

i) Define the terms: *data*, *relation*, *relational schema*, *relational algebra*, and *Database Management System* (*DBMS*).

(5 marks)

ii) Explain *primary key* and *foreign key* in relational model with an example. Using the same example, explain *entity integrity* and *referential integrity*.

(5 marks)

	Do not write in this column
i)	
Data: facts and statistics collected together for reference or analysis.	
Relation: A relation is a table with rows and columns.	
Relational schema: A relation schema can be thought of as the basic information describing a table or relation.	
Relational algebra: Relational algebra is formal language associated with the relational Model.	
Database management system: A software system that enables users to define, create, maintain, and control access to the database.	
ii)	
Example of relation: 	

branchNo is the primary key of the relation Branch. or	
staffNo is the primary key of the relation Staff.	
Example of foreign key is branchNo in Staff relation, referencing the branchNo in the base relation of Branch.	
Entity integrity: In a base relation, no attribute of a primary key can be null.	
branchNo and staffNo can't be null.	
Referential integrity: If foreign key exists in a relation, either foreign key value must match a candidate key value of some tuple in its home relation or foreign key value must be null.	
	10 marks

b) In a time of coronavirus, restaurants can only provide limited set menu types, which include Meat, Vegetarian, Vegan, and Halal. Considering the following relational schema:

```
Restaurant (rid: integer, rname: string, address: string)
SetMenu (sid: integer, main_ingredient: string, type: string)
Order (rid: integer, sid: integer, cost: real)
```

Where primary key attributes are underlined, and the domain of each attribute is listed after the attribute name. Therefore `rid` is the primary key for `Restaurant`, `sid` is the primary key for `SetMenu`, and `rid` and `sid` together form the primary key for `Order`. The `Order` relation lists all available orders of set menus from each `Restaurant`.

Formulate the following queries in **relational algebra**. **[9 marks]**

i) List all restaurant information. **(2 marks)**

ii) List the names of restaurants that can provide Vegetarian and Vegan options. **(3 marks)**

iii) List the IDs of restaurants that supply all Vegetarian option. (Hint: Each type of set menu has different ingredient lists, for example, the main ingredient of Meat option could be pork or beef.) **(4 marks)**

	Do not write in this column
i)	

Restaurant.		
Or $\Pi_{rid, rname, address}$ (Restaurant).		
ii)		
$\Pi_{rname}(\Pi_{rid} (\Pi_{sid}(\sigma_{type = \text{“Vegetarian”} \wedge type = \text{“Vegan”}} SetMenu) \bowtie Order) \bowtie Restaurant)$.		
iii)		
$(\Pi_{rid, sid} Order) \div (\Pi_{sid} (\sigma_{type = \text{“Vegetarian”}} SetMenu))$		
		9 marks

c) For the relational schema in part b) above, formulate the following queries using **SQL**.
[6 marks]

i) Find the names of restaurants who provide Halal option. **(2 marks)**
ii) Find the sid of the most expensive order sold by the restaurant named Hakkasan. **(4 marks)**

	Do not write in this column
i)	
SELECT R.rname	
FROM Restaurant R, SetMenu S, Order O	
WHERE S.type = “Halal” AND O.sid = S.sid AND O.rid = R.rid	
ii)	
SELECT O.sid	
FROM Order O, Restaurant R	
WHERE R.rname = “Hakkasan” AND O.rid = R.rid	
AND O.cost >= ALL (Select O2.cost	
From Order O2, Restaurant R2	

WHERE R2.rname = “Hakkasan”	
AND O2.rid = R2.rid)	

Question 2

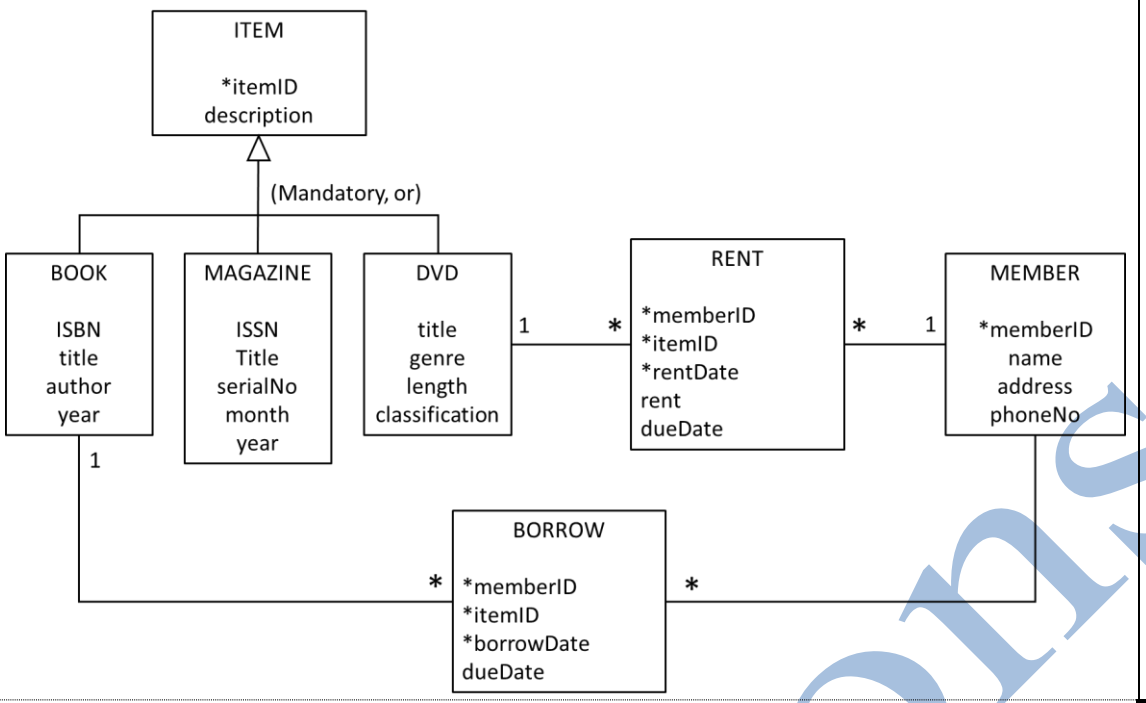
a) You are asked to design a relational database for a local library, and the following information is given by the library:

- The library has a number of items available to the library members. Items are identified by an *item ID* and have an attribute of *description*. There are three different types of items: books, magazines and DVDs. Books are described by *ISBN number*, *title*, *author*, and *year of publication*. Magazines are described by *ISSN number*, *title*, *serial number*, *month* and *year of publication*. DVDs are described by *title*, *genre*, *length*, *classification*. (We assume that each item in the library only has one copy of the same item, and each item belongs to only one type.)
- Members are identified by member ID and have descriptive data of name, address, and phone number.
- A member may borrow a book for free, and the library needs to record the date a book is borrowed and the due date.
- A member may rent a DVD, and the library needs to record the rental fee, date of the rent and due date.
- Magazines are available for the library members, but not available for borrowing or loans.

Draw an ***Enhanced Entity-Relationship(EER) model*** for the library database. Your EER diagram must include details of the entities, attributes and relationships and *label* them clearly. Clearly *state the assumptions* if you make any.

[18 marks]

	Do not write in this column
The design below assumes rent and borrow are both m:m relationships.	



18 marks

b) Answer the following questions. **[7 marks]**

i) Explain how 1:m relationship in ER model can be mapped into relations in relational model.

(3 marks)

ii) Derive ***relational schema*** for the entities and relationships in the ER model in Figure 1 below. Indicate the primary key for each relation in the relational schema. **(4 marks)**

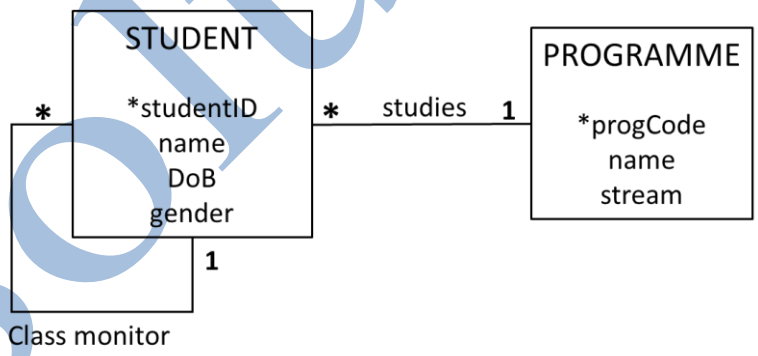


Figure 1

	Do not write in this column
i) Mapping 1:m relationship into relational model: Post a copy of the primary key attribute(s) of one-side entity into the relation representing the many-side, to act as a foreign key	
ii)	
Student(studentID, name, DoB, gender, monitorID, progCode)	
Programme (ProgCode, name, stream)	

		7 marks

Question 3

a) For a relational schema R (A, B, C, D, E, F), there exist functional dependencies of

fd1: A, B → C, D, F
fd2: B → D
fd3: C → E

Answer the following questions: **[9 Marks]**

i) What is *Boyce–Codd normal form* (*BCNF*)? **(2 Marks)**

ii) Is relation R in BCNF? If not, **normalize** R to BCNF and **explain** all steps of normalization. **(7 Marks)**

	Do not write in this column
i) A relation is in BCNF if and only if every determinant is a candidate key. ii) R is not in BCNF. Primary key of R is A, B. So fd2 is partial dependencies on primary key and fd3 is transitive dependencies on primary key. Remove partially dependent attributes from the relation by placing the attributes in a new relation along with a copy of their determinant.	
We have R1 (A, B, C, E, F), R2 (B, D)	
R1 still has transitive dependency fd3, so R1 normalise to R3 (A, B, C, F), R4 (C, E).	
So final normalised relations are R2, R3, R4.	
	9 marks

b) The table below provides sample data for an agency called Hotel Services, which supplies part-time/temporary staff to hotels, and the number of hours worked by each staff at various hotels. Table 1 is susceptible to update anomalies. Provide examples of how insertion, deletion, and modification anomalies could occur on this table.

[6 marks]

Table 1. Contracts

contractNo	hours	staffNo	staffName	hotelNo	hotelLoc
C1024	16	S102	Smith, J.	H25	Haidian
C1025	16	S267	Green, D.	H4	Chaoyang
C1026	28	S267	Green, D.	H25	Haidian
C1027	16	S102	Smith, J.	H4	Chaoyang
C1028	25	S267	Green, D.	H15	Changping
C1027	25	S408	Crowe, M.	H25	Haidian

	Do not write in this column
Insertion anomaly: to insert details of a new staff with no contract, the contract number would be null which violate entity integrity.	
Deletion anomaly: if we delete staff Green D, then hotel H15 will be lost from the table too.	
Modification anomaly: if we need to change location of a certain hotel, then all tuples with that hotel must be changed, otherwise we will have different location information for the same hotel.	
	6 marks

c) Describe how ***two-phase locking protocol*** (*2PL*) works.

[4 marks]

	Do not write in this column
• To read a resource a transaction needs to have a shared lock on that resource. A transaction cannot acquire a shared lock on any resource that has an exclusive lock on it by another transaction. • To write a resource a transaction needs an exclusive lock on it. A transaction cannot acquire an exclusive lock if any other transaction has any lock on that resource.	

- In the two-phase locking protocol, a transaction needs to acquire all the necessary locks (and waits if the locks are not available), this is known as the growing phase.
- A transaction only releases the locks when no new locks are needed (shrinking phase).

d) Figure 2 illustrates a number of transactions (T1, T2, T3, T4 and T5) being processed from time t_0 until t_f when system had a power failure. At time t_c a checkpoint was done. Explain for each transaction in Figure 2 below how recovery is done.

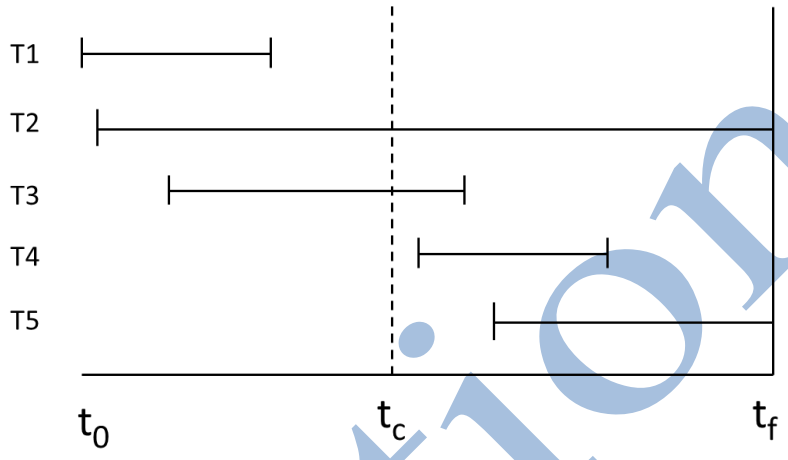


Figure 2

	Do not write in this column
Checkpoint at time t_c, changes made by T_1 have been written to secondary storage, so no need to do any recovery for T1 Thus: only redo T_3 and T_4, undo transactions T_2 and T_5	
	6 marks

Question 4

a) Consider the following statement:

```
<xs: element name = "STAFFNO" type = "xs: string"/>

    <xs: element ref = "STAFFNO"/>
```

Interpret the meaning of the above XML Schema (XSD) code and explain the advantages of using XSD over Document Type Definition (DTD).

[6 marks]

	Do not write in

	this column
a)	
The first line defines an attribute STAFFNO to be "string" type.	
The second line is used whenever a STAFFNO element is required.	
Compared with DTD, XSD has the following advantages:	
1. XSD is more comprehensive method of defining content model of an XML document.	
2. XSD tries to overcome XML deficiencies by using additional expressiveness to allow Web applications that exchange XML data more robustly.	
	6 marks

b) Answer the following questions about Distributed Database Management Systems (DDBMS). **[12 marks]**

i) Briefly explain the difference among distributed database, Distributed DMBS (DDBMS), and distributed processing. **(6 marks)**

ii) Blockchain is a distributed ledger that enables secure and transparent transactions. Using your knowledge on datamining and blockchain, discuss if data mining can be integrated into blockchain technology. **(6 marks)**

	Do not write in this column
i) Distributed database is a logically interrelated collection of shared data (and a description of this data), physically spread over a computer network.	
Distributed DBMS is a software system that permits the management of the distributed database and makes the distribution transparent to users.	
A DDBMS consists of a single logical database that is split into fragments. These fragments are distributed over a computer network.	
In distributed processing there is a centralised database. The centralised database can be accessed over a computer network.	
ii)	

[This is an open question, so students can answer either yes or no. Marks can be given as long as they have reasonable explanations.] Data mining techniques can be applied in the blockchain application.	
1. Before encapsulating transactions into a block, potential data mining techniques could be applied to analyse the authenticity. 2. Transaction analysis: data mining techniques could be used to analyse the traffic and transaction status so that it can identify the chain activity, which is extremely useful in blockchain evaluation.	
Data mining techniques cannot be applied in the blockchain application. As blockchain is just a distributed ledger that records transactions. Users can read and add information on the chain. So it is nothing but a shared ledger. There is few application or possibility of applying data mining technologies in Blockchain.	
	12 marks

c) Big data explosion is the main catalyst behind the growth of popularity of NoSQL. Answer the following questions. **[7 marks]**

i) Define NoSQL. **(2 marks)**

ii) Instagram uses Cassandra (a NoSQL database management system) as a general key-value storage service, to support the users photo feed. Briefly explain why Cassandra is adopted by the Instagram application rather than relational database. **(5 marks)**

	Do not write in this column
i) NoSQL: Not only SQL or "Not Relational".	
ii)	
1. Instagram has a large number of users that generate enormous images and contents all the time.	
2. In the mean time, it requires storing large amount of data that need fast scalability while maintaining high performance. 3. There is no strict ACID constrains (Atomicity, consistency, isolation, durability) on a social media platform such as Instagram.	

		7 marks