School of Computing and Information Systems

INFO90002

Database Systems and Information Modelling

Practice Exam 2

Semester 1 2022

Reading Time: 15 minutes

Writing Time: 120 minutes (2 Hours)

This exam has 8 pages including this page

ATTEMPT ALL QUESTIONS IN ANY ORDER

Authorised Materials:

- None

While you are undertaking this assessment, you MUST NOT

- make use of any messaging or communication technology
- record, screenshot, stream, upload or in any known format duplicate this document
- record, screenshot, stream, upload or in any known format duplicate your solutions
- make use of any world wide web or internet based resources such as wikipedia, github, stackoverflow, google, Weichat or any known search engine / messaging services
- act in a manner that could be regarded as providing assistance to a student who is undertaking this assessment or in the future will be undertaking this assessment
- seek assistance from any other student who is undertaking this assessment or in the future will be undertaking this assessment

Instructions to Students:

- This exam is in 8 sections. Attempt all questions in all sections.
- We recommend using pencil and paper for modelling questions to save time
- The total for this exam is 100 marks representing 50% of your final assessment
- Attempt **all** questions which are of unequal marks value
- **PLEASE DO NOT USE RED font colour**

IMPORTANT

- **Answer questions in the allocated space for answers.**
- **The official exam language is English. Sections of the submission in languages other than English will NOT be assessed and will be marked as 0.**
- Before submitting your solution document, check that the diagram(s) is/are readable. It is your responsibility to ensure your answers are readable and make sense to the marker.

The work you submit **must be based on your own knowledge and skills** and without the assistance of any other person. You MUST NOT directly copy work that you have not authored (e.g. slide notes, websites, other student's study notes).

Q1. ER Modelling (30 Marks)

TeaTime

TeaTime is a platform for tea lovers. To use the system, users must install the TeaTime app on their phone and set up a user profile that contains: a login name (their email address), password (encrypted and stored as a string of 64 characters), short screen name e.g. "DonkeyKong64" (which other users will see), profile photo, and a "My Cuppa" entry - their favourite tea type, chosen from a list of standard drink names such as "Peppermint Tea" and "Black Tea with Honey". We plan to have up to 2 million customers (end users) accessing the system.

While a user's phone is switched on, the TeaTime app sends the phone's current location to the server once per minute. We store these, so that the system knows where a given user is "now", as well as the history of where the user has been.

All locations in this system are recorded as a pair of numbers representing latitude and longitude. Latitudes are between -90 and 90 degrees (south pole to north pole) while longitudes are between -180 and 180 degrees (west or east of the prime meridian in Greenwich). We will use a precision of 4 decimal places, which is about 11 metres at the equator. For example, the Doug McDonell building at UniMelb is at latitude -37.7989, longitude 144.9627.

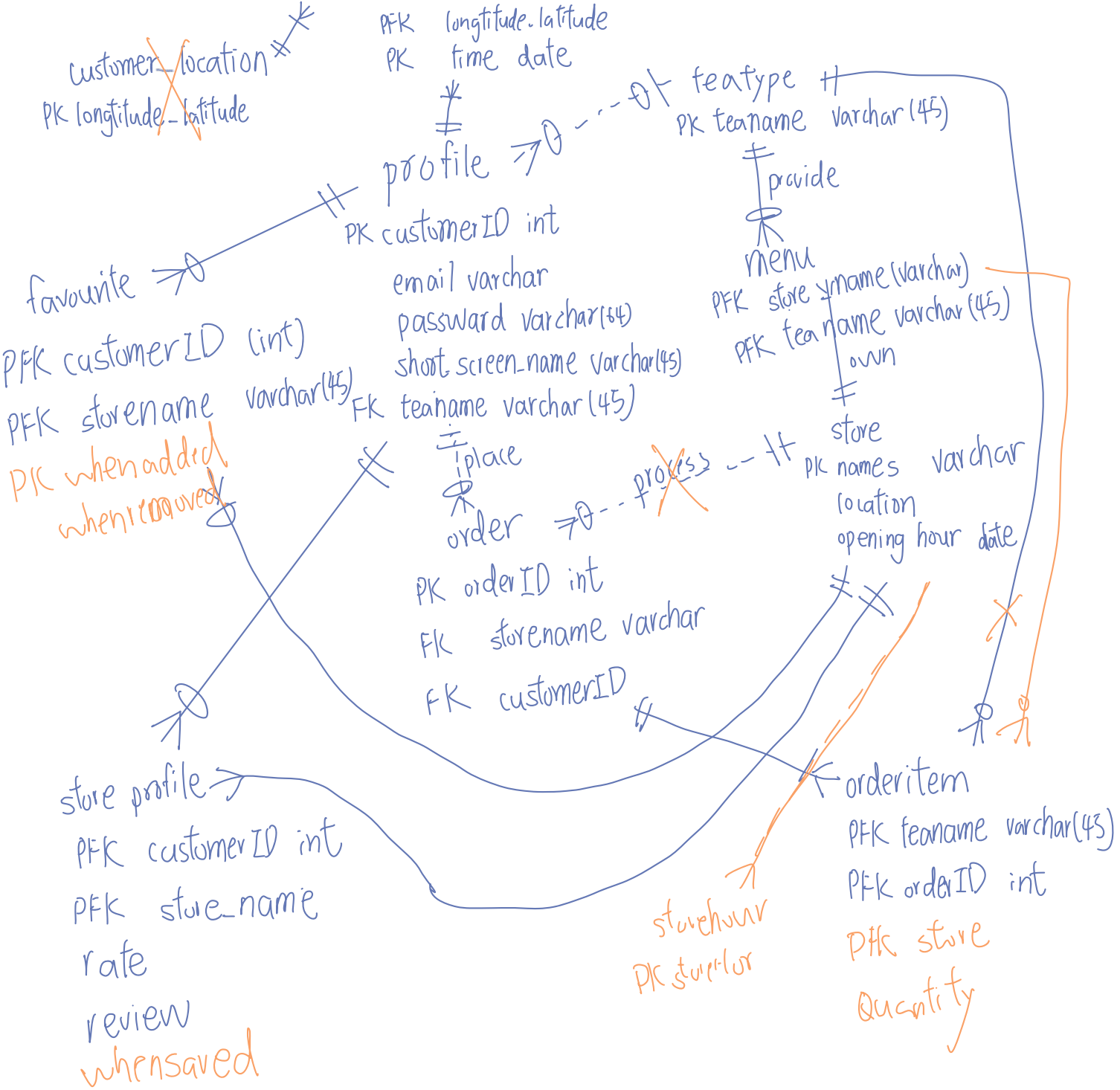
The names, locations and opening hours of about 600 tea stores are available via the app. Users can browse stores either in an alphabetical list or via a map. When viewing a store's profile, users can see reviews and photos uploaded by other users, and the store's average rating. The store's opening hours are recorded as an opening and closing time for each day of the week (for example, a café might open Mondays 8am to 6pm, Tuesdays 9:30am to 7pm, and so on). Each café stores a menu of the teas it sells: these must be drawn from our standard list mentioned above.

Users can rate stores. A rating consists of a whole number between 0 and 10, along with an optional piece of text (up to about 30 words). A given user can only rate a given tea store once. Users can mark particular stores as favourites. These can be viewed in a list. Users can later "unfavourite" the tea store if they wish, and yet later "favourite" it again. We keep a history of these favourites and unfavourites. Customers can use the app to order drinks from a tea store. To do this a user first selects a tea store, then chooses how many of each tea type(s) they want from the store's menu. We keep track of when orders are placed and when customers later pick up the order.

Q.1. Draw a physical model in Crow's Foot notation for this case study. Be sure to write down any assumptions you make.

(30 marks)

customer_location_profile
PFK customerID int


customer_location
PK longtitude_latitude
PFK longtitude.latitude
PK time date
profile
PK customerID int
email varchar
password varchar(64)
short screen_name varchar(45)
FK teaname varchar(45)
teatype
PK teaname varchar(45)
provide
menu
PFK storeyname(varchar)
PFK teaname varchar(45)
own
store
PK names varchar
location
opening hour date
favourite
PFK customerID (int)
PFK storename varchar(45)
PK whenadded
whenremoved
place
order
PK orderID int
FK storename varchar
FK customerID
store profile
PFK customerID int
PFK store_name
rate
review
whensaved
orderitem
PFK teaname varchar(45)
PFK orderID int
PFK store
Quantity
storehour
PK storehour

Q2. SQL (20 marks)

Given the schema in Figure 2, write a single SQL statement to correctly answer each of the following questions (2A – 2D). DO NOT USE VIEWS to answer questions.

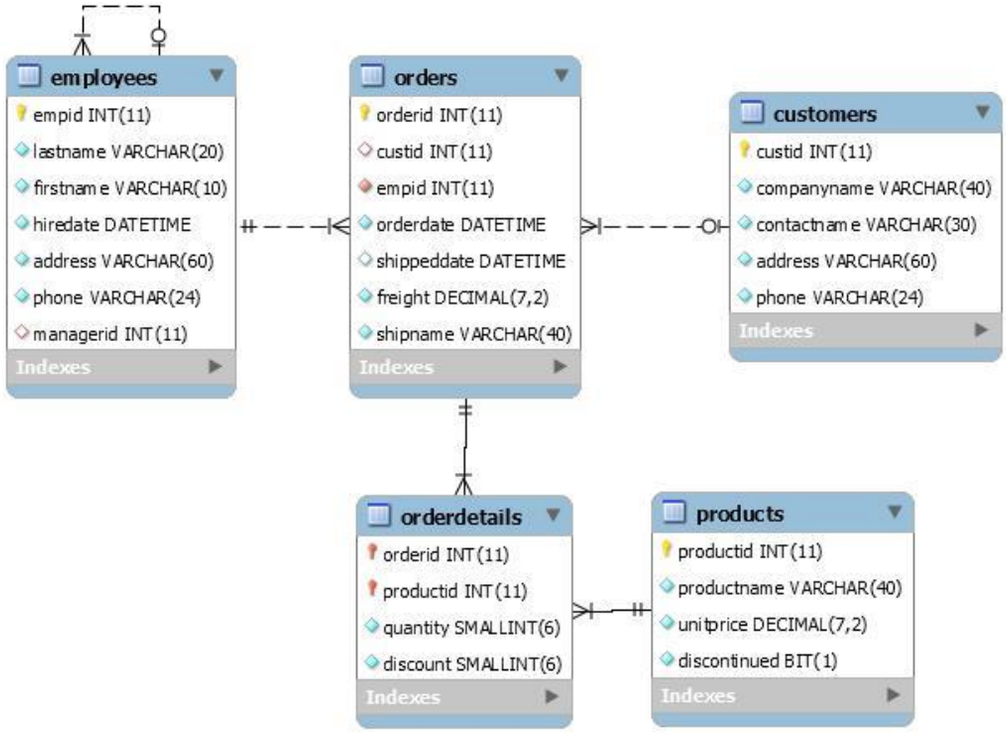


Figure 1 Company Schema

Q.2A. Write a query that returns customers (company names) and the details of their orders (orderid and orderdate), including customers who placed no orders.

(3 marks)

Q.2B. Write a query that returns the first name and last name of employees whose manager was hired prior to 01/01/2002.

(4 marks)

Q.2C. Write a query that returns customers (customer ID) whose company name is 'Google', and for each customer return the total number of orders and total quantities for all products that were not discontinued ('1' means discontinued, '0' not discontinued).

(5 marks)

Q.2D. Write a query that returns the ID and company name of customers who placed orders in 2007 but not in 2008.

(8 marks)

Q2A:

```
select  customer.companyname, order.orderid ., oder.orderdate
from    customer
left outer join  order
on  customer.custumerID= order.customerID
```

Q2B:

```
select  emp.firstname, emp.lastname
from  employees emp
inner  join  employee manager
on    emp.employeeID= manager.employeeID
where   manager.employeeID < '2002-01-01'
```

Q2C:

```
select distinct customer.customerID. count(orders.orderID), sum(orderdetails.quantity)
from   customer
inner  join  order
on   customer.custid = orders.custid
inner  join  orderdetails
on   orders.orderID = orderdetails.orderID
inner  join  products
on   orderdetails.productid = products.productID
where  products.discountinued = 0 and  customers.companyname = 'Grougle';
group by  customer.customerID
```

Q2D:

```
select  customers.custId, customers.companyname
from  customers.
inner  join  order
on  customer.custid =  orders.custid
```

```
where orders.orderdate in
    (select orderdate
     fium orders
     where orderdate = 2007)
and orders.orderdate not in
    (select orderdate
     fum orders
     where ordate = 2008)
```

Q3. Normalisation (16 marks)

A car rental company keeps its records in a spreadsheet which looks like a table below. Note a car cannot be rented for part of the day. CostPerDay is determined on the day of rent by some business rules.

RegNo	Make	Model	DateRented	DateReturned	CostPerDay	CustID	CustName	Mobile
2LB 8NG	Toyota	Camry	1/2/2024	5/2/2024	80	2453	Jay Lee	0411199422
2LB 8NG	Toyota	Camry	6/2/2024	16/2/2024	80	7812	Lou Stone	0466121260
2LA 4TX	Hyundai	i30	4/2/2024	11/2/2024	70	3357	Jason Goh	0431441187

Represent the table as a relation, identify candidate key.

Produce 1NF, 2NF and 3NF.

Legend: PK – underline, FK – italics, PFK – underline+italics

1NF

Car (RegNo (PK), Make, Model)

rent (RegNo (PFK), DateRented (PK), DateReturned, CostPerDay, CustID, CustName, Mobile)

2NF

Car (RegNo (PK), Make, Model)

rent (RegNo . DateRented . DateReturned . CostPerDay . CustID . CustName . Mobile
PFK PK

3NF

Car (RegNo, Make, Model)

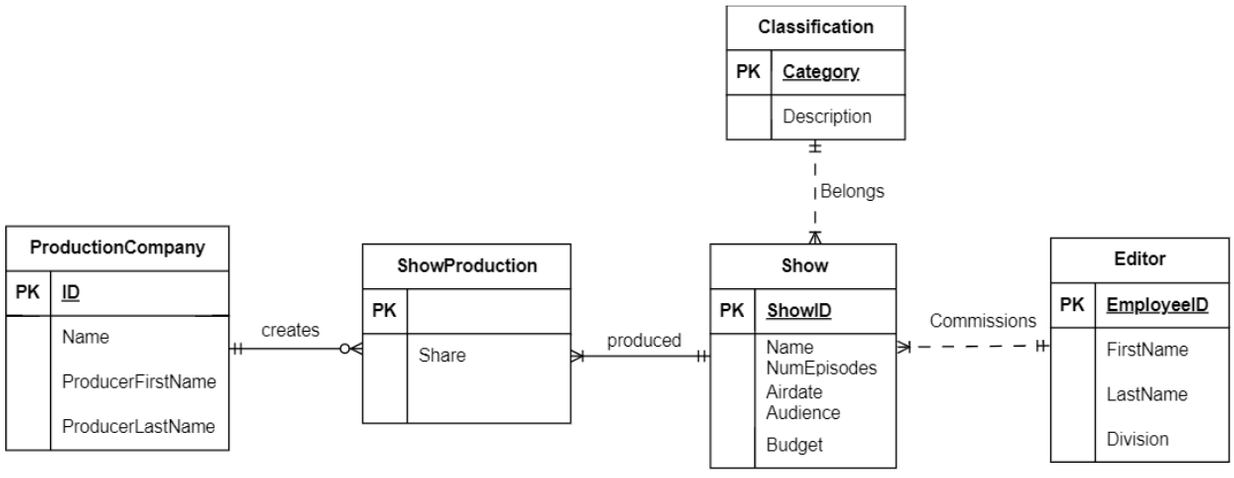
rent (RegNo . DateRented . CustID DateReturned, CostPerDay)
PFK PFK FK

Customer (CustID , CustName . Mobile)
PK

Q4. SQL DDL (10 marks)

The British Broadcasting Corporation (BBC) has a number of commissioning editors (e.g. Martin Davidson) who have cast a number of landmark shows ("Life on Earth", "A History of Britain") from independent production companies such as names like Hartswood films (Sherlock, Coupling) with producer Sue Vertue. Sometimes more than one production company is involved in the production of a series. Every Show is given a name (e.g. "Sherlock"), a number of episodes (between 6 and 24), a budget (up to several hundred thousand pounds) and an estimated air date for the first episode. The Producers must also nominate who is the target audience (Audiences are graded by age and income demographic) and classify the show (e.g. Documentary, Drama, Comedy, Light Entertainment, Game Show, Reality, News and Current Affairs, Special Event).

Q4. Write the SQL DDL for the following conceptual model. You can use whichever data types you think are appropriate for a MySQL relational database. You need to use NOT NULL to specify mandatory participation.



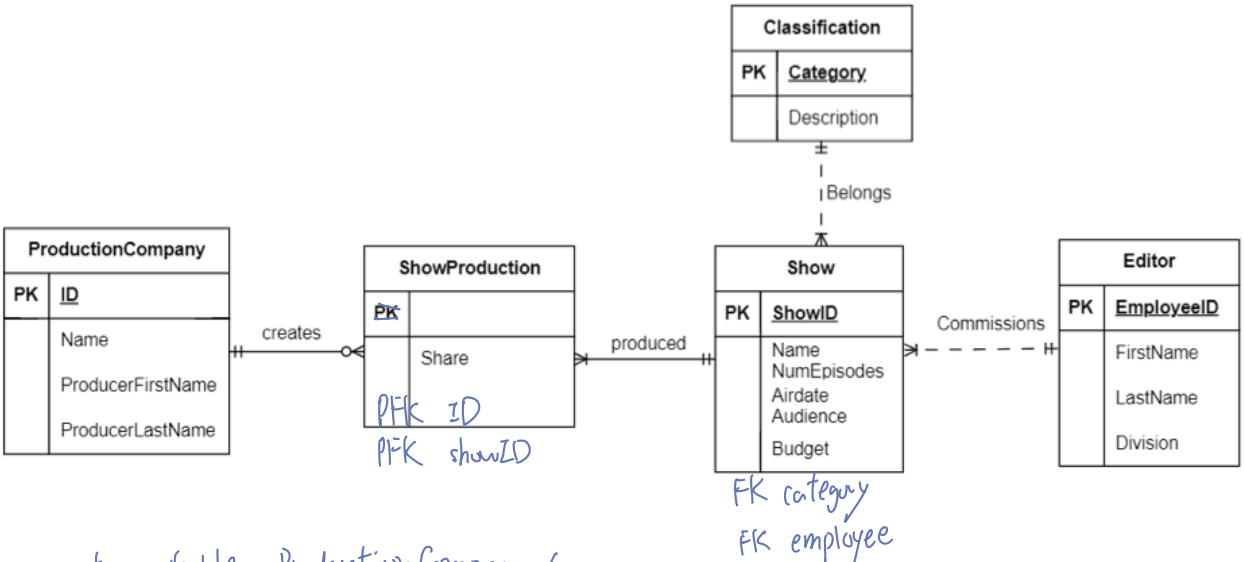
Q5. Data Warehousing (10 marks)

Q5A. Transurban operates toll roads in Australia and the United States of America. Traffic data analysis helps determine road maintenance, RFID reader maintenance, gantry maintenance and capacity management to reduce inefficiencies in its tollway network.

Each part of the tollway is identified by a section (E3), and multiple sections (E1, E2, E3) make up a chunk, (C3) and multiple Chunks make up an entire roadway ("Eastlink"). Many different vehicles on toll roads every day. Each vehicle falls into one category: motorcycles, passenger vehicles, 4WD's, vans, trucks, prime movers, buses, trailers, recreational and miscellaneous machinery (e.g. tractors, cranes, street sweepers, back hoes).

Transurban's management wants to understand the vehicle trips on its road network. They need to understand the number of trips, trip length, trip duration, at different times of the day, on different days and during different period (e.g. Easter holidays, school holidays).

Q4:




```
create table ProductionCompany (
    ID int,
    Name varchar(40),
    ProductFirstName (varchar(40) not null,
    ProductLastName Varchar(40) not null,
    primary key (ID)
);

create table Editor (
    employeeID int,
    firstName varchar(40) not null,
    LastName varchar(40) not null,
    Division varchar(40),
    Primary key (EmployeeID);

create table classification (
    category char(3),
    Description varchar(60),
    primary key (category));

create table Show (
    showID int,
    Name varchar(40),
    NumEpisode int,
    Audience varchar(40),
```

```
    Airdate  Date,
    Budget  decimal(8,2),
    category  char(3) not null,
    employeeID  int not null,
    Primary key (showID),
    Foreign key (category) reference classification((category),
    Foreign key (employeeID) reference editor ( employeeID) );
```

Foreign key need to have not null

each statement seperate by coma.

foreign is also one of attribute.

```
create table ShowProduction (
    share int,
    ID int,
    showID int,
    primary key (ID, showID),
    foreign key (ID) reference ProductionCompany (ID),
    foreign key (showID) reference Show(showID)
);
```

Draw a *star schema* to support the design of this data warehouse, showing the attributes in each table. You do not need to select data types. Clearly display the legend for Primary Key, Foreign Key and Primary Foreign Key.

(8 marks)

Q.5B. Why are star schemas preferred over relational database designs to support decision making?

(2 marks)

Q6. Security and Backups (10 marks)

Q6A. One of the technical safeguards of database systems is access control. Which type of access is based on the "need to know" principle? Explain how this principle works with an example.

discrement access control, it's based on user, action and object. The data owner have access to determine who can use these data. **(4 marks)**

Q6B. Learning Management Systems use Role-based Access Control (RBAC). Explain this statement.

The people in the same group, all have the access for a particular data. **(2marks)**

Q6C Illustrate the concept of an SQL injection. Describe a scenario and a string a hacker could enter that would jeopardise database security.

Hackers will write some queries to minic the SQL query and figure out what is the structure of database. Once they success, they will write in some queries that might affect the whole database. **(4 marks)**

Q7. Transactions (10 marks)

Q7A. What is the inconsistent retrieval problem? Describe the problem and use your own example to demonstrate the answer.

(4 marks)

Q7B. A sales company decided to increase salaries by 2% across all jobs. Write a transaction to increase annual salary as specified.

The Jobs table was created using DDL:

CREATE TABLE Jobs (JobCode int, JobTitle char(15), AnnualSalary decimal(10,2), primary key(JobCode));

Hint: it is a better practice to use variables instead of hardcoded values.
Syntax example: SET @name='John';

(6 marks)

```
start transaction;
set @increase_percent = 1.02;
update Jobs
set annualsalary = annualSalary * @increase_percent;
```

commit;

Q8. NoSQL (4 marks)

Q8. Domain integrity can be violated in NoSQL databases whereas Relational databases will report a domain integrity violation. Discuss the benefits and risks of violating domain integrity.

(4 marks)

END OF EXAM