THE UNIVERSITY OF MELBOURNE

School of Computing and Information Systems
End of Semester 2 2017 Examination

INFO90002 DATABASE SYSTEMS & INFORMATION MODELLING

Reading time: 15 minutes

Writing time: 180 minutes

This paper has 6 pages, including this page.

Authorised Materials:	No materials are authorised.

Instructions to Invigilators:

The examination paper IS TO REMAIN in the examination room.

Students are to be provided with standard script books.

Instructions to Students:

The total mark for this paper is 70 marks, representing 70% of your final assessment.

Ensure your student number is written on all script books during writing time.

This exam paper has 9 questions, some with multiple parts.

Attempt all questions.

Answer all questions on the right-hand lined pages of the script book.

Start the answer to each question on a new page in the script book.

The left-hand unlined pages of script books are for draft working and notes and will not be marked.

Write legibly in blue or black pen.

Textual answers can be in point form.

All electronic devices (including mobile phones and phone alarms) must be switched off and remain under your desk until you leave the examination venue. No items may be taken to the toilet.

Q1 – Database Design (16 marks)

SnacksNow

You are designing a database to support "SnacksNow", a new startup in the food delivery business. SnacksNow allows people to order snacks from local shops and have them delivered by a bike rider. SnacksNow will only operate in the Melbourne area and does not plan to have more than one million customers.

How SnacksNow works

The SnacksNow smartphone app presents the customer first with a list of shops. The customer chooses a shop. Then the app shows the snacks available from that shop. The customer chooses which snacks they want from that shop: for example "3 chocolate bars and 1 bottle of water". The phone sends this order information, along with the phone's current GPS coordinates: this is the location to where the snacks will be delivered.

When a customer order is received, SnacksNow broadcasts a job offer to all riders who are within 1 km of the customer. Riders see the offer pop up on their app, and can press "accept" or "no thanks". If several riders press "Accept", our algorithm chooses one rider and gives them the job. We need to track these rider offers and responses. If no rider accepts the offer within 5 minutes, another offer is broadcast, and so on, until a rider has taken the job.

The chosen rider now goes to the shop, picks up the snacks, and delivers them to the customer. We record when the order is delivered to the customer. (You do not need to worry about the payment mechanism.)

People

If people want to order snacks, they need to first register as a customer, giving us their first and last name and mobile phone number.

Snack shops must register their name and address, and tell us the name and price of each snack they want to sell.

If someone wants to be a delivery rider, they need to register, giving us their first and last name, mobile phone number, and date of birth. Whenever a rider is on duty, their app sends us the rider's GPS coordinates about once per minute, allowing us to keep track of each rider's location. We must store a history of all rider locations.

Location coordinates

All locations 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). We use a precision of 4 decimal places, which is about 11 metres at the equator (smaller in Melbourne). For example, the Doug McDonell building is at latitude -37.7989, longitude 144.9627.

Question 1 instructions:

For the above scenario, draw a ***physical*** data model. (You must show the data-types of columns.)

Use crows-foot notation for relationships, and join the lines to the related columns. Show the cardinalities of relationships and whether they are optional or mandatory.

You do not need to add names to relationships. You do not need to write a data dictionary. If you wish, you may explain the reasoning behind any design decisions or assumptions you made, but this is not needed. You only need to submit the ER diagram.

Note that marks are largely based on a workable model that enforces all the constraints stated in the case. Marks may be lost for incorrect entities, relationships, cardinalities, attributes, data types or notation, lack of detail or internal contradictions.

Q2 – SQL (17 marks)

Consider the following database for a chain of 5 supermarkets in Melbourne. Customers take a collection of products to a checkout for scanning: this is one "Purchase". Each item within the purchase is a "PurchaseItem". Some customers identify themselves at the checkout by scanning a loyalty card. The checkouts are numbered within each store: so a given store n has checkouts n-1, n-2 etc. Each product is classified within a particular category. There are several payment methods. The tables are illustrated in the following printouts.

Store

id	address	postcode
1	1 One st	3053
2	2 Two st	3056
3	3 Three st	3010
4	4 Four st	3053
5	5 Five st	3123

Customer

id	firstName	lastName	dateofbirth	postcode
1	Andy	Alpha	1981-01-01	3053
2	Beyonce	Bravo	1982-02-02	3056
3	Cynthia	Charlie	1983-03-03	3010
4	Dave	Delta	1984-04-04	3053
5	Evie	Echo	1985-05-05	3123
6	Fred	Foxtrot	1986-06-06	3053
7	Gary	Golf	1987-07-07	3056
8	Harriet	Hotel	1988-08-08	3010
9	Irene	India	1989-09-09	3053
10	Jack	Juliet	1990-01-01	3123
NULL	NULL	NULL	NULL	NULL

Product

id	name	category	price	tax
1	apple	f	1.11	0.00
2	bread	f	2.22	0.00
3	cabbage	f	3.33	0.00
4	dates	f	4.44	0.00
5	eggs	f	5.55	0.00
6	fan	h	6.66	0.67
7	grater	h	7.77	0.78
8	hairdryer	h	8.88	0.89
9	iron	h	9.99	1.00
10	jacket	c	10.00	1.00

ProductCategory

id	name
c	clothing
f	food
h	household
NULL	NULL

PaymentMethod

id	name
c	credit
e	eftpos
m	money

Purchase

id	when	store	checkout	paymethod	customer
1	2017-06-01 01:01:01	1	1	m	1
2	2017-06-02 02:02:02	2	2	m	1
3	2017-06-03 03:03:03	3	3	e	1
4	2017-06-04 04:04:04	4	1	m	NULL
5	2017-06-05 05:05:05	5	2	m	2
6	2017-06-06 06:06:06	1	2	m	2
7	2017-06-07 07:07:07	2	1	m	NULL
8	2017-06-08 08:08:08	3	2	m	NULL
9	2017-06-09 09:09:09	4	3	m	3
10	2017-06-10 10:10:10	1	3	m	4
11	2017-06-11 11:11:11	2	5	c	NULL
12	2017-06-12 12:12:12	3	6	c	5
13	2017-06-13 13:13:13	4	3	c	5
14	2017-06-14 14:14:14	1	5	c	5
15	2017-06-15 15:15:15	2	5	c	NULL
16	2017-06-16 16:16:16	3	6	c	1
17	2017-06-17 17:17:17	1	6	c	7
18	2017-06-18 18:18:18	2	4	e	8
19	2017-06-19 19:19:19	3	2	e	NULL
20	2017-06-20 20:20:20	1	6	e	9
NULL	NULL	NULL	NULL	NULL	NULL

PurchaseItem

purchase	product	quantity
1	1	1
1	2	2
1	5	3
2	2	4
2	4	5
3	7	9
3	8	8
3	9	7
3	10	6
4	4	10
5	5	9
6	5	8
6	6	7
7	6	6
8	2	5
8	4	4
8	6	3
9	1	2
9	2	1
9	5	2
9	7	3
9	9	4

Checkout

store	id
1	1
1	2
1	3
1	4
1	5
1	6
2	1
2	2
2	3
2	4
2	5
3	1
3	2
3	3
3	4
3	5
3	6
3	7
4	1
4	2
4	3
5	1
5	2
5	3

Tasks:

Questions 2A-E require you to write one single SQL statement per question. Do not use Views or temporary tables. Format code for ease of reading. Ensure user-friendly output by ordering and renaming columns where appropriate.

A) In which postcodes do we have more than one store?
For each such postcode, show the postcode and the number of stores. (2 marks)

B) Show, for each payment type, the total payments in dollars made using that type. (3 marks)

C) What percentage of purchases have been made by customers born before 1990? (4 marks)

D) List any customers who have made purchases at the same store more than once.
Show customer id, store id, and the number of times they visited that store. (4 marks)
GROUP BY customer, store

E) List any customers who have made purchases at a store located in the same postcode they live in.
Show the customers' first and last names, the store address, and postcode. (4 marks)

Q3 – Normalisation (3 marks)

Excel Airlines operates flights between major Australian cities. It currently keeps its schedule in a spreadsheet: a sample of the spreadsheet is shown below. In this sample, flight 123 flies daily at 11am from Melbourne to Sydney, while flight 44 flies on weekends at 2pm from Sydney to Melbourne.

flightNumber	timeTakeoff	fromAirportCode	fromAirportName	toAirportCode	toAirportName	dateTakeoff	pilotId	pilotName
123	11am	MEL	Melbourne	SYD	Sydney	Fri 3rd Nov	1	Dave
123	11am	MEL	Melbourne	SYD	Sydney	Sat 4th Nov	2	Marie
123	11am	MEL	Melbourne	SYD	Sydney	Sun 5th Nov	3	Ibrahim
44	2pm	SYD	Sydney	MEL	Melbourne	Sat 4th Nov	1	Dave
44	2pm	SYD	Sydney	MEL	Melbourne	Sun 5th Nov	3	Ibrahim

The company would like to start storing their data in a relational database. Their initial data model, based on the spreadsheet, is expressed in relation notation as:

Flight(flightNumber, timeTakeoff, fromAirportCode, fromAirportName, toAirportCode, toAirportName, (dateTakeoff, pilotId, pilotName))

However this data model is not in 3rd normal form. Your job is to convert it to 3rd normal form.
You don't need to show intermediate normal forms – just the 3rd normal form you end up with.
Mark your primary keys with a solid underline, and your foreign keys with a dotted underline.
(Any attributes that are both primary and foreign keys should get both underlines.)

Q4 – Physical Design (6 marks)

A) For each of the following columns in the "supermarket" database shown in **question 2**,
write the column's correct MySQL data type, including the width in the case of string columns.

The following sub-questions of 4A are worth 1/2 mark each. (4 marks)

i) Store.id SMALLINT
ii) Customer.id INT
iii) Customer.lastname VARCHAR(50)
iv) Customer.dateofbirth DATE
v) Customer.postcode CHAR(4)
vi) Purchase.when TIMESTAMP
vii) Product.price DECIMAL(5,2)
viii) PurchaseItem.quantity TINYINT

B) For the same "supermarket" database, it has been found that the most common queries being run against the Product table are the following:

```
select price from Product where id = @input;

select price, tax from Product where name = @input;
```

Based on this evidence, which columns in the Product table should we index? (2 marks)

Both id and name appear in the WHERE clause, so they should be indexed. id is primary key, they are automated indexed in most database.

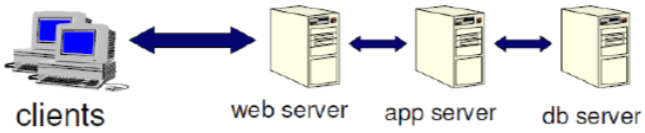
Q5 – Applications (6 marks)

A) Describe two advantages and two disadvantages of using the web to give users access to a database. (4 marks)

Ad: Everyone has a web browser, Client software not needed, HTML and HTTP are widely supported;

Dis: More complex in the middle tier, Simple standards mean hard to make complex applications, Global access potentially raise security risks

B) Draw a diagram of a 4-tier application architecture. Label each component. (2 marks)



Q6 – Distributed Databases (5 marks)

A) Describe the three (3) main options for distributing data in a distributed database. (3 marks)

Data replication; Horizontal partition; Vertical Partition

B) Describe one advantage and also one disadvantage of synchronous updates in a distributed database. (2 marks)

Ad: Data is continuously kept up to date; Dis: result in slow response time and high network usage

Q7 – Transactions (7 marks)

A) Explain, using a timeline, how locking can prevent the "Lost Update" problem. (3 marks)

the lock is not released until the statement is completed

B) Describe the four properties of transactions referred to by the acronym "ACID". (4 marks)

Atomicity; Consistency; Isolation; Durability

Q8 – Database Administration (6 marks)

A) How does the Buffer Pool work, and what is its purpose? fast retrieval (2 marks)

Detect data in BP or not; yes: process,no: read data from disk to BP; after processed, write updates(original, new, changes) back to BP and then write back to disk.

B) How does an 'SQL Injection' attack work, and how can we protect against it? (2 marks)

insert malicious codes in sql query; prepared statements, stored procedures

C) Explain the difference between physical and logical backups. (2 mark)

content, speed, output size, log, machine, recovery command

Q9 – NoSQL (4 marks)

A) Describe two (2) problems with relational databases that NoSQL databases aim to address. (2 marks)

To improve programmer productivity (or mismatch);

To handle larger data volumes and throughput (big data)

B) What does "schema-less" mean? Explain using MongoDB as an example. (2 marks)

END OF EXAM