School of Computing and Information Systems

INFO90002
Database Systems and Information Modelling
Practice Exam 2

Semester 1 2021

Reading Time: 30 minutes

Writing Time: 120 minutes (2 Hours)

This exam has 8 pages including this page

ATTEMPT ALL QUESTIONS IN ANY ORDER

Authorised Materials:

While you are undertaking this assessment, you are permitted to

- make use of any textbook, lecture slides (including soft copies)
- MySQL Workbench is supported for E.R. modelling questions
- Draw IO (diagrams.net) for Chen conceptual models or any suitable modelling equivalent
- Any lecture notes, books, laptop, PC
- You are free to use the course materials and your laptop/PC in this exam.

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
- This exam is a timed assessment which must be completed within **150 minutes** of official commencement time
- Questions can be answered in any order (please number your attempts)
- Start a new question on a new page
- **PLEASE DO NOT USE RED font colour**
- You must not communicate with other students whilst taking this exam, e.g. using messaging, chat rooms or email

IMPORTANT

- **Your file upload must be a single PDF document before the elapsed time.**

- **No other document format will be assessed (e.g. Pages, doc, txt, .SQL, etc).**

- **Email submissions will not be assessed.**

- **Every question attempt must be numbered (e.g. Q2C, Q1, Q6) to ensure it is assessed.**

- If you choose to **upload photos**, ensure that they are clear, **compressed to < 1mb**, and importantly, check them once you have uploaded. It is your responsibility to ensure your file and any images has been submitted successfully.

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)

Q2. SQL   (20 marks)

Given the schema in Figure 2, write a single SQL statement to correctly answer each of the following questions (3A – 3D). DO NOT USE VIEWS to answer questions.
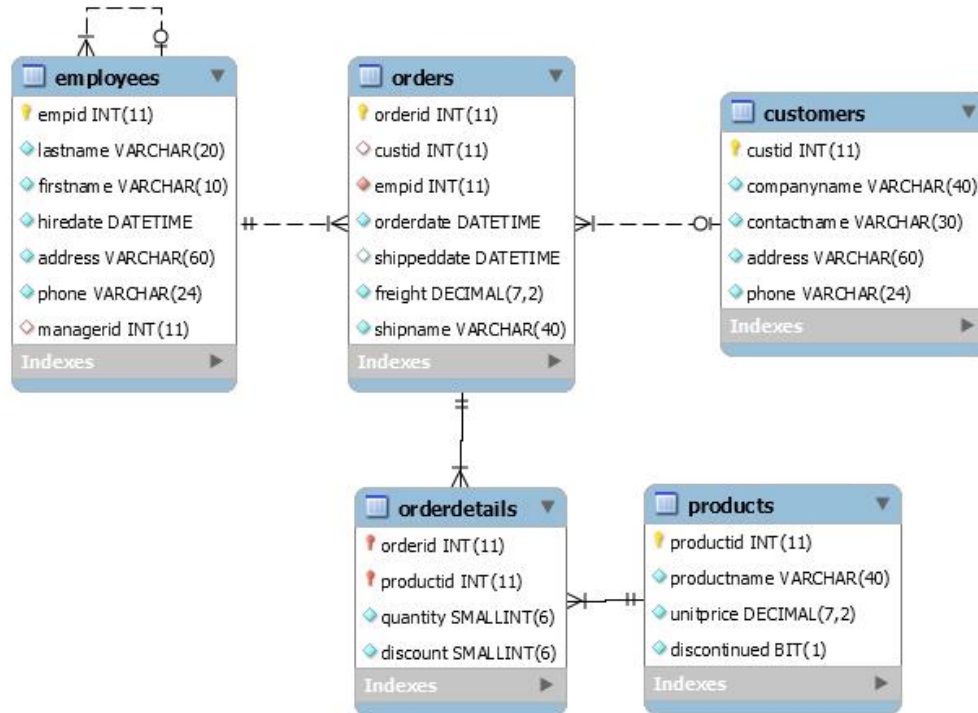

Figure 1 Company Schema

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)

Q3. Normalisation   (16 marks)

Q.3A. The table shown below is part of an office inventory database. Identify the design problems and draw a revised table structure in 3rd Normal Form (3NF) that corrects those problems. For each step explicitly identify and discuss which normal form is violated.

(Key: PK = Bold FK = Italic PFK = Bold + Italic)

Inventory (**ItemID**, Description, Quan, Cost/Unit, Dept, Dept Name, Dept Head)

ItemID is the candidate key for this table.

The following functional dependencies hold:

Dept ⟶ Dept Name and Dept Head

Quan, Cost/Unit ⟶ Inventory Value

ItemID	Description	Dept	Dept Name	Dept Head	Qunatity	Cost/Unit	Inventory Value
4011	1.4m Desk	MK	Marketing	Jane Thompson	5	200	1000
4020	Filing Cabinet	MK	Marketing	Jane Thompson	10	75	750
4005	Executive chair	MK	Marketing	Jane Thompson	5	100	500
4036	1.2m Desk	ENG	Engineering	Ahmad Rashere	7	200	1400

Table 1. The Inventory table

(10 marks)

Q.3B. Given the following relation (Inventory), and its functional dependencies - is it possible to demonstrate Armstrong's Axioms of Reflexivity, Augmentation and Transitivity?

Inventory (**ItemID**, Description, Quan, Cost/Unit, Dept, Dept Name, Dept Head(

ItemID is the candidate key for this table.

The following functional dependencies hold:

Dept ⟶ Dept Name and Dept Head

Quan, Cost/Unit ⟶ Inventory Value

(6 marks)

Q4. Data Warehousing   (10 marks)

Q4A. 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: motor cycles, 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,

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.4B. Why are star schemas preferred over relational database designs to support decision making?

(2 marks)

Q5. Storage and Indexes   (12 marks)

Q5A. Explain the components and storage hierarchy of a DBMS data file.

(4 marks)

Q5B. Sort the following values into a hash index where *h=5*.

{12, 5, 1, 34, 19, 13, 87, 2, 7, 15, 45, 8, 10, 101, 244, 910}

(2marks)

Q5.C An OLTP database has a large volume of users, with each user running a large volume and variety of concurrent DML and SELECT statements within the database (schema).

What recommendations would you make about file organization, index choice and index classification, given the volume and variety of transactions in OLTP databases? Justify your recommendations.

(6 marks)

Q6. Transactions (8 marks)

Q6A. What is the inconsistent retrieval problem? Describe the problem and use your own example to demonstrate the answer.

(4 marks)

Q6B. Describe how DBMS solve concurrency issues?

(4 marks)

Q7. NoSQL (4 marks)

Q7. 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)

Good Luck!

END OF EXAM