



The University of Melbourne
School of Computing and Information Systems

INFO90002

Database Systems and Information Modelling

EXAM

End of Semester 2, 2021

Reading Time: 15 minutes

Writing Time: 120 minutes (2 Hours)

Upload Time: 30 minutes

Authorised Materials:

While you are undertaking this assessment you are permitted to

- make use of any textbook, lecture slides (including soft copies)
- Any lecture notes and recommended texts
- You are free to use the course materials and your laptop/PC in this exam.

While you are undertaking this assessment you MUST NOT

- Copy material without attribution (including slide notes & shared study notes)
- Plagiarise
- Collude with any other person in any form
- Make use of any messaging or communication technology
- Make use of any world wide web or internet-based resources such as Wikipedia, Stackoverflow, Google, social media, or any search engine 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
- Record, broadcast, stream or distribute in any way your exam attempt

Instructions to Students:

- This exam is in 10 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 **120 minutes** of official commencement of writing 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 – YOU MUST READ THIS SECTION

- 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)

Question 1. – ER Modelling

(20 Marks)

The Residential Land Authority

The Residential Land Authority (RLA) tracks all new residential property developments in Australia. The RLA collects information from local government areas (LGA) on real estate developers and registered builders. Each LGA approves the real estate developers and registered builders who operate in their local government area.

LGA is a generic term which captures a variety of city, suburban, regional, rural, and remote areas (the level below State governments in Australia). About each LGA we store its name, population, the name of the land department ("Urban planning", "Land use development committee"), the department phone number and email address.

Real estate developers buy parcels of land in an LGA which they then apply to be rezoned as 'residential' (we do not need to track the zoning process). Real estate developers are registered with the RLA and must have a valid registration number (e.g. 2AWS390) which uniquely identifies each real estate developer. About each land parcel we store its total size in hectares, and the proposed name of the residential development (e.g. "Winter Valley Rise", "Newington Estate").

Once rezoned by the local government area, the developer subdivides the land into smaller land lots. Each residential land lot is associated with a unique land title (e.g. 11235689), and has lot dimensions (length, and width) in metres (e.g. W 18.75m L 33.20), street number, street name, street type (road, street, court, lane, etc.), suburb name and postcode. About each real estate developer we store the trading name, Australian business number (ABN) and the physical address (the structure of this field is not important) of the head office including its phone number and email. Real estate developers must provide a nominated representative to each LGA in which they operate including the representative's name and email.

Registered builders must be registered within each LGA they wish to construct residences. The RLA needs to track which builder(s) carried out construction on any given land parcel. Multiple registered builders may build on any of the land parcels owned by developers. About registered builders we store their trading name, the name of the company officer, their phone, address, and email. It is important we store the date the LGA registered the builder.

Q1. You have been asked to provide a **logical** Entity Relationship model in Crows foot notation of the Residential Land Authority case study.

(20 Marks)

Question 2. – SQL

(20 Marks)

St Clemens' Grocery

St Clemens' Grocery is an online only grocery store that specialises in fresh produce delivery (primarily fruit and vegetables) to local restaurants ('trade') in the Brunswick, Carlton, Parkville, Coburg and Essendon area. During the Covid19 pandemic they also branched out into contactless delivery for residential ('retail') customers.

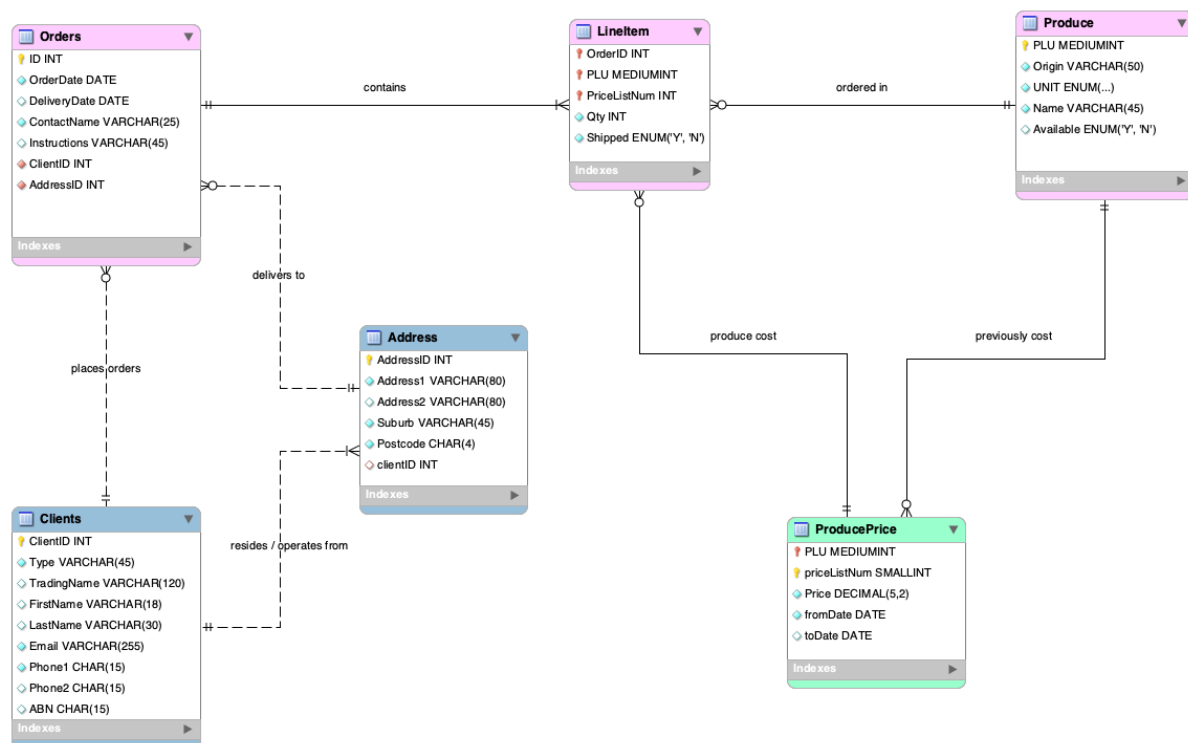


Figure 1. The Entity Relationship Model for St Clemens' Grocery

St Clemens' Business Rules

Price Lists

St Clemens' update their price list regularly. While fruit and vegetable produce may come in and out of the season, the price list does not change. The currently available produce is available in the Produce table. If the available flag is set to 'N' then the produce is not currently available to order. The ProducePrice table stores the all price lists including for currently available produce. The only way to identify if produce was not available for a previous period is that its price is set to NULL (empty). To date there have been 7 price list periods (indicated by the column pricelistnum in the ProducePrice table) from August 2020 until February 2021. Some produce stays the same price year-round whereas some other produce prices change as it comes into (and out of) the standard Australian growing season. It is important when writing queries using previous orders or time periods you ensure the correct price is applied to previous orders.

Clients

Clients can be retail or trade. Retail clients do not have an ABN (Australian Business Number) nor do they have a trading name. Clients can be in St Clemens's grocery system but have never placed an order.

Q2A. List the total number of trade orders per suburb. List the suburb, and order count. Order the result by suburb alphabetically.

(2 marks)

Q2B List the total value sold of each citrus (oranges, limes, lemons, mandarins, and grapefruits). List the name of the citrus and the total sale amount.

(3 marks)

Q2C. List the month name, year, and the amount where the total of **retail** orders has been less than \$10,000. The output should be in the form: April 2020 \$123, April 2021 \$456 etc.

(4 marks)

Q2D. Xinhue Zhang is both a retail and trade customer of St Clemens's. Write a query that lists the total cost in trade orders, total cost in retail orders and total cost of all orders, made by Xinhue Zhang. Be sure to include customers name in the output.

(7 marks)

Q2E. St Clemens' has decided to add St David's Dairy products to its current price list. Please write **the two** necessary **SQL INSERT** statement to add these rows to the St Clemens' database.

From Date	PLU	PRICE	Origin	Unit	Name
5-Nov-2021	5802	6.49	Fitzroy	ea	St David Dairy Full Cream Milk 2L
5-Nov-2021	5801	3.25	Fitzroy	ea	St David Dairy Full Crem Milk 1L
5-Nov-2021	5902	6.40	Fitzroy	ea	St David Dairy Low Fat Milk 2L
5-Nov-2021	5901	3.25	Fitzroy	ea	St David Dairy Low Fat Milk 1L

(2 + 2= 4 marks)

Question 3. – Normalisation

(10 Marks)

The Melbourne Eastern Health Network needs to normalise the vaccination records of all vaccination centres in the eastern suburbs of Melbourne. Registered nurses work at each vaccination venue and vaccinate patients with batches of specific vaccines (Pfizer, Astra Zeneca, Moderna).

Vaccination (**registeredNurse**, nurseName, venue, address, (**batch**, vaccineName, **patientID**, gender, date)

batch, registeredNurse, patientID is the candidate key for this relation.

The following functional dependencies hold

registeredNurse \rightarrow nurseName

venue \rightarrow address

batch \rightarrow vaccineName

patientID \rightarrow gender

Q3. Please convert the non-normalised Vaccination relation into 3rd normal form (3NF). Please identify the intermediate steps (1NF, 2NF) .

(10 marks)

HINT: Clearly identify the legend which identifies primary key (PK), foreign key (FK), and primary foreign key (PFK).

Question 4. – Data Warehousing (8 marks)

T2 Corporation

Q4A. T2 Corporation wants to track information about the selling of specialist teas, and tea making equipment such as diffusers, teapots, teacups, and other giftware via physical retail stores located throughout Australia. You need to design a data warehouse to report information about sales of teas and tea giftware. You need to store the retail store (store identifier, store location, delivery address and state) and product (product ID, name, category, price). The sales managers want to find the number of items sold, the revenue and profit. The information needs to be available by retail store, postcode, state, and product name and for different times (day, week, month, quarter, half year and year).

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

(8 marks)

Question 5. – Transactions (8 marks)

Q5A What is the difference between *repeatable read* and *read committed* transaction states?

(4 marks)

Q5B. Explain using an example how shared locks work to support concurrent access to data.

(4 marks)

Question 6. – No SQL (8 marks)

Q6A. Describe a “document database”, explaining how it differs from other aggregate orientated NoSQL databases

(3 marks)

Q6B. Describe a “graph database”, explaining how it differs from other types of NoSQL databases.

(2 marks)

Q6C. What are the characteristics of NoSQL transactions?

(3 marks)

Question 7. Web Applications

(8 marks)

Inspect the following HTML code

```
<HTML>
<HEAD>
<title>The American Film Institute Top 10 Westerns</title>
</HEAD> <BODY bgcolor="#333333">
<H2>The Top 10 Westerns </H2>
<OL>
  <LI>The Searchers</LI>
  <LI>High Noon</LI>
  <LI>Shane</LI>
  <LI>Unforgiven</LI>
  <LI>Red River</LI>
  <LI>The wild Bunch</LI>
  <LI>Butch Cassidy and the Sundance Kid</LI>
  <LI>McCabe and Mrs Miller</LI>
  <LI>Stagecoach</LI>
  <LI>Cat Ballou</LI>
</OL>
<H3>The Directors</H3>
<UL>
  <LI>John Ford</LI>
  <LI>Fred Zimmerman</LI>
  <LI>George Stevens</LI>
  <LI>Clint Eastwood</LI>
  <LI>Howard Hawks</LI>
  <LI>Sam Peckinpah</LI>
  <LI>George Hill Roy</LI>
  <LI>Robert Altman</LI>
  <LI>John Ford</LI>
  <LI>Elliot Silverstein</LI>
</UL>
</BODY>
<HTML>
```


Q7A. Reviewing the HTML can you identify the director of 'Red River' from this HTML? Use a paragraph (3-5 sentences) to justify your answer.

(2 marks)

Q7B. What will appear as the name of the browser tab?

(2 marks)

Q7C. How is the rendering effect different for the HTML tags and ?

(2 marks)

Q7D. Draw and label an example of a 4-tier **Web** architecture labelling and explaining each tier's role.

(2 marks)

Question 8. - Storage & Indexes (8 marks)

Q8A. Explain the difference between a clustered and non-clustered index

(2 marks)

Q8B. Explain how all database management systems (DBMS) – hierarchical, relational and NoSQL – must support three capabilities of file access.

(3 marks)

Q8C. Explain the characteristics of a dense index

(3 marks)

Question 9. – Distributed Databases (5 marks)

'UlikeMe2' is a social media company seeking to launch their social media application in English, Swedish, Norwegian, French, and Spanish speaking territories in March 2022. Customers of 'UlikeMe2' will be able to target advertising to different user profiles, and, in different languages. Users will be able to share microblogs, photos, videos, and animated gifs.

'UlikeMe2' knows that they need a NoSQL Graph *distributed* database.

Q9A. Explain what recommendations would you make about the type of distributed database should be implemented for 'UlikeMe2'?

(3 marks)

Q9B. What are the problems with using a centralised relational DBMS for this kind of application?

(2 marks)

Question 10. – VLDB & Data Analytics (5 marks)

Q10A. Why are relational databases like MySQL and Oracle **not** suited for real time VLDB data analysis such as understanding real time Twitter sentiment?

(4 marks)

Q10B. What are the requirements for any external data being shared with the AURION platform?

(1 mark)

GOOD LUCK!

END OF EXAM

One Drive File Upload

Please ensure your student ID and name is on the file and use the following link to upload your PDF document:

[One Drive File Upload Link](#)