

*Programme Code:* TU856/TU857/TU858  
*Module Code:* CMPU 4003

**TECHNOLOGICAL UNIVERSITY DUBLIN**  
Grangegorman

---

TU856 - BSc in Computer Science  
TU857 – BSc in Computer Science (Infrastructure)  
TU858 – BSc in Computer Science (International)

*Year 4*

---

*SAMPLE PAPER – Guide to answering*

---

***CMPU 4003 Advanced Databases***

**SAMPLE PAPER -Guide to answering**

**Instructions To Candidates:** Answer any **THREE (3)** Questions  
All questions carry equal marks

**Exam Duration:** 2 hours

**Special Instructions /Handouts/ Materials Required:** NA

### **General Guidance**

- You have 2 hours to tackle this exam paper – 120 mins.
- How long should you spend on each question?
  - A general heuristic is to spend the same number of minutes answering a question as there are marks allocated.
    - In this case you have to answer 3 questions. Each is worth 33 marks so a general guide would be to spend 33 minutes answering each (this includes thinking time).
    - If we round that up to 35 marks per question. That works out at 105 minutes total.
  - If you adopt the heuristic above when you reach 35 mins for a question stop and move on to another question.
    - Why? The more questions you answer the better chance you have of gaining marks.
    - E.g.
      - If you answer two full questions that is a total of 66 marks available. If you do really well you may get 90% of those marks which will give you a result of 59.4% overall.
      - If you answer three questions there will be 100 marks available. If you answer 66% of each question then this is a better result.

### **General Guidance**

- What should I do with the rest of the time?
  - If you follow the heuristic above you are left with 15 mins.
  - A suggestion would be to use this time as follows:
    - At the start of the exam spend 5 mins reading the paper. Read every question and decide which questions you are going to answer.
    - When you are finished answering all the questions spend 10 mins revisiting each question.
      - If you have left parts of the question unfinished quickly try to finish the ones that you think will be doable.
      - If you have finished answering all the questions read the questions again and make sure you have answered all parts of each question.
- How should I tackle answering a question?
  - Read the question.
  - Decide your answer.
  - Write your answer.
  - Read your answer.

1. (a) Explain *THREE (3)* possible advantages and *THREE (3)* possible disadvantages of SQL compliant relational databases.

**(6 x 2 marks)**

This is a straight-forward question. You will be able to find advantages and disadvantages in the notes.

The key here is the word explain – this means you cannot simply state an advantage or disadvantage you need to provide a bit more, explain why it is an advantage or disadvantage.

Your answer must be in your own words.

1. (b) Describe *THREE (3)* possible advantages and *THREE (3)* disadvantages of NoSQL databases.

**(6 x 2 marks)**

This is a straight-forward question. You will be able to find advantages and disadvantages in the notes.

The key here is the word explain – this means you cannot simply state an advantage or disadvantage you need to provide a bit more, explain why it is an advantage or disadvantage.

Your answer must be in your own words.

1. (c) Suppose that you are tasked with implementing a distributed data solution for a retail enterprise who wishes to achieve the following:

- a. Store details of customers, their accounts, transactions against these accounts ensuring that all data is secure and consistent. The enterprise operates in several global regions. Customers are associated with a particular global region.
- b. Implement a chat utility for customer support which is available 24 x 7. Chat participants need to be able to view a full thread of chat conversations.
- c. Explore expansion into new regional markets, predicting expected profit levels, potential challenges to stock management using not only data owned by the retail enterprise but external sources such as regional regulatory information, taxation etc.

For each of the above, state whether you would implement a SQL or NoSQL solution.

Justify your answer.

**(3 x 3 marks)**

This requires you to consider the answers you provided to part a and part b. Work out for each scenario which you think would work best SQL or NoSQL. You are making an overall choice – there may be some aspects of SQL that suit a scenario and some of NoSQL that suit a scenario. You need to make a judgment call, state it and discuss how you came to the conclusion.

2. (a) (i) Explain *FOUR* (4) key characteristics of a data warehouse.

**(4 x 2 marks)**

- (ii) Briefly compare a data warehouse and relational DBMS considering data design, data structure and access pattern.

**(7 marks)**

- (i) This is a straight-forward question. You will be able to find the characteristics in the lecture notes. The key here is the word explain – this means you cannot simply state a characteristic you need to provide a bit more about each characteristic. Your answer must be in your own words.
- (ii) This is a straight-forward question. You will be able to work this out from the lecture notes. You need to construct your answer based on the things you are asked to consider – data design, data structure, and access pattern. For each of these discuss the issue for a data design then discuss the issue for a relational database – you can comment on similarities and differences.

2. (b) (i) Explain the ACID and BASE transaction models and when you would use each.

**(12 marks)**

- (i) This is a straight-forward question. You will be able to find the details of each model in the lecture notes. The section of the course where they are discussed will allow you work out when you would use each. The key here is the word explain – this means you cannot simply state the properties for each model, you need to explain in your own words in a way that makes sense for the examples of use you provide.

2. (b) (ii) Consider the following scenarios:

- a. Departmental managers in a retail company want to identify buying patterns of individual customers and different types of customers, analyse the impact of special sales promotions and determining future pricing policy for different products.
- b. A small marketing company wants to store data from social networks and conduct sentiment analysis on this data to explore the impact of its marketing campaigns, in particular involving TV advertising during prime time. Analysis will be differentiated between weekday and weekend sentiment.

For each, state whether you consider the ACID or BASE transaction model most suitable.

**(2 x 3 marks)**

- (ii) This requires you to consider the answer you provided to part b. Work out for each scenario which you think would work best ACID or BASE. You are making an overall choice – there may be some aspects of ACID that suit a scenario and some of BASE that suit a scenario. You need to make a judgment call, state it and discuss how you came to the conclusion.

3. (a) (i) Explain what a secondary index is and its purpose in a database.

**(4 marks)**

- (ii) Suppose you are trying to improve query performance in a SQL compliant database and in a NoSQL data store. Should you add multiple indexes to the tables involved? Justify your answer.

**(5 marks)**

- (i) This is a straight-forward question. You will be able to find the answer in the lecture notes. The key here is the word explain – this means you cannot simply provide a definition. You need to explain in your own words what an index is and why you would use one. This is a straight-forward question. You will be able to work this out from the lecture notes. You need to construct your answer based on the things you are asked to consider – data design, data structure, and access pattern. For each of these discuss the issue for a data design then discuss the issue for a relational database – you can comment on similarities and differences.
- (ii) This is a straight-forward question. You will be able to find the answer in the lecture notes. The question asks for you to make a judgement but you should have a short discussion first. Why would you add multiple indexes? What would the impact be – positive and negative? What would you recommend – do it or not or do it with care? Your answer must be in your own words.

- .3 (b)** Suppose you are implementing a chat system and are designing the database for a collection of messages. Each message has an author name, recipient name, content, sequence number and timestamp.

Explain how you would implement a secondary index to facilitate pattern matching on chat content in each of the following:

- PostgreSQL
- MongoDB
- Apache Cassandra

In your answer you must explain:

- The most appropriate type of index
- The potential disadvantages

You are not required to write any code.

**(3 x 4 marks)**

This is a straight-forward question. You will be able to find the answer in the lecture notes and you have experience of doing this in the lab classes/CA. Pay attention to the detail of what you are being asked – the most appropriate type of index and the potential disadvantages. You are being asked specifically about PostgreSQL, MongoDB and Apache Cassandra so be specific about index types in those databases. Your answer must be in your own words.



3. (c) Suppose you are implementing a chat system and are designing the database for a collection of messages. Each message has a sender id, recipient id, chat content, sequence number and timestamp.

(i) Other than using indexes, explain *TWO (2)* approaches which could be used to improve performance in a document NoSQL database and *TWO (2)* approaches which could be used in a distributed NoSQL wide column database.

**(4 x 2 marks)**

(ii) Provide examples of situations in which you would use each approach.

**(4 x 1 mark)**

You do not need to provide any code.

- (i) This is a pretty straight-forward question. You can find the answer in the lecture notes. Here you are looking at MongoDB and Apache Cassandra from your experience. Explain in your own words what else you can do to improve performance in each of these.
- (ii) Using your answer to part (i) provide examples of how you would use each approach.

4. (a) What is partitioning?

How can vertical and horizontal partitioning be used to improve database performance?

**(10 marks)**

This is a pretty straight-forward question. You can find the answer in the lecture notes.

Explain what partitioning is then explain what vertical and horizontal partitioning are (this will explain the difference) then consider how each can improve database performance – so think about what could be partitioned in each approach.

(b) Explain the difference between partition and replication.

**(4 marks)**

This is a pretty straight-forward question. You can find the answer in the lecture notes.

Explain what replication is first then then explain what is different to partitioning.

(c) (i) Explain the CAP theorem.

**(6 marks)**

(ii) Why is the PACELC extension to CAP important?

**(3 marks)**

(i) This is a pretty straight-forward question. You can find the answer in the lecture notes.

Explain in your own words what the CAP theorem is and then explain the the C, A and P.

(ii) This is pretty straight-forward question. You can find the answer in the lecture notes.

4. (d) Suppose that you are tasked with implementing a distributed data solution for a retail enterprise who wishes to achieve the following:
- a. Store details of customers, their accounts, transactions against these accounts ensuring that all data is secure and consistent. Online retail applications must be available 24 x 7.
  - b. Implement a chat utility for customer support which is available 24 x 7. Chat participants need to be able to view a full thread of chat conversations.
  - c. Explore expansion into new regional markets, predicting expected profit levels, potential challenges to stock management using not only data owned by the retail enterprise but external sources such as regional regulatory information, taxation etc.

Discuss the implications of each property of the CAP theorem for each scenario above and for each identify which properties are most important. Justify your answer.

**(3 x 3 marks)**

This requires you to consider the answer you provided to part c. Work out for each scenario which properties are most important. You are making an overall choice – there may be some aspects of each property that suit a scenario. You need to make a judgment call, state it and discuss how you came to the conclusion.