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

CMPU 4003 Advanced Databases

SAMPLE PAPER

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

Exam Duration: 2 hours

Special Instructions /Handouts/ Materials Required: NA

1. (a) Explain *THREE (3)* possible benefits of denormalization and *THREE (3)* possible disadvantages of denormalization.

(6 x 2 marks)

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

(6 x 2 marks)

(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)

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)

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

(12 marks)

(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)

3. (a) **(i)** Explain what a secondary index is and its purpose in a data store.

(4 marks)

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

(5 marks)

(b) Suppose you are implementing a chat system and are designing the data store 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

You <u>are not</u> required to write any code.

In your answer you must explain:

- The most appropriate type of index
- The potential disadvantages

(3 x 4 marks)

(c) Suppose you are implementing a chat system and are designing the data store for a collection of messages. Each message has an author name, recipient name, 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 <u>do not</u> need to provide any code.

4. (a) What is partitioning?

How can vertical and horizontal partitioning be used during the physical data design phase to improve performance?

(10 marks)

(b) Explain the difference between partition and replication.

(6 marks)

(c) **(i)** Explain the CAP theorem.

(5 marks

(ii) Why is the PACELC extension important?

(3 marks)

(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.

(iii) 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)