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

SEMESTER 1 EXAMINATIONS 2022/23

CMPU 4003 Advanced Databases

Internal Examiner(s):
Dr. Deirdre Lawless
Dr. Paul Doyle

External Examiner(s):
Ms. Sanita Tifenale
Ms. Caroline McEnroy

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) (i) Explain what is meant by the term denormalization.

(4 marks)

(ii) Explain *TWO (2)* situations when you could use denormalization.

(2 x 3 marks)

(iii) Explain *FOUR (4)* possible benefits of denormalization.

(4 x 2 marks)

(b) (i) Explain *THREE (3)* key differences between SQL and NoSQL databases.

(3 x 2 marks)

(ii) Suppose that for each of the following scenarios you are implementing a distributed data solution. For each, state whether you would implement a SQL or NoSQL solution. Justify your answer.

a. A financial organisation wants to store details of customers, their accounts, transactions against these accounts and generate monthly and annual statements for both customers and the organisation as a whole. Customers conduct business 24 hours a day and 7 days a week and may access their accounts using different applications.

b. A university wants to provide its students with a chat application that can be used during class time to interact with peers and lecturers. Students need to be able to access it during class time. Messages need to be in order and those involved in a chat need to see all messages.

c. A city planning department working for a city transport authority want to use a range of different sources with varying structure to explore alternate traffic planning solutions, predicting expected reductions in car traffic, potential carbon emission reductions and potential health improvements.

(3 x 3 marks)

2. **(a)** Explain what is meant by the following terms:

- Database
- Data Warehouse
- Data Lake.

(3 x 2 marks)

(b) Compare and contrast a database, data warehouse and data lake considering the following criteria:

- Type of data stored
- Preparing data for use
- Data Freshness
- Typical users

(4 x 3 marks)

Question 2 continues on the next page

2. (c) Consider the following scenarios:

a. An insurance company wants to analyze data patterns across different geographical regions, analyse customer trends, and to track market movements quickly.

b. A telecommunications company wants to store details of customers, their call records, their data usage and generate monthly bills and an annual report.

c. Environmental scientists working for a city transport authority want to use a range of different sources to explore alternate traffic planning solutions, predicting expected reductions in car traffic, potential carbon emission reductions and potential health improvements.

d. 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.

e. A large oil and gas production company was an early adopter of multiple disruptive technologies, in particular cloud computing and Internet of Things. The company wishes to use large volumes of historical data to optimise drilling, lower operating costs, stay compliant with regulatory requirements and investigate the impact more stringent economic and regulatory requirements would have on their operating costs.

For each scenario state whether you consider a database, data warehouse or data lake to be the most appropriate for the identified purpose(s). Justify your choice.

(5 x 3 marks)

3. **(a)** Explain the objective of a query optimizer in a relational DBMS.

(4 marks)

(b) **(i)** Explain the difference between a primary and secondary index.

(6 marks)

(ii) Explain the difference between a sparse and dense index.

(6 marks)

(iii) State clearly whether primary and secondary indexes can be implemented as a sparse or dense index in a relational database. Justify your answer

(5 marks)

(c) Suppose you are implementing a blogging system and are designing the data store for a collection of posts. Each post has an author name, title, content and timestamp.

(i) Other than using indexes, explain *TWO (2)* approaches which could be used to improve performance in MongoDB providing an example of how you would use each approach.

(2 x 3 marks)

(ii) Other than using indexes, explain *TWO (2)* approaches which could be used to improve performance in Apache Cassandra providing an example of how you would use each approach.

(2 x 3 marks)

4. (a) (i) Explain what is meant by partitioning in the context of distributed data management and why it is needed.

(4 marks)

(ii) Compare and contrast how and when vertical and horizontal partitioning can be used during the physical data design phase to improve performance.

(6 marks)

(iii) Explain why partitioning is always combined with replication.

(4 marks)

(b) (i) Explain the CAP theorem and its PACELC extension.

(10 marks)

(ii) Suppose that for each of the following scenarios you are implementing a distributed data solution.

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

a. A financial organisation wants to store details of customers, their accounts, transactions against these accounts and generate monthly and annual statements for both customers and the organisation as a whole. Customers conduct business 24 hours a day and 7 days a week and may access their accounts using different online and in person.

b. A university wants to provide its students with a chat application that can be used during class time to interact with peers and lecturers. Students need to be able to access it during class time. Messages need to be in order and those involved in a chat need to see all messages.

c. 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.

(3 x 3 marks)