***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***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

*SUPPLEMENTAL EXAMINATIONS 2022/23*

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

***CMPU 4003 Advanced Databases***

**Internal Examiner(s):**

Dr. Deirdre Lawless

Dr. Paul Doyle

**External Examiner(s):**

Ms. Sanita Tifenale

Ms. Pauline Martin

**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) | 1. Explain *THREE (3)* key characteristics of a data warehouse.   (3 x 2 marks)   1. Briefly compare a data warehouse and a relational DBMS considering data design, data structure and access pattern.   (6 marks)   1. Explain what denormalization is and why denormalization is beneficial in a data warehouse.   (6 marks) |
|  | (b) | 1. Explain the differences between hot, warm and cold data.   (6 marks)   1. Identify the most appropriate data storage solution hot, warm and cold data. Justify your answer.   (3 marks) |
|  | (c) | Consider the following scenarios:   1. A smart city, an early adopter of multiple disruptive technologies, in particular cloud computing and Internet of Things, wants to support and encourage greater use of IoT devices in the private and public sector. Using a range of data, mainly historical, in a variety of formats from difference sources, planners want to investigate how best to deliver improved connectivity, reduce carbon emissions, remaining compliant with regulatory requirements, and to investigate the impact more stringent environmental regulation would have on their plans. 2. A retail company with outlets in a number of cities wants to store details of their stores and employees, products for sale, record all sales made and generate monthly sales performance reports for each store and an annual company sales report.   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.  (2 x 3 marks) |

|  |  |  |
| --- | --- | --- |
| 2. | (a) | 1. Describe *THREE (3)* situations where NOSQL has still substantial drawbacks compared to SQL relational DB.   (3 x 2 marks) |
|  |  | 1. Compare and contrast the ACID and BASE models of transaction management.   (9 marks) |
|  | (b) | 1. Explain the terms Database as a Service (DbaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).   (3 x 2 marks)   1. Suppose a small marketing company wants to store data from social networks and conduct sentiment analysis on this data. To date the company has had a single employee managing their technical environment. They are considering using a cloud database to support their sentiment analysis project.   Discuss the pros and cons of using an IaaS solution and the pros and cons of using a PaaS solution.  (8 marks)   1. Based on your discussion in part b (ii), make a recommendation to the marketing company. Justify your answer.   (4 marks) |

|  |  |  |
| --- | --- | --- |
| 3. | (a) | 1. Explain what a B-tree index is and how it can be used to improve a database performance.   (6 marks)   1. If you are trying to improve query performance in a relational database, should you add multiple indexes to the tables involved? Justify your answer.   (3 marks) |
|  | (b) | Suppose we are developing a blogging system. You are tasked with improving the performance of pattern matching for queries involving posts.  Explain how you would implement an index to address this in each of the following:   * PostgreSQL * MongoDB * Apache Cassandra.   In your answer you must explain:   * The most appropriate type of index * The potential disadvantages   Note: You do not need to write any code.  (3 x 4 marks) |
|  | (c) | 1. Other than using indexes, explain *TWO (2)* approaches which could be used to improve query performance in MongoDB.   (2 x 3 marks)   1. Other than using indexes, explain *TWO (2)* approaches which could be used to improve performance in Apache Cassandra.   (2 x 3 marks) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4. | **(a)** | | 1. Explain what partitioning is and explain briefly when you would use either vertical or horizontal partitioning.   (5 marks)   1. Explain what replication is and when you would use it.   (5 marks)   1. Explain why partitioning is always combined with replication.   (4 marks) | |
|  | | (b) | 1. Explain the CAP theorem and its PACELC extension.   (10 marks) |
|  | |  | 1. Suppose that for each of the following scenarios you are implementing a distributed data solution. For each state which properties of the CAP theorem are most important the solution. Justify your answer. 2. 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. 3. 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. 4. 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.   (3 x 3 marks) |