

**TECHNOLOGICAL UNIVERSITY DUBLIN
TALLAGHT CAMPUS**

**Higher Diploma in Science in Computing
Higher Diploma in Science
Bachelor of Science (Honours)
Bachelor of Science
Computing
Computing with Languages
Computing with Software Development
Computing with IT Management
Computing with Data Analytics**

Full Time and ACCS

Semester Five : January 2022

Big Data Technologies

Internal Examiners

Mr. Seán McHugh

Mr. Phillip Fitzpatrick

External Examiners

Mr. Thomas Davis

Day Thursday

Date 13th January 2022

Time 09.30-11.30

Instructions to Candidates

Please answer any three out of four questions on the paper.

Note Question 1 carries 34 marks. All other questions carry 33 marks.

Please start each question on a new page.

QUESTION 1 (34 Marks)

- a) Oracle Database provides object relational features. Explain the main features of an **object type** in the database. Provide an example to demonstrate your answer.

With the aid of an example, demonstrate how you would create inheritance hierarchy in this environment. What advantage does this give when storing objects in the database?

(12 Marks)

- b) It has been decided to rationalise the relational tables below.

Module Exam Result table is to be removed (maximum 60 exam results per student) and data is to be stored as a column in the Student table. Briefly explain the object relational collection type options available to you. Discuss their similarities and differences. Which one is the most suitable? Justify your answer.

(10 Marks)



Student(Id, StudentName, DOB, DegreeProgramme)

ModuleExamResult(moduleName, ExamResult, ResultDate)

- c) Explain the use of **REF** in Oracle and why one would use it. Contrast this approach to using primary key/foreign key relationships in a relational database. Why would one use **SCOPE** in defining a REF column?

Provide examples to demonstrate your answer.

(12 Marks)

QUESTION 2 (33 Marks)

- a) Discuss THREE main differences between NoSQL database and a Relational database. Give examples to support your answer.

(9 Marks)

- b) Briefly explain **Master/Slave replication**. Provide a relevant diagram to demonstrate your answer. As part of your answer, highlight its advantages and disadvantages. Explain what happens when the Master fails.

(12 Marks)

- c) Riak's Key-Value store architecture uses a logical ring. Briefly discuss the structure of the ring. Provide a diagram to demonstrate your answer.

As part of your answer, highlight the significance of Consistent Hashing and the Hinted Hand-Off in the operation of the ring.

(12 Marks)

QUESTION 3 (33 Marks)

Data warehouse data can be described as a subject-oriented, integrated, time-variant, and a non-volatile collection of data in support of management's decision-making process (Inmon, 1993).

- a) Explain the main characteristics of data in a data warehouse as defined in the above definition. Provide examples to demonstrate each characteristic.

(10 Marks)

- b) With respect to Data Warehouse design, describe star, snowflake and star flake schemas indicating how they differ. Provide examples to demonstrate your answer.

(13 Marks)

c)

- i) Explain the term Conformed Dimensions and the circumstances under which they facilitate the "drill-across" operation.

(4 Marks)

- ii) Explain the three types of Slowly Changing Dimensions (SCDs). Support your answers with examples.

(6 Marks)

QUESTION 4 (33 Marks)

- a) Big Data is often characterised in the context of “the 5 Vs”. Discuss these characteristics in the context of NoSQL databases.

(10 Marks)

- b) Briefly explain how data is stored in Hadoop Distributed File System (HDFS). Explain the roles of the NameNode, the DataNode and the SecondaryNameNode in the Hadoop Cluster.

(15 Marks)

- c) How does the NameNode know when a DataNode fails? What action does it take?

(8 Marks)