TECHNOLOGICAL UNIVERSITY DUBLIN
TALLAGHT CAMPUS

Bachelor of Science (Honours)
Bachelor of Science
Higher Diploma in Science

IT Management

Computing
Data Analytics

Software Development

Full Time and ACCS

Semester Five : January 2020

Big Data Technologies (Advanced Databases)

Internal Examiners
Mr. Phillip Fitzpatrick
Mr. Seán McHugh

External Examiners
Mr. Thomas Davis

Day **Tuesday**
Date **14th January 2020**
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 how you would create an **object type** in the database. With the aid of an example, demonstrate how you would create an inheritance hierarchy in this environment. What advantage does this give when storing objects in the database?

(9 Marks)

b) With the aid of examples, briefly explain the TWO options in Oracle to store data in the database against an existing Object Type definition.

(8 Marks)

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?

(9 Marks)

d) With regard to a REF value, explain **Implicit** and **Explicit** Dereferencing. Give examples to demonstrate your answer.

(8 Marks)

QUESTION 2 (33 Marks)

The general consensus is that there are specific characteristics that define Big Data. In most Big Data circles, they are called the **Four** V's: **Volume**, **Variety**, **Velocity**, and **Veracity**.

a) Briefly explain the FOUR V's identified above? Give examples to support your answer

(12 Marks)

b) Briefly explain Peer to Peer (masterless) Replication? Use a diagram to demonstrate your answer. What advantages does it bring? What are its main disadvantages?

The Hinted-Handoff Protocol is used in Peer-Peer NoSQL environments like Cassandra. Explain this protocol by providing an example (and supporting diagram) to demonstrate the protocol in action.

(13 Marks)

c) Briefly discuss the forms of Consistency supported in a NoSQL database. Gives examples to demonstrate your answer. How does this compare to the Consistency provided in a Relational DBMS?

(8 Marks)

QUESTION 3 (33 Marks)

a) In the context of Hadoop Distributed File System (HDFS) AND YARN, briefly explain the main daemons and their roles in a Hadoop cluster.

(12 Marks)

b) Briefly explain how data is stored in Hadoop Distributed File System(HDFS). Explain how a write operation is supported in HDFS. Provide a diagram to support your answer.

(15 Marks)

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

(6 Marks)

QUESTION 4 (33 Marks)

As a Data Warehouse Consultant, you have been asked to brief a client company whose experience has been in online transaction processing-based systems and are embarking on a data warehouse project.

a) Highlight a common Data Warehouse architecture briefly discussing its main components. Provide a diagram to support your answer.

(10 Marks)

b) What are **Independent Data Marts** and **Dependant Data Marts**? Which one would you recommend and why?

(5 Marks)

c) Dimensional modelling is a logical design technique used to design star and derivative schemas.

i) Compare and contrast Star and Snowflake schemas explaining when you would use one over the other. Use diagrams to support your answer.

(8 Marks)

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

(4 Marks)

d) Explain the **three** types of Slowly Changing Dimensions (SCD's). Support your answers with examples.

(6 Marks)