Section B
Answer Section B questions in Answer Book B

B4.

a) Explain how the following **THREE** traits of Big Data help to differentiate Big Data processing from other data processing:

i) Variety;
ii) Veracity;
iii) Value.

(15 marks)

b) Explain why a traditional Relational database management system is generally considered unsuitable as the primary platform for processing the large volume, high velocity, unstructured data associated with a Big Data processing task.

(10 marks)

B5.

a) Explain the benefits of a NoSQL database system for Big Data processing.

(10 marks)

b) Describe the following **THREE** properties of Brewer's CAP theorem:

i) Consistency;
ii) Availability;
iii) Partition tolerance.

(9 marks)

c) Explain why only **TWO** of the **THREE** properties of the CAP theorem can be simultaneously supported in a distributed database cluster.

(6 marks)

B6.

a) Describe the principal goals of machine learning.

(5 marks)

b) Briefly explain learning bias in a machine learning algorithm.

(5 marks)

c) Explain, with an example, how a supervised machine learning algorithm can be used in a data classification task.

(15 marks)

End of Examination

BCS THE CHARTERED INSTITUTE FOR IT

BCS HIGHER EDUCATION QUALIFICATIONS
BCS Level 5 Diploma in IT

BIG DATA MANAGEMENT

Thursday 6th May 2021 - Morning

Answer **any** FOUR questions out of SIX. All questions carry equal marks.

Time: TWO hours.

Answer any Section A questions you attempt in Answer Book A
Answer any Section B questions you attempt in Answer Book B

The marks given in brackets are **indicative** of the weight given to each part of the question.

Calculators are **NOT** allowed in this examination.

Section A
Answer Section A questions in Answer Book A

A1.

a) Explain the defining characteristics of the following **TWO** data types:

i) Structured data;
ii) Unstructured data.

(10 marks)

b) Explain the main issues to be considered in the processing of large volumes of fast real time streamed data.

(5 marks)

c) Describe the basic components of the Kafka event data streaming platform.

(10 marks)

A2.

a) Describe **TWO** advantages and **TWO** disadvantages to outsourcing a Big Data project to an external supplier.

(10 marks)

b) Big Data network infrastructure requirements have identified properties that are crucial to effective handling of Big Data. Explain the following **THREE** network properties and state how they can be optimised for handling Big Data processing:

i) Network resilience;
ii) Network partitioning;
iii) Network application awareness.

(15 marks)

A3.

a) Describe the Comprehensive R Archive Network (CRAN).

(5 marks)

b) It is a common view that the R platform is unsuited to dealing directly with Big Data. Briefly explain why this view **might** be taken.

(5 marks)

c) A vector of nine numbers is created in R, by the following script:

```
xvar <- c (-20,7,3.5, -7,16,31, -1,11,30).
```

For this vector `xvar`. Write R scripts using base R functions to compute the following statistics:

i) The Median of `xvar`;
ii) The Mean of `xvar` using the Trim option to remove the leading and trailing two numbers in the vector.

(5 marks)

d) Write an R script that implements your own user function to compute the statistical Mode of a data vector. Make use of any other base R functions within your script.

(10 marks)

[Turn Over]