

BCS THE CHARTERED INSTITUTE FOR IT
BCS HIGHER EDUCATION QUALIFICATIONS
BCS Level 5 Diploma in IT

BIG DATA MANAGEMENT

SAMPLE PAPER

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 how each of the following **THREE** characteristics of Big Data define the differences between Big Data and other data processing:

- i) Volume;
- ii) Velocity;
- iii) Variety.

(12 marks)

b) Describe **FOUR** major sources of machine generated data that might be found in a Big Data processing task.

(13 marks)

A2.

a) Describe **THREE** important skillsets that need to be considered in building a Big Data science team.

(9 marks)

b) Give **TWO** examples of the way in which international privacy laws might affect the use of personally identifiable information.

(10 marks)

c) Compare and contrast data privacy and data protection.

(6 marks)

A3.

a) Describe with the aid of a diagram, the Hadoop distributed file system architecture, and explain how it stores data across a cluster of machines.

(10 marks)

b) Explain how the MapReduce pipeline is used for program execution in the Hadoop system.

(15 marks)

[Turn Over]

Section B

Answer Section B questions in Answer Book B

B4.

- a) It is widely acknowledged that Relational Database Management Systems cannot always support the rapid growth in data storage. Give brief comments on why this view is often taken.

(5 marks)

- b) Explain why the use of a Hierarchical Data File Store (HDFS) in products like Hadoop can offer significant advantages when processing Big Data.

(10 marks)

- c) One method of storing Big Data is within key-value pairs. Explain key-value pairs and give two scenarios in which the use of key-value pairs would be applicable.

(10 marks)

B5.

- a) Gartner's ascendency model provides **FOUR** levels of analytics for information. Explain the type of analytics required at **EACH** level, and state the most appropriate tools to use at each level.

(15 marks)

- b) Give an example of an Artificial Intelligence method that uses an unsupervised learning technique and describe the value that this type of method provides to the analyst and business.

(10 marks)

B6.

- a) In Big Data computing, proprietary distributed file systems which are optimised for maximising efficiency for a particular set of applications can be used. The Google File System (GFS) cluster is an example of such a distributed file system.

With the aid of a diagram, explain the architecture of the GFS cluster and suggest **TWO** ways in which this architecture is optimal for some type of applications.

(15 marks)

- b) In a Big Data initiative, you might consider the use of Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) from a public cloud provider. Explain what these services provide and give **TWO** advantages and **TWO** disadvantages you might expect from each service.

(10 marks)

END OF EXAMINATION