

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

BIG DATA MANAGEMENT

Wednesday 5th October 2022 – 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) What are the main costs to an organisation in the management of Big Data?
(3 marks)
- b) Describe the main issues for an organisation in each of the following?
- i) Data Gathering **(3 marks)**
 - ii) Data Storage **(3 marks)**
 - iii) Data Quality **(3 marks)**
 - iv) Data Resilience to growth. **(3 marks)**
- c) What is personal data, and what are the challenges of managing personal data?
(5 marks)
- d) Explain the importance of external compliance regulations.
(5 marks)

A2.

- a) Describe what you understand by cloud storage and onsite storage methods.
(5 marks)
- b) What are the main advantages and disadvantages for an organisation in each of the Cloud storage methods?
(10 marks)
- c) Describe and explain the primary impacts of the distributed file system Hadoop on big data processing?
(5 marks)
- d) Describe the financial and security implications of cloud-based technology?
(5 marks)

A3.

- a) Describe and explain Gartner's maturity model. **(7 marks)**

- b) Describe the data analysis process and how it can help provide valuable information. **(5 marks)**

- c) Explain and provide **THREE** examples of the use of AI to analyse data. **(7 marks)**

- d) Provide **THREE** examples of neural networks. **(3 marks)**

- e) Describe and explain the functionality of Bloom Filters. **(3 marks)**

[Turn Over]

Section B
Answer Section B questions in Answer Book B

B4.

- a) Explain why the following three characteristics of Big Data support the view that Big Data processing differs from other forms of data processing.
- i) Volume **(3 marks)**
 - ii) Velocity **(3 marks)**
 - iii) Variety. **(3 marks)**
- b) Explain what cluster analysis is and state why it is particularly useful in the processing of large volumes of data. **(7 marks)**
- c) Give **TWO** examples of applications where cluster analysis would be useful. **(5 marks)**
- d) Describe the syntax of the following standard R function code for performing clustering.
- ```
kmeans(x, centers, iter.max = 10, nstart = 1, algorithm =
c("Hartigan-Wong", "Lloyd", "Forgy", "MacQueen"),
trace=FALSE)
```
- (4 marks)**

**B5.**

- a) Explain why Big Data is sometimes regarded as a challenge for data processing and storage management. **(7 marks)**
- b) Explain how some consider Apache Spark to offer an advantage compared to Apache MapReduce in the following three areas.
- i) Data Processing **(6 marks)**
  - ii) Programming tasks **(6 marks)**
  - iii) Business resource costs. **(6 marks)**

**B6.**

- a) Explain the following three characteristics of Brewer's CAP theorem.
- i) Consistency **(3 marks)**
  - ii) Availability **(3 marks)**
  - iii) Partition tolerance. **(3 marks)**
- b) Explain how the following two NoSQL database systems are classified according to the CAP characteristics they support.
- i) Apache MongoDB
  - ii) Apache Cassandra. **(16 marks)**

**End of Examination**