

B6.

a) Describe the following four NoSQL database stores and state the type of data **each** would typically contain:

- i. Column oriented
- ii. Document oriented
- iii. Key-value oriented
- iv. Graph oriented.

(12 marks)

b) Describe the following three types of databases which are classified according to Brewer's CAP theorem:

- i. CP database
- ii. AP database
- iii. CA database.

(9 marks)

c) It is often stated that a NoSQL distributed database system running on a cluster cannot be a CA database. Explain this statement with reference to Brewer's CAP theorem.

(4 marks)

END OF EXAMINATION

BCS THE CHARTERED INSTITUTE FOR IT

BCS HIGHER EDUCATION QUALIFICATIONS
BCS Level 5 Diploma in IT

BIG DATA MANAGEMENT

Monday 15th April 2024 - Afternoon

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 any Section A questions you attempt in Answer Book A

- A1.**
- a) What should be considered when formulating strategies for Big Data? (8 marks)
 - b) What are considered to be the main advantages in the use of big data management systems? (7 marks)
 - c) Explain Privacy by Design and why it is important. (7 marks)
 - d) In which circumstances is personal data **not** covered by the General Data Protection Regulation (GDPR)? (3 marks)
- A2.**
- a) Explain what you understand by cloud and onsite storage method. (7 marks)
 - b) What are the advantages and disadvantages of Cloud storage? (7 marks)
 - c) Describe data modelling for the Entity Relationship (E-R) Model and UML (Unified Modelling Language) methods. (7 marks)
 - d) Identify and explain the stages in data modelling. (4 marks)
- A3.**
- a) Describe the seven dimensions mapped against **each** stage of Gartner's ascendancy model. (7 marks)
 - b) Give reasons for the introduction of the four stages in the Gartner's maturity model. (5 marks)
 - c) Describe the benefits of data analysis for an organisation. (7 marks)
 - d) Provide **TWO** examples of basic machine learning algorithms. (3 marks)
 - e) Briefly describe how Artificial Intelligence (AI) is used to analyse data. (3 marks)

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

- B4.**
- a) Explain, with an example, **each** of the following **THREE** defining characteristics of big data:
 - i. Volume
 - ii. Variety
 - iii. Veracity.(12 marks)
 - b) Explain why it is suggested that a distributed real-time or near real-time data processing system can only ever simultaneously support two of the three big data requirements for high-speed high-volume and highly consistent data processing. (10 marks)
 - c) Give an example of the type of data analytics you might carry out with an R k-means clustering function. (3 marks)
- B5.**
- a) Explain why using MapReduce is often considered an advantage compared to Apache Spark for the following two big data attributes:
 - i. Security of data
 - ii. Hardware costs of processing data.(10 marks)
 - b) Explain why Apache Spark is often considered superior in performance to MapReduce for speed of processing data. (5 marks)
 - c)
 - i. Describe **TWO** types of big data application where the Map Reduce framework is considered most suited.
 - ii. Describe **TWO** types of big data application where the Apache Spark framework is considered most suited. (10 marks)

[Turn Over]