

BCS THE CHARTERED INSTITUTE FOR IT

BCS HIGHER EDUCATION QUALIFICATIONS
BCS Level 5 Diploma in IT

BIG DATA MANAGEMENT

Friday 4th October 2024 - 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 challenges for management in the implementation of Big Data processing. **(8 marks)**
 - b) Describe the term 'unstructured data' and provide **three** examples of it. **(5 marks)**
 - c) Provide an assessment of the challenges of managing personal data. **(8 marks)**
 - d) Provide **four** examples of personal data. **(4 marks)**
- A2.**
- a) Describe the main challenges for a company when managing Big Data. **(7 marks)**
 - b) Describe the key requirements for the storage of Big Data. **(5 marks)**
 - c) Outline the main security risks of cloud-based storage. **(5 marks)**
 - d) Describe the Hadoop distributed file system and explain how it is used for Big Data processing. **(8 marks)**
- A3.**
- Understanding the tools and techniques that may be deployed to assist in the processing of Big Data.
- a) Provide a comparison of the three types of data storage – Object, File, and Block. Explain the pros and cons of **each** of their uses. **(15 marks)**
 - b) Explain what you understand by the term 'neural networks'. **(5 marks)**
 - c) Explain the term 'data mining'. **(5 marks)**

Section B
Answer Section B questions in Answer Book B

- B4.**
- a) Describe how the following five Big Data characteristics are used to differentiate Big Data from other forms of data. **(15 marks)**
 - i. Volume.
 - ii. Velocity.
 - iii. Variety.
 - iv. Veracity.
 - v. Value.
 - b) Explain, with a real-world example, the type of Big Data you would expect from the following two categories of data source. **(10 marks)**
 - i. Machines.
 - ii. People.
- B5.**
- a) Explain the following two Big Data processing terms associated with real time streamed data. **(10 marks)**
 - i. Event stream processing.
 - ii. Complex event processing.
 - b)
 - i. Explain how the Mapper process and the Reducer process **each** contribute to making MapReduce generally unsuitable for real time Big Data processing.
 - ii. Describe **two** strategies that could be adopted so that MapReduce can be used for near real time processing. **(15 marks)**
- B6.**
- a) Explain how the Hadoop Distributed Filesystem (HDFS) manages Big Data sets in secondary storage. **(9 marks)**
 - b) Explain the following four database transaction management design principles and state why some of the principles might pose a particular problem in a Big Data environment. **(16 marks)**
 - i. Atomicity.
 - ii. Consistency.
 - iii. Isolation.
 - iv. Durability.

END OF EXAMINATION