

BCS Higher Education Qualification

Diploma

MAY 2021

EXAMINERS' REPORT

BIG DATA MANAGEMENT

General comments

Very few candidates attempted section A of the examination. Section (B) of the paper contained the most popular questions. It covered the defining characteristics of Big Data – Laney's v's and later extensions, volume, velocity, variety, valence, and value, typical measures of magnitude and the challenges of Big Data – storage, infrastructure, technologies, processing.

Question number: A1

Syllabus area:

- 1 Big Data fundamentals
- 1.3 Challenges of Big data – storage, infrastructure, technologies, processing
- 3 Storage, data and programming interface models for Big Data
- 3.3 Data modelling – structured data, unstructured data and streamed data
- 3.4 Modelling real time data – pipelines with Kafka and similar technologies

Total marks allocated: 25

Examiners' Guidance Notes

Most candidates achieved reasonably high marks for this question. For part a the essential distinction between structured and unstructured data was answered. Part b was generally well answered although some candidates did not address consistency as a main issue. Part c was well answered with the key areas of Kafka comprehensively described by many. A few attempts showed difficulty in going beyond a simple publish and subscribe definition and tended to gain lower marks for this part of the question.

Question number: A2

Syllabus area:

- 2 Management issues for Big Data
- 2.1 Staffing, skillsets, strategies, total cost of ownership
- 5 Big Data infrastructures
- 5.1 Networks, cloud based – in house v outsourced

Total marks allocated: 25

Examiners' Guidance Notes

Part a of this question was generally less well answered. Many answers tended to confuse the costs and benefits between being an advantage or disadvantage and few answers mentioned any advantage to having flexibility in development methods. For part b most answers indicated a good appreciation of the three network properties but did not state how they can be optimised for Big Data processing in particular. This reduced the potential to gain higher marks for this part of the question.

Question number: A3

Syllabus area:

7 Introduction to R

7.2 R programming language – language and ecosystem basic syntax

7.3 Basic statistics in R – mode, mean, variance, moments

Total marks allocated: 25

Examiners' Guidance Notes

No attempts were made for this question

Question number: B4

Total marks allocated: 25

Examiners' Guidance Notes

This was a very well answered question with most candidates achieving high marks. A small number of candidates did not provide sufficient depth in their answers

Question number: B5

Total marks allocated: 25

Examiners' Guidance Notes

Candidates were able to demonstrate good knowledge on these topics. Nearly all provided satisfactory or better answers

Question number: B6

Total marks allocated: 25

Examiners' Guidance Notes

A reasonably well answered question. Again more depth in the answers would have helped