This question paper consists of 4 printed pages each of which is identified by the Code Number COMP5111M01

This is an open book examination.
This means that you may take with you any written or printed material into the examination room. You are unlikely to win credit in any question for reproduction or simple rephrasing of notes.

School of Computing
May/June 2018
COMP5111M01
Big Data Systems

Answer all three questions

Time allowed: 2 hours

Question 1

(a) The term “Big Data” is widely used but often misunderstood.

(i) Explain the concept of Big Data and provide an example application. **[2 marks]**

(ii) Describe the main characteristics, referring to the V’s, of a Big Data system along with examples for each. **[8 marks]**

(b) IBM proposed a layered architecture for Big Data systems.

(i) How do the business layers relate to Smart Cites? **[4 marks]**

(ii) How do the technical layers relate to a Big Data system that is part of a smart city infrastructure? **[4 marks]**

(c) Within the financial sector Big Data systems can have significant implications. Discuss one example of where this has gone wrong in industry. **[2 marks]**

[question 1 total: 20 marks]

Question 2

(a) One of the most commonly used tools for a Big Data system is Hadoop. Describe how the architecture of Hadoop, which is deployed at the messaging and storage layer of a Big Data system, relates to the analysis layer and to database technologies.

[6 marks]

(b) A major challenge in Big Data systems is managing data flow across the network and through the necessary systems.

(i) One approach is the use of stream processing. Describe one or more technologies that could be used to manage streams and manipulate data as it flows through the system. Discuss how this relates to the Hadoop framework.

[5 marks]

(ii) The AMQP protocol is often used. Outline the benefits of using this or a related protocol and explain how such an implementation could be used as part of a Big Data system. **[4 marks]**

(c) Data governance is already a huge issue with Big Data systems. Discuss some of the challenges that organisations will face with the introduction of GDPR and how any technology could be used to help enforce it. **[5 marks]**

[question 2 total: 20 marks]

Question 3

(a) Fault tolerance and dependability are important concepts in computer systems. Discuss approaches of at least two Big Data systems such as Hadoop for dealing with failures. **[4 marks]**

(b) Blockchain has received a lot of interest since its conception in 2007. It is now being applied to a wide range of application domains. Consider the domain of pensions where you have:

- A pension fund manager
- Members
- Investments

The following transactions can occur:

- Member contribution, paid monthly
- Calculation of annual return on investments
- Payments to pensions

Discuss how blockchain could be used for managing such a system, how smart contracts could automate some of the transactions, and how permissions could be managed. Explain why blockchain makes sense for this application? **[8 marks]**

(c) One of the most challenging areas of blockchain and other distributed ledger technologies is managing consensus. Discuss why this is challenging making reference to at least two protocols. **[6 marks]**

(d) Explain how permissioning could be used to make the consensus management easier. **[2 marks]**

[question 3 total: 20 marks]
[grand total: 60 marks]

END