THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

This exam paper must not be removed from the venue

Venue _____
Seat Number _____
Student Number |__|__|__|__|__|__|__|__|
Family Name _____
First Name _____

School of Information Technology and Electrical Engineering

EXAMINATION

Semester One Final Examinations, 2021

DATA7201 Data Analytics at Scale

This paper is for St Lucia Campus students.

Examination Duration: 120 minutes

Reading Time: 10 minutes

Exam Conditions:

This is a Closed Book examination - specified written materials permitted
No calculators permitted
During reading time - write only on the rough paper provided
This examination paper will be released to the Library

Materials Permitted In The Exam Venue:
(No electronic aids are permitted e.g. laptops, phones)

One A4 sheet of handwritten or typed notes single sided is permitted

Materials To Be Supplied To Students:

None

Instructions To Students:

Additional exam materials (e.g. answer booklets, rough paper) will be provided upon request.

Students to answer Q1-Q7 in the space provided in the question paper.

Total marks = 100

For Examiner Use Only

Question	Mark

Total _____

Question 1. (14 marks)

Explain the benefit of the Shuffle phase in Map/Reduce.

Question 2. (14 marks)

Discuss why Apache Spark can be used for different big data problems (e.g., volume, variety, velocity).

Question 3. (14 marks)

Critically compare the functioning of Apache Storm and Apache Kafka.

Question 4. (14 marks)

Critically compare Apache Giraph and Spark GraphX for large data graph management.

Question 5. (14 marks)

Discuss the advantages and disadvantages of unsupervised and supervised opinion mining approaches also including scalability issues.

Question 6. (15 marks)

Discuss the following scenario describing the type of data infrastructure you would adopt and why. The scenario description below does not provide all the necessary details. You will need to describe your assumptions on the scenario to complement the information given to you. Describe the assumptions you are making in terms of data availability, analytics queries of interest, user expertise and requirements. 1) discuss your assumptions, 2) outline the design of your data infrastructure solution (i.e., which data, which systems, which users, etc.) and, 3) justify your solution.

Scenario: A metropolitan hospital needs to store data about their patients, staff, and physical resources like medical equipment, furniture, and other assets. The aim is to design and deploy a data solution that can support analytics over such data and may inform decision making processes (e.g., need for more ICU beds, more staff required to be available at a given point in time).

Question 7. (15 marks)

Discuss the following scenario describing the type of data infrastructure you would adopt and why. The scenario description below does not provide all the necessary details. You will need to describe your assumptions on the scenario to complement the information given to you. Describe the assumptions you are making in terms of data availability, analytics queries of interest, user expertise and requirements. 1) discuss your assumptions, 2) outline the design of your data infrastructure solution (i.e., which data, which systems, which users, etc.) and, 3) justify your solution.

Scenario: A national government needs to decide an investment strategy to support public health; they are looking for data that can inform their decision about how much to invest from a fixed budget into a) hospital infrastructure; b) healthy lifestyle campaigns (prevention); c) research on new treatments (correction).

END OF EXAMINATION