北京郵電大學

Queen Mary University of London

For examiners' use only

1	
2	
3	
4	
5	
6	
7	
8	
Total	

EBU750U A

Joint Programme Examinations 2018/19

EBU750U Cloud Computing

Paper A

Time allowed 2 hours

Answer ALL questions

Complete the information below about yourself very carefully.

QM student number										
BUPT student number										
Class number										

NOT allowed: Electronic dictionaries.
Allowed: Electronic calculators

INSTRUCTIONS

1. **You must NOT take answer books, used or unused, from the examination room.**
2. Write only with a black or blue pen **and in English**.
3. Do all rough work in the answer book – **do not tear out any pages**.
4. If you use Supplementary Answer Books, tie them to the end of this book.
5. Write clearly and legibly.
6. **Read the instructions on the inside cover.**

Examiners

Dr Gokop Goteng, Dr Gareth Tyson, Dr Richard G. Clegg

Filename: 1819_EBU750U_A No answer book required

Instructions

Before the start of the examination

1) Place your BUPT and QM student cards on the corner of your desk so that your picture is visible.

2) Put all bags, coats and other belongings at the back/front of the room. All small items in your pockets, including wallets, mobile phones and other electronic devices must be **placed in your bag in advance**. **Possession of mobile phones, electronic devices and unauthorised materials is an offence.**

3) Please ensure your mobile phone is switched off and that no alarm will sound during the exam. **A mobile phone causing a disruption is also an assessment offence.**

4) Do not turn over your question paper or begin writing until told to do.

During the examination

1) You must not communicate with or copy from another student.

2) If you require any assistance or wish to leave the examination room for any reason, please raise your hand to attract the attention of the invigilator.

3) If you finish the examination early you may leave, but not in the first 30 minutes or the last 10 minutes.

4) For 2 hour examinations you may **not** leave temporarily.

5) For examinations longer than 2 hours you **may** leave temporarily but not in the first 2 hours or the last 30 minutes.

At the end of the examination

1) You must stop writing immediately – **if you continue writing after being told to stop, that is an assessment offence.**

2) Remain in your seat until you are told you may leave.

Question 1

a) An Amazon AWS (Amazon Web Services) EC2 (Elastic Compute Cloud) cluster consists of 300 vCPUs (virtual Central Processing Units). If 80% of these processors are used to perform computational and processing activities in parallel, calculate:

[15 marks]

i) The system efficiencies (TWO separate calculations) of this AWS EC2 cluster using "fixed workload" and then "scaled workload".

(10 marks)

ii) The AWS EC2 cluster described above also has an AWS S3 (Simple Storage Service) which provides high availability (HA) to the cluster. If the cluster has a total Mean Time To Failure (MTTF) of 500 days and an average Mean Time To Repair (MTTR) of 2.5 days, calculate the System Availability of the cluster (show all calculation steps).

(5 marks)

	Do not write in this column

	15 marks

b) Describe the AWS Trusted Advisor.

[3 marks]

	Do not write in this column
	3 marks

c) Describe the FOUR ways to manage a Virtual Cluster.

[4 marks]

	Do not write in this column
	4 marks

d) Describe THREE differences between private and public cloud deployment models.

[3 marks]

	Do not write in this column

		3 marks

Question marking: $\frac{}{15} + \frac{}{3} + \frac{}{4} + \frac{}{3} = \frac{}{25}$

Question 2

a) This question is about cloud economics:

[15 marks]

i) Describe with the help of a diagram TWO arithmetic/maths models for calculating the cost of traditional IT costs and cloud computing costs respectively.

(10 marks)

ii) If the total cost of a traditional IT system is £500,000 and out of this, its upfront capital cost is £300,000, what is the cost of using cloud computing instead of traditional IT model.

(5 marks)

	Do not write in this column

	15 marks

b) This question is about Graphics Processing Units (GPU) and Compute Unified Device Architecture (CUDA).

[8 marks]

i) Describe the THREE Ways to program GPUs using CUDA platform.

(6 marks)

ii) Describe TWO properties of heterogeneous computing in GPU CUDA computing or programming.

(2 marks)

	Do not write in this column

	8 marks

c) Describe TWO uses of Amazon AWS CloudFront using your knowledge of "edge caching".

[2 marks]

	Do not write in this column

		2 marks

Question marking: $\frac{}{15} + \frac{}{8} + \frac{}{2} = \frac{}{25}$

Question 3

a) You operate an online event booking website that allows users to reserve seats in music concerts. Your website records a dataset of all the events (i.e. concerts) that have been booked by users, and your website supports various categories of music event, e.g. classical, rock, pop, jazz. The dataset contains one log entry for each time a user has interacted with the website. Therefore, each log entry records the following information [types in brackets]:

`eventID [String], userID[String], eventCategory [String], action [String], time [Date]`

`action` is a fixed **`String`** that can be either "`visit`", when the user first visits the event's webpage when browsing, or "`reserve`", when the user actually books the event. **`eventID`** is a unique identifier that is assigned to each individual event (i.e. each concert). **`eventCategory`** is a **`String`** representing the category of the event (e.g. classical, jazz).

This data should be fed as input into a **Map/Reduce (Hadoop) job** as a set of key:value pairs (**`String eventID, BookingReservation request`**). The *key* is a **`String`** object representing the unique identifier of the event being booked. The *value* is a **`BookingReservation`** object, containing details of the booking (and methods to access each of the fields). For example, **`request.getEventID()`** will return the **`eventID`** field of the input row, **`request.geUserID()`** will return the **`userID`** *etc.*

You must **write** a Map/Reduce Java program for the given input that computes what is the **most popular** event for each event category (i.e. the most frequently booked classical concert, jazz concert, pop concert *etc*). Include comments to explain what the code does. You can also use pseudocode to write the specification, or a diagram to illustrate the data flow between input, map, reduce, and output blocks. **Note:** you can assume that there is a method called **`computeMax(List<Pair> values)`**, which returns the most frequently seen event in a list. This should be used in the **`reduce`** method.

[12 marks]

	Do not write in this column

		12 marks

b) Suppose the previous job is executed on a dataset of **1,000,000** log entries, and the event booking website has **500,000** unique users and **5,000** different music events. The job is executed in a datacentre where Hadoop uses **10** Mappers and **2** Reducers to complete the computation. Based on this information, how many intermediate key:value pairs are emitted by each Mapper? How many unique keys are fed to each Reducer?

[5 marks]

	Do not write in this column

		5 marks

c) This question is about **resilience** in distributed processing systems. Answer the following questions: **[8 marks]**

i) What is **"five nines availability"** in the context of a distributed system? Explain how this relates to the concept of a "single point of failure", and how this might negatively affect a distributed system.

(3 marks)

ii) How does the Hadoop Distributed File System (HDFS) detect when data blocks have been corrupted?

(1 mark)

iii) Will a Map/Reduce job complete if one of its Map tasks fails? How does the Application Master and Node Manager detect and react to the failure of a Map task?

(4 marks)

	Do not write in this column

	8 marks

Question marking: $\frac{}{12} + \frac{}{5} + \frac{}{8} = \frac{}{25}$

Question 4

a) This question is about **distributed graph processing.** Answer the following questions:

[9 marks]

i) **Explain** the way that **Pregel** uses the "*think like a vertex*" model when it parallelises graph computations. Give an **example** of a graph algorithm that is suitable for this model.

(4 marks)

ii) What is **graph partitioning**? Why is it necessary? Discuss the role of graph partitioning in distributed graph processing systems.

(3 marks)

iii) What is the relationships between graph partitioning and performance? Would a *bad* partitioning decision results in worse performance? If so, **why**?

(2 marks)

	Do not write in this column

		9 marks

b) This question is about **Casandra**, which is a NoSQL database. Answer the following questions:

[6 marks]

i) Explain what the **replication factor** refers to on a Cassandra ring. What impact does this have on the resilience of Cassandra's data storage?

(2 Marks)

ii) Imagine you manage a Cassandra database and you face scalability problems, i.e. the current set of Cassandra nodes are insufficient to handle your application's demand. How can you increase the capacity of your Cassandra database? Explain this is terms of **elasticity**, and the impact this will have on performance.

(2 Marks)

iii) Does Cassandra have a single point of failure? Explain your answer.

(2 Marks)

`

	Do not write in this column

		6 marks

c) This question is about **distributed cloud databases.** Answer the following questions:

[10 marks]

i) Explain the difference between **NoSQL** databases and traditional **relational** databases. Explain this within the context of **ACID** transactional properties. What are the benefits of using a **NoSQL** database?

(2 marks)

ii) Name two reasons for using a NoSQL database (e.g. Cassandra) instead of using a traditional SQL relational database.

(2 marks)

iii) What is **Brewer's CAP Theorem**? Explain the three properties of CAP.

(4 marks)

iv) Do **NoSQL** databases fulfil all three properties of CAP? If not, **explain** why not, and what the benefits are of relaxing these constraints.

(2 marks)

	Do not write in this column
	10 marks

Question marking: $\frac{}{9}+\frac{}{6}+\frac{}{10}=\frac{}{25}$

Do not write in this column

2017-2018

Rough Working

Page 19 of 21

Do not write in this column

2017-2018
Rough Working
Page 20 of 21

Do not write in this column

2017-2018
Rough Working
Page 21 of 21