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 Electrical Engineering & Computer Science

Semester Two Examinations, 2023

INFS3208 Cloud Computing

This paper is for St Lucia Campus students.

Examination Duration: 120 minutes

Planning Time: 10 minutes

Exam Conditions:

•This is a Closed Book examination - no written materials permitted
•Any calculator permitted - unrestricted
•During Planning Time - Students are encouraged to review and plan responses to the exam questions
•This examination paper will be released to the Library

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

None

Materials to be supplied to Students:
Additional exam materials (e.g. answer booklets, rough paper) will be provided upon request.

1 x Gradescope Bubble Sheet

Instructions to Students:
If you believe there is missing or incorrect information impacting your ability to answer any question, please state this when writing your answer.

Please answer all the MCQ questions (Part A) on the Gradescope Bubble Sheet and answer all the short-answer questions (Part B) on the examination paper. Total Marks: 50

For Examiner Use Only

Question	Mark

Total _____

Part A – Multiple choice questions (15 marks, 0.5 marks per question)
Answer all questions on the Gradescope Bubble sheet.

1. What are the business drivers for Cloud Computing?

   a. Capacity planning

   b. Cost reduction

   c. Organisational agility

   d. All of the above

2. Which of the following are the pre-existing technologies that heavily impacted Cloud Computing?

   a. Clustering, Big Data, Artificial Intelligence

   b. Database, Grid Computing, Docker

   c. Clustering, Grid Computing, Virtualisation

   d. Artificial Intelligence, Big Data, 5G technology

3. What is the difference between the Cloud and the Internet?

   a. Cloud has a clear and finite boundary.

   b. Cloud is usually private and offers metered IT resources, while the Internet provides open access to many Web-based IT resources.

   c. Cloud often provides back-end processing capabilities and user-based access to these capabilities, while the Internet is dedicated to the access of content-based IT resources published via the World Wide Web.

   d. All of the above.

4. What computing resources do cloud users need to manage in Platform-as-a-service?

   a. Networking and storage

   b. Operation system

   c. Application and data

   d. None of the above

5. What are the cloud delivery models for Virtual Machine and Google App Engine, respectively?

   a. SaaS and PaaS

   b. IaaS and PaaS

   c. IaaS and SaaS

   d. None of the above

6. Which one of the following is FALSE about load balancing?

   a. Load balancing improves the distribution of workloads across multiple computing resources.

   b. Load balancing aims to optimize resource use, maximize throughput, minimize response time, and avoid overload of any single resource.

   c. Using multiple components with load balancing instead of a single component may increase reliability and availability through redundancy.

   d. None of the above.

7. Given two servers A and B, which of the following load balancing algorithms should be considered if A has more CPU cores than B?

   a. Round Robin

   b. Weighted Round Robin

   c. Least Connections

   d. Random

8. What are the features of docker?

   a. Easy modelling

   b. Version Control

   c. Application Isolation

   d. All of the above

9. Which of the following statements about the container is NOT correct?

   a. An image cannot be changed after it was created.

   b. The container can share the kernel of the host operation system but can provide user isolation.

   c. From one image you can create multiple containers.

   d. When constructing an image, each layer in the image can be changed after it has been constructed.

10. Which of the following tools can help you install Docker Engine on multiple virtual hosts?

    a. Dockerfile

    b. Docker-compose

    c. Docker machine

    d. Docker-swarm and Kubernetes

11. How to run docker-compose in detached mode?

    a. Use --daemon flag with docker-compose up command

    b. Use docker-compose run --background command

    c. Use -d option with docker-compose up command

    d. None of the above

12. Which of the following statements about dockerfile is correct?

    a. Dockerfile is a text file that defines the environment inside the container.

    b. Dockerfile is a collection of instructions and commands.

    c. Dockerfile is transparent to users as it tells what is contained in the image.

    d. All of the above.

13. Which of the following tools can help you orchestrate multiple containers on many virtual machines in the cloud?

    a. Dockerfile

    b. Docker-compose

    c. Docker machine

    d. Docker-swarm and Kubernetes

14. Deployments use _____ to provide self-healing and scaling.

    a. ReplicaSets

    b. Pods

    c. Services

    d. Controller

15. What is the Kubernetes default Service type?

    a. NodeIP

    b. PortIP

    c. PodIP

    d. ClusterIP

16. The command to imperatively create a Kubernetes Service is:

    a. kubectl apply

    b. kubectl set

    c. kubectl expose

    d. kubectl deploy

17. Which of the following statements is WRONG?

    a. NoSQL databases require that schemas be defined before you can add data.

    b. NoSQL databases are built to allow the insertion of data without a predefined schema.

    c. Transaction properties (ACID) in traditional relational databases are not suitable to be a set of properties to describe NoSQL databases.

    d. All of the above.

18. What kind of database system is MongoDB?

    a. Document store

    b. Distributed database

    c. SQL relational database

    d. Both A and B

19. According to the CAP (Consistency, Availability, and Partition Tolerance) Theorem, which one of the following options is NOT possible to be satisfied for a NoSQL database?

    a. CA

    b. AP

    c. CP

    d. CAP

20. In GFS (Google File System) files are divided into _____ chunks.

    a. Variable size

    b. Fixed size

    c. Both fixed size and variable size

    d. None of the above

21. Apache Spark has language APIs in _____

    a. Scala

    b. Java

    c. Python

    d. All of the above

22. Which Cluster Manager does Spark support?

    a. Standalone Cluster Manager

    b. MESOS

    c. YARN

    d. All of the above

23. Which of the following is the reason for Spark being faster than MapReduce?

    a. DAG execution engine and in-memory computation

    b. Support for different language APIs like Scala, Java, Python and R

    c. RDDs are immutable and fault-tolerant

    d. None of the above

24. What are the features of Spark RDD?

    a. In-memory computation

    b. Lazy evaluations

    c. Fault Tolerance

    d. All of the above

25. Which of the following is true for Spark SQL?

    a. It is the kernel of Spark.

    b. Provides an execution platform for all the Spark applications

    c. It enables users to run SQL / HQL queries on the top of Spark.

    d. Enables powerful interactive and data analytical applications across live-streaming data

26. Which of the following operations is NOT an action operation in Spark?

    a. collect

    b. take(n)

    c. top

    d. reduceByKey

27. Fault Tolerance in RDD is achieved using:

    a. Immutable nature of RDD

    b. Lineage graph

    c. Lazy-evaluation

    d. None of the above

28. Caching is an optimizing technique that saves interim partial results so they can be reused in subsequent stages.

    a. True

    b. False

29. Which one of the following is FALSE about Hadoop MapReduce?

    a. It is a distributed framework.

    b. The main algorithm used in it is MapReduce.

    c. It runs with commodity hardware.

    d. Hadoop MapReduce cannot provide fault-tolerant features.

30. Which of the following security threats can be countered by an encryption mechanism?

    a. Traffic eavesdropping

    b. Malicious intermediary

    c. Insufficient authorisation

    d. All of the above

Part B – Short answer questions (35 marks)
There are 10 questions in this part. Please answer all questions in the spaces provided on this examination paper.

Question 1: (**4 marks**) What are the cloud delivery models and their corresponding characteristics?

Question 2: (**4 marks**) Describe the differences in working mechanisms and performance between Layer 4 and Layer 7 load balancers. Provide a usage case for each.

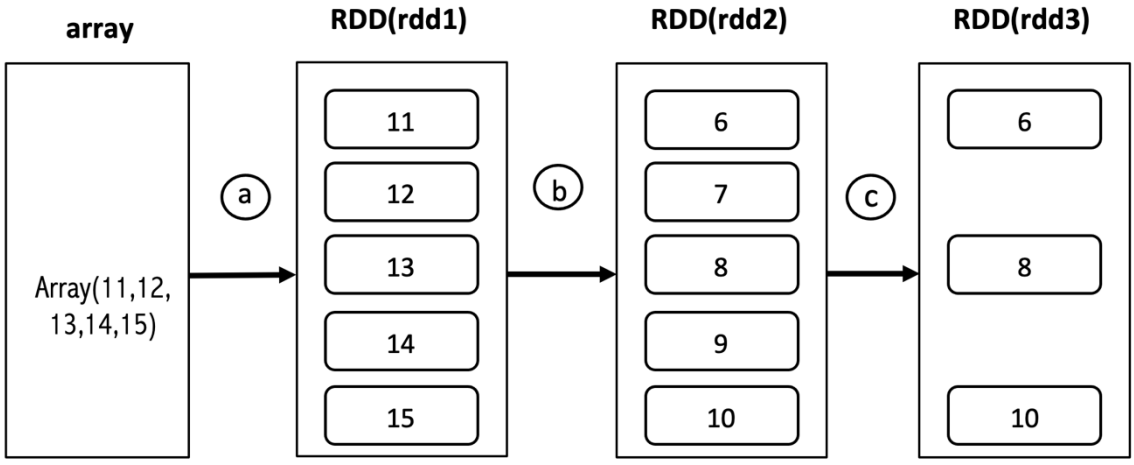
Question 3: (**3 marks**) What are Docker Image, Docker Container, and Docker Registries?

Question 4: (**3 marks**) Describe two popular partitioning methods used in distributed Database Management Systems (DBMS) and explain their main characteristics.

Question 5: (**4 marks**) Please explain the CAP theorem and describe how you choose the right distributed database for a "Likes" counter in a social media scenario based on the CAP theorem.

Question 6: (**3 marks**) Please briefly describe how a digital signature works when sending a message from a sender to a receiver.

Question 7: (**3 marks**) Please use Scala to fill in the appropriate RDD transformation operations in steps a, b, and c.



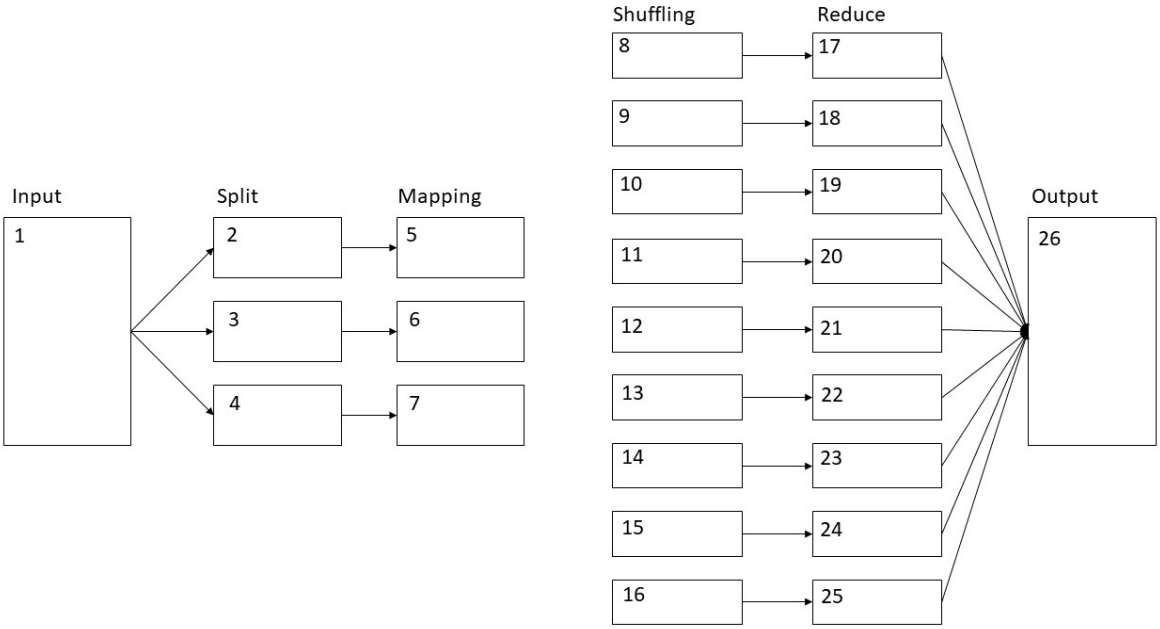
a. _____

b. _____

c. _____

Question 8: (**5 marks**) Given the following text input, please use the below Map-Reduce framework to perform word count (case insensitive). Note that a combiner should be in use if necessary. Please write down the answers according to the box numbers below.

Text Input:
Cloud computing is popular.
Google Cloud Platform is a popular cloud platform.
MapReduce model is a popular computing model.



Answer Sample:

Input
1: Cloud Computing is popular. Google Cloud Platform is a popular cloud platform. MapReduce model is a popular computing model.

Please write the answers in the below blanks:

Split		Mapping	
2		5	
3		6	
4		7	

Shuffling		Reduce	
8		17	
9		18	
10		19	
11		20	
12		21	
13		22	
14		23	
15		24	
16		25	

Output	
26	

Question 9: (**2 marks**) Please briefly describe what etcd in Kubernetes is.

Question 10: (**4 marks**) Please write down each object's name in Kubernetes given its function in the table below.

Objects	Functions
Answer Example: 1. Pods	the smallest deployable units of computing that you can create and manage in Kubernetes. It has a group of one or more containers, with shared storage and network resources.
2. _____	contains a small amount of sensitive data such as a password, a token, or a key.
3. _____	maintains a sticky identity for each of its Pods. These pods are created from the same spec, but each pod has a persistent identifier that it maintains across any rescheduling.
4. _____	manages external access to the services in a cluster and may provide load balancing, SSL termination etc.
5. _____	stores non-confidential data in key-value pairs.
6. _____	maintains a stable set of replica Pods running at any given time and thus provides scalability and self-healing
7. _____	provides declarative updates for Pods and ReplicaSets and supports rolling updates and rollback features.
8. _____	exposes a network application that is running as one or more Pods in your cluster.
9. _____	attaches a physical storage facility to persist the data.

END OF EXAMINATION