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, 2024
INFS3208 Cloud Computing

This paper is for St Lucia Campus students.

Examination Duration: 120 minutes

Planning Time: 10 minutes

Exam Conditions:

•No written or printed material permitted
•Casio FX82 series or UQ approved and labelled calculator only
•During Planning Time - Students are encouraged to review and plan responses to the exam questions

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. Which of the following is NOT the business driver of Cloud Computing?

   a. Ease of use

   b. Capacity planning

   c. Cost reduction

   d. Organisational agility

2. Which of the following are the pre-existing technologies that heavily impacted on 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 are the differences between Clustering and Grid Computing?

   a. Nodes in Grid Computing must have the same type of hardware and operating system, while nodes in Clustering could have different types.

   b. Cluster nodes are often located close to each other, while nodes in Grid may be located at a huge distance from each other.

   c. Cluster nodes are connected through a relatively low-speed network (e.g. Internet), while Grid nodes are usually internally connected via a high-speed network (e.g. fibre cables, routers, and switches).

   d. Both a and b are correct.

4. 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

5. What is a public cloud?

    a. A publicly accessible and metered cloud environment owned by a third-party cloud provider.

    b. A computing service that can only be accessed from a publicly shared computer.

    c. A publicly accessible computing environment free of charge.

    d. The cloud infrastructure that is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units)

6. Which of the following statements about regions and zones in cloud computing are correct?

    a. There are many regions and zones for both GCP and AWS.

    b. In each region, there will be a number of zones. In each zone, there will be one or more discrete data centres with redundant power, networking, and connectivity.

    c. Normally, zones in a region are interconnected with high-bandwidth, low-latency networking, over fully redundant, dedicated metro fibre providing high-throughput, low-latency networking between AZs.

    d. All 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. 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.

9. Which of the following docker commands is to delete an existing image?

    a. docker rmi

    b. docker kill

    c. docker rm

    d. docker pause

10. Which of the following tools can help you manage multiple containers on one single machine?

    a. Dockerfile

    b. Docker-compose

    c. Docker machine

    d. None of the above

11. Decoupling MySQL, Redis, Nginx, and PHP services into multiple containers makes it easier to scale horizontally and reuse the containers.

    a. True

    b. False

12. Which of the following commands can be used to scale down a service named as "frontend" running with 6 containers to 3 containers?

    a. docker service frontend scale down 3

    b. docker services frontend scale=3

    c. docker service scale frontend=3

    d. None of the above

13. Which of the following commands will POST pod.yml to the API server?

    a. kubectl apply -f pod.yml.

    b. kubectl post -f pod.yml

    c. kubectl app -f pod.yml

    d. kubectl manifest -f pod.yml

14. What is the Kubernetes default Service type?

    a. NodeIP

    b. PortIP

    c. PodIP

    d. ClusterIP

15. Which of the following is a NoSQL database type?

    a. SQL

    b. Document databases

    c. JSON

    d. All of the above

16. 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 the traditional relational databases are not suitable to be a set of properties to describe distributed NoSQL databases.

    d. All of the above.

17. Which of the following is a wide-column store?

    a. Cassandra

    b. Redis

    c. MongoDB

    d. MySQL

18. Which of the following is a key characteristic of a vector database?

    a. Support for SQL queries

    b. Handling ACID transactions

    c. Optimized for similarity searches in high-dimensional spaces

    d. Designed for data warehousing

19. Which of the following Faiss Index types is most suitable for brute-force search in a small dataset?

    a. IndexIVFFlat

    b. IndexFlatL2

    c. IndexIVFPQ

    d. IndexHNSW

20. Which group of the following operations are supported by GFS architecture?

    a. Read, write, create, delete

    b. Update, append, read, write

    c. Snapshot, append, update, delete

    d. None of the above

21. Which of the following statements about HDFS Version 2 is correct?

    a. HDFS needs to run on expensive commodity hardware and cannot deal with failures.

    b. HDFS is used to handle large files and block size is 128 MB by default.

    c. HDFS supports multi-users to write one file simultaneously.

    d. In HDFS, low latency is more important than high sustained bandwidth.

22. Apache Spark has language API's in _____

    a. Scala

    b. Java

    c. Python

    d. All of the above

23. Which of the following is NOT a feature of Spark?

    a. Supports in-memory computation

    b. Fault-tolerance

    c. It does not support interactive processes.

    d. Compatible with other file storage systems

24. Which of the following is true for RDD?

    a. RDD is a programming paradigm.

    b. RDD in Apache Spark is an immutable collection of objects.

    c. It is a database.

    d. None of the above.

25. Which of the following operation is NOT a transformation in Spark?

    a. flatmap

    b. filter

    c. reduce

    d. reduceByKey

26. Fault Tolerance in RDD is achieved using:

    a. Immutable nature of RDD

    b. Lineage graph

    c. Lazy evaluation

    d. None of the above

27. The shortcomings of Hadoop MapReduce are overcome by Spark RDD by:

    a. Lazy evaluation

    b. DAG

    c. In-memory processing

    d. All of the above

28. Which of the following is the fundamental data structure of Spark?

    a. RDD

    b. DataFrame

    c. Dataset

    d. None of the above

29. What are the advantages of 3x replication schema in Hadoop?

    a. Fault tolerance

    b. High availability

    c. Reliability

    d. All of the above

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: (**3 marks**) What are the characteristics of cloud computing?

Question 2: (**3 marks**) What are Docker Image, Docker Container, and Docker Registries?

Question 3: (**4 marks**) Please briefly compare the differences between RDD and DataFrame.

Question 4: (**3 marks**) What are the differences between SQL and NoSQL?

Question 5: (**2 marks**) Please briefly describe each step in the READ operation in Google File System (GFS).

Question 6: (**4 marks**) Please explain the CAP theorem. Based on CAP, how would you choose the right distributed database for a flight booking system, and why?

Question 7: (**4 marks**) Please answer the below questions about vector databases.

1) Briefly describe the **Inverted File Product Quantization Index** used in vector databases (1 mark).

2) Fill in the missing components in the steps of the **Inverted File Product Quantization (IVF-PQ) Index** training process (3 marks):

Training Process:

1. An ____a.____ is constructed by dividing the set of vectors into k Voronoi partitions.
2. Inside each k Voronoi partition, each vector is subtracted by the coordinates of its ____b.____ .
3. The ____c.____ algorithm is run on vectors from all the partitions:
   o Splitting high-dimensional vectors into smaller ____d.____ .
   o Quantizing these ____d.____ into a finite number of ____e.____ , each represented by ____f.____ .
   o Using a codebook for each sub-vector to enable efficient ____g.____ and approximate distance calculation.
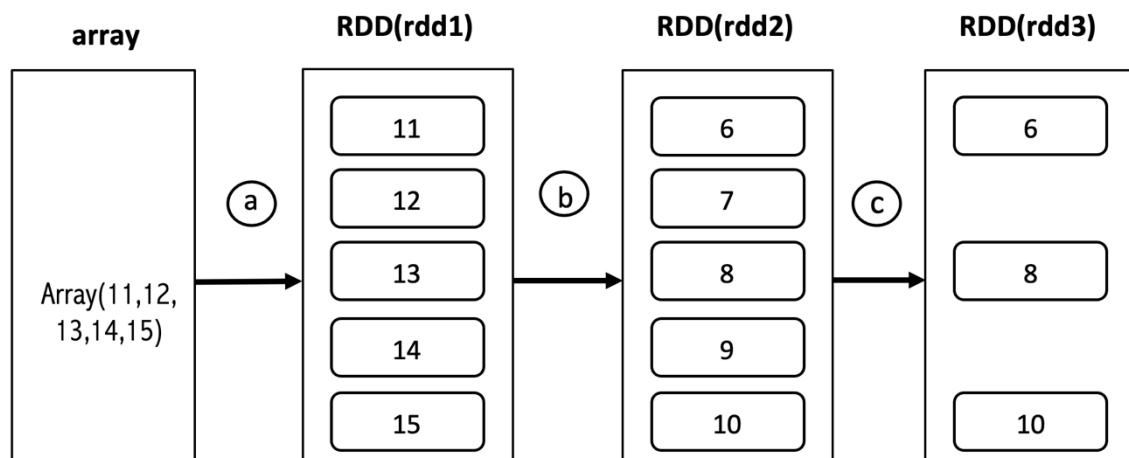
Answer Example:

a. _inverted file index_

b._____   c._____   d. _____

e._____   f._____   g. _____

Question 8: (**3 marks**) Please briefly describe how a digital signature works when sending a message from a sender to a receiver.

Question 9: (**3 marks**) Please use Scala to fill in the appropriate RDD transformation operations in steps a, b, and c.

array	RDD(rdd1)	RDD(rdd2)	RDD(rdd3)
Array(11,12, 13,14,15)	11, 12, 13, 14, 15	6, 7, 8, 9, 10	6, 8, 10

a ⟶ (between array and rdd1)
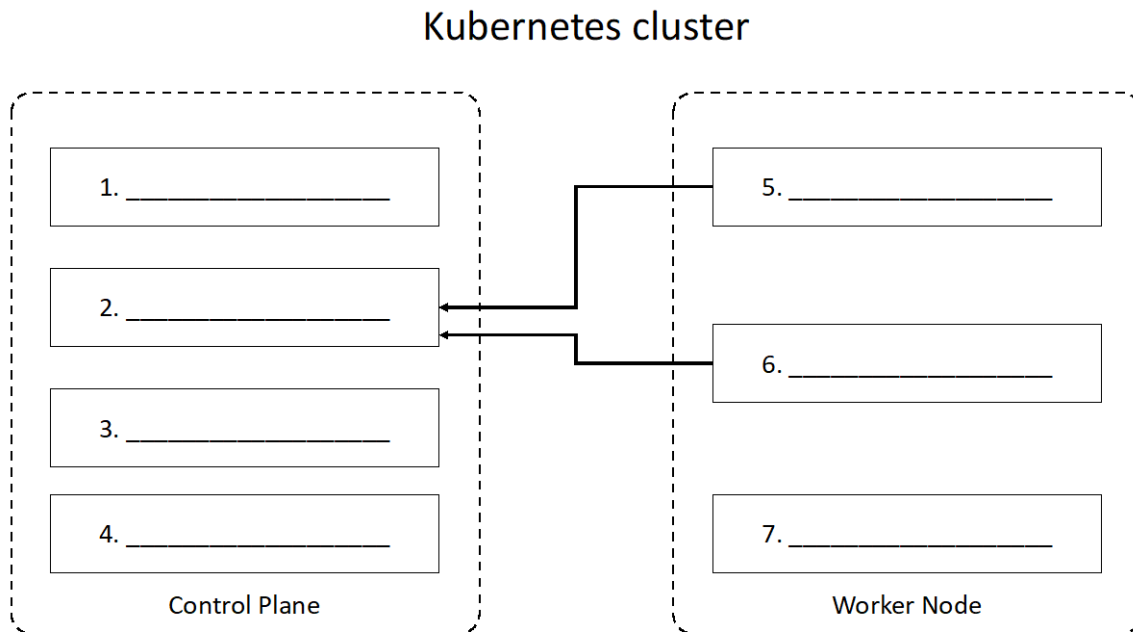b ⟶ (between rdd1 and rdd2)
c ⟶ (between rdd2 and rdd3)

a. _____

b. _____

c. _____

Question 10: (**6 marks**) Given the following conceptual framework of Kubernetes that contains a Control Plane node and a Worker node, please write down each component's name and its function label below. The function labels are listed in the table below.

Kubernetes cluster


```
Control Plane                                    Worker Node

1. _____          5. _____

2. _____          6. _____

3. _____          7. _____

4. _____
```

Label	Function
(A)	makes sure that containers are running in a Pod.
(B)	watches for newly created Pods with no assigned node, and selects a node for them to run on.
(C)	validates and configures data for the API objects which include pods, services, replication controllers, and others.
(D)	is the software that is responsible for running containers.
(E)	is a daemon that embeds the core control loops shipped with Kubernetes.
(F)	provides a consistent and highly available key-value store for all the cluster data.
(G)	maintains network rules on nodes.

Answer Sample:

Component Name	Function Label
1. *Controller Manager*	8. *E*

Write down the component name and the respective function label below:

Component Name	Function Label
2.	9.
3.	10.
4.	11.
5.	12.
6.	13.
7.	14.

END OF EXAMINATION