CMI/2022/Final-Semester Exam May 2022

Chennai Mathematical Institute

Distributed Computing and Big Data

DURATION: 2 HOURS MAX MARKS: 35.

ROLL NO.: _____ DATE: 12/05/2022

NAME: _____

Instructions

- Please remember to mention your name and roll number in your answer sheet.
- You are allowed to bring one handwritten A4 sheet with notes. You will not be allowed to borrow or lend notes from or to other students.
- No electronic devices (calculators, laptops, etc) are allowed in the exam hall.
- No negative marks.

Section 1: All questions carry one mark each.

Question 1. Which of the following depicts the correct order by the factor of increasing size? Choose the best answer.

(1) petabytes, exabytes, zettabytes
(2) exabytes, petabytes, zettabytes
(3) exabytes, zettabytes, petabytes
(4) terabytes, exabytes, petabytes

Question 2. To store an object in Amazon S3, you create a _____ in one of the AWS regions and then upload the object to it.

Question 3. Which CAP category does Apache Cassandra belong to?

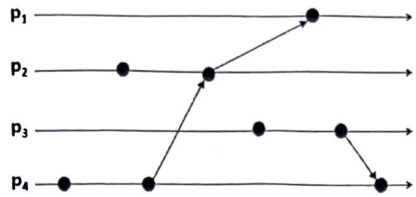
Question 4. Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. True/False?

Question 5. Azure regions are physically separate locations within each Azure availability zones that are tolerant to local failures. True/False?

Section 2: All questions carry three marks each.

Question 6. If only 80% computation can be executed in parallel, and if we have 8 processors, what is the best speed up achievable as per Amdahl's law?

Question 7. Annotate the following space-time diagram with vector time.



Question 8. CMI was excited to go digital by streaming its lectures on its own platform. Three other universities decided to subscribe to these lectures. Draw a kafka-based publish subscribe architecture to implement this use-case. Take any example within the scope of the given use-case to explain *topics* and *partitions*. You are not expected to write any code.

Question 9. What will be the output of the following pig script if the input contains a single line of $(1,3,5)(2,4,6)$?

```
A = LOAD 'data' AS (
 t1:tuple(t1a:int, t1b:int,t1c:int),
 t2:tuple(t2a:int,t2b:int,t2c:int)
 );
X = FOREACH A GENERATE t1.t1a,t2.$1;
DUMP X;
```

Question 10. Assume a disk size of 4 Terabyte with block size of 4 KB. How much space will you need to store the free space bitmap?

Section 3: All questions carry 5 marks each.

Question 11. Thomas Cook is a travel management consultancy. In the post-covid era, they want to automate their travel booking systems. Design a RESTful web service for the travel booking scenario. You may scope your answer to two resources. Your answer must cover at least one idempotent method assignment and one non-idempotent method assignment.

Question 12. Describe a map-reduce design pattern to join two tables from an RDBMS.

Question 13. Provide distributed algorithms for the following:

(1) To elect three processes as coordinators from a group of n independent processes. (2 marks)

(2) To elect m processes as coordinators from a group of n independent processes where $m < n$. (3 marks)