

## DISTRIBUTED COMPUTING AND BIG DATA

Chennai Mathematical Institute

DURATION: 120 MINS.

MAX: 25 MARKS.

---

**Instructions**

- This is a closed book exam.
  - This is an individual task. Do not discuss with anyone.
  - No electronic devices are allowed. Wherever heavy calculation is involved, you need not evaluate it to the final number unless it is explicitly asked for. For example, it is acceptable to leave the answer as  $\frac{1}{1+\frac{5}{32}}$ . You need not evaluate it to 0.865.
  - Clearly mention your name and roll number in your answer sheet.
- 

**Section 1: Correct answers carry 1 mark each. Answer True/False.  
Wrong answers carry -0.5 marks each.**

**Question 1.** Schema-on-write is an approach that is commonly used in data warehouses whereas Schema-on-read is commonly used in data lakes. True/False?

**Question 2.** Amdahl's law states that the number of transistors on a microchip doubles every two years, though the cost of computers is halved. True/False?

**Question 3.** CUDA enables developers to use GPUs for compute-intensive applications. True/False?

**Question 4.** YARN is Hadoop's framework for job scheduling and cluster resource management. True/False?

**Question 5.** Scalar time is not strongly consistent. True/False?

**Question 6.** Docker helps in application isolation. True/False?

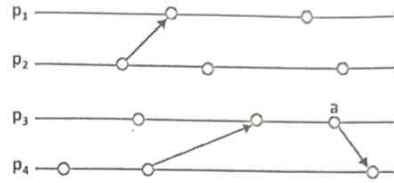
**Question 7.** In the distributed computing model, causal dependencies between events are transitive in nature. True/False?

**Section 2: Correct answers carry 2 marks each. No negative marks.**

**Question 8.** Assume disk size = 1 TB and block size = 32 KB. How much space (in MB) will we need to store the free space bitmap?

**Question 9.** As per Amdahl's law, What is the best achievable speed up if only 20% of the job can be parallelized, and we have 8 processors?

**Question 10.** If we were to annotate the following space-time execution diagram with scalar time stamp, how would we annotate the event marked as 'a'?



**Question 11.** For the same diagram given above, annotate the events in  $p_3$  with vector time.

**Question 12.** For the same diagram given above, identify a consistent cut not involving the event marked as 'a' i.e., 'a' must be in the future of the cut.

**Question 13.** In the muddy children puzzle, as discussed in the class, what would the children say during the first and second rounds if  $n=4$  and  $k=2$ ? i.e., there are four children, and two of them have muddy forehead and to start with, they are told that at least one of them have muddy forehead. Assume that each child can only say 'yes', 'no' or 'dont know'. Use the following format to provide your answer.

Round1: <1st child response>, <2nd ...>, <3rd ...>, <4th ...>

Round2: <1st child response>, <2nd ...>, <3rd ...>, <4th ...>

**Section 3: Question carries 3 marks. No negative marks.**

**Question 14.** Draw a space-time diagram which has an event 'a' annotated in matrix time as  $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 1 & 2 & 2 \end{bmatrix}$ .

**Question 15.** You are given a large text file. You need to find all the distinct words that are longer than 5 characters. Describe an approach using map reduce logic. No need to write any code.

$$\begin{aligned}
 1TB &= 2^{10} GB \\
 &= 2^{10} \cdot 2^{10} MB \\
 &= 2^{10} \cdot 2^{10} \cdot 2^{10} KB \\
 &= 2^{10} \cdot 2^{10} \cdot 2^{10} \cdot 2^{10} B
 \end{aligned}$$

