

**Instructions**

- Please remember to mention your name and roll number in your answer sheet.
- This is an individual task. Do not discuss with anyone.
- This is a closed book exam. You are not allowed to carry books or cheatsheets.
- No electronic devices (calculators, laptops, etc) are allowed in the exam hall. 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.
- First section has negative marks. No negative marks for the rest of the sections.

**Section 1: All questions carry one mark each. -0.5 for wrong answers.  
Answer in True/False.**

**Question 1.** One Petabyte space is enough to store 4 Million ebooks of size 1 MB each. **T**

**Question 2.** Based on Amdahl's law, we can expect a linear increase in speed-up for a specific job as we increase the number of processors. **F**

**Question 3.** Yet Another Resource Negotiator, is a resource management and job scheduling technology for the Hadoop distributed processing framework **T**

**Question 4.** One of the items in pig philosophy is that pigs live anywhere. **T**

**Question 5.** As per the principles of object-oriented programming, an object has an identity while a class does not. **T**

**Question 6.** An impedance mismatch occurs in relational databases when a relational database needs to be transformed into an object-oriented model. **T**

**Question 7.** BSON is a binary serialization format used to store documents in MongoDB. **T**

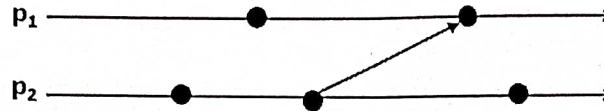
**Question 8.** NoSQL datastores are not ACID compliant. **T**

**Question 9.** Changing a block in a blockchain makes all the following blocks invalid. **T**

**Question 10.** A carrier hotel is a facility strategically based in a location closer to users that houses networks and cloud services. **T**

Section 2: All questions carry 2 marks each.

Consider the following space-time execution diagram while answering the questions in this Section.



- ☒ Question 11. List all the happens-before relationships.
- ☒ Question 12. Annotate the events using scalar time.
- ☒ Question 13. Annotate the events using vector time.
- ☒ Question 14. Annotate the events using matrix time.
- ☒ Question 15. Identify an inconsistent cut.

Section 3: All questions carry 2 marks each.

☒ Question 16. Consider the following Pig script.

```
Lines = LOAD 'file1' USING PigStorage() as (line:chararray);
Words = FOREACH Lines GENERATE FLATTEN(TOKENIZE(line)) AS word;
Groups = GROUP Words BY word;
Counts = FOREACH Groups GENERATE group, COUNT(Words) as Cnt;
Results = ORDER Counts BY Cnt ASC;
Dump Results;
```

Assume that the input file 'file1' contains the following two lines:

~~cmi is the best~~  
~~the best in chennai is cmi~~

What does the pig script output?

☒ Question 17. The following pig script was written to find the most expensive iphone. However, it has errors. Identify the errors and correct them.

Pig Script:

```
A = LOAD 'file2' USING PigStorage(',') AS (year:int,product:chararray,cost:int);
B = GROUP A BY $2;
C = FOREACH B GENERATE MIN(A.cost);
DUMP C;
```

CREATE  
GET  
PUT  
DELETE

cmi  
cmi  
is  
is  
the  
the  
best  
best  
in  
in  
chenn  
chenn



Input File ('file2' contains year,product,cost):

2022, iphone, 50000

2023, iphone, 65000

2024, iphone, 72000

Expected output is 72000.

~~Question 18.~~ How many nodes are created when the following three statements are executed by Neo4j?

1. CREATE (p:Person{name:'Venkatesh'})-[:Teaches]->(c:Course{name:'BigData'})
2. CREATE (p:Person {name:'Raj'})-[:StudentOf]->(o:Org{name:'CMI'})
3. MATCH (a:Person),(b:Org) WHERE a.name = 'Venkatesh' AND b.name = 'CMI'  
CREATE (a)-[:FacultyAt]->(b)

Section 4: All questions carry 3 marks each.

~~Question 19.~~ Describe a map-reduce design for computing median of a large list of numbers. Assume that the input file contains 2 Million lines. Each line contains an integer ranging between 1 and 1000.

**Question 20.** Assume that Indian Railways wants to store the train running schedule (past and live status) information in MongoDB. Provide a database design along with at least one or two queries as example to indicate how you would query the data.

~~Question 21.~~ Design a RESTful web service for a learning management system such as moodle. Include at least three items in your object model.

