CMI/2022/Mid-Semester Exam Mar 2022

Chennai Mathematical Institute

Distributed Computing and Big Data for Uploading. Max Marks: 20. Duration: 60 Mins + 30 Mins

---

Roll No.: _____ Date: 17/03/2022

Name: _____

Instructions

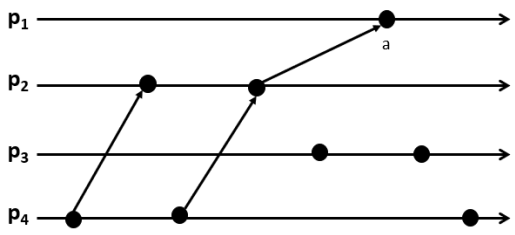
- Submit a single pdf file carrying your answers on moodle under "Mid Term" assignment. For any reason, if you cannot upload to moodle, email your work to vvtesh.cmi@gmail.com.
- A penalty of 1 mark applies for every two minutes of late submission.
- This is an individual task. Do not discuss with anyone.
- Please stop writing after 60 minutes. Uploading may take time. We apply late penalty strictly. If you make several submissions, the last submission will be taken for grading.

Section 1: Questions carry 3 marks each.

Question 1. Ramesh bought a hard disk with rotational delay of 3ms. With what RPM does the disk spin? If it had 20 sectors per track, what is its read time?

Question 2. Ram bought a new hard disk and configured it so that the number of bytes per inode is r. Prem too bought a new hard disk of same size and configured it to have the number of bytes per inode as p. Ram and Prem stored f number of files each in their respective disks. Given that $r > p$ and Ram had relatively smaller sized files, is it possible that Ram ran out of disk space while Prem did not? Explain with an example.

Question 3. If we were to annotate the following space-time execution diagram with matrix time stamps, how would we annotate the event marked as 'a'?

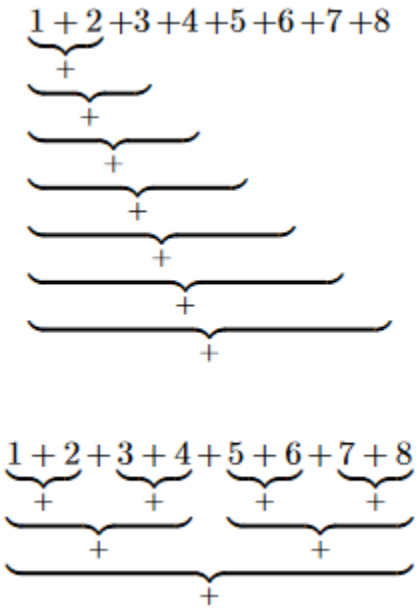


Question 4. In the muddy children puzzle, as discussed in the class, what would the children say during the first two rounds if n=4 and k=3 i.e., there are four children and they are told that at least three of them have muddy forehead? Assume that each child can only say 'yes', 'no' or 'dont know'. Use the following format to provide your answer.

```
Round1: <1st child response>,<2nd ...>,<3rd ...>,<4th ...>
Round2: <1st child response>,<2nd ...>,<3rd ...>,<4th ...>
```

Section 2: Questions carries 4 marks each.

Question 5. Our ability to write parallelizable programs decides the speed up we can achieve through scaling. Consider two programs to add numbers. The first program adds the first two numbers, remembers the result, and adds that result to the next number. It continues doing this until the end of the list is reached. The second program adds two numbers at a time in parallel. It recursively does so until the final results are arrived at. The following figure explains their logic.



As per Gustafson's law, assuming each addition is an operation, how much speedup (approximately) can these programs achieve if there are four processors?

Question 6. You are provided with a large text file containing the names of millions of chess players and their world rank. The file format is as shown below:

```
Viswanathan Anand, 15
Magnus Carlsen, 1
Venkatesh Vinayakarao, 1029388
...
```

Provide the design of a map-reduce job to pick top 10 players from this file. You do not need to write code. Explain clearly, the logic behind the mappers and reducers.
