CMI/2021/Final-Semester Exam July 2021

Chennai Mathematical Institute

Distributed Computing and Big Data

Duration: $2\frac{1}{2}$ hours + 30 mins for uploading. Max Marks: 35.

Roll No.: _____ Date: 07/07/2020

Name: _____

Instructions

- Submit a single pdf file carrying your answers on moodle under "Final" assignment. For any reason, if you cannot upload to moodle, email your work to vvtesh.cmi@gmail.com.
- This is an individual assessment. Do not discuss with anyone.
- Please stop writing after $2\frac{1}{2}$ hours.
- Please remember to mention your name and roll number in your answer sheet.
- Late submissions will attract penalty.

Section 1: All questions carry 5 marks each.

Consider the following version of the problem of muddy children puzzle.

Suppose that there are a total of N children, where M ∈ {1, ... , N} of them has mud on their forehead. Each child can observe whether another child (but not himself or herself) has mud on their forehead. The muddy children protocol goes in rounds. Before the first round, the father declares that there is at least a muddy child (i.e. with mud on their forehead). In each round, he asks the children whether they know if they are muddy, to which the children can answer yes or no. The children are perceptive, intelligent, and truthful.

Question 1. Answer the following:

(1) If $N = 4$ and $M = 2$, enumerate what each child would have mentioned in each round till all the responses converge. For each round, clearly explain the reasoning behind every child's answer. (2.5 Marks)

(2) If $N = 4$, $M > 0$, and M is even, enumerate what each child would have mentioned in each round till all the responses converge. For each round, clearly explain the reasoning behind every child's answer. (2.5 Marks)

Section 2: All questions carry 5 marks each.

The Ministry of Health and Family Welfare (MoHFW) needs to maintain the COVID vaccination status. Therefore, it requires the hospitals and other camps where vaccines are administered, to collect and upload the data related to the person who is getting vaccinated.

Question 2. Design a RESTful web service for maintaining COVID vaccination records for the Ministry of Health. 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 3. Explain a strategy for hospitals to securely transfer the vaccination data through web services using a sequence diagram.

Question 4. Developers nowadays commonly make use of Object-Relation Mapping (ORM) frameworks to provide a conceptual abstraction between objects in Object-Oriented Languages and data records in the underlying database. Provide a partial class-diagram with at least two classes for the above mentioned MoHFW scenario. Explain how ORM can be used in this scenario.

Section 3: All questions carry 5 marks each.

Question 5. An N x N adjacency matrix describes if a person is a friend of another person. You may visualize this as a directed graph where nodes represent users and edges represent friendship. Note that X is a friend of Y does not mandate that Y needs to be a friend of X. We want to associate an influencer score for each user. The score is any integer directly proportional to the number of users that they are connected to, either directly or transitively. Assuming that the matrix is very large, explain a *map-reduce design pattern* to accomplish this task.

Question 6. To implement the vector clocks, we need to fix the number of bits for each integer element of the vector. If the bit count is too large, we need more space to store them and if the count is too small, we may end up in a overflow situation. For example, a two bit clock implementation can only describe three events (1, 2 and 3. Recall that 0 is needed as a special case) in a process. Assuming that each process in our distributed systems may have several billions (or even more) events, how can we decide the optimal size for the elements of a vector clock implementation?

Question 7. Provide distributed algorithms for the following:

(1) To elect two 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)