



**CSL4030: Data Engineering (July-December, 2023)**  
**Minor-2 Question Paper, October 16, 2023**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology Jodhpur, Rajasthan, India 342030**

**Duration: 1 hour**

**Full Marks: 32**

\*\*\*\*\*

**Question 1.**

[2+3+1+1+6+2]=15

- i. What are the two major demands of data-intensive applications from distributed databases?
- ii. Is there a condition under which both the demands can be met together?
- iii. How many times a persistent message is guaranteed to be delivered?
- iv. Is there a way to distinguish a link failure from a network partition?
- v. If you become the CTO of Instagram, which algorithm would you recommend for handling network partitions as well as coordinator failures? Please discuss along with a pseudocode.
- vi. Which properties your aforementioned algorithm need to honor?

**Question 2.**

[1+3+2]=6

- i. Is it possible to generate structured data from unstructured data?
- ii. What are the subtypes of the graph-like data model? Please provide examples of one DBMS and one query language for each subtype.
- iii. Can you please explain the in-memory sparse index scheme with an example?

**Question 3.**

[1+1+4+5]=11

- i. Why do not we use regular distributed database protocols for managing network directories?
- ii. "Referrals are the key component that help organize a distributed collection of directories into an integrated system." ~ Korth et al. What is a referral?
- iii. The Board of Control for Cricket in India (BCCI) has released a request for proposal (RFP) for developing an e-commerce platform for Cricket supplies under the 'Make in India' banner. All the BCCI users (i.e. players and administrators across India) will be required to make their official purchases only through this platform. You and your friends want to submit a proposal. For that purpose, you have started working on registering your team as a start-up and prototyping a web app called 'CricGem' (the cricket governance e-commerce marketplace of India). One of the specifications is that the players are frequent travelers whereas the administrators only move when they are transferred from one location to another. What kind of data fragmentation and/or replication strategy would you consider for each type of users?
- iv. Another specification is that different state cricket associations have been managing their purchase records with different database management systems. In that situation, what types of strategies would you consider for detecting local and global deadlocks? (To help you visualize the situation, the following

diagram shows a use case where there is a local deadlock in site S2 as well as a global deadlock between sites S1 and S2. Transaction T1 is initiated at S1 whereas transactions {T2, T3} are initiated at S2. Three data items are represented with Q's. An incoming edge to a data item represents a request to lock the data item. On the other hand, an outgoing edge from a data item represents that the data item has been locked by the corresponding transaction.)

