



**CSL4030: Data Engineering (July-December, 2023)**  
**Minor-1 Question Paper, September 6, 2023**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology Jodhpur, Rajasthan, India 342030**

**Duration: 1 hour**

**Full Marks: 32**

\*\*\*\*\*

**Question 1.**

[10]

Prakriti, an IITJ student, is planning to develop an attendance tracker app for the institute. With some initial thoughts, she has come up with the following relational schema.

Relation 1: STUDENT (RegNo, RollNo, FName, LName, DeptID, DeptName, Active(Y/N))

RollNo is a multivalued attribute. A student can have at most two roll numbers (an old one and a current one in case of a department change) and at least one roll number.

Relation 2: ATTENDANCE (RegNo, CourseID, CourseName, Date, Present(Y/N))

The functional dependencies (FDs) are as follows:

FD1: RegNo -> {RollNo, FName, LName, DeptID, DeptName}

FD2: DeptID -> DeptName

FD3: {RegNo, CourseID, Date} -> {CourseName, Present}

FD4: CourseID -> CourseName

Can you please help Prakriti normalize the relations to the highest normal form possible, showing one normalization step at a time?

**Question 2.**

[10]

Prakriti, after realizing her interest in developing AI apps, has applied for a department change from ME to CSE(AI). As a result, Prakriti's records (RollNo, DeptID, etc.) need to be updated. To that cause, the Office of Academics issues a database update query to the ACAD computer server. However, necessary details must be simultaneously updated at the ME and CSE department servers so that Prakriti could smoothly transition to the new program. In other words, a database transaction with the following steps need to be executed where Prakriti's RegNo is 007.

-----  
lock(007) at ACAD;  
update(007) at ACAD;  
lock(007) at ME;  
update(007) at ME;  
lock(007) at CSE;  
update(007) at CSE;  
unlock(007) at ACAD;  
unlock(007) at ME;  
unlock(007) at CSE;

-----  
Using one column for each server (site), can you please implement a commit protocol that can handle site failures and network partitions?

**Question 3.**

[2 + 2 + 8] = 12

CERN's Large Hadron Collider produces hundreds of petabytes of sensors data that scientists analyze, visualize, and discuss about on their internal forums. These activities are supported by a data management and analytics software system called ROOT. What might be the key responsibilities of the Data Engineers at ROOT? Does this data qualify as a Big Data? If you were to design ROOT, what kind of data model(s) would you implement and why?