Term: Fall 2023   **Subject:** Computer Science & Engineering (CSE)   **Number:** 512

Course Title: Distributed **D**atabase **S**ystems (CSE 512)

GROUP PROJECT REQUIREMENTS

Project Overview:

In this semester-long project, students will collaboratively design and implement a distributed database system, applying various concepts and techniques learned throughout the course. Each part of the project corresponds to different aspects of distributed databases.

Implementation Tools and Languages:

Students have the freedom to choose their preferred implementation tools and languages based on their expertise and project requirements for each part.

PART 1: Design and Implementation of a Distributed Database System [10 Points]

Problem Statement	Design and implement a distributed database system for a topic of your choice, capable of handling real-time data updates and queries while ensuring data consistency and availability.
Topics	Including but not limited to E-commerce, Healthcare, Finance, Inventory and Supply Chain, Social Networking, Energy Management, Education, Transportation and Logistics, Media and Entertainment, Manufacturing, IoT (Internet of Things), Agriculture, Government and Civic Services, Gaming, Human Resources
To-do Tasks	Schema design, Table creation, Data distribution, Data Insertion, Data Retrieval
Deliverables	1. Distributed Database Schema: Documentation with entity-relationship diagrams and table definitions. 2. Database Tables: Code or scripts for the creation of necessary tables with attributes, keys, and constraints. 3. Possible Data Distribution Plan: Documentation of the chosen data distribution strategy. 4. Data Insertion Mechanism: Code or scripts for efficient initial data insertion. 5. Data Retrieval Proof: Evidence (Code, script, snapshots) of successful basic data retrieval queries.
Possible Tools	PostgreSQL (or) - MySQL (or) CockroachDB (**or) any other of your choice.**

PART 2: Fragmentation and Replication Techniques   [20 Points]

Problem Statement	Implement fragmentation and replication techniques for optimizing the performance of the distributed database system in your chosen topic.
Techniques/ Topics	Fragmentation – Horizontal and Vertical, Replication Strategies
To-do Tasks	1. Horizontal Fragmentation: Split tables into subsets based on specific criteria 2. Vertical Fragmentation: Divide tables into smaller subsets based on columns to optimize data retrieval. 3. Replication Setup: Configure replication models such as master-slave or peer-to-peer replication to enhance data availability.
Deliverables	Code/Script, Snapshots, Documentation
Possible Tools	PostgreSQL (or) - MySQL (or) CockroachDB **(or) any other of your choice.**

PART: 3 Query Processing and Optimization Techniques   [20 Points]

Problem Statement	Implement query processing and optimization techniques to enhance the performance of the distributed database system with your chosen topic.
Techniques/ Topics	Query optimization, Distributed indexing
To-do Tasks	1. Query Optimization: Analyze and optimize queries for efficient data retrieval, utilizing features provided by the chosen database system (e.g., PostgreSQL, MySQL). 2. Distributed Indexing: Implement distributed indexing strategies to improve query performance in a distributed environment.
Deliverables	Code/Script, Snapshots, Documentation
Possible Tools	PostgreSQL or MySQL with Query Optimization: Utilize built-in query optimization capabilities in PostgreSQL or MySQL to optimize queries. (or) Apache Calcite: Explore the open-source Calcite framework for building custom query optimization solutions **(or) any other of your choice.**

PART 4: Distributed Transaction Management   **[20 Points]**

Problem Statement	Implement distributed transaction management to ensure data consistency and concurrency control in your chosen topic.
Techniques/ Topics	ACID properties in distributed databases, Concurrency control, and Distributed coordination for transactions.
To-do Tasks	1. Implement ACID-compliant distributed transactions using the chosen database system (e.g., PostgreSQL, CockroachDB). 2. Propose concurrency control mechanisms to handle simultaneous transactions. – Explore any tools for distributed coordination and managing distributed transactions.
Deliverables	Code/Script, Snapshots, Documentation
Possible Tools	Apache Ignite **(or) any other of your choice.**

PART 5: Distributed NoSQL Database Systems Implementation   **[20 Points]**

Problem Statement	Implement a distributed NoSQL database system to understand data storage and retrieval process with your chosen domain.
Techniques/ Topics	NoSQL data modeling
To-do Tasks	1. Choose an appropriate NoSQL database system based on your project's topic and data requirements. Define and document the data schema and data model tailored to your topic's data storage needs. 2. Implement basic CRUD (Create, Read, Update, Delete) operations for the domain-specific data. 3. Create sample queries and data retrieval operations to showcase the functionality of your NoSQL database for your chosen topic.
Deliverables	Code/Script, Snapshots, Documentation
Possible Tools	MongoDB, Cassandra, Redis, Couchbase, **(or) any other of your choice.**

PART 6: A 3-Minute Video Demo   **[10 Points]**

Problem Statement	Create a 3-minute video presentation to showcase and explain the key aspects and achievements of your distributed database project.
To-do Tasks	1. Prepare a script or outline for the video presentation, highlighting the main components of the project. 2. Create an engaging and informative video that visually demonstrates the distributed database system, its design, implementation, and functionality. 3. Narrate (voice-over) or add subtitles to the video to explain the project's objectives, challenges, and solutions. 4. Emphasize the significance of the project in the context of the chosen topic. 5. Include snapshots, diagrams, or animations to illustrate key concepts and results.
Deliverables	A 3-minute video presentation that effectively communicates the project's goals, execution, and outcomes.
Possible Tools	**Any Tools** - Video editing software, animation software, narration tools.

Submission Instruction:

1. **Project Organization:**
   - Create six folders, each dedicated to one of the project parts (Part 1 through Part 6).
   - Name these folders accordingly: "Part 1," "Part 2," "Part 3," "Part 4," "Part 5," and "Part 6."
2. **Folder Contents:**
   - In each folder, include the following:
     - Relevant documentation in PDF format (reports, schemas, diagrams, etc.).
     - Source code and scripts associated with the respective project part.
     - Part 6, Include the video presentation file.
3. **Naming Conventions:**
   - Name your files and documents descriptively to ensure clarity.
4. **Final Folder:**
   - Create one final folder to hold all the individual project part folders.
5. **Zip the Submission**
   - Submit the ZIP file containing all project parts on the specified submission platform.