Course Title: Distributed Database Management System **Credit:** 3
Course No: CSIT.424.2 **Number of period per week:** 3+3
Nature of the Course: Theory + Lab **Total hours:** 45+45
Year: Fourth Semester: Eight
Level: B. Sc. CSIT

1. Course Introduction

This course is designed to develop acquaintance with fundamental concepts of distributed databases. The course starts with the basic concepts and also includes distributed database design, distributed query processing, distributed transaction management, distributed concurrency control, distributed recovery, and introduction to parallel databases.

2. Objectives

On completion of this course, students will be able to develop knowledge in different basic to advanced concepts of distributed databases and fundamental concepts of parallel databases.

3. Specific Objectives and Contents

Specific Objectives	Contents
• Understand the concepts of distributed data processing • Know about distributed database systems • Comparison between distributed and centralized database systems • Understand the benefits of distributed databases • Gaining knowledge about different design issues of distributed databases • Know about different architectures of distributed database management systems	**Unit One: Introduction [7 Hrs.]** 1.1. Distributed Data Processing 1.2. What is a Distributed Database System? 1.3. Distributed vs. Centralized Database System 1.4. Promises of DDBSs: Transparent Management of Distributed and Replicated Data, Reliability Through Distributed Transactions, Improved Performance, Easier System Expansion 1.5. Design Issues: Distributed Database Design, Distributed Directory Management, Distributed Query Processing, Distributed Concurrency Control, Distributed Deadlock Management, Reliability of Distributed DBMS, Replication, Relationship among Problems, Additional Issues 1.6. Distributed DBMS Architecture: ANSI/SPARC Architecture, A Generic Centralized DBMS Architecture, Architectural Models for Distributed DBMSs, Autonomy, Distribution, Heterogeneity, Architectural Alternatives, Client/Server Systems, Peer-to-Peer Systems, Multidatabase System Architecture
• Understand about the top down design process of distributed	**Unit Two: Distributed Database Design [8 Hrs.]** 2.1. Top-Down Design Process

Databases • Know about different design issues of distribution of data • Gaining knowledge about fragmentation • Know about allocation • Understanding the concepts of data replication and different replication protocols	2.2. Distribution Design Issues: Reasons for Fragmentation, Fragmentation Alternatives, Degree of Fragmentation, Correctness Rules of Fragmentation, Allocation Alternatives, Information Requirements 2.3. Fragmentation: Horizontal Fragmentation, Vertical Fragmentation, Hybrid Fragmentation 2.4. Allocation: Allocation Problem, Information Requirements, Allocation Model, Solution Methods 2.5. Data Replication and Replication Protocols
• Know about query processing problem in distributed databases • Understanding objectives of distributed query processing • Know the complexity of relational algebra operations • Know about different query processing characterization • Know about different layers of query processing	**Unit Three: Overview of Query Processing [7 Hrs.]** 3.1. Query Processing Problem 3.2. Objectives of Query Processing 3.3. Complexity of Relational Algebra Operations 3.4. Characterization of Query Processors: Languages, Types of Optimization, Optimization Timing, Statistics, Decision Sites, Exploitation of the Network Topology, Exploitation of Replicated Fragments, Use of Semijoins 3.5. Layers of Query Processing: Query Decomposition, Data Localization, Global Query Optimization, Distributed Query Execution
• Know the concept of transaction • Know the proprieties of transaction • Understand different types of transactions	**Unit Four: Introduction to Transaction Management [5 Hrs.]** 4.1. Definition of a Transaction: Termination Conditions of Transactions, Characterization of Transactions, Formalization of the Transaction Concept 4.2. Properties of Transactions: Atomicity, Consistency, Isolation, Durability 4.3. Types of Transactions: Flat Transactions, Nested Transactions, Workflows
• Know about serializability concepts • Understand different lock-based concurrency control algorithms • Understand different timestamp-based algorithms • Know about optimistic algorithms • Know to handle deadlock in distributed databases	**Unit Five: Distributed Concurrency Control [8 Hrs.]** 5.1. Serializability Theory 5.2. Locking-Based Concurrency Control Algorithms: Centralized 2PL, Distributed 2PL 5.3. Timestamp-Based Concurrency Control Algorithms: Basic TO Algorithm, Conservative TO Algorithm, Multiversion TO Algorithm 5.4. Optimistic Concurrency Control Algorithms 5.5. Deadlock Management: Deadlock Prevention, Deadlock Avoidance, Deadlock Detection and

	Resolution
• Know about reliability concepts • Know about failures in distributed databases • Know about local reliability • Know about distributed reliability protocols • Know to deal with site failures • Know the concept of network partitioning	**Unit Six: Distributed DBMS Reliability [6 Hrs.]** 6.1. Reliability Concepts and Measures: System, State, and Failure; Reliability and Availability; Mean Time between Failures/Mean Time to Repair 6.2. Failures in Distributed DBMS: Transaction Failures; Site (System) Failures; Media Failures; Communication Failures 6.3. Local Reliability Protocols: Architectural Considerations; Recovery Information; Execution of LRM Commands; Checkpointing; Handling Media Failures 6.4. Distributed Reliability Protocols: Components of Distributed Reliability Protocols; Two-Phase Commit Protocol; Variations of 2PC 6.5. Dealing with Site Failures: Termination and Recovery Protocols for 2PC, Three-Phase Commit Protocol 6.6. Network Partitioning: Centralized Protocols, Voting-based Protocols
• Know about parallel database architectures • Understand about placement of parallel data in database • Know the concept of parallel query processing • Know about load balancing	**Unit Seven: Parallel Database Systems [4 Hrs.]** 7.1. Parallel Database System Architectures: Objectives; Functional Architecture; Parallel DBMS Architectures 7.2. Parallel Data Placement 7.3. Introduction to Parallel Query Processing 7.4. Load Balancing: Parallel Execution Problems; Intra-Operator Load Balancing; Inter-Operator Load Balancing; Intra-Query Load Balancing

Evaluation System

Undergraduate Programs							
External Evaluation	**Marks**	**Internal Evaluation**	**Weightage**	**Marks**	**Practical**	**Weightage**	**Mark**
End semester examination	60	Assignments	20%	20	Practical Report copy	25%	20
(Details are given in the separate table at the end)		Quizzes	10%		Viva	25%	
		Attendance	20%		Practical Exam	50%	
		Internal Exams	50%				
Total External	60	Total Internal	100%	20		100%	20

Full Marks 60+20+20 = 100

External evaluation

1. **End semester examination:**
   It is a written examination at the end of the semester. The questions will be asked covering all the units of the course. The question model, full marks, time and others will be as per the following grid.

2. **External Practical Evaluation:**
   After completing the end semester theoretical examination, practical examination will be held. External examiner will conduct the practical examination according to the above mentioned evaluation. There will be an internal examiner to assist the external examiner. Three hours time will be given for the practical examination. In this examination Students must demonstrate the knowledge of the subject matter.

Full Marks: 100, Pass Marks: 45, Time: 3 Hrs

Nature of question	Total questions to be asked	Total questions to be answered	Total marks	Weightage
Group A: multiple choice*	20	20	20×1 = 20	60%
Group B: Short answer type questions	7	6	6×8 = 48	60%
Group C: Long answer type question/long menu driven programs	3	2	2×16 =32	60%
			100	100%

Each student must secure at least 50% marks in internal evaluation in order to appear in the end semester examination. Failed student will not be eligible to appear in the end semester examinations.

Internal evaluation

Assignment: Each student must submit the assignment individually. The stipulated time for submission of the assignment will be seriously taken.

Quizzes: Unannounced and announced quizzes/tests will be taken by the respective subject teachers. Such quizzes/tests will be conducted twice per semester. The students will be evaluated accordingly.

Attendance in class: Students should regularly attend and participate in class discussion. Eighty percent class attendance is mandatory for the students to enable them to appear in the end semester examination. Below 80% attendance in the class will signify NOT QUALIFIED (NQ) to attend the end semester examination.

Presentation: Students will be divided into groups and each group will be provided with a topic for presentation. It will be evaluated individually as well as group-wise. Individual students have to make presentations on the given topics.

Mid-term examination: It is a written examination and the questions will be asked covering all the topics in the session of the course.

Discussion and participation: Students will be evaluated on the basis of their active participation in the classroom discussions.

Instructional Techniques: All topics are discussed with emphasis on real-world application. List of instructional techniques is as follows:

- Lecture and Discussion
- Group work and Individual work
- Assignments
- Presentation by Students
- Quizzes
- Guest Lecture

Students are advised to attend all the classes and complete all the assignments within the specified time period. If a student does not attend the class(es), it is his/her sole responsibility to cover the topic(s) taught during that period. If a student fails to attend a formal exam/quiz/test, there won't be any provision for re-exam.

Laboratory Work

The laboratory work develops practical knowledge on different concepts of Distributed database design. Students should be able to design distributed database and distributed query to retrieve, from distributed database.

Prescribed Text:

1. Principles of Distributed Database Systems, **Özsu**, M. Tamer, **Valduriez**, Patrick, Third Edition.

References:

1. Distributed Database Management Systems: A Practical Approach, Saeed K. Rahimi, Frank S. Haug.
2. Distributed Database Systems, Chhanda Ray