

GUJARAT TECHNOLOGICAL UNIVERSITY

SUBJECT NAME: Distributed DBMS

SUBJECT CODE: 2170714

B.E. 7th SEMESTER

Type of course: Elective

Prerequisite: Database Management Systems & Networking

Rationale: Students are familiar with Centralized DBMS. This subject will give introduction to Distributed DBMS and associated problems. Students will be able to understand various algorithms and techniques for managing distributed database.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks						Total Marks
L	T	P		Theory Marks			Practical Marks			
			ESE (E)	PA (M)		ESE (V)		PA (I)		
				PA	ALA	ESE	OEP			
3	0	2	5	70	20	10	20	10	20	150

Content:

Sr. No.	Content	Total Hrs	% Weightage
1	Introduction: Distributed Data Processing, Distributed Database Systems, Promises of DDBSs, Complicating factors, Problem areas	03	7
2	Overview of RDBMS: Concepts, Integrity, Normalization	02	5
3	Distributed DBMS Architecture : Models- Autonomy, Distribution, Heterogeneity DDBMS Architecture – Client/Server, Peer to peer, MDBS	03	7
4	Data Distribution Alternatives: Design Alternatives – localized data, distributed data Fragmentation – Vertical, Horizontal (primary & derived), hybrid, general guidelines, correctness rules Distribution transparency – location, fragmentation, replication Impact of distribution on user queries – No Global Data Dictionary(GDD), GDD containing location information, Example on fragmentation	05	15
5	Semantic Data Control : View Management, Authentication – database authentication, OS authentication, Access Rights, Semantic Integrity Control – Centralized & Distributed , Cost of enforcing semantic integrity	03	10

6	Query Processing : Query Processing Problem, Layers of Query Processing Query Processing in Centralized Systems – Parsing & Translation, Optimization, Code generation, Example Query Processing in Distributed Systems – Mapping global query to local, Optimization,	04	10
7	Optimization of Distributed Queries: Query Optimization, Centralized Query Optimization, Join Ordering Distributed Query Optimization Algorithms	06	10
8	Distributed Transaction Management & Concurrency Control: Transaction concept, ACID property, Objectives of transaction management, Types of transactions, Objectives of Distributed Concurrency Control, Concurrency Control anomalies, Methods of concurrency control, Serializability and recoverability, Distributed Serializability, Enhanced lock based and timestamp based protocols, Multiple granularity, Multi version schemes, Optimistic Concurrency Control techniques	08	18
9	Distributed Deadlock & Recovery Deadlock concept, Deadlock in Centralized systems, Deadlock in Distributed Systems – Detection, Prevention, Avoidance, Wait-Die Algorithm, Wound-Wait algorithm Recovery in DBMS - Types of Failure, Methods to control failure, Different techniques of recoverability, Write- Ahead logging Protocol, Advanced recovery techniques- Shadow Paging, Fuzzy checkpoint, ARIES, RAID levels, Two Phase and Three Phase commit protocols	08	18

Suggested Specification table with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
10	15	15	15	10	5

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Reference Books:

1. Principles of Distributed Database Systems, Ozsu, Pearson Publication
2. Distributed Database Management Systems, Rahimi & Haug, Wiley
3. Distributed Database Systems, Chanda Ray, Pearson Publication
4. Distributed Databases, Sachin Deshpande, Dreamtech

Course Outcome:

After learning the course the students should be able to:

- Understand what is Distributed DBMS
- Understand various architectures of DDBMS
- Apply various fragmentation techniques given a problem
- Understand and calculate the cost of enforcing semantic integrity control
- Understand the steps of query processing
- How optimization techniques are applied to Distributed Database
- Learn and understand various Query Optimization Algorithms
- Understand Transaction Management & Compare various approaches to concurrency control in Distributed database
- Understand various algorithms and techniques for deadlock and recovery in Distributed database

List of Experiments:

- [1] Create two databases either on single DBMS and Design Database to fragment and share the fragments from both database and write single query for creating view.
- [2] Create two databases on two different computer systems and create database view to generate single DDB.
- [3] Create various views using any one of examples of database and Design various constraints.
- [4] Write and Implement algorithm for query processing using any of Example in either C /C++ /Java / .NET
- [5] Using any of example, write various Transaction statement and show the information about concurrency control [i.e. various lock's from dictionary] by executing multiple update and queries.
- [6] Using Transaction /commit rollback, Show the transaction ACID properties.
- [7] Write java JDBC program and use JTA to show various isolation level's in transaction.
- [8] Implement Two Phase Commit Protocol
- [9] Case study on noSQL
- [10] Case study on Hadoop

Design based Problems (DP)/Open Ended Problem:

1. Countrywide drug supplier chain operates from five different cities in the country and it maintains following database.
Shop(ds-id, ds-city, ds-contactno)
Medicine(med-id, med-name,manuf-id)
Manufacturer(manuf-id, manuf-name, manuf-city)
Order(med-id, ds-id,qty)
Suggest fragmentation and allocation schema considering following frequent queries”
 - (a) List manufacturer names who belong to the same city in which the drug shop that has placed an order resides.
 - (b) How many orders are generated from a city say “Ahmedabad”?
Justify your design and mention assumptions if any clearly.

2. Consider relations EMP(eno,ename,title) and ASG(eno,pno, resp,dur). Write down suitable queries in SQL-like syntax and in relational algebra for finding the names of employees who are managers of any project. Is the query optimized? If not, optimize it.

Major Equipment:

Networking of computers, RDBMS

List of Open Source Software/learning website:

1. https://docs.oracle.com/cd/B10501_01/server.920/a96521/ds_concepts.htm
2. <https://cs.uwaterloo.ca/~tozsu/ddbook/presentation-slides.php>
3. http://www.tutorialspoint.com/distributed_dbms/distributed_dbms_databases.htm

ACTIVE LEARNING ASSIGNMENTS: Preparation of power-point slides, which include videos, animations, pictures, graphics for better understanding theory and practical work – The faculty will allocate chapters/ parts of chapters to groups of students so that the entire syllabus to be covered. The power-point slides should be put up on the web-site of the College/ Institute, along with the names of the students of the group, the name of the faculty, Department and College on the first slide. The best three works should submit to GTU.