**FR. Conceicao Rodrigues College Of Engineering**

Father Agnel Ashram, Bandstand, Bandra-west, Mumbai-50

**Department of Information Technology**

**T.E. (IT) (semester V)  (2019-2020)**

**Lesson Plan**

**Subject: Advanced Database Management Technologies(TE ITC503)**

 **Credits-4**

SYLLABUS

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr.****No.** | **Module**  | **Detailed Content**  | **CO****Mapping** |
| 00 | Prerequisite | Reviewing basic concepts of aRelational database, SQL concepts |  |
| 01 | Query Processingand Optimization: | Overview, Measures of Query CostSelection Operation, Sorting, JoinOperation, Other OperationsEvaluation of Expressions.Query Optimization Overview,Transformation of RelationalExpressions Estimating Statistics ofExpression Results Choice ofEvaluation Plans | CO1 |
| 02 | TransactionsManagement andConcurrency: | Transaction concept, Transactionstates, ACID properties,Implementation of atomicity anddurability, Concurrent Executions,Serializability, Recoverability,Implementation of isolation,Concurrency Control: Lock-based,Time-stamp based Deadlockhandling, Recovery System: FailureClassification, Storage structure,Recovery & atomicity, Log basedrecovery, Checkpoints, ShadowPaging, ARIES Algorithm. | CO2 |
| 03 | Advanced DataManagementtechniques | Advanced Database Access protocols:Discretionary Access Control Basedon Granting and Revoking Privileges;Mandatory Access Control and Role-Based Access Control.Overview of Advanced Databasemodels like Mobile databases,Temporal databases, Spatial databases. | CO3CO4 |
| 04 | DistributedDatabases | Introduction : Distributed DataProcessing, What is a DistributedDatabase System? Design Issues .Distributed DBMS Architecture.Distributed Database Design : Top-DownDesign Process, Distribution DesignIssues, Fragmentation , Allocation .Overview of Query Processing : QueryProcessing Problem, Objectives of QueryProcessing, Complexity of RelationalAlgebra Operations, Characterization ofQuery Processors, Layers of QueryProcessing, Query Optimization inDistributed Databases; Overview of Transaction Management inDDB;Overview of Concurrency Control inDDB;Overview of Recovery in DDB | CO4 |
| 05 | Data Warehousing,DimensionalModeling andOLAP | The Need for Data Warehousing; DataWarehouse Defined; Benefits of DataWarehousing ; Features of a DataWarehouse; Data WarehouseArchitecture; Data Warehouse andData Marts; Data Warehousing DesignStrategies.Dimensional Model Vs ER Model;The Star Schema; How Does a QueryExecute? The Snowflake Schema;Fact Tables and Dimension Tables;Factless Fact Table; Updates ToDimension Tables, Primary Keys,Surrogate Keys & Foreign Keys;Aggregate Tables; Fact ConstellationSchema or Families of StarNeed for Online AnalyticalProcessing; OLTP vs OLAP; OLAPOperations in a cube: Roll-up, Drilldown,Slice, Dice, Pivot ; OLAPModels: MOLAP, ROLAP, HOLAP. | CO5 |
| 06 | ETL Process | Challenges in ETL Functions; DataExtraction; Identification of DataSources; Immediate Data Extraction,Deferred Data Extraction; DataTransformation: Tasks Involved inData Transformation, Techniques ofData Loading, Loading the FactTables and Dimension Tables | CO6 |

**Text Books:**

1. Korth, Slberchatz,Sudarshan, :”Database System Concepts”, 6th Edition, McGraw – Hill

2. Elmasri and Navathe, “Fundamentals of Database Systems”, 6th Edition, PEARSON

Education.

3. Theraja Reema, “Data Warehousing”, Oxford University Press, 2009.

4. Raghu Ramakrishnan and Johannes Gehrke, “Database Management Systems” 3rd Edition -

McGraw Hill

References:

1. Paulraj Ponniah, “Data Warehousing: Fundamentals for IT Professionals”, Wiley India.

2. Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom “Database System

Implementation”, Pearson Ltd. 1/ e

3. Thomas M. Connolly Carolyn Begg, Database Systems : A Practical Approach to Design,

Implementation and Management, 4/e, Pearson Ltd.

4. Ralph Kimball, Margy Ross, “The Data Warehouse Toolkit: The Definitive Guide To

Dimensional Modeling”, 3rd Edition. Wiley India.

5. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition.

**Internal Assessment for 20 marks:**

Consisting of **Two Compulsory Class Tests**

Approximately 40% to 50% of syllabus content must be covered in First test and remaining 40% to 50% of syllabus contents must be covered in second test.

**CO-Statements:**

|  |  |
| --- | --- |
| **Sr.No.** | **Course Outcome Statement** |
| **TEITC503.1** | Explain and understand the concept of a transaction and how ACIDproperties are maintained when concurrent transactions occur in database. |
| **TEITC503.2** | Measure query cost and design alternate efficient paths for query execution. |
| **TEITC503.3** | Apply sophisticated access protocols to control access to the database |
| **TEITC503.4** | Implement alternate models like distributed databases and design applications uisng advanced models like mobile, spatial databases. |
| **TEITC503.5** | Organize strategic data in an enterprise and build a data Warehouse. |
| **TEITC503.6** | Analyze data using OLAP operations so as to take strategic decisions. |
|  |  |

**CO-PO-PSO Mapping**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course****Name** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO****11** | **PO****12** | **PSO1** | **PSO2** |
| CO1 | 2 | 1 |  |  |  |  |  |  |  |  |  |  |  | 2 |
| CO2 | 2 | 1 |  |  | 1 |  |  |  |  |  |  |  | 1 | 2 |
| CO3 | 2 |  |  |  | 1 |  |  |  |  |  |  |  | 1 | 1 |
| CO4 | 2 |  |  |  | 1 |  |  |  |  |  |  |  |  | 2 |
| CO5 | 1 | 1 | 2 |  | 1 |  |  |  |  |  |  |  | 1 | 2 |
| CO6 | 1 | 1 | 2 |  | 1 |  |  |  |  |  |  |  | 1 | 2 |

**CO Assessment Tools**

|  |  |  |
| --- | --- | --- |
|  | **Direct Methods** | **Indirect Methods** |
|  | **Test1** | **Assig1** | **Lab Work** | **Test2** | **Assig2** | **University Theory Exam** | **University Practical Exam** | Course Exit Survey |
| CO1 | 20% | 20% | **15%** |  |  | **20%** | 25% | 100% |
| CO2 | 20% | 20% | **15%** |  |  | **20%** | 25% | 100% |
| CO3 |  |  |  | 30% | **25%** | **20%** | 25% | 100% |
| CO4 |  |  | **25%** | 15% | **15%** | **20%** | 25% | 100% |
| CO5 |  |  | **10%** | **25%** | **20%** | **20%** | 25% | 100% |
| CO6 |  |  | **10%** | **25%** | **20%** | **20%** | 25% | 100% |

**Content beyond syllabus**

**Information Package diagram**

**Lecture Plan:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lecture no | Topic | Planned date  | Actual Date | Mode of teaching |
| 1 | Reviewing basic concepts of a Relational database, SQL concepts, Overview, Measures of Query Cost  | 2/7/2019 | 3/7/2019 | White board marker |
| 2 | Selection Operation, | 3/7/2019 | 4/7/2019 | White board marker |
| 3 | Sorting | 4/7/2019 | 7/7/2019 | White board marker |
| 4 | Join Operation, Other Operations Evaluation of Expressions. | 5/7/2019 | 9/7/2019 | White board marker |
| 5 | Query Optimization Overview, Transformation of Relational Expressions  | 9/7/2019 | 10/7/2019 | White board marker |
| 6 | Estimating Statistics of Expression Results Choice of Evaluation Plans | 10/7/2019 | 11/7/2019 | White board marker |
| 7 | Estimating Statistics of Expression Results Choice of Evaluation Plans | 11/7/2019 | 12/7/2019 | White board marker |
| 8 | Introduction : Distributed Data Processing, What is a Distributed Database System? | 12/7/2019 | 16/7/2019 | PPT, White board marker |
| 9 | Design Issues . Distributed DBMS Architecture.  | 15/7/2019 | 17/7/2019 | PPT, White board marker |
| 10 | Recoverability, Implementation of isolation,  | 16/7/2019 | 19/7/2019 | PPT, White board marker |
| 11 | Concurrency Control: Lock based,  | 17/7/2019 | 22/7/2019 | PPT, White board marker |
| 12 | Time stamp based ,  | 19/7/2019 | 23/7/2019 | PPT, White board marker |
| 13 | Deadlock handling, | 22/7/2019 | 29/7/2019 | PPT, White board marker |
| 14 | Recovery System: Failure Classification,  | 23/7/2019 | 26/7/2019 | PPT, White board marker |
| 15 | Storage structure, Recovery & atomicity,  | 24/7/2019 | 29/7/2019 | PPT, White board marker |
| 16 | Log based recovery,  | 26/7/2019 | 30/7/2019 | PPT, White board marker |
| 17 | Checkpoints, Shadow Paging, | 29/7/2019 | 31/7/2019 | PPT, White board marker |
| 18 | ARIES Algorithm. | 30/7/2019 | 21/7/2019 | White board marker |
| 19 | ARIES Algorithm. | 31/7/2019 | 6/7/2019 | black board and chalk |
| 21 | ARIES Algorithm. | 2/8/2019 | 7/8/2019 | White board marker |
| 21 | introduction : Distributed Data Processing, What is a Distributed Database System? | 5/8/2019 | 9/8/2019 | PPT, White board marker |
| 22 | Design Issues . Distributed DBMS Architecture.  | 6/8/2019 | 19/8/2019 | PPT, White board marker |
| 23 | Allocation, Overview of Query Processing : Query Processing Problem, Objectives of Query Processing, | 7/8/2019 | 20/8/2019 | PPT, White board marker |
| 24 | Complexity of Relational Algebra Operations, Characterization of Query Processors, | 9/8/2019 | 20/8/2019 | PPT, White board marker |
| 25 | Layers of Query Processing, Query Optimization in Distributed Databases; | 19/8/2019 | 21/8/2019 | PPT, White board marker |
| 26 | Query Processing, Query Optimization in Distributed Databases; | 20/8/2019 | 23/8/2019 | PPT, White board marker |
| 27 | Overview of Transaction Management in DDB;  | 21/8/2019 | 23/8/2019 | PPT, White board marker |
| 28 | Overview of Concurrency Control in DDB; | 23/8/2019 | 26/8/2019 | PPT, White board marker |
| 29 | Overview of Recovery in DDB | 26/8/2019 | 27/8/2019 | PPT, White board marker |
| 30 | The Need for Data Warehousing; Data Warehouse Defined; Benefits of Data Warehousing ; Features of a Data Warehouse; Data Warehouse Architecture; Data Warehouse and Data Marts; Selection Operation, Data Warehousing Design Strategies. | 27/8/2019 | 28/8/2019 | White board marker |
| 31 | Data Warehouse Architecture;  | 28/8/2019 | 30/8/2019 | White board marker |
| 32 | Data Warehouse and Data Marts; Selection Operation, Data Warehousing Design Strategies. | 30/8/2019 | 9/9/2019 | White board marker |
| 33 | Dimensional Model Vs ER Model; The Star Schema; How Does a Query Execute?  | 9/9/2019 | 11/9/2019 | White board marker |
| 34 | The Snowflake Schema; Fact Tables and Dimension Tables; Factless Fact Table;  | 11/9/2019 | 13/9/2019 | White board marker |
| 35 | Aggregate Tables; Fact Constellation Schema or Families of Star, Need for Online Analytical Processing; OLTP vs OLAP;  | 13/9/2019 | 16/9/2019 | White board marker |
| 36 | OLAP Operations in a cube: Rollup, Drilldown, Slice, Dice, Pivot ;  | 16/9/2019 | 17/9/2019 | White board marker |
| 37 | OLAP Operations in a cube: Rollup, Drilldown, Slice, Dice, Pivot | 17/9/2019 | 20/9/2019 | White board marker |
| 38 | OLAP Models: MOLAP, ROLAP, HOLAP. | 18/9/2019 | 23/9/2019 | White board marker |
| 39 | Challenges in ETL Functions; Data Extraction; Identification of Data Sources;Immediate Data Extraction, Deferred Data Extraction;  | 20/9/2019 | 24/9/2019 | White board marker |
| 40 | Data Transformation: Tasks Involved in Data Transformation,Techniques of Data Loading, Loading the Fact Tables and DimensionTables | 23/9/2019 | 27/9/2019 | White board marker |
| 41 | Advanced Database Access protocols: Discretionary Access Control Based on Granting and Revoking Privileges;  | 24/9/2019 | 2/10/2019 | PPT, black board and chalk |
| 42 | Mandatory Access Control,  | 25/9/2019 | 4/10/2019 | PPT, black board and chalk |
| 43 | Role based Access control | 27/9/2019 | 7/10/2019 |  |
| 44 | Overview of Advanced Database models like Mobile databases, | 30/9/2019 | 9/10/2019 | White board marker |
| 45 | Temporal databases, Spatial databases | 1/9/2019 | 10/10/2019 | White board marker |

**Lab Plan for OLAP Lab**

**Lab Outcomes:**

LO1 -Implement simple query optimizers and design alternate efficient paths for query execution**.**

LO2-Simulate the working of concurrency protocols, recovery mechanisms in a database

LO3-Design applications using advanced models like mobile, spatial databases**.**

LO4-Implement query processing and transaction processing mechanisms.

LO5- Design Star schema, Snowflake schema and Fact constellation Schema.

LO6- Analyze data using OLAP operations so as to take strategic decisions

Lab Plan: ADMT

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No | Topic | Week No  | Lab outcome |
| 1 | To execute complex SQL queries in Posgresql | Week1 |  |
| 2 |  To implement cost estimation for different Join operations | Week2 | LO1 |
| 3 | To implement query cost optimization | Week3 | LO1 |
| 4 | To implement concurrency control algorithm | Week 4 | LO2 |
| 5 | To implement ARIES recovery algorithm | Week5 | LO2 |
| 6 | To implement Query Processing for distributed Databases | Week6 | LO4 |
| 7 | To implement Data Fragmentation | Week7 | LO4 |
| 8 | Case study on Data warehouse construction | Week8 | LO5 |
| 9 | Implementation of OLAP queries | Week9 | LO6 |
| 10 | Case study on Mobile, Temporal and Spatial databases | Week10 | LO3 |

**Assignment Plan:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assignment No** | **Date** | **Questions** | **CO/LO** |
| **1** | **16/8/2019** |  | **CO1,CO2** |
| **2** | **25/9/2019** |  | **CO2,CO3,CO4,CO5,CO6** |

  **Assignment 1 Questions**

1) Given this database schema, convert into relational algebra queries [TEITC503.2]

Schema:-

Product (pid, name, price)

Purchase (pid, cid, store)

Customer (cid, name, city)

1. Select distinct x.store from purchase x, customer y where x.cid=y.cid and y.city= ‘Seattle’
2. Select z.city, sum(x.price) from product x, purchase y, customer z where x.pid=y.pid and y.cid=z.cid and y.store= ‘Walmart’ groupby z.city having count(\*) >100

2) Draw a query tree for the following SQL query [TEITC503.2]

Select P.Pnumber, P.Dnum, E.Lname, E.address, E.bdate from Project as P, Department as D, Employee as E where P.Dnum= D.Dnumber and D.Mgr\_ssn= E.ssn and P.location = ‘Mumbai’

3) Consider the following schedule S of transactions T1, T2, T3, T4: Identify if S is conflict serializable and recoverable.[TEITC503.1]



4) Describe the timestamp based locking protocol. And Thomas write rule.[TEITC503.1]

**Assignment 2 Questions**

1) Consider a data warehouse for a hospital, where there are three dimensions: Doctor, Patient and Time. And two measures Count and charge.

Draw a star schema and snowflake schema for the given hospital. **[TEITC503.5]**

2) Using the above question, describe the OLAP operations: Rollup, DrillDown, Slice, Dice, Pivot.

 **[TEITC503.6]**

3) Analyze the log after crash and briefly answer the following questions: **[TEITC503.2]**

|  |  |
| --- | --- |
| 0 | BEGIN CHECKPOINT |
| 5 | END CHECKPOINT  |
| 10 | T1 UPDATE P1(OLD:YYY, NEW:ZZZ) |
| 15 | T1:UPDATE P2(OLD:WWW, NEW XXX) |
| 20 | T1:COMMIT |

What is done during Analysis, Redo, and Undo (Be precise about the points where each pahse begin and end, and describe the contents of tables used in Analysis)

4) Consider the following database that has to be distributed. **[TEITC503.4]**

PROJ ( **PNO**, PNAME, BUDGET)

PAY ( **TITLE**, SALARY)

EMP ( **ENO**, ENAME, TITLE)

ASG ( **ENO, PNO**, RESPONSIBILITY, DURATION)

Show two examples of Horizontal fragmentation, one example of derived, and one example of vertical.

5) Explain in short mobile database **[TEITC503.3]**

**Term Work:**

Term Work shall consist of at least 10 to 12 practical’s based on the above list. Also Term work

Journal must include at least 2 assignments.

**Term Work Marks:** 25 Marks (Total marks) = 15 Marks (Experiment) + 5 Marks (Assignments) + 5 Marks (Attendance)

**Oral Exam:** An Oral exam will be held based on the above syllabus.