

GUJARAT TECHNOLOGICAL UNIVERSITY

Master in Computer Application (Integrated MCA)

Year IV (Semester-VII) (W.E.F. June 2016)

Subject Name: Data Warehousing

Subject Code: 4470622

Learning Objectives:

- To understand the need and applications of Data Marts, and Data Warehouses
- To understand the schema and architecture of Data Mart, and Data Warehouse
- Ability to identify and finalize Facts and Dimensions for various different requirements
- Ability to use OLAP operations and understand the role of Meta Data
- Ability to implement a simple Data Mart for Examination Result Analysis
- To understand the need for and various techniques used for preprocessing
- To understand basic concepts and methodology of Extract, Transform, Load (ETL) operations

Prerequisites: Basics of Database Management System, SQL

Contents:

- 1. Introduction, the Data Warehouse Environment (6 Lect.)**
 - Introduction, Four Salient Characteristics of Data Warehouse: Subject Orientation, Integration, Non-volatility, Time Variance;
 - The Structure of Data Warehouse; Typical Process of Building a Data Warehouse; Granularity: Issues, Benefits, Examples;
 - Partitioning as a Design Approach; Structuring Data in the Data Warehouse; Data Homogeneity & Heterogeneity
- 2. Data Warehouse and Design (11 Lect.)**
 - Begin with Operational Data; Process and Data Models, and Architected Environment
 - Data Warehouse and Data Models: Mid-level Data Model, Physical Data Model; The Data Model and Iterative Development; Normalization and De-normalization
 - Meta-data; Cyclicity of Data
 - Transformation and Integration
 - Direct and Indirect Access of Data Warehouse Data: Examples; Star Joins; Data Marts; Zachman Framework;
 - Granularity in Data Warehouse: Basic Aspects, Example: Levels of Granularity – Banking Environment
 - Migration to the Architected Environment: Migration Plan; Strategic Considerations; Data-driven Development Methodology
- 3. Data Warehouse and Technology (6 Lect.)**
 - Managing Large Amounts of Data and Multiple Media;
 - Interfaces to Many Technologies; Programmers and Designer Control of Data Placement; Parallel Storage and Management of Data: Meta-data Management; Efficient Loading of

- Data;
- Indexing and Monitoring Data; Efficient Index Utilization; Compaction of Data; Compound Keys; Variable Length Data; Lock Management; Index-only Processing; Fast Restore; Other Technological Features
- DBMS Types and the Data Warehouse; Changing DBMS Technology; Multi-dimensional DBMS and the Data Warehouse;
- Data Warehousing across Multiple Storage Media; Role of Meta-data in Data Warehouse Environment;
- Context and Content: Three Types of Contextual Information; Capturing and Managing Contextual Information; Looking at the Past
- Refreshing Data Warehouse; Testing

4. Executive Information Systems and the Data Warehouse (6 Lect.)

- OLAP Operations: Drill-down, Roll-up, Slice, Dice, Pivot
- Executive Information System (EIS): Introduction; Potential of EIS; Examples; Drill-down analysis;
- Data Warehouse as a Basis for EIS: Event Mapping, Detailed versus Summarized Data and EIS; Examples

5. The Relational and Multi-dimensional Model; Web Log Clickstream Data Warehouse (5 Lect.)

- The Relational Model; The Multi-dimensional Model, Snowflake Structures; Difference between Star Join and Relational Structure
- Independent Data Marts, Building Independent Data Marts
- Data Warehouse and the Web: Granularity Manager to Process Web Log Clickstream Data;
- Operational Data Store (ODS) to Provide Information Flow between Data Warehouse and the Web
- Moving Data from the Web to the Data Warehouse and vice versa

6. Unstructured Data and the Data Warehouse (4 Lect.)

- Structured versus Unstructured Data; Handling Textual Data: Matching Text Across Environments: A Probabilistic Match, A Themed Match, Linkage through Themes and Themed Words; Linkage through Abstraction and Meta-data
- A Two-tiered Data Warehouse, Dividing Unstructured Data Warehouse; Visualizing Unstructured Data;
- A Self-organizing Map (SOM), The Unstructured Data Warehouse
- Fitting the Unstructured and Structured Environments Together

Text Books:

1. William H. Inmon, "Building the Data Warehouse", 4th Edition, Wiley India, (2012), Rs. 499/-

Reference Books:

1. Ralph Kimball, Margy Ross, Warren Thornwite, Joy Mundy, Bob Becker, "The Data Warehouse Life-cycle Toolkit", Wiley India, 2nd Edition, (2008), Rs. 699/-
2. Ralph Kimball, Joy Caserta, "The Data Warehouse ETL Toolkit", Wiley India, 1st Edition, (2004), Rs. 640/-
3. Jiawei Han, Micheline Kamber, Jain Pei, "Data Mining: Concepts and Techniques", Elsevier, 3rd Edition, 2011, Rs. 560/-

4. Alex Berson, Stephon J Smith, "Data Warehousing, Data Mining and OLAP", McGraw-Hill, Rs. 715/-
5. Soumendra Mohanty, "Data Warehousing - Design, Development and best Practices", McGraw-Hill, Rs. Rs. 495/-
6. S. Nagabhushan, "Data Warehousing OLAP and Data Mining", New Age International, 1st Edition, 2006, Rs. 275/-

Chapter wise coverage from the Text Books:

Unit-1: Ch-1, 2; Unit-2: Ch-3, 4; Unit-3: Ch-5; Unit-4: Ch-7;
Unit-5: Ch-13, 14 (Partially); Unit-6: Ch-11

Accomplishment of the student after completing the course:

The student will be able to identify Fact and Dimensions and develop schema for variety of problem statements. They will be able to decide on appropriate OLAP operations to carryout desired analysis.