Data Warehousing & Data Mining
CS704C
Contracts: 3L
Credits- 3

Module 1: Overview and Concepts of Data Warehousing (Lectures : 9)
Overview of Data warehousing
Strategic information and the need for Data warehousing, Defining a Data warehouse, Evolution of Data warehousing, Data warehousing and Business Intelligence

The Building Blocks of Data warehouse
Defining features - Subject-oriented data, Integrated data, Time-variant data, Nonvolatile data, Data granularity
Data warehouses and Data marts
Architectural Types - Centralized, Independent data marts, Federated, Hub-and-Spoke, Data mart bus
Overview of components - Source Data, Data Staging, Data Storage, Information Delivery, Metadata, and Management and Control components

Business Requirements and Data warehouse
Dimensional nature of Business data and Dimensional Analysis, Dimension hierarchies and categories, Key Business Metrics (Facts), Requirement Gathering methods and Requirements Definition Document (contents)
Business Requirements and Data Design - Structure for Business Dimensions and Key Measurements, Levels of detail
Business Requirements and the Architecture plan
Business Requirements and Data Storage Specifications
Business Requirements and Information Delivery Strategy

Module 2 : Data warehouse Architecture and Infrastructure (Lectures : 8)
Architectural components
Concepts of Data warehouse architecture - Definition and architecture in the areas of Data acquisition, Data storage, and Information delivery
Distinguishing characteristics - Different objectives and scope, Data content, Complex analysis for faster response, Flexible and Dynamic, Metadata-driven etc
Architectural Framework - supporting flow of data, and the Management and Control module
Technical architecture - Data acquisition, Data storage, and Information delivery
Overview of the components of Architectural Types introduced in Module 1.

Infrastructure for Data warehousing
Distinction between architecture and infrastructure, Understanding of how data warehouse infrastructure supports its architecture
Components of physical infrastructure, Hardware and Operating systems for data warehouse, Database Software, Collection of Tools,
Data warehouse Appliances - evolution and benefits

The role of Metadata
Understanding the importance of Metadata
Metadata types by functional areas - Data acquisition, Data storage, and Information delivery
Business Metadata - overview of content and examples
Technical Metadata - overview of content and examples
Metadata Requirements, Sources of Metadata, Metadata management - challenges, Metadata Repository, Metadata integration and standards

Module 3 : Data Design and Data Preparation (Lectures : 9)
Principles of Dimensional Modeling
Data Design - Design decisions, Basics of Dimensional modeling, E-R modeling versus Dimensional modeling
The STAR schema - illustration, Dimension Table, Fact Table, Factless Fact Table, Data granularity
STAR schema keys - Primary, Surrogate, and Foreign
Advantages of the STAR schema, STAR schema examples

Data Extraction, Transformation, and Loading
Overview of ETL, Requirements of ETL and steps
Data extraction - identification of sources and techniques
Data transformation - Basic tasks, Transformation types, Data integration and consolidation, Transformation for dimension attributes
Data loading - Techniques and processes, Data refresh versus update, Procedures for Dimension tables, Fact tables : History and incremental loads
ETL Tool options

Data Quality
Importance of data quality, Challenges for data quality, Data quality tools, Data cleansing and purification, Master Data Management

Module 4 : Information access and delivery (Lectures : 10)
Matching information to classes of users
Information from Data warehouse versus Operational systems, Users of information - their needs and how to provide information
Information delivery - queries, reports, analysis, and applications
Information delivery tools - Desktop environment, Methodology and criteria for tool selection, Information delivery framework, Business Activity Monitoring, Dashboards and Scorecards

OLAP in Data warehouse
Overall concept of Online Analytical Processing (OLAP), OLAP definitions and rules, OLAP characteristics
Major features and functions of OLAP - General features, Dimensional analysis, Hypercubes, Drill Down and Roll Up, Slice and Dice, Rotation, Uses and Benefits
Familiarity with OLAP models - Overview of variations, MOLAP, ROLAP, HOLAP, DOLAP, Database OLAP, Web OLAP

Data Warehouse and the web
Web-enabled Data Warehouse - adapting data warehouse for the web
Web-based information delivery - Browser technology for data warehouse and Security issues
OLAP and Web - Enterprise OLAP, Web-OLAP approaches, OLAP Engine design

Data Mining
Overview of Data mining - Definition, Knowledge Discovery Process (Relationships, Patterns, Phases of the process), OLAP versus Data mining
Some aspects of Data mining - Association rules, Outlier analysis, Predictive analytics etc)
Concepts of Data mining in a Data warehouse environment
Major Data Mining techniques - Cluster Detection, Decision Trees, Memory-based Reasoning, Link Analysis, Neural Networks, Genetic Algorithms etc
Data Mining Applications in industry - Benefits of Data mining, Discussion on applications in Customer Relationship Management (CRM), Retail, Telecommunication, Biotechnology, Banking and Finance etc

Books Recommended:
1. Data Warehousing Fundamentals for IT Professionals, Second Edition by Paulraj Ponniah, Wiley India

References:
2. Data Warehousing, Data Mining, & OLAP - Second Edition by Alex Berson and Stephen J. Smith, Tata McGraw Hill Education
3. Data warehouse Toolkit by Ralph Kimball, Wiley India