



C. U. SHAH UNIVERSITY
Wadhwan City

FACULTY OF:-Computer Science
DEPARTMENT OF:-Master of Computer Applications
SEMESTER:- V
CODE:- 5CS05MDW1
NAME:- DATA WAREHOUSING & DATA MINING (DWDM)

Teaching and Evaluation Scheme

Subject Code	Name of the Subject	Teaching Scheme (Hours)				Credits	Evaluation Scheme								
		T h	T u	P r	To tal		Theory				Practical (Marks)				To tal
							Sessional Exam		University Exam		Internal		University		
							Ma rks	H rs	Ma rks	H rs	Pr/ Viva	T W	Pr		
5CS05 MDW1	DATA WAREHOUSING & DATA MINING (DWDM)	4	0	0	4	4	30	1.5	70	3	----	---	---	100	

Objectives:-

- To understand the need of Data Warehouses over Databases, and the difference between
- Usage of operational and historical data repositories.
- To be able to differentiate between RDBMS schemas & Data Warehouse Schemas.
- To understand the concept of Analytical Processing (OLAP) and its similarities & differences with respect to Transaction Processing (OLTP).
- To conceptualize the architecture of a Data Warehouse and the need for pre-processing.
- To understand the need for Data Mining and advantages to the business world. The validating criteria for an outcome to be categorized as Data Mining result will be understood. To get a clear idea of various classes of Data Mining techniques, their need, scenarios (situations) and scope of their applicability.
- To learn the algorithms used for various type of Data Mining problems.

Prerequisites:-

- Knowledge of RDBMS and OLTP

Course outline:-

Sr. No.	Course content	Hours
1	<p>Introduction to Data Warehousing, A Multi-dimensional Data Model & Schemas, OLAP Operations & Servers</p> <p>An overview and definition along with clear understanding of the four key-words appearing in the definition. Differences between Operational Database Systems and Data Warehouses; Difference between OLTP & OLAP • Overview of Multi-dimensional Data Model, and the basic differentiation between Fact and Dimension; Multi-dimensional Cube Concept Hierarchies of Dimensions Parameters: Examples and the advantages Star, Snowflakes, and Fact Constellations Schemas for Multi-dimensional Databases Measures: OLAP Operations in Multi-dimensional Data Model: Roll-up, Drill-down, Slice & Dice, Pivot (Rotate)Indexing OLAP Data; Type of</p>	6



C. U. SHAH UNIVERSITY
Wadhwan City

	OLAP Servers: ROLAP versus MOLAP versus HOLAP ,Metadata Repository	
2	Data Warehouse Architecture; Further Development of Data Cube & OLAP Technology The Design of A Data Warehouse: A Business Analysis Framework; The Process of Data Warehouse Design .A 3-Tier Data Warehouse Architecture; Enterprise Warehouse, Data mart, Virtual Warehouse	3
3	Data Mining: Introduction An Overview; What is Data Mining; Data Mining – on What Kind of Data .Data Mining Functionalities – What Kind of Patterns Can be Mined; Concept/Class Description: Characterization & Discrimination; Mining Frequent Patterns, Associations, and Correlations; Classification & Prediction; Cluster Analysis; Outlier Analysis .Classification of Data Mining Systems. Data Mining Task Primitives.	8
4	Data Pre-processing The need for Pre-processing, Descriptive Data Summarization. Data Cleaning: Missing Values, Noisy Data, Data Cleaning as a Process. Data Integration & Transformation. Data Cube Aggregation; Attribute Subset Selection. Dimensionality Reduction: Basic Concepts only Numerosity Reduction: Regression & Log-linear Models, Histograms, Clustering, Sampling	5
5	Mining Frequent Patterns, Associations, and Correlations Basic Concepts: Market Basket Analysis; Frequent Itemsets, Closed Itemsets, and Association Rules; Frequent Pattern Mining: A Roadmap Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation; Generating Association Rules from Frequent Itemsets; Improving the Efficiency of Apriori From Association Mining to Correlation Analysis.	5
6	Classification & Prediction Introduction to Classification and Prediction; Basics of Supervised & Unsupervised Learning; Preparing the Data for Classification and Prediction; Comparing Classification and Prediction Methods. Classification by Decision Tree Induction, Attribute Selection Measures; • Rule-based Classification: Using IF-THEN Rules for Classification; Bayesian Classification: Bayes’ Theorem, Naïve Bayesian Classification; Bayesian Belief Networks.	12
7	Cluster Analysis Introduction to Cluster Analysis; Types of Data in Cluster Analysis; A Categorization of major Clustering Methods Partitioning Methods; Centroid-Based Technique: K-Means Method.	6
8	Data Mining Applications Financial Data Analysis, The Retail Industry, The Telecommunication Industry	5

Learning Outcomes:

- Ability to create a Star Schema for a given Data warehousing requirements
- Ability to decide the number & levels of pre-computed Data Cubes, the corresponding Metadata and the appropriate OLAP operation
- Ability to apply pre-processing on existing operational & historical data for creation of Data Warehouse
- Ability to apply Apriori algorithm for Association Mining
- Ability to apply Decision Tree and Bayesian algorithms for Classification

Teaching & Learning Methodology:

- Using Whiteboard & Multimedia or OHP

Books Recommended:

1. Data Mining: Concepts & Techniques 3rd ed. , **Jiawei Han & Micheline Kamber**, Morgan



C. U. SHAH UNIVERSITY **Wadhwan City**

Kaufmann Publishers (2006)

2. Building the Data Warehouse, **W. H. Inmon**, Wiley Dreamtech India Pvt. Ltd.
3. Data Warehousing: Design, Development and Best Practices, **Mohanty, Souendra**, Tata McGraw Hill (2006)
4. Data Mining, Addison-Wesley, **Pieter Adriaans & Dolf Zentinge**, Pearson (2000)
5. Data Mining Methods & Models, **Daniel T. Larose**, Wiley-India (2007)
6. Data Mining, **Vikram Pudi & P. Radhakrishnan**, Oxford University Press (2009)
7. Data Warehousing, **Alex Berson & Stephen J. Smith**, Data Mining & OLAP, Tata McGraw-Hill (2004)
8. Data Mining Techniques, **Michael J. A. Berry & Gordon S. Linoff**, Wiley-India (2008)
9. Data Mining – a Tutorial-based Primer, **Richard J. Roiger & Michael W. Geatz**, Pearson Education (2005)
10. Data Mining: Introductory and Advanced Topics, **Margaret H. Dunham & S. Sridhar**, Pearson Education (2008)
11. Introduction to Data Mining with Case Studies, **G. K. Gupta**, PHI (2006)