

Data Warehousing and Data Mining

MIE 212

Semester: Third
Credit Hour: 3

Full Marks: 75
Internal: 30
Final Exam: 45

General Objectives:

This course provide the overview techniques and development on data warehousing and data mining. It focuses on providing information regarding establishment of data warehouse and Online Analytical Processing (OLAP). It introduces broad research areas for further development.

Specific Objectives:

- to make students understand the principle of data mining and warehousing,
- to orient students about implementation technique of data mining,
- to orient students about mining primitives, language and system architecture,

Unit- 1

4 Hrs.

The Evolution and brief history of Data Warehousing today's Development Environment. Principles Of Data Warehousing (Architecture And Design Techniques) : Types of Data and Their Uses, Conceptual Data Architecture, Design Techniques, introduction to the Logical Architecture.

Unit- 2

5 Hrs.

Creating the Data Asset: Business Data Warehouse Design, Populating the Data warehouse, Unlocking the Data Asset for End Users: Designing, Business Information Warehouses, Populating Business information Warehouses, User Access to Information, Information Data in Context.

Unit- 3

4 Hrs.

Implementing The Warehouse (Managing the Project and Environment): Obstacles to Implementation, Planning your Implementation, Justifying the Warehouse, Organization Implications of Data Warehousing, The data Warehouse in your Organization, Data Warehouse Management, Looking to the Future.

Unit- 4

6 Hrs.

Differences between Operational Database Systems and Data Warehouses, a multidimensional Data Model, Data warehouse and OLAP technology, multidimensional data models and different OLAP operations, OLAP Server: ROLAP, MOLAP and HOLAP. Data warehouse implementation, efficient computation of data cubes, processing of OLAP queries, indexing OLAP data.

Unit- 5**5 Hrs.**

Data Mining Primitives, Languages, and System Architectures, graphical user interfaces. Concept Description: Characterization and Comparison, Data generalization and summarization-based characterization, Analytical characterization, analysis of attribute relevance, mining class comparisons, and mining descriptive statistical measures in large databases.

Unit- 6**8 Hrs.**

Mining Association Rules in Large Databases, Mining single-dimensional Boolean association rules from transactional databases, mining multilevel association rules from transaction databases, Mining multidimensional association rules from relational databases and data warehouses, From association mining to correlation analysis, Constraint-based association mining.

Unit- 7**4 Hrs.**

Classification and prediction, issues, classification by decision induction, Bayesian classification, classification by back propagation, classification based on concepts from association rule mining other classification methods.

Unit- 8**4 Hrs.**

Cluster Analysis Introduction : Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Partitioning Methods, Density-Based Methods, Grid-Based Methods, Model-Based Clustering Methods, Outlier Analysis.

Unit-9**5 Hrs.**

Mining Complex Types of Data: Multi-Dimensional Analysis and Descriptive Mining of Complex Data Objects, Mining Spatial databases, Mining Multimedia databases, Mining Time-Series and Sequence data, Mining Text databases, mining the World Wide Web.

References:

1. Data Mining Concepts and Techniques, Morgan Kaufmann J. Han, M Kamber Second Edition ISBN: 978-1-55860-901-3
2. Data Warehousing in the Real World – Sam Anahory and Dennis Murray, Pearson Edition Asia.