MCA-14-43 DATA WAREHOUSING AND MINING

Maximum marks: 100 (External: 80, Internal: 20) Time: 3 hours

Note: Examiner will be required to set NINE questions in all. Question Number 1 will consist of objective type/short-answer type questions covering the entire syllabus. In addition to question no. 1, the examiner is required to set eight more questions selecting two from each unit. Student will be required to attempt FIVE questions in all. Question Number 1 will be compulsory. In addition to compulsory question, student will have to attempt four more questions selecting one question from each Unit. All questions will carry equal marks.

UNIT – I

Data Warehouse: Basic concepts, The Data Warehouse - A Brief History, Characteristics, Difference between Operational Database Systems and Data Warehouse, Architecture for a Data Warehouse, Fact and Dimension Tables, Data Warehouse Schemas, Data Cube : A Multidimensional Data Model, Data Cube Computation Methods, Typical OLAP Operations, Data Warehouse Design and Usage, Data Warehouse Implementation, Data Generalization by Attribute Oriented Induction.

UNIT – II

Data Mining: Introduction: Motivation, Importance, Knowledge Discovery Process, Data Mining Functionalities, Interesting Patterns, Classification of Data Mining Systems, Major issues, Data Objects and Attribute Types. Data Preprocessing: Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization. Data Mining Models: Directed Data Mining Models, Directed Data Mining Methodology. Data Visualization. Outliers, Types of Outliers and Challenges of Outlier Detection.

UNIT – III

Data Mining Classical Techniques: Statistics – Similarity Models, Steps for Designing Similarity Models, Table Lookup Model. Clustering- Requirement for Cluster Analysis, Clustering Methods- Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of Clustering. Nearest Neighborhood- Memory Based Reasoning, Challenges of Memory Based Reasoning,

UNIT – IV

Data Mining Next Generation Techniques: Decision Tree- Decision Tree Induction, Attribute Selection Measures, Tree Pruning. Association Rule Mining- Market Basket Analysis, Frequent Itemset Mining using Apriori Algorithm, Improving the Efficiency of Apriori, Neural Network- Bayesian Belief Networks, Classification by Backpropagation. Data Mining Applications, Data Mining Trends and Tools.

Reference Books:

1. J Hanes, M. Kamber, “Data Mining Concepts and Techniques”, Elsevier India.
2. G.S. Linoff, M.J.A. Berry, “Data Mining Techniques”, Wiley India Pvt. Ltd.
3. A. Berson, S.J. Smith, “Data Warehousing, Data Mining & OLAP”, Tata McGraw- Hill.