Subject Name: Big Data Analytics

Subject Code: IT 801

Teaching Scheme (Credits and Hours)

Teaching scheme				Total Credit	Evaluation Scheme					
L	T	P	Total		Theory		Mid Sem Exam	CIA	Pract.	Total
Hrs	Hrs	Hrs	Hrs		Hrs	Marks	Marks	Marks	Marks	Marks
04	00	02	06	5	3	70	30	20	30	150

Learning Objectives:

The main objectives for offering the course Data Warehousing and Data Mining are:

- To introduce students to basic applications, concepts, and techniques of Data Warehousing & mining.
- Understand the fundamental processes, concepts and techniques of data mining and develop an appreciation for the inherent complexity of the data- mining task.
- To develop skills for using recent data mining software to solve practical problems in a variety of disciplines.
- To gain experience doing independent study and research.

Outline of the Course:

Sr. No	Title of the Unit	Minimum Hours
1	Introduction to Data Warehousing	6
2	Concepts and techniques in Data Warehousing	5
3	Introduction to data mining (DM)	5
4	Data Preprocessing	6
5	Concept Description & Association Rule Mining	8
6	Classification and Prediction	8
7	Advance topics	8
8	Introduction to Big Data, MapReduce and Hadoop, Hadoop Implementation and Deployment	14

Total hours (Theory): 60

Total hours (Lab): 30

Total hours: 90

Detailed Syllabus

Sr. No	Topic	Lecture Hours	Weightage(%)
1	**Overview and concepts Data Warehousing** What is data warehousing - The building Blocks, Defining Features – Data warehouses and data marts, Overview of the components, Metadata in the data warehouse, Need for data warehousing, Basic elements of data warehousing, Trends in data warehousing	6	10
2	**Concepts and techniques in Data Warehousing** OLAP (Online analytical processing) definitions, Difference between OLAP and OLTP, Dimensional analysis - What are cubes?, Drill-down and roll-up - slice and dice or rotation, OLAP models, ROLAP versus MOLAP, defining schemas: Stars, snowflakes and fact constellations	5	10
3	**Introduction to Data Mining (DM)** DM Functionalities, Classification of DM Systems, Issues in DM – KDD Process	5	10
4	**Data Preprocessing** Why to preprocess data?, Data cleaning: Missing Values, Noisy Data, Data Integration and transformation, Data Reduction: Data cube aggregation, Dimensionality Reduction, Data Compression, Numerosity Reduction, Data Mining Primitives, Languages and System Architectures: Task relevant data, Kind of Knowledge to be mined, Discretization and Concept Hierarchy	6	10
5	**Concept Description and Association Rule Mining** Introduction to Concept description, Data Generalization and summarization-based Characterization, Analytical Characterization, Class Comparisons, Descriptive Statistical Measures, Market basket analysis- basic concepts ,Association Rule Mining, The Apriori Algorithm, Mining Multilevel Association Rule Mining, Mining Multidimensional Association Rule Mining	8	12
6	**Introduction to Classification and Prediction** Introduction to classification and prediction, Issues regarding Classification, Classification using Decision trees, Bayesian Classification, Classification by Backpropagation, Prediction Classification Accuracy	8	12

7	**Advance topics** Introduction of Clustering, Spatial mining, Web mining, Text mining	6	10
8	**Introduction to Big Data, MapReduce and Hadoop:** What Is Big Data?, Driving the growth of Big Data, Differentiating between Big Data and traditional enterprise relational data, Challenges of Bid Data, Hadoop, MapReduce Why Is MapReduce Necessary?, How Does MapReduce Work?, Real-World MapReduce Examples	8	14
9	**Hadoop Implementation and Deployment:** Introducing Hadoop, Hadoop cluster components, Hadoop Architecture, Hadoop Ecosystem, Evaluation criteria for distributed MapReduce runtimes, Enterprise-grade Hadoop Deployment, Hadoop Implementation	8	12
	Total	**60**	**100**

Instructional Method and Pedagogy:

- At the start of course, the course delivery pattern, prerequisite of the subject will be discussed.
- Lectures will be conducted with the aid of multi-media projector, black board, OHP etc.
- Attendance is compulsory in lecture and laboratory which carries 10 marks in overall evaluation.
- One internal exam will be conducted as a part of internal theory evaluation.
- Assignments based on the course content will be given to the students for each unit and will be evaluated at regular interval evaluation.
- Surprise tests/Quizzes/Seminar/tutorial will be conducted having a share of five marks in the overall internal evaluation.
- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.
- Experiments shall be performed in the laboratory related to course contents.

Learning Outcome:

Upon completion of this course, students will be able to do the following:

- Students will able to understand important of data mining and its various concepts like data preprocessing, various classification algorithms etc.
- Student will be able to develop a reasonably sophisticated data mining application.
- Student will be able to develop a reasonably sophisticated data mining application.
- Student is able to select methods and techniques appropriate for the task
- Student is able to develop the methods and tools for the given task

Text Books:

- J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
- Paulraj Ponnian, "Data Warehousing Fundamentals", John Willey.
- Robert D. Schneider , Hadoop for Dummies, Wiley India.

Reference Books:

- M. Kantardzic, "Data mining: Concepts, models, methods and algorithms, John Wiley &Sons Inc.
- M. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson
- Pieter Adriaans, Dolf Zantinge , "Data Mining", Pearson Education Asia

List of experiments:

	Name of Experiment
1	Design and Create Cube by identifying measures and dimensions for Star Schema, Snowflake
2	Design and Create Cube by identifying measures and dimensions for Design storage for cube using storage
3	Process Cube and Browse Cube Data 1. By replacing a dimension in the grid, filtering and drilldown using cube browser 2. Browse dimension data and view dimension members, member properties, member property values 3. Create calculated member using arithmetic operators and member property of dimension Member
4	Create and use Excel Pivot Table Report based on data cube
5	Design and Create data mining models using Analysis Service of SQL server 2005
6	Design and Build targeted mailing data mining model using analysis service of SQL server 2005 and compare their predictive capabilities using the Mining Accuracy Chart View and Create predictions using Prediction Query Builder.
7	Perform various steps of Preprocessing on the given relational database / warehouse.
8	To implement Data Mining Extensions (DMX) language and MDX query language
9	Perform various steps of Preprocessing using WEKA software.
10	Creating Data Mining Structure & Predictive Models (Neural Networks and Decision Tree) using the Excel Add-In for SQL Server 2008.
11	Case Study: To study research papers on the given topic and prepare the report on it.
12	To setup Hadoop.
13	To run sample program using hadoop.