SEMESTER 4

DATA WAREHOUSING

Course Code	PEAIT412	CIE Marks	40
Teaching Hours/Week (L: T:P: R)	3:0:0:0	ESE Marks	60
Credits	3	Exam Hours	2 Hrs 30 Mins
Prerequisites (if any)	None	Course Type	Theory

Course Objectives:

1. To provide a comprehensive understanding of data warehousing concepts, architecture, and the role of dimensional modeling in designing data warehouses.
2. To equip students with practical skills in ETL processes, OLAP operations, and optimizing data warehouses for real-world applications.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	**Introduction to Data Warehousing :-** Overview of Data Warehousing - Need, Definition, Characteristics, Benefits, Operational vs. Analytical Systems, Applications in Industry. Architecture of Data Warehousing - Data Sources, Staging Area, Data Marts, ETL (Extract, Transform, Load) Processes, Components of a Data Warehouse (Data Models, OLAP, Metadata); Data Warehouse Models - Enterprise Data Warehouse, Virtual Warehouse, Data Mart. **[Text 1 : Ch 1 -3]**	9
2	**Dimensional Modeling & Design :-** Dimensional Modeling Basics - Facts, Dimensions, Star Schema, Snowflake Schema; Steps in Dimensional Modeling; Fact Tables, Dimension Tables; Data Warehouse Design-Top-Down & Bottom-Up, Fact Table Granularity; Slowly Changing Dimensions; Advances in Dimensional Modeling - Aggregates & Roll-Up, Time Dimension, Junk Dimensions, Degenerate Dimensions. **[Text 2 : Ch 1-5]**	9
3	**ETL Processes & Tools :-** ETL Concepts and Process - Extracting Data from Multiple Sources, Data Transformation and Cleansing, Loading Data; ETL Tools and Platforms - Common ETL Tools (Informatica, Talend, Microsoft SSIS), Scheduling and Automation of ETL Processes; Data Quality and Data Governance in ETL - Handling Incomplete and Inconsistent Data, Ensuring Data Accuracy, Consistency, and Integrity. **[Text 1 : Ch 4-7]**	9

KTUSPECIAL

4	**Data Warehouse Optimization & OLAP :-** Query Optimization in Data Warehouses - Indexing Strategies, Partitioning, Performance Tuning; OLAP - Architectures: MOLAP, ROLAP, HOLAP; Operations: Drill-Down, Roll-Up, Slicing, Dicing; Data Warehouse Maintenance - Data Loading Automation, Backup, and Recovery, Data Warehouse Evolution and Scalability. **Text 1 : Ch 8-10, Text 2 : Ch 7-9]**	9

Course Assessment Method
(CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment/ Microproject	Internal Examination-1 (Written)	Internal Examination- 2 (Written)	Total
5	**15**	**10**	**10**	**40**

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

KTU SPECIAL

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3 =24 marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 subdivisions. **(4x9 = 36 marks)**	**60**

Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the fundamental concepts of data warehousing, including dimensional modeling, star and snowflake schemas, and their role in organizing facts and dimensions.	K2
CO2	Apply ETL (Extract, Transform, Load) process and its significance in ensuring data quality and integrity in a data warehouse.	K3
CO3	Apply data warehouse performance optimization using indexing, partitioning and other query optimization techniques for efficient data retrieval and processing.	K3
CO4	Utilize OLAP tools and techniques to perform multidimensional data analysis, including drill-down, roll-up, and slicing operations on large datasets.	K3

Note*: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create*

KTU SPECIAL

CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	PO2	PO3	PO4	PO 5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	**3**	**2**										
CO2	**2**	**3**	**3**	**2**	**2**							**2**
CO3	**3**	**3**	**3**	**2**								**3**
CO4	**2**	**3**	**3**	**3**						**3**		**2**

Note: *1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High), -: No Correlation*

Text Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Data Warehousing fundamentals for IT Professionals	Paulraj Ponniah	Wiley	2/e, 2010
2	The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling	Ralph Kimball, Margy Ross	Wiley	3/e, 2013

Reference Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	DW 2.0: The Architecture for the Next Generation of Data Warehousing.	Derek Strauss, Genia Neushloss, W.H. H. Inmon	Morgan Kaufman	1/e, 2008
3	Data Warehousing and Data Mining	Express Learning [eBook]	ITL Education	1/e, 2012
4	Building the Data Lakehouse	Bill Inmon	Technics Pub.	1/e, 2021

Video Links (NPTEL, SWAYAM…)	
Module No.	**Link ID**
1	**https://archive.nptel.ac.in/courses/106/106/106106093/**

KTU SPECIAL