

Course Title: ETL Tool	Course Code: 24MJCS4S
Total Contact Hours: 26	No. of Credits: : 02
L:T:P: 1-0-2	Duration of SEE: 01
Internal Assessment Marks: 10	
Semester End Exam Marks: 40	

Course Outcomes (COs):

<ul style="list-style-type: none"> • Data Integration Skills: Students will be able to integrate data from multiple sources, such as databases, files, and applications.
<ul style="list-style-type: none"> • Data Transformation Skills: Students will be able to transform data into a standardized format, including data cleaning, data aggregation, and data formatting.
<ul style="list-style-type: none"> • Data Loading Skills: Students will be able to load data into a target system, such as a data warehouse or a database.
<ul style="list-style-type: none"> • Data Warehousing Skills: Students will be able to design and implement data warehouses using ETL tools.

Unit	Description	Hours
1	Introduction: Overview of Data Warehousing, Introduction to ETL (Extract, Transform, Load), Importance of ETL in Data Warehousing, Types of ETL Tools (On-Premise & Cloud-Based), ETL vs ELT, Architecture of ETL Process	7
2	Understanding Data Sources (Databases, Files, APIs, etc.), Types of Data Extraction, Connecting to Different Data Sources, Change Data Capture (CDC), Real-time vs Batch Extraction	6
	<ol style="list-style-type: none"> 1. Develop an ETL program to extract data from various sources, transform it into a standardized format, and load it into a data warehouse. 2. Create an ETL program to migrate data from an old system to a new system, transforming the data as needed to match the new system's requirements. 3. Develop an ETL program to synchronize data between two or more systems, ensuring that data is consistent and up-to-date across all systems. 4. Create an ETL program to cleanse data by removing duplicates, handling missing values, and correcting errors. 5. Develop an ETL program to aggregate data from multiple sources, summarizing data by grouping, filtering, and sorting. 6. Create an ETL program to validate data against a set of rules, ensuring that data meets specific criteria before loading it into a target system. 7. Develop an ETL program to load data into a database, handling issues such as data type conversions, indexing, and constraint checking. 8. Create an ETL program to load data into a data warehouse, handling issues such as data partitioning, indexing, and aggregation. 9. Develop an ETL program to load data into a big data system, handling issues such as data ingestion, processing, and storage. 10. Develop ETL programs using Informatica PowerCenter to extract, transform, and load data into various target systems. 	13

References:

1. Ralph Kimball, Margy Ross – *The Data Warehouse Toolkit*
 2. Mark Adelhart – *ETL Architecture for Data Integration*
- Jiawei Han, Micheline Kamber – *Data Mining: Concepts and Technique*