Data Engineering & Analytics

(Industry-Oriented Program – 140 Hours)

Course Overview

This course is designed to train technical freshers (B.Tech / MCA / BCA) in industry-oriented Data Engineering and Data Analytics using Python and SQL. The program focuses on practical implementation of data pipelines, database engineering, ETL workflows, and business intelligence systems aligned with current market requirements.

Students will gain hands-on experience in data processing, warehouse design, reporting systems, and analytics dashboards. The course also includes an overview of how modern AI systems depend on structured data pipelines, preparing learners for future AI-driven data roles.

This program prepares students for entry-level roles such as:

- Junior Data Engineer
- Data Analyst
- BI Developer
- SQL Engineer
- ETL Developer

Module 1: Python Programming for Data Engineering

Python syntax and structured programming
Functions and modular coding
Object-Oriented Programming fundamentals
Data structures (lists, dictionaries, sets, tuples)
File handling (CSV, JSON)
Exception handling and logging
Working with APIs for data ingestion
Writing clean and maintainable data processing code

Data Libraries:

- Python ecosystem
- Pandas
- NumPy
- Matplotlib

Hands-on Implementation:

- Data cleaning and preprocessing
- Transforming raw datasets
- Automating data processing scripts

Module 2: SQL & Relational Database Engineering

Relational database fundamentals
Database architecture concepts
ER diagram design
Normalization techniques
Transactions and data integrity

Advanced SQL:

- Joins (inner, left, right, full)
- Subqueries
- Grouping and aggregations
- Window functions
- Indexing and query optimization
- Execution plan basics
- Views and stored procedures

Practical Implementation:

- Designing a reporting database
- Writing optimized analytical queries
- Business reporting use cases

Module 3: Data Processing & Database Integration

Connecting Python applications with relational databases
Executing analytical queries using Python
Automating ETL scripts
Batch data processing techniques
Data validation and transformation rules
Error handling and logging in data workflows

Mini Project:

- Automated Data Processing System
  (File/API → Transform → Database → Reporting Table)

Module 4: Data Warehousing & ETL Fundamentals

Understanding OLTP vs OLAP systems
Data warehouse architecture
Star schema and Snowflake schema
Fact and Dimension tables
Data modeling for analytics
ETL lifecycle and workflow design
Data quality and validation strategies

Introduction (Concept + Demonstration):

- Apache Airflow

Practical Assignment:

- Designing a small data warehouse system

Module 5: Data Analytics & Business Intelligence

Exploratory Data Analysis (EDA)
Business metrics and KPI design
Data storytelling principles
Data visualization best practices
Dashboard structuring and reporting standards

Tool-Based Implementation:

Power BI dashboard development

Projects:

- Sales Analytics Dashboard
- Marketing Performance Dashboard
- HR Analytics Dashboard

Module 6: Data Engineering in AI Systems (Overview Module)

Understanding the role of data pipelines in AI systems
Training data lifecycle and preparation
Feature engineering fundamentals
Structured vs unstructured data handling
Embeddings and vectorization concepts
Overview of Retrieval-Augmented Generation (RAG) architecture

Conceptual Exposure To:

- OpenAI API
- LangChain

Focus:

Architecture-level understanding of how Data Engineering enables AI systems.
(No deep AI model development included.)

Capstone Project (Industry-Focused)

Students will design and implement an end-to-end Data Engineering & Analytics system:

Data Ingestion → ETL → Data Warehouse → Reporting Queries → Business Dashboard

Final Deliverables:

- Structured database design
- Optimized SQL queries
- Automated data processing scripts
- Interactive analytics dashboard
- Architecture documentation