

Practical Data Engineering Course - 40 Training Hours

This course is a hands-on professional training program focused on practical Data Engineering skills. It covers Python-based data ingestion, SQL and NoSQL data storage, data processing with Apache Spark, ETL pipeline development, workflow orchestration, and data quality practices.

Course Objectives

1. Understand Data Engineering responsibilities.
2. Design data pipelines
3. Ingest data using Python
4. Work with SQL and NoSQL
5. Process data using Spark
6. Build automated ETL pipelines
7. Orchestrate workflows using Airflow
8. Apply validation and optimization

Prerequisites

1. Basic Python
2. Basic SQL

Tools & Technologies

- Python
- SQL
- PostgreSQL
- MongoDB
- Apache Spark
- Apache Airflow
- Docker

Module 1: Data Engineering Fundamentals – 3 hours

- ✓ Pipeline components
- ✓ ETL/ELT
- ✓ Data flow design

Module 2: Data Ingestion with Python – 5 hours

- ✓ CSV/JSON ingestion
- ✓ APIs
- ✓ Cleaning

Module 3: SQL & Relational Databases – 5 hours

- ✓ Schema design
- ✓ Joins
- ✓ Indexing

Module 4: NoSQL Databases – 4 hours

- ✓ NoSQL concepts
- ✓ Document modeling
- ✓ CRUD

Module 5: Apache Spark – 7 hours

- ✓ DataFrames
- ✓ Transformations
- ✓ Aggregations

Module 6: ETL Pipeline Development – 5 hours

- ✓ Design
- ✓ Error handling
- ✓ Logging

Module 7: Workflow Orchestration – 5 hours

- ✓ DAGs
- ✓ Scheduling
- ✓ Dependencies

Module 8: Data Quality & Optimization – 3 hours

- ✓ Validation
- ✓ Checks
- ✓ Optimization

Module 9: Capstone Project – 3 hours

- ✓ End-to-end pipeline implementation.