multisoft
Systems

# IBM Data Engineering Professional Certificate Training

*COURSE CONTENT*


GET IN TOUCH

Multisoft Systems
B - 125, Sector - 2, Noida

(+91) 9810-306-956

info@multisoftsystems.com

www.multisoftsystems.com

## About Multisoft

Train yourself with the best and develop valuable in-demand skills with Multisoft Systems. A leading certification training provider, Multisoft collaborates with top technologies to bring world-class one-on-one and certification trainings. With the goal to empower professionals and business across the globe, we offer more than 1500 training courses, which are delivered by Multisoft's global subject matter experts. We offer tailored corporate training; project Based Training, comprehensive learning solution with lifetime e-learning access, after training support and globally recognized training certificates.

## About Course

The IBM Data Engineering Professional Certificate training by Multisoft Systems is designed to equip learners with the essential skills and knowledge required to excel in the field of data engineering. This comprehensive program covers the entire data engineering lifecycle, from data ingestion and transformation to storage, governance, and analysis.

## Module 1: Introduction to Data Engineering

- ✓ What is Data Engineering?
- ✓ Role of a Data Engineer
- ✓ Overview of Data Ecosystem
- ✓ Key Tools and Technologies for Data Engineers
- ✓ Career Path in Data Engineering

## Module 2: Python for Data Engineering

- ✓ Introduction to Python Programming
- ✓ Data Structures in Python
- ✓ Working with APIs and JSON
- ✓ Python Libraries for Data Engineering (Pandas, NumPy)
- ✓ Data Manipulation and Transformation
- ✓ Writing Data Pipelines in Python

## Module 3: Relational Databases and SQL

- ✓ Introduction to Relational Databases
- ✓ Core SQL Concepts (DDL, DML, DCL, TCL)
- ✓ Advanced SQL Queries (Joins, Subqueries, Window Functions)
- ✓ Indexing and Query Optimization
- ✓ Data Modeling Concepts
- ✓ Hands-on: SQL with IBM Db2 or PostgreSQL

## Module 4: ETL and Data Pipelines

- ✓ What is ETL (Extract, Transform, Load)?
- ✓ ETL Architecture and Patterns
- ✓ Building ETL Pipelines using Python and SQL
- ✓ Introduction to Apache Airflow
- ✓ Data Pipeline Orchestration

- ✓ Testing and Monitoring Pipelines

## Module 5: Data Warehousing

- ✓ Concepts of Data Warehousing
- ✓ OLTP vs OLAP Systems
- ✓ Star and Snowflake Schema
- ✓ Data Warehouse Architecture
- ✓ IBM Db2 Warehouse and IBM Netezza
- ✓ Querying Data Warehouses

## Module 6: Big Data and Distributed Computing

- ✓ Introduction to Big Data Ecosystem
- ✓ Hadoop and HDFS
- ✓ Introduction to Apache Spark
- ✓ PySpark for Data Engineering
- ✓ Distributed Data Processing
- ✓ Hands-on Labs: Processing Large Datasets

## Module 7: NoSQL Databases

- ✓ Introduction to NoSQL Concepts
- ✓ Types of NoSQL Databases: Key-Value, Document, Columnar, Graph
- ✓ Working with MongoDB
- ✓ Hands-on: CRUD Operations in MongoDB
- ✓ Data Modeling for NoSQL
- ✓ Use Cases and Limitations

## Module 8: Data Integration & Data Streaming

- ✓ Data Ingestion Patterns
- ✓ Introduction to Apache Kafka

- ✓ Streaming Data Pipelines
- ✓ Kafka Producers and Consumers
- ✓ Real-Time Data Processing with Spark Streaming
- ✓ Monitoring Streaming Applications

## Module 9: Data Lakes

- ✓ Introduction to Data Lakes
- ✓ Data Lake vs Data Warehouse
- ✓ IBM Cloud Object Storage for Data Lakes
- ✓ Organizing Data in Data Lakes
- ✓ Data Governance in Data Lakes

## Module 10: Data Governance, Security, and Privacy

- ✓ Data Governance Principles
- ✓ Data Quality and Data Lineage
- ✓ Data Security Fundamentals
- ✓ Encryption and Masking Techniques
- ✓ Regulatory Compliance (GDPR, HIPAA, etc.)
- ✓ IBM Tools for Data Governance

## Module 11: Cloud Data Engineering

- ✓ Cloud Platforms for Data Engineering (IBM Cloud, AWS, Azure, GCP)
- ✓ Cloud Storage Options
- ✓ Cloud Databases (Cloudant, Db2 on Cloud, Amazon RDS, BigQuery)
- ✓ Cloud ETL Tools
- ✓ CI/CD for Data Pipelines
- ✓ IBM Cloud Pak for Data

## Module 12: Capstone Project

- ✓ Designing an End-to-End Data Engineering Solution
- ✓ Building ETL Pipelines with Airflow
- ✓ Ingesting Data into Data Warehouse or Data Lake
- ✓ Processing with Spark / PySpark
- ✓ Visualization and Reporting
- ✓ Presenting Project Results