Certificate Program In Data Engineering

Duration – 80 Hours

Program Description

This program equips learners with the knowledge and skills to design, build, and manage scalable data pipelines and platforms that enable advanced analytics and AI. Covering modern data engineering tools, frameworks, and cloud ecosystems, it prepares professionals to work with structured, semi-structured, and unstructured data. Participants will gain hands-on experience in ingestion, transformation, storage, and orchestration of data at scale, aligned with industry practices.

Learning Goals

- Understand the role of data engineering in modern data-driven organizations
- Learn data modeling, ETL/ELT pipelines, and workflow orchestration
- Gain expertise in working with relational, NoSQL, and cloud-native databases
- Use distributed systems (Hadoop, Spark) for large-scale data processing
- Learn streaming data processing with Kafka and real-time frameworks
- Understand data security, governance, and compliance principles
- Apply DevOps/MLOps practices for deploying and monitoring pipelines
- Demonstrate skills through a real-world data engineering project

Course Topics

- Introduction to Data Engineering & Ecosystem
- Data Modeling & Warehousing (Star Schema, Snowflake, OLAP/OLTP)
- ETL/ELT Pipelines & Data Ingestion Strategies
- Distributed Data Processing (Hadoop, Apache Spark)
- Real-time Data Streaming (Apache Kafka, Flink)
- Cloud Data Platforms (AWS, Azure, GCP – BigQuery, Redshift, Synapse)
- Workflow Orchestration (Apache Airflow, Prefect, Dagster)
- Data Lakes, Lakehouse Architectures & Delta Lake
- Data Security, Governance, and Quality Management
- DevOps & MLOps Practices in Data Engineering
- Capstone Project: Building & Deploying an End-to-End Data Pipeline