**Advanced Data Engineering with Databricks: Data Engineer Professional**

**Duration: 40 Hours**

**Prerequisites:**

* Proficiency in Python and SQL.
* Hands-on experience with Databricks or Apache Spark.
* Understanding of data engineering workflows.
* Familiarity with Delta Lake (recommended).

**Scope of the Training:**

* Equip participants with advanced skills to handle enterprise-scale data engineering projects.
* Provide expertise in optimizing, securing, and automating data pipelines.
* Enable participants to design scalable data architectures using Databricks.
* Prepare participants for the **Databricks Data Engineer Professional Certification**.

**Course Content**

**Module 1: Advanced Concepts in Databricks**

* Understanding Databricks Architecture for Enterprise Scale
* Workspace Management and Advanced User Permissions
* Databricks Runtime: Tuning for Advanced Workloads

**Module 2: Data Ingestion and Transformation**

* Optimized Delta Lake Ingestion Patterns
* Advanced Transformation Techniques with PySpark
* Managing Schema Evolution in Delta Lake

**Module 3: Streaming Pipelines**

* Building Fault-Tolerant Streaming Pipelines
* Advanced Window Functions and Watermarking
* Integrating with Message Queues (Kafka, Event Hubs)

**Module 4: Advanced Delta Lake**

* Time Travel and Change Data Capture (CDC)
* Implementing Delta Lake in Multi-Hop Architectures
* Advanced Delta Table Optimizations and Constraints

**Module 5: Orchestration and Automation**

* Building and Managing Workflows with Databricks Workflows
* Integration with Airflow and Step Functions
* Automating Jobs with CI/CD Pipelines

**Module 6: Advanced Performance Tuning**

* Optimizing Storage and Compute Costs
* Adaptive Query Execution (AQE)
* Managing Shuffles and Broadcast Joins

**Module 7: Security and Governance**

* Implementing Role-Based Access Control (RBAC)
* Enabling Data Lineage and Auditing with Unity Catalog
* Advanced Encryption and Network Security

**Module 8: Hands-On Projects**

* Enterprise-Scale ETL Pipelines
* Real-Time Fraud Detection System
* Building a Data Lakehouse for Big Data Analytics