**Data Engineering with Databricks: Data Engineer Associate**

**Duration: 40 Hours**

**Prerequisites:**

* Familiarity with Python programming.
* Basic understanding of SQL.
* General knowledge of data processing and analytics.

**Scope of the Training:**

* Provide participants with foundational skills for building data engineering solutions.
* Enable participants to work effectively with Databricks and Spark for ETL workflows.
* Prepare participants for the **Databricks Data Engineer Associate Certification**.
* Build confidence in handling real-world data engineering challenges.

**Course Content**

**Module 1: Introduction to Data Engineering on Databricks**

* Data Engineering Fundamentals
* Overview of Databricks Platform
* Setting Up and Managing Databricks Clusters

**Module 2: ETL Development**

* Loading and Transforming Data with DataFrames
* Managing Schema and Data Quality
* Optimizing ETL Workflows with Delta Lake

**Module 3: Batch and Stream Processing**

* Batch Data Processing Concepts
* Introduction to Structured Streaming
* Unified Batch and Stream Processing in Databricks

**Module 4: Delta Lake Essentials**

* Introduction to Delta Lake
* Building Reliable and Scalable Data Pipelines
* Implementing ACID Transactions and Time Travel

**Module 5: Orchestration**

* Using Databricks Workflows for ETL Pipelines
* Scheduling Jobs and Managing Dependencies
* Integration with Orchestration Tools (Airflow, ADF)

**Module 6: Security and Compliance**

* Unity Catalog for Access Control
* Managing Permissions in Databricks
* Ensuring Data Security and Compliance

**Module 7: Optimization and Monitoring**

* Performance Optimization Techniques
* Monitoring Jobs and Diagnosing Failures
* Cost Management for Databricks Clusters

**Module 8: Hands-On Projects**

* Building Data Pipelines for Data Warehousing
* Real-Time Analytics with Structured Streaming
* Implementing a Unified Batch and Stream Architecture