

Azure Data Engineering Training Curriculum

Professional Training Program (2 Months)

Course Overview

Duration: 2 Months

Mode: Online / Offline

Prerequisites: Basic Computer Knowledge

Training Type: Practical Hands-On with Real-Time Projects

Tools Covered: Azure, ADF, ADLS Gen2, SQL Server, Python, Spark, Databricks, Delta Lake, Synapse Analytics, Microsoft Fabric

Projects Included: 5 End-to-End Industry Projects

Interview Preparation: SQL, Spark, Databricks, Synapse, Fabric, Scenario-Based Questions

Module 1: Data Engineering Fundamentals

- What is Data Engineering?
- Roles and Responsibilities of a Data Engineer
- Modern Data Platform Architecture
- Batch Processing vs Streaming Processing
- Data Lake vs Data Warehouse
- Azure Data Engineering Overview
- Assignment: Design a Modern Data Platform Architecture

Module 2: Azure Fundamentals

- Create Azure Free Account
- Overview of Azure Portal
- Resource Groups
- Azure Storage Account
- ADLS Gen2 Overview
- RBAC and IAM Concepts
- Access Control Management
- Hands-On Lab

Module 3: Azure Data Factory (ADF)

- ADF Architecture
- Integration Runtime (IR)
- Linked Services and Datasets
- Pipelines and Copy Activity
- SQL Server Integration
- SQL Server to Parquet Conversion
- Parameters, Variables and Expressions
- Mapping Data Flow
- Azure Key Vault Integration
- Monitoring and Error Handling
- Industry Project: End-to-End Ingestion Framework

Module 4: SQL Foundation

- SELECT Statements
- Filtering and Sorting

- Joins and Group By
- Window Functions
- CTE
- Real-Time SQL Assignments

Module 5: Python Foundation

- Variables and Data Types
- Operators and Conditions
- Loops and Functions
- File Handling
- Pandas Basics
- Data Manipulation

Module 6: Apache Spark

- Spark Architecture
- DAG, Job, Stage and Task
- RDD, DataFrame and Dataset
- Narrow and Wide Transformations
- Spark Actions
- Partitioning, Cache and Persist
- Broadcast Join and AQE
- Spark SQL

Module 7: Azure Databricks

- Workspace and Cluster Creation
- Mount ADLS using Access Key, SAS and Service Principal
- Create DataFrames
- Read CSV, JSON and Parquet Files
- Azure Key Vault Integration
- Industry Project: Bronze-Silver-Gold Lakehouse

Module 8: Delta Lake

- Delta Tables and ACID Transactions
- MERGE, UPDATE and DELETE
- Time Travel
- OPTIMIZE and Z-Ordering
- Liquid Clustering
- Change Data Feed (CDF)
- Industry Project: Incremental Data Processing Framework

Module 9: Azure Synapse Analytics

- Synapse Studio and Spark Pools
- Dedicated SQL Pool
- Fact and Dimension Tables
- Star and Snowflake Schema
- Distribution and Clustering
- SCD Type 1 and Type 2
- Monitoring
- Industry Project: Retail Data Warehouse

Module 10: Microsoft Fabric

- Fabric Account Creation
- OneLake Architecture
- Lakehouse Concepts
- Gateway and Copy Activity
- Data Pipelines
- Fabric Notebooks
- Industry Project: End-to-End Fabric Data Engineering Solution

Interview Preparation

- SQL Interview Questions
- Spark Interview Questions
- Databricks Interview Questions
- Synapse Interview Questions
- Microsoft Fabric Interview Questions
- Resume Building and Mock Interviews