Learn Azure Data Engineering

Trainer Name: Abhishek Agrawal
Exp: 8+ years Azure Data Engineering and Big Data
Location: Dusseldorf, Germany (Currently working on ALDI SUD)

Course Description : This comprehensive training is designed to enhance your skills in both Azure Data Engineering, helping you advance your career. Throughout the course, you'll work on 3 real-world projects, gaining hands-on experience with Azure Data Engineering. We start from the basics, so no prior knowledge is required. In addition to mastering these technologies, you'll receive guidance on completing projects, preparing for interviews, building a strong resume, and certification support for exams like the Azure Fundamentals, Azure Data Fundamentals, Azure Data Engineer Associate (DP-203).

Azure Data Engineering

- Day to Day Explanation on **Azure** for each topic on live session based on
- the Business Requirements.
- Working with Azure Data Factory, Azure Databricks, Azure Synapse
- Analytics, Azure SQL Server, Azure DevOps a.
- Hands On Exercises of Basic to Advanced
- **Doubts will clear on live sessions**.
- **Interview questions** on each topic will be covered on daily basis and
- provided to the candidate
- Daily Practice at Home
- **Complete End to End** Practical sessions.
- **3 Realtime Project with Explanation** using Azure account
- Realtime Project Document Explanation
- **Resume building**
- **Certification and Interview Preparation**
- **Mock Interviews**

Course Duration: 1.5 - 2 Months (60 mins of Live sessions + 30 Mins Doubts clearance)

Software Installation: MySQL, MS SQL server

Trial Accounts: Azure

Take Away: Daily notes with Sample Data Model and sample case study

Recordings : **Life-time recording access**

Class Delivery: On-Line (Interactive Web Based LIVE Sessions)

Azure Data Engineering Curriculum

Table of Contents:

Azure Basics

- Basics of Cloud
- Public, Private and Hybrid Cloud
- Iaas, Saas, and Paas
- Azure Entra
- Tenant
- Subscription
- Management Group
- Resource Group
- Azure Portal Overview

Azure Storage Fundamentals

- Azure Storage Account
- Azure Blob Storage
- Azure Data Lake
- ACL and Permissions

Azure Data Factory Components

- Azure Data Factory Overview
- Integration Run Time
- Linked Service
- Pipeline, Dataset and Activity
- Source and Sink Configuration
- Copy Data Activity
- Recursive, Wildcard, File Listing, Mapping, User Properties.
- Parameterization of Copy Data Activity
- Sequential and Parallel Copy

Data Loading in Azure SQL Server

- Upsert, Pre-Copy Script, and AutoCreate Table Options.
- Handling Multiple Files with Lookup
- Full Load Pipeline
- New Watermark and Old watermark Concept
- Delta Laod for Single File.
- Delta Load for Multiple File

Pipeline Monitoring

- Monitoring of Activity
- Monitoring of Pipeline
- Execute Pipeline Activity

Azure DevOps and Git

- Azure DevOps Overview
- Azure DevOps Integration with ADF
- Azure Data Factory Integration with Git
- Git Pull, Git Push
- Cherry Pick, Git Revert.

Azure Key Vault and Security

- Azure Key Vault
- Managing Keys and Security
- Azure Key Vault integration with ADF
- Azure Key Vault Integration with Linked Service

Data Modelling and Design

- Fact and Dimension Tables.
- Star and Snowflake Schemas
- SCD Types and Implementation
- Creating Stored Procedures for Data Modelling
- Trigger Stored Procedure from ADF

Azure Logic Apps and Notifications

- Azure Logic App
- Sending Email Alerts using Logic App

Project 1 - Migrating Data from MS SQL Server to Cloud

- Full Load and Delta Load with Monitoring and Warehousing

Azure Data Factory Advance

- Event-Based, Scheduled, and Tumbling Window Triggers
- Data flow and Data Transformation Activity
- If-Else, Metadata, and Web Activities
- Global Parameter
- Parameterizing Triggers

Azure Databricks

- Overview of Databricks
- Spark and Spark Architecture
- Data Lake and Delta Lake
- Delta Table and Features
- Processing CSV, JSON AND XML file using PySpark.
- Integration of Azure Key Vault with Databricks.
- Secret Scope and JDBC connector
- Writing Data to Azure SQL server.

Project 2 - KPI Dashboarding with Azure Databricks

- Data processing using Pyspark and creating Delta Lake

Azure Synapse Analytics

- Overview of Azure Synapse
- Difference between Synapse and Data Factory
- Synapse Dedicated Pool
- Synapse Serverless Pool
- Polybase Copy
- Synapse Spark Pool

Project 3 - Dashboarding using Azure Synapse

- orchestrated the secure migration of data from on-premise servers to Azure Data Lake using Synapse Analytics pipelines

Microsoft Fabric

- Microsoft Fabric Introduction and Signup process.
- OneLake and Lakehouse in Fabric.
- Synapse Fabric Data Warehouse and Engineering
- Spark in Fabric and Data Processing.
- Real time Analytics in Fabric.
- Microsoft Fabric with Power BI.

By the end of this course, you will be able to:

- Migrate data from on-premise systems to Azure using Data Factory and Synapse.
- Create full load and delta load pipelines with watermarking.
- Schedule and automate pipelines using triggers.
- Create end-to-end data pipelines in Azure Synapse.
- Monitor and track pipeline performance in ADF.
- Secure data pipelines using Azure Key Vault.
- Manage source control with Azure DevOps and Git.

Certification:

- Microsoft Certified: Azure Fundamentals
- Microsoft Certified: Azure Data Fundamentals
- Microsoft Certified: Azure Data Engineer Associate