VECTOR E

AZURE DATA ENGINEERING TRAINING

AZURE DATABRICKS,AZURE DATA FACTORY, SQL DATA WAREHOUSE, ADLS GEN2, & AZURE SQL DB CURRICULUM

WHY TO CHOOSE AZURE DATA ENGINEERING?

- High Demand
- High paying jobs
- Best Career in 2024
- Place in top companies like MANG or Fortune 100

WHY TO CHOOSE VECTOR E?

- Best curriculum in the market
- resoanable & affordable price
- course offered by industry experts
- 24/7 support

Azure Data Factory

Introduction to Azure Data Factory

• Overview of Azure Data Factory (ADF)

• Key features and benefits

• Use cases and scenarios

• Hands-on Lab: Setting up your Azure Data Factory environment

Different Kinds of Integration Runtimes

• Definition and purpose of Integration Runtimes (IR)

• Types: Azure, Self-hosted, and Azure-SSIS Integration Runtime

• When and how to use each type

• Hands-on Lab: Setting up different types of Integration Runtimes

Linked Services

• What are Linked Services?

• How to create and manage Linked Services

• Examples of commonly used Linked Services

• Best practices for securing Linked Services

• Hands-on Lab: Creating and managing Linked Services

Data Sets

• Definition and role of Data Sets in ADF

• How to create and configure Data Sets

• Examples and best practices

• Hands-on Lab: Creating and managing Data Sets

Pipelines

• Introduction to Pipelines

• Components of a Pipeline

• How to create and manage Pipelines

• Best practices for designing efficient Pipelines

• Hands-on Lab: Building your first Pipeline

Parameters vs Variables

• Differences between Parameters and Variables

• How and when to use Parameters and Variables

• Practical examples

• Hands-on Lab: Using Parameters and Variables in a Pipeline

Copy Data

- Overview of the Copy Data tool

- Step-by-step guide to copying data

- Common scenarios and troubleshooting

- Hands-on Lab: Copying data between different data stores

Data Flow Activity

- Introduction to Data Flow in ADF

- Building and managing Data Flows

- Examples of Data Flow transformations

- Best practices for optimizing Data Flows

- Hands-on Lab: Creating a Data Flow with transformations

Configure Azure Integration Runtime

- Setting up Azure Integration Runtime

- Configuration options and best practices

- Common issues and resolutions

- Hands-on Lab: Configuring Azure Integration Runtime

Configure SSIS Integration Runtime

• Overview of SSIS IR

• Steps to configure and manage SSIS IR

• Migration tips and tricks

• Hands-on Lab: Configuring SSIS Integration Runtime

System Assigned and User Assigned Managed Identities

• Explanation of Managed Identities

• Differences between System Assigned and User Assigned

Managed Identities

• Implementation and use cases

• Hands-on Lab: Implementing Managed Identities in ADF

Lookup and Metadata Activity

• Introduction to Lookup and Metadata Activity

• How to configure and use these activities

• Practical examples

• Hands-on Lab: Using Lookup and Metadata Activity in Pipelines

For Each Activity

• Understanding For Each Activity

• Configuration and use cases

• Best practices and examples

• Hands-on Lab: Implementing For Each Activity in a Pipeline

Fail Over, Validate & more Activities

• Strategies for Failover in ADF

• Validation mechanisms and best practices

• Examples and common pitfalls

• Hands-on Lab: Implementing failover and validation in Pipelines

IF, Switch, Until

• Conditional activities in ADF: IF, Switch, Until

• How to configure and use each activity

• Practical scenarios and examples

• Hands-on Lab: Using IF, Switch, and Until activities in Pipelines

Set and Append Variable

• Using Set and Append Variable activities

• Configuration and best practices

• Examples and use cases

• Hands-on Lab: Setting and appending variables in a Pipeline

Stored Procedure

• Executing Stored Procedures in ADF

• Configuration and usage

• Best practices and troubleshooting

• Hands-on Lab: Executing a Stored Procedure in a Pipeline

Different Kinds of Triggers

• Overview of Triggers in ADF

• Types: Schedule, Tumbling Window, Event-based

• How to create and manage Triggers

• Best practices for trigger management

• Hands-on Lab: Creating and managing different types of Triggers

Monitoring and Managing Pipelines

• Tools for monitoring Pipelines

• Debugging and troubleshooting techniques

• Hands-on Lab: Monitoring and managing Pipeline runs

Advanced Topics

• Incremental data loading

• Handling schema changes

• Integration with other Azure services (e.g., Azure Synapse, Databricks)

• Hands-on Lab: Implementing an end-to-end data integration solution

Power Query Activity

• Introduction to Power Query in ADF

• Building and managing Power Query transformations

• Practical examples and use cases

• Best practices for using Power Query

• Hands-on Lab: Implementing Power Query transformations in a Pipeline

AZURE SYANAPSE ANALYSTICS(SQL DATA WAREHOUSE)

Data Warehouse Crash course

- What is Data Warehouse?
- Why we need Data Warehouse?
- Ideal Data Warehouse Solution
- SQL Server (OLTP) vs Data Warehouse (OLAP)
- Basics of Dimensional Modeling
- Facts and Fact table
- Dimensions and Dimension table
- Star vs Snowflake vs Galaxy Schema

Data warehouse Internals and Architecture
- Azure Synapse MPP Architecture
- Data Distribution and Distribution Keys

- Partitioning
- Best practices for Fact and Dimensional tables

Data Migration

- Create Dedicated SQL Pool
- Create External Tables & Tables
- Data Loading with Copy Data
- Data Loading with Polybase

Azure Key Vault

- Introduction to Azure key vault
- Why Key vault?
- How to create Secrete for SQL Server connection string?
- How to create Key vault linked service in ADF?

Apache Spark

Spark architecture

spark drivers and executers

Direct Acyclic Graphs(DAG)

Spark Jobs, stages, tasks

Spark Operations

Spark Transformations(Narrow & wide)

Spark Shuffling(sort-merge/hash)

Data Skewness

Repartition

Different File formats

Catalyst Optimizer

Pyspark Window Functions

Pyspark Joins, With column, concat, concat_ws, lit, coalesce etc.,

Broadcasting

Try catch and alert mechanisms

Spark Optimization

Logic Apps

- Create Logic Apps
- First Logic App in action
- Create Project
- Send Email Notifications using Logic Apps from ADF

Azure Databricks

Introduction to Databricks

- Creating Azure Databricks Service
- Azure Databricks Architecture Overview
- Databricks User Interface Overview

Databricks Cluster

- What is Databricks cluster?
- Cluster Types and Configure Databricks cluster

Databricks Notebooks

- Notebooks usage
- Magic commands
- Databricks Utilities

Read & Write data

- Accessing ADLS Gen2 account keys, SAS Token, Service Principal
- Securing Access to ADLS using Key Vault
- Databricks File System(DBFS)
- Databricks Mount Overview
- Data Ingestion - CSV
- Data Ingestion - JSON
- Data Ingestion - Multiple Files

Data Processing

- Actions,
- Wide & narrow transformation
- Lazy & Eager evaluations
- Optimization in Azure Databricks

Data Transformations

- Dataframes column-level transformations, such as Joins, sorts, filters Aggregations.
- Advanced DataFrame functions operations to manipulate data
- Perform date and time operations

ADF Integration:

- Create Pipeline - Transformation Notebooks
- Debugging a Pipeline
- Create ADF Trigger

Delta Lake

- Describe Azure Databricks Delta Lake architecture
- How to use Delta Lake?
- Create, Append, & upsert data

Unity Catalog

- Overview
- Unity Catalog Set up
- Cluster Configurations for Unity Catalog
- Unity Catalog Object Model Overview

CICD (Azure DevOps)

- Azure DevOps Repos
- Building Release pipelines

Real-world Case Studies

• Case study 1: Data migration from on-premises to cloud

• Case study 2: Data Loading to data warehouse with ADF

• Case study 3: Integrating ADF with other Azure services

• Case study 4 Implement SCD Type 1 using ADF with Error handling,
Status integration, Data Quality checks for banking or Finance Data.

Best Practices

• Design best practices for ADF and Databricks

• Security best practices

• Performance optimization tips

• Troubleshooting common issues

Additional Resources

• Documentation and tutorials

• Blogs and articles

• Community forums and support

Q&A Sessions

• Regular Q&A sessions to address participant questions

• Dedicated time for troubleshooting and discussion

Certification Preparation

• Overview of relevant Azure certifications

• Study tips and resources

• Practice exams and sample questions

This comprehensive curriculum covers all essential aspects of Azure Data Factory, including advanced topics and hands-on practice.

Contact: 8096677247;   email: info@vectoretech.com