TRENDY TECH
UPLIFT YOUR CAREER!

Azure
Data Engineering
Masters Program

By Sumit Mittal

Azure Data Engineering Curriculum

Azure Data Engineering Fundamentals

Cloud and On-Premise

Characteristics of Cloud

IAAS, PAAS, SAAS

Cloud Deployment Models- Public, Private, Hybrid

Azure Microsoft Services

Resource, Resource Group, Subscription

Data Center

Azure Regions

Azure Availability Zones

Zonal Services & Zone-Redundant Services

Handling Datacenter Failures

Region Pair

Virtual Machines

When to use VM

Virtual Machine Scale Set(VMSS)

Create and Manage Multiple VMs
Load Balancer
Availability Set Vs Availability Zone
Fault Domain & Update Domain

Azure Data Storage

Storage Account - Blob, Table, File, Queue services
Access Tiers -Data Accessibility
Locally Redundant Storage(LRS)
Zone Redundant Storage(ZRS)
Geo Redundant Storage(GRS)
Read Access Geo Redundant Storage
Geo-Zone Redundant Storage
Read Access Geo-Zone Redundant Storage
Introduction to DataLake
Azure DataLake Storage Gen2 (ADLS gen2)
Azure Storage Account Features
Access Control List (ACL)
Access Tiers - Hot, Cold/Cool, Archive

Azure Data Factory

Overview of Different Azure Services

Complete End to End Datapipeline – Usecase

Data Ingestion

Azure Data Factory Introduction

Data Transfer (Source to Sink)

Data Transformation – Data Flow

Workflow Orchestration

Data Transfer from RDBMS to ADLS Gen2

Azure SQL Databses

Data Transfer from Azure SQL to ADLS Gen2

Author, Monitor & Manage

Data Integration Service (ADF)

Usecases where ADF can be used

Data Ingestion

Data Transformation

Data Orchestration

Data Flow Mapping

Data transfer from external URL to ADLS – Usecase

Linked Services for Source and Sink

Select Transformation

Practice Assignments

ADF Primary Usage

Tansfer data from Blob to Datalake – Usecase

Blob Connector

Http Connector

Datalake Instance

Data Factory Instance

Linked Service Creation – Blob & Datalake

Dataset for Blob and Datalake

Complete Pipeline setup

Pipeline Parameterization

Key Vault

Scheduled Triggers

Tumbling Window Triggers

Storage & Custom Events

Trigger Pipeline on Custom Event – Usecase

Data Ingestion from 2 Sources (Blob & Amazon S3) to ADLS Gen2

Azure Synapse Analytics

Need for Datawarehouse

Best tool for Data Analysis

What is Azure Synapse Analytics

Synapse Usecase

Ingestion – Synapse Pipeline & Mapping Dataflow

Computation – SQL pools & Mapping Dataflow

Dedicated SQL Pool

Serverless SQL Pool

Apache Spark Pool

External Table & Normal Table

How to Create External Tables

Usecase for Serverless SQL Pool

OPENROWSET & External Table

2 ways to query data in ADLS Gen2

OPENROWSET & External Table

Data Lakehouse Architecture

Limitations of Datawarehouse

Limitations of Datalake

Objectives of Data Lakehouse

When to use Dedicated SQL Pool

Datawarehouse Architecture

Datawarehousing Units (DWU)

Fact and Dimension Tables

Table Distributions – Round Robin, Hash, Replicate

2 Options to Load Data from Datalake to Dedicated SQL Pool

Polybase

Copy Command

Distribution Types in detail

Processing the data in dedicated SQL Pool with Spark

Processing the spark table through serverless SQL Pool

Azure Synapse Summary

Serverless SQL Pool – Openrowset, external tables

Dedicated SQL Pool – Control, Compute, Distribution, Polybase, Copy

Spark SQL Pool – Spark Tables, Dedicated SQL

Azure Databricks

What is Databricks

Why Databricks

Databricks Pricing – Infrastructure and Software Charges

Different Cloud Providers offering Databricks

Databricks Features

3 ways to create cluster

All Purpose Cluster

Job Cluster

Cluster Pool

Cluster Modes – Single Node, Standard, High Concurrency

When to use the Different Cluster Modes

Databricks Benefits

Different optimized Cluster types

Memory Optimized, Storage Optimized, Compute Optimized, General Purpose, GPU Accelerated

Databricks File System (DBFS)

Databricks Architecture – Control and Data Plane

Databricks Community Edition

DBFS in detail

Object Store – Blob, Datalake Gen2

Filesystem utility– dbutils

Data Utility

Notebook Utility

Widgets Utility

Parameter passing from one Notebook to another

Mount Point

How to create Mount Point

Databricks Workspace

Databricks CLI

Ways to access Storage Account

Access Key/ Account Key

SAS Key

Service Principal

Secret Scope

Azure Key vault Backed Secret Scope

Databricks Backed Secret Scope

Azure Delta Lake

What is Datalake

Advantages and Challenges of Datalake

Example Usecases

Why Delta Lake

Updates and Deletes in Delta Lake

Azure Portal Setup

Cluster creation in Databricks

Delta Table Creation

Inserting values into Delta Table - Insert, Append, Copy Commands

Schema Evolution

Azure Lakehouse Architecture

Datalake Benefits and Challenges

2-tier modern Datawarehouse Architecture

Challenges of 2-tier Architecture

Lakehouse Architecture

Azure Delta Engine

Delta Cache

2 ways to enable Delta Cache

Optimizations

Data Skipping using Stats

Delta Cache

Small File Problem

Compaction/Bin-Packing

Optimize Command

Z-Ordering

Partitioning Bucketing

Vacuum Command

Auto Optimization and Auto Compaction

Optimized Writes

Photon Query Engine

Azure Delta Architecture

Medallian Architecture

Change Data Feed

(Additional Modules that will be developed shortly)

Azure HDInsight

Azure Cosmos DB

Azure Event Hub

Azure Logic App

2 End to End Real-Time Projects

(Sumit Sir Will be open to add more topics as per the Industry relevance)

Practice Environment

-Create your own Azure Account and you could use the Free Credits available for a month. After which, you will be charged a nominal amount for the rest of your practice.
-First few modules like Fundamentals, Data Storage and Data Factory will not charge you too much and most part of Databricks can be practiced on community edition which is free!
-Synapse Service has to be used cautiously as it is a slightly expensive service. (If used in the right way as explained by Sumit Sir in the videos, you will not be charged more than 20$ for the entire course practice)

Note : The course access is valid for 1.5 years from the first course release date.

TRENDY TECH
UPLIFT YOUR CAREER!

Thank You

Contact : hello@trendytech.in

https://trendytech.in/