Azure Data Engineering with Data Factory and Databricks

Day01: Azure Data Factory and Activity

Overview of Different Types Azure Storage -- Overview

ADLS Gen2

Azure SQL Database | Azure SQL Server

Key Component of Azure Data Factory

Copy Data Flow Activity in ADF

How to Ingest Data from On-prem SQL Server To Azure Data Factory Will show Demo but Participants may not be testing due to time constraints

Control Flow Activity

Get Metadata Activity

Filter Activity

If Activity

Append Activity – High Level Overview

Wait Activity – High Level Overview

Lookup Activity

Dataflow Transformation

Source Transformation

Sink Transformation

Day02: Azure Data Factory and Transformation

Derived Column Transformation

Select Transformation

Join Transformation

Pivot Transformation

Execute Pipeline

Store Procedure Activity

File format Conversion using copy activity (csv to parquet) -- High Level Overview

Pipeline Parameters

Email Notification Activity – High Level Overview

Event Based Trigger

Scheduling Pipeline Through Scheduling Window

CI/CD using GitHub /DevOps

Day03: Azure Databricks and Notebook

Overview of Apache Spark and Databricks

Databricks Architecture

Azure implementation

Getting started with Databricks

Creating and configuring your Cluster

Uploading data

Creating a Table

Connecting to a Spark data source

Previewing your Table

Columns and Datatypes basics

View aggregates

Perform Joins

Visualisations & Data Frames

Plots

- Choosing Chart types

- Chart Toolbar

- Layout and styling considerations

Day04: Azure Databricks Jobs

Databricks Jobs

Creating a Job

View Jobs and Job details

Running your first Job

Scheduling Jobs

Setting Parameters

Viewing completed jobs

Managing Dependencies

Setting up Alerts

Delta Lake and Delta Tables

Getting data into Delta Lake

Delete, update, merge

Constraints

Versioning

Concurrency – High Level Overview

Delta Commands

Day05: Integrating Azure Data Factory and Databricks

Describe how Azure Databricks notebooks can be run in a pipeline.

Create an Azure Data Factory linked service for Azure Databricks.

Use a Notebook activity in a pipeline.

Pass parameters to a notebook

Create Spark catalogue tables for Delta Lake data. – High Level Overview

Use Delta Lake tables for streaming data – High Level Overview

Azure Databrick connection With PowerBI -- Overview

Devops/Github