CONTENTS

AZURE DATA FACTORY

Chapter 3: Overview of Different Azure Services & Introduction to Azure Storage:

- *Different Azure Services*
- *What is Microsoft Azure Subscription and different types of Subscriptions?*
- *Azure free Account Creation and Portal Walk-Through*
- *What are Resource Groups & Resources? and their creation:*
- *What is Azure storage service and different Azure Storage data services?*
- *Benefits of Azure Storage|Azure Storage Account|Types of Storage Accounts*
- *Introduction to Blob Storage Service*
  - *Block Blobs*
  - *Append Blobs*
  - *Page Blobs*
- *Create an Azure Storage Account*
- *Azure Storage Performance Tiers*
  - *Standard*
  - *Premium Performance*
- *What is a Region/Azure region?*
- *Understanding Data Redundancy/Replication*
  - *Data Redundancy and Azure Redundancy storages*
  - *Types of Redundant Storages*
    - *LRS (Locally Redundant Storage)*
    - *ZRS (Zone Redundant Storage)*
    - *GRS (Geo Redundant Storage)*
    - *GZRS (Geo-Zone Redundant Storage)*

Chapter 4: Azure Data Lake Storage & Blob Storage:

- *What is Data Lake?*
- *What is Azure Data Lake?*
- *Azure Data Lake Storage (ADLS)*
  - *ADLSGen1*
  - *ADLSGen2*
- *Differences Between ADLSGen1 & ADLSGen2 Storages*

- *Blob Storage Vs Data Lake Storage*
- *Creation of Blob and ADLSGen2 Storage Account*
- *Azure Storage-Access Tiers*
  - *Hot*
  - *Cold*
  - *Archive*

Chapter 5: Introduction to Azure Data Factory (ADF) & Data Movement with Copy Activity:

- *What is Azure Data Factory (ADF)?*
- *Azure Data Factory Top level concepts :*
  - *Pipelines*
  - *Activities*
  - *Datasets*
  - *Linked Services*
  - *Triggers*
  - *Integration Runtimes*
  - *Data Flows*
- *Azure Data Factory Service creation & Walk-Through*
- *Integration of Azure Storage Account with ADF*
- *Creation of Pipeline with Copy activity : Linked Services|Datasets*
- *Copy files between different storage accounts & Processing Multiple Files (diff: format files) from Source to Sink*
- *Debugging Pipeline and Monitoring*

Chapter 6: Automation of Pipeline run Copy Activity| Behaviour|Performance Tuning in ADF :

- *What is a Trigger in ADF?*
- *Different Types of Triggers in ADF*
  - *Schedule trigger*
  - *Tumbling window trigger*
  - *Event-based trigger*
    - *Storage event trigger*
    - *Custom event trigger*

- *Creation of a Trigger/Publishing Trigger/Monitoring Trigger Runs*
- *Deactivation and Deletion of Triggers*
- *Differences between Schedule and Tumbling window trigger?*
- ***When to Opt for What Trigger?***

- *Copy Behaviour*
  - *Preserve Hierarchy*
  - *Flatten Hierarchy*
  - *Merge Files*
- *Copy Activity Performance Tuning*
  - *Data Integration Units (DIUs)*
  - *Degree of Copy Parallelism*

Chapter 07: ***Activities, Parameters and Variables in ADF :***

- *Get Metadata*
- *Filter*
- *For Each*
- *If Condition*
- *Delete*
- *Until*
- *Wait*
- *Validate*
- *Execute Pipeline*
- *Set Variable*
- *Append Variable*
- *Lookup*
- *Switch Activity*
- *Web Activity*

- *What is a variable and parameter in ADF?*
- *Difference between a Variable and a Parameter*
- *Different types of Parameters in ADF*
  - *Dataset Parameters*
  - *Linked Service Parameters*
  - *Pipeline Parameters*
  - *Global Parameters*

Chapter 08: Introduction Azure SQL Database :

- *What is Azure SQL Database*
- *How to create Azure SQL Server*
- *Creation of Database in Azure SQL Server*
- *Different kinds of Pricing Tiers in Azure SQL*
- *SQL Concepts*
  - *Data Definition Language (DDL: Create/Alter/Drop)*
  - *Data Manipulation Language (DML: Select/Insert/Update/Delete)*
  - *Joins*
  - *Constraints*
  - *Views*
  - *Stored Procedures*

Chapter 09: Integration of Azure SQL Database with ADF & Data Ware Housing (DWH) Concepts :

- *Creation of Azure SQL Database*
- *Copying Data from Blob Storage to Azure SQL DB and Vice-Versa*
- *Loading Data to Azure SQL DB from ADLSGen2 and Vice-Versa*
- *Loading Multiple Tables Data from Azure SQL DB to Azure SQL DB*
- *How to Load Data from On-premises SQL Server to Azure SQL DB using Self-Hosted Integration Runtime*
  (To Query SQL Server Data from On-premise SSMS and Cloud)
- *Incremental Load or Delta load from Azure SQL DB to Blob Storage in Azure Data factory*
  - *Watermark Tables*
  - *Lookup Activity*
  - *Stored Procedure Activity*
- *Pipeline Execution Details Logging to SQL Table (Audit Logging)*
- *Integration of SnowFlake with ADF (Loading data from Azure SQLDB to Snowflake DW)*

Data Ware Housing (DWH) Concepts :

- *What is Data Warehouse/Data Warehousing?*
- *Why is the need for a Data Warehouse?*
- *OLTP (vs) OLAP*

- *Dimensional Modelling*
- *Fact and Dimensional Tables*
- *Types of Facts*
  - *Additive Facts*
  - *Semi Additive Facts*
  - *Non-Additive Facts*
- *Dimensional Modelling-Schemas*
  - *Star Schema*
  - *Snowflake Schema*
- *What are Slowly Changing Dimensions?*
- *Types of Slowly Changing Dimensions (SCD)*
  - *SCDType1 Dimension*
  - *SCDType2 Dimension*
  - *SCDType3 Dimension*

Chapter 10: Azure Key Vault and Azure Logic Apps :

- *What is Azure Key Vault?*
- *Create a Key Vault*
- *Create a Key Vault Secret*
- *Grant Access to the Key Vault*
- *Create a Key Vault ADF Linked Service*

- *Logic Apps usage*
- *Creation of Workflow using Logic Apps in Azure:*
- *Sending an email using Logic app in Azure Data Factory(Web Activity)*
- *Triggering Logic Apps*

Chapter 11: Azure Data Flows :

- *What are data flows in ADF?*
- *Azure Data Flows Introduction|Setup Integration Runtime for Data Flows*
- *Build a Data Flow|Enable Data Flow Debugging| Data Flow Execution*
- *Data Flow Transformations*
  - *Filter Transformation*
  - *Sort Transformation*
  - *Select Transformation*
  - *Derived column Transformation*

- *Conditional split Transformation*
- *Rank Transformation*
- *Surrogate key Transformation*
- *Exists Transformation*
- *Union Transformation*
- *Join Transformation*
- *Lookup Transformation*
- *Aggregate Transformation*
- *Alter row Transformation*
- *Flatten Transformation*
- *Pivot Transformation*
- *Unpivot Transformation*
- *Source Transformation*
- *Sink Transformation*

▪ *Slowly Changing Dimension Type2 Implementation*
▪ *Data Quality Checks*

Chapter 12: ***Azure DevOps - CI/CD :***

▪ *Azure DevOps Introduction*
▪ *What is Azure Git Repository?*
▪ *Creation of the Git Repository in Azure Repos*
▪ *How to connect to a Git Repository |Linking the Azure Data Factory to the Git Repository*
▪ *What is Feature Branch and creation & Workflow|Pull Request Creation*
▪ *What are ARM Templates?*
▪ *Deployment using ARM Templates*
▪ *Deployment of Azure Data Factory Pipelines in Dev/Test/Prod Environments*

AZURE DATABRICKS

Chapter 01: Azure Databricks Overview :

- Big Data Introduction
  - What is Big Data & Types of Big Data
  - Sources of Big Data
  - Characteristics of Big Data (5 V's)
  - What is Big Data Analytics and Advantages of it ?
  - Implementation of Big Data Analytics?
- Azure Databricks Introduction
- Creating Azure Databricks Service
- Azure Databricks Architecture

Chapter 02: Computation Management :

- Databricks Clusters
  - Azure Databricks Cluster Types and Cluster Configuration
  - Creating Azure Databricks Cluster
  - Azure Databricks Pricing and Azure Cost Control
  - Azure Databricks Cluster Pool and Cluster Policy (Auto scaling)
- Jobs and Workload
  - Creating jobs from notebooks and assigning types of clusters for jobs
  - Monitoring jobs and managing loads:

Chapter 03: Data Management & Security :

- Databricks Notebooks
  - Azure Databricks Notebooks Introduction
  - Magic Commands
- Databricks Utilities

- *Databricks File System - DBFS commands copy and manage files using DBFS*

- *Database - Creating database, tables and managing databases and tables*
- *Table - Creating Tables, dropping tables, loading data:*
- *Metastore - managing metadata and delta table creation, managing delta tables*
- *Creation of users and groups*
- *Managing access to users and groups*

Chapter 04: *Introduction to Spark,Python & Using Python Libraries in Azure Databricks :*

- *Python Introduction and Installation for Local development*
- *Data Types, Functions*
- *Python Modules and Packages*
- *File Operations, Errors and Exceptions, Regular Expressions*
- *Spark Cluster Architecture*
- *RDDs, Dataframe and Data Source API Overview*
- *Using Python Libraries in Azure Databricks*
    - *Installing Python libraries in Azure Databricks*

Chapter 05: *Reading and Writing Data from and to Various Azure Services and File Formats :*

- *Mounting ADLS Gen2 and Azure Blob storage to Azure DBFS*
- *Reading and writing data from and to Azure Blob storage*
- *Reading and writing data from and to ADLS Gen2*
- *Reading and writing data from and to an Azure SQL database using native connectors*
- *Intergration with Azure Synapse :*
  Reading and writing data from and to Azure Synapse SQL (dedicated SQL pool) using native connectors
- *Reading and writing data from and to Azure Cosmos DB*
- *Reading and writing data from and to CSV and Parquet*

- *Reading and writing data from and to JSON, including nested JSON*

Chapter 06: *Integrating with Azure Key Vault, App Configuration and Log Analytics :*

Creating an Azure Key Vault to store secrets using the UI

- *Creating an Azure Key Vault to store secrets using ARM templates*
- *Using Azure Key Vault secrets in Azure Databricks*
- *Creating an App Configuration resource*
- *Using App Configuration in an Azure Databricks notebook*
- *Creating a Log Analytics workspace*
- *Integrating a Log Analytics workspace with Azure Databricks*

Chapter 07: *Spark SQL*

- *Introduction and Spark SQL Architecture*
- *Databases, Managed Tables and External Tables*
- *Conversion of different sources to Tables*
- *Functions (Aggregation/Window/Ranking) and Joins*
- *Using query parameters and filters*
- *Introduction to visualizations in Databricks SQL*
- *Creating dashboards in Databricks SQL*
- *Connecting Power BI to Databricks SQL*

Chapter 08: *Understanding Spark Query Execution :*

- *Introduction to jobs, stages, and tasks*
- *Checking the execution details of all the executed Spark queries via the Spark UI*
- *Deep diving into schema inference*
- *Looking into the query execution plan*
- *How joins work in Spark*
- *Learning about input partitions and output partitions*
- *Learning about shuffle partitions*
- *Storage benefits of different file types*

Chapter 09: *Exploring Delta Lake in Azure Databricks :*

- *Pitfalls of Data Lakes and Data Lakehouse Architecture*
- *Exploring Delta Lake in Azure Databricks*
- *Delta table operations – create, read, and write and Delta table data format*
- *History, Time Travel, Vacuum and Delta Lake Transaction Log*
- *Handling concurrency and Delta table performance optimization*
- *Constraints in Delta tables and Versioning in Delta tables*

Chapter 10: *DevOps Integrations and Implementing CI/CD for Azure Databricks :*

- *DevOps Integrations and Implementing CI/CD for Azure Databricks*
- *How to integrate Azure DevOps with an Azure Databricks notebook*
- *Using Azure DevOps Git for Azure Databricks notebook version control*
- *Understanding the CI/CD process for Azure Databricks*
- *How to set up an Azure DevOps pipeline for deploying notebooks and Deploying notebooks to multiple environments*
- *Enabling CI/CD in an Azure DevOps build and release pipeline*
- *Deploying an Azure Databricks service using an Azure DevOps release pipeline:*

ADD-ONs:

- ***Session on Azure Synapse***
- ***Attendees Tasks Addressal***
- ***Interview Perspective Session***
- ***Mock Interviews***