Azure Data Engineer (ADE) Course Content

Introduction to Cloud Computing

- Advantages of Cloud Computing
- Types of Cloud Deployment Models
- Types of Cloud Services
- Characteristics of Cloud Computing
- Economies of Scale
- CAPEX vs OPEX

Azure Fundamentals

- History of Azure
- Azure Infrastructure
- Security Privacy and Trustworthy In Azure
- Azure Account & Subscription
- Azure Services Overview
- Creating Azure Account
- Navigating through Portal
- Creating Dashboards
- Resources, Resource groups and Resource manager (ARM Templates)

Implementing and Designing Storage Services in Azure

- Creation of Azure Storage account
- Types of Azure Storage
- Access Tiers
- Standard vs Premium Storage
- Storage Redundancy
- Azure Storage Explorer

Azure Blob Storage Service for Unstructured Data

- Working With Blob Storage
- Types of Blobs
- Blob Containers and Public Access Levels
- Creating Blobs and Exploring Blob Features
- Azure Blob Storage Life Cycle Management
- Data protection features for Blob Storage
- Hosting Static Website
- Securing Blob Storage

Implementing Data Storage solutions using Azure Data Lake

- Introduction to Azure Data Lake Storage
- Road to Azure Data Lake Gen 2
- Architecture and Features of ADLS Gen 2
- Data Ingestion from Multiple Sources
- Ingesting Data to ADLS Gen2 from AWS S3 Using Azure Data Factory
- Ingesting Data to ADLS Gen2 from Azure SQL Database Using Azure Data Factory
- Visualization Using Power BI using ADLS Sources
- Implementing Structured Storage Solutions
- Introduction to Azure SQL DB
- Provisioning Azure SQL Database
- Loading data into Azure SQL Database
- Controlling User Access
- Configuring Firewall Access
- Implementing Geo Replication
  - Implementing Backups using PITR and LITR
  - Migrating data from On-prem to Azure SQL Database
  - Connecting Applications to Azure SQL Database

Introduction to Cosmos DB

- What is No SQL Database
- Creating Cosmos DB Account
- Planning Capacity and Costs
- Tools and SDK
- Creating Cosmos Database Account
- Creating a Container
- Creating Documents
- Presenting Cosmos DB API's
- Multiple API's and Data Modules
- Describing Data Security Components
- Enabling Global Distribution
- Enabling Multi Master Databases

Data Ingesting Using Azure Data Factory

- Introduction to Azure Data Factory
- ETL vs ELT
- Traditional tools vs Azure Data factory
- Components of Azure Data factory
- Setting up Azure Data Factory Work space
- Overview of Azure Data Factory User Interface
  - Integration Run times
  - Linked Services
  - Data sets
- Pipelines
- Triggers

- Labs
- Copy Data Activity
- From Unstructured to Structured Storage
  - Blob Storage to Azure SQL database
  - ADLS to Azure SQL database
  - Onpremises to Azure SQL database
- Orchestrating Pipelines
- Monitoring the Pipeline Activities
- Debugging Pipelines
  - Parameterization in Azure Data Factory
- Variables
- ForEach Loops
- Lookups
- Labs:
- Copying Multiple files from container to Azure SQL Database
- Understanding Pricing

Implementing Mapping Data Flows

- Transformations using Mapping Data flows
- Source
- Sync
- Sort
- Filter
- Lookup
- Select
- Derived Column
- Conditional Split
- SSIS Packages using Azure Datafactory
- Azure SSIS Runtime
- Migrating SSIS Packages in ADF
- Running SSIS Packages in ADF

Introduction to Azure Data Bricks

- Introduction to Databricks
- Fundamentals of Azure Data bricks
- Apache Spark Clusters
- Creating an Azure Databricks Workspace
- Azure Databricks Clusters
- Azure Databricks Workspace
- Creating Azure Databricks Notebooks and Jobs
- Working with Storage Options
- Security with Azure Databricks
- Basics of ETL process
- Ingestion and Extracting data using Azure Data bricks
- Transforming Data in Azure Data bricks
- Loading Data in Azure Databricks
- Creating data frames using Pyspark
- Loading and reading data using Pyspark
- Data Exploration with Azure Databricks
- Visualization in PowerBI

Azure Synapse Analytics

- Introduction to Data Lakehouse and Synapse Analytics
- Setting up Synapse Work Space
- Working with Dedicated SQL Pools
- Transforming data with synapse Spark pools
- Configuring Spark pools
- Working with Notebooks
- Loading data in spark tables
- Implementing transformations
- Data Ingestion and Orchestration with Synapse pipelines
- Using Copy data activity
- Running and Monitoring Pipelines
- Working with Serverless SQL Pools

Introducing Azure Synapse Analytics

- Why Azure Synapse Analytics?
- Components of Synapse Analytics
  - Storage, ingestion, transformation and orchestration
  - Data storage in Synapse Analytics
  - Options for data transformation
  - Visual ingestion and orchestration with Pipelines

The Synapse Workspace

- Create a Synapse workspace
  - Basics
  - Security
  - Networking
  - Tags
  - Review + Create
- Explore Synapse Studio
  - Navigation Header and Sidebar
  - Data Hub
  - Develop Hub
  - Integrate Hub
  - Monitor Hub
  - Manage Hub

Data Ingestion and Orchestration with Synapse Pipelines

- Integration components
- Data ingestion
  - The Copy activity
  - Integration runtimes
  - The Copy Data tool
  - Orchestration

- Executing transformations
  - Dynamic activity properties
  - Controlling execution flow
  - Creating pipeline runs
- Sample dataset
  - HTTP linked service
  - Source dataset
  - Sink dataset
  - Copy pipelines

Batch Data Processing with Apache Spark

- Apache Spark
  - Apache Spark pools
  - Notebooks
  - Sessions
- Transforming data using notebooks
  - Data transformation with Python
  - Delta format
  - Lake databases
  - Data transformation with C#
  - Microsoft Spark Utilities
- Low Code Data Engineering
- Execution and orchestration
  - Notebook activity
  - Data flow activity