Course Curriculum for MS Azure + SQL + Azure Data Engineering

Introduction to Cloud Computing:

- Understanding different Cloud Models
- Advantages of Cloud Computing
- Different Cloud Services
- Different Cloud vendors in the market

Microsoft Azure Platform:

- Introduction to Azure
- Azure cloud computing features
- Azure Services for Data Engineering.
- Introduction of Azure Resources/Services with examples
- Azure management portal
- Advantage of Azure Cloud Computing
- Managing Azure resources with the Azure portal
- Overview of Azure Resource Manager
- Azure management services.

- What is Azure Resource Groups
- Configuration and management of Azure Resource groups for hosting Azure services

Introduction to Azure Resource Manager & Cloud Storage Services

- Completed walkthrough of the Azure Portal with all the features.
- What is Resource Groups and why we need RG's in Azure cloud computing platform to host resources??
- Different types of Storage Accounts provisioning in Cloud computing with different storage services
- (i)Container/Blob storage service,

- (ii)File share storage service,
- (iii)Table storage service &
- (iv)Queue storage service
- Details explanation & understanding of different Blob/container storage services…
- (i)Page Blob.
- (ii)Append Blob &
- (iii)Block Blob
- Creating and managing the data in container storage services with Public and Private accesses as per the need of a project.
- Implementation of Snapshots for Blob storage services and File share storage service
- Generating SAS for different storage services to make the storage content browseable across all the globe or Publicly.
- What is Standard Storage Account and Premium Storage account and which to use accordingly as per the real time scenarios.
- Detail explanation and implementation of Data Lake storage Gen2 Storage Account to store the unstructured data in cloud storage services.
- All the features/properties(Overview, activity log, Tags, Access control(IAM), Storage browser…etc) of Azure Storage Accounts.
- Maintenance and management of Storage keys and connection string for Azure Storage services.
- Implementing different levels of access(Reader, contributor, owners…etc) to the Azure Storage accounts

Migration of storage contents across Public & Private Clouds

- Moving the storage account with storage content across different Resources Groups based on real time scenarios.
- Migrating the data from On-prem(Private cloud) to Azure Storage account (Public cloud) using Az copy(forward migration).
- Migrating the data from public cloud to Private cloud(revers migration).

- Implementing the Az copy commands to migrate the data.
- (i)On-prem to Azure cloud storage services
- (ii)cloud storage services to On-prem
- (iii)Cloud to Cloud
- Moving the SA & its content from one Resource Group to another.

Replication of Storage Accounts Authentication & Authorization of Storage Accounts & Azure Storage Explorer

- Azure Storage explorer for creating, managing, and maintaining the Azure storage services data.
- Installation of Azure Storage Explorer and what is the purpose of this tool for Azure Storage accounts(its Purpose & benefits with real time scenarios)
- Generate Shared Access Signature(SAS) in Azure Storage Explorer(ASE) for security implementation of Storage account content.
- Managing of Access keys & connection strings of SA with Azure Storage Explorer
- Configuration of Authentication and Authorization for Storage Account via Azure Active Directory.

- Hosting File share Storage services to On prem servers or Cloud Servers as shared drive for File share servers.

Provisioning of SQL DB's in Private & Public cloud computing:

- Introduction to SQL DB's
- Creation of new SQL DB's & Sample SQL DB's both in On-prem and Cloud computing.

- Planning and deploying Azure SQL Database
- Implementing and managing Azure SQL Database
- Managing Azure SQL Database security
- Planning and deployment of SQL DB's in Azure cloud computing with real time scenarios.

- Different DB's Deployment options.
- Databases purchasing models.(VCore & DTU's)
- Visualization of cloud DB server, Database, and validation of data from on-prem(private cloud)
- Implementation of Firewall security rules on Azure DB servers to access and connect from on-prem SSMS.
- Creation of Database in on-premises and synch with azure cloud

SQL DB Migrations:

- Migrating SQL DB's from On-premises to Azure cloud computing using Microsoft Data migration assistant.
- Restoring SQL DB's from On-prem to cloud computing.
- Migration of Specific DB objects from on-prem to cloud based upon base upon project requirements.
- Implementation of RSV and scheduling the backups of SQL DB's and Azure Storage Account file share services on schedule, on demand based upon real time scenarios.

Introduction to SQL Server & SQL Queries from basics to Advance(till ADE Services):

- Introduction to SQL DB Queries
- Below SQL queries detail explanations, syntax & execution based upon real time scenarios.
  - Select queries.
  - Distinct queries
  - Where queries
  - And or not queries.
  - Order By queries
  - Insert into queries.
  - Null values queries
  - Update queries
  - Delete queries.

- Select Top queries.
- Min & Max queries
- Count, Avg, Sum queries.
- Like queries.
- Wildcards queries.
- In queries
- Between queries.
- Aliases queries.
- Joins(Inner join, Left join, Right join, Full join, Self-join...etc)
- Union queries.
- Group By queries.
- Having queries.
- Exists queries.
- Any All queries.
- Select into queries.
- Insert into select queries.
- Store procedures queries.

What is Azure Data Factory(ADF):

- Deep understanding and implementation of concepts/Components of ADF
  - Pipelines
  - Activities
  - Datasets
  - Linked Services
- Building blocks of Azure Data Factory
  - Triggers
  - Integration runtime
  - Dataflow
- Complete features and walk through of Azure Data factory studio.
- Different triggers and their implementation in ADF
  - Scheduled trigger
  - Tumbling window trigger

    - Event trigger
- What is integration run time and different types of integration run time in ADF.
    - Azure
    - Azure – SSIS
    - Self-hosted
- When to use ADF.
- Why to use ADF.
- Different types of ADF pipelines
    - Dynamic pipelines
    - Parameterized pipelines
    - Automated pipelines
- Pipelines in ADF
- Different types of Activities in ADF
    - (i)Data movement activities
    - (ii)Data transformation activities
    - (iii)Data control activities.
- Datasets in Azure Data factory
- Linked services in ADF.

Controls/Activities of Azure Data Factory(ADF) for copying the DATA across various sources to Azure IAAS & PAAS Services:

- Copying the data from Blb Storage account to ADL's Gen2 Storage account.
- Copying of zip files(.csv) from Blob SA to ADL's Gen2 SA using ADF
- Implementation and explanation of Metadata control in ADF to find the structure before copying the data.
- Implementation and explanation of Validation and If Condition
- Implementation of Get Metadata control, filter control & For Each Control or activities in ADF.
- Implementation & execution to copy the data from GitHub platform to Azure Storage services with variables and parameters.

- Implementation of Foreach control, copy data control and Set variable to dynamically load the data from source to target using ADF.
- Creating Dynamic pipelines with lookup activity to copy multiple .csv files data picking form Json format data in Azure Storage services.
- Copying the files from GitHub Dynamically with the use of Dynamic parameters allocation-AUTOMATION PROCESS:
- Copying the data from different files formats(.csv, .xlsx, .txt, .Parquet, .Json, .SQL…etc) using suitable ADF controls/activities.
- Implementation and execution of Loading the data from Blb SA to SQL DB single table & multiple tables using copy data activity, ForEach activity,
- Executing multiple pipelines in parallel with Execute pipeline activity.

Scheduling Triggers for automation of Dataflow/Datacopy to various sources and destinations in ADF:

- Implementation of Schedule based triggers for different ADF pipeline containing different activities.
- Implementation of Event based triggers for different ADF pipeline containing different activities.
- Implementation of Thumbling window-based triggers for different ADF pipeline containing different activities.
- Implementation and execution of storage and Event based triggers.

What is Azure Keyvault, purpose of using Keyvault, Storing the SA keys, connection string in Azure KV with Access policies:

- Detail explanation & implementation of Azure Keyvaults,
- Making the SQL DB connection string to store in Keyvault to enhance the security for SA content and SQL DB

- Generating the secrets inside the Azure keyvault and granting access by implementing the access policies for different users.

Integrating Azure Data Factory with GitHub Portal:

- Detail walk through of GitHub portal
- Creating an account, repo's, in GitHub portal
- Integrating Azure Data Factory with GitHub Portal as per project requirements.
- Placing, maintaining and executing the source code via GitHub portal for Azure Data Factory.
- Creating master branch, practice branches in GitHub portal to merge the newly created code via Pull Requests.
- Setting up the Repo for ADF pipelines and converting to live mode from GitHub portal covering with real time scenarios.

Data Flows Transformations in Azure Data Factory:

- Designing new Data flows
- Designing and implementing transformations like
- 1)Source transformation
- 2)Join transformations
- Inline Datasets in data flow source control
- Designing and implementing of Data flow with Source transformations, Filter transformations & Sink transformations in ADF with inline Datasets
- Implementation of Select transformations with Data flows for various source controls.
- Implementation of Dataflows using Aggregate & Sink transformation:
- Implementation of Dataflow with conditional split & Sink transformation with copy data activity:
- Implementation of Dataflow with Exists & Sink transformation:
- Implementation of Azure Dataflows for Derived column transformation with Source & Sink transformation:

- Implementation of Azure Dataflows to connect to SQL DB with Source & Sink transformation:
- Implementation of Azure Dataflows to connect to SQL DB with Source & Sink transformation.

Azure Data Bricks & Apache Spark:

- What is Apache Spark, details explanation and implementation of Apache Spark.
- Illustration and Elaboration of Apache Spark Architecture.
- Explanation of
- What are worker nodes and slaves nodes in Azure Data Bricks clusters
- Implementation of Azure Databricks cluster by considering different worker nodes and slave nodes.
- Different features and properties of Azure Data Bricks clusters

  - Single node
  - Multi node
  - Photon acceleration
  - Auto turn off Azure Data bricks cluster after a defined time.
  - Autoscaling of cluster
  - Configuration provisioning of Azure Data Bricks clusters

Azure Data Bricks & Apache Spark clusters features:

- Creating single node and multi nodes clusters
- Creation of Pyspark notebooks in Databricks cluster to fulfil different business requirements.
- Creation of folder hierarchies, notebooks in Azure Databricks workspace.
- Onboarding users, data files in Azure Databricks workspace
- Writing pyspark scripts to fetch the data from source system in Azure Databricks
- Mounting the Storage accounts with Azure Databricks to fetch the data from different source systems.

- Extracting the data from web portal by writing the pyspark scripts
- Connecting Azure Databricks to different API's to write the scripts in SQL & Pyspark scripting.
- Converting the python code to SQL scripts in Azure Databricks
- Onboarding source files in Azure Databricks workspace DBFS.
- Importing files, folders, extracting data from files in Azure

Azure Databricks Notebooks :

- Databricks Files System(DBFS):
- Importing raw data files into DBFS, reading and analysing the file data with Pyspark scripts:
- Mount points in Azure Databricks with Blob Storage & Data Lake Storage services.
- Installing Databricks CLI & configuring with Azure Databricks Workspace
- Installing python package in local laptop to connect with Azure Databricks workspace
- Generating Access token in Databricks workspace to integrate with python package.

File System Utilities:

- mkdirs
- ls
- cp
- Copying a File
- Copying a Folder
- mv
- Moving a file

- Moving a Folder
- rm • Removing a File
- Removing a Folder
- head
- put

Widgets utilities in Azure Databricks:

- Combobox
- Dropdown
- Multiselect
- Text
- Remove
- Removeall

Azure Synapse Analytics:

- What is Azure Synapse Analytics
- (i)What is Synapse workspace used for
- (iii)What is Synapse SQL
- (iv)Apache Spark for Synapse
- (v)How to design Pipelines in Azure Synapse
- Implementation of Linked Services/Datasets in Synapse Analytics:
- Implementation of dedicated SQL Pool inside Synapse Analytics
- Implementation of serverless SQL Pool inside Synapse Analytics
- Creation of Apache spark pool in Azure Synapse Analytics.
- Writing SQL Script in Azure Synapse analytics to get the result set in tabular and chart formats.
- Visualizing the data in Synapse analytics in variety of different charts (like pie charts, line charts, bar charts…. etc)
- Designing of Synapse Analytics pipelines by considering various activities as per the business requirements.

- Creation of Datasets, Linked services for Synapse Analytics pipelines.
- Data analysis with serverless spark pools in Azure Synapse Analytics
- What is Apache spark in azure synapse analytics.
- Designing and development of Apache spark pool in Azure synapse
- Creating Spark Databases and tables to load the data from source system and analysing the data in Synapse analytics.

Azure Stream Analytics:

- What is Azure Stream Analytics
- Purposes and usage of Stream Analytics in Azure cloud computing
- Benefits and advantages of stream analytics
- Architecture diagram of data flow in Azure stream analytics with other cloud services.
- Understanding & usage of browser-based Raspberry Pi simulator.
- Deployment of IoT Hub services as an input for Stream analytics jobs
- Implementation & execution of stream analytics jobs and designing inputs and outputs for IoT Hub and Datalake Gen2.
- Writing SQL scripts to generate live streaming data and loading it in destination.