



Technical Guftgu

(Established under Ministry of Micro, Small and Medium Enterprises, Govt. of India)

Contact- +91-9870663188

+91-8527556109

Website: www.technicalguftgu.in

- ✓ Certificate provided
- ✓ Recordings Provided
- ✓ Training In Hindi/Urdu
- ✓ Expert Trainers

Azure Data Engineer (DP – 203)

Course Curriculum –

Introduction - Microsoft Azure is a powerful cloud computing platform offering scalable solutions for data storage, computing, and analytics. SQL on Azure provides robust, cloud-based database services for managing and querying relational data efficiently. Azure Data Engineering empowers businesses to build, transform, and manage big data pipelines using tools like Azure Data Factory, Databricks, and Synapse Analytics. Together, they enable seamless data integration, advanced analytics, and scalable data solutions for modern enterprises.

Module 1: Introduction to Cloud Computing:

- I. Understanding different Cloud Models
- II. Advantages of Cloud Computing
- III. Different Cloud Services
- IV. Different Cloud vendors in the market

Module 2: Microsoft Azure Platform:

- I. Introduction to Azure
- II. Azure cloud computing features
- III. Azure Services for Data Engineering
- IV. Introduction of Azure Resources/Services with examples
- V. Azure management portal
- VI. Advantage of Azure Cloud Computing
- VII. Managing Azure resources with the Azure portal
- VIII. Overview of Azure Resource Manager
- IX. Azure management services
- X. What is Azure Resource Groups
- XI. Configuration and management of Azure Resource groups for hosting Azure services

Module 3: Introduction to Azure Resource Manager & Cloud Storage Services:

I. Completed walkthrough of the Azure Portal with all the features

II. What is Resource Groups and why we need RG's in Azure cloud computing platform to host resources?

III. Different types of Storage Accounts provisioning in Cloud computing with different storage services:

- Container/Blob storage service
- File share storage service
- Table storage service
- Queue storage service

IV. Detailed explanation & understanding of different Blob/container storage services:

- Page Blob
- Append Blob
- Block Blob

V. Creating and managing the data in container storage services with Public and Private accesses as per the need of a project

VI. Implementation of Snapshots for Blob storage services and File share storage service

VII. Generating SAS for different storage services to make the storage content browseable across all the globe or Publicly

VIII. Explanation of Standard Storage Account and Premium Storage account and their use in real-time scenarios

IX. Detailed explanation and implementation of Data Lake storage Gen2 Storage Account to store unstructured data in cloud storage services

X. Features/properties of Azure Storage Accounts (Overview, activity log, Tags, Access control (IAM), Storage browser, etc.)

XI. Maintenance and management of Storage keys and connection strings for Azure Storage services

XII. Implementing different levels of access (Reader, Contributor, Owner) to the Azure Storage accounts

Module 4: Migration of storage contents across Public & Private Clouds:

I. Moving the storage account with storage content across different Resource Groups based on real-time scenarios

II. Migrating data from On-prem (Private cloud) to Azure Storage account (Public cloud) using Az copy (forward migration)

III. Migrating data from Public cloud to Private cloud (reverse migration)

IV. Implementing Az copy commands to migrate data:

- On-prem to Azure cloud storage services
- Cloud storage services to On-prem
- Cloud to Cloud

V. Moving the Storage Account & its content from one Resource Group to another

Module 5: Authentication & Authorization of Azure Storage Services:

I. Configuration, Authentication, and Authorization for Storage Account via Azure Active Directory

II. Implementation of IAM for different storage accounts to perform read, write, modify, and upgrade operations

III. Verification and validations of access granted to different storage accounts for users

IV. Implementing Reader, Contributor, Owner access to Storage accounts as per business needs

V. Hosting File Share Storage services to On-prem servers or Cloud Servers as a shared drive to access and generate new files and data

Module 6: Provisioning of SQL Databases in Private & Public Cloud Computing:

I. Introduction to SQL Databases

II. Creation of new SQL Databases & Sample SQL Databases both in On-prem and Cloud computing

III. Planning and deploying Azure SQL Database

IV. Implementing and managing Azure SQL Database

V. Managing Azure SQL Database security

VI. Planning and deployment of SQL Databases in Azure cloud computing with real-time scenarios

VII. Different Database Deployment options

VIII. Database purchasing models (VCore & DTU's)

IX. Visualization of cloud Database server, Database, and validation of data from on-prem (private cloud)

X. Implementation of Firewall security rules on Azure Database servers to access and connect from on-prem SSMS

XI. Creation of Database in on-premises and synchronization with Azure cloud

Module 7: Introduction to SQL Server & SQL Queries from Basics to Advanced:

I. Introduction to SQL Database Queries

II. Detailed explanations, syntax & execution of the following queries based on 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)
- Union queries
- Group By queries
- Having queries
- Exists queries
- Subqueries
- Common Table Expressions/Temporary Tables
- Select Into queries
- Insert Into Select queries
- Stored Procedures queries

Module 8: What is Azure Data Factory (ADF):

I. Deep understanding and implementation of ADF concepts/components:

- Pipelines
- Activities
- Datasets
- Linked Services

II. Building blocks of Azure Data Factory:

- Triggers
- Integration runtime
- Dataflow

III. Complete features and walkthrough of Azure Data Factory studio

IV. Implementation of different triggers in ADF:

- Scheduled trigger
- Tumbling window trigger
- Event trigger

V. Integration runtime and its types:

- Azure
- Azure-SSIS
- Self-hosted

VI. Why to use cases for ADF

VII. Different types of ADF pipelines:

- Dynamic pipelines

- Parameterized pipelines
- Automated pipelines

Pipelines in ADF

VIII. Types of activities in ADF:

- Data movement activities
- Data transformation activities
- Data control activities

IX. Datasets in Azure Data factory

X. Linked Services in ADF.

Module 9: Controls/Activities of Azure Data Factory (ADF) for copying the DATA across various sources to Azure IAAS & PAAS Services:

- I. Copying the data from Blob Storage account to ADL's Gen2 Storage account.
- II. Copying of zip files (.csv) from Blob SA to ADL's Gen2 SA using ADF.
- III. Implementation and explanation of Metadata control in ADF to find the structure before copying the data.
- IV. Implementation and explanation of Validation and If Condition.
- V. Implementation of Get Metadata control, Filter control & For Each Control or activities in ADF.
- VI. Implementation & execution to copy the data from GitHub platform to Azure Storage services with variables and parameters.
- VII. Implementation of Foreach control, copy data control, and Set variable to dynamically load the data from source to target using ADF.
- VIII. Creating Dynamic pipelines with lookup activity to copy multiple .csv files data picking from JSON format data in Azure Storage services.
- IX. Copying the files from GitHub dynamically with the use of Dynamic parameters allocation - AUTOMATION PROCESS.
- X. Copying the data from different file formats (.csv, .xlsx, .txt, .Parquet, .Json, .SQL) using suitable ADF controls/activities.
- XI. Implementation and execution of loading the data from Blob SA to SQL DB single table & multiple tables using copy data activity, ForEach activity.
- XII. Executing multiple pipelines in parallel with Execute pipeline activity.

Module 10: Automation of Dataflow/Datacopy to various sources and destinations in ADF:

- I. Implementation of Schedule-based triggers for different ADF pipelines containing different activities.
- II. Implementation of Event-based triggers for different ADF pipelines containing different activities.
- III. Implementation of Tumbling window-based triggers for different ADF pipelines containing different activities.
- IV. Implementation and execution of storage and Event-based triggers.

Module 11: What is Azure Keyvault, purpose of using Keyvault, Storing the SA keys, connection string in Azure KV with Access policies:

- I. Detailed explanation & implementation of Azure Keyvaults.

- II. Making the SQL DB connection string to store in Keyvault to enhance the security for SA content and SQL DB.
- III. Generating the secrets inside the Azure Keyvault and granting access by implementing the access policies for different users.

Module 12: Data Flows Transformations in Azure Data Factory:

- I. Designing new Data flows.
- II. Designing and implementing transformations:
 - Source transformation
 - Join transformations
- III. Inline Datasets in data flow source control
- IV. Designing and implementing Data flow with Source transformations, Filter transformations & Sink transformations in ADF with inline Datasets.
- V. Implementation of Select transformations with Data flows for various source controls.
- VI. Implementation of Dataflows using Aggregate & Sink transformation.
- VII. Implementation of Dataflow with Conditional Split & Sink transformation with copy data activity.
- VIII. Implementation of Dataflow with Exists & Sink transformation.
- IX. Implementation of Azure Dataflows for Derived Column transformation with Source & Sink transformation.
- X. Implementation of Azure Dataflows to connect to SQL DB with Source & Sink transformation.

Module 13: Provisioning of Azure Data Engineering resources/cloud services with Artificial Intelligence (AI/Copilot):

- I. Overview of AI and how we can use AI to provision new resources (IAAS, PAAS & SAAS) in Azure cloud computing as an Azure Data Engineer.
- II. Key Features of Artificial Intelligence/Copilot.
- III. Benefits of AI/Copilot in Azure Cloud Computing.
- IV. Use Cases of AI & How AI works in Azure cloud computing platform.
- V. Execution of different prompts in AI to provision new resources for Azure Data Engineer projects in Azure cloud computing platform for automations, billings, operational tasks, alerts & configurations.
- VI. Future of AI with Azure Data Engineering cloud services.
- VII. Effectively utilizing AI to maintain, manage, provision, and automate Azure cloud computing resources using Artificial Intelligence.

Module 14: Azure Databricks & Apache Spark:

- I. What is Apache Spark.
- II. What is Azure Databricks.
- III. Why we are using Azure Databricks.
- IV. Detailed explanation of Databricks architecture and its components.
- V. What is Control Plane & Data Plane in Databricks architecture.
- VI. Detailed explanation and understanding of control plane components in Azure Databricks.
- VII. Illustration of Data plane components in Databricks architecture.

- VIII. Detailed explanation and provisioning of Databricks workspace features and clusters implementation to associate with Pyspark notebooks.
- IX. Implementation of Azure Databricks cluster by considering single node & multi-node as per business needs.
- X. Different features and properties of Azure Databricks clusters:
 - Single node
 - Multi-node
 - Photon acceleration
 - Auto turn off Azure Databricks cluster after a defined time
 - Autoscaling of cluster
 - Configuration provisioning of Azure Databricks clusters

Module 15: Azure Data Bricks & Apache Spark clusters features:

- I. Creating single node and multi nodes clusters
- II. Creation of Pyspark notebooks in Databricks cluster to fulfil different business requirements.
- III. Creation of folder hierarchies, notebooks in Azure Databricks workspace.
- IV. Onboarding users, data files in Azure Databricks workspace
- V. Writing pyspark scripts to fetch the data from source system in Azure Databricks
- VI. Mounting the Storage accounts with Azure Databricks to fetch the data from different source systems.
- VII. Extracting the data from web portal by writing the pyspark scripts
- VIII. Connecting Azure Databricks to different API's to write the scripts in SQL & Pyspark scripting.
- IX. Converting the python code to SQL scripts in Azure Databricks o Onboarding source files in Azure Databricks workspace DBFS.
- X. Importing files, folders, extracting data from files in Azure

Module 16: Azure Databricks Notebooks :

- I. Databricks Files System(DBFS):
- II. Importing raw data files into DBFS, reading and analysing the file data with Pyspark scripts:
- III. Mount points in Azure Databricks with Blob Storage & Data Lake Storage services.
- IV. Installing Databricks CLI & configuring with Azure Databricks Workspace
- V. Installing python package in local laptop to connect with Azure Databricks workspace
- VI. Generating Access token in Databricks workspace to integrate with python package.
- VII. Copying files from source system to target system using
- VIII. Databricks pyspark commands o Copying specific types of file from source system to target systems based upon the business requirements.
- IX. **Automations** in Azure Data bricks to copy the data from source to target in timely manner to fulfil the business requirements.
- X. Azure Databricks backed secret scope to read the files without exposing the keys of source systems.

XI. Detail explanation and differentiation between Azure Data Factory & Azure Data bricks **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

Module 17: Widgets utilities in Azure Databricks:

- I. Combobox
- II. Dropdown
- III. Multiselect
- IV. Text
- V. Remove
- VI. Removeall

Module 18: Slowly Changing Dimensions(SCD Type2) in Azure Databricks Notebook with Pyspark:

- I. Different types of SCD's
- II. Details explanation of SCD Type1, Type2 & Type3
- III. Generation of sample data for SCD Type2 Concept.
- IV. Implementation of SCD Type2 in Azure Databricks Notebook with Pyspark commands.

Module 19: Connecting Azure Databricks Cluster to Azure SQL DB to extract/read the data from SQL DB tables:

- I. Integrating Azure Databricks cluster with SQL DB using JDBC.SQLServer driver in our Pyspark notebook.
- II. Reading the data from SQL DB tables by declaring variables in Pyspark notebooks.
- III. Retrieving the data from multiple tables using different filters and joins.
- IV. Creating dataframe as per the business requirement to display the data from SQL tables in Azure databricks Pyspark notebook.

Module 20: Azure Synapse Analytics:

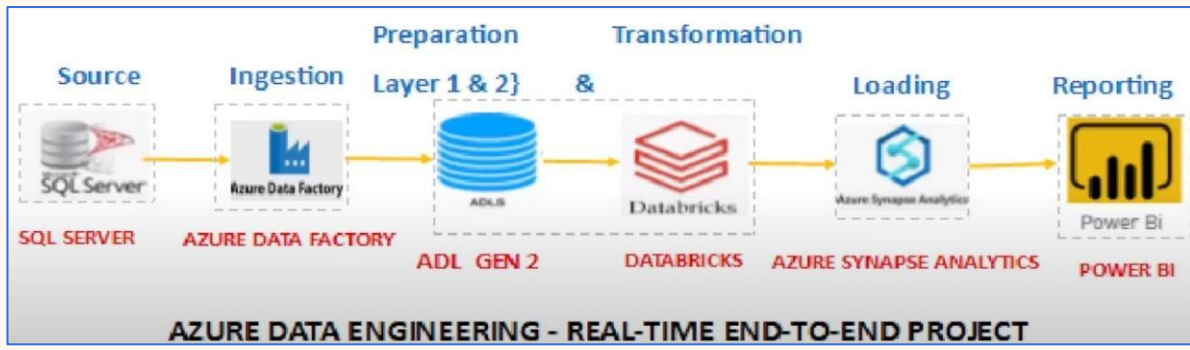
- I. What is Azure Synapse Analytics.
- II. Synapse workspace use cases.
- III. Synapse SQL overview.
- IV. Apache Spark for Synapse.
- V. Designing Pipelines in Azure Synapse.
- VI. Implementation of Linked Services/Datasets in Synapse Analytics.
- VII. Implementation of dedicated SQL Pool inside Synapse Analytics.
- VIII. Implementation of serverless SQL Pool inside Synapse Analytics.
- IX. Creation of Apache Spark pool in Azure Synapse Analytics.
- X. Writing SQL Script in Azure Synapse Analytics to generate results in tabular and chart formats.
- XI. Data visualization in Synapse Analytics using various charts (pie charts, line charts, bar charts, etc.).
- XII. Designing Synapse Analytics pipelines with various activities as per business requirements.
- XIII. Creation of Datasets, Linked services for Synapse Analytics pipelines.
- XIV. Data analysis with serverless Spark pools in Azure Synapse Analytics.

Module 21: Azure Stream Analytics:

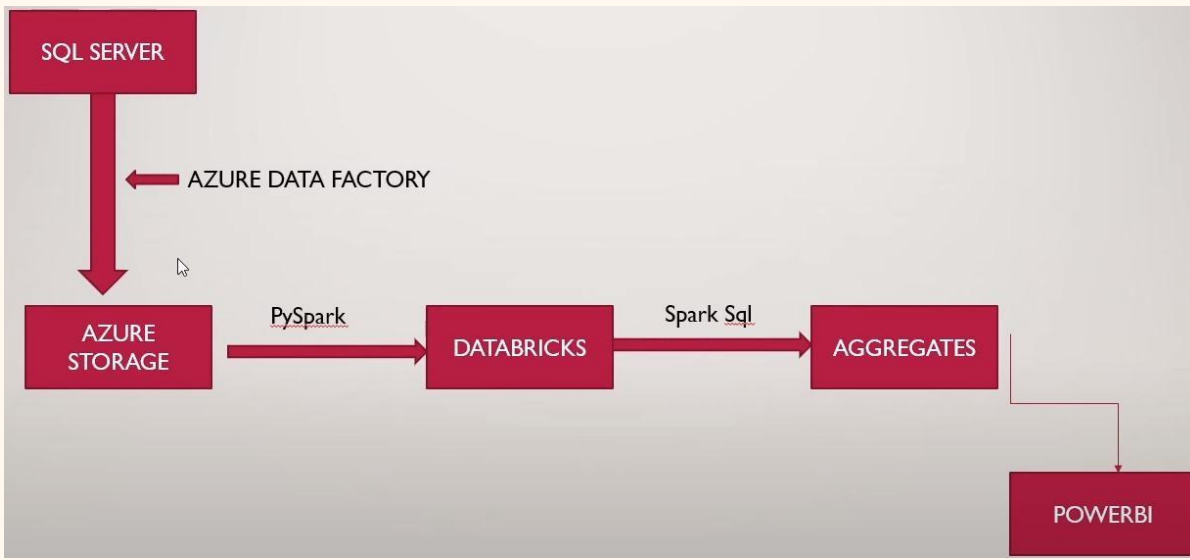
- I. What is Azure Stream Analytics.
- II. Purpose and usage of Stream Analytics in Azure cloud computing.
- III. Benefits and advantages of Stream Analytics.
- IV. Architecture diagram of data flow in Azure Stream Analytics with other cloud services.
- V. Understanding & usage of browser-based Raspberry Pi simulator.
- VI. Deployment of IoT Hub services as input for Stream Analytics jobs.
- VII. Implementation & execution of Stream Analytics jobs and designing inputs and outputs for IoT Hub and Data Lake Gen2.
- VIII. Writing SQL scripts to generate live streaming data and load it into the destination.

Module 22: Capstone Projects (An absolute Real-time course end projects):

- Project 1: Data analysis and populating updated data in Power BI dashboard with just one refresh for end users and Board of Directors.



- Project 2: Performing data analysis to present updated data in Power BI dashboard with just one refresh for end users and Board of Directors.



KEY HIGHLIGHTS OF THIS TRAINING PROGRAM:

- ✓ Entire training programme is in Hindi Language for Better understanding.
- ✓ Special focus on Non- technical and Fresher candidates.
- ✓ Resume Preparation for Fresher's and Experienced Both.
- ✓ Provides Recording of each live session which you can access from anywhere anytime for One year.
- ✓ Interview Cracking tips during live sessions.
- ✓ Provide complete notes and e-books for preparation.
