



Technical Guftgu

Established Under Ministry of Micro, Small and Medium Enterprises, Govt. of India

Contact - **+91 9870663188** or **8527556109**

Website: www.technicalguftgu.in

- ✓ Certificate provided
- ✓ Recordings Provided
- ✓ In Hindi
- ✓ Expert Trainers

Azure Data Engineering (DP-203)

Course Content

Module 1: SQL Fundamentals

- DDL, DML, DCL
- OLAP vs. OLTP
- Procedures & Functions
- Joins
- Aggregation
- Analytical Functions
- Subqueries
- Views
- Schemas and Instances

Lab

- Syntax & Query designing for SQL

Module 2: Python & Spark

- Tuple Array
- Pandas
- Spark architecture
- Python function
- Modules
- Key components of spark
- Pyspark

Lab

- Architecture, syntax code & programming

Module 3: Azure Fundamentals

- What are different cloud concepts?
- Types of Clouds (Public, Private, Hybrid)
- What is Azure?
- IaaS, PaaS, SaaS
- What is Regions and Region Pairs
- What is Resource Groups & Subscriptions

- What is Role-Based Access Control (RBAC)
- Azure Policy
- Role based Access Control (RBAC)
- Creating and managing Azure Resource Manager (ARM)
- Azure Powershell
- CLI
- Active Directory
- What is resource locks & functionality and usage of tags
- Hierarchy of service

Lab

- Setting up free account for Azure
- Functionality and usage of the Azure Portal, Azure PowerShell, Azure CLI
- Create Resource group
- Apply Tags or locks
- ARM
- Overview of the services
- Azure services provisioning with portal or ARM

Module 4: Azure Storage

- Different types of storage
- Access Tiers & Redundancy
- Datalake storage
- Different authentication & security option
- Understand Life cycle management
- Working with Azure Storage Containers
- Serving layer
- Soft delete

Lab

- Deploying Azure Storage

- Apply ACL & Lifecycle managements
- Implement File partitioning
- Maintain different serving layer on ADLS for storing data from multiple sources

Module 5: Azure SQL

- Different types of azure sql & understand the use cases
- Different service tiers for sql
- Understand dynamic data masking
- Row level security
- Encryption & threat protection

Lab

- Deploying Azure Deploy Azure SQL
- Enable AD Authentication
- Implement Dynamic Data Masking & Row level Security

Module 6: Azure Data Factory

- Azure Data Factory Architecture
- Linked Service, Dataset & Activities
- Pipeline Executions & Scheduling
- Pipeline Trigger Schedules, Modifications
- Debugging: ADF Managed Executions
- Dataflow debug session
- Cluster allocation for dataflowdebug
- Understand different types of IR
- Setup self hosted IR
- Onboard data from on-prem to azure with Selfhosted
- Handling different files format
- Work within multiple serving layer
- Table_Schema for Column Mapping
- Writing Expressions For Dynamic Loads
- Transformation Editor and Parameters
- Rerun adf pipelines
- Error handling in ADF
- Incremental Loads
- Incremental Load Pipeline Design in ADF
- Stored Procedures, Loops in ADF Pipelines
- Configure ETL Sources, Pre-Copy Scripts

Lab

- Deploy ADF With key components
- Setting up source & sink for ADF
- Build pipeline for data ingestion

- Onboard data from on-prem to azure
- Dynamic data load
- Incremental Data load
- Email configuration with ADF

Module 7: Azure Databricks

- Introduction to Azure Databricks
- working with various note books like python, scala, spark etc.
- Read and write data in Azure Databricks
- Data processing in Azure Databricks
- Work with DataFrames in Azure Databricks
- Platform architecture, security, and data protection in Azure Databricks
- Azure Databricks Configuration with Datalake
- Datalake mounting
- Process data
- Understande dataframes
- Understand function in databrick
- Handle data transformation scenerio with ADB
- Work with duplicate data
- Handle different time format with ADB
- Work with schema in adb

Lab

- Read and write data in Azure Databricks
- Data processing in Azure Databricks
- Work with DataFrames in Azure Databricks
- Write code for converting csv files record into JSON
- Work with different functions in ADB

Module 8: Azure Synapse

- Introduction to Azure Synapse
- Dedicated pool & Serverless Pool
- data partitioning
- spark pool
- synapse node distribution architecture
- Create external table In serverlesspool
- running notebook in Apache soark pool
- Partitioning, Indexing, Distribution
- Mapping Data Flow
- SCD
- Implement SCD with
- Lookup Transformation

- Exist Transformation
- Derived column Transformation
- materialized view
- difference between datawarehouse & datalake & database

Lab

- Deploying Synapse
- Deploye dedicated sql & spark pool
- create external tables
- Implement SCD With Synapse/ADF pipeline

Module 9: Eventhub & stream analytics job

- Evethub & eventhubnamespace
- partion key
- consumer group & connectivity
- send events to eventhub
- send real time data with stream analytics job
- Windowing function
- Setup input & output & query in SJ

Lab

- Send streaming data to eventhub

Module 10: Authentication

- RBAC, ACCESS TOKEN, Mnaged identity, keyvault, service principle

Lab

- Use keyvault secret for handling credentials
- Design pipeline while storing the credentials in Keyvault

Module 11: Auditing

- Log analytics & azure monitor, application insight

Lab

- Configure log analytics & azure montior for auditing

Module 12: Best practices

- Microsoft recommendation, Azure DevOps, Committing code in branch, best practices, recap & QnA

Lab

- Best Practices

KEY HIGHLIGHTS OF THIS TRAINING PROGRAM:

- ✓ *Case Study Based Session. High on interactivity and Q&A.*
- ✓ *Architect Drawing Representation of concept.*
- ✓ *Official Microsoft Slide Decks for Revision and Learning, Shared on Google Drive.*
- ✓ *95% more hands-on training.*
