

ONLINE COURSE

FULL STACK AZURE DATA ENGINEERING

(Developers Guidance Program)



Mr. Dinesh Kumar
ex-TCS

Python



Mr. Arun Kumar
*Ex-TCS, Ex-Tredence,
Ex-Harman*
Azure



Mr. Amit Kaushik
At WNS

SQL

TOPICS COVERED

1. Python
2. SQL
3. Azure Data Engineering
 - Azure Databricks
 - Azure Storage
 - Azure DevOps
 - Azure Data Factory
 - Pyspark
 - Azure Key Vault
 - Azure Logic Apps
 - Azure SQL Database

28 OCTOBER

2023 ONWARDS

Register Now!

Contact Us

Phone:

+91- 78709 70617

+91- 70420 97759

E-Mail:

inquiry@gbdrift.com

PYTHON SYLLABUS

FUNDAMENTAL OF PYTHON

- Learn about Language fundamental.
- Know about What, Why and How to use.
- Basic syntax.
- Identifiers
- Reserved Words
- Data Types
- Type Casting

PYTHON OPERATORS

- Arithmetic Operators
- Relational Operators
- Equality operators
- Logical Operators
- Assignment Operators
- Ternary Operator
- Special operators
- Math Module

FLOW CONTROL AND I/O

- Dynamic input from the keyboard
- Dynamic input from the keyboard
- Command Line Arguments
- Output statements
- Conditional Statements
- Iterative Statements
- Transfer Statements

DATA TYPES IN PYTHON

- String data type
- List data type
- Set data type
- Dictionary data structure

MODULES, PACKAGES

- Module aliasing
- Various possibilities of import
- The Special variable
- Packages in details

EXCEPTION HANDLING

- What is Exception
- Default Exception Handling
- Customized Exception Handling
- Types of Exceptions

FUNCTIONS AND IT' S FEATURE

- What is function
- Parameters
- Return Statement
- Types of functions
- Recursive Functions
- Nested Functions
- Generators

REGULAR EXPRESSIONS

- What is regular exp
- Character classes
- Quantifiers
- Important functions of re module

PYTHON LOGGING

- Python Logging
- Ways to implements logging.

ForumDE
WE RISE BY LIFTING OTHERS

OOPS CONCEPT

- Learn about Class,
- Method and it's types.
- Different types of pillars in Opps and its significance.
- Abstraction
- Encapsulation
- Inheritance
- Polymorphism

MULTI-THREADING, FILE HANDLING

- Learn about Input and output for handle files.
- What is multithreading?
- How to do multithreading

DATABASE PROGRAMMING IN PYTHON

- Learn about database programming using python.
- Open, fetch data and close the connection from database.

ENROLLMENT: Contact: 9608703348, 7870970617 | Website: <https://forumde.in/>

SQL SYLLABUS

INTRO

- Database, Dataware house and DataLake
- Type of databases
- RDBMS
- Logical Schema Designing
- Data modeling
- Star Schema
- Fact & Dimension table
- Normalization
- Normal Forms
- Sequence Diagram
- E-R Diagram
- Flowcharts
- Physical Database Design
- BusinessConstraints–Primarykey,ForeignKey,CandidateKey,CompositeKey,SurrogateKey,Unique Key, Check, Default, Not Null
- Index-Clustered and Non-Clustered

SQL->T-SQL/ORACLE

- Download and Install SSMS, SQL Server
- SQL commands-DDL,DML,DCL,TCL
- Select,INSERT,Update,Delete,Merge,Create,Alter,Truncate,Rename,Grant,Revoke,Commit,Rollback,Savepoint, Drop
- Query Execution sequence/Order
- Clause, Operators, Predicates in SQL
- LIKE,IN,LIMIT,TOP,CASE,INSTRING,SUBSTRING,AVERAGE,COUNT,MAX,MIN,SUMUPPER,LOWER,EXIST,EXTRACT,WildCardOperatorsEtc.
- SUBQUERIES-Nested Sub queries, Correlated Sub queries, In line View
- JOINS–Self Join, Inner Join, Left Outer Join, Right Outer Join, Cross Join
- SET Operators–Union, Union All, Minus, Intersect
- CTE(Common Table Expression)
- WindowFunction–Rank,DenseRank,Rownumber,Lead,Lag,FirstValue,Last_Value,PercentRank
- Pivot and Un pivot
- Rollup, Cube, Grouping Sets
- Views, Materialized View

AZURE DATA ENGINEERING SYLLABUS



ETL CONCEPT

- Lake house Concept
- Data Warehouse Concept
- Difference between Database, Data Warehouse and Lake house?
- OLAP vs OLTP systems Concept
- Fact Table and Dimension Table
- Data Modeling-Star Schema vs Snowflake Schema Concept

SPARK

- Architecture of Spark
- Job, Stages and Task in Spark
- Pyspark concepts
- Spark SQL concepts
- Transformation and Action
- Lineage

- Lazy Evaluation
 - rdd vs dataframe vs dataset
 - Immutability Concept
 - operations on RDD
 - operations on dataframe
 - Difference file formats like parquet, avro, orc and their differences
 - Understanding Shuffling in Spark
 - Narrow Transformation
 - Wide Transformation
 - Accumulator
 - Broadcast variable
 - Partition By
 - repartition and coalesce
 - partition pruning
 - cache vs persist
 - User Defined Functions(UDF)
-

AZURE

- Furnishing basic information about services of Azure with perspective of Data Engineering and how it is used in industry and real time scenario.

AZURE DATABRICKS

- Clusters
- Widgets
- Databricks CLI overview
- Service Principal
- Mount Point
- JDBC connection with Azure SQL Server
- Read and write data in Azure Databricks
- Data processing in Azure Databricks
- Introductory notebooks
- Working with Dataframes in Azure Databricks
- Working with different types of file formats like CSV, Parquet, Delta, Avro
- dbutils commands
- Databricks Runtime
- Concept of Delta Table
- Get started with Delta Lake
- Time Travel in Delta Lake
- Schema Enforcement and Schema Evolution
- MERGE statement
- Ingest data into Data Lake
- Table batch reads and writes
- Table delete, update, and merge
- Handling Bad Records
- Scheduling jobs in Databricks Notebook
- Medallion architecture

ENROLLMENT: Contact: 9608703348, 7870970617 | Website: <https://forumde.in/>

AZURE DATA LAKE GEN2 (ADLSGEN2)

- How to provision ADLS Gen2
- Theoretical concept of Storage Azure Data Lake Gen2
- How do we store data in ADLS Gen2
- How to read/write data in ADLS Gen2 from databricks
- Hot, Cold, Cool and Archive Access Tiers
- Difference between ADLS Gen2 vs ADLS Gen1 vs Blob Storage
- ZRS,LRS,GRS,GZRS

AZURE DATA FACTORY (ADF)

- Pipelines
- Activities
- Datasets
- Linked Services
- Integration Runtimes
- Auto Resolve Infrastructure Runtime
- Self Hosted Infrastructure Runtime(SHIR)
- SSIS Infrastructure Runtime
- Pipeline Run
- Scheduling and Trigger of pipeline
- Parameters
- Variables
- Orchestration of Pipeline in ADF
- Triggers
- Schedule Trigger and its System Variables
- Event Based Trigger and its System Variables
- Tumbling Widow Trigger and its System Variables
- Data Flow Pipeline
- Pipeline Level Parameters
- Global Parameters
- Creating Dynamic Pipeline creation
- How to monitor ADF data pipelines?
- ARM Template
- Copy one table from My SQL (on-premise) to Azure SQL Server.
- Copy filtered data of one table from My SQL (on-premise) to Azure SQL Server.
- Copy All tables from My SQL Server(on-premise) to Azure SQL Server
- Copy selected tables from My SQL Server(on-premise) to Azure SQL Server
- Pipeline for Incremental Load
- Copy all tables from Azure SQL Database to ADLS Gen2
- Copy selected tables from Azure SQL Database to ADLS Gen2
- Error Logging of ADF pipelines in Azure SQL Server.

DEPLOYMENT

- Deploy Azure Databricks and Azure Data Factory pipelines using Azure DevOps.

AZURE KEY-VAULT SERVICE

- How to securely store Keys and Password in Azure Key-Vault
- How to access Keys and Password for Azure Key-Vault
- How to connect AKV from ADF(Azure Data Factory) and read the secrets of AKV.
- How to connect AKV from ADB(Azure Databricks) and read the secrets of AKV.

AZURE LOGIC APP

- How to send mail alert for pipeline failure and success using logic app.

AZURE SQL SERVER

- Connect Azure SQL server from databricks
- Connect Azure SQL server from Azure Data Factory

SOFTWARE INSTALLATIONS

- Installation of My SQL
- Installation of SSMS and connection with Azure SQL Server
- Installation of Azure Data Studio and connection with Azure SQL Server
- Installation of Azure Storage Explorer and connection with ADLS Gen2
- Installation of Notepad++

ForumDE
WE RISE BY LIFTING OTHERS

APART FROM ABOVE COURSE DETAILS, WE WILL ALSO DELIVER BELOW: -

- In addition to this we will do one end to end data pipeline on Azure using all the major components of Azure using Python as programming language.
- We will guide you what kind of questions can be asked in interview.
- We will help you in building your resume.
- We will take mock interview to prepare you for real time interview.

BATCH DETAILS:

The classes will commence on **28 October 2023** and the **expected duration** of the course will be **3 months**.

The classes will be **Instructor led online training**.

The classes will be held on **Python (Friday, Saturday and Sunday)**

SQL (Monday, Tuesday and Wednesday)

AZURE (Monday, Tuesday and Wednesday)

The timing of classes will be **8:00 pm to 9:30 pm**.

Classroom video recording will be provided to all the candidates till **1 year** from the date of course completion.

Sharing of videos to another person is strictly prohibited.

Contact us For Admission or Queries: -



+91-7870970617



+91-7042097759