

VERSION IT

Since 2001



AZURE DATA ENGINEERING & FABIC COURSE

(Azure Data Factory + Azure Databricks +
Python Basics + SQL Basics + Microsoft Fabric)



Best Training And Placements Institute



211, 2nd Floor, Anapurna Block,
Adithya Enclave, Ameerpet.



+91 9848015399
+91 9391237284

www.versionit.co.in

About Version IT

Version IT is not a mere software training institute, a team of IT professionals developed it as the best knowledge centre for hundreds of career-building conscious young people. Our training academy is the best training institute in Hyderabad offering various software courses with aptly placement orientation. We proudly announce that we achieved 100% placements in every batch we have taken up in the past two decades. Version IT Academy's strength is our academic excellence with which we have been placed in the top position among the software training institute in Hyderabad.



Corporate Training



Class Room Training



Online Training

Why Choose Us!

Training By Certified Instructors



Mock Interviews



Weekly Assignments



Project Training



Interview Cracking tips



Resume Preparation

A. Foundations of Data Engineering

1. Fundamentals of Cloud Computing

- What is Cloud Computing?
- Cloud Deployment Models
 - Private Cloud
 - Public Cloud
 - Hybrid Cloud
- Cloud Service Models
 - IaaS (Infrastructure as a Service)
 - PaaS (Platform as a Service)
 - SaaS (Software as a Service)

2. Overview of Major Cloud Providers

- Introduction to Leading Cloud Platforms
 - Microsoft Azure
 - Amazon Web Services (AWS)
 - Google Cloud Platform (GCP)

3. Getting Started with Microsoft Azure

- Introduction to Azure and Its Ecosystem
 - Navigating the Azure Portal
 - Understanding Subscriptions
 - Creating and Managing Resource Groups
- Azure Resources Overview

4. Core Azure Services for Data Engineering

- Azure Data Factory
- Azure Databricks
- Azure Blob Storage & Data Lake Gen1/Gen2
- Azure SQL Server & SQL Database
- Azure Key Vault for Secrets Management

5. Introduction to Big Data Concepts

- | | |
|--|-------------------|
| • What is Data vs Big Data? | • Types of Data: |
| • Common Sources of Big Data | • Structured |
| • The 5 V's of Big Data: | • Semi-Structured |
| • Volume, Variety, Velocity, Veracity, Value | • Unstructured |

6. Python Programming Essentials

- **Python Syntax and Structure**
- **Core Concepts:**
 - **Variables and Data Types**
 - **Operators**
 - **Collections (Lists, Tuples, Sets, Dictionaries)**
 - **Functions and Parameters**
 - **If-Else, For & While Loops**

7. SQL for Data Handling

- **SQL Language Fundamentals**
 - **DQL: SELECT**
 - **DDL: CREATE, ALTER, DROP, TRUNCATE**
 - **DML: INSERT, UPDATE, DELETE, MERGE**
- **Filtering and Conditions**
 - **WHERE, AND, OR, NOT, IN, BETWEEN, LIKE, IS NULL, CASE WHEN**
- **Sorting and Limiting**
 - **ORDER BY, ASC, DESC, LIMIT, FETCH FIRST**
- **Working with Joins**
 - **INNER, LEFT, RIGHT, FULL, CROSS, SELF JOIN**
- **Aggregate Functions**
 - **SUM, AVG, COUNT, MIN, MAX**
- **Grouping and Filtering**
 - **GROUP BY, HAVING**

B. Azure Data Factory (ADF)

1. Introduction to ADF

- What is Azure Data Factory?
- Use Cases and Benefits
- ADF Architecture Overview
- Navigating the ADF Interface
- Key Concepts:
 - Pipelines, Activities, Linked Services
 - Datasets, Triggers, Data Flows
 - Integration Runtimes

2. Building Pipelines

- Purpose and Usage
- Creating Linked Services for:
 - Azure BLOB, SQL, ADLS Gen1/Gen2, Oracle, PostgreSQL
- Parameterizing Linked Services

3. Linked Services

- Purpose and Usage
- Creating Linked Services for:
 - Azure BLOB, SQL, ADLS Gen1/Gen2, Oracle, PostgreSQL
- Parameterizing Linked Services

4. Working with Datasets

- Creating Datasets for File Formats:
 - Avro, Binary, CSV, Excel, JSON, ORC, Parquet, XML
- Datasets for Tabular Sources:
 - SQL Server, Azure SQL, Oracle
- Parameterization in Datasets

5. Activities in ADF

- Core Activities
 - Wait, Variable (Create/Set/Append)
- Copy Activity
 - General Settings, Source/Sink, Mapping
 - File Copy: Single, Multiple, Recursive
 - Format Conversions (CSV, Parquet, Avro)

- **Integration Activities**
 - Databricks, Azure Functions, Stored Procedures
 - Lookup, Get Metadata, Delete, Execute Pipeline
- **Iteration & Conditionals**
 - Filter, ForEach, If Condition, Switch, Until

6. Triggers

- **Introduction to Triggers**
- **Trigger Types:**
 - Schedule
 - Tumbling Window
 - Event-Based (Blob Events)
- **Trigger Parameterization**

7. Integration Runtime (IR)

- **Azure Auto-Resolve IR**
- **Managed Virtual Network IR**
- **Self-Hosted IR and Linked IR**

8. Source Control and CI/CD

- **Integrating with Git (Azure DevOps/GitHub)**
- **Using ARM Templates: Exporting, Importing**
- **Source Control Best Practices**

9. Global Parameters

- **Creating and Using Global Parameters**

10. Monitoring and Alerts

- **Monitoring Pipelines and Activities**
- **Setting Up Alerts and Notifications**

11. Data Flows

- **Creating Mapping Data Flows**
- **Using Data Flow Debug**
- **Transformations:**
 - Filter, Aggregate, Join
 - Conditional Split, Derived Column
 - Exists, Union, Lookup, Sort
 - GroupBy, Pivot, Unpivot
 - Flatten, Parse, Stringify
 - AlterRow, Assert, Flowlets
- **Schema Validation and Drift Handling**
- **Duplicate Removal**

C. Azure Databricks & PySpark

1. Introduction to Apache Spark

- Spark Architecture and Internals
- RDDs and DataFrames
- Spark Streaming Basics
- Spark vs Hadoop Comparison

2. Getting Started with Databricks

- What is Databricks?
- Architecture and Workspace Overview
- Using the Notebook Interface
- Introduction to DBFS
- File Handling with dbutils

3. Core Spark with Databricks – RDDs

- RDD Programming and Transformations
 - Narrow vs Wide Transformations
 - Lazy Evaluation
- Key-Value RDDs

4. Spark SQL & DataFrames

- Creating and Transforming DataFrames
- DataFrame Actions and Queries
- User Defined Functions (UDFs)
- Execution Internals

5. File Formats in Databricks

- Reading/Writing:
 - CSV, JSON, Parquet, Excel, ORC
- Schema Inference and Handling

6. Databricks Utilities

- dbutils Usage:
- Credentials, FS, Notebooks, Secrets, Widgets

7. Cluster Management

- **Creating & Configuring Clusters**
- **Cluster Modes:**
 - **Job vs All-Purpose**
 - **Standard vs High Concurrency**
- **Autoscaling and Runtime Versions**

8. Batch Processing Workloads

- **Historical vs Incremental Loads**
- **Data Transformations, Joins, Aggregations**
- **Using Unions and Window Functions**

9. Integration with Azure Services

- **Blob Storage, ADLS Gen2**
- **Azure SQL, Synapse, Key Vault**

10. Structured Streaming

- **Streaming Architecture and Use Cases**
- **PySpark Stream Processing**
- **Handling Bad Records, Stream to Table**

11. Delta Lake & Lakehouse

- **Delta Architecture & Features**
- **Delta Table Creation, Merge, DML**
- **SCD Type 1 and Type 2**
- **Deduplication and Streaming Integration**

12. Medallion Architecture

- **Bronze: Raw Data**
- **Silver: Cleansed/Enriched**
- **Gold: Curated for Analytics**

13. Job Orchestration

- **Using Databricks Workflows**
- **Job Scheduling and Monitoring**

14. Unity Catalog

- **Implementing Data Governance with Unity Catalog**

D. Microsoft Fabric

1. Introduction to Microsoft Fabric

- What is Microsoft Fabric?
- Key Features & Architecture
- Core Components:
 - Data Engineering
 - Data Factory
 - Synapse Integration
 - OneLake Storage
- Fabric vs Azure Synapse vs Databricks
- Creating a Fabric Workspace

2. Data Ingestion

- Batch vs Streaming
- Connecting to Sources:
 - Azure Blob, ADLS, SQL, REST APIs
- ETL with Fabric Pipelines
- Real-Time Ingestion with Eventstreams
- Handling Structured and Unstructured Data \

3. OneLake – Unified Storage

- What is OneLake?
- OneLake vs ADLS
- Lakehouses and Delta Table Management
- Security and Access Control

4. Spark in Fabric

- Spark Engine Overview
- Creating and Running Notebooks
- PySpark Transformations
- Performance Tuning Best Practices
- Job Monitoring and Scheduling

5. Dataflows (Low-Code)

- Dataflows Gen2 Introduction
- Power Query and M Language
- Managing and Automating Dataflows
- Integration with Fabric Data Factory
- Debugging and Monitoring

6. SQL Analytics & Modeling

- **Fabric SQL Data Warehouse Overview**
- **Query Writing and Optimization**
- **Implementing SCD Type 1 & 2**
- **Materialized Views and Incremental Processing**

7. Orchestration, Automation & CI/CD

- **Microsoft Fabric Data Factory Overview**
- **Pipeline Scheduling and Monitoring**
- **Integrating with Synapse & Power BI**
- **Error Handling and Logging**
- **CI/CD Deployment Best Practices**

Our Alumni Work At



+91 9848015399 +91 9391237284



www.versionit.co.in

Our Other Courses

Development Technologies

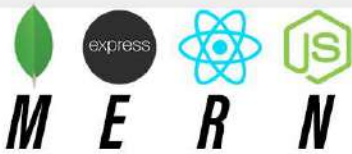
Java Full stack



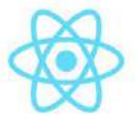
Python Full stack



.Net Full stack



React



Cloud Technologies



Dev Ops



Salesforce



servicenow

Triending Technologies

Data Science



Data Analytics



Cyber Security



Azure Data Engineer



Aws Data Engineer



GCP Data Engineer



Our Infrastructure



Our Branches

Address -1

HMDA Maitrivanam,
Plot #46, SAP St,
Gayatri Nagar, Ameerpet,
Hyderabad, Telangana
500016

Address -2

Amrutha Arcade, SAP St,
behind Maitrivanam , 1st Left,
Srinivasa Nagar, Ameerpet,
Hyderabad, Telangana 500082

Address -3

Flat No.211, 2nd Floor, ADITYA
ENCLAVE, Annapurna Block,
Ameerpet, Hyderabad,
Telangana 500016



+91 9848015399 +91 9391237284



www.versionit.co.in