



AZURE + MICROSOFT FABRIC DATA ENGINEERING COURSE

Best Training and Placements Institute



Address : 604/B ,6th floor ,Nilgiri Block , Adithya Enclave,
Ameerpet. Hyderabad, Telangana 500016.



+91 - 7995138210
+91 - 93925 30568



contact@nextitcareer.com

About Next IT Career

Next IT Career Best Software Coaching Center in Hyderabad Corporate training is a function of human resource management that aims to provide the organization's employees with the knowledge and skills required to be successful. In turn, the growth of employees also contributes to the success of the business. In recent years, Its also seen as a long-term strategic assignment rather than a cost center for the organization. The majority of companies have embraced 'continual learning' in it to promote employee growth both professionally and personally and acquire a highly skilled workforce as a result.



Corporate Training



Classroom Training



Online Training

Why Choose Us?



**Training by
Certified Instructors**



Weekly Assignments



Project Training



Resume Preparation



Mock Interviews



**Interview cracking
tips**

AZURE + MICROSOFT FABRIC DATA ENGINEERING COURSE

1. INTRODUCTION TO DATA ENGINEERING

- Overview of Data Engineering
- Definition and importance
- Key roles and responsibilities
- Introduction to Data Pipelines
- What is a data pipeline?
- Components of a data pipeline
- Examples of data pipelines in the industry

2. AZURE DATA FACTORY

- Introduction to ADF
- Different ways to work with ADF
- Pipelines and Activities in ADF
- Linked Services and Data Sets
- Triggers in ADF
- Schedule Trigger in ADF
- Tumbling Window Trigger
- Event Based Triggers
- Integration Runtime
- Azure Integration Runtime
- Self-Hosted Integration Runtime
- Derived Column transformations
- Derived Column transformations in dataflows
- Exists Transformations
- Union Transformations
- Lookup Transformations
- Sort Transformations
- New Branch in Mapping Dataflows
- Select Transformations
- Pivot Transformations

- Unpivot Transformations
- Surrogate Key Transformations
- Window Transformations
- Alter Row Transformations
- Flatten Transformations
- Parameterize Mapping Dataflow
- Validate Schema Mapping Data flow
- Schema Drift Mapping Dataflow
- Wrangling Dataflow
- Merge Queries in wrangling Dataflow
- Groupby in wrangling dataflow
- Author Modes
- Set Up Git hub Code repo
- Setup Azure devops Git code repo
- Use Azure Key vault secrets
- CI/CD in ADF
- How to Read Json Output from 1 Activity to Another
- Annotations in ADF
- Templates Overview in ADF
- Global Params in ADF
- Rank Transformations in ADF
- Cache Sink and cached Lookup
- Session Log in copy activity
- Write cache sink to activity output
- Parse Transformation in mapping Dataflow
- Fail Activity
- Inline Dataset
- Stringy Transformation in Dataflows
- Assert Transformations
- Flow lets
- Script Activity
- UDFs in Dataflows
- Fuzzy Joins in Dataflows

- Parameterize Linked Services
- Cast Transformation in Dataflows
- Extract Data from table of website Pages
- Per pipeline Billing view
- Create Alert Rules
- Pipeline return value in set variable
- Copy activity pagination rules
- ADF Real Time Scenarios - 20

3. PYTHON FOR DATA ENGINEERING

- Python Fundamentals
- Variables and Data Types
- Basic Operators
- Control Flow (if-else, loops)
- Functions and Modules
- Defining and calling functions
- Importing and using modules
- Working with Data Structures
- Lists, Tuples, Dictionaries, and Sets
- File Handling
- Reading and writing files
- Working with CSV and JSON files
- Python Libraries for Data Engineering
- Overview of Pandas, NumPy, and Matplotlib

4. SQL FOR DATA ENGINEERING

Introduction to SQL and Databases

- What is SQL?
- Types of SQL commands (DDL, DML, DCL, TCL)
- Tables, rows, and columns basics

Data Definition Language (DDL)

- Creating tables using CREATE TABLE
- Modifying tables using ALTER TABLE
- Deleting tables with DROP TABLE

Data Manipulation Language (DML)

- Inserting data using INSERT INTO
- Updating records with UPDATE
- Deleting records using DELETE

Data Retrieval using SELECT Statement

- Basic SELECT queries
- Using WHERE clause to filter data
- Sorting with ORDER BY, limiting with LIMIT or TOP

Filtering and Pattern Matching

- AND, OR, NOT conditions
- Using LIKE, IN, BETWEEN, IS NULL

Aggregate Functions and Grouping

- Functions like COUNT(), SUM(), AVG(), MAX(), MIN()
- Grouping data using GROUP BY
- Filtering groups with HAVING

5. DATA BRICKS AND PYSPARK

Introduction to Databricks

- Overview of Databricks
- What is Databricks?
- Databricks vs traditional data platforms
- Setting Up Databricks Environment
- Creating a Databricks account
- Navigating the Databricks workspace

Data Ingestions and Transformation with Databricks

- Data Ingestion Techniques
- Reading data from various sources(CSV, JSON, Parquet)
- Connecting to databases
- Data Transformation
- Basic transformations using Databricks
- Using SQL in Databricks
- Handling Missing Data and Duplicates
- Techniques for dealingwith missing values
- Removing and handlingduplicates

Apache Spark and PySpark

- Overview of Apache Spark
- What is ApacheSpark?
- Spark ecosystem and components
- Introduction to PySpark
- Setting up PySpark in Databricks
- PySpark vs Pandas

PySpark Basics

- PySpark Data Frames
- Creating Data Frames
- Performing basic operations on Data Frames
- Data Frame Transformations and Actions
- Common transformations (select, filter, group By, etc.)
- Actions (collect, show, count, etc.)

Advanced PySpark Concepts

- Working with Spark SQL
- Using SQL queriesin PySpark
- Integrating SQL and DataFrame API

- User Defined Functions (UDFs)
- Creating and using UDFs
- Performance considerations

Data Aggregation and Analysis with PySpark

- Aggregation Functions
- Grouping and aggregating data
- Window functions
- Data Joins in PySpark
- Different types of joins (inner, outer, etc.)
- Best practices for joins

Optimizing and Managing Spark jobs

- Performance Tuning
- Caching and persistence
- Partitioning and shuffling
- Spark Job Monitoring and Debugging
- Using Spark UI for monitoring
- Debugging common issues

Advanced Topics in Data bricks and PySpark

- Delta Lake and Data bricks Delta
- Introduction to Delta Lake
- Implementing Delta Lake in Data bricks
- Realtime Data Processing with Structured Streaming
- Basics of Structured Streaming
- Building and managing streaming pipelines
- Unity CatLog and its implementation

6. MICROSOFT FABRICFOR DATA ENGINEERING

Module 1: Introduction to Microsoft Fabric

- What is Microsoft Fabric? Overview of its capabilities
- Key Components: Data Engineering, Data Factory, Synapse,OneLake, etc.
- Comparing Microsoft Fabric with Azure Synapse and Databricks
- Understanding Fabrics Unified Data Lake (One Lake)
- Setting up a Microsoft Fabric Workspace

Module 2: Data Ingestion in Fabric

- Data Ingestion Methods:Batch vs. Streaming
- Connecting to Data Sources: Azure Blob, ADLS, SQL, APIs
- Using Fabric Pipelines for ETL (Extract, Transform, Load)
- Working with Eventstreams for real-time data
- Handling structured and unstructured data ingestion

Module 3: Storage and Management with One Lake

- Introduction to One Lake Storage in Microsoft Fabric
- One Lake vs. ADLS (Azure Data Lake Storage)
- Creating and Managing Lake houses in Fabric
- Delta Tables: Format, Transactions, and Versioning
- Data Security and Access Control in One Lake

Module 4: Data Processing with Spark and Notebooks

- Introduction to Apache Spark in Fabric
- Setting up and running Spark Notebooks
- Data Transformation using Py Spark
- Optimizing Spark Performance in Fabric
- Managing Spark Jobs and Scheduling

Module 5: Data Transformation with Dataflows and Pipelines

- Introduction to Dataflows Gen2
- Creating and Managing Dataflows
- Data Transformation with Power Query and M Language
- Automating Data Pipelines using Fabric Data Factory
- Debugging and Monitoring Fabric Pipelines

Module 6: Data Modelling and SQL Analytics in Fabric

- Understanding Fabric's Data Warehouse
- Writing SQL Queries for Data Analysis
- Performance Optimization in Fabric SQL Engine
- Implementing Slowly Changing Dimensions (SCDs)
- Materialized Views and Query Optimization

Module 7: Orchestration & Automation

- Understanding Microsoft Fabric Data Factory
- Creating and Scheduling Pipelines
- Integrating Data Factory with Synapse and Power BI
- Error Handling and Logging Mechanisms
- CI/CD Deployment in Microsoft Fabric

7. END TO END REAL TIME PROJECTS

- Domain - Healthcare Revenue Cycle Management (RCM)
- Domain – NYC Taxi – Real Time Data
- Real Time news API Project- Microsoft Fabri

Our Other Courses

Trending Software Testing courses

Manual Testing 

Selenium Automation 

Tocsa 

ETL Testing 

Performance Testing 

API Testing 

Load runner 

J-meter 

Neo load 

Most Trending Courses

Data Science 

Data Analytics 

Scrum Master 

Most Demand Courses

AWS DevOps 

Azure DevOps 

Power BI 

Java Full Stack 

Python Full Stack 

Cyber Security 

Our Recruitment Partners



Our Branch



Address: 604/B ,6th floor ,Nilgiri Block , Adithya Enclave,
Ameerpet. Hyderabad, Telangana 500016.



+91 - 7995138210
+91 - 7013123485



contact@nextitcareer.com

Our Infrastructure

