



NewbieGet

INNOVATION THROUGH LEARNING

UDYAM MSME · GEM REGISTERED

PROFESSIONAL TRAINING PROGRAM

Azure Data Engineering

[Spark · Databricks · Fabric]

A hands-on, Azure-focused programme that takes you from Spark fundamentals to a production-style cloud data pipeline - built, governed and monitored end to end.

38

HOURS

11

MODULES

1

CAPSTONE

10

TOOLS COVERED

TOOLS & PLATFORMS COVERED

PySpark

Azure Data Factory

ADLS Gen2

Azure Databricks

Delta Lake

Unity Catalog

Microsoft Fabric

Event Hubs

Spark SQL

Structured Streaming

Live

INSTRUCTOR LED

Lab

HANDS-ON GUIDANCE

Project

CAPSTONE REVIEW

Certificate

ON COMPLETION

Enrol for the upcoming batch

Live instructor-led · Weekday & weekend batches

+91 87912 95762

contact@newbieget.com

newbieget.tech

Built for the Azure data stack

This programme is deliberately **practical**. You start with the distributed-processing ideas that make Spark necessary, learn PySpark properly, then spend the bulk of your time inside the Azure services teams actually ship on – **Data Factory, ADLS Gen2, Databricks, Delta Lake, Unity Catalog and Microsoft Fabric** – finishing with a capstone pipeline you can walk an interviewer through.

DURATION**38 hours****PROGRAMME****14 Weeks****FOCUS****Azure-first, practical****LEVEL****Beginner →
Intermediate****Who this is for**

- Developers and analysts moving into data engineering
- SQL / ETL professionals modernising onto Azure
- Graduates targeting Azure data engineer roles
- Teams standardising on Databricks or Fabric

What you will master

- PySpark: RDDs, DataFrames, Spark SQL, joins and schemas
- Orchestration with Azure Data Factory, CDC and triggers
- Databricks Lakehouse, Delta Lake and MERGE/UPSERT
- Governance with Unity Catalog and Key Vault
- Streaming with Event Hubs and Structured Streaming

OUTCOMES

What you walk away with

01**A working pipeline**

An end-to-end Azure pipeline – ingestion through to an analytics-ready dataset – you built yourself.

02**Interview-ready depth**

Architecture reasoning: batch vs streaming, managed vs external tables, Fabric vs Databricks.

03**Completion certificate**

A NewbieGet certificate recognised by our industry partners on finishing the programme.

Foundations & PySpark

01

Big Data & Data Engineering Foundations

2.5 hrs

[Big Data concepts](#) · [Hadoop overview](#) · [Distributed processing](#)

Volume, Variety, Velocity, Veracity; structured vs unstructured data; traditional systems vs big data; HDFS and MapReduce overview; why Spark is preferred for modern processing.

02

Apache Spark Essentials

3 hrs

[Spark architecture](#) · [Setup](#) · [Application lifecycle](#) · [Cluster modes](#)

Spark ecosystem; Spark vs Hadoop; Driver, Executors, Cluster Manager; SparkSession; local mode; Spark UI; basic Spark application lifecycle.

03

PySpark Core Concepts

4 hrs

[RDDs](#) · [DataFrames](#) · [Transformations](#) · [Actions](#) · [Lazy evaluation](#)

RDD creation; transformations and actions; lineage and fault tolerance; DataFrame creation; schemas; type-safe concepts at overview level; RDD and DataFrame interoperability.

04

Spark SQL & Data Processing

4 hrs

[Spark SQL](#) · [Joins](#) · [Aggregations](#) · [Data sources](#) · [Functions](#)

Read CSV, JSON and Parquet; select, filter, groupBy, join, union; SQL queries with spark.sql(); temporary views; schema inference and manual schemas; dates and timestamps; arrays and structs.

05

Azure Data Engineering Fundamentals

2.5 hrs

[Azure storage and analytics services](#)

Azure regions and availability zones; Entra ID and RBAC basics; ADLS Gen2 and Blob Storage; Azure SQL; Synapse Analytics overview; Event Hubs overview; batch vs streaming; Azure data engineering architecture.

06

Azure Data Factory

4.5 hrs

[Pipelines](#) · [Data movement](#) · [Orchestration](#) · [Monitoring](#)

ADF Studio; linked services; datasets; integration runtime; Copy Activity; pipelines and activities; parameters and variables; triggers and scheduling; incremental loads and CDC concepts; Mapping Data Flows; Databricks integration; monitoring, debugging and alerts; CI/CD overview.

Databricks, Fabric & Streaming

07

Azure Databricks

5.5 hrs

[Lakehouse](#) · [Clusters](#) · [Notebooks](#) · [Delta Lake](#) · [PySpark ETL](#)

Workspace and UI; notebooks; compute and cluster types; job clusters and scheduling; connecting to ADLS Gen2; PySpark ETL; Delta Lake; managed vs external tables; MERGE/UPSERT; partitioning; Auto Loader overview; Databricks SQL; Unity Catalog integration; practical Azure data pipeline.

08

Unity Catalog & Data Governance

1.5 hrs

[Data governance](#) · [Access control](#) · [Managed and external data](#)

Unity Catalog concepts; catalogs, schemas and tables; RBAC and permissions; managed vs external tables; basic secrets and Key Vault integration; governance and lineage overview.

09

Microsoft Fabric for Data Engineering

3.5 hrs

[OneLake](#) · [Lakehouse](#) · [Data Factory](#) · [Notebooks](#) · [Pipelines](#)

Fabric architecture; OneLake; Lakehouse vs Warehouse; Fabric Data Factory; pipelines; Dataflows Gen2; Spark notebooks; shortcuts; semantic model overview; Eventstream overview; monitoring; governance; positioning Fabric vs Azure Databricks.

10

Streaming Data Engineering

2.5 hrs

[Streaming concepts](#) · [Event Hubs](#) · [Structured Streaming](#)

Batch vs streaming; Azure Event Hubs; Structured Streaming; windows; checkpointing and fault tolerance; streaming ingestion into ADLS and Databricks; Fabric Eventstream awareness; simple real-time pipeline design.

MODULE 11 · 4.5 HOURS

End-to-End Capstone Project

Build a practical Azure data pipeline using ADF, ADLS Gen2 and Databricks/Delta Lake. Transform data with PySpark, apply basic governance, publish an analytics-ready dataset, then monitor it and present a full project walkthrough.

Ingestion



Storage



Transformation



Analytics

Start with the **early-bird batch**

★ **EARLY BIRD SALE**

~~₹39,999~~ **₹4,999** + 18% GST / \$59 USD

Limited early-bird seats · EMI available · Corporate and group rates on request. Fees include all course material, lab guidance and the capstone review.

Hands-on projects

Real-world work that mirrors industry challenges and builds a portfolio you can show.

Industry mentors

Taught by practitioners working on production systems every day.

Practical curriculum

Built around the tools teams actually deploy - not textbook theory.

Certificate on completion

Recognised by our industry partners, issued when you finish the capstone.

ENROL NOW

Let's get you started.

Call or write to us for the next batch dates, the detailed session plan, or a corporate training proposal for your team.

CALL

+91 87912 95762

EMAIL

contact@newbieget.com

WEB

newbieget.tech/training

Government of India **Udyam-registered MSME** · **GeM-registered** technology company