

Azure Data Engineer Training Curriculum

Designed For: Final-year students, fresh graduates, working professionals

Hands-on Data Projects, Cloud Experience, Industry-Level Data Engineering Skills

Mode: Online | Duration: 2 to 2.5 Months | Includes Live Projects

STAGE 1: PySpark & Big Data Processing

Work with large-scale datasets using Apache Spark via Python.

Modules:

- Spark Architecture & Concepts •
- PySpark DataFrames and SQL •
- ETL with PySpark
- Aggregations, joins, transformations

Hands-on:

ETL Pipeline with PySpark for batch processing

STAGE 2: Azure Data Services

Learn Azure cloud-based data storage, management, and access.

Modules:

- Azure SQL Database
- Table creation, indexing, procedures
- Cosmos DB (NoSQL)
- Partitioning, consistency models

Hands-on:

Querying, indexing, and partitioning in Azure SQL & Cosmos DB

STAGE 3: Azure Synapse Analytics

Build analytics pipelines using Synapse Studio.

Modules:

- Synapse Studio Overview
- Serverless vs Dedicated SQL Pools
- Integration with Azure Data Lake
- Query Performance Tuning

Hands-on:

End-to-end analytics workflow using Synapse + Data Lake

STAGE 4: Azure Data Factory (ADF)

Orchestrate and automate data workflows in the cloud.

Modules:

- ADF Pipelines: structure & triggers
- Data flows, Copy activity
- Linked services, Integration runtime
- Monitoring & Debugging

Hands-on:

Automated data movement using ADF pipelines

STAGE 5: Azure Databricks

Scalable data science and machine learning on Azure.

Modules:

- Databricks Notebooks & Clusters •
Delta Lake for storage & updates
- Integration with ADF
- ML pipelines using PySpark in Databricks

Hands-on:

Data exploration and ML model training in Databricks

STAGE 6: Stream Analytics & Real-Time Processing

Real-time data ingestion and event stream processing.

Modules:

- Azure Stream Analytics Basics
- Access control with Azure Active Directory
- Key Vault Integration

Hands-on:

Real-time pipeline with live dashboard output

STAGE 7: Version Control & DevOps Basics

Learn to manage code and collaborate in team environments.

Modules:

- Git & GitHub basics
- Branching, commits, collaboration workflow

Hands-on:

Versioned data pipeline scripts with Git

BONUS MODULE: SQL Fundamentals for Data Engineers

Deep SQL skills for querying and optimization.

Topics:

- Joins, Aggregations, Window functions
- Stored Procedures & Triggers
- Performance tuning techniques

Real-world SQL query challenges and tuning exercises

What You Will Achieve:

- 8+ Cloud Data Engineering Projects
- End-to-End Azure Data Pipeline experience
- GitHub Portfolio + Resume-ready Projects
- Hands-on in PySpark, ADF, Synapse, Databricks
- Placement-ready for Data Engineer & Azure roles

Empowering careers with real-world IT training—learn, build, and grow with Brolly Academy.

TOOLS AND PLATFORMS:

