Azure Data Engineer with Fabric Data

What is a Data Engineer?

A **Data Engineer** is someone who **collects, organizes, and prepares data** so that companies can use it to make decisions, build reports, and train AI/ML models.

Think of a Data Engineer like the person who:

- Brings data from different sources .
- Cleans and arranges the data so it makes sense.
- Builds pipelines to move data to storage systems.
- Makes the data available in a ready-to-use format for **analysts, data scientists, and business teams**.

In short 👉:

Data Engineers don’t just store data — they make sure the right data reaches the right place, in the right format, at the right time.

Why is Data Engineering Important?

Every company today generates a **huge amount of data –** from apps, websites, sensors, payments, customer activity, and more.
But **raw data is messy**. It can't be used directly. That's where **Data Engineers** come in.

Importance of Data Engineering:

- **Organizes data** → turns raw data into clean, structured information.
- **Makes data fast & accessible** → so companies can get real-time insights.
- **Enables decision making** → managers, analysts, and AI models rely on well-prepared data.
- **Supports AI & ML** → without good data pipelines, AI systems can't learn properly.
- **Drives business growth** → companies use data to improve products, cut costs, and serve customers better

In today's world, data is like "oil," and Data Engineers are the ones who refine it into fuel.

Why Choose Azure Cloud for Data Engineering?

Microsoft is a global leader in enterprise solutions — the same technology that powers Office 365, LinkedIn, Xbox, and Teams is available for companies through **Azure** and the modern **Microsoft Fabric** platform.

Why Azure + Fabric for Data Engineering:

- **Azure Synapse Analytics** → scalable cloud data warehouse for batch and interactive analytics.
- **Microsoft Fabric (Lakehouse + Warehouse)** → unified data platform combining **data lake + data warehouse** with real-time analytics.
- **Azure Data Factory & Fabric Data Pipelines** → powerful ETL/ELT tools for orchestrating both batch and streaming pipelines.
- **Azure Databricks + Delta Lake** → enterprise-scale Spark platform for big data processing and machine learning.
- **Azure Event Hub & Stream Analytics** → real-time ingestion and streaming analytics for IoT and event-driven systems.
- **Azure SQL, Cosmos DB, Synapse, and OneLake in Fabric** → databases for every type of workload (transactional, NoSQL, analytical).
- **Enterprise Security & Governance** → Azure Purview, RBAC, Key Vault, and Managed Identity for compliance and data protection.
- **Global Azure Cloud Infrastructure** → high availability, low latency, and trusted by Fortune 500 enterprises worldwide.

👉 Simply put:

If you want to design **modern, enterprise-ready data pipelines at scale**, Azure with **Microsoft Fabric** provides the most **integrated and future-proof ecosystem** for data engineering and analytics.

Career Demand & Salaries for Data Engineers

- One of the **fastest-growing roles worldwide**, as companies move their analytics and reporting to **Azure + Microsoft Fabric**.
- Needed across every industry — finance, healthcare, e-commerce, entertainment.
- Average Salaries:
  - India: **₹12 – 25 LPA** (mid-level), **₹30 – 45 LPA+** (experienced).
  - US & Global: **$110K – $160K per year**.
- At i27Academy, many students in our Cloud & DevOps tracks have secured **30–60 LPA packages**. With Data Engineering, the opportunities are even bigger.
- With the arrival of **Microsoft Fabric**, demand for **Fabric-ready Data Engineers** is surging — making this one of the **hottest skills for the next decade**.

"If Cloud is the backbone of IT, Data Engineering is the brain that powers decisions"

Why i27Academy for Azure Data Engineer (with Microsoft Fabric)?

- **Cutting-Edge Start** → We built our Azure Data Engineer program around **Microsoft Fabric**, the newest and most in-demand data platform.
- **Industry Expertise** → Program designed and delivered by **real experts working on Azure Data projects day-to-day**.
- **Specialized Focus** → We don't do generic data training — our track is **laser-focused on Azure Data + Fabric**.
- **Trusted Career Path** → Our students already work in top MNCs with **30–60 LPA packages**, and ADE with Fabric opens even bigger opportunities.
- **Proven Methodology** → The same structured training that made i27Academy a leader in **Cloud & DevOps** is now applied to **Azure Data Engineering**.

👉 At i27Academy, we don't just teach Fabric and Azure services — **we prepare you to design and run production-grade data pipelines from Day 1**.

Who Can Join This Course?

This program is ideal for professionals and aspirants such as:

- **Database Engineers** → looking to move beyond traditional RDBMS into Azure SQL, Synapse, and Fabric Lakehouse
- **Big Data / Hadoop Engineers** → wanting to upskill into Azure Databricks, Delta Lake, and Fabric pipelines
- **ETL / Data Warehouse Engineers** → ready to design scalable pipelines with Azure Data Factory & Fabric Dataflows

- **Application Programmers** → interested in working with real-time data (Event Hub, Stream Analytics) and analytics
- **Test Engineers** → exploring a career transition into high-growth Data Engineering roles on Azure
- **Data Analysts / Power BI Developers** → who want to step into end-to-end engineering and automation with Fabric

👉 Even **fresh graduates** with interest in **Cloud & Data** can join, as we start from fundamentals and build step by step.

Course Content:

Module 1: Data Engineering Introduction:

- What is data engineering
- Introduction to Azure
- Comparison of Azure vs other cloud services
- Bigdata introduction

Module 2: SQL for Data Engineers

- SQL Fundamentals
  i. DDL (CREATE, ALTER, DROP)
  ii. DML (INSERT, UPDATE, DELETE)
  iii. TCL (COMMIT, ROLLBACK, SAVEPOINT)
  iv. DCL (GRANT, REVOKE)
- Filtering & Expressions
  i. WHERE, ORDER BY, GROUP BY, HAVING
  ii. String, Date, Format & Cast Functions
  iii. Conditional Expressions (CASE, COALESCE, NULLIF)
- Joins
  i. INNER JOIN
  ii. LEFT JOIN / RIGHT JOIN
  iii. FULL OUTER JOIN
  iv. CROSS JOIN
  v. SELF JOIN
  vi. ANTI JOIN / SEMI JOIN
- Advanced Queries
  i. Subqueries (Scalar, Correlated, Nested)
  ii. Common Table Expressions (CTEs)
  iii. Set Operators (UNION, UNION ALL, INTERSECT, EXCEPT/MINUS)
- Aggregations
  i. COUNT, SUM, AVG, MIN, MAX
  ii. GROUPING SETS, ROLLUP, CUBE
- Window Functions
  i. RANK, DENSE_RANK, ROW_NUMBER
  ii. PARTITION BY & ORDER BY in Window Functions
- Performance & Optimization
  i. Indexes & Partitioning
  ii. Query Optimization Techniques
  iii. Explain Plans & Execution Strategies

Module 3: Python

- Python Overview
  i. History & features of Python
  ii. Python vs Other Programming Languages

iii. First Python Program
iv. Python basic syntax
v. Python Development Tools & Packages

- Python Environment Setup
  i. Installing Python
  ii. Verify & Setup Python environment
  iii. Online editors
- Python DataTypes
  i. Variables
  ii. Data Types
  iii. Strings
  iv. Type Casting
- Python Operators
  i. Arithmetic Operators
  ii. Relational Operators
  iii. Logical Operators
  iv. Bitwise Operators
  v. Assignment Operator
- Python Flow Control & Loops
  i. if, if else, if else if
  ii. while, do while loops
  iii. for loops
- Python Functional Programming
  i. Function Declarations
  ii. Calling Functions
  iii. Functions Call-by-Name
  iv. Functions with Named Arguments
  v. Functions with Variable Arguments
  vi. Functions with Default Parameter Values
  vii. Lambda Functions
- Python Collections
  i. Lists
  ii. Tuples
  iii. Sets
  iv. Dictionaries
- Python Files I/O
  i. Reading input from console
  ii. Reading data from Files
  iii. Writing data to Files
- Python Object Oriented Programming
  i. Python Classes
  ii. Simple class
  iii. Class objects
  iv. Inheritance

Module 4: Azure Data Factory:

- Introduction to Azure Data Factory
- Top level Concepts in Azure Data Factory
- Create your First Azure Data Factory
- Different ways to work with Azure Data Factory
- Pipelines and Activities
- Linked Services and Datasets
- Triggers Schedule Trigger
- Tumbling Window Trigger
- Tumbling Window Trigger
- Dependency Event based Triggers
- Integration runtime
- Azure Integration runtime
- Self-Hosted Integration runtime
- Setting up Self Hosted Integration runtime
- Shared Self Hosted Integration runtime
- Parameterize Linked Services
- Parameterize Datasets
- Parameterize Pipelines
- System Variables
- Connectors Overview
- Supported File Formats
- Copy Data Activity
- Monitor Copy Data Activity
- Delete Activity
- Variables
- Set Variable Activity
- Append Variable Activity
- User Properties
- Execute Pipeline Activity
- Filter Activity
- ForEach Activity
- Get Metadata Activity
- If Condition Activity
- Wait Activity
- Until Activity
- Web Activity
- WebHook Activity
- Switch Activity
- Validation Activity
- Lookup Activity
- Transform Data Activities Overview
- Stored Procedure Activity Data flow
- Mapping Data Flow
- Data Flow Activity
- Mapping Data Flow Debug Mode

- Filter Transformation in Mapping Data Flow
- Aggregate Transformation in Mapping Data Flow
- JOIN Transformation in Mapping Data Flow
- Conditional Split Transformation in Mapping Data Flow
- Derived Column Transformation in Mapping Data Flow
- Exists Transformation in Mapping Data Flow
- Union Transformation in Mapping Data Flow
- Lookup Transformation in Mapping Data Flow
- Sort Transformation in Mapping Data Flow
- New Branch in Mapping Data Flow
- Select Transformation in Mapping Data Flow
- Pivot Transformation in Mapping Data Flow
- Unpivot Transformation in Mapping Data Flow
- Surrogate Key Transformation in Mapping Data Flow
- Window Transformation in Mapping Data Flow
- Alter Row Transformation in Mapping Data Flow
- Flatten Transformation in Mapping Data Flow
- Parameterize Mapping Data Flow
- Validate Schema in Mapping Data Flow
- Schema Drift in Mapping Data Flow
- Wrangling Data Flow Overview
- Merge Queries in Wrangling Data Flow
- Group By in Wrangling Data Flow
- Different Author Modes
- Setup GitHub Code Repository for Azure Data Factory
- Setup Azure DevOps Git Code Repository
- Use Azure Key Vault Secrets
- Continuous integration and deployment
- How to read JSON output of one Activity in to another Activity
- Annotations
- Templates Overview
- Global Parameters
- Rank Transformation in Mapping Data Flow
- Cache Sink and Cached lookup in Mapping Data Flow
- Session log in Copy Activity | Log Copied File names in Copy Activity
- Write Cache Sink to Activity Output
- Parse Transformation in Mapping Data Flow
- Fail Activity
- Inline Dataset
- Stringify transformation in Mapping Data Flow
- Assert Transformation in Mapping Data Flows
- Flowlets in Mapping data flow
- Script Activity in Azure Data Factory or Azure Synapse
- User defined Functions in Mapping data flows
- Fuzzy Joins Using mapping data flows
- Parameterize Linked Services using
- Cast Transformation in Mapping data flows

- Extract Data from table of website page
- Per Pipeline Billing View for Azure Data factory
- Time To Live(TTL) Setting in Azure IR to reduce cluster spin up time for dataflows
- Create Alert rules in Azure Data factory for Pipeline or activity Failures
- Pipeline return value in Set variable

Module 5: Azure Synapse Analytics:

- Introduction to Azure Synapse Analytics
- Create Azure Synapse Analytics Workspace
- Basic Concepts in Azure Synapse Analytics
- Analyze data with a server less SQL pool and dedicated SQL Pool
- Analyze data with Server less Spark Pool
- Analyze data in Storage Account in Azure Synapse Analytics
- Integrate Pipelines in Azure Synapse Analytics
- Monitor your Azure Synapse Analytics Workspace
- Add an Administrator to your Azure Synapse Workspace
- Azure Synapse SQL Architecture Distributions(Hash, Round Robbin & Replicate)
- Server less SQL Pool Overview
- Create External Data source
- Create External File Format
- CETAS with Synapse SQL CTAS with Synapse SQL
- External Tables with Synapse SQL
- Create and query external tables from a file in ADLS
- Types of External Tables(Hadoop & Native) in Synapse SQL
- Administrative accounts in Synapse SQL
- Create Login and User for Server less SQL Pool
- Create Login and User for Dedicated SQL Pool

Module 6: Azure Data Bricks:

- Introduction to Azure Databricks
- Create an Azure Databricks Workspace using Azure Portal
- Create Databricks Community Edition Account
- Workspace in Azure Databricks
- Workspace assets in Azure Databricks
- Working with Workspace Objects in Azure Databricks
- Create and Run Spark Job in Databricks
- Azure Databricks architecture overview
- Databricks File System(DBFS) overview in Azure Databricks
- Databricks Utilities(dbutils) in Azure Databricks
- Data Utility(dbutils.data) in Azure Databricks in Databricks utilities
- File System utility(dbutils.fs) of Databricks Utilities in Azure Databricks
- exit() command of notebook utility(dbutils.notebook) in Azure Databricks
- run() command of notebook utility(dbutils.notebook) in Databricks Utilities in Azure Databricks
- Widgets utility(dbutils.widgets) of Databricks Utilities in Azure Databricks
- Pass values to notebook parameters from another notebook using run command in Azure Databricks

- Parameterize SQL notebook using widgets in Azure Databricks | Widgets in SQL in Azure Databricks
- Create Mount point using dbutils.fs.mount() in Azure Databricks
- Mount Azure Blob Storage to DBFS in Azure Databricks
- Delete or Unmount Mount Points in Azure Databricks
- mounts() & refreshMounts() commands of File system Utilities in Azure Databricks
- Update Mount Point(dbutils.fs.updateMount()) in Azure Databricks
- Secret Scopes Overview in Azure Databricks
- Install Databricks CLI and configure with your workspace | Azure Databricks
- Create an Azure Key Vault backed secret scope using the UI in Azure Databricks
- Create a Databricks backed secret scope in Azure Databricks
- Secrets Utility(dbutils.secrets) of Databricks Utilities in Azure Databricks
- Access ADLS Gen2 storage using Account Key in Azure Databricks
- Configure access to Azure storage with an Azure Active Directory service principal
- Access Data Lake Storage Gen2 or Blob Storage with an Azure service principal in Azure Databricks
- Access ADLS Gen2 or Blob Storage using a SAS token in Azure Databricks

Module 7: Spark

- Spark Introduction
- Spark Overview
- Spark features
- Spark vs Hadoop MapReduce
- Programming Language choices in Spark
- Spark History
- Spark use cases

Module 8: Spark Components or modules

- Spark Core
- Spark SQL
- Spark Streaming

Module 9: Spark Architecture

- Spark Application flow
- Spark Driver, Executors
- Spark Context, Spark Session
- Spark dependency on Cluster Managers
- Spark execution modes
- Standalone cluster mode
- Spark on YARN mode

Module 10: Spark Core

- Spark's main Data Abstraction
- About RDD
- RDD Features
- Creating RDDs
- Saving Files
- Data Manipulation using RDDs
- Transformations & Actions
- RDD Partitions & Coalesce
- Memory Management: cache & persist
- Data Loading and Saving through RDDs
- Aggregations, Joins through RDDs
- RDD Advanced concepts – Accumulators, Broadcast variables

Module 11: Spark Structured APIs

- DataFrames
- Columns
- Rows
- Spark Types
- Performance optimization with Spark Structured APIs
- Logical planning, Physical planning, and Execution

Module 12: Spark DataFrames

- DataFrames basics
- Creating DataFrames
- Schemas
- DataFrame Operations
- Column wise operations
- Row wise operations
- Aggregations DataFrames
- Joins using DataFrames

Module 13: Spark Data Sources

- Reading & writing different files formats
- CSV Files
- JSON Files
- Parquet Files
- ORC Files
- SQL Databases
- TextFiles

Module 14: Spark SQL

- Bigdata & SQL: Apache Hive vs Spark SQL
- Catalog, Tables, Views, Databases
- Data selection and manipulation using Spark SQL
- User Defined Functions
- Spark SQL integration with Hive

Module 15: Spark Streaming

- About Batch vs Streaming processing
- About Spark DStreams
- About Spark Structured streaming
- Transformation of Streams data
- Streaming sources & sinks
- Event Time Stateful processing

Module 16: Azure Storage and Data Lake:

- Detail knowledge of Blob
- Create a container
- Upload a blob to Azure Storage
- Create a Blob List of the blobs in a container
- Delete a container
- Download the blob to your local computer
- Detail knowledge of Data Lake
- Create Azure Data Lake Gen 2 Account
- Create Folders
- Upload data
- Secure data
- Delete Azure Data Lake

Module 17: Azure Key Vault:

- Introduction to Azure Key Vault
- Store Secrets in Azure Key Vault using Azure Portal

Module 18: Delta Lake usage in Databricks:

- Architecture
- Storage Understanding
- Table creation and API options
- DML Operations usage.
- Partitions
- Schema Enforcement
- Schema Evolution
- Versions
- Time Travel
- Vacuum
- Delta Lake Merge (SCD Type 1 and SCD Type2)

Module 19: Microsoft Fabric:

- Introduction to Microsoft Fabric
- create the workspace in Fabric and Build a lake house in fabric.
- Install One Lake Explorer &Data Studio
- Create Your First warehouse in Fabric | | Lakehouse vs Warehouse
- Microsoft Fabric: Introduction to Data Factory Experience | Pipelines & Dataflows
- Apache spark in Fabric
- Work with Delta Lake tables in Microsoft Fabric
- Copy data from REST API to Lakehouse in Fabric using Bearer Token
- Moving Multiple Folders from Data Lake Storage Gen2 to Lakehouse
- Ingest Data from On Premise SQL Server to Lakehouse| Data Gateway
- Microsoft Fabric: Copy data from Azure SQL DB to Lakehouse Table | Query | Stored proc
- Microsoft Fabric: Copy Multiple tables from Azure SQL DB to Lakehouse dynamically
- Ingest data from on premise SQL server to Warehouse Incrementally
- Microsoft Fabric: Logging Fabric Pipeline Logs to Warehouse table using stored proc
- End-to-End Scenario with Data Factory Pipelines and Dataflows Gen2

Module 20: Microsoft Certified: Fabric Data Engineer Associate Preparation

\

Capstone Projects Implementation

- Mini projects and Capstone projects on Azure data engineer and fabric data

Career Readiness & Next Steps

Job Profiles to Target

- Azure Data Engineer
- Fabric Data Engineer
- Big Data Engineer (Azure/Fabric focus)
- Data Platform Engineer (Lakehouse + Synapse/Fabric)

Interview Preparation

- Azure + Fabric Data Engineer Interview Questions topic wise discussion
- Scenario-based real-world discussions (which we face in realworld)

All The Best!