Upskill ACADEMY
SKILLUP, STANDOUT

CLOUD AND BIG DATA ENGINEER
MASTER PROGRAM

Become a Data Engineer with in-demand tools and Land high-paying jobs at top tech Companies

INDEX

3 Our Mission & Visions

Our Mission

Our mission is simple — to help you upgrade your knowledge and stay relevant in the fast-changing IT industry. As technology evolves, staying updated is not just important, it's necessary to survive and grow. With our program, you'll build the right skillsets to sustain your career confidently.
A simple rule we follow: The more you learn, the faster you earn.

Our Vision

Our vision is to UPSKILL Professionals and prepare them for real-world industry needs. We want to build a strong, skilled workforce that not only grows in their careers but also contributes to the growth of the country. Our goal is to become a trusted name in job-oriented, practical training that makes a real difference in people's lives.

4 About the Program

The Data Engineer Job-Oriented Program is a complete training course designed to get you job-ready for the most in-demand roles in the data industry.

You'll learn top tools like Microsoft Azure, Fabric, Spark, PySpark, Python, SQL, Databricks, Data Factory, and Synapse Analytics — all through hands-on labs and real-world projects.
We don’t just teach — we guide. You'll get:

01 **1-on-1 mentorship from experienced Data Engineers**

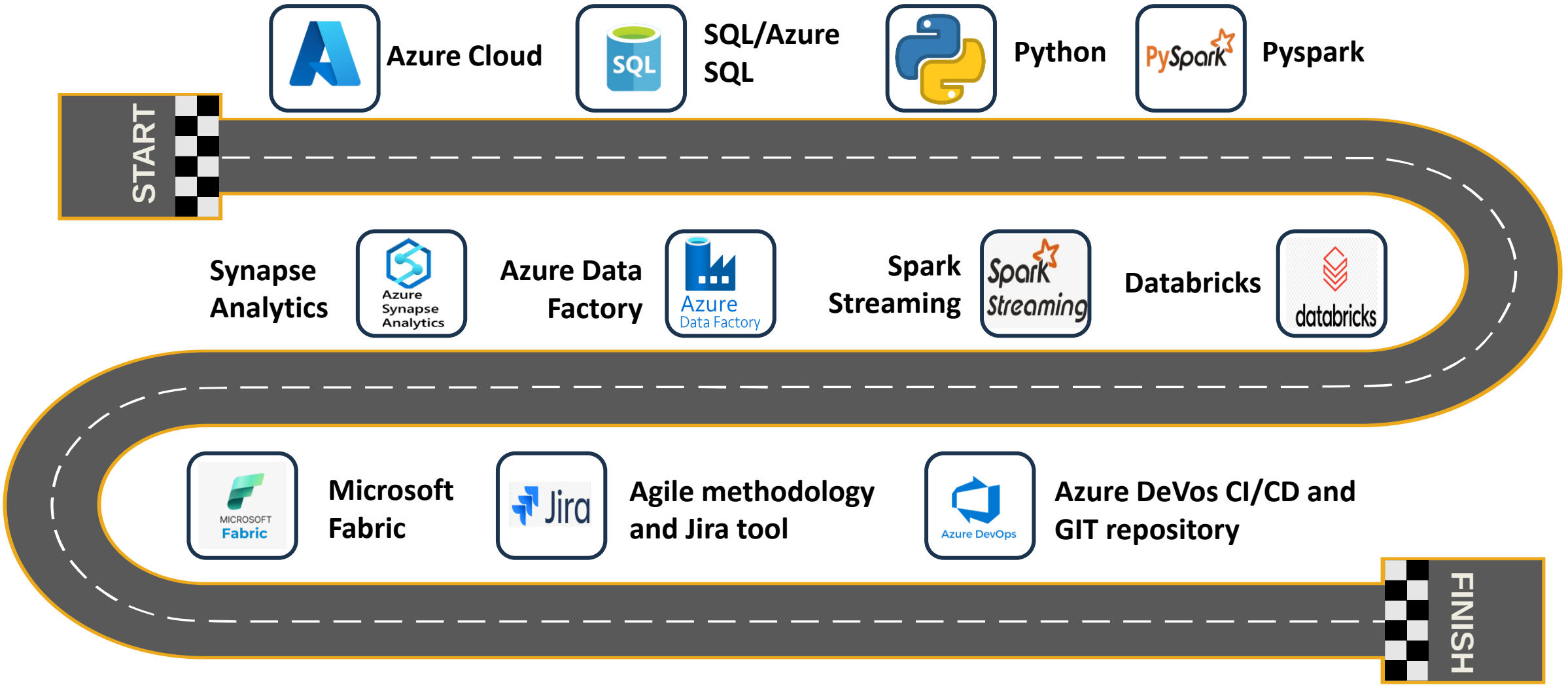
02 **Support with your resume, LinkedIn profile and interview preparation**

03 **Experience working on live, cloud-based data projects**

04 **Skills for roles like Azure Data Engineer, Data Integration Specialist, Cloud Data Engineer Data Architect, and more**

This program is your step-by-step path to a high-growth career in data. Want to learn more or get started? Book a free call today!

5 Roadmap


Azure Cloud
SQL/Azure SQL
Python
Pyspark
START
Synapse Analytics
Azure Synapse Analytics
Azure Data Factory
Azure Data Factory
Spark Streaming
Spark Streaming
Databricks
databricks
Microsoft Fabric
MICROSOFT Fabric
Agile methodology and Jira tool
Jira
Azure DevOps
Azure DevOps
Azure DeVos CI/CD and GIT repository
FINISH

06 Tools Covered

Azure cloud

Azure SQL

Python

PySpark

Databricks

Spark Streaming

Azure Data Factory

Azure Synapse Analytics

Microsoft Fabric

Jira

Azure DevOps

07 DATA ENGINEER CURRICULUM

Module 1: Introduction to Big Data

1.1-What is Data

1.2-What is database

1.3 Types of databases

1.4-What is Big Data

1.5- 5 V’s of Big Data

1.6- Types of data (structure, semi-structure, unstructured)

1.7- Store and process big data

Module 2: Azure Cloud Fundamentals

2.1-What is cloud computing?

2.2 Categories of cloud services:

2.3- Features of cloud:

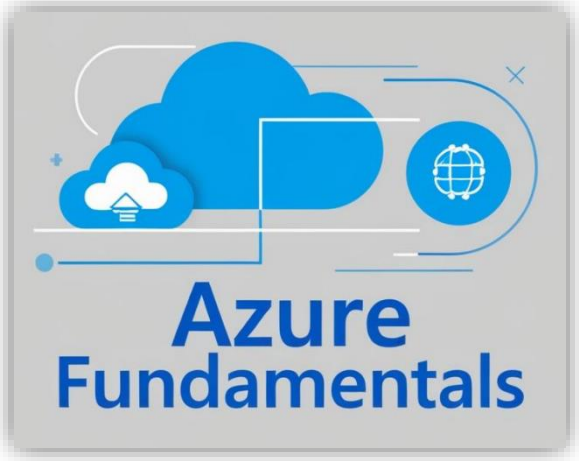
2.4- Different cloud providers, Azure, AWS, GCP

2.5- Steps to create an Azure account

2.6-Azure Subscription

2.7- Accessing the Azure account:

2.8- Create Azure Resource Group Manager



2.9- Overview of Azure Resources / Services

- Data Factory
- Azure Data bricks
- BLOB Storage, Data Lake Storage Gen1 and Gen2
- Azure SQL Server, SQL Database
- Key Vault
- Logic Apps
- Synapse Analytics

Module 3: Data Storage on Azure Cloud

3.1- Creating a storage account in Azure

3.2- Create Blob Storage and a container, upload files

3.3-Create Data lake Gen2 and upload files

3.4-Difference between blob storage and ADLS GEN2

3.5-Access tiers

3.6- Data Replication Policies

3.7- Set up an Azure storage account with a different tier

3.8- Maintain a version of files in the storage account whenever a file is modified.

3.9- exclusive access to the file for one Application

3.10 Version control of files in Blob and ADLS

Module 4- SQL and Azure SQL

4.1-What is SQL

4.2- Installation of Azure SQL Database

4.3- Connect Azure SQL from local SQL Server Management Studio

4.4- DQL Commands (select)

4.5-DDL commands (create, alter, drop, truncate)

4.6-DML Commands (insert, update, delete, merge)

4.7-Joins in SQL

4.8-Window functions

4.9-Aggregate functions

4.10-CTE (Common table expression)

Module 5- Python

5.1-What is python

5.2-Why python for Data Engineers

5.3-Advantage of Python

5.4-Installation of Python

5.5-Variables

5.6-DataTypes

5.7- Collections (List, Tuple, Set, Dictionary)

5.8- If-Else statements

5.9- Loops

5.10- Functions

5.11-Operators

Module 6- Introduction to Hadoop

6.1- What is Hadoop?

6.2- How Hadoop overcomes big data challenges

6.3-Hadoop Architecture

6.4-Hadoop Daemons

6.5-HDFS

6.6-YARN

6.7-MapReduce

6.8- Hadoop Ecosystem Components

Module 7-Introduction to Spark

7.1-What is Spark

7.2-What is pyspark

7.3 Challenges of Hadoop

7.4-Why spark needed

7.5-Spark Architecture & Internals7.4-Spark internals

7.6- Spark Data Frame and datasets

Module 8-Introduction To Databricks

8.1- What is Databricks?

8.2- Difference between Spark vs Databricks

8.3- Databricks Architecture

8.4- Creating a Databricks community account

8.5- Creating an Azure Databricks account

8.6- Databricks UI walkthrough

8.7- Creating a notebook in Databricks

8.8- Creating a compute cluster in Databricks

Module 9- Working with Databricks Filesystem – DBFS

9.1-Understanding Databricks File System (DBFS)

9.2-DBFS commands – mkdirs, cp, mv, head, put, rm, rmdir

9.3-How to handle multiple files in DBFS

9.4- Loading data in Databricks

9.5- Running the first notebook in Databricks

Module 10: Spark Data Frame Reading APIs and Databricks

10.1-Introduction to Spark DataFrames

10.2-What is Spark DataFrame APIs?

10.3- Reading Files in Spark and Databricks

10.4- Reading CSV files:

- Without header
- With header
- With schema
- Multiline CSV files

10.5- Reading JSON files:

- Standard JSON
- Nested JSON

10.6-Reading Parquet files

10.7-DataFrame Read Modes

- Permissive
- Drop Malformed
- Fail Fast

10.8- Data frame Modes

- error or errorifexists
- Append
- Overwrite
- ignore

10.9-Schema Management

10.10-Challenges of schema inference

10.11- Introduction to schema enforcement

- Schema DDL
- Struct type

Module 11: Spark Data Frame Transformation and Databricks

11.1-Data Transformation

- Adding new columns (with Column)
- Renaming columns
- Dropping columns
- Derived columns

11.2-Handling Nulls

- Dropping null values
- Filling null values

11.3-Filtering and Sorting

- Accessing columns in multiple ways

- Filtering data using where/filter
- Select vs select expression
- Sorting data

11.4-Grouping and Aggregations

- Grouping data
- Deduplication
- Union

11.5-String and Date Manipulation

- Handling date manipulation in Spark
- Handling string manipulation in Spark
- Handling timestamp manipulation in Spark

11.6- Types of joins:

- Left join
- Right join
- Full outer join
- Inner join

11.7-Displaying Data

- Show vs display
- Limit/take

Module 12: Spark SQL in Databricks

12.1-Introduction to Spark SQL APIs in Apache Spark

12.2- Creating temporary views on DataFrames

12.3- Using Spark SQL for transformations

12.4-Global temporary views:

12.5-Difference between temporary views vs global temporary views

12.6-Creating Spark SQL tables from DataFrame and vice versa

12.7-Creating Spark tables

12.8- Types of tables:

12.9- Use cases of DataFrames & Spark SQL

12.10- Understanding metastore in detail

Module 13: Mounting Storage in Databricks

13.1-What is mounting?

13.2- Why is mounting needed?

13.3-How to mount ADLS to Databricks?

13.4- Checking existing mount points

13.5- Mounting via SAS token

13.6- Mounting via registered application

13.7-Unmounting storage

13.8- Secret Scope in Databricks

Module 14: Databricks Utilities and Widgets

14.1- Reading data from ADLS in Databricks

14.2- Writing data from Databricks to ADLS

14.3- Databricks magic commands:

14.4-Databricks widgets:

14.5-Databricks utilities commands

14.6-Databricks help commands

14.7-DBFS file explorer

Module 15- Notebook Integration & Connectivity

15.1- Calling one notebook from another

15.2- Calling a Databricks notebook from Azure Data Factory

15.3- Connecting Azure SQL Database to Databricks

Module 16: Databricks Cluster Management

16.1- Azure Databricks cluster:

16.2- Databricks architecture - control and data plane

16.3- Pricing in Azure Databricks

16.4- When to use different cluster modes

16.5- Databricks benefits

16.6- Optimized cluster types:

16.7- Adding external libraries in Databricks

16.8- Setting Spark configurations in Databricks cluster

16.9- Deleting the Databricks cluster

Module 17: Spark Join Optimization and Performance Tuning

17.1- Why avoid shuffling?

17.2- Hash partitioner

17.3-Local aggregation

17.4 Challenges of wide transformations

17.5- Data skew and resolving skewed data

17.6-Repartitioning vs coalesce

17.7- AQE (Adaptive Query Execution)

17.8-Optimizing joins:

17.9- Optimizing the join of two large tables using bucketing

17.10- Spark optimization techniques:

17.11-Spark executors:

Module 18: Spark Performance & Memory Management

18.1 Memory management in Apache Spark

18.2-Sort aggregate vs hash aggregate

18.3- Various plans in Apache Spark:

18.4-Catalyst optimizer

18.5-Introduction to file formats & compression techniques:

18.6-Compression techniques:

18.7- Accessing Spark UI and resource manager

18.8-Understanding Spark UI

18.9-Understanding serialization & deserialization

18.10-Importance of caching

18.11- Practical applications of caching

18.12-Cache technique & persist in Spark

18.13- Storage levels provided by persist

18.14- Spark submit command understanding

Module 19: Delta Lake & Data Lakehouse

19.1-Understanding data lake

19.2-Understanding Delta Lake

19.3-Understanding data lakehouse

19.4-Difference between Data Warehouse vs Data Lake vs Data Lakehouse

19.5- Creating tables in Delta Lake

19.6- Inserting data into the Delta table

19.7- Updating data into the Delta table

19.8- Deleting data from the Delta table

19.9-Time travel in Delta Lake

- History based on version
- History based on timestamp

19.10-Writing DataFrame as files in Delta Lake format

19.11- Explanation of three-layer / medallion architecture

- Bronze layer
- Silver layer
- Gold layer

19.12- Optimization techniques in Delta Lake

- Compaction
- Z-ordering
- Vacuum command

19.13- Restoring data to an older version

Module 20: Databricks New Features

20.1-Introduction to Delta Live Tables

20.2- SQL Warehouse in Databricks

20.3-Scheduling Databricks notebooks

20.4- Git integration in Databricks

20.5-DevOps for Databricks

20.7- Databricks cost optimization strategies

20.8- End-to-end integration of Databricks with Power BI

Module 21: Spark Streaming

21.1-Introduction to Spark Streaming

21.2- Streaming data using files

21.3- End-to-end practical implementation of Spark Streaming

Module 22: Unity Catalog in Databricks

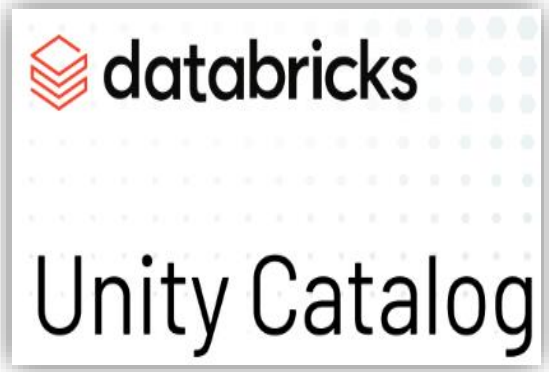
22.1- Unity catalog introduction

22.2- Key features of Unity Catalog

22.3- Unity Catalog object model

22.4- Internal vs. External tables

22.5- End-To-End Implementation of Unity Catalog



Module 23- Databricks Workflows

23.1- Introduction of Databricks workflows

23.2- Jobs in Databricks Workflows

23.3- Cluster types in workflows

Module 24-Azure Data Factory

24.1- Introduction

- Understanding Azure Data Factory
- Creating an Azure Data Factory account

24.2-Data Factory Components

24.3-First Pipeline: Azure SQL DB to ADLS

24.4-Dynamic Configurations

24.5-Copy Activity Details

24.6-Activities

24.7-Incremental Pipeline (High Water Mark) using data flow

24.8- SCD Types in Data Flows

24.9-Debugging & Monitoring

24.10 Triggers in Data Factory

24.11-Variables & Parameters

24.12-Integration Runtime

24.13-Production & Optimization

24.14-DevOps & ADF

24.15-Alerts & Monitoring

Module 25: Azure Key Vaults

25.1-Understanding the need for key vaults

25.2-How key vaults work

25.3- Creating a key vault account

25.4- Setting secrets in key vault

25.5- Accessing key vaults via ADF and Databricks

Module 26- Data Warehousing

26.1- What is a Data Warehouse

26.2- OLTP VS OLAP

26.3- Extract, Transform, Load (ETL)

26.4- Extract, Load, Transform(ELT)

26.5- Data Lake VS Data Warehouse

26.6- Star Schema

26.7- Snowflake Schema

26.8-Fact Table

26.9- Dimension Table

26.10- Data Marts

Azure Synapse Analytics

26.11- Slowly changing dimensions(SCD)

26.12-What is data modeling?

26.13-Understanding dimensional modeling

Module 27: Microsoft Fabric

27.1-What is Microsoft Fabric?

27.2- Creating a Fabric account

27.3- Building a data pipeline in Fabric

27.4-Creating tables in Fabric

27.5-Hands-on Fabric exercises

Module 28- Introduction to Agile Methodology

28.1-What is Agile Methodology?

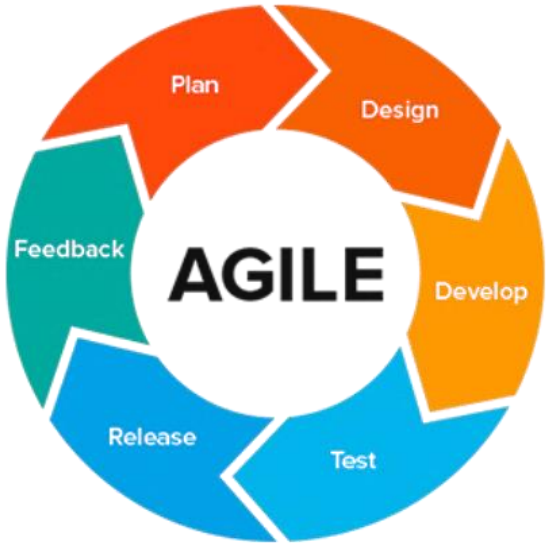
28.2- Why Agile is Important in Modern Data Projects

28.3-Early Software Development Models: Waterfall Approach

28.4- Limitations of Waterfall in Data Projects

28.5-Agile Events and meetings

- Sprint Planning
- Daily Stand-ups (Scrum Meetings)
- Sprint Review
- Sprint Retrospective
- Grooming call



Module 29- Introduction to JIRA for Agile

29.1-What is JIRA?

29.2-Basic JIRA Concepts

29.3- JIRA Workflow Stages

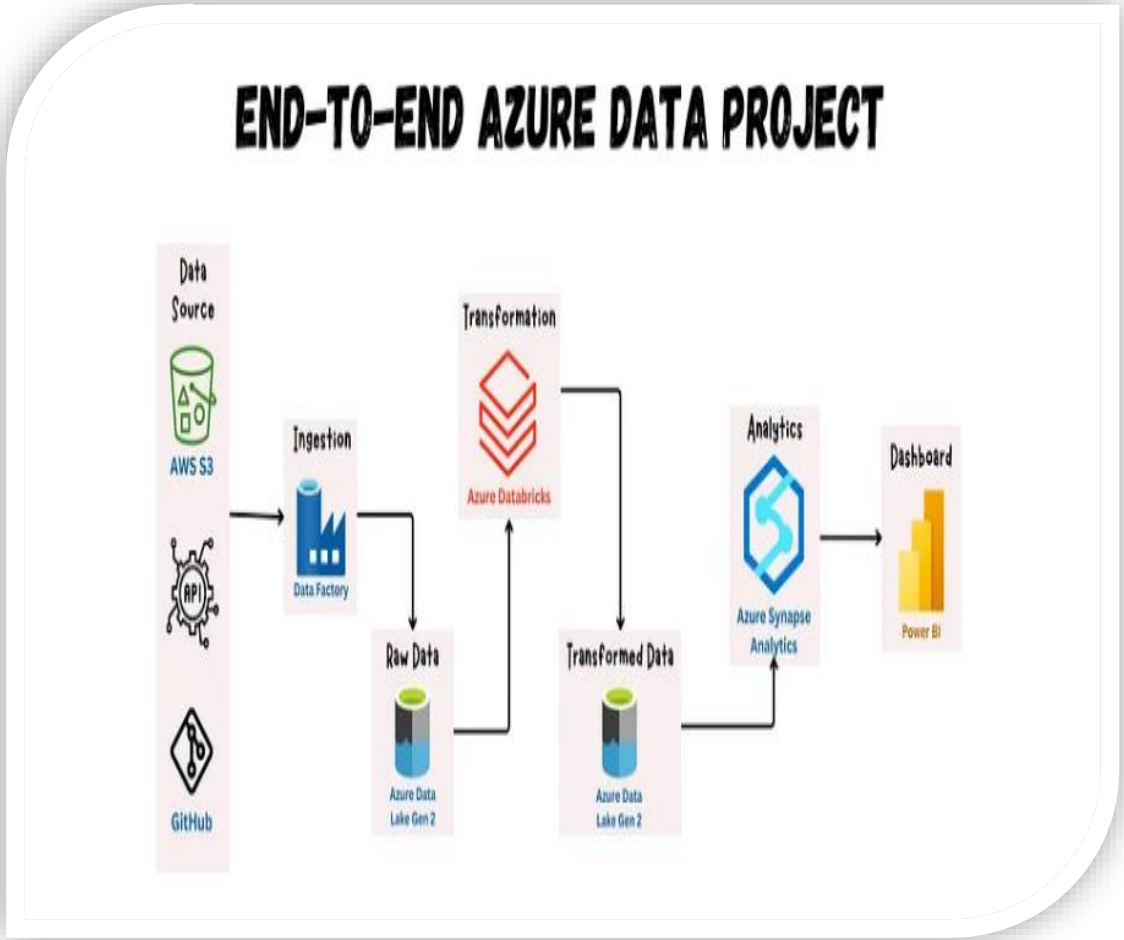
29.4-Sprint Board Overview

29.5- Using JIRA for Data Engineering

8 PROJECT

End-to-End Data Engineering Project Overview

- Data Lakehouse & Medallion Architecture Introduction
- Data Ingestion using Azure Data Factory (ADF) from Web URL
- Connect ADLS with Azure Databricks
- Advanced Transformations with Databricks
- Clean and Store Data in Silver Layer
- Aggregations and Business Logic for Gold Layer
- Data Modelling using Synapse Serverless SQL Pool
- Build Dashboards with Power BI



09 PLACEMENT SUPPORT

- LinkedIn/Naukri profile optimization
- Resume-building strategies
- Resume optimization for job applications
- Interview preparation guidance
- Salary and package negotiation strategies
- Handling common interview questions
- Handling HR interview questions
- Handling Managerial Round Interview Strategies
- Strategies and guidance for handling career gap interviews.

10 KNOW YOUR MENTOR

Sr. Data Engineer

Hi, I'm **Gaurav Kurhekar**, a Senior Data Engineer with 9+ years of experience in the IT industry, working across diverse domains and real-world data projects.

I'm passionate about helping aspiring professionals **kickstart their Data Engineering** journey by turning complex topics into simple, hands-on concepts through a practical approach.

I'm a **Microsoft and Databricks certified** professional with credentials in:

- Microsoft Certified: **Azure Fundamentals**
- **Databricks Fundamentals**
- **Databricks Certified Data Engineer Associate**

What sets my teaching apart?

End-to-End Ownership: I guide you through every stage of the project—from data ingestion to visualization—ensuring a clear understanding of how all tools work together.

Real-World Integration: Unlike other programs with multiple tutors for different topics, I teach every tool and technique myself. Why? Because in real jobs, **a Data Engineer handles the entire stack**, and I want you to be job-ready with the same mindset.

Career-Focused Learning: My goal is simple—**transform your career** with relevant, high-quality education and give you the confidence to enter the data engineering world.

Let's build your future

Linked id- www.linkedin.com/in/gaurav-kurhekar

CONTACT US

Email ID- mail@upskillacademy.co.in
Mobile - 7020269389
Website : upskillacademy.co.in

THANKS!

Thank You for Your Time! We Appreciate Your Attention!