Azure Data Engineering Mastery
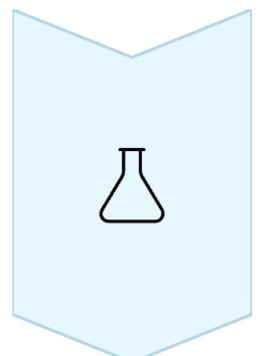
Transform your career with comprehensive Python, PySpark, and Azure cloud training

Two-Phase Learning Journey

Phase 1

Master Python fundamentals and PySpark for big data processing

Phase 2

Azure cloud services: ADLS Gen2, Delta Lake, Databricks, Synapse Analytics, ADF

DRI
DATA RISE INSIGHT
Rising in career through analytics

Phase 1: Python Foundations

Core Concepts

Data types, collections (mutable & immutable), string interpolation, control statements

Advanced Features

Functions as first-class citizens, Map/FlatMap/Filter/Reduce operations
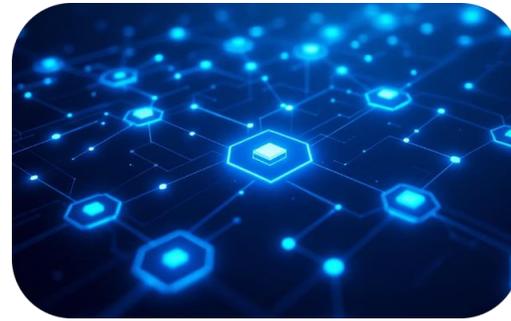
OOP & Production

Classes, objects, inheritance, wrapper code, error handling, real-time code preparation

PySpark Deep Dive

Spark Architecture

Ecosystem components, RDD concepts, DAG, lineage, lazy transformations, narrow vs wide operations

Spark Submit

Client, cluster, standalone modes. Dynamic memory allocation, broadcast variables, accumulators

Optimization

Repartition vs coalesce, caching strategies, serialization, memory levels, production parameters

Spark SQL Mastery

1

DataFrames

Creation from CSV, JSON, AVRO, XML, Parquet. Schema design and write modes

2

Operations

DSL operations, joins, broadcast joins, aggregations, window functions

3

Advanced

Complex data (Struct/Array), UDFs, date-time functions, performance optimization
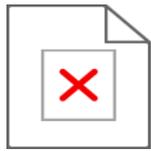
4

Production

Config files, null handling, error records, memory optimization, end-to-end pipelines
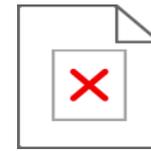
Phase 2: Azure Data Engineering

Azure Storage

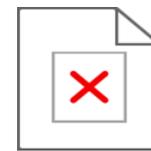**ADLS Gen2, BLOB service with access tiers, redundancy features, Storage Explorer integration**

Databricks

Cluster types, notebooks, mount points, CLI deployments, pipeline creation, architecture patterns

Synapse Analytics

Dedicated and serverless pools, pipeline creation, Spark integration, SQL pool analysis

Delta Lake

Schema evolution, version control, ACID transactions, SCD Type 1, Unity catalog for metadata

Real-World Project Architecture

On-Prem Ingest

Collect data from legacy systems

Processing

Transform with Databricks or Synapse

Azure Storage

Store raw data in ADLS Gen2

Analytics & Serving

Serve models and BI workloads
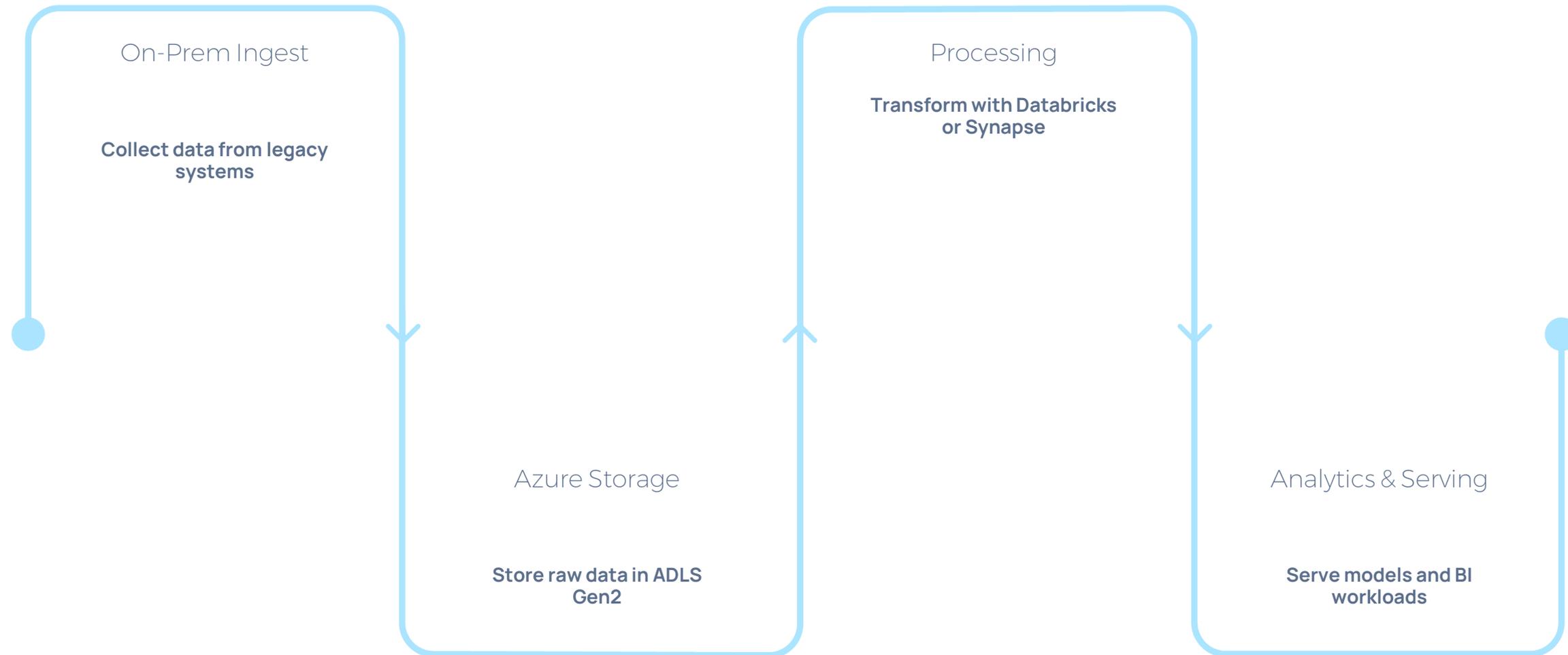
Learn production-grade architectures used in enterprise environments, from on-premises systems to cloud-native Azure solutions.

End-to-End Pipeline

Integrate Sqoop, Hive, Spark, HBase, and RDBMS in complete data workflows

Multi-Cloud Patterns

Understand use cases across different cloud platforms and migration strategies

Program Details & Support

Schedule

Weekend batch: 8:15-10:45 AM IST with 20-min break. 4-month duration starting January 31, 2026

Class Size

Limited to 30 enrollments per batch for personalized attention and quality learning

Complete Career Support

○ Resume Preparation

Professional guidance to showcase your new skills effectively

○ Session Recordings

Access all training materials for review and reinforcement

○ Interview Prep

Mock interviews and data engineer perspective preparation

○ Practice Sessions

Weekly Wednesday sessions for hands-on reinforcement

Ready to Transform Your Career?

Join us for comprehensive, real-time Azure Data Engineering training with complete career support.

Thank you!