

DATA ENGINEERING

INTERMEDIATE

5 DAYS

Data Engineering on Azure

Build a strong foundation in Azure Data Engineering, from lakehouse architecture to enterprise ingestion and transformation workflows. Learn how to build end-to-end pipelines using ADLS, Data Factory, Databricks, and Synapse while applying security, monitoring, and production-ready operational practices.

Why Azure Data Engineering?

Enterprise-ready data platforms

Build governed pipelines using Azure-native security and identity controls.

Better compliance readiness

Azure governance tools support regulated and enterprise workloads.

Faster integration with Microsoft ecosystem

Strong alignment with Power BI, Synapse, and ADLS.

Accelerated analytics delivery

Ship dashboards and insights faster with integrated cloud services.

Improved operational resilience

Use managed services for stable scaling and reliability.

What you'll be able to do

- ✓ Understand modern Data Engineering architecture on Azure and service selection for workloads.
- ✓ Build a complete data platform using Azure services across ingestion, storage, transformation, orchestration, and serving/analytics.
- ✓ Design and implement a data lake foundation using Azure Data Lake Storage Gen2 (ADLS) with raw/cleansed/curated zones.
- ✓ Build ingestion and orchestration pipelines using Azure Data Factory (ADF) with linked services, datasets, triggers, and monitoring.
- ✓ Transform data using Azure Databricks with Spark fundamentals and Delta Lake patterns.
- ✓ Implement batch and incremental pipelines including full load and watermark-based processing.
- ✓ Build analytics datasets and data warehouse style tables and views for reporting.
- ✓ Apply security and governance practices including RBAC, managed identities, Key Vault integration, and encryption.
- ✓ Implement monitoring and operational practices including pipeline logs, failure handling, and alerting.
- ✓ Deliver an end-to-end capstone project on Azure with validated outputs and documentation.

Who this is for

- Data Engineers
- Analytics Engineers
- Cloud Engineers supporting data platforms
- Data Platform Engineers
- DevOps engineers working with Azure data services
- BI engineers moving into data engineering

Curriculum

01 Azure Data Engineering Architecture Overview

- Data engineering lifecycle (ingest → store → transform → serve)
- Modern platform patterns (Data Lake, Data Warehouse, Lakehouse)
- Batch vs streaming overview
- Choosing Azure services per workload
- **Hands-on:** Activity: Design an Azure reference architecture for an analytics use case

02 ADLS Gen2, Data Formats, and Lake Zones

- Storage account fundamentals
- Containers and folder strategy
- Lake zones (raw, cleansed, curated)
- Naming and partitioning best practices
- CSV vs JSON vs Parquet
- **Hands-on:** Lab/Exercise: Upload datasets + define partition folders and choose partition keys

03 Security Basics: RBAC, Managed Identity, and Key Vault

- RBAC basics for storage
- Managed Identity introduction
- Data access patterns (engineers vs analysts, least privilege)
- Key Vault overview
- Why secrets should never be hardcoded
- **Hands-on:** Lab: Connect ADF to Key Vault for secret management

04 Azure Data Factory Fundamentals and Copy Activity

- What ADF solves (ingestion automation, scheduling, dependency management)
- Key components (linked services, datasets, pipelines, triggers)
- Activity types (Copy Activity, Validation activity intro, ForEach intro)
- Copy from HTTP source
- Copy from Blob/ADLS source
- **Hands-on:** Lab: Build ingestion pipeline (source → ADLS raw) + parameterize dataset path with dynamic folders

05 Incremental Loads, Monitoring, and ADF Operations

- Full load vs incremental load
- Watermark concepts (updated_at, ingestion timestamp)
- Using pipeline parameters for incremental runs
- Monitor pipeline runs
- Activity logs and failure reasons
- **Hands-on:** Lab: Simulate pipeline failure, troubleshoot, and add retry + failure handling strategy

06 Databricks Fundamentals and Spark Transformations

- What is Databricks (Spark-based processing, notebooks, jobs)
- Clusters and compute basics
- Reading from ADLS securely
- Data cleaning (null handling, schema casting, deduplication)
- Joins and aggregations
- **Hands-on:** Lab: Transform raw orders to cleansed zone + build curated datasets (daily revenue, top customers)

07 Delta Lake, Jobs, and Data Quality Checks

- Why Delta in a lakehouse (ACID transactions, schema enforcement, scalable merges)
- Delta operations (overwrite vs append)
- Merge/upsert concept
- Time travel introduction
- Convert notebook into job
- **Hands-on:** Lab: Add validation and stop pipeline on quality failure

08 Synapse Fundamentals and Serverless Lake Queries

- Query choices on Azure (Databricks queries, Synapse serverless SQL overview)
- When you need a warehouse vs lake queries
- Synapse components overview (SQL pools, Spark pools concept, pipelines concept)
- Serverless SQL vs Dedicated SQL (when to use each)
- External tables concept
- **Hands-on:** Lab: Query curated zone + build analytics views (revenue by region, top customers)

09 Warehouse Modeling and ADF + Databricks + Synapse Flow

- Star schema basics (fact + dimension)
- KPI reporting dataset design
- Transformations for warehouse tables
- End-to-end architecture (ingest with ADF, transform with Databricks, serve via Synapse)
- Automation and reliability approach
- **Hands-on:** Lab: Create orchestrated pipeline concept (ADF triggers Databricks job, curated data queryable in Synapse)

10 Performance, Cost, Security, and Governance

- Databricks cluster cost drivers
- Synapse query cost awareness
- Storage cost optimization basics
- RBAC for storage and compute
- Managed identities in pipelines
- **Hands-on:** Workshop: Define governance model for data lake access and ownership

11 Observability, Reliability, and Production Practices

- Monitoring across services (ADF logs, Databricks job logs, storage metrics)
- SLA and freshness tracking
- Alerting and incident readiness
- Idempotent pipeline design
- Backfill strategies
- **Hands-on:** Activity: Build production readiness checklist for Azure pipelines

12 Capstone: End-to-End Azure Data Engineering Pipeline

- Build ADLS zones (raw/cleansed/curated)
- ADF pipeline for ingestion
- Databricks transformation pipeline (Delta output)
- Synapse analytics queries/views
- Data quality checks + incremental load logic
- **Hands-on:** Capstone deliverables (architecture diagram, notebooks, curated datasets, KPI queries, documentation)