Where trainees meet real Support

Azure Data Engineering Course

Azure, SQL & Azure Data Factory (ADF) Masterclass

8 Weeks Duration • 40 Total Lectures (5 Lectures/Week) • Beginner to Advanced

About the Course

- **Structured Learning Journey:** From cloud basics & SQL to enterprise Azure Data Factory (ADF) ETL engineering
- **Career-Oriented:** Designed specifically for aspiring Azure Data Engineers, ETL developers, and DBAs
- **Comprehensive Scope:** Covers ADLS Gen2, Azure SQL, ADF pipelines, Mapping Data Flows, and CI/CD
- **Hands-On Methodology:** 100% practical labs building live automated data ingestion pipelines in Azure Portal
- **Portfolio Capstone:** Architect an end-to-end lakehouse ETL pipeline loading business data into reporting warehouses

Who Can Attend?

- Aspiring Azure Data Engineers & ETL Developers
- SQL Developers & DBAs upgrading to Cloud Lakehouse platforms
- Data Analysts & BI Professionals scaling to Big Data pipelines
- Software Engineers transitioning into Cloud Data Engineering

**No previous Azure experience required. Starting points are tailored to learner background.*

Course Outline (Part 1)

Phase 1: Basics of Cloud Computing & Azure Fundamentals (Week 1)

- **What is Cloud Computing?** Key benefits, scalability, elasticity, high availability & cost efficiency
- **Cloud Deployment Models:** Private Cloud, Public Cloud, and Hybrid Cloud architectures
- **Cloud Service Models:** IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service)
- **Introduction to Azure:** Create an Azure Account, navigation & walkthrough of the Azure Portal
- **Overview & Practical Implementation of Core Services:** Subscriptions, Resource Groups, Blob Storage, Data Lake Storage (ADLS Gen2), Azure SQL Server & Database, Azure Data Factory (ADF), Azure Databricks, Azure Key Vaults, Azure Logic Apps, and GitHub Repositories
- **Outcome:** Navigate Azure Portal, manage resources, and understand core cloud data services.

Phase 2: Introduction to Big Data & Modern Data Warehousing (Week 2)

- **Big Data Fundamentals:** What is Data? What is Big Data? Challenges of traditional RDBMS at scale
- **Types of Data:** Structured (Tables/RDBMS), Semi-Structured (JSON, XML, Avro, Parquet), and Unstructured (Audio, Video, Logs)
- **Characteristics of Big Data (5 V's):** Variety, Velocity, Volume, Veracity, and Value with industry use cases
- **Data Sources of Big Data:** Ingesting from OLTP databases, REST APIs, enterprise applications, IoT & file feeds
- **Data Warehouse Concepts & Architecture:** What is a Datawarehouse? OLTP vs. OLAP architectures, staging areas, dimensional modeling (Star & Snowflake schemas), and Lakehouse architecture paradigms
- **Outcome:** Evaluate Big Data patterns and design scalable enterprise data warehouse architectures.

Phase 3: Azure Data Lake Storage Gen2 (ADLS) & Cloud Storage (Week 3)

- **Storage Account Architecture:** Create a Storage account, understand types of storage accounts & performance tiers
- **Azure Data Lake Storage (ADLS Gen2):** Create ADLS Gen2 accounts, enable Hierarchical Namespace & manage containers
- **Access Control & Security:** Configure Access to ADLS via Azure RBAC, Access Control Lists (ACLs), SAS tokens & firewall rules
- **Data Operations:** Ingest & load data to ADLS; read and write data programmatically, via Azure Storage Explorer & CLI
- **Reliability & Governance:** Configure Backup, Redundancy (LRS, GRS, ZRS), lifecycle management & Disaster Recovery
- **Outcome:** Build secure, partitioned, enterprise-grade cloud data lakes on ADLS Gen2.

Phase 4: Introduction to Azure SQL Database & SQL Basics (Week 4)

- **Provisioning & Management:** Create Azure SQL Server and Database; configure Compute resources & Elastic Pools
- **Security & Connectivity:** Configure Access & Security, firewall rules, private endpoints, connection strings & authentication
- **SQL Basics Mastery:** DDL, DML, DQL statements, data types, constraints, and relational database schema design
- **Advanced SQL Querying:** JOINS (Inner, Left, Right, Full Outer, Cross), Subqueries & Correlated Subqueries, Common Table Expressions (CTEs), Set Operations (UNION, UNION ALL, INTERSECT, EXCEPT)
- **Analytical Functions & Automation:** Window Functions (ROW_NUMBER, RANK, DENSE_RANK, NTILE, LEAD, LAG), aggregate functions with OVER(PARTITION BY), Stored Procedures & table indexing for ETL performance
- **Outcome:** Write optimized complex analytical SQL queries and manage cloud relational databases.

SUPPORT ME
BY THE DEVELOPERS GROUP

Course Outline (Part 2 - Azure Data Factory Deep Dive)

Phase 5: Azure Data Factory (ADF) - Architecture, Runtimes & Connectors (Week 5)

- **ADF Introduction:** Cloud ETL vs. ELT orchestration, ADF Studio UI walkthrough & core architectural components
- **Integration Runtime (IR):** Azure Auto Integration Runtime, Self-Hosted Integration Runtime (SHIR for on-premises connectivity), and Azure-SSIS Integration Runtime for legacy SSIS package migration
- **Linked Services:** Create Linked Services for BLOB Storage, Azure Data Lake Storage (ADLS), Azure SQL & On-Premises Servers
- **Datasets:** Create Datasets for CSV, Parquet, Excel, Avro, JSON, Azure SQL Tables & On-Premises SQL Tables
- **Pipeline Management:** Create new pipelines, execute pipelines via REST API, and interactive pipeline debugging

Phase 6: ADF Orchestration Activities & Control Flow Logic (Week 6)

- **Data Movement & Utility Activities:** 1. Copy Data (Source/Sink, schema mapping, partitioning), 2. Delete, 3. Stored Procedures, 4. Get-Meta Data (validation & child items retrieval)
- **Control Flow Activities:** 5. Lookup, 6. For Each (batch/sequential), 7. IF Condition & Switch, 8. Until, 9. Wait, 10. Fail
- **Transform & Pipeline Execution Activities:** 11. Data Flow, 12. Set Variable & Append Variable, 13. Databricks Notebook (cluster execution), 14. Execute Pipelines (child/parent pipeline invocation)

Phase 7: ADF Triggers, Parameters & Dynamic Orchestration (Week 7)

- **Pipeline Triggers:** 1. Schedule Trigger (cron-based), 2. Tumbling Window Trigger (dependency handling & backfill), 3. Storage Events (Blob/ADLS created & deleted triggers)
- **Dynamic Parameters:** Dynamically manage multiple ADLS accounts, SQL Servers, datasets, pipelines & triggers; use parameters during pipeline execution; Create Global Parameters across factory environments
- **Expression Language:** Expression Language usage in ADF, system variables, string/date manipulation & dynamic SQL passing

Phase 8: Mapping Data Flows, Logic Apps, Key Vault & Git CI/CD (Week 8)

- **Transformations (Mapping Data Flows):** 1. Create & debug Data Flow on Spark, 2. Filter, 3. Select & Sort, 4. Aggregate & GroupBy, 5. Joins, 6. Lookup & Exists, 7. Union, 8. Alterrow (Insert/Update/Upsert/Delete), 9. Rank, 10. Pivot & UnPivot, 11. Use Flowlet to avoid redundant steps
- **Enterprise Operations:** Monitor Jobs & pipeline run metrics; Send Failure notifications using Azure Logic Apps
- **Security & Change Management:** Manage credentials securely using Azure Key Vault; Repository Change Management (1. Create GitHub Repository integration, 2. Migrate ARM templates using Git, 3. ARM Templates – Export/Import Manually)

Hands-on Capstone Projects

- **Enterprise Hybrid Big Data Pipeline:** Ingest multi-format data (CSV/JSON/Parquet) from On-Premises SQL & Blob into ADLS Gen2 using Self-Hosted IR, apply complex Mapping Data Flow transformations, and execute Databricks Notebooks.
- **Metadata-Driven Dynamic ETL Engine:** Build a generic, scalable ADF pipeline dynamically processing 50+ tables using Lookup, ForEach, Global Parameters, and Stored Procedures with zero hardcoded paths.
- **Production Monitoring, Security & CI/CD:** Secure connection strings with Azure Key Vault, automate instant incident alert emails via Logic Apps on pipeline failure, and implement Git ARM template automated release pipelines.

Tools & Technologies

Microsoft Azure, Azure Data Factory (ADF v2), Azure SQL Database, Azure Data Lake Storage (ADLS Gen2), Blob Storage, Azure Databricks, Azure Key Vault, Azure Logic Apps, GitHub, ARM Templates, SQL (T-SQL), JSON, Parquet, Avro, CSV, Integration Runtimes (SHIR)

High-Value POC Categorization

- **Beginner:** ADLS Gen2 setup, SQL queries & basic ADF Copy pipeline
- **Intermediate:** Tumbling window triggers, Data Flows & Key Vault
- **Advanced:** Dynamic metadata pipeline, SHIR, Databricks & ARM CI/CD

Why This Program?

- 100% hands-on cloud labs with real-world enterprise architectures
- Complete mastery of SQL querying & ADF data orchestration
- Production-ready CI/CD, Git version control & Logic Apps alerting
- Targeted interview prep for Azure Data Engineer roles

How This Fits Your Full Learning Stack

- Frontend → UI • Backend → Java / Python / Node.js • Cloud → AWS / Azure • Data → Azure Data Factory / Databricks
- **Final Result: Full Stack + Cloud + Data + AI Engineer**

SUPPORT ME
— BY THE DEVELOPERS' GROUP