



1. Course Outline

Module 1: Introduction to ADF

- Historical background: SSIS, ADF v1, other ETL/ELT tools
- Key capabilities and benefits of ADF v2
- Recent feature updates and enhancements

Module 2: Core Architectural Components

- Connectors: Azure services, databases, NoSQL, files, generic protocols, services & apps, custom
- Pipelines
- Activities: data movement, data transformation, control flow
- Datasets: source, sink
- Integration Runtimes: Azure, Self-Hosted, Azure-SSIS

Module 3: Building and Executing Your First Pipeline

- Creating ADF v2 instance
- Creating a pipeline and associated activities
- Executing the pipeline
- Monitoring execution
- Reviewing results

Module 4: Data Movement

- Copying Tools and SDKS
- Copy Data Tool/Wizard
- Copy activity
- SDKs: Python, .NET
- Automation: PowerShell, REST API, ARM Templates

Copying Considerations

- File formats: Avro, binary, delimited, JSON, ORC, Parquet
- Data store support matrix
- Write behavior: append, upsert, overwrite, write with custom logic
- Schema and data type mapping
- Fault tolerance options

Module 5: Data Transformation

- Transformation with Mapping Data Flows
- Introduction to mapping data flows
- Data flow canvas
- Debug mode
- Dealing with schema drift
- Expression builder & language
- Transformation types: Aggregate, Alter row, Conditional split, Derived column, Exists, Filter, Flatten, Join, Lookup, New branch, Pivot, Select, Sink, Sort, Source, Surrogate key, Union, Unpivot, Window

Transformation with External Services

- Databricks: Notebook, Jar, Python
- HDInsight: Hive, Pig, MapReduce, Streaming, Spark
- Azure Machine Learning service
- SQL Stored procedures
- Azure Data Lake Analytics U-SQL
- Custom activities with .NET or R

Module 6: Control Flow

- Purpose of activity dependencies: branching and chaining
- Activity dependency conditions: succeeded, failed, skipped, completed

- Control flow activities: Append Variable, Azure Function, Execute Pipeline, Filter, ForEach, Get Metadata, If Condition, Lookup, Set Variable, Until, Wait, Web

Module 7: Runtime and Operations

- Debugging
- Monitoring: visual, Azure Monitor, SDKs, runtime-specific best practices
- Scheduling execution with triggers: event-based, schedule, tumbling window
- Performance, scalability, tuning
- Common troubleshooting scenarios in activities, connectors, data flows and integration runtimes

Module 8: DevOps with ADF

- Quick introduction to source control with Git
- Integration with GitHub and Azure DevOps platforms
- Environment management: Development, QA, Production
- Iterative development best practices
- Continuous Integration (CI) pipelines
- Continuous Delivery (CD) pipelines

Module 9: Promoting Reuse

- Templates: out-of-the-box and organizational
- Parameters
- Naming convention

Module 10: Security

- Data movement security
- Azure Key Vault
- Self-hosted IR considerations
- IP address blocks
- Managed identity

