

**Applied Data Pipelines using Azure Data Factory**

**Course Number:** AZR-154WA
**Duration:** 2 days

**Overview**

In this Azure Data course, participants explore Azure Data Factory (ADF), Microsoft's cloud-based data integration service. Participants learn ETL (Extract-Transform-Load) fundamentals, pipeline building, and external service integration. Through hands-on exercises, participants master data transformation techniques, orchestration, and monitoring. They also explore the similarities and differences among ADF, Synapse Pipelines, and Fabric.

**Prerequisites**

Participants should have a basic understanding of data concepts and some experience with data handling. Familiarity with cloud computing and the Azure portal is beneficial, but prior experience with Azure Data Factory is not required.

**Materials**

All students receive comprehensive courseware covering all topics in the course.

**Software Needed on Each Student PC**

Attendees will not need to install any software on their computers for this class. The class will be conducted in a remote environment that Accelebrate will provide; students will only need a local computer with a web browser and a stable Internet connection. Any recent version of Microsoft Edge, Mozilla Firefox, or Google Chrome will work well.

**Objectives**

* Understand Azure Data Factory's role in modern data integration.
* Design and build data pipelines with Azure Data Factory.
* Integrate ADF with external services like Azure Blob Storage and Azure SQL.
* Master data orchestration and workflow management.
* Learn to monitor, manage, and optimize data pipelines for efficiency.

**Outline**

* Azure Data Factory Overview
	+ Understanding ETL
	+ Understanding Data Pipeline and Data Flow
	+ Evolution of data integration from on-premises to cloud-based solutions
	+ Understanding Azure Data Factory
	+ Key concepts in Azure Data Factory: pipelines, activities, datasets, triggers
* Building Data Pipelines in Azure Data Factory
	+ Introduction to Data Movement activities
	+ Using Copy Data activity
	+ Using Data Flow activity
	+ Using Custom activity
	+ Introduction to Data Transformation activities
	+ Using Mapping Data Flow
	+ Using Wrangling Data Flow
	+ Using Stored Procedure activity
	+ Working with Variables
	+ Working with Lookup
	+ Working with Wait
* Integration with External Services
	+ Integration with Azure services: Azure Blob Storage, Azure SQL, and Azure Synapse Analytics
	+ Integration with external services: Amazon S3, Google Cloud Storage, Salesforce
* Data Transformation Techniques
	+ Working with Multiple inputs/outputs
	+ Working with Schema modifier
	+ Working with Formatters
	+ Mapping Data Flow
	+ Wrangling Data Flow
	+ Implementing data cleansing, enrichment, and aggregation
* Data Orchestration and Workflow Management
* Managing dependencies and scheduling in Azure Data Factory
* Implementing parameterized pipelines for dynamic data processing
* Configuring triggers
* Working with parameters
* Using dynamic content in data pipelines
* Working with Iteration & Conditionals
	+ Using ForEach and Until
	+ Using Filter
	+ Using If Condition & Switch
* Monitoring and Management in Azure Data Factory
	+ Monitoring data pipelines using Azure Data Factory Monitoring Hub
	+ Understanding pipeline run status, triggers, and alerts
	+ Managing data factory resources, including linked services, datasets, and pipelines
	+ Implementing best practices for performance optimization and cost management
* Synapse Pipelines Integration
	+ Leveraging Azure Synapse Analytics for scalable data processing
	+ Integrating Synapse pipelines with Azure Data Factory for end-to-end data workflows
	+ Understanding Synapse Pipelines and their role in data processing
	+ Configuring Data Movement activities between Azure Data Factory and Azure Synapse Analytics
	+ Utilizing Synapse Linked Services and Datasets in Azure Data Factory pipelines
* Pipelines in Fabric
	+ Understanding and implementing Pipelines in Fabric for large-scale data processing
	+ Leveraging Data Flows in Pipelines in Fabric for complex data transformations
	+ Creating and managing Data Marts within Pipelines in Fabric for optimized data storage and retrieval
	+ Utilizing Datasets in Pipelines in Fabric for defining data structures and schemas
	+ Implementing best practices for designing and orchestrating Pipelines in Fabric
	+ Hands-on exercises: Building and optimizing data pipelines using Pipelines in Fabric
* Best Practices in Azure Data Factory
	+ Performance tuning for data pipelines
	+ Parallel execution and partitioning strategies for improved throughput
	+ Data skew management and load balancing techniques
	+ Pipeline design best practices: modularization, reusability, and maintainability
	+ Error handling and exception management strategies
	+ Version control and deployment best practices
	+ Resource optimization
	+ Leveraging serverless compute options and auto-scaling capabilities.