

# Azure Data Engineer Learning Pathway (1/2)

[www.aka.ms/pathways](http://www.aka.ms/pathways)

## Getting started

Learn how to implement and manage data engineering workloads on Microsoft Azure, using Azure services such as Azure Synapse Analytics, Azure Data Lake Storage Gen2, Azure Stream Analytics, Azure Databricks, and others.

Explore common data engineering tasks such as orchestrating data transfer and transformation pipelines, working with data files in a data lake, creating and loading relational data warehouses, capturing and aggregating streams of real-time data, and tracking data assets and lineage.

### Audience Profile:

The primary audience for this course is data professionals, data architects, and business intelligence professionals who want to learn about data engineering and building analytical solutions using data platform technologies that exist on Microsoft Azure.

- [New to the Cloud or Azure? Start with Azure Fundamentals](#)
  - [New to data solutions on Azure? Build your knowledge with Data Fundamentals](#)
  - [Intro to data classification and protection](#)
- Get started with data engineering on Azure**
- [Introduction to data engineering on Azure](#)
  - [Introduction to Azure Data Lake Storage Gen2](#)
  - [Introduction to Azure Synapse Analytics](#)

## Microsoft Learn/Documentation

### Design and implement data storage

- [Understand Azure Data Lake Storage Gen2](#)
- [Access tiers for Azure Blob Storage](#)
- [Storage considerations when using Azure Synapse serverless SQL pools](#)
- [Query a Parquet file using Azure Synapse serverless SQL pools](#)
- [Dynamic file pruning](#)
- [Understand table distribution design](#)
- [Partitioning tables in dedicated SQL pool](#)
- [Understand table distribution design](#)
- [Best practices for dedicated SQL pools in Azure Synapse Analytics](#)
- [Star Schema](#)
- [Multidimensional Schemas and Data](#)
- [Manage retention of historical data in system-versioned temporal tables](#)
- [Getting started with temporal tables](#)
- [Create and configure a self-hosted integration runtime](#)
- [Manage self-hosted integration runtime](#)
- [Choosing an analytical data store in Azure](#)
- [Synapse Analytics shared metadata tables](#)
- [When do you use Apache Spark pools?](#)
- [Data Compression](#)
- [Exercise - Use table distribution and indexes to improve performance](#)
- [Change storage account is replication](#)
- [Slowly Changing Dimension Transformation](#)
- [Populate slowly changing dimensions](#)
- [Create external tables in Azure Synapse serverless SQL pools](#)

### Design and implement data storage

- [Views in Synapse serverless SQL pools](#)
- [Tutorial: Load data to Azure Synapse Analytics SQL pool](#)
- [Create, develop, and maintain Synapse notebooks in Azure Synapse Analytics](#)

### Design and develop data processing

- [Common practices for data loading](#)
- [Tutorial: Extract, transform, and load data by using Azure Databricks](#)
- [Understand the Streaming Analytics Workflow](#)
- [Handling bad records and files](#)
- [Prepare and transform data with Azure Synapse Analytics](#)
- [Analyse complex data types in Azure Synapse Analytics](#)
- [Understand data store models](#)
- [Prepare and transform data](#)
- [Define a modern data warehouse architecture](#)
- [Choosing a batch processing technology](#)
- [Manage source data files](#)
- [Copy activity in Azure Data Factory](#)
- [MERGE \(Transact-SQL\)](#)
- [Continuous integration and delivery for Azure Synapse workspace](#)
- [Handle SQL truncation error rows in Data Factory mapping data flows](#)

## Role based Certification

### DP-203: Azure Data Engineer

#### Skills Measured

- Design and implement data storage
- Develop data processing
- Secure, monitor, and optimize data storage and data processing

#### Self Study:

- [Get started with data engineering on Azure](#)
- [Build data analytics solutions using Azure Synapse serverless SQL pools](#)
- [Perform data engineering with Azure Synapse Apache Spark Pools](#)
- [Work with Data Warehouses using Azure Synapse Analytics](#)
- [Transfer and transform data with Azure Synapse Analytics pipelines](#)
- [Work with Hybrid Transactional and Analytical Processing Solutions using Azure Synapse Analytics](#)
- [Implement a Data Streaming Solution with Azure Stream Analytics](#)
- [Govern data across an enterprise](#)
- [Data engineering with Azure Databricks](#)

[Exam Study Guide](#)
[Course Page](#)
[Exam Page](#)
[30 Day Challenge](#)
[Practice Assessment](#)
[Video on Demand \(soon\)](#)
[Azure Data Architecture Guide](#)

# Azure Data Engineer Learning Pathway (2/2)

[www.aka.ms/pathways](http://www.aka.ms/pathways)

## Additional Study

### Design and develop data processing

- [Backup and restore in Azure Synapse Dedicated SQL pool](#)
- [Implement workload management](#)
- [Use extended Apache Spark history server to debug and diagnose Apache Spark applications](#)
- [Enterprise Data Warehouse Architecture](#)
- [Stream processing with Azure Databricks](#)
- [Azure Synapse Analytics](#)
- [Monitoring for performance efficiency](#)
- [Work with windowing functions](#)
- [Schema drift](#)
- [Time handling in Stream Analytics](#)
- [Checkpoint and replay concepts in Azure Stream Analytics jobs](#)
- [Scale an Azure Stream Analytics job to increase throughput](#)

### Design and develop data processing

- [Use repartitioning to optimize processing](#)
- [Azure Stream Analytics output error policy](#)
- [Stream Analytics output to Cosmos DB](#)
- [Stream processing with Stream Analytics](#)
- [Data Loading best practices](#)
- [Get Started with Synapse Analytics](#)
- [Monitor your Synapse Workspace](#)

### Design and implement data security

- [Implement encryption](#)
- [Data ingestion security considerations](#)
- [Configure authentication](#)
- [Access control lists \(ACLs\) in Azure Data Lake Storage Gen2](#)
- [Synapse access control](#)
- [Column-level security](#)
- [Manage authorization through column and row level security](#)
- [Manage user permissions](#)
- [Auditing for Azure SQL Database and Azure Synapse Analytics](#)
- [Retention Policy on storage accounts](#)
- [Understand network security options](#)
- [Dynamic Data Masking](#)
- [Secure a dedicated SQL pool](#)

### Monitor and optimize data storage and data processing

- [Monitor and Alert Data Factory by using Azure Monitor](#)
- [Exercise – Implement workload management](#)
- [Monitor your Azure Synapse Analytics dedicated SQL pool workload using DMVs](#)
- [Collect custom logs with Log Analytics agent](#)
- [Use Synapse Studio to monitor your workspace pipeline runs](#)
- [Deploying Apache Airflow in Azure to build and run data pipelines](#)

### Monitor and optimize data storage and data processing

- [Auto Optimize in Azure Databricks](#)
- [Modify user-defined functions](#)
- [Designing distributed tables](#)
- [Data spillage scenario - Search and purge](#)
- [Quickstart: Create an Azure Synapse workspace using an ARM template](#)
- [Indexing dedicated SQL pool tables](#)
- [Performance tuning with result set caching](#)
- [Optimize Apache Spark jobs](#)
- [Troubleshoot library installation errors](#)
- [Debug data factory pipelines](#)



## Microsoft Applied Skills

Targeted validation for real-world scenarios. Demonstrate proficiency in specific, scenario-based skill sets so you can make a bigger impact on every project, at your organization, and in your career

[Explore Applied Skills](#)

## 30 days to Learn it Challenge

30 Days to Learn It can help you build skills and start your preparation for Microsoft Certifications for AI, DevOps, Microsoft 365, low code, IoT, data science, cloud development, and more. Select your challenge below, work through learning modules, and exchange ideas with peers through a global community forum.

[Explore the challenges](#)