## ****What is Azure Data Factory?****

Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

ADF does not store any data itself. It allows you to create data-driven workflows to orchestrate the movement of data between supported data stores and then process the data using compute services in other regions or in an on-premise environment. It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.

## **Azure Data Factory use cases**

ADF can be used for:

* Supporting data migrations
* Getting data from a client’s server or online data to an Azure Data Lake
* Carrying out various data integration processes
* Integrating data from different ERP systems and loading it into Azure Synapse for reporting

## ****How does Azure Data Factory work?****

The Data Factory service allows you to create data pipelines that move and transform data and then run the pipelines on a specified schedule (hourly, daily, weekly, etc.). This means the data that is consumed and produced by workflows is time-sliced data, and we can specify the pipeline mode as scheduled (once a day) or one time.

Azure Data Factory pipelines (data-driven workflows) typically perform three steps.

### **Step 1: Connect and Collect**

Connect to all the required sources of data and processing such as SaaS services, file shares, FTP, and web services. Then,  move the data as needed to a centralized location for subsequent processing by using the Copy Activity in a data pipeline to move data from both on-premise and cloud source data stores to a centralization data store in the cloud for further analysis.

### ****Step 2: Transform and Enrich****

Once data is present in a centralized data store in the cloud, it is transformed using compute services such as HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Machine Learning.

### ****Step 3: Publish****

Deliver transformed data from the cloud to on-premise sources like SQL Server or keep it in your cloud storage sources for consumption by BI and analytics tools and other applications.

## ****Data migration activities with Azure Data Factory****

By using Microsoft Azure Data Factory, data migration occurs between two cloud data stores and between an on-premise data store and a cloud data store.

Copy Activity in Azure Data Factory copies data from a source data store to a sink data store. Azure supports various data stores such as source or sink data stores like [Azure Blob storage](https://cloudacademy.com/course/intro-to-azure-storage/course-introduction-1/" \t "https://cloudacademy.com/blog/what-is-azure-data-factory/_blank), [Azure Cosmos DB](https://cloudacademy.com/course/introduction-azure-cosmos-db/" \t "https://cloudacademy.com/blog/what-is-azure-data-factory/_blank) (DocumentDB API), Azure Data Lake Store, Oracle, Cassandra, etc. For more information about Azure Data Factory supported data stores for data movement activities, refer to Azure documentation for [data movement activities](https://docs.microsoft.com/en-us/azure/data-factory/data-factory-introduction" \t "https://cloudacademy.com/blog/what-is-azure-data-factory/_blank).

Azure Data Factory supports transformation activities such as Hive, MapReduce, Spark, etc that can be added to pipelines either individually or chained with other activities. For more information about ADF-supported data stores for data transformation activities, refer to the following Azure Data Factory documentation: [Transform data in Azure Data Factory](https://docs.microsoft.com/en-us/azure/data-factory/data-factory-data-transformation-activities" \t "https://cloudacademy.com/blog/what-is-azure-data-factory/_blank).

If you want to move data to/from a data store that Copy Activity doesn’t support, you should use a .NET custom activity in Azure Data Factory with your own logic for copying/moving data. To learn more about creating and using a custom activity, [check the Azure documentation](https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-use-custom-activities" \t "https://cloudacademy.com/blog/what-is-azure-data-factory/_blank) and see “Use custom activities in an Azure Data Factory pipeline”.

## ****Azure Data Factory key components****

Azure Data Factory has four key components that work together to define input and output data, processing events, and the schedule and resources required to execute the desired data flow:

* ****Datasets represent data structures within the data stores.**** An input dataset represents the input for an activity in the pipeline. An output dataset represents the output for the activity. For example, an Azure Blob dataset specifies the blob container and folder in the Azure Blob Storage from which the pipeline should read the data. Or, an [Azure SQL](https://cloudacademy.com/course/introduction-to-azure-storage-solutions/course-intro-14/" \t "https://cloudacademy.com/blog/what-is-azure-data-factory/_blank) Table dataset specifies the table to which the output data is written by the activity.
* ****A pipeline is a group of activities.**** They are used to group activities into a unit that together performs a task. A data factory may have one or more pipelines. For example, a pipeline could contain a group of activities that ingests data from an Azure blob and then runs a Hive query on an HDInsight cluster to partition the data.
* ****Activities define the actions to perform on your data.**** Currently, Azure Data Factory supports two types of activities: data movement and data transformation.
* ****Linked services define the information needed for Azure Data Factory to connect to external resources.**** For example, an Azure Storage linked service specifies a connection string to connect to the Azure Storage account.

## Performing copy activity







