**<https://k21academy.com/microsoft-azure/dp-203/adf-interview-questions/#question_5>**

Azure Data Factory doesn’t store any data itself; it lets you produce workflows that orchestrate the movement of data between supported data stores and data processing. You can monitor and manage your workflows using both programmatic and UI mechanisms. Apart from that, it is the best tool available today for ETL processes with an easy-to-use interface. This shows the need for Azure Data Factory.

Azure Data Factory is a cloud-based integration service offered by Microsoft that lets you create data-driven workflows for orchestrating and automating data movement and data transformation over cloud. Data Factory services also offer to create and run data pipelines that move and transform data and then run the pipeline on a specified schedule.

**What are the different components used in Azure Data Factory?**

Azure Data Factory consists of several numbers of components. Some components are as follows:

* **Pipeline:** The pipeline is the logical container of the activities.
* **Activity:** It specifies the execution step in the [Data Factory pipeline](https://k21academy.com/microsoft-azure/create-azure-data-factory-pipeline/), which is substantially used for data ingestion and metamorphosis.
* **Dataset:** A dataset specifies the pointer to the data used in the pipeline conditioning.
* **Mapping Data Flow:** It specifies the data transformation UI logic.
* **Linked Service:** It specifies the descriptive connection string for the data sources used in the channel conditioning.
* **Trigger:** It specifies the time when the pipeline will be executed.
* **Control flow:** It’s used to control the execution flow of the pipeline activities.

**What is the key difference between the Dataset and Linked Service in Azure Data Factory?**

The dataset specifies a source to the data store described by the linked service. When we put data to the dataset from an SQL Server instance, the dataset indicates the table’s name that contains the target data or the query that returns data from dissimilar tables.

Linked service specifies a definition of the connection string used to connect to the data stores. For illustration, when we put data in a linked service from a SQL Server instance, the linked service contains the name for the SQL Server instance and the credentials used to connect to that case.

**How many types of triggers are supported by Azure Data Factory?**

Following are the three types of triggers that Azure Data Factory supports:

1. **Tumbling Window Trigger:** The Tumbling Window Detector executes the [Azure Data Factory pipelines](https://k21academy.com/microsoft-azure/azure-pipelines-azure-devops-pipeline-concept-uses-its-advantages/) over cyclic intervals. It’s also used to maintain the state of the pipeline.
2. **Event-based Trigger:** The Event-based Trigger creates a response to any event related to blob storage. These can be created when you add or cancel blob storage.
3. **Schedule Trigger:** The Schedule Trigger executes the Azure Data Factory pipelines that follow the wall clock timetable.

**What are the different rich cross-platform SDKs for advanced users in Azure Data Factory?**

The Azure Data Factory V2 provides a rich set of SDKs that we can use to write, manage, and watch pipelines by applying our favourite IDE. Some popular cross-platform SDKs for advanced users in Azure Data Factory are as follows:

* Python SDK
* C# SDK
* PowerShell CLI
* Users can also use the documented REST APIs to interface with Azure Data Factory V2

**What is the difference between Azure Data Lake and Azure Data Warehouse?**



|  |  |
| --- | --- |
| **Azure Data Lake** | **Data Warehouse** |
| Data Lake is a capable way of storing any type, size, and shape of data. | Data Warehouse acts as a repository for already filtered data from a specific resource. |
| It is mainly used by **Data Scientists**. | **Business professionals** tend to use it more often. |
| It is **highly accessible** with quicker updates. | Modifying the Data Warehouse can become a **challenging and expensive endeavor.** |
| It defines the schema after the data is stored successfully. | Datawarehouse defines the schema before storing the data. |
| It uses the **ELT** (Extract, Load, and Transform) process. | It uses**the ETL** (Extract, Transform, and Load) process. |
| It is an ideal platform for doing **in-depth analysis.** | It stands out as the top choice for **operational users**. |

**Difference between Data Lake Storage and Blob Storage.**



|  |  |
| --- | --- |
| **Data Lake Storage** | **Blob Storage** |
| It is an optimized storage solution for **big data analytics workloads**. | Blob Storage is **general-purpose storage** for a wide variety of scenarios. It can also do Big Data Analytics. |
| It follows a **hierarchical file system**. | It utilizes an object store with a straightforward **namespace structure**. |
| In Data Lake Storage, data is stored as**files** inside **folders**. | Blob storage lets you create a **storage account**. Storage account has **containers** that store the data. |
| It can be used to store Batch, interactive, stream analytics,  and machine learning data. | We can use it to store text files, binary data, media storage for streaming, and general-purpose data. |