**Q1) What is Azure Data Factory?**

Azure Data Factory is an integration and ETL service offered by Microsoft. You can create**data-driven workflows** to orchestrate and automate data movement. You can also transform the data over the [cloud](https://cloudkeeda.com/). It lets you create and run data pipelines that can help move and transform data and run scheduled pipelines.

**Q2) Why do we need Azure Data Factory?**

As the world moves to the cloud and big data, data integration and migration remain an integral part of enterprises in all industries. [ADF](https://cloudkeeda.com/azure-data-factory/) helps solve both of these problems efficiently by focusing on the data and planning, monitoring, and managing the ETL / ELT pipeline in a single view.

**The reasons for the growing adoption of Azure Data Factory are:**

* Increased value
* Improved results of business processes
* Reduced overhead costs
* Improved decision making
* Increased business process agility

**Q3) What do we understand by Integration Runtime?**

Integration runtime is referred to as a **compute infrastructure** used by Azure Data Factory. It provides integration capabilities across various network environments.

**A quick look at the Types of Integration Runtimes:**

* **Azure Integration Runtime** – Can copy data between cloud data stores and send activity to various computing services such as SQL Server, Azure HDInsight, etc.
* **Self Hosted Integration Runtime** – It’s basically software with the same code as the Azure Integration Runtime, but it’s installed on your local system or virtual machine over a virtual network.
* **Azure SSIS Integration Runtime** – It allows you to run SSIS packages in a managed environment. So when we lift and shift SSIS packages to the data factory, we use Azure SSIS Integration Runtime.

**Q4) What is the difference between Azure Data Lake and Azure Data Warehouse?**

|  |  |
| --- | --- |
| **Azure Data Lake** | **Data Warehouse** |
| Data Lakes are capable of storing data of any form, size, or shape. | A Data Warehouse is a store for data that has previously been filtered from a specific resource. |
| Data Scientists are the ones who use it the most. | Business professionals are the ones who use it the most. |
| It is easily accessible and receives frequent changes. | Changing the Data Warehouse becomes a very strict and costly task. |
| When the data is correctly stored, it determines the schema. | Before storing the data, the data warehouse defines the schema. |
| It employs the ELT (Extract, Load, and Transform) method. | It employs the ETL (Extract, Transform, and Load) method. |
| It’s an excellent tool for conducting in-depth research. | It is the finest platform for operational users. |

***Check out:***[***Azure Free Trial Account***](https://cloudkeeda.com/azure-free-account/)

**Q5) What is the limit on the number of Integration Runtimes?**

There is no restriction on the number of integration runtime instances that can be used. However, the number of VM cores used by Integration runtime for SSIS package execution is limited to one per subscription.

**Q6) What is Blob Storage in Azure?**

Blob storage is specially designed for**storing a huge amount of unstructured data** such as text, images, binary data. It helps make your data available public globally. The most common use of blob storage is to stream audios and videos, store data for backup, analysis etc. You can also work with data lakes to perform analytics using blob storage.

|  |  |
| --- | --- |
| **Data Lake Storage** | **Blob Storage** |
| It’s a big data analytics workload-optimized storage solution. | Blob Storage is a type of general-purpose storage that can be used in a variety of situations. It’s also capable of Big Data Analytics. |
| A hierarchical file system is used. | It’s based on a flat namespace object store. |
| Data is saved in Data Lake Storage as files within folders. | You can create a storage account with Blob Storage. The data is stored in containers in the storage account. |
| Batch, interactive, stream analytics, and machine learning data can all be stored in it. | Text files, binary data, media storage for streaming, and general-purpose data can all be stored on it. |

**Q7) Difference between Data Lake Storage and Blob Storage.**

**Q8) Describe the process to create an ETL process in Azure Data Factory?**

You can create an ETL process with a few steps.

* Create a service for linked data store i.e. SQL Server Database.
* Let’s consider you have a dataset for vehicles.
* Now for this dataset, you can create a linked service for the destination store i.e. Azure Data Lake.
* Then create a Data Set for Data Saving.
* The next step is to create a pipeline and copy activity. When you are done with creating a pipeline, schedule a pipeline with the use of an added trigger.

**Q9) What is the difference between Azure HDInsight and Azure Data Lake Analytics?**

**Q 10) What are the top-level concepts of Azure Data Factory?**

There are four basic top-level Azure Data Factory concepts:

* **Pipeline** – It acts as a transport service where many processes take place.
* **Activities** – It represents the stages of processes in the pipeline.
* **Datasets** – This is the data structure that holds our data.
* **Linked Services** – These services store information needed when connecting other resources or services. Let’s say we have a SQL server, so we need a connection string that is connected to an external device and we will mention its source and destination.

1. **What is Azure Data Factory?**

Azure Data Factory (ADF) is a cloud-based data integration service that allows users to create, schedule, and manage data pipelines to move and transform data across different platforms and services.

1. **What are the different types of activities in Azure Data Factory?**

There are three types of activities in Azure Data Factory: data movement activities, data transformation activities, and control activities. Data movement activities are used to move data from one source to a destination, data transformation activities are used to transform data in various ways, and control activities are used to control the flow of data within a pipeline.

1. **Explain the difference between a pipeline and an activity in Azure Data Factory.**

A pipeline is a logical grouping of activities that are used to move and transform data, while an activity is a specific operation within a pipeline. A pipeline can contain multiple activities, and each activity can perform a different type of operation.

1. **How do you monitor and troubleshoot Azure Data Factory pipelines?**

Azure Data Factory provides various tools for monitoring and troubleshooting pipelines. These include pipeline runs, activity runs, and data integration performance metrics. In addition, users can use Azure Monitor, Azure Log Analytics, and Azure Data Factory Management API to monitor and troubleshoot pipelines.

1. **What is the role of Integration Runtime in Azure Data Factory?**

Integration Runtime (IR) is a data integration infrastructure that is used to perform data movement and transformation activities in Azure Data Factory. IR provides a scalable and secure way to move and transform data across different platforms and services, and supports various connectivity options, including on-premises, cloud, and hybrid scenarios.

**6: How do you deploy an Azure Data Factory pipeline?**

To deploy an Azure Data Factory pipeline, you can use the Azure portal, Azure PowerShell, or Azure CLI. Here are the general steps:

1. Create an Azure Data Factory instance in your Azure subscription.
2. Create or import the pipeline definition in the Azure Data Factory instance.
3. Configure the pipeline properties, including the data source and destination connections.
4. Validate the pipeline by testing it against sample data.
5. Publish the pipeline to the Azure Data Factory instance.
6. Schedule the pipeline to run on a regular basis or trigger it manually.

**7: What is the difference between a pipeline and an activity in Azure Data Factory?**

A pipeline in Azure Data Factory is a logical grouping of activities that together perform a specific data integration task, such as copying data from one source to another. An activity is a specific operation within a pipeline, such as a data copy or transformation. A pipeline can contain one or more activities, and the activities are executed in the order specified in the pipeline definition.

**8: What is the difference between a linked service and a dataset in Azure Data Factory?**

A linked service is a connection to a data store or service, such as Azure SQL Database, Azure Blob Storage, or Amazon S3. It contains the information required to connect to the data store, such as the server's name, the database name, the authentication method, and the credentials. A dataset is a representation of a data structure in a data store, such as a table, a file, or a folder. It contains the metadata required to read or write data.

**9.What is the difference between ADF and SSIS?**

ADF is a cloud-based data integration service, while SSIS is an on-premises data integration tool. ADF is designed for big data and cloud scenarios, while SSIS is more suited for traditional data warehousing scenarios.

**10. What are the benefits of using ADF for data integration?**

Answer: ADF provides a number of benefits for data integration, including the ability to move data between on-premises and cloud data stores, support for multiple data sources and destinations, and built-in data transformation capabilities. ADF also includes monitoring and logging features that allow users to track the progress of data integration operations.

**11. What is an Integration Runtime in ADF?**

Answer: An Integration Runtime is a compute infrastructure that is used to execute data integration tasks in ADF. It provides the environment for executing data flows, executing SSIS packages, and running custom code.

**12. How do you monitor and troubleshoot data pipelines in ADF?**

ADF provides various monitoring and troubleshooting tools, such as the pipeline run logs, the Data Factory Monitor, and Azure Monitor, to help identify and resolve any issues in your pipelines.

**13. What are the different types of data integration patterns supported by Azure Data Factory?**

Azure Data Factory supports three main data integration patterns: Extract, Transform, Load (ETL), Extract, Load, Transform (ELT), and Extract, Load, Transform, Aggregate (ELTA). Each pattern is used for different data integration scenarios and requirements.

**14. What is a data pipeline in Azure Data Factory, and what are its components?**

A data pipeline in Azure Data Factory is a logical grouping of activities that perform a specific data integration task. Its components include inputs and outputs (represented by linked services and datasets), data transformation activities, and control activities (such as branching and looping).

**15. What is Azure Databricks, and how is it used in conjunction with Azure Data Factory?**

Azure Databricks is a cloud-based analytics platform that provides a unified workspace for data engineering, data science, and machine learning. It can be used in conjunction with Azure Data Factory to perform advanced data transformation and processing tasks, and to build machine learning models for data analysis and prediction.

**16. What is an SSIS package, and how is it different from an ADF pipeline?**

Answer: An SSIS package is a collection of tasks that can be used to extract, transform, and load data. An ADF pipeline is a set of interconnected activities that move and transform data from source to destination.

**17. What are some best practices for designing a data warehouse?**

Answer: Best practices for designing a data warehouse include using a star or snowflake schema, using fact and dimension tables, and designing for scalability and flexibility. You should also consider performance optimization, data quality, and security.

**Q 22) Can we monitor and manage Azure Data Factory Pipelines?**

Yes, we can monitor and manage ADF Pipelines using the following steps:

* Go to the Data factory tab and click on the **monitor and manage.**
* Now click on the **resource manager**.
* You will be able to see pipelines, datasets, and linked services in a tree format.

**Q 23) An Azure Data Factory Pipeline can be executed using three methods. Mention these methods.**

Methods to execute Azure Data Factory Pipeline:

* Debug Mode
* Manual execution using trigger now
* Adding schedule, tumbling window/event trigger

**Q 24) If we need to copy data from an on-premises SQL Server instance using a data factory, which integration runtime should be used?**

Self-hosted integration runtime should be installed on the **on-premises machine** where the SQL Server Instance is hosted.