Azure Data Factory Questions

1. What is Azure Data Factory (ADF)?

Azure Data Factory (ADF) is a cloud-based data integration service that enables you to create data-driven workflows for orchestrating and automating data movement and data transformation. ADF provides a managed service that is continuously monitored and updated, and it provides built-in security features, such as data encryption, identity and access management, and data privacy.

2. What are the core components of Azure Data Factory (ADF)?

The core components of Azure Data Factory (ADF) are:

- **Pipelines:** A pipeline is a logical grouping of activities that perform a specific task.
- **Activities:** An activity is a unit of work within a pipeline.
- **Datasets:** A dataset is a named view of data that is used by activities.
- **Linked services:** A linked service is a connection to a data store or a compute service.
- **Triggers:** A trigger is a mechanism that starts a pipeline run.

3. What are the different types of activities in Azure Data Factory (ADF)?

The different types of activities in Azure Data Factory (ADF) are:

- **Data movement activities:** Data movement activities move data from one location to another.
- **Data transformation activities:** Data transformation activities transform data from one format to another.
- **Control activities:** Control activities control the flow of a pipeline.

4. What is a linked service in Azure Data Factory (ADF)?

A linked service in Azure Data Factory (ADF) is a connection to a data store or a compute service. Linked services are used to connect to various data stores, such as Azure Blob Storage, Azure Data Lake Storage, and Azure SQL Database.

5. What is a dataset in Azure Data Factory (ADF)?

A dataset in Azure Data Factory (ADF) is a named view of data that is used by activities. A dataset represents the input or output of an activity.

6. What is a pipeline in Azure Data Factory (ADF)?

A pipeline in Azure Data Factory (ADF) is a logical grouping of activities that perform a specific task. A pipeline can contain one or more activities, and it can be triggered manually or scheduled to run at a specific time.

7. What is a trigger in Azure Data Factory (ADF)?

A trigger in Azure Data Factory (ADF) is a mechanism that starts a pipeline run. Triggers can be scheduled to run at a specific time, or they can be triggered by an event, such as the arrival of a new file in a data store.

8. What is the difference between a tumbling window and sliding window trigger in Azure Data Factory (ADF)?

A tumbling window trigger in Azure Data Factory (ADF) triggers a pipeline run at a fixed interval, while a sliding window trigger triggers a pipeline run at a sliding interval. For example, a tumbling window trigger might trigger a pipeline run every hour, while a sliding window trigger might trigger a pipeline run every 30 minutes.

9. What is the difference between a single-node and an integrated runtime in Azure Data Factory (ADF)?

A single-node runtime in Azure Data Factory (ADF) is a standalone runtime that is used for data integration tasks, while an integrated runtime is a runtime that is integrated with Azure Data Factory. An integrated runtime provides additional features, such as support for custom activities and integration with Azure DevOps.

10. What is the difference between a tumbling window and a sliding window in Azure Data Factory (ADF)?

A tumbling window in Azure Data Factory (ADF) is a fixed-size window that moves data at regular intervals, while a sliding window is a moving window that moves data based on a specific time interval.

For example, a tumbling window might move data every hour, while a sliding window might move data every 30 minutes, but also include the previous 30 minutes of data.

11. What is the difference between a dataset and a linked service in Azure Data Factory (ADF)?

A dataset in Azure Data Factory (ADF) is a named view of data that is used by activities, while a linked service is a connection to a data store or a compute service. A dataset represents the input or output of an activity, while a linked service is used to connect to various data stores or compute services.

12. What is the difference between a pipeline and a trigger in Azure Data Factory (ADF)?

A pipeline in Azure Data Factory (ADF) is a logical grouping of activities that perform a specific task, while a trigger is a mechanism that starts a pipeline run. A pipeline can contain one or more activities, and it can be triggered manually or scheduled to run at a specific time, while a trigger starts a pipeline run based on a specific event or schedule.

13. What is the difference between a data flow and a mapping data flow in Azure Data Factory (ADF)?

A data flow in Azure Data Factory (ADF) is a data transformation that is executed in a managed runtime, while a mapping data flow is a data transformation that is executed in a Spark runtime. A data flow provides a visual interface for data transformation, while a mapping data flow provides a code-first interface for data transformation.

14. What is the difference between a tumbling window and a tumbling window trigger in Azure Data Factory (ADF)?

A tumbling window in Azure Data Factory (ADF) is a fixed-size window that moves data at regular intervals, while a tumbling window trigger is a trigger that starts a pipeline run at regular intervals. A tumbling window trigger might start a pipeline run every hour, while a tumbling window moves data every hour.

15. What is the difference between a sliding window and a sliding window trigger in Azure Data Factory (ADF)?

A sliding window in Azure Data Factory (ADF) is a moving window that moves data based on a specific time interval, while a sliding window trigger is a trigger that starts a pipeline run based on a specific time interval. A sliding window trigger might start a pipeline run every 30 minutes, while a sliding window moves data every 30 minutes, but also include the previous 30 minutes of data.