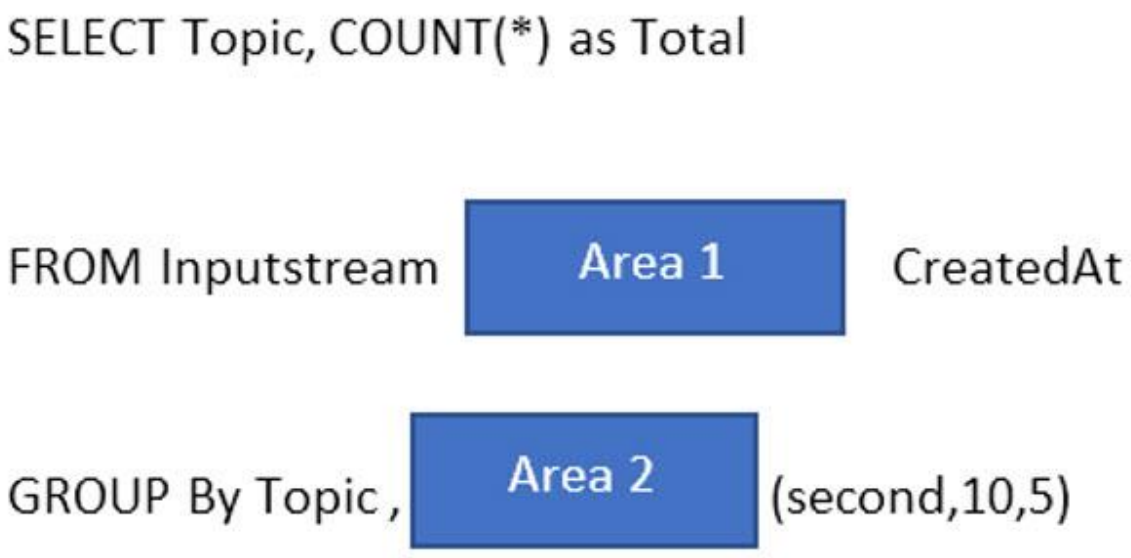
Question 1: **Incorrect**
Your data engineering team needs to develop a solution that will make use of Azure Stream Analytics. The stream will take in data from Azure Event Hubs. Azure Event Hubs has been configured to take in the streaming data from twitter.

You have to write the SQL statement that will be part of the Stream Analytics job. Below is the key requirement of the job

"Provide the count of tweets every 5 seconds over the last 10 seconds"

Below is the SQL statement that needs to be completed



Which of the following will go into Area 1?

- **ORDER BY**

  (Incorrect)

- **COUNT**

- **TIMESTAMP BY**

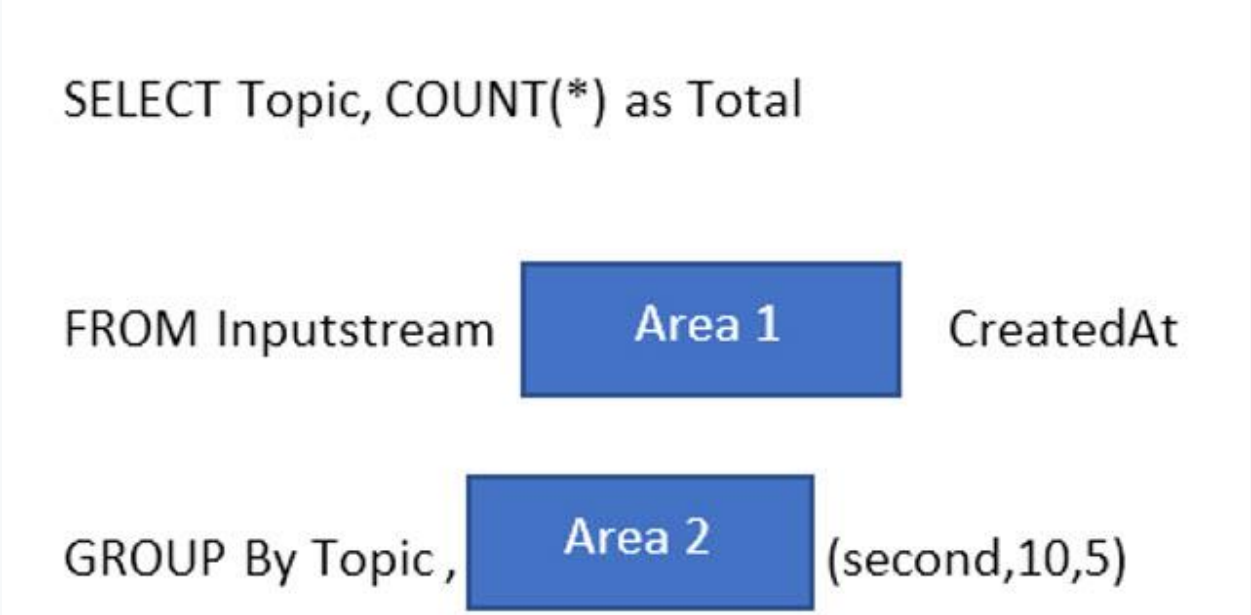
(Correct)

Explanation
Here we are ensuring the events are sorted by the timestamp for each window.
Question 2: **Incorrect**
Your data engineering team needs to develop a solution that will make use of Azure Stream Analytics. The stream will take in data from Azure Event Hubs. Azure Event Hubs has been configured to take in the streaming data from twitter.

You have to write the SQL statement that will be part of the Stream Analytics job. Below is the key requirement of the job

"Provide the count of tweets every 5 seconds over the last 10 seconds"

Below is the SQL statement that needs to be completed



Which of the following will go into Area 2?

- **TumblingWindow**

  (Incorrect)

- **HoppingWindow**

(Correct)

- **SlidingWindow**

Explanation

Here we need to use the Hopping window. Here each window has a size of 10 seconds. And every 5 seconds a new window is created.

This reference has been taken from the below link

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

Question 3: **Correct**

Your team has an Azure Synapse Analytics workspace in place. It also has a Dedicated SQL pool

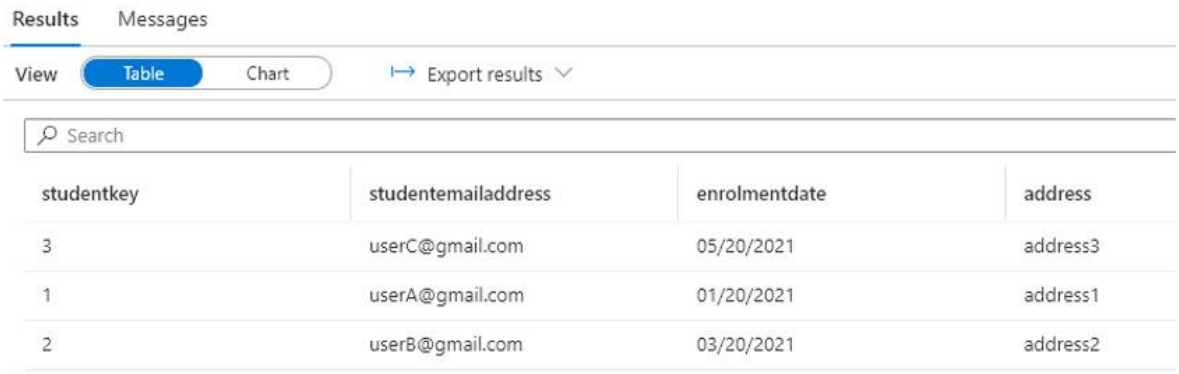
The following table has been created in a Dedicated SQL pool

```
CREATE TABLE [dbo].[DimStudent]
(
    studentkey int NOT NULL,
    studentemailaddress nvarchar(500),
    enrolmentdate   date,
    address   nvarchar(100)
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
)
```

When you issue the following statement

SELECT studentkey,studentemailaddress,format(enrolmentdate,'MM/dd/yyyy') AS enrolmentdate,address FROM [dbo].[DimStudent]

You get the following values



Two users have been defined in the database that are given below

Username	Database role
companyusrA	Server admin
companyusrB	db_datareader

The following masking rules have been applied to the table

Masking rules

Schema	Table	Column	Mask Function
dbo	DimStudent	enrolmentdate	Default value (0, xxxx, ...
dbo	DimStudent	studentemailaddress	Email (aXXX@XXXX.co...
dbo	DimStudent	studentkey	Default value (0, xxxx, ...

What is the value that will be shown to companyusrA when they issue the following statement?

SELECT studentemailaddress FROM [dbo].[DimStudent]

WHERE studentkey=1

- **userA@gmail.com**

  (Correct)

- **uXXX@XXXX.com**

-

userA@XXXX.com

Explanation

Since here the user is an admin level user for the database , the masking rules will not apply

For more information on Dynamic data masking , one can visit the below URL

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

Question 4: **Incorrect**

Your team has an Azure Synapse Analytics workspace in place. It also has a Dedicated SQL pool

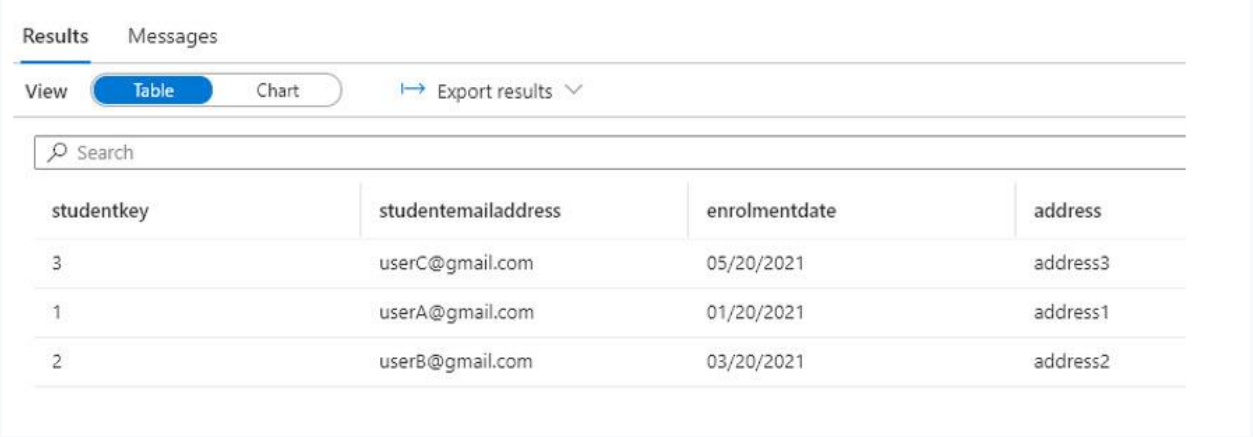
The following table has been created in a Dedicated SQL pool

```
CREATE TABLE [dbo].[DimStudent]
(
    studentkey int NOT NULL,
    studentemailaddress nvarchar(500),
    enrolmentdate   date,
    address   nvarchar(100)
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
)
```

When you issue the following statement

SELECT studentkey,studentemailaddress,format(enrolmentdate,'MM/dd/yyyy') AS enrolmentdate,address FROM [dbo].[DimStudent]

You get the following values



Two users have been defined in the database that are given below

Username	Database role
companyusrA	Server admin
companyusrB	db_datareader

The following masking rules have been applied to the table

Masking rules

Schema	Table	Column	Mask Function
dbo	DimStudent	enrolmentdate	Default value (0, xxxx, ...
dbo	DimStudent	studentemailaddress	Email (aXXX@XXXX.co...
dbo	DimStudent	studentkey	Default value (0, xxxx, ...

What is the value that will be shown to companyusrB when they issue the following statement?

SELECT studentemailaddress FROM [dbo].[DimStudent]

WHERE studentkey=1

- **userA@gmail.com**

  (Incorrect)

-

uXXX@XXXX.com

(Correct)

- **userA@XXXX.com**

Explanation
Here since the Email masking function has been used , the constant prefix of uXXX@XXXX.com will be returned.

For more information on Dynamic data masking , one can visit the below URL

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

Question 5: **Correct**
Your team has an Azure Synapse Analytics workspace in place. It also has a Dedicated SQL pool

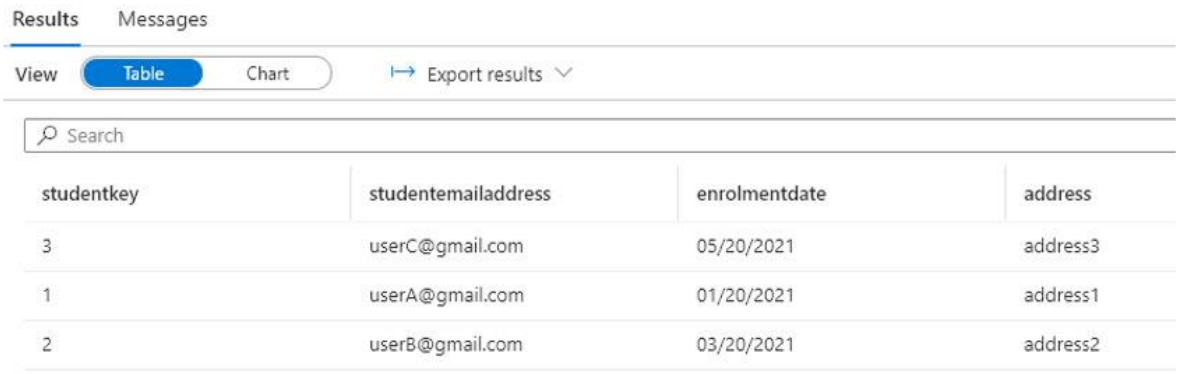
The following table has been created in a Dedicated SQL pool

```
CREATE TABLE [dbo].[DimStudent]
(
    studentkey int NOT NULL,
    studentemailaddress nvarchar(500),
    enrolmentdate   date,
    address   nvarchar(100)
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
)
```

When you issue the following statement

SELECT studentkey,studentemailaddress,format(enrolmentdate,'MM/dd/yyyy') AS enrolmentdate,address FROM [dbo].[DimStudent]

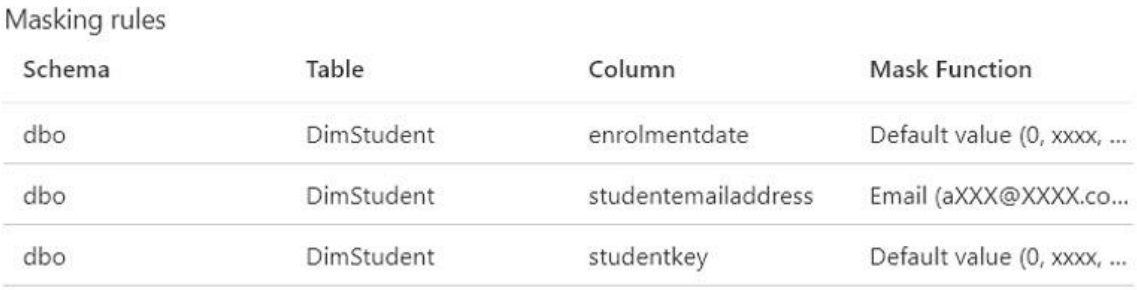
You get the following values



Two users have been defined in the database that are given below

Username	Database role
companyusrA	Server admin
companyusrB	db_datareader

The following masking rules have been applied to the table



What is the value that will be shown to companyusrB when they issue the following statement?

SELECT studentkey FROM [dbo].[DimStudent]

WHERE studentkey=1

- **0**

  (Correct)

- **1**

-

Null value

Explanation

Here since the Default masking function has been used , and since the data type is an integer , the default value of 0 will be returned.

For more information on Dynamic data masking , one can visit the below URL

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

Question 6: **Incorrect**

You currently have a table named logdata in an Azure Dedicated SQL pool within an Azure Synapse Analytics workspace. Currently the table is partitioned based on the Time column in the table. You need to move a partition from the logdata onto a new table. You need to complete the below SQL statements for this requirement

Which of the following would come in Area 1?

- **CREATE**

  (Correct)

- **UPDATE**

  (Incorrect)

- **ALTER**

- **SWITCH**

Explanation
The first step is to create a new table. We then need to switch the partition from the original table to the new table

For more information on Table Partitions , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

Question 7: **Incorrect**
You currently have a table named logdata in an Azure Dedicated SQL pool within an Azure Synapse Analytics workspace. Currently the table is partitioned based on the Time column in the table. You need to move a partition from the logdata onto a new table. You need to complete the below SQL statements for this requirement

Which of the following would come in Area 2?

- **CREATE**
- **UPDATE**

  (Incorrect)
- **ALTER**

  (Correct)
- **SWITCH**

Explanation

Here we need to ALTER the existing table to switch the partitions.

For more information on Table Partitions , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

Question 8: **Incorrect**

You currently have a table named logdata in an Azure Dedicated SQL pool within an Azure Synapse Analytics workspace. Currently the table is partitioned based on the Time column in the table. You need to move a partition from the logdata onto a new table. You need to complete the below SQL statements for this requirement

Which of the following would come in Area 3?

- CREATE
- UPDATE

  (Incorrect)

ALTER

- **SWITCH**

  (Correct)

Explanation
Here we need to use the SWITCH command to switch the partition from the original to the new table.

For more information on Table Partitions , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

Question 9: **Incorrect**
Your team currently has the following resources defined on Azure

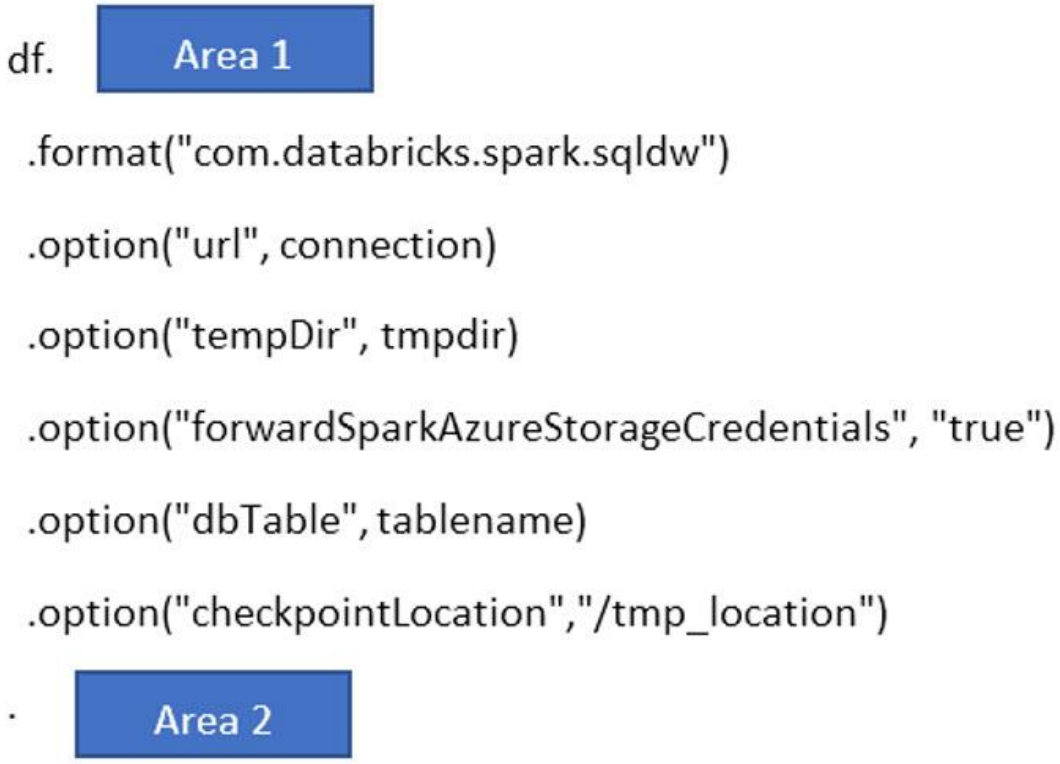
1) An Azure Event Hub namespace and an Event Hub. The Event Hub is used to stream data from an external data source

2) An Azure Databricks cluster

3) An Azure Synapse Analytics workspace that contains a Dedicated SQL Pool

A Notebook is being developed in Scala in Azure Databricks. The Notebook will be used to take in the streaming data from Azure Event Hubs and add it on a table in the Dedicated SQL Pool

Below is a snippet of the code that needs to be completed



Which of the following needs to go into Area 1?

- **start()**
- **update()**

  (Incorrect)
- **writeStream**

  (Correct)
- **write**

Explanation

When it comes to streaming, we must use the writeStream() method.

For more information on using Azure Databricks and Azure Synapse , one can visit the below URL

https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/synapse-analytics

Question 10: **Incorrect**
Your team currently has the following resources defined on Azure

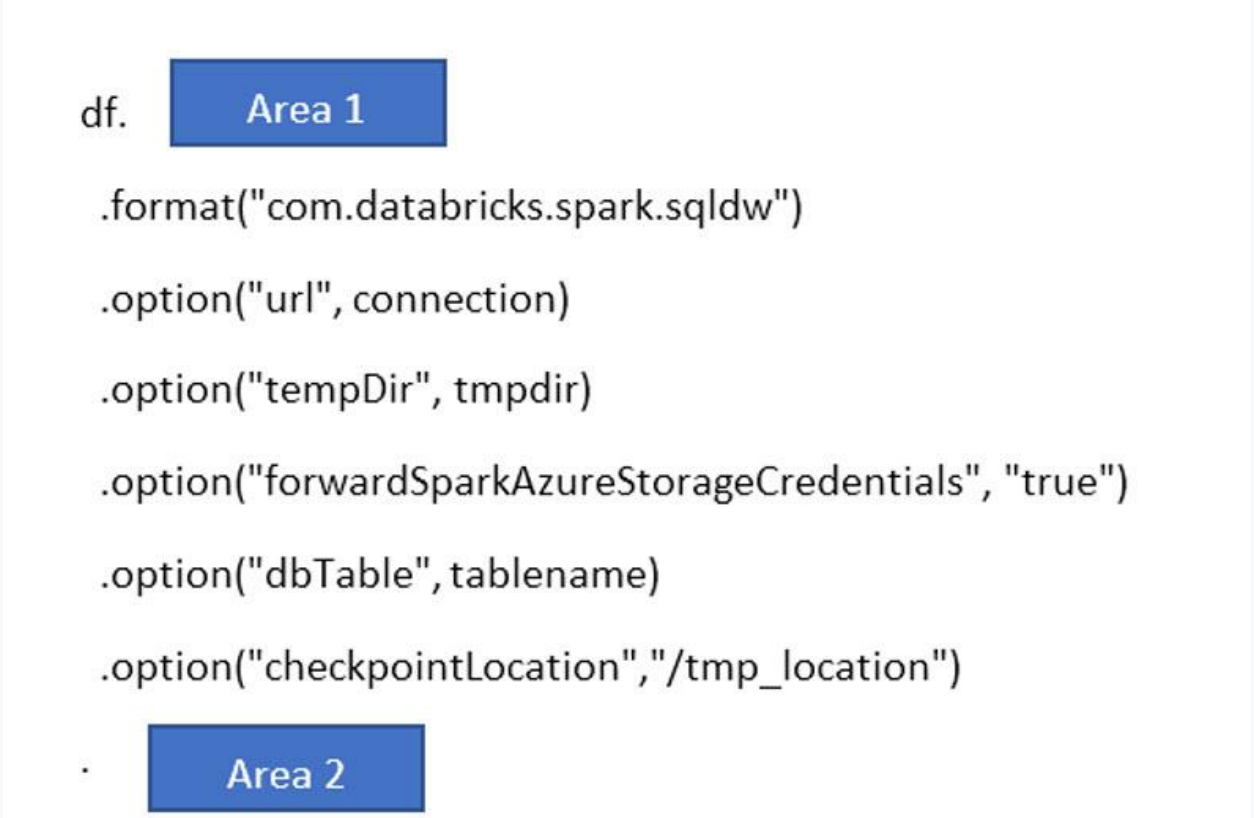
1) An Azure Event Hub namespace and an Event Hub. The Event Hub is used to stream data from an external data source

2) An Azure Databricks cluster

3) An Azure Synapse Analytics workspace that contains a Dedicated SQL Pool

A Notebook is being developed in Scala in Azure Databricks. The Notebook will be used to take in the streaming data from Azure Event Hubs and add it on a table in the Dedicated SQL Pool

Below is a snippet of the code that needs to be completed



Which of the following needs to go into Area 2?

- **start()**

  (Correct)

- **update()**

  (Incorrect)

- **writeStream**

- **write**

Explanation
When it comes to streaming, we must use the start() method to start the streaming process

For more information on using Azure Databricks and Azure Synapse , one can visit the below URL

https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/synapse-analytics

Question 11: **Incorrect**
Your company currently has an on-premises environment. Here they have a Microsoft SQL Server that hosts the database for an ecommerce application. They have also setup an Azure subscription and an Azure Synapse Analytics workspace.

You have to implement the following

1) Create an Orders table in the Azure Synapse Analytics pool. The Orders table will have the following schema

Column name	Column Type
OrderID	int
Quantity	int
Price	decimal
OrderDate	int
Region	varchar(100)

2) Have Azure Data Factory transfer transactional data from the on-premises SQL Server to the Orders table on a weekly basis.

3) The Orders table will contain around 2 billion rows initially

4) Queries on the Orders table will be carried out using the OrderID. It needs to be ensured that queries performing joins with the Orders table completes as quickly as possible

Which of the following would you choose as the table distribution for the Orders table?

- **Hash**

  (Correct)

- **Round-Robin**

  (Incorrect)

- **Replicated**

Explanation

Since this is a large table and a Fact-based table, we should choose Hash distribution for the underlying table.

For more information on designing distributed tables , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute

Question 12: **Incorrect**

Your company currently has an on-premises environment. Here they have a Microsoft SQL Server that hosts the database for an ecommerce application. They have also setup an Azure subscription and an Azure Synapse Analytics workspace.

You have to implement the following

1) Create an Orders table in the Azure Synapse Analytics pool. The Orders table will have the following schema

Column name	Column Type
OrderID	int
Quantity	int
Price	decimal
OrderDate	int
Region	varchar(100)

2) Have Azure Data Factory transfer transactional data from the on-premises SQL Server to the Orders table on a weekly basis.

3) The Orders table will contain around 2 billion rows initially

4) Queries on the Orders table will be carried out using the OrderID. It needs to be ensured that queries performing joins with the Orders table completes as quickly as possible

Which of the following would you choose as the Integration runtime in Azure Data Factory for the pipeline?

- **Azure Integration Runtime**

  (Incorrect)

- **Self-Hosted Runtime**

  (Correct)

- **SSIS-Runtime**

Explanation

Here data needs to be transferred from the on-premises Microsoft SQL Server. Hence you need to install the Integration runtime on the on-premises server itself.

For more information on the self-hosted runtime , one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime?tabs=data-factory

Question 13: **Correct**

Your company currently has an on-premises environment. Here they have a Microsoft SQL Server that hosts the database for an ecommerce application. They have also setup an Azure subscription and an Azure Synapse Analytics workspace.

You have to implement the following

1) Create an Orders table in the Azure Synapse Analytics pool. The Orders table will have the following schema

Column name	Column Type
OrderID	int
Quantity	int
Price	decimal
OrderDate	int
Region	varchar(100)

2) Have Azure Data Factory transfer transactional data from the on-premises SQL Server to the Orders table on a weekly basis.

3) The Orders table will contain around 2 billion rows initially

4) Queries on the Orders table will be carried out using the OrderID. It needs to be ensured that queries performing joins with the Orders table completes as quickly as possible

Which of the following would you choose as the distribution column for the Orders table?

- **OrderID**

  (Correct)

- **OrderDate**

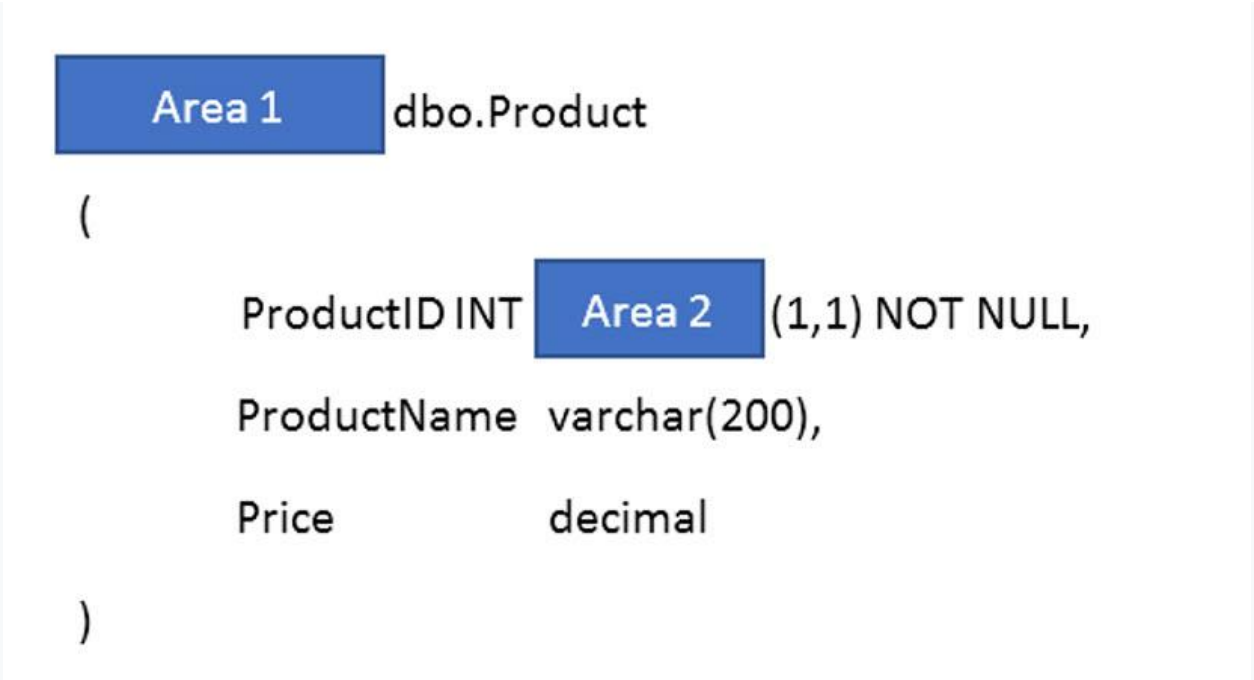
- **Region**

Explanation
Here since the JOINS are being performed on the OrderID, we need to ensure this is chosen as the distribution column.

For more information on designing distributed tables , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute

Question 14: **Correct**
Your company currently has an Azure Synapse Analytics workspace. They want to create a table that will hold product-related data. The data will be stored in the SQL data warehouse itself. The table needs to also have a surrogate key named ProductID in place. You have to create the table. Below is the script that you need to complete for this requirement



Which of the following will come in Area 1?

- **CREATE TABLE**

  (Correct)

- **CREATE EXTERNAL TABLE**

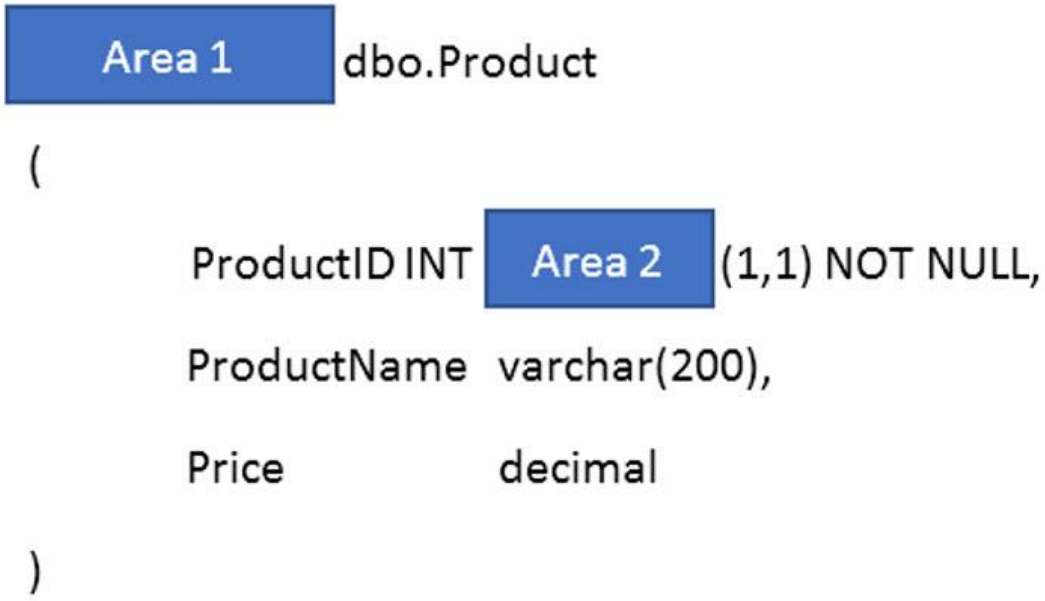
- **CREATE VIEW**

Explanation

Here we need to create a table since the data needs to reside within the SQL data warehouse itself.

Question 15: **Incorrect**

Your company currently has an Azure Synapse Analytics workspace. They want to create a table that will hold product-related data. The data will be stored in the SQL data warehouse itself. The table needs to also have a surrogate key named ProductID in place. You have to create the table. Below is the script that you need to complete for this requirement



Which of the following will come in Area 2?

- **SURROGATE**

  (Incorrect)

- **UNIQUE**

- **IDENTITY**

  (Correct)

Explanation
For implementing surrogate keys in Azure Synapse, we can make use of the IDENTITY column.

For more information on using the IDENTITY column , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity

Question 16: **Incorrect**
Your team has an Azure Data Lake Gen2 storage account named datalakestore. Data is streamed onto the storage account with the use of Azure Event Hubs. You have to create a lifecycle rule that will be based on the following requirements

1) Should be applied to base blobs in a container named data

2) Ensure that if objects are not modified since the last 30 days, they are moved to the Cool tier

3) Ensure that if objects are not modified since the last 90 days, they are moved to the Archive tier

You have to complete the below lifecycle policy for this requirement

Which of the following would go into Area 1?

- “datalakestore/data”

  (Incorrect)

- “data”

- **"tierToCool"**

  (Correct)

- **"tierToArchive"**

Explanation
Here since we need to ensure that if objects are not modified since the last 30 days , they are moved to the Cool Access tier, we need to choose the tier as tierToCool

For more information on Azure Blob storage lifecycle management , one can visit the below URL

https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview

Question 17: **Incorrect**
Your team has an Azure Data Lake Gen2 storage account named datalakestore. Data is streamed onto the storage account with the use of Azure Event Hubs. You have to create a lifecycle rule that will be based on the following requirements

1) Should be applied to base blobs in a container named data

2) Ensure that if objects are not modified since the last 30 days, they are moved to the Cool tier

3) Ensure that if objects are not modified since the last 90 days, they are moved to the Archive tier

You have to complete the below lifecycle policy for this requirement

Which of the following would go into Area 2?

- “datalakestore/data”

  (Incorrect)

- “data”

- "tierToCool"
- "tierToArchive"

  (Correct)

Explanation
Here since we need to ensure that if objects are not modified since the last 90 days , they are moved to the Cool Access tier, we need to choose the tier as tierToArchive

For more information on Azure Blob storage lifecycle management , one can visit the below URL

https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview

Question 18: **Incorrect**
Your team has an Azure Data Lake Gen2 storage account named datalakestore. Data is streamed onto the storage account with the use of Azure Event Hubs. You have to create a lifecycle rule that will be based on the following requirements

1) Should be applied to base blobs in a container named data

2) Ensure that if objects are not modified since the last 30 days, they are moved to the Cool tier

3) Ensure that if objects are not modified since the last 90 days, they are moved to the Archive tier

You have to complete the below lifecycle policy for this requirement

Which of the following would go into Area 3?

- **"datalakestore/data"**

  (Incorrect)

- **"data"**

(Correct)

- **"tierToCool"**
- **"tierToArchive"**

Explanation
Since we need to apply this to the data container, this should be the filter used in the lifecycle policy.

For more information on Azure Blob storage lifecycle management , one can visit the below URL

https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview

Question 19: **Incorrect**
Your team is designing a data engineering solution for your company. Data is going to be ingested via the use of Azure Stream Analytics. The data will then be written as files in an Azure Data Lake Gen2 storage account. Azure Databricks will be used to analyze the data in the storage account. You have to decide on the file format for the files that will be stored in the storage account. Some of the queries used within Azure Databricks will only target certain columns of data within the files. You have to ensure that the queries fired are optimized for speed and efficiency. Which of the following would you choose as the underlying file format for the files?

- **CSV**
- **TSV**

  (Incorrect)
- **JSON**
-

Parquet

(Correct)

Explanation
In such a scenario , Parquet is the ideal file format. It's a columnar-based file format. So querying for just certain columns is very efficient.

For more information on Parquet file format and its benefits , one can visit the below links

https://databricks.com/glossary/what-is-parquet

https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/read-parquet

Question 20: **Incorrect**
You have to design a table in a Dedicated SQL pool in Azure Synapse Analytics. This will be a Fact table that will contain order information. You need to decide on the optimal number of partitions for the table. The table will contain around 2.1 billion rows. What should be the ideal number of partitions set for the table?

- **24**
- **36**

  (Correct)
- **360**
- **4000**

  (Incorrect)

Explanation
This is based on the below Microsoft documentation link

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

Here for optimal performance, you should have a minimum of 1 million rows per distribution and partition. Since there are 60 distributions by default, you already have 60 million rows. Hence 2.1 billion/60 million will give you 36 partitions.

Question 21: **Incorrect**
Your team is planning on using Azure Data Factory to design several ETL workflows. You need to store the pipeline-run data for 60 days. Which of the following would you configure for this requirement?

- **Integration Runtime**
- **Diagnostic settings**

  (Correct)
- **Mapping Data Flow**

  (Incorrect)

Explanation
In order to retain logs for a duration more than 45 days, you need to enable the diagnostic settings for Azure Data Factory.

For more information on monitoring aspects for Azure Data Factory , one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor

Question 22: **Correct**
Your team is planning on using Azure Data Factory to design several ETL workflows. You need to store the pipeline-run data for 60 days. Which of the following could you create for this requirement?

- **Azure SQL Database**
- **Azure Synapse Analytics**
- **Log Analytics workspace**

  (Correct)

Explanation
With Diagnostics setting , you can stream the data onto a storage account, Azure Event Hubs or a Log Analytics workspace.

For more information on monitoring aspects for Azure Data Factory , one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor

Question 23: **Incorrect**
Your data engineering team wants to develop a series of Notebooks for data analytics. Which of the below two services have support to develop and debug Notebooks?

- **Azure Data Factory**
- **Azure Synapse Analytics**

  (Correct)
- **Azure Databricks**

  (Correct)
- **Azure Stream Analytics**

(Incorrect)

Explanation
You can develop and debug Notebooks in Azure Synapse Analytics and Azure Databricks

For more information on using Notebooks in either service , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-development-using-notebooks?tabs=classical

https://docs.microsoft.com/en-us/azure/databricks/notebooks/

Question 24: **Incorrect**
Your company is planning on setting an Azure Synapse Analytics workspace. The workspace will contain a set of SQL data warehouses. The data is going to be ingested into these SQL data warehouse from a data source that has an external IP address of 56.10.2.10. You have to ensure that only this data source can initially connect to the data warehouses. Which of the following would you configure for this requirement?

- **Server-Level IP firewall rule**

  (Correct)

- **Virtual Network service endpoint**

- **Transparent Data Encryption**

  (Incorrect)

Explanation
Here you can configure a Server-Level IP firewall rule to ensure that only a certain IP can create a connection to the data warehouses stored in the Azure Synapse Analytics workspace.

For more information on configuring firewall rules , one can visit the below URL

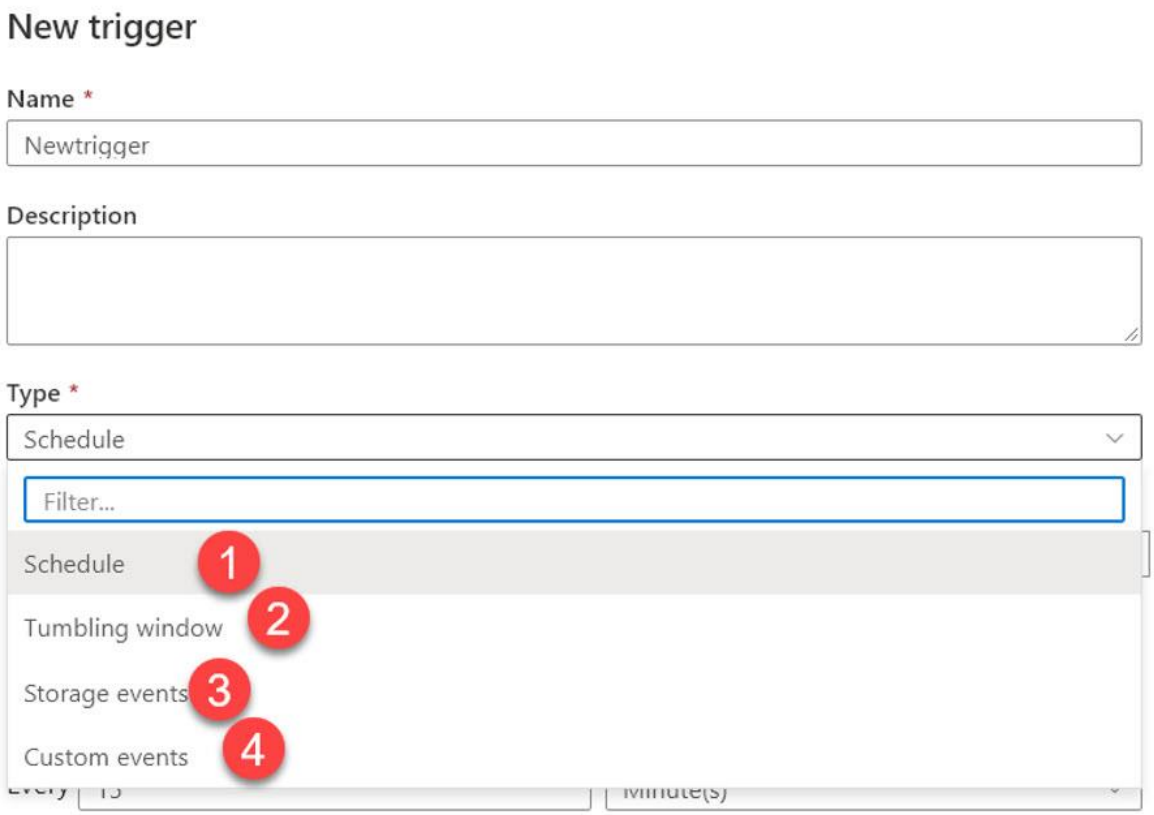
https://docs.microsoft.com/en-us/azure/azure-sql/database/firewall-configure

Question 25: **Correct**
Your team is currently developing an Azure Data Factory pipeline. You have to design a trigger for the pipeline. The pipeline will be used to ingest data from an Azure Data Lake Gen2 Storage account. Below are the key requirements

1) Ensure that data is loaded from the storage account every 20 minutes

2) Cater to late arriving events

Which of the following should you define as the trigger type?



- **Schedule**
- **Tumbling**

  (Correct)

- **Storage events**
- **Custom events**

Explanation

Here the requirement is to continuously load events based on a schedule. But we need to additional functionality to load late-arriving events. This can be achieved with the use of the Tumbling window trigger.

For more information on the tumbling window trigger , one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger?tabs=data-factory

Question 26: **Incorrect**

Your team is currently developing an Azure Data Factory pipeline. You have to design a trigger for the pipeline. The pipeline will be used to ingest data from an Azure Data Lake Gen2 Storage account. Below are the key requirements

1) Ensure that data is loaded from the storage account every 20 minutes

2) Cater to late arriving events

Which of the following would you configure for the late arriving events?

- **Delay**

  (Correct)
- **Max concurrency**
- **Retry policy: count**

  (Incorrect)

Explanation
Here you need to configure the Delay for late arriving events

For more information on the tumbling window trigger , one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger?tabs=data-factory

Question 27: **Incorrect**
Your team needs to develop an Azure Stream Analytics job. The job will ingest data from Azure Event Hubs. Here log data is being ingested. The log data then needs to be matched for records based on a CSV file which will be stored in an Azure Data Lake Gen2 storage account.

Which of the following would you choose as the input type for Azure Event Hubs?

- **Stream**

  (Correct)

- **Reference**

  (Incorrect)

Explanation
Here we need to add the input type as stream type since the data is being streamed from Azure Event Hubs.
Question 28: **Correct**
Your team needs to develop an Azure Stream Analytics job. The job will ingest data from Azure Event Hubs. Here log data is being ingested. The log data then needs to be matched for records based on a CSV file which will be stored in an Azure Data Lake Gen2 storage account.

Which of the following would you choose as the input type for Azure Data Lake Gen2 storage account?

- **Stream**

- Reference

  (Correct)

Explanation
Here we need to add the input type as reference type since the data is reference data being stored in the Azure Data Lake Gen2 storage account.

For more information on reference data , one can visit the below URL

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

Question 29: **Incorrect**
Your team is developing a system that will make use of an Azure Stream Analytics job. You need to add a reference stream to the Stream Analytics job. Which of the following can you add as reference input? Choose 2 answers from the options given below

- Azure SQL Database

  (Correct)

- Azure Data Lake Gen 2 storage accounts

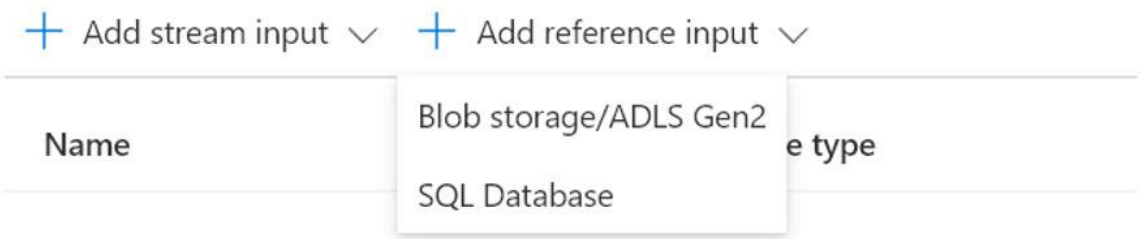
  (Correct)

- Azure Synapse Analytics

- Azure Databricks

  (Incorrect)

Explanation
You can add reference input as shown below in an Azure Stream Analytics job



For more information on reference data , one can visit the below URL

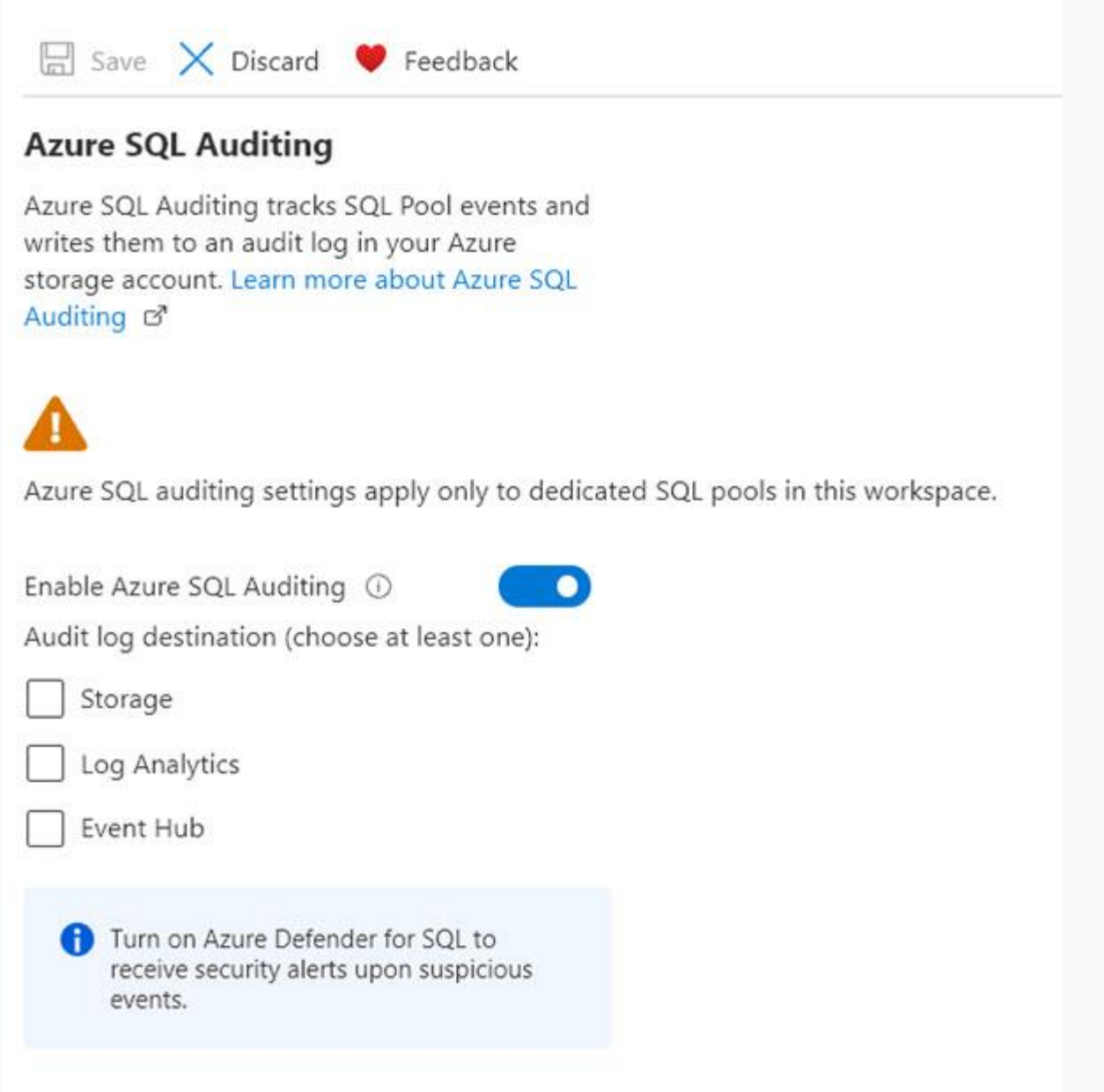
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

Question 30: **Correct**
Your team is currently having an Azure Synapse Analytics workspace in place. A set of SQL data warehouses are going to be hosted in the Synapse workspace. The IT Security team wants to gain insights on the activities happening on the SQL data warehouses. Which of the following should be enabled on the workspace for this requirement?

- **Private endpoint connections**
- **Encryption**
- **Azure SQL Auditing**

  (Correct)

Explanation
For this you should enable Azure SQL Auditing. It gives you insights onto the various SQL data warehousing activities.



For more information on SQL Auditing , one can visit the below URL

https://docs.microsoft.com/en-us/azure/azure-sql/database/auditing-overview

Question 31: **Incorrect**

Your team is currently having an Azure Synapse Analytics workspace in place. A set of SQL data warehouses are going to be hosted in the Synapse workspace. The IT Security team wants to gain insights on the activities happening on the SQL data warehouses. Which of the following needs to be additionally created for this requirement?

- **Azure Log Analytics workspace**

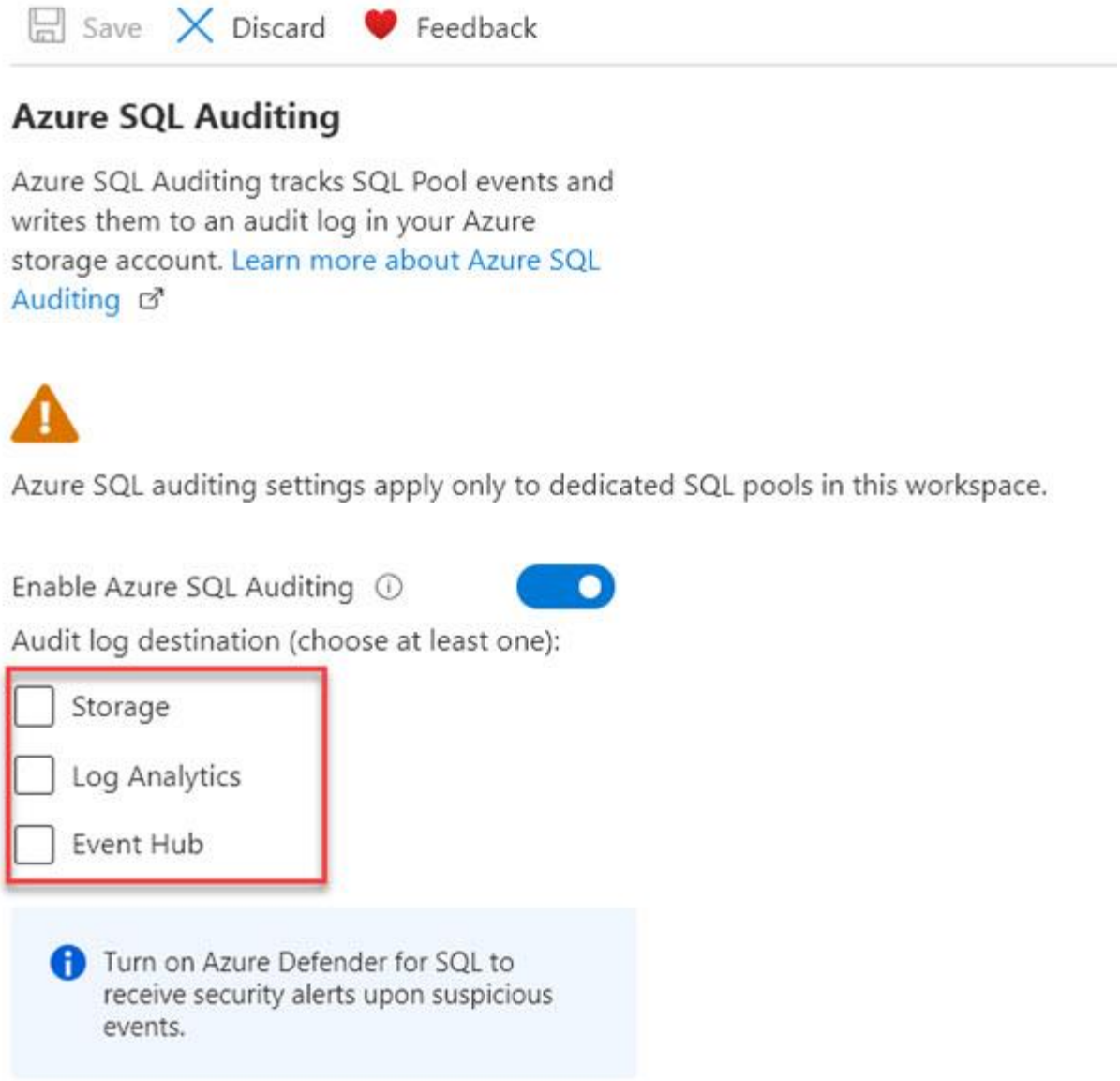
  (Correct)

- **Azure SQL Database**

- **Azure Databricks**

  (Incorrect)

Explanation
Here you need to make use of Azure SQL auditing. And you can stream the audit logs either onto a storage account , a Log Analytics workspace or Azure Event Hubs.



For more information on SQL Auditing , one can visit the below URL

https://docs.microsoft.com/en-us/azure/azure-sql/database/auditing-overview

Question 32: **Incorrect**
Your team has a Dedicated SQL Pool in Azure Synapse Analytics. You have to create a table named Orders_staging that will be used as a staging table. Around a million rows will be loaded every day into the staging table. You have to ensure that rows can be loaded quickly into the staging table.

You have to also ensure that queries against the OrderID are processed faster.

Complete the following script to designing the best way to achieve this

```
CREATE TABLE [dbo].[Orders_staging]
(
    OrderID int NOT NULL,
    OrderName varchar(1000),
    price decimal,
    quantity int
)
WITH
Area 1

CREATE Area 2 OrderIDIndex ON [dbo].[Staging] (OrderID);
```

Which of the following would go into Area 1?

- **HASH DISTRIBUTION**

  (Incorrect)

- **CLUSTERED INDEX**

- **HEAP**

  (Correct)

Explanation
Since this is a staging table, the ideal recommendation would be to create the table as a Heap table.
Question 33: **Incorrect**
Your team has a Dedicated SQL Pool in Azure Synapse Analytics. You have to create a table named Orders_staging that will be used as a staging table. Around a million rows will be loaded every day into the staging table. You have to ensure that rows can be loaded quickly into the staging table.

You have to also ensure that queries against the OrderID are processed faster.

Complete the following script to designing the best way to achieve this

```
CREATE TABLE [dbo].[Orders_staging]
(
    OrderID int NOT NULL,
    OrderName varchar(1000),
    price decimal,
    quantity int
)
WITH
Area 1

CREATE Area 2 OrderIDIndex ON [dbo].[Staging] (OrderID);
```

Which of the following would go into Area 2?

- **CLUSTERED INDEX**

  (Incorrect)

- **INDEX**

  (Correct)

- **PRIMARY INDEX**

Explanation

For the heap table , we can use the INDEX command to create a non-clustered index on the OrderID column of the table. This can help improve the query performance.

For more information on Table Indexes , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-index

Question 34: **Incorrect**

You have an Azure Synapse workspace in place. You have to use the Serverless SQL pool to query for data from an Azure Data Lake Gen2 storage account. The file is a JSON-based file. You need to complete the below script for this requirement

```
SELECT
    TOP 20 *
FROM
    Area 1 (
            bulk 'https://storage10000.blob.core.windows.net/data/order.json',
        format = 'csv',
        fieldterminator ='0x0b',
        fieldquote = '0x0b'
    ) with (doc nvarchar(max)) as rows
```

Which of the following would come in Area 1?

- WHERE
- JOIN

  (Incorrect)
- OPENROWSET

  (Correct)

Explanation
Here we can use the OPENROWSET command to query data in JSON-based files.

For more information on querying JSON-based files , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files

Question 35: **Correct**
You have an Azure Data Lake Gen2 storage account and an Azure Data Factory resource. You need to create a pipeline that will pick up data from the Azure Data Lake Gen2 storage account. You want to ensure that authorization from Azure Data Factory to the storage account is done via Azure Active Directory. Which of the following would you use as the authentication type when setting up the connector?

- **Access Keys**
- **Shared Access Signature**
- **Managed Identity**

  (Correct)

Explanation
You can use Managed Identities to authenticate onto the storage accounts. Here the identities are managed by Azure Active Directory.

For more information on connecting Azure Data Lake Gen2 in Azure Data Factory , one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

Question 36: **Incorrect**
Your team needs to work with an Azure Stream Analytics job. Here the job would ingest data from Azure Event Hubs. The data being ingested is log data from an application. You need to write the query for the job that looks at the growth of CPU percentage log data that is being streamed in a one hour duration. You need to complete the below query for this requirement

```
SELECT metricID,
    change = CPUpercent -
        Area 1 (CPUpercent) OVER (PARTITION BY metricID Area 2 (hour, 1))
FROM input
```

Which of the following would go into Area 1?

- **WHERE**

- **LAG**

  (Correct)

- **LAST**

  (Incorrect)

- **LIMIT DURATION**

Explanation
Here we need to use the LAG operator to look at the previous CPUpercent reading.

This question is based on the example in the below Microsoft documentation link

https://docs.microsoft.com/en-us/stream-analytics-query/lag-azure-stream-analytics

Question 37: **Incorrect**
Your team needs to work with an Azure Stream Analytics job. Here the job would ingest data from Azure Event Hubs. The data being ingested is log data from an application. You need to write the query for the job that looks at the growth of CPU percentage log data that is being streamed in a one hour duration. You need to complete the below query for this requirement

```
SELECT metricID,
    change = CPUpercent -
        Area 1 (CPUpercent) OVER (PARTITION BY metricID Area 2 (hour, 1))
FROM input
```

Which of the following would go into Area 2?

- **WHERE**

LAG

- **LAST**

  (Incorrect)

- **LIMIT DURATION**

  (Correct)

Explanation
Since we need to see the data over an hour, we can use the LIMIT DURATION within the SELECT statement

This question is based on the example in the below Microsoft documentation link

https://docs.microsoft.com/en-us/stream-analytics-query/lag-azure-stream-analytics

Question 38: **Incorrect**
Your team has workloads running in a Dedicated SQL Pool in Azure Synapse Analytics. You need to find the top 5 longest running queries. Which of the following view can help to get this information?

- **sys.dm_pdw_exec_requests**

  (Correct)

- **sys.dm_pdw_request_steps**

- **sys.dm_pdw_dms_workers**

  (Incorrect)

Explanation

You can get this information from the sys.dm_pdw_exec_requests view

For more information on monitoring your SQL data warehouse , one can visit the below URL

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-manage-monitor

Question 39: **Incorrect**
You currently need to process the following JSON document in Azure Databricks

```
{
  "Customers":[
    {
  "customerid": 1,
  "customername": "UserA",
  "registered": true,
  "courses": [
    "AZ-900",
    "AZ-500",
    "AZ-303"
  ] }, {
  "customerid": 2,
  "customername": "UserB",
  "registered": true,
  "courses": [
    "AZ-104",
    "AZ-500",
    "DP-200"
  ]}]}
```

You need to complete the below Notebook script for this requirement

```
import org.apache.spark.sql.functions._

val dfjson = spark.read.format("json")

.option("multiline","true")

.load("/FileStore/json/customer01.json")


val customerjson=dfjson.[Slot1] ( [Slot2] (col("Customers")).alias("Customers"))

val
coursesjson=customerjson.select(col("Customers.customerid").alias("CustomerId"),col("Customers.c
ustomername").alias("CustomerName"),

explode(col("Customers.courses")).alias("Courses"))
```

Which of the following would go into Slot 1?

- **expand**

  (Incorrect)

- **explode**

- **select**

  (Correct)

- **alias**

Explanation

Here we first need to select the column we want to perform the explode function on.

Question 40: **Incorrect**

You currently need to process the following JSON document in Azure Databricks

```
{
   "Customers":[
      {
   "customerid": 1,
   "customername": "UserA",
   "registered": true,
   "courses": [
      "AZ-900",
      "AZ-500",
      "AZ-303"
   ] }, {
   "customerid": 2,
   "customername": "UserB",
   "registered": true,
   "courses": [
      "AZ-104",
      "AZ-500",
      "DP-200"
   ]}]}
```


You need to complete the below Notebook script for this requirement

```
import org.apache.spark.sql.functions._

val dfjson = spark.read.format("json")

.option("multiline","true")

.load("/FileStore/json/customer01.json")


val customerjson=dfjson.[Slot1] ( [Slot2] (col("Customers")).alias("Customers"))

val
coursesjson=customerjson.select(col("Customers.customerid").alias("CustomerId"),col("Customers.c
ustomername").alias("CustomerName"),

explode(col("Customers.courses")).alias("Courses"))
```

Which of the following would go into Slot 2?

- **expand**

  (Incorrect)

- **explode**

  (Correct)

- **select**

- **alias**

Explanation

Then we need to use the explode function to give the elements of the array.

Question 41: **Correct**

You currently have an Azure Databricks workspace. You need to stream data from files that are added onto an Azure Data Lake Gen2 storage account. You need to ensure this is done seamlessly and also has the ability to infer any sort of schema drift. Which of the following would you consider for this requirement?

- **External tables**
- **Using the COPY INTO command**
- **Using Auto Loader**

  (Correct)

Explanation
With the use of Auto Loader , you can directly infer the schema when copying data into Azure Databricks.

Question 42: **Incorrect**
You need to process the following document in Azure Databricks.

```
{
   "Customers":[
      {
   "customerid": 1,
   "customername": "UserA",
   "registered": true,
   "courses": [
      "AZ-900",
      "AZ-500",
      "AZ-303"
   ] }, {
   "customerid": 2,
   "customername": "UserB",
   "registered": true,
   "courses": [
      "AZ-104",
      "AZ-500",
      "DP-200"
   ]}]}
```


Which of the following would you use to retrieve the Customer and courses information in the nested JSON?

- filter

  (Incorrect)

-

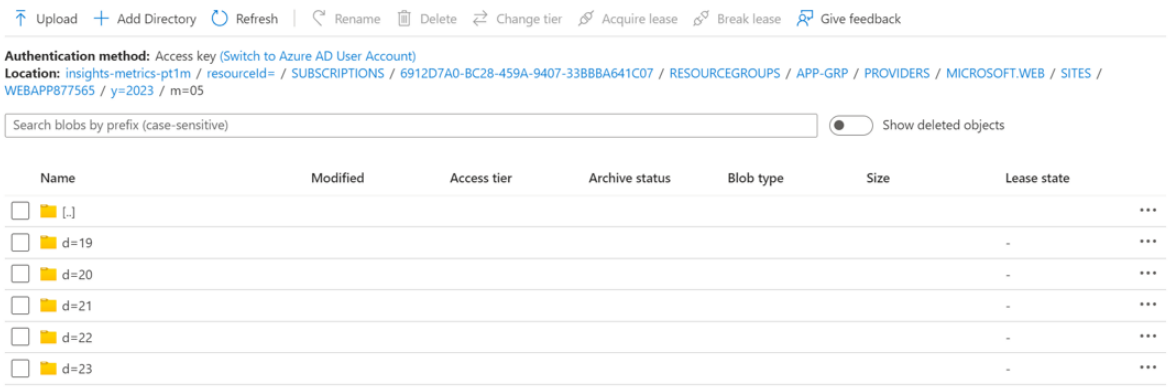
expand

- **explode**

  (Correct)

Explanation
We need to use the explode spark function to retrieve the elements in the JSON document.
Question 43: Incorrect
Your team needs to use Azure Data Factory to copy files every hour from an Azure Data Lake Gen 2 storage account. The Azure Data Lake Gen 2 storage account is used to store metric data.

An example of the folder structure of data is given below



Which of the following trigger would you recommend in Azure Data Factory for the pipeline for the copying of files?

- **Schedule**

- **Storage events**

  (Incorrect)

-

Tumbling Window

(Correct)

Explanation
Since we are dealing with time series events of data, its better to use the Tumbling window so that Azure Data Factory can incrementally copy the data every hour.

Question 44: **Incorrect**
You have the following Azure Data Lake Gen2 account. You need to give access to the data within the account for a certain amount of time. Which of the following should be used ideally to provide access in such a situation?

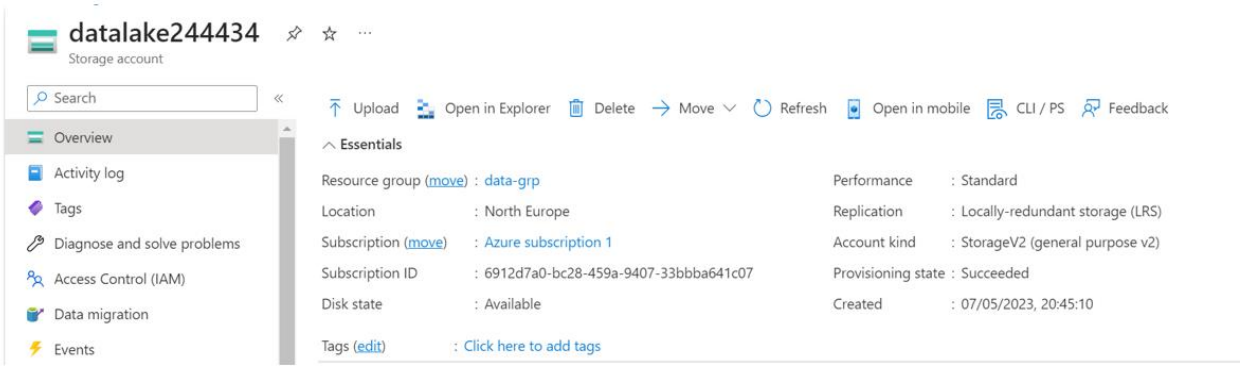
- **Use of Account Access keys**
- **Use of Shared Access Signatures**

  (Correct)
- **Use of Managed Identities**

  (Incorrect)

Explanation
When you want to provide access for a short duration of time, you need to consider the use of Shared Access Signatures.

Question 45: **Correct**
Your team currently has the following Azure Data Lake Gen2 storage account in place.



The storage account has a virtual network service endpoint in place. You need to use Azure Data Factory to copy data from the storage account using Polybase. Which of the following should you use to authorize Azure Data Factory to pick up data from the storage account?

- **Access Keys**
- **Shared Access Signature**
- **Managed Identity**

  (Correct)

Explanation
If you have a virtual network service endpoint enabled and you need to use Polybase , then you need to use Managed Identities to authorize the use of the Azure Data Lake Gen 2 storage account.

One can refer to the below link for more information on the Azure Data Lake storage connector in Azure Data Factory.

https://learn.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage?tabs=data-factory

Question 46: **Incorrect**
Your team has a table defined in a dedicated SQL pool. They want to implement row-level and column-level security for the table.

Which of the following statement is used when it comes to implementing row-level security?

- **Using the GRANT statement**
- **Using the SECURITY POLICY statement**

  (Correct)
- **Using the CREATE TABLE statement**
- **Using the UPDATE statement**

  (Incorrect)

Explanation
When it comes to implementing row-level security we need to use the SECURITY POLICY statement

For more information on row-level security one can refer to the below URL

https://learn.microsoft.com/en-us/sql/relational-databases/security/row-level-security?view=azure-sqldw-latest&preserve-view=true

Question 47: **Incorrect**
Your team has a table defined in a dedicated SQL pool. They want to implement row-level and column-level security for the table.

Which of the following statement is used when it comes to implementing column-level security?

- **Using the GRANT statement**

  (Correct)

- **Using the SECURITY POLICY statement**
- **Using the CREATE TABLE statement**
- **Using the UPDATE statement**

(Incorrect)

Explanation
When it comes to implementing column-level security we need to use the GRANT statement

For more information on column-level security one can refer to the below URL

https://learn.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/column-level-security

Question 48: **Incorrect**
You have a dedicated SQL pool in an Azure Synapse workspace. You need to ensure that data is encrypted at rest when it comes to the dedicated SQL pool. Which of the following would you configure for this requirement?

- **Dynamic Data Masking**
- **Auditing**
- **Transparent Data Encryption**

(Correct)

- 

  Row-Level security

  (Incorrect)

Explanation
For this we need to make use of Transparent Data Encryption
Question 49: **Incorrect**
You are planning on setting up an Azure Synapse Analytics workspace. You want to enable double-encryption when it comes to the workspace. Which of the following is required for this implementation?

- 

  Azure Data Factory

- 

  Azure Key Vault

  (Correct)

- 

  Azure Stream Analytics

- 

  Azure Event Hubs

  (Incorrect)

Explanation
You need to have an Azure Key vault that has purge protection enabled. You need to create an encryption key in the key vault.
Question 50: **Incorrect**
Your team has a table in a dedicated SQL pool. You need to complete the below script to implement row-level security

```
CREATE FUNCTION Security.securitypredicate(@Agent AS nvarchar(50))
   RETURNS TABLE
WITH [Slot 1]
AS
   RETURN SELECT 1 AS securitypredicate_result
WHERE @Agent = USER_NAME() OR USER_NAME() = 'Supervisor';

CREATE [Slot 2] Filter
ADD FILTER PREDICATE Security.securitypredicate(Agent)
ON [dbo].[Orders]
WITH (STATE = ON);
GO
```

Which of the following would go into Slot 1?

- **SECURITY POLICY**
- **SECURITY FILTER**
- **SCHEMABINDING**

  (Correct)

TABLEBINDING

(Incorrect)

Explanation
Here we need to use SCHEMABINDING
Question 51: **Incorrect**
Your team has a table in a dedicated SQL pool. You need to complete the below script to implement row-level security

```
CREATE FUNCTION Security.securitypredicate(@Agent AS nvarchar(50))
   RETURNS TABLE
WITH [Slot 1]
AS
   RETURN SELECT 1 AS securitypredicate_result
WHERE @Agent = USER_NAME() OR USER_NAME() = 'Supervisor';

CREATE [Slot 2] Filter
ADD FILTER PREDICATE Security.securitypredicate(Agent)
ON [dbo].[Orders]
WITH (STATE = ON);
GO
```

Which of the following would go into Slot 2?

- **SECURITY POLICY**

  (Correct)

-

SECURITY FILTER

(Incorrect)

- **SCHEMABINDING**

- **TABLEBINDING**

Explanation
Here we need to use the SECURITY POLICY

Question 52: **Incorrect**
Your team has an Azure Databricks workspace. You need to create a local table from a DataFrame named df. Which of the following statement can be used to create the local table?

- **df.write.saveAsTable("Orders")**

  (Correct)

- **df.write("Orders")**

- **df.createLocalTable("Orders")**

  (Incorrect)

Explanation
You can create a local table with the write method and saveAsTable method.

For more information on creating tables in Azure Databricks , one can visit the below URL

https://docs.microsoft.com/en-us/azure/databricks/data/tables

Question 53: **Correct**
Your team has set up an Azure Stream Analytics job that reads metric data from an Azure Web App resource. You have to formulate a query that would find the last value for a metric within a defined time window. You have to ensure the metric values don't overlap over each time window. How would you complete the query below for this requirement.

```
WITH LastEvent AS (
        SELECT
                MAX(time) AS LastWindowTime
        FROM
                webinput TIMESTAMP BY time
        GROUP BY
                Area 1 (minute, 5)
)
```

```
SELECT
   webinput.time,
   webinput.metricName,
   webinput.average
INTO
   [weboutput]
FROM
   [webinput]
   TIMESTAMP BY time
        INNER JOIN LastEvent
        ON [Area 2] (minute, webinput, LastEvent) BETWEEN 0 AND 5
        AND webinput.time = LastEvent.LastWindowTime
```

Which of the following would go into Area 1?

- **TumblingWindow**

  (Correct)

- **SessionWindow**

- **HoppingWindow**

Explanation

Here since we need to ensure that the events don't overlap, we need to use the TumblingWindow.

Question 54: **Correct**

Your team has set up an Azure Stream Analytics job that reads metric data from an Azure Web App resource. You have to formulate a query that would find the last value for a metric within a defined time window. You have to ensure the metric values don't

overlap over each time window. How would you complete the query below for this requirement.

```
WITH LastEvent AS (
        SELECT
                MAX(time) AS LastWindowTime
        FROM
                webinput TIMESTAMP BY time
        GROUP BY
                Area 1 (minute, 5)
)
```

```
SELECT
   webinput.time,
   webinput.metricName,
   webinput.average
INTO
   [weboutput]
FROM
   [webinput]
   TIMESTAMP BY time
        INNER JOIN LastEvent
        ON  Area 2  (minute, webinput, LastEvent) BETWEEN 0 AND 5
        AND webinput.time = LastEvent.LastWindowTime
```

Which of the following needs to go into Area 2?

- **TIMESTAMP BY**
- **CASE**
- **DATEDIFF**

  (Correct)

Explanation
Here we need to use the DATEDIFF function to find the difference in the time between the events.

Question 55: **Incorrect**
Your team has to generate a set of files that would contain log-based data. Each file would consist of around 20 columns of data. The data engineers would need to query

data from around 5-6 columns of data from the files. You need to decide on the best file format for the files. The read times to the files need to be minimized. And storage size needs to be optimized. Which of the following would you recommend as the file format?

- **JSON**
- **CSV**
- **Parquet**

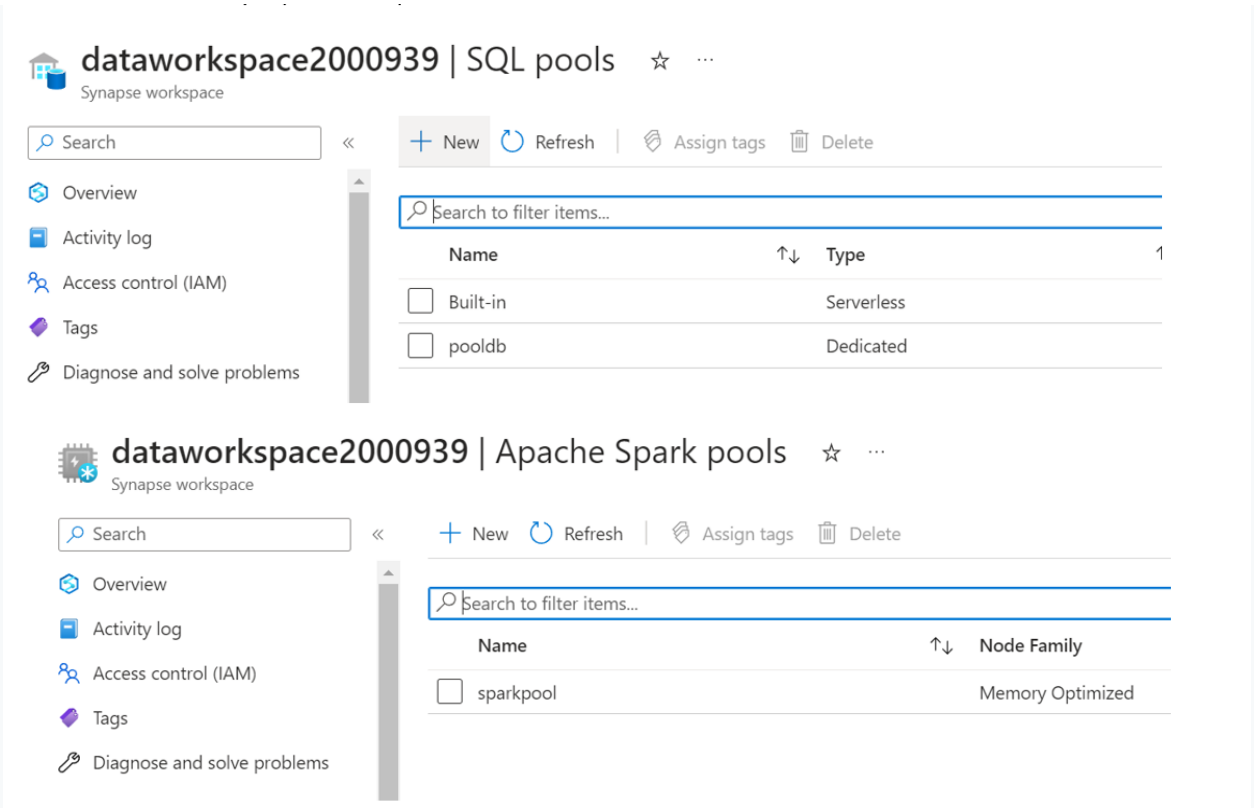
  (Correct)
- **Avro**

  (Incorrect)

Explanation
JSON and CSV are not efficient when it comes to storage size.

Remember that Avro is a row-based file format. And parquet is column-based. Hence when you want to query just certain columns of data, parquet is the best option to choose.

Question 56: Incorrect
You need to create a lake database in Azure Synapse. You currently have the following pools in an Azure Synapse workspace.



Where would you create the table for the lake database?

- **Built-in pool**
- **pooldb**

  (Incorrect)
- **sparkpool**

  (Correct)

Explanation
For the lake database, the tables need to be created in the spark pool.

Question 57: **Correct**

You have a data frame in a spark session in an Azure Spark pool. You need to write the data onto a table in the dedicated SQL pool. You need to complete the below statement for this requirement.

```
import com.microsoft.spark.sqlanalytics

from com.microsoft.spark.sqlanalytics.Constants import Constants

df.write \

   .option(Constants.SERVER,"dataworkspace2000939.sql.azuresynapse.net") \

   .option(Constants.USER,"sqladminuser") \

   .option(Constants.PASSWORD,"sqlpassword@123") \

   .option(Constants.DATA_SOURCE,"pooldb") \

   .option(Constants.TEMP_FOLDER,"abfss://staging@datalake244434.dfs.core.windows.net") \


.option(Constants.STAGING_STORAGE_ACCOUNT_KEY,"dilbGv2rof6G4emB0qWgVwAOOexu/bIpvJiUn
fal7+klHqCsKLB+JkQzMfRlgu0fm14iUFNHXPeU+AStZZXK2w==") \
```

Slot 1 ("overwrite") \

Slot 2 ("pooldb.dbo.logdata")

Which of the following would need to go into Slot 1?

- **mode**

  (Correct)

- **synapsesql**

- **write**

- **read**

Explanation

Here we are specifying the mode as overwrite for the table.

Question 58: **Incorrect**

You have a data frame in a spark session in an Azure Spark pool. You need to write the data onto a table in the dedicated SQL pool. You need to complete the below statement for this requirement.

```
import com.microsoft.spark.sqlanalytics

from com.microsoft.spark.sqlanalytics.Constants import Constants

df.write \

   .option(Constants.SERVER,"dataworkspace2000939.sql.azuresynapse.net") \

   .option(Constants.USER,"sqladminuser") \

   .option(Constants.PASSWORD,"sqlpassword@123") \

   .option(Constants.DATA_SOURCE,"pooldb") \

   .option(Constants.TEMP_FOLDER,"abfss://staging@datalake244434.dfs.core.windows.net") \


.option(Constants.STAGING_STORAGE_ACCOUNT_KEY,"dilbGv2rof6G4emB0qWgVwAOOexu/bIpvJiUn
fal7+klHqCsKLB+JkQzMfRlgu0fm14iUFNHXPeU+AStZZXK2w==") \
```

Slot 1 ("overwrite") \

Slot 2 ("pooldb.dbo.logdata")

Which of the following would need to go into Slot 2?

- **mode**

  (Incorrect)

- **synapsesql**

  (Correct)

- **write**

- **read**

Explanation
We need to use the synapsesql connector to write to a table in the dedicated SQL pool.

Question 59: **Correct**
You are planning on using Azure Synapse database templates. Which of the following is used to group the tables in the template?

- **Functional areas**
- **Facts and Dimensions**
- **Topology**
- **Business Areas**

  (Correct)

Explanation
They are grouped by Business Areas.

Question 60: **Correct**
You are planning on using Azure Databricks to take in streaming data and write the data onto a table. You need to map the right mode when it comes to the requirement of writing data to the table.

“Here only new rows need to be added to the result table”

Which mode would you use for this requirement?

-

- **Complete Mode**

- **Append Mode**

  (Correct)

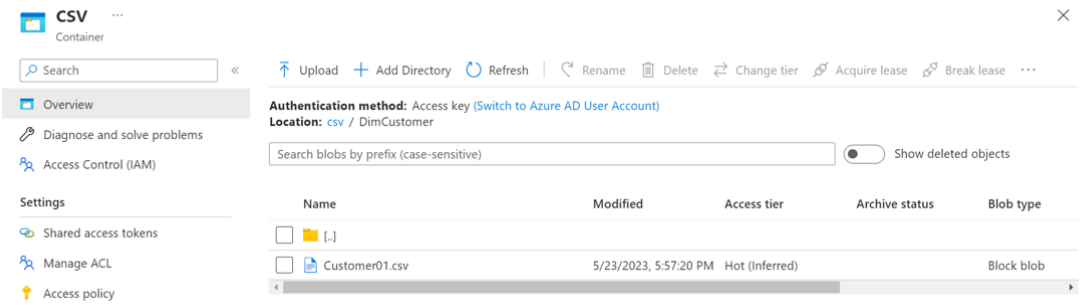
- **Update Mode**

Explanation
When only new rows need to be added to the table , we need to use the Append Mode.

You can refer to the below URL when it comes to the different modes

https://www.databricks.com/spark/getting-started-with-apache-spark/streaming

Question 61: **Incorrect**
You need to provide permissions to a user defined in Azure Active Directory the permission to append data to a blob named Customer01.csv as shown below. The blob is in a DimCustomer folder in a csv container in an Azure Data Lake Gen 2 storage account. You need to provide permissions based on ACL's. You need to provide the least amount of permissions.



Which permission would you grant on the csv container?

- **Read**

  (Incorrect)

-

Write

- **Execute**

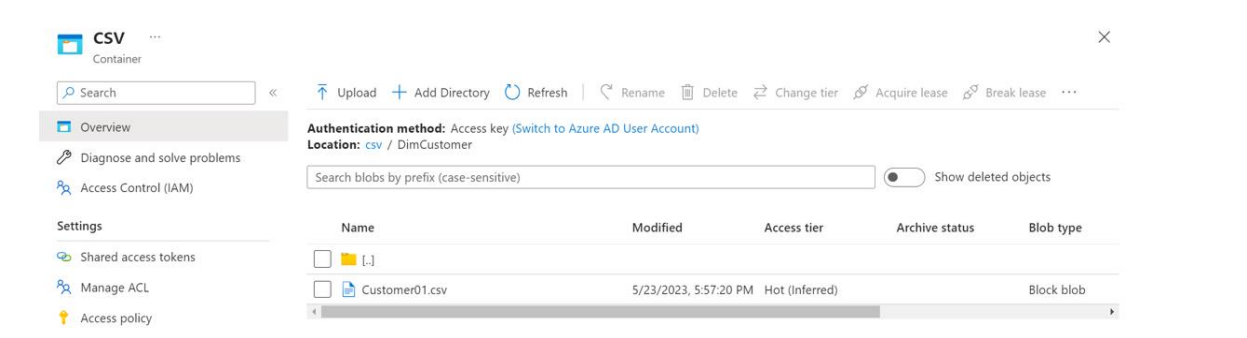
(Correct)

Explanation
In order to read the blob, we need to at least give permissions on the parent container/folders.

An example of this is also given in the Microsoft documentation.

https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control

Question 62: **Incorrect**
You need to provide permissions to a user defined in Azure Active Directory the permission to append data to a blob named Customer01.csv as shown below. The blob is in a DimCustomer folder in a csv container in an Azure Data Lake Gen 2 storage account. You need to provide permissions based on ACL's. You need to provide the least amount of permissions.



Which permission would you grant on the DimCustomer folder?

- **Read**

- **Write**

(Incorrect)

- **Execute**

  (Correct)

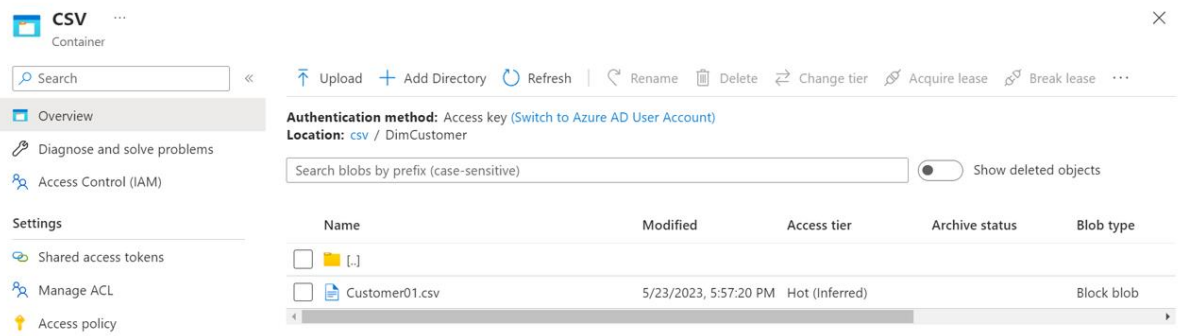
Explanation
In order to read the blob, we need to at least give permissions on the parent container/folders.

An example of this is also given in the Microsoft documentation.

https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control

Question 63: **Incorrect**
You need to provide permissions to a user defined in Azure Active Directory the permission to append data to a blob named Customer01.csv as shown below. The blob is in a DimCustomer folder in a csv container in an Azure Data Lake Gen 2 storage account. You need to provide permissions based on ACL's. You need to provide the least amount of permissions.

permissions.



Which permission would you grant on the Customer01.csv blob?

- **Read**

  (Incorrect)

- **Write**

(Correct)

- **Execute**

Explanation
In order to append data to the blob, we need to give the Execute permission.

Question 64: **Incorrect**
Your team is using an Azure Data Lake Gen2 storage account for the storage of log data. The retention period for the log data is 75 days. You need to ensure costs are minimized when it comes to storing log data. Which of the following would you consider for the tier for the storage of log data?

- **Hot**
- **Cool**

(Correct)

- **Archive**

(Incorrect)

Explanation
Here since the retention period is 75 days, we can consider the Cool tier for the logs. This would save on costs. We are not considering the Archive tier , because you need to ensure that Blobs are retained for at least 180 days.