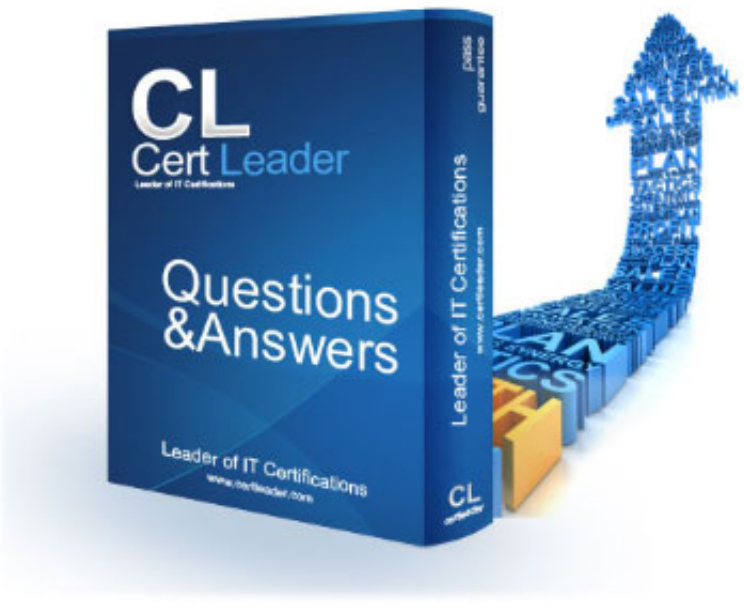
DP-700 Dumps

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

https://www.certleader.com/DP-700-dumps.html



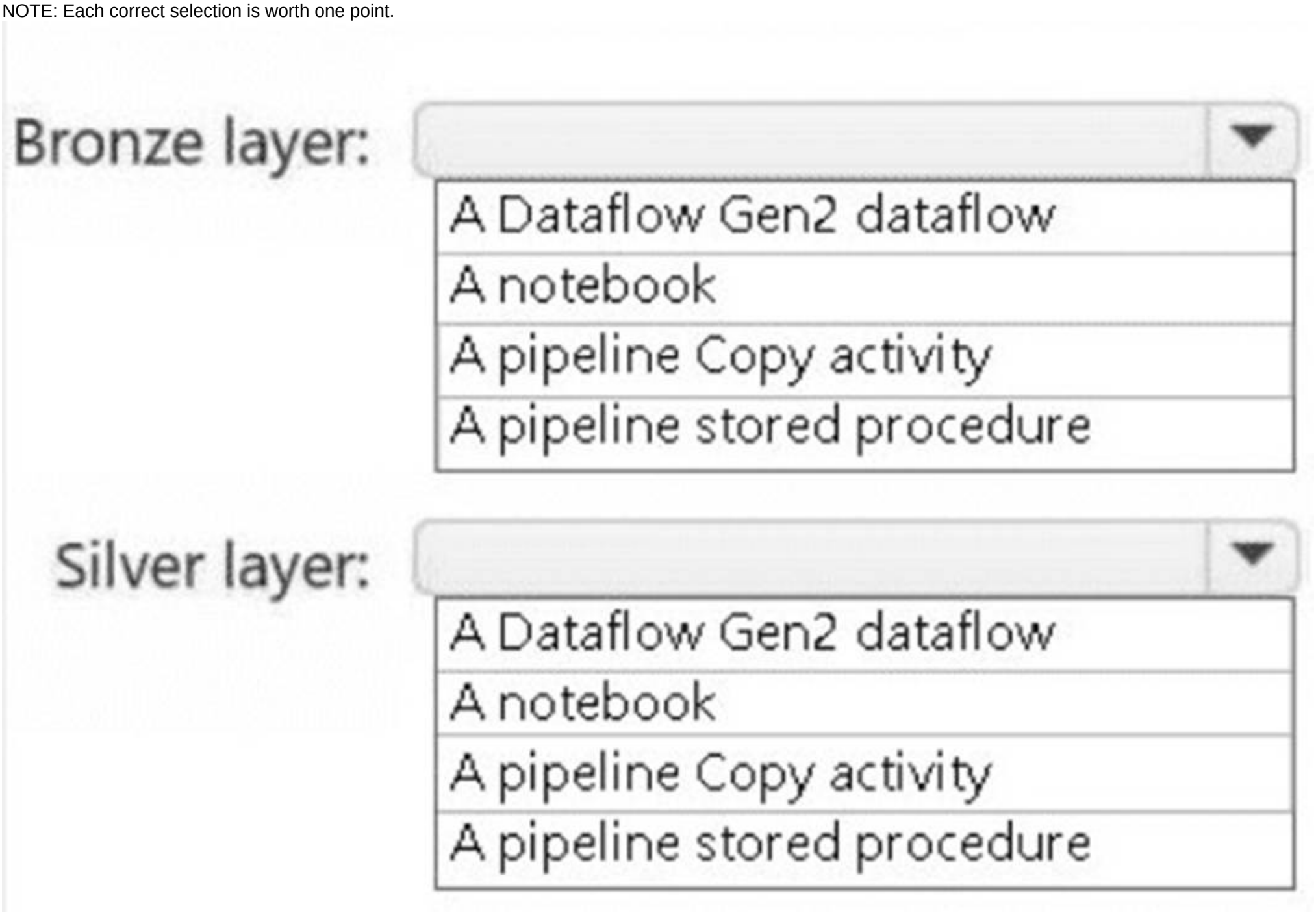
NEW QUESTION 1
- (Topic 1)
You need to ensure that usage of the data in the Amazon S3 bucket meets the technical requirements.
What should you do?

A. Create a workspace identity and enable high concurrency for the notebooks.
B. Create a shortcut and ensure that caching is disabled for the workspace.
C. Create a workspace identity and use the identity in a data pipeline.
D. Create a shortcut and ensure that caching is enabled for the workspace.

Answer: B

Explanation:
To ensure that the usage of the data in the Amazon S3 bucket meets the technical requirements, we must address two key points:
Minimize egress costs associated with cross-cloud data access: Using a shortcut ensures that Fabric does not replicate the data from the S3 bucket into the lakehouse but rather provides direct access to the data in its original location. This minimizes cross-cloud data transfer and avoids additional egress costs.
Prevent saving a copy of the raw data in the lakehouses: Disabling caching ensures that the raw data is not copied or persisted in the Fabric workspace. The data is accessed on- demand directly from the Amazon S3 bucket.

NEW QUESTION 2
HOTSPOT - (Topic 1)
You need to recommend a method to populate the POS1 data to the lakehouse medallion layers.
What should you recommend for each layer? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

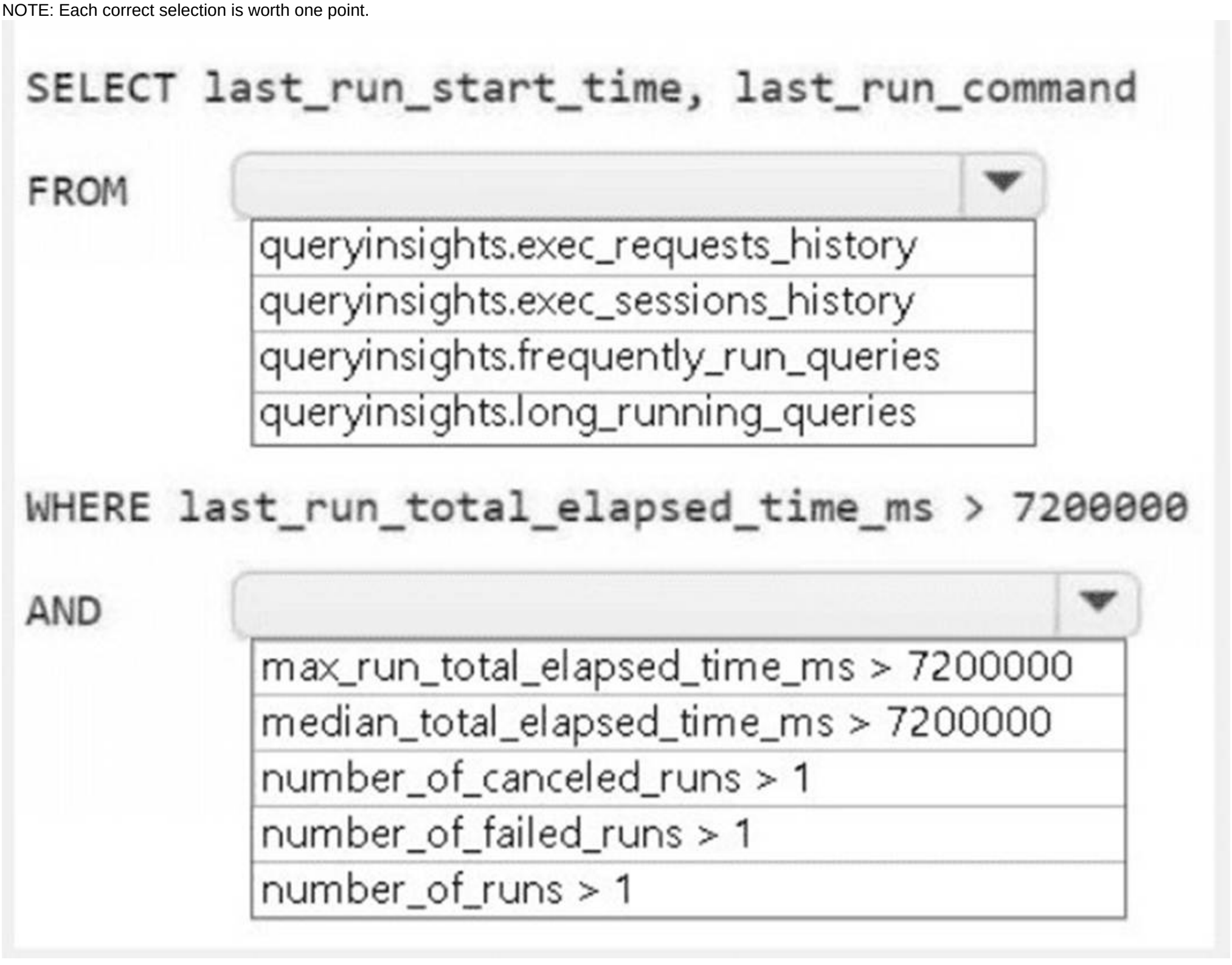


A. Mastered
B. Not Mastered

Answer: A

Explanation:
Bronze Layer: A pipeline Copy activity
The bronze layer is used to store raw, unprocessed data. The requirements specify that no transformations should be applied before landing the data in this layer. Using a pipeline Copy activity ensures minimal development effort, built-in connectors, and the ability to ingest the data directly into the Delta format in the bronze layer.
Silver Layer: A notebook
The silver layer involves extensive data cleansing (deduplication, handling missing values, and standardizing capitalization). A notebook provides the flexibility to implement complex transformations and is well-suited for this task.

NEW QUESTION 3
HOTSPOT - (Topic 2)
You need to troubleshoot the ad-hoc query issue.
How should you complete the statement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
SELECT last_run_start_time, last_run_command: These fields will help identify the execution details of the long-running queries.
FROM queryinsights.long_running_queries: The correct solution is to check the long- running queries using the queryinsights.long_running_queries view, which provides insights into queries that take longer than expected to execute.
WHERE last_run_total_elapsed_time_ms > 7200000: This condition filters queries that took more than 2 hours to complete (7200000 milliseconds), which is relevant to the issue described.
AND number_of_failed_runs > 1: This condition is key for identifying queries that have failed more than once, helping to isolate the problematic queries that cause failures and need attention.

NEW QUESTION 4
- (Topic 3)
You have an Azure event hub. Each event contains the following fields: BikepointID
Street Neighbourhood
Latitude Longitude No_Bikes No_Empty_Docks
You need to ingest the events. The solution must only retain events that have a Neighbourhood value of Chelsea, and then store the retained events in a Fabric lakehouse.
What should you use?

A. a KQL queryset
B. an eventstream
C. a streaming dataset
D. Apache Spark Structured Streaming

Answer: B

Explanation:

An eventstream is the best solution for ingesting data from Azure Event Hub into Fabric, while applying filtering logic such as retaining only the events that have a Neighbourhood value of "Chelsea." Eventstreams in Microsoft Fabric are designed for handling real-time data streams and can apply transformation logic directly on incoming events. In this case, the eventstream can filter events based on the Neighbourhood field before storing the retained events in a Fabric lakehouse. Eventstreams are well-suited for stream processing, such as this case where you need to filter out only specific data (events with a Neighbourhood of "Chelsea") before storing it in the lakehouse.

NEW QUESTION 5
- (Topic 3)
You have a Fabric workspace named Workspacel that contains the following items:
• A Microsoft Power Bl report named Reportl
• A Power Bl dashboard named Dashboardl
• A semantic model named Modell
• A lakehouse name Lakehouse1
Your company requires that specific governance processes be implemented for the items. Which items can you endorse in Fabric?

A. Lakehouse1, Modell, and Dashboard1 only
B. Lakehouse1, Modell, Report1 and Dashboard1
C. Report1 and Dashboard1 only
D. Model1, Report1, and Dashboard1 only
E. Lakehouse1, Model1, and Report1 only

Answer: B

NEW QUESTION 6
- (Topic 3)
You have a Fabric workspace that contains an eventhouse and a KQL database named Database1. Database1 has the following:
A table named Table1 A table named Table2
An update policy named Policy1
Policy1 sends data from Table1 to Table2.
The following is a sample of the data in Table2.

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-18 12:45:17.16524	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6" }]
2024-05-18 12:45:21.76423	bb664e1e-02aa-4e17-8c8a-116cd4458d52	[{ "index": 0, "eventid": "782222b2-fbcb-43c0-82d6-ecd49a99dbf5 }]
2024-05-18 12:45:23.98642	717bfe7d-0e5d-498f-9f21-e60aaf258056	[{ "index": 0, "eventid": "d5730286-0da4-41f8-8e59-f75e209310a9 }]

Recently, the following actions were performed on Table1:
An additional element named temperature was added to the StreamData column. The data type of the Timestamp column was changed to date.
The data type of the DeviceId column was changed to string. You plan to load additional records to Table2.
Which two records will load from Table1 to Table2? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.
A)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-18	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 40, "eventid": "729afca2-be30-4559-bb5e-59feade642f3", "temperature": 32 }]

B)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-21	81416f30	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-5werade642f6", "temperature": 27 }]

C)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-23	81416f3060a24e759b192a84ea05973532dhdyte3	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6" }]

D)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-24	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6" }]

A. Option A
B. Option B
C. Option c
D. Option D

Answer: BD

Explanation:
Changes to Table1 Structure:
StreamData column: An additional temperature element was added. Timestamp column: Data type changed from datetime to date. DeviceId column: Data type changed from guid to string.
Impact of Changes:
Only records that comply with Table2??s structure will load.
Records that deviate from Table2??s column data types or structure will be rejected.
Record B:
Timestamp: Matches Table2 (datetime format). DeviceId: Matches Table2 (guid format).
StreamData: Contains only the index and eventid, which matches Table2. Accepted because it fully matches Table2??s structure and data types.
Record D:
Timestamp: Matches Table2 (datetime format). DeviceId: Matches Table2 (guid format). StreamData: Matches Table2??s structure.
Accepted because it fully matches Table2??s structure and data types.

NEW QUESTION 7
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a notebook named Notebook1.
In Workspace1, you create a new notebook named Notebook2.
You need to ensure that you can attach Notebook2 to the same Apache Spark session as Notebook1.
What should you do?

A. Enable high concurrency for notebooks.

B. Enable dynamic allocation for the Spark pool.
C. Change the runtime version.
D. Increase the number of executors.

Answer: A

Explanation:
To ensure that Notebook2 can attach to the same Apache Spark session as Notebook1, you need to enable high concurrency for notebooks. High concurrency allows multiple notebooks to share a Spark session, enabling them to run within the same Spark context and thus share resources like cached data, session state, and compute capabilities. This is particularly useful when you need notebooks to run in sequence or together while leveraging shared resources.

NEW QUESTION 8
- (Topic 3)
You have an Azure SQL database named DB1.
In a Fabric workspace, you deploy an eventstream named EventStreamDBI to stream record changes from DB1 into a lakehouse.
You discover that events are NOT being propagated to EventStreamDBI.
You need to ensure that the events are propagated to EventStreamDBI. What should you do?

A. Create a read-only replica of DB1.
B. Create an Azure Stream Analytics job.
C. Enable Extended Events for DB1.
D. Enable change data capture (CDC) for DB1.

Answer: D

NEW QUESTION 9
- (Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

Column name	Data type
Timestamp	Datetime
GeoLocation	Dynamic
Temperature	Decimal
DeviceId	Int

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

Both tables contain millions of rows.
You have the following KQL queryset.

```
01 Stream
02 | extend lat = todecimal(GeoLocation.Latitude), long = todecimal(GeoLocation.Longitude)
03 | join kind=inner Reference on DeviceId
04 | project Timestamp, lat, long, Temperature, DeviceName
05 | filter Temperature >= 10
06 | render scatterchart with (kind = map)
```

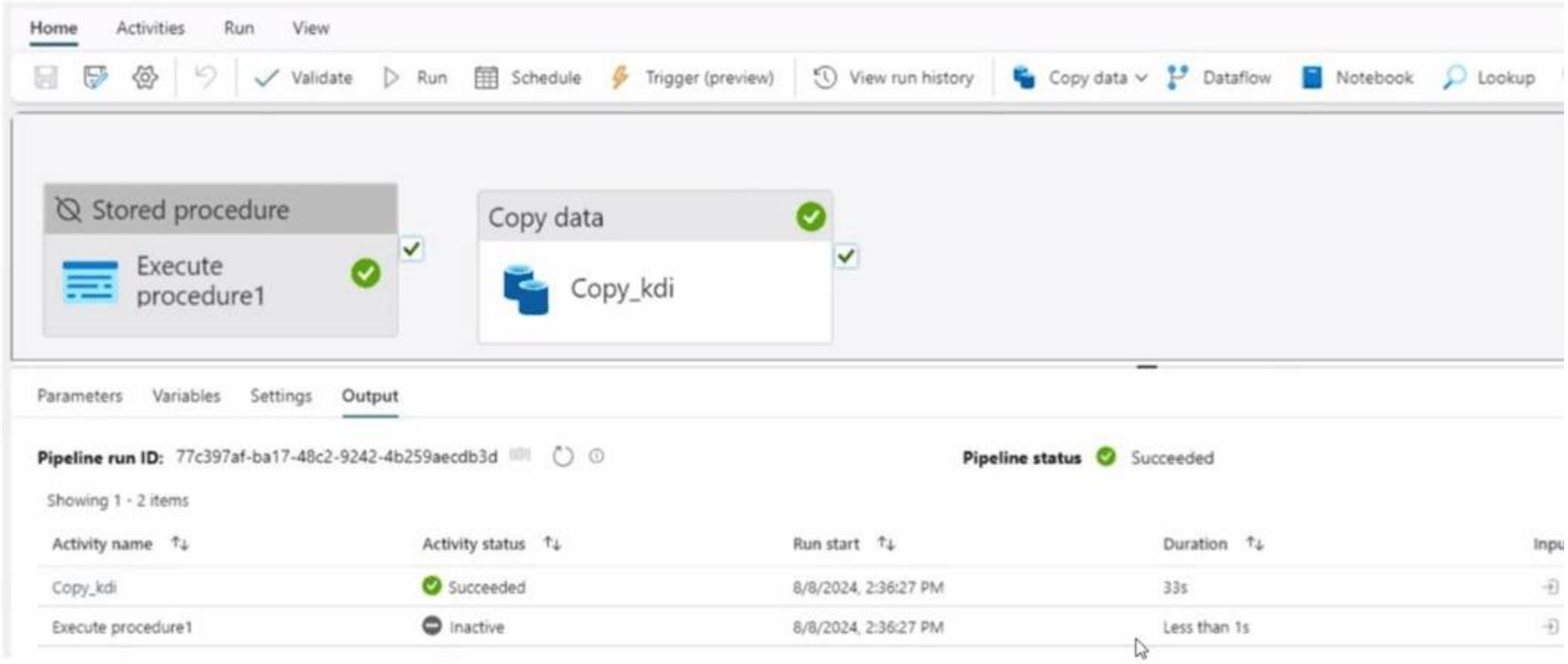
You need to reduce how long it takes to run the KQL queryset. Solution: You change the join type to kind=outer.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
An outer join will include unmatched rows from both tables, increasing the dataset size and processing time. It does not improve query performance.

NEW QUESTION 10
- (Topic 3)
You have a Fabric workspace that contains a data pipeline named Pipeline! as shown in the exhibit.



What will occur the next time Pipelinel tuns?

A. Both activities will run simultaneously.
B. Both activities will be skipped.
C. Execute procedurel will run and Copy_kdi will be skipped.
D. Copy.kdi will run and Execute procedurel will be skipped.
E. Execute procedure1 will run first, and then Copy_kdi will run.
F. Copy.kdi will run first, and then Execute procedurel will run.

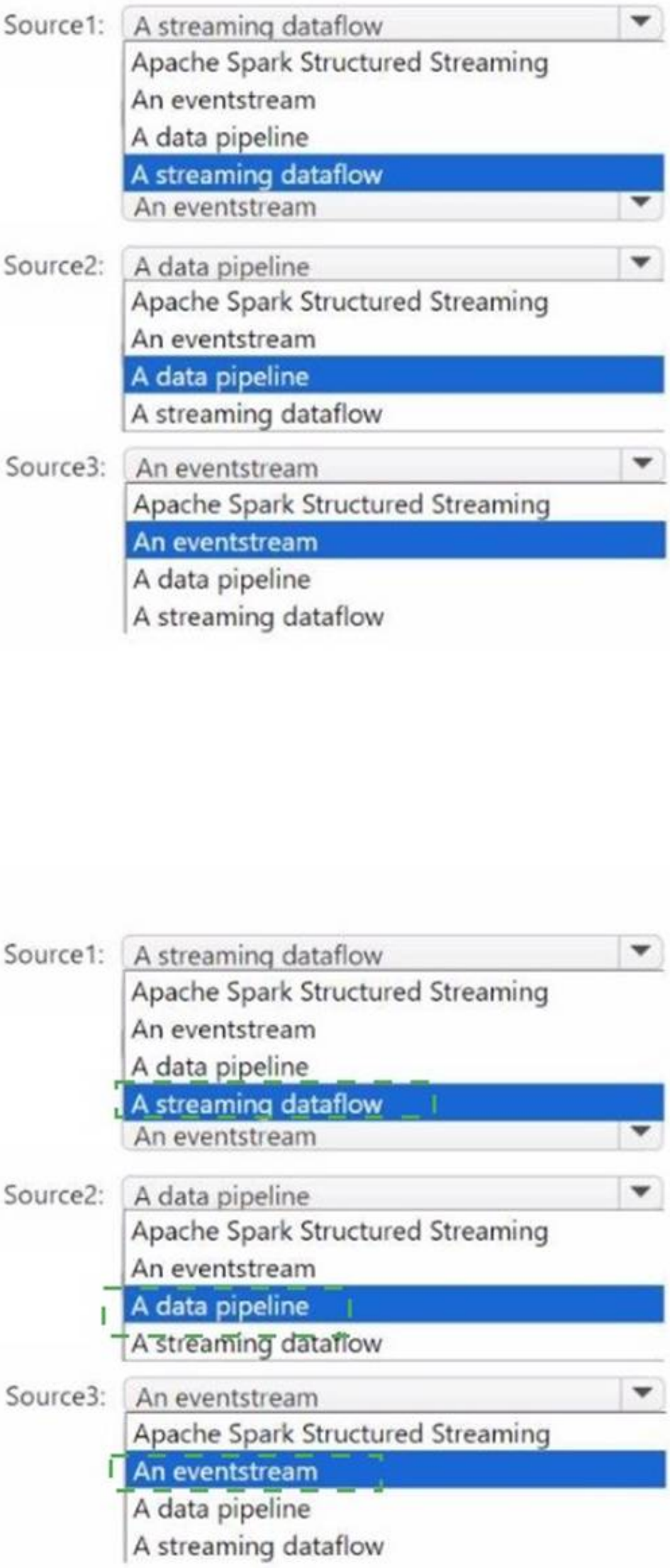
Answer: A

NEW QUESTION 10
HOTSPOT - (Topic 3)
You need to recommend a Fabric streaming solution that will use the sources shown in the following table.

Name	Message size	Description
Source1	10 MB	Contains semi-structured data that has a bigint column in the messages
Source2	25 MB	Contains structured data that has 19 columns
Source3	5 MB	Contains unstructured data that has images in the messages

The solution must minimize development effort.
What should you include in the recommendation for each source? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area



A. Mastered
B. Not Mastered

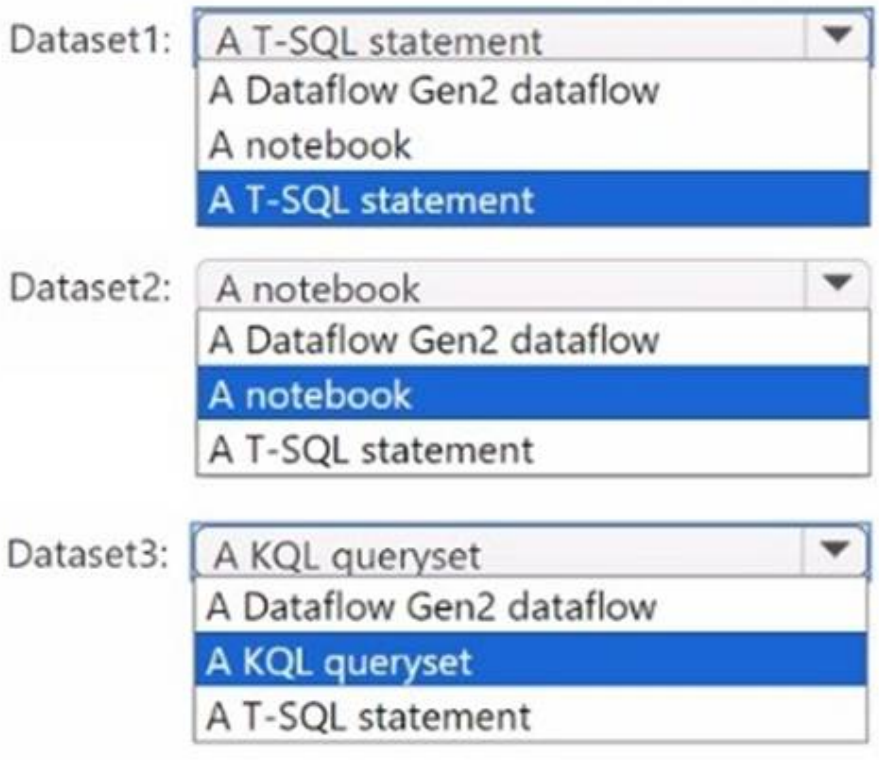
Answer: A

Explanation:

Answer Area

NEW QUESTION 15
HOTSPOT - (Topic 3)
You plan to process the following three datasets by using Fabric:
• Dataset1: This dataset will be added to Fabric and will have a unique primary key between the source and the destination. The unique primary key will be an integer and will start from 1 and have an increment of 1.
• Dataset2: This dataset contains semi-structured data that uses bulk data transfer. The dataset must be handled in one process between the source and the destination. The data transformation process will include the use of custom visuals to understand and work with the dataset in development mode.
• Dataset3. This dataset is in a takehouse. The data will be bulk loaded. The data transformation process will include row-based windowing functions during the loading process.
You need to identify which type of item to use for the datasets. The solution must minimize development effort and use built-in functionality, when possible. What should you identify for each dataset? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

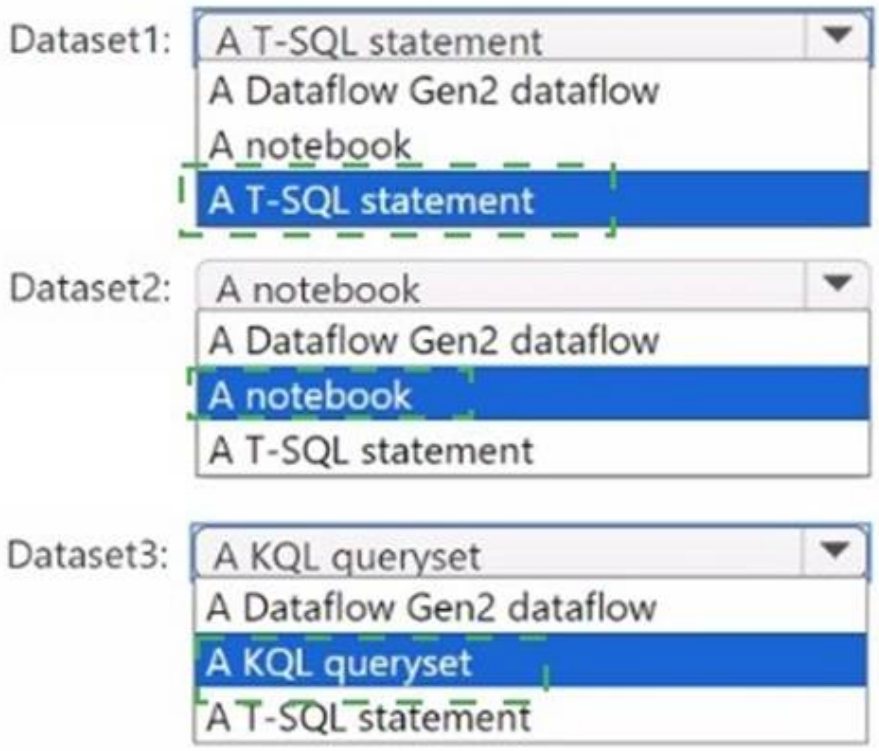


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 20
- (Topic 3)
You have a Fabric workspace named Workspace1. You plan to integrate Workspace1 with Azure DevOps.
You will use a Fabric deployment pipeline named deployPipeline1 to deploy items from Workspace1 to higher environment workspaces as part of a medallion architecture. You will run deployPipeline1 by using an API call from an Azure DevOps pipeline.
You need to configure API authentication between Azure DevOps and Fabric. Which type of authentication should you use?

A. service principal
B. Microsoft Entra username and password
C. managed private endpoint
D. workspace identity

Answer: A

Explanation:
When integrating Azure DevOps with Fabric (Workspace1), using a service principal is the recommended authentication method. A service principal provides a way for applications (such as an Azure DevOps pipeline) to authenticate and interact with resources securely. It allows Azure DevOps to authenticate API calls to Fabric without requiring direct user credentials. This method is ideal for automating tasks such as deploying items through a Fabric deployment pipeline.

NEW QUESTION 21
- (Topic 3)

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table in a lakehouse.
You need to remove files that are older than seven days and are no longer in use. Which command should you run?

A. VACUUM
B. COMPUTE
C. OPTIMIZE
D. CLONE

Answer: A

Explanation:
VACUUM is used to clean up storage by removing files no longer in use by a Delta table. It removes old and unreferenced files from Delta tables. For example, to remove files older than 7 days:
VACUUM delta.`/path_to_table` RETAIN 7 HOURS;

NEW QUESTION 24
- (Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

Column name	Data type
Timestamp	Datetime
GeoLocation	Dynamic
Temperature	Decimal
DeviceId	Int

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

Both tables contain millions of rows. You have the following KQL queryset.

```
01 Stream
02 | extend lat = todecimal(GeoLocation.Latitude), long = todecimal(GeoLocation.Longitude)
03 | join kind=inner Reference on DeviceId
04 | project Timestamp, lat, long, Temperature, DeviceName
05 | filter Temperature >= 10
06 | render scatterchart with (kind = map)
```

You need to reduce how long it takes to run the KQL queryset. Solution: You change project to extend.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Using extend retains all columns in the table, potentially increasing the size of the output unnecessarily. project is more efficient because it selects only the required columns.

NEW QUESTION 27
- (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse1.
You have an on-premises Microsoft SQL Server database named Database1 that is accessed by using an on-premises data gateway.
You need to copy data from Database1 to Warehouse1. Which item should you use?

A. an Apache Spark job definition
B. a data pipeline
C. a Dataflow Gen1 dataflow
D. an eventstream

Answer: B

Explanation:
To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Fabric, a data pipeline is the most appropriate tool. A data pipeline in Fabric is designed to move data between various data sources and destinations, including on-premises databases like SQL Server, and cloud-based storage like Fabric warehouses. The data pipeline can handle the connection through an on-premises data gateway, which is required to access on-premises data. This solution facilitates the orchestration of data movement and transformations if needed.

NEW QUESTION 31
DRAG DROP - (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1.
In Warehouse1, you create a table named DimCustomer by running the following statement.

```
CREATE TABLE dbo.DimCustomer (
CustomerKey VARCHAR(255) NOT NULL,
Name VARCHAR(255) NOT NULL,
Email VARCHAR(255) NOT NULL
);
```

You need to set the Customerkey column as a primary key of the DimCustomer table. Which three code segments should you run in sequence? To answer, move the appropriate
code segments from the list of code segments to the answer area and arrange them in the correct order.

Code Segments

- DROP CONSTRAINT PK_DimCustomer
- ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)
- NOT ENFORCED
- ALTER TABLE dbo.DimCustomer
- ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)
- ENFORCED

Answer Area

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Code Segments

```
DROP CONSTRAINT PK_DimCustomer
ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED
(CustomerKey)
NOT ENFORCED
ALTER TABLE dbo.DimCustomer
ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
(CustomerKey)
ENFORCED
```

Answer Area

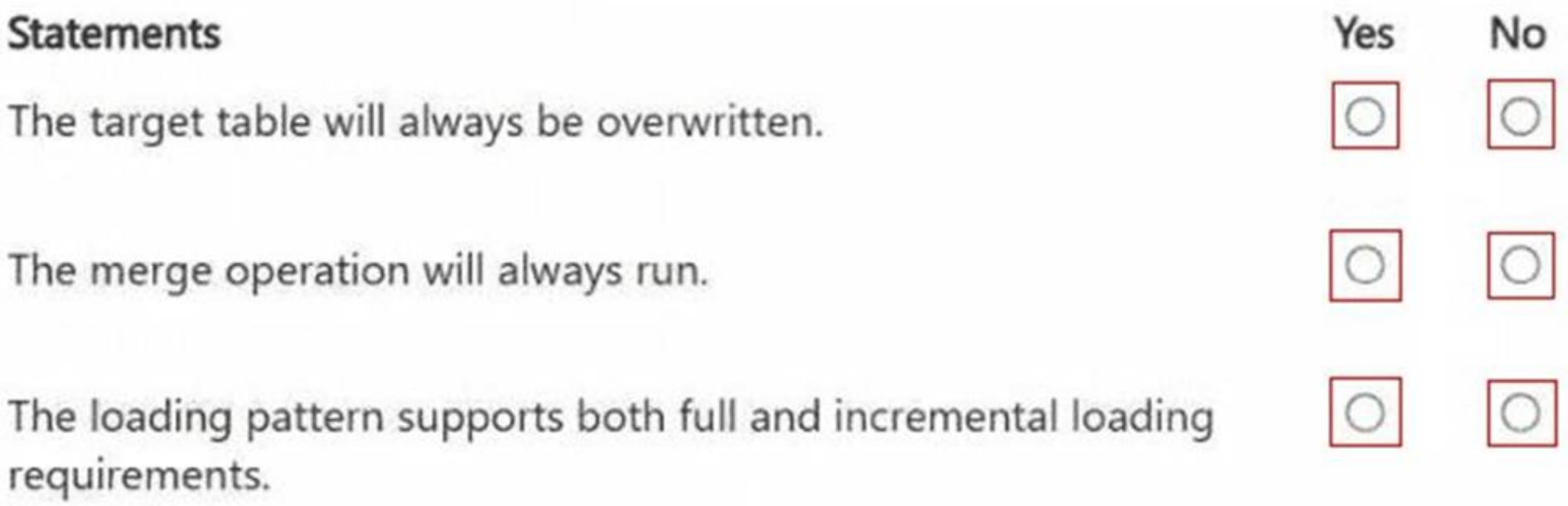
```
ALTER TABLE dbo.DimCustomer
ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
(CustomerKey)
ENFORCED
```

NEW QUESTION 34
HOTSPOT - (Topic 3)
You are building a data loading pattern for Fabric notebook workloads. You have the following code segment:

```
def loading_pattern_sample(df_source):
    try:
        deltaTable = DeltaTable.forName(spark, target_table)
    except Exception:
        try:
            df_source.write.format('delta').mode('overwrite').saveAsTable(f"{target_table}")
        except Exception as e:
            print(f':Load for table {target_table} failed with error: {str(e)}')
            raise
        return

    try:
        change_detection_columns = [col for col in df_source.columns if col not in candidate_key]

        match_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate_key])
        update_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change_detection_columns])

        update_expr = {col: f'source.{col}' for col in df_source.columns}

        merge_operation = deltaTable.alias('target').merge(
            source=df_source.alias('source'),
            condition=match_condition
        ).whenMatchedUpdate(
            condition=update_condition,
            set=update_expr
        ).whenNotMatchedInsertAll()

        merge_operation.execute()
    except Exception as e:
            print(f'Insert operation for table {target_table} failed with error: {str(e)}')
    return
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Answer Area

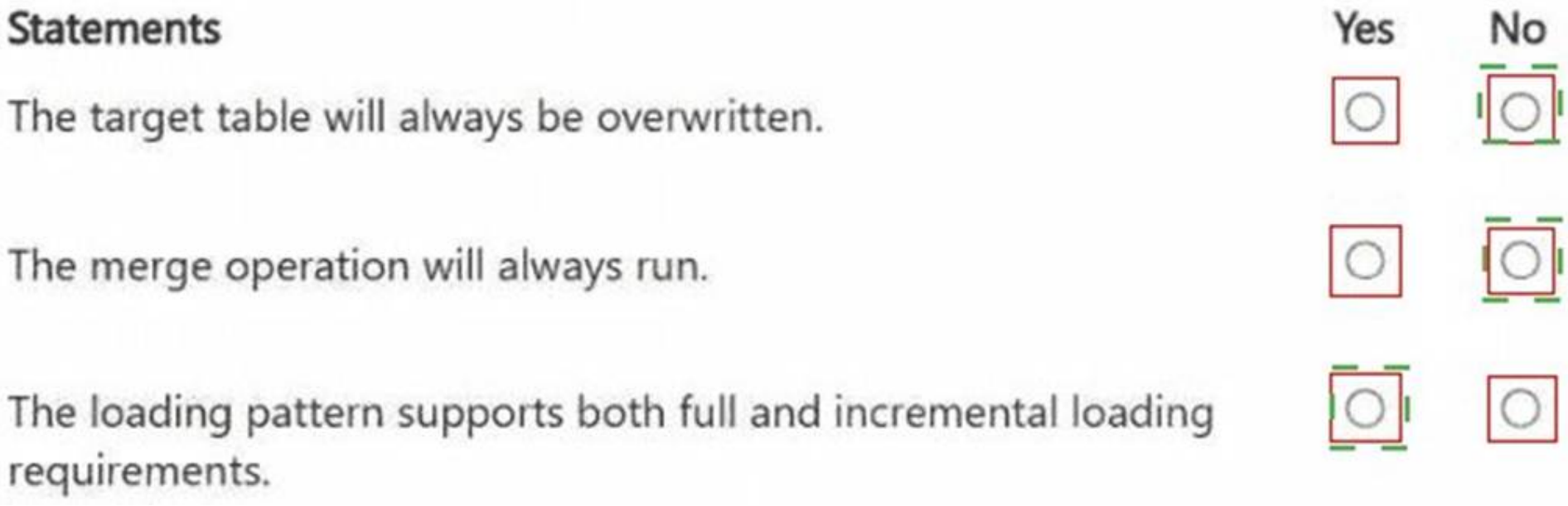


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 37

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse!. Warehousel contains a table named DimCustomers. DimCustomers contains the following columns:

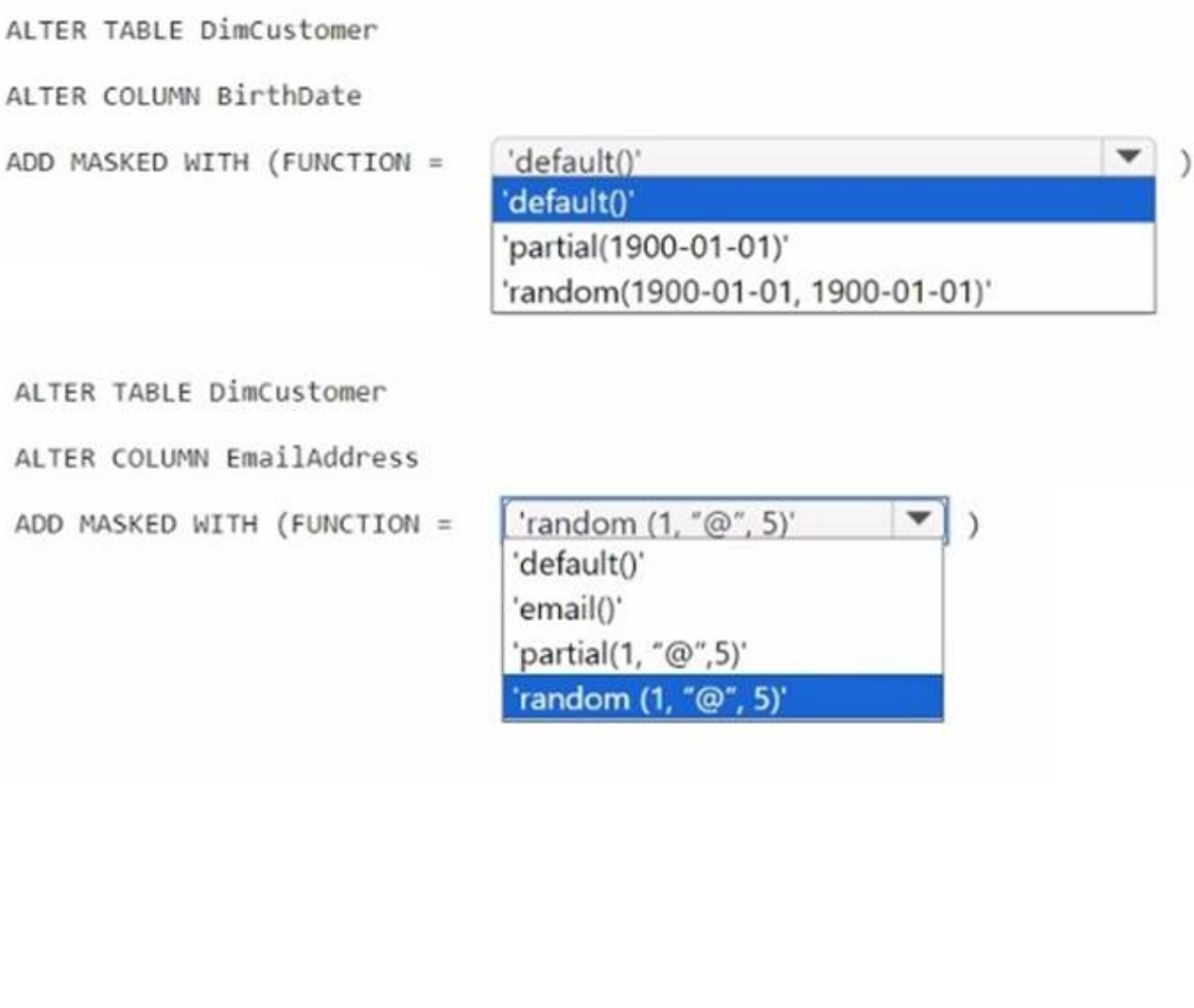
- CustomerName
- CustomerlD
- BirthDate
- Email

You need to configure security to meet the following requirements:

- BirthDate in DimCustomer must be masked and display 1900-01-01.
- Email in DimCustomer must be masked and display only the first leading character and the last five characters.

How should you complete the statement? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

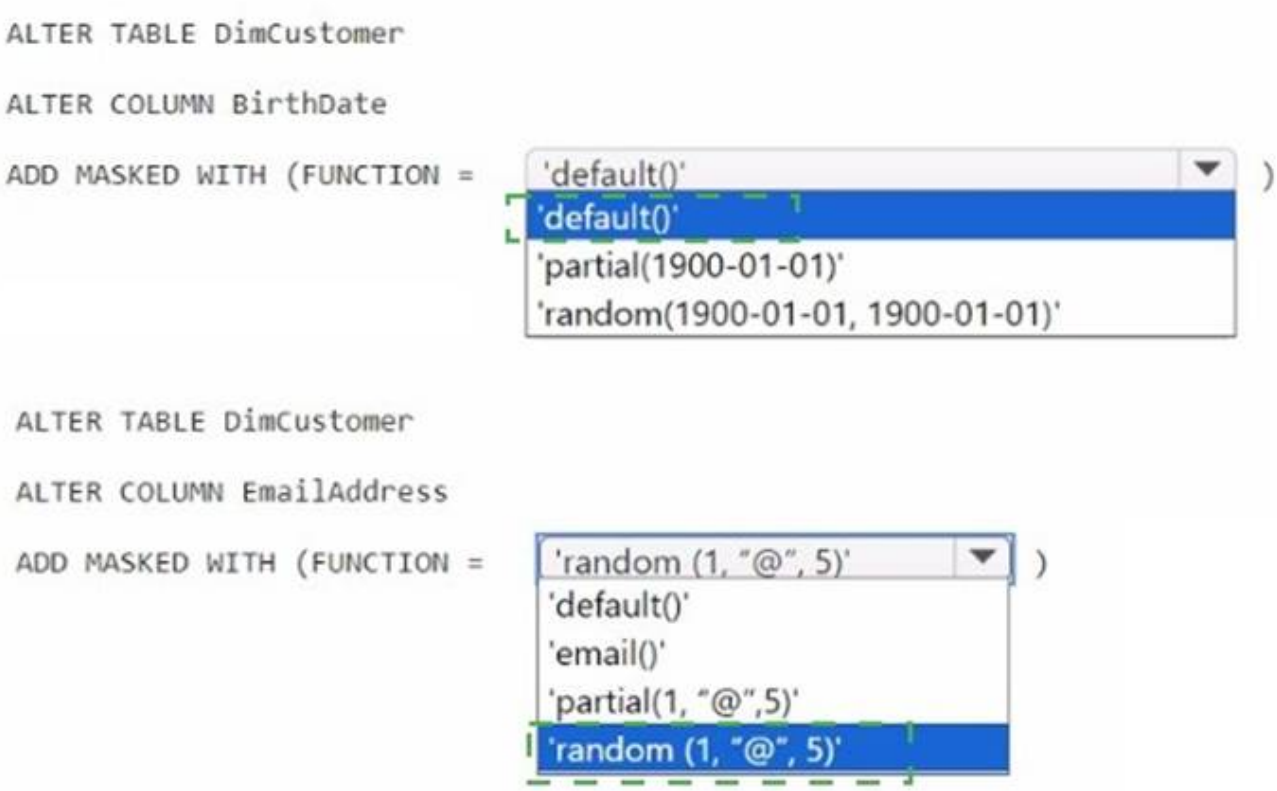


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

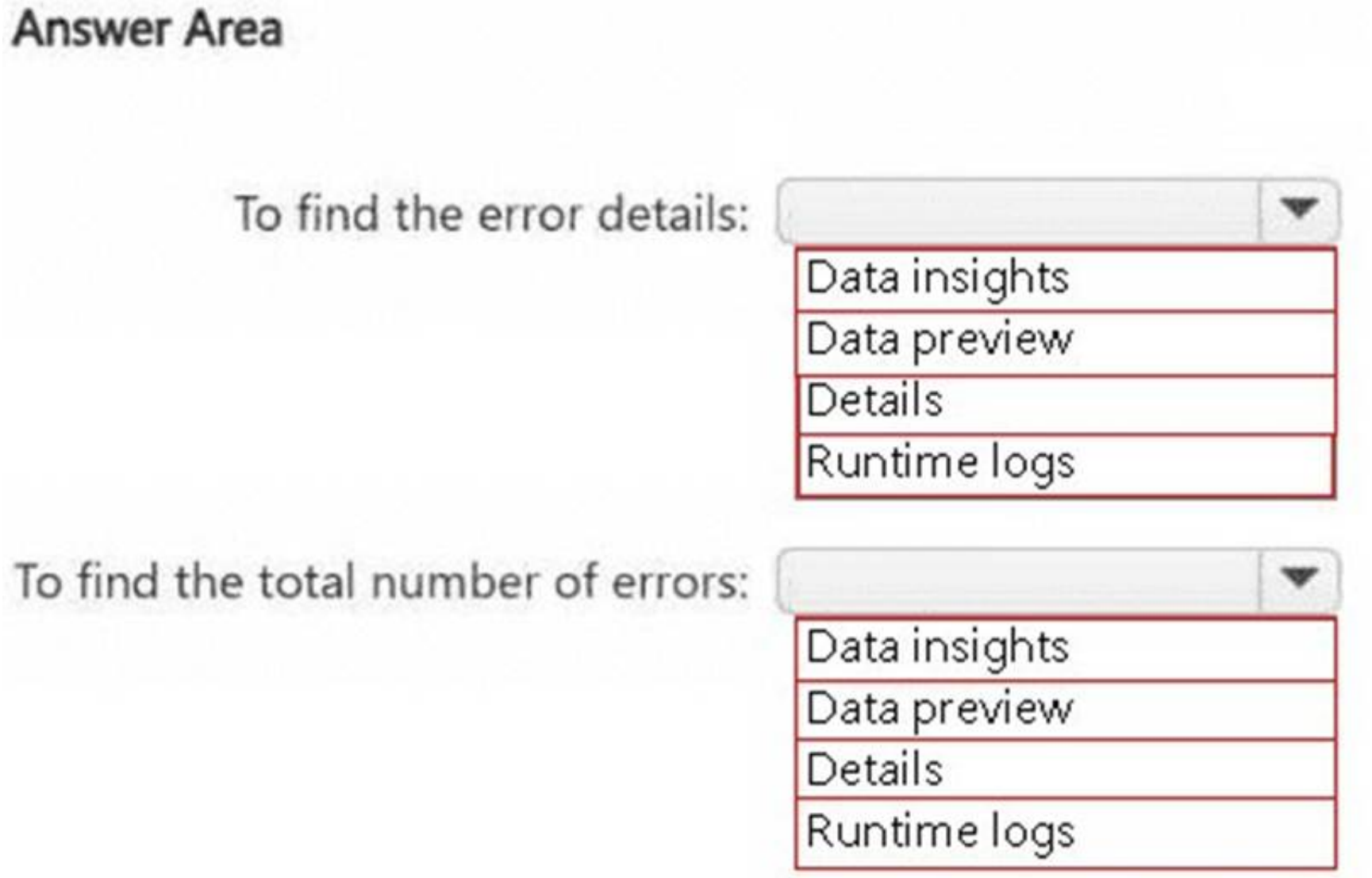


NEW QUESTION 38
- (Topic 3)
You have an Azure key vault named KeyVaultl that contains secrets.
You have a Fabric workspace named Workspace!. Workspace! contains a notebook named Notebookl that performs the following tasks:
• Loads stage data to the target tables in a lakehouse
• Triggers the refresh of a semantic model
You plan to add functionality to Notebookl that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVaultl to generate the authentication token. Solution: You use the following code segment:
Use notebookutils. credentials.getSecret and specify key vault URL and the name of a linked service.
Does this meet the goal?

A. Yes
B. No

Answer: B

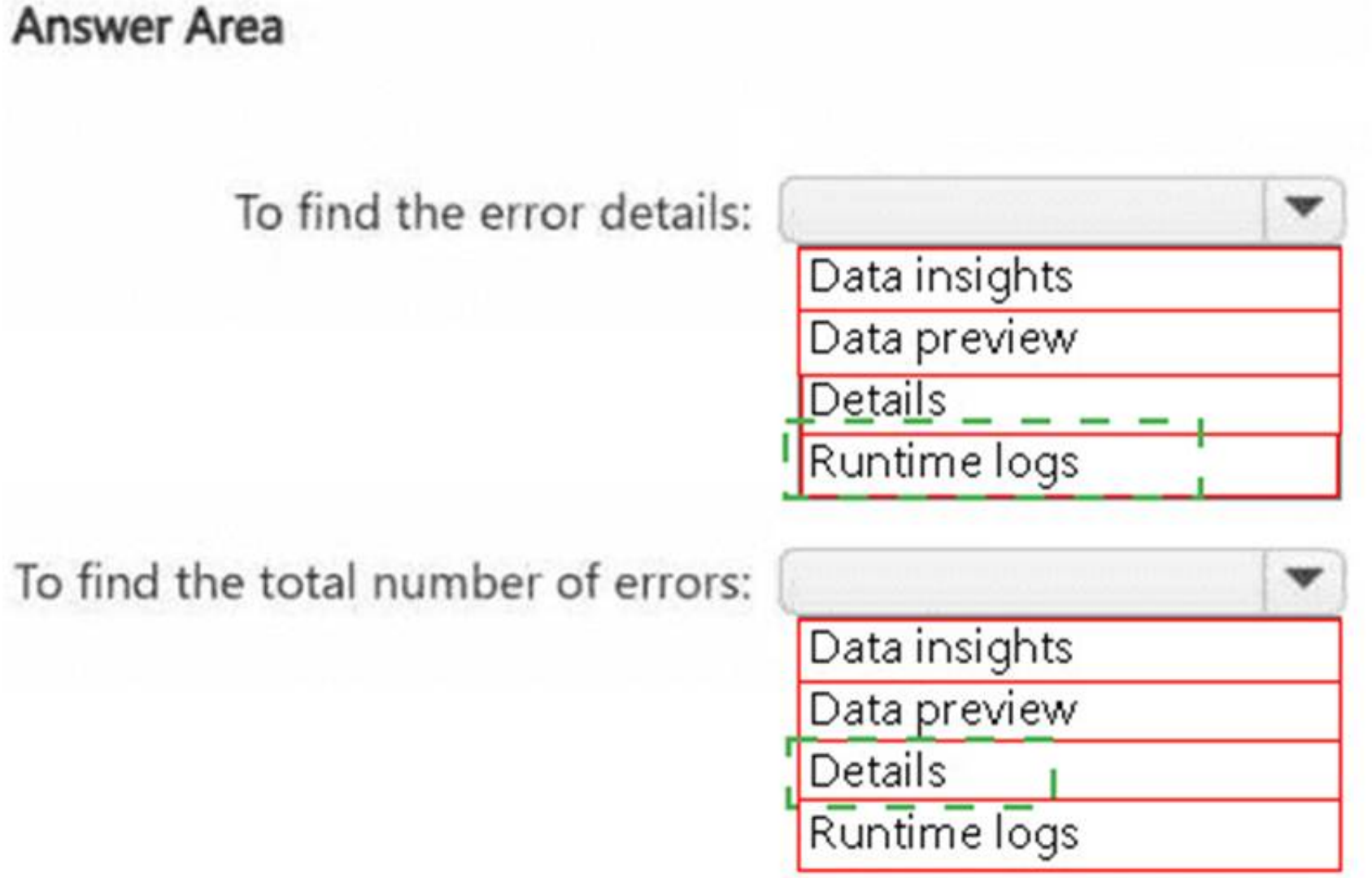
NEW QUESTION 39
HOTSPOT - (Topic 3)
You have a Fabric workspace that contains an eventstream named EventStream1. You discover that an EventStream1 transformation fails.
You need to find the following error information: The error details, including the occurrence time The total number of errors
What should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 41
HOTSPOT - (Topic 3)
You have a Fabric workspace named Workspace1_DEV that contains the following items: 10 reports
Four notebooks Three lakehouses Two data pipelines
Two Dataflow Gen1 dataflows Three Dataflow Gen2 dataflows
Five semantic models that each has a scheduled refresh policy
You create a deployment pipeline named Pipeline1 to move items from Workspace1_DEV to a new workspace named Workspace1_TEST.
You deploy all the items from Workspace1_DEV to Workspace1_TEST.
For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.
Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	○	○
The Dataflow Gen1 dataflows will be deployed to the target stage.	○	○
The scheduled refresh policies will be deployed to the target stage.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	○	○
The Dataflow Gen1 dataflows will be deployed to the target stage.	○	○
The scheduled refresh policies will be deployed to the target stage.	○	○

NEW QUESTION 46
- (Topic 3)
You have a Fabric workspace that contains a semantic model named Model1. You need to dynamically execute and monitor the refresh progress of Model1. What should you use?

A. dynamic management views in Microsoft SQL Server Management Studio
B. Monitoring hub
C. dynamic management views in Azure Data Studio
D. a semantic link in a notebook

Answer: D

Explanation:
Semantic models in Microsoft Fabric are part of Power BI datasets and require refreshes to stay updated with the latest data.
Dynamically executing and monitoring the refresh progress requires a tool or approach that integrates with Fabric's capabilities for semantic models.

NEW QUESTION 47
- (Topic 3)

You have a Google Cloud Storage (GCS) container named storage1 that contains the files shown in the following table.

Name	Size
ProductFile.parquet	8 MB
StoreFile.json	500 MB
TripsFile.csv	99 MB

You have a Fabric workspace named Workspace1 that has the cache for shortcuts enabled. Workspace1 contains a lakehouse named Lakehouse1. Lakehouse1 has the shortcuts shown in the following table.

Name	Source	Last accessed
Products	ProductFile	12 hours ago
Stores	StoreFile	4 hours ago
Trips	TripsFile	48 hours ago

You need to read data from all the shortcuts. Which shortcuts will retrieve data from the cache?

A. Stores only
B. Products only
C. Stores and Products only
D. Products, Stores, and Trips
E. Trips only
F. Products and Trips only

Answer: C

Explanation:
When reading data from shortcuts in Fabric (in this case, from a lakehouse like Lakehouse1), the cache for shortcuts helps by storing the data locally for quick access. The last accessed timestamp and the cache expiration rules determine whether data is fetched from the cache or from the source (Google Cloud Storage, in this case).
Products: The ProductFile.parquet was last accessed 12 hours ago. Since the cache has data available for up to 12 hours, it is likely that this data will be retrieved from the cache, as it hasn't been too long since it was last accessed.
Stores: The StoreFile.json was last accessed 4 hours ago, which is within the cache retention period. Therefore, this data will also be retrieved from the cache.
Trips: The TripsFile.csv was last accessed 48 hours ago. Given that it's outside the typical caching window (assuming the cache has a maximum retention period of around 24 hours), it would not be retrieved from the cache. Instead, it will likely require a fresh read from the source.

NEW QUESTION 49
- (Topic 3)
You have a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Table1.
You analyze Table1 and discover that Table1 contains 2,000 Parquet files of 1 MB each. You need to minimize how long it takes to query Table1.
What should you do?

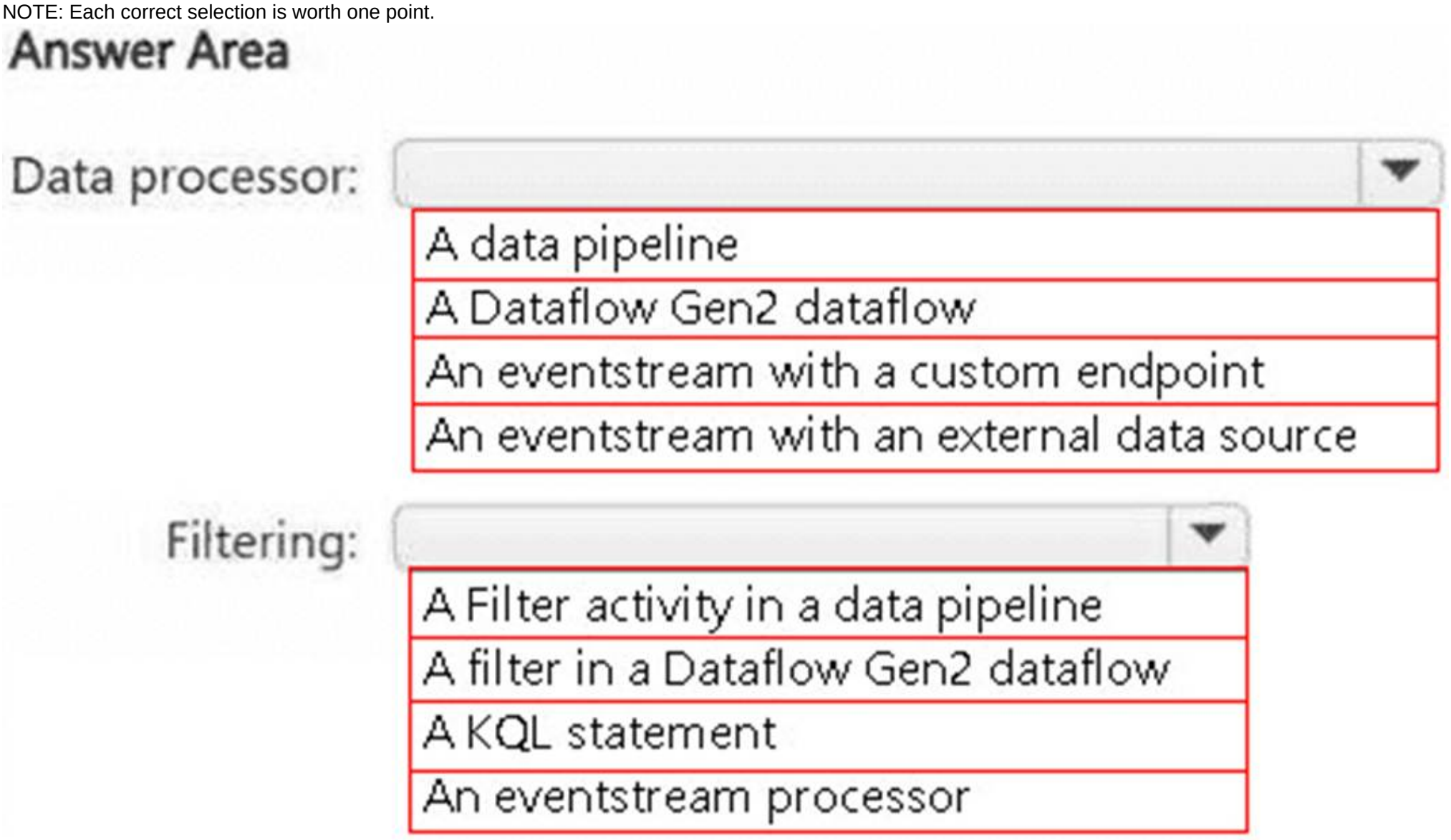
A. Disable V-Order and run the OPTIMIZE command.
B. Disable V-Order and run the VACUUM command.
C. Run the OPTIMIZE and VACUUM commands.

Answer: C

Explanation:
Problem Overview:
Table1 has 2,000 small Parquet files (1 MB each).
Query performance suffers when the table contains numerous small files because the query engine must process each file individually, leading to significant overhead.
Solution:
To improve performance, file compaction is necessary to reduce the number of small files and create larger, optimized files.
Commands and Their Roles: OPTIMIZE Command:
- Compacts small Parquet files into larger files to improve query performance.
- It supports optional features like V-Order, which organizes data for efficient scanning. VACUUM Command:
- Removes old, unreferenced data files and metadata from the Delta table.
- Running VACUUM after OPTIMIZE ensures unnecessary files are cleaned up, reducing storage overhead and improving performance.

NEW QUESTION 54

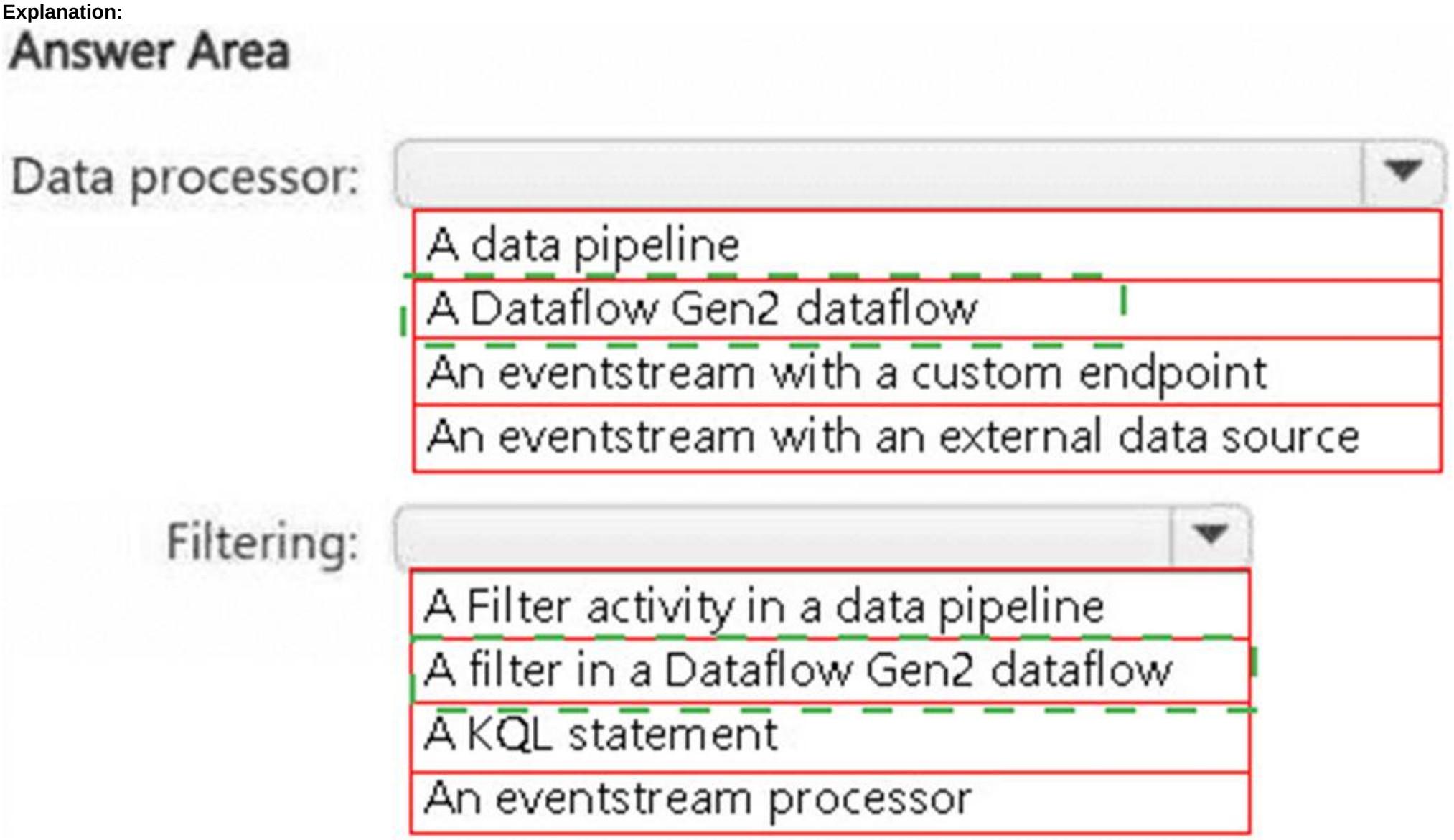
HOTSPOT - (Topic 3)
You have an Azure Event Hubs data source that contains weather data.
You ingest the data from the data source by using an eventstream named Eventstream1. Eventstream1 uses a lakehouse as the destination.
You need to batch ingest only rows from the data source where the City attribute has a value of Kansas. The filter must be added before the destination. The solution must minimize development effort.
What should you use for the data processor and filtering? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

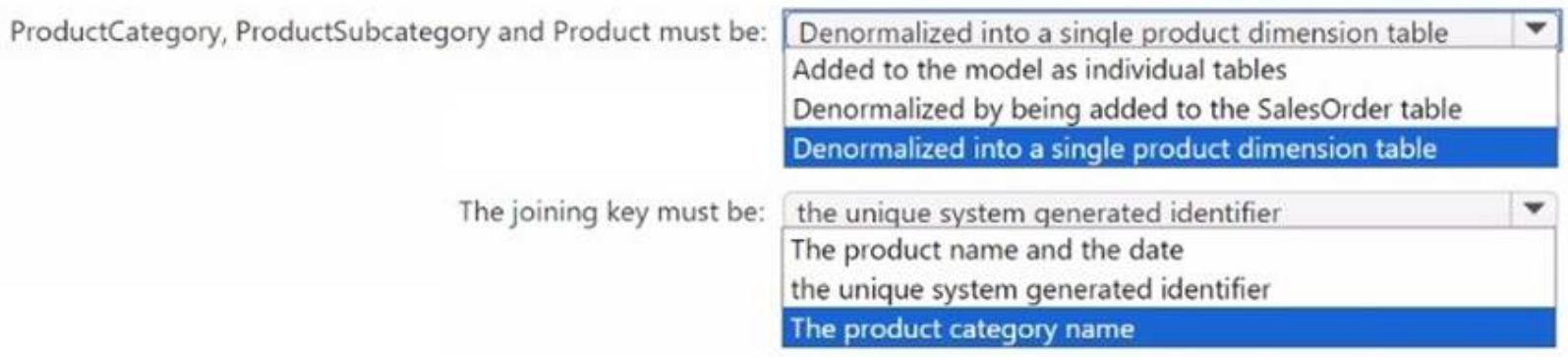
Answer: A

Explanation:



NEW QUESTION 57
HOTSPOT - (Topic 3)
You have a Fabric warehouse named DW1 that contains four staging tables named ProductCategory, ProductSubcategory, Product, and SalesOrder.
ProductCategory, ProductSubcategory, and Product are used often in analytical queries.
You need to implement a star schema for DW1. The solution must minimize development effort.
Which design approach should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

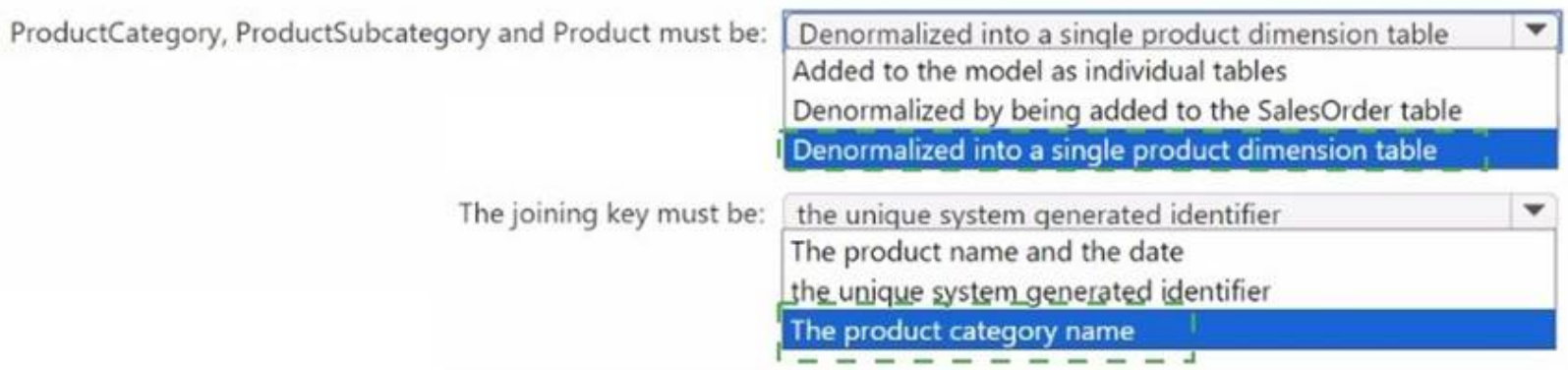


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 61
HOTSPOT - (Topic 3)
Your company has three newly created data engineering teams named Team1, Team2, and Team3 that plan to use Fabric. The teams have the following personas:
• Team1 consists of members who currently use Microsoft Power BI. The team wants to transform data by using by a low-code approach.
• Team2 consists of members that have a background in Python programming. The team wants to use PySpark code to transform data.
• Team3 consists of members who currently use Azure Data Factory. The team wants to move data between source and sink environments by using the least amount of effort.
You need to recommend tools for the teams based on their current personas.
What should you recommend for each team? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

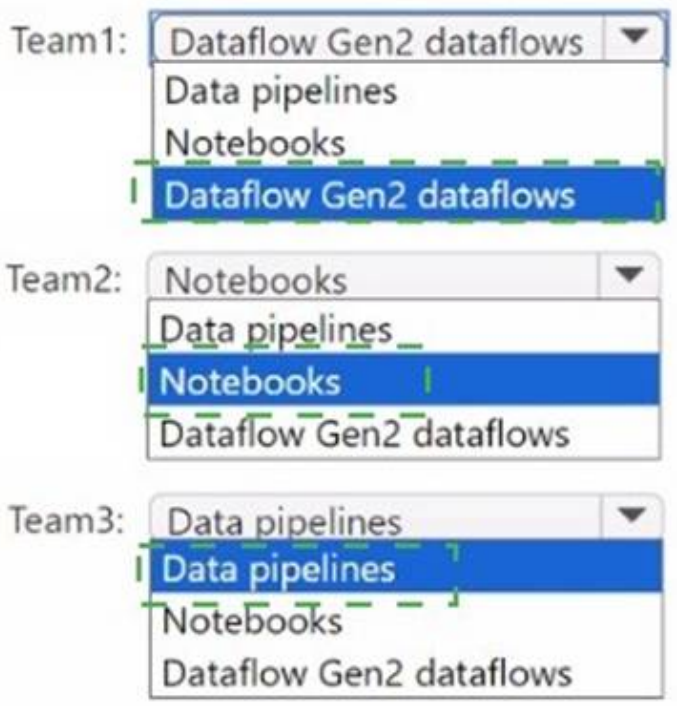
Answer Area

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



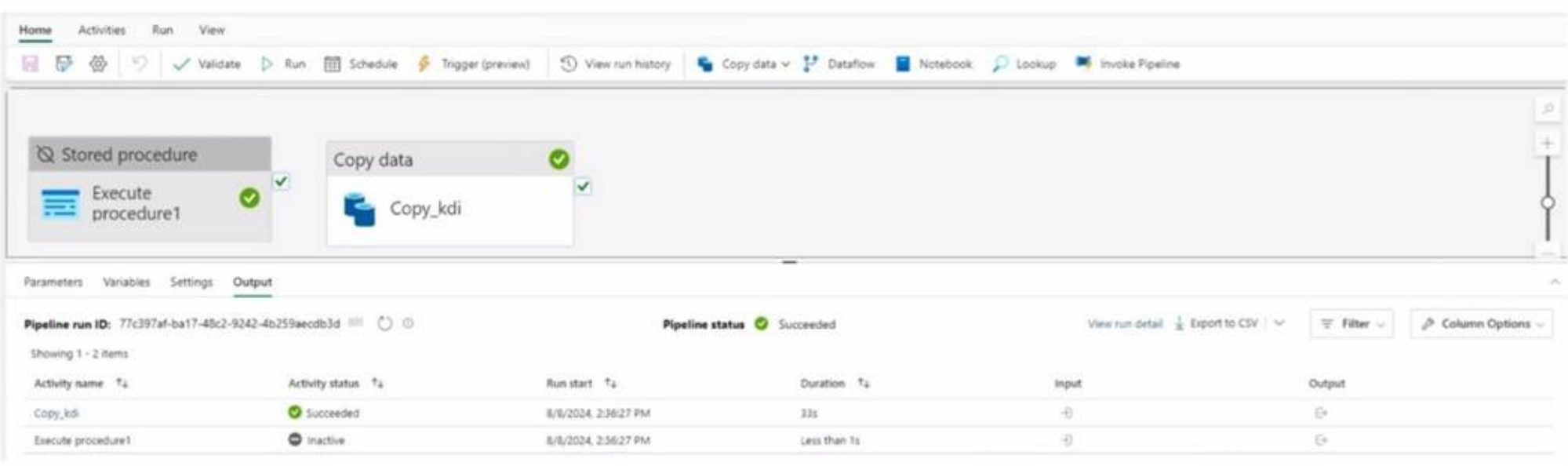
NEW QUESTION 66
- (Topic 3)
You have a Fabric deployment pipeline that uses three workspaces named Dev, Test, and Prod.
You need to deploy an eventhouse as part of the deployment process. What should you use to add the eventhouse to the deployment process?

A. GitHub Actions
B. a deployment pipeline
C. an Azure DevOps pipeline

Answer: B

Explanation:
A deployment pipeline in Fabric is designed to automate the process of deploying assets (such as reports, datasets, eventhouses, and other objects) between environments like Dev, Test, and Prod. Since you need to deploy an eventhouse as part of the deployment process, a deployment pipeline is the appropriate tool to move this asset through the different stages of your environment.

NEW QUESTION 70
- (Topic 3)
Exhibit.



You have a Fabric workspace that contains a write-intensive warehouse named DW1. DW1 stores staging tables that are used to load a dimensional model. The tables are often read once, dropped, and then recreated to process new data.
You need to minimize the load time of DW1. What should you do?

A. Disable V-Order.
B. Drop statistics.
C. Enable V-O-der.
D. Create statistics.

Answer: C

NEW QUESTION 71
HOTSPOT - (Topic 3)
You are processing streaming data from an external data provider. You have the following code segment.

```
datatable (Location:string, Company:string, UnitsSold:long)
[
   "New York", "Contoso", 300,
   "New York", "Litware", 1000,
   "New York", "Relecloud", 300,
   "New York", "Fabrikam", 200,
   "Seattle", "Contoso", 300,
   "Seattle", "Litware", 100,
   "Seattle", "Fabrikam", 100,
   "San Francisco", "Relecloud", 500,
   "San Francisco", "Litware", 500,
   "Washington DC", "Litware", 300,
   "Washington DC", "Contoso", 400
]
| sort by Location desc, UnitsSold desc
| extend Rank=row_rank_dense(UnitsSold, prev(Location) != Location)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.
Answer Area

Statements	Yes	No
Litware from New York will be displayed at the top of the result set.	○	○
Fabrikam in Seattle will have value = 2 in the Rank column.	○	○
Litware in San Francisco will have the same value in the Rank column as Litware in New York.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Litware from New York will be displayed at the top of the result set – Yes
The data is sorted first by Location in descending order and then by UnitsSold in descending order. Since "New York" is alphabetically the last Location, it will appear first in the result set. Within "New York", Litware has the highest UnitsSold (1000), so it will be displayed at the top.
Fabrikam in Seattle will have value = 2 in the Rank column – No
The row_rank_dense function assigns dense ranks based on UnitsSold within each location. In "Seattle":
Contoso has UnitsSold = 300 Rank 1 Litware has UnitsSold = 100 Rank 2
Fabrikam also has UnitsSold = 100, so it shares the same rank (2) as Litware.
Litware in San Francisco will have the same value in the Rank column as Litware in New York – No
The rank is calculated separately for each location. In "San Francisco":
Both Relecloud and Litware have UnitsSold = 500, so they share the same rank (1). In "New York", Litware has the highest UnitsSold = 1000 Rank 1.
Since ranks are calculated independently for each location, Litware in San Francisco does not share the same rank as Litware in New York.

NEW QUESTION 72

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-700 Exam with Our Prep Materials Via below:

https://www.certleader.com/DP-700-dumps.html