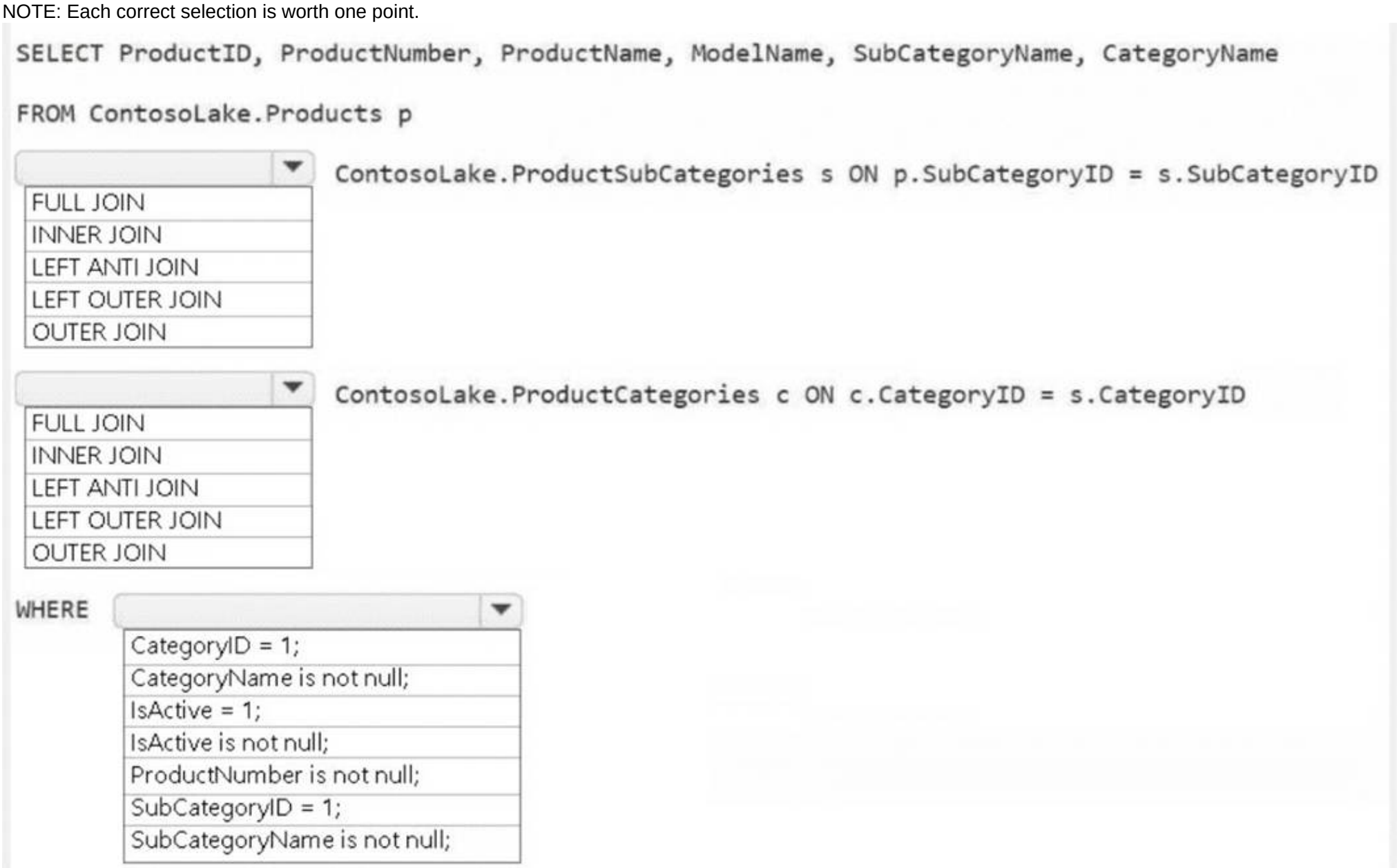
Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

https://www.2passeasy.com/dumps/DP-700/

NEW QUESTION 1
HOTSPOT - (Topic 1)
You need to create the product dimension.
How should you complete the Apache Spark SQL code? 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:
Join between Products and ProductSubCategories: Use an INNER JOIN.
The goal is to include only products that are assigned to a subcategory. An INNER JOIN ensures that only matching records (i.e., products with a valid subcategory) are included.
Join between ProductSubCategories and ProductCategories: Use an INNER JOIN.
Similar to the above logic, we want to include only subcategories assigned to a valid product category. An INNER JOIN ensures this condition is met.
WHERE Clause Condition: IsActive = 1
Only active products (where IsActive equals 1) should be included in the gold layer. This filters out inactive products.

NEW QUESTION 2
- (Topic 1)
You need to populate the MAR1 data in the bronze layer.
Which two types of activities should you include in the pipeline? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. ForEach
B. Copy data
C. WebHook
D. Stored procedure

Answer: AB

Explanation:
MAR1 has seven entities, each accessible via a different API endpoint. A ForEach activity is required to iterate over these endpoints to fetch data from each one. It enables dynamic execution of API calls for each entity.
The Copy data activity is the primary mechanism to extract data from REST APIs and load it into the bronze layer in Delta format. It supports native connectors for REST APIs and Delta, minimizing development effort.
You need to schedule the population of the medallion layers to meet the technical requirements.
What should you do?
* A. Schedule a data pipeline that calls other data pipelines.
* B. Schedule a notebook.
* C. Schedule an Apache Spark job.
* D. Schedule multiple data pipelines.

* Answer: A
The technical requirements specify that:
Medallion layers must be fully populated sequentially (bronze silver gold). Each layer must be populated before the next.
If any step fails, the process must notify the data engineers. Data imports should run simultaneously when possible.
Why Use a Data Pipeline That Calls Other Data Pipelines?
A data pipeline provides a modular and reusable approach to orchestrating the sequential population of medallion layers.
By calling other pipelines, each pipeline can focus on populating a specific layer (bronze, silver, or gold), simplifying development and maintenance.
A parent pipeline can handle:
- Sequential execution of child pipelines.
- Error handling to send email notifications upon failures.
- Parallel execution of tasks where possible (e.g., simultaneous imports into the bronze layer).

NEW QUESTION 3
- (Topic 1)
You need to ensure that the data analysts can access the gold layer lakehouse. What should you do?

A. Add the DataAnalyst group to the Viewer role for WorkspaceA.
B. Share the lakehouse with the DataAnalysts group and grant the Build reports on the default semantic model permission.
C. Share the lakehouse with the DataAnalysts group and grant the Read all SQL Endpoint data permission.
D. Share the lakehouse with the DataAnalysts group and grant the Read all Apache Spark permission.

Answer: C

Explanation:
Data Analysts' Access Requirements must only have read access to the Delta tables in the gold layer and not have access to the bronze and silver layers.
The gold layer data is typically queried via SQL Endpoints. Granting the Read all SQL Endpoint data permission allows data analysts to query the data using familiar SQL-based tools while restricting access to the underlying files.

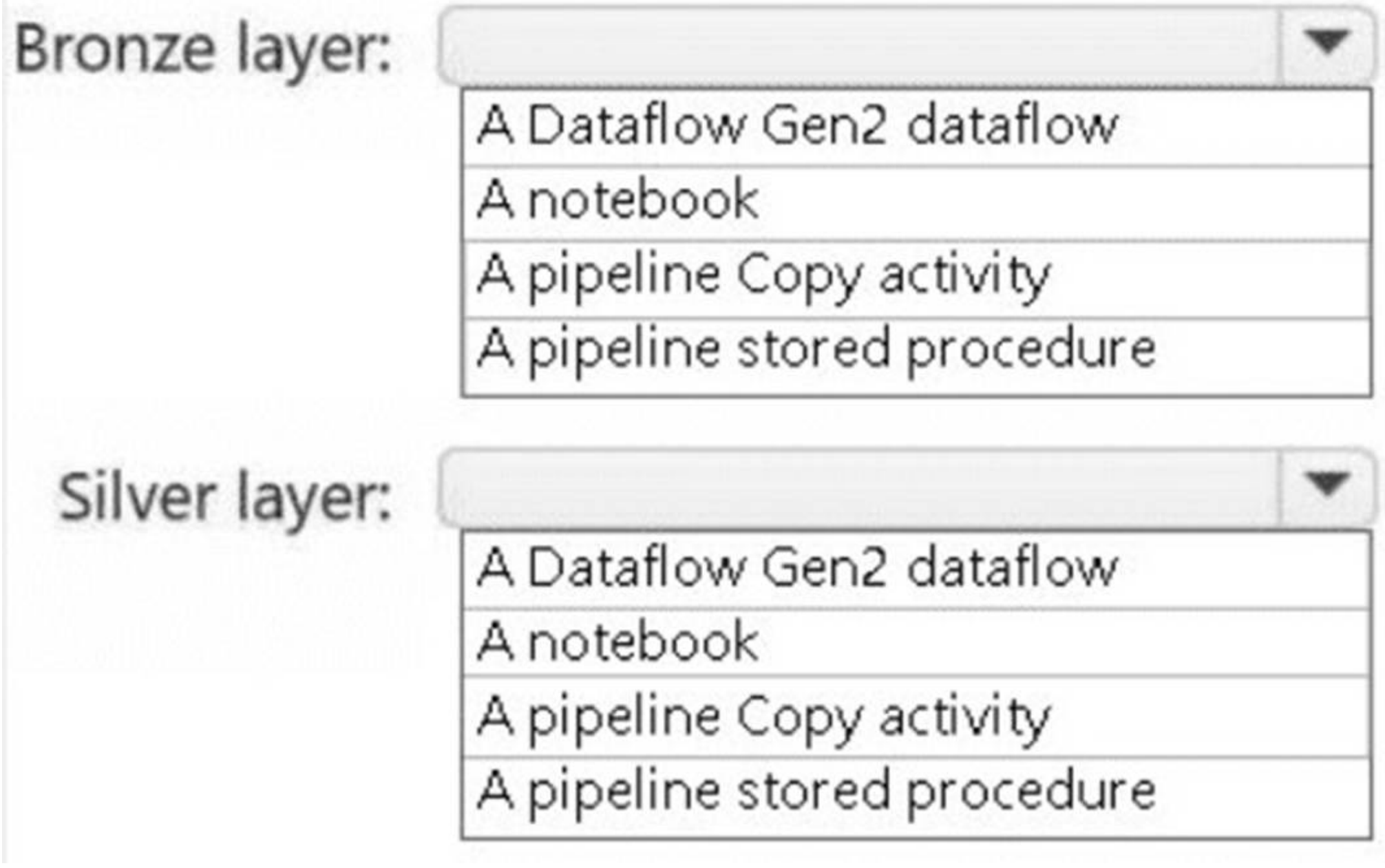
NEW QUESTION 4
- (Topic 1)
You need to ensure that usage of the data in the Amazon S3 bucket meets the technical requirements.
What should you do?

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 5
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.

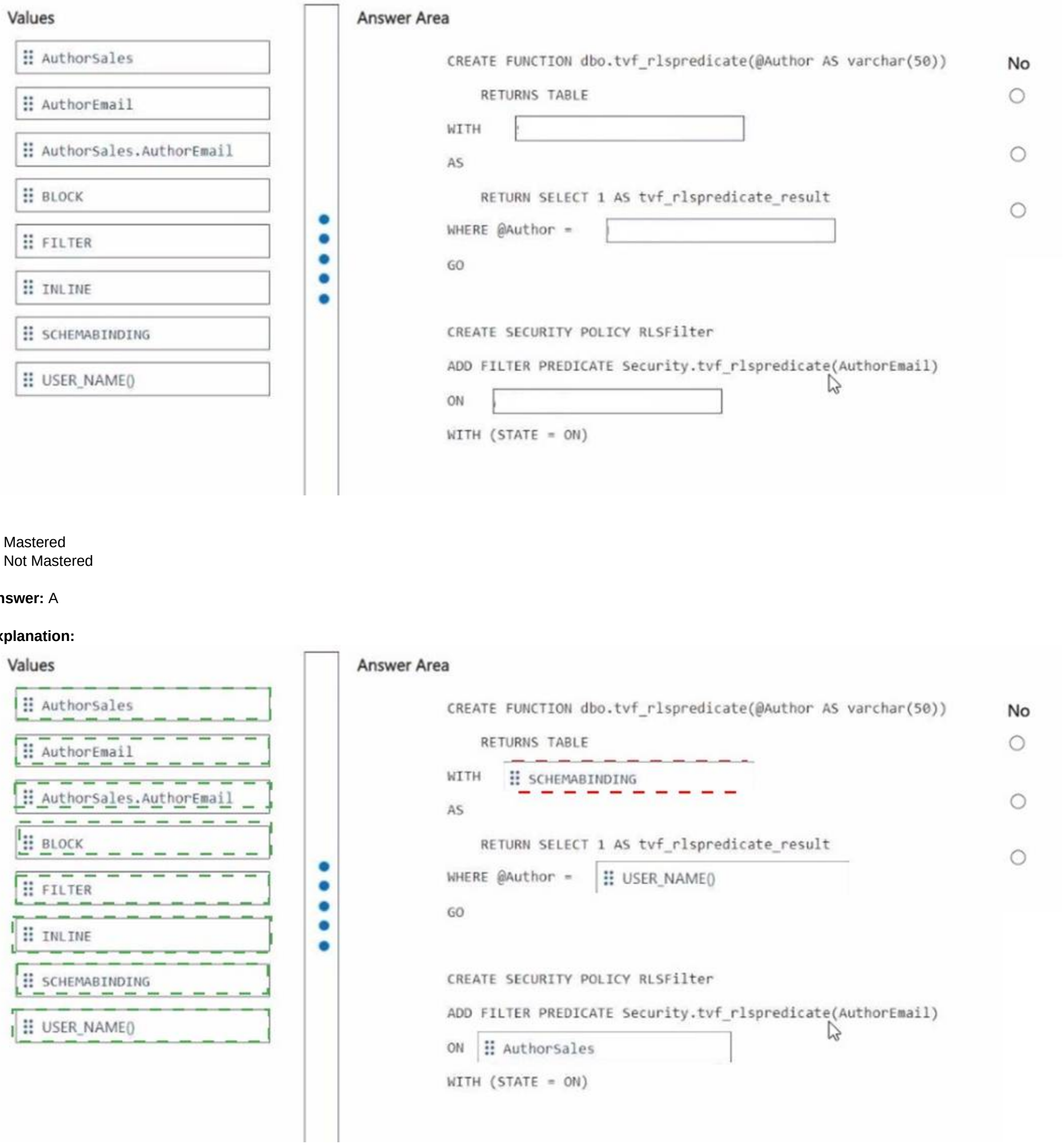


A. Mastered
B. Not Mastered

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 6
DRAG DROP - (Topic 2)
You need to ensure that the authors can see only their respective sales data.
How should you complete the statement? To answer, drag the appropriate values the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 7
- (Topic 2)
You need to resolve the sales data issue. The solution must minimize the amount of data transferred.
What should you do?

A. Spilt the dataflow into two dataflows.
B. Configure scheduled refresh for the dataflow.
C. Configure incremental refresh for the dataflo
D. Set Store rows from the past to 1 Month.
E. Configure incremental refresh for the dataflo
F. Set Refresh rows from the past to 1 Year.
G. Configure incremental refresh for the dataflo
H. Set Refresh rows from the past to 1 Month.

Answer: E

Explanation:

The sales data issue can be resolved by configuring incremental refresh for the dataflow. Incremental refresh allows for only the new or changed data to be processed, minimizing the amount of data transferred and improving performance.
The solution specifies that data older than one month never changes, so setting the refresh period to 1 Month is appropriate. This ensures that only the most recent month of data will be refreshed, reducing unnecessary data transfers.

NEW QUESTION 8
- (Topic 2)
You need to implement the solution for the book reviews.
Which should you do?

A. Create a Dataflow Gen2 dataflow.
B. Create a shortcut.
C. Enable external data sharing.
D. Create a data pipeline.

Answer: B

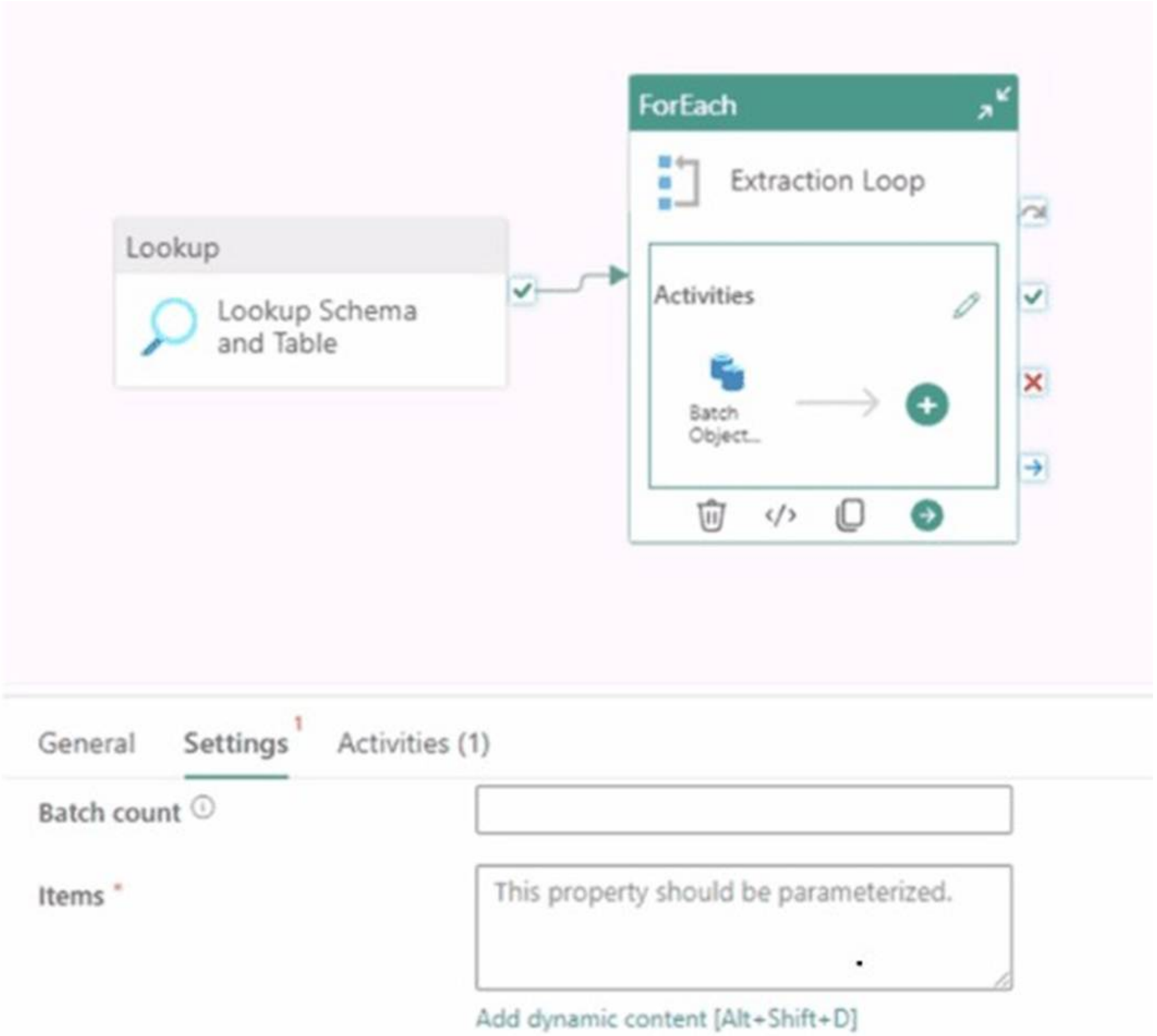
Explanation:
The requirement specifies that Litware plans to make the book reviews available in the lakehouse without making a copy of the data. In this case, creating a shortcut in Fabric is the most appropriate solution. A shortcut is a reference to the external data, and it allows Litware to access the book reviews stored in Amazon S3 without duplicating the data into the lakehouse.

NEW QUESTION 9
- (Topic 2)
What should you do to optimize the query experience for the business users?

A. Enable V-Order.
B. Create and update statistics.
C. Run the VACUUM command.
D. Introduce primary keys.

Answer: B

NEW QUESTION 10
HOTSPOT - (Topic 3)
You are building a data orchestration pattern by using a Fabric data pipeline named Dynamic Data Copy as shown in the exhibit. (Click the Exhibit tab.)



Dynamic Data Copy does NOT use parametrization.
You need to configure the ForEach activity to receive the list of tables to be copied. How should you complete the pipeline expression? 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 10
- (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1. Data is loaded daily into Warehouse1 by using data pipelines and stored procedures.
You discover that the daily data load takes longer than expected.
You need to monitor Warehouse1 to identify the names of users that are actively running queries.
Which view should you use?

A. sys.dm_exec_connections
B. sys.dm_exec_requests
C. queryinsights.long_running_queries
D. queryinsights.frequently_run_queries
E. sys.dm_exec_sessions

Answer: E

Explanation:
sys.dm_exec_sessions provides real-time information about all active sessions, including the user, session ID, and status of the session. You can filter on session status to see users actively running queries.

NEW QUESTION 11
- (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 16
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse1.
You plan to deploy Warehouse1 to a new workspace named Workspace2.
As part of the deployment process, you need to verify whether Warehouse1 contains invalid references. The solution must minimize development effort.
What should you use?

A. a database project

B. a deployment pipeline
C. a Python script
D. a T-SQL script

Answer: C

Explanation:
A deployment pipeline in Fabric allows you to deploy assets like warehouses, datasets, and reports between different workspaces (such as from Workspace1 to Workspace2). One of
the key features of a deployment pipeline is the ability to check for invalid references before deployment. This can help identify issues with assets, such as broken links or dependencies, ensuring the deployment is successful without introducing errors. This is the most efficient way to verify references and manage the deployment with minimal development effort.

NEW QUESTION 18
- (Topic 3)
You have a Fabric workspace. You have semi-structured data.
You need to read the data by using T-SQL, KQL, and Apache Spark. The data will only be written by using Spark.
What should you use to store the data?

A. a lakehouse
B. an eventhouse
C. a datamart
D. a warehouse

Answer: A

Explanation:
A lakehouse is the best option for storing semi-structured data when you need to read it using T-SQL, KQL, and Apache Spark. A lakehouse combines the flexibility of a data lake (which can handle semi-structured and unstructured data) with the performance features of a data warehouse. It allows data to be written using Apache Spark and can be queried using different technologies such as T-SQL (for SQL-based querying), KQL (Kusto Query Language for querying), and Apache Spark (for distributed processing). This solution is ideal when dealing with semi-structured data and requiring a versatile querying approach.

NEW QUESTION 20
- (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

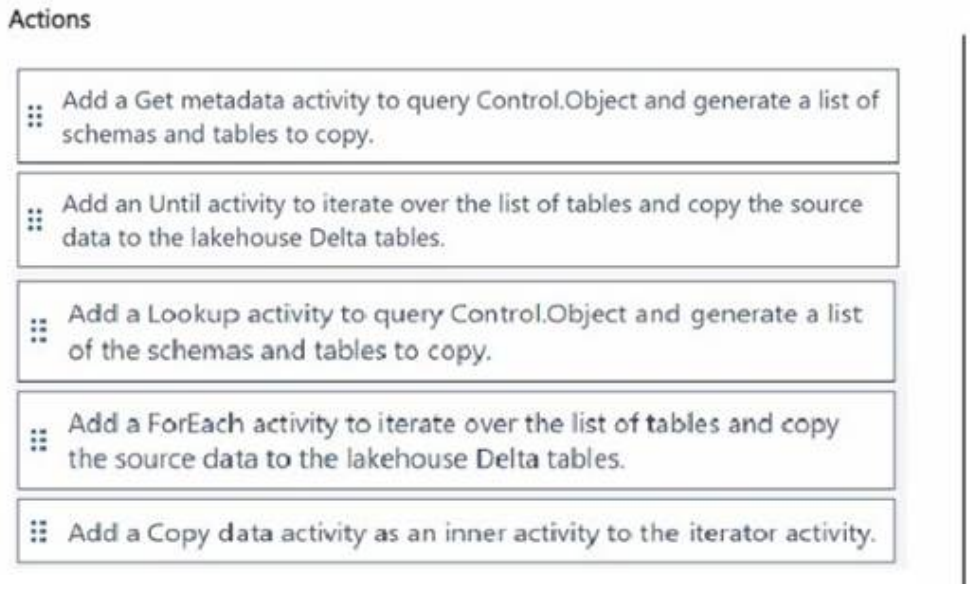
Explanation:
An outer join will include unmatched rows from both tables, increasing the dataset size and processing time. It does not improve query performance.

NEW QUESTION 23
- (Topic 3)
You need to develop an orchestration solution in fabric that will load each item one after the other. The solution must be scheduled to run every 15 minutes. Which type of item should you use?

A. warehouse
B. data pipeline
C. Dataflow Gen2 dataflow
D. notebook

Answer: B

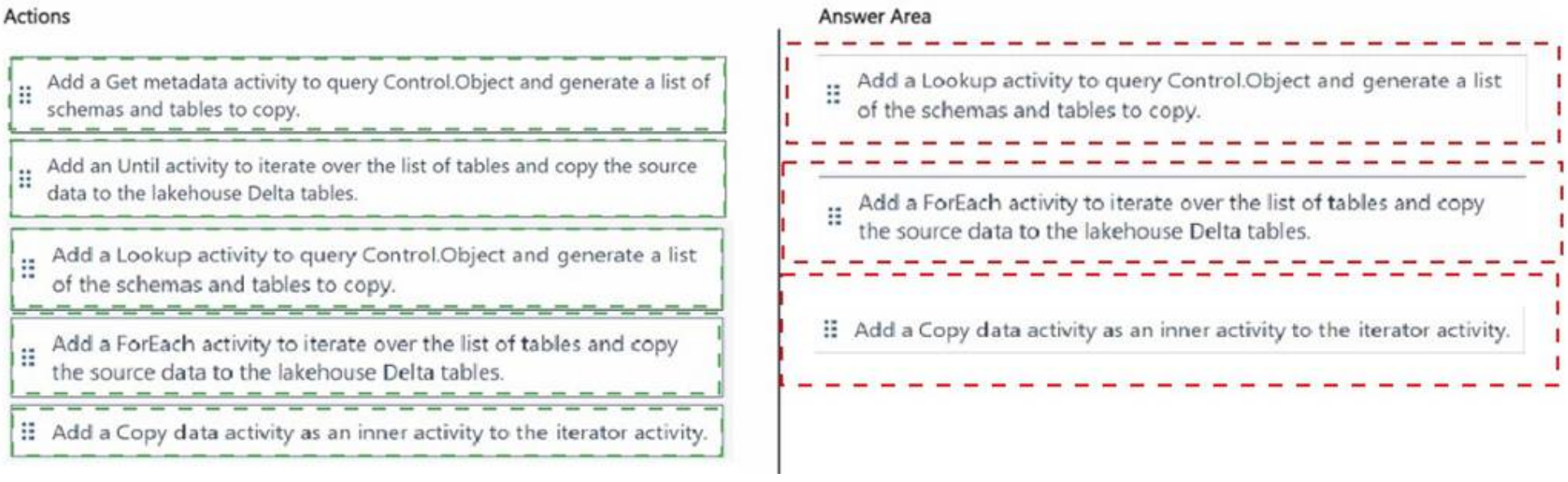
NEW QUESTION 25
DRAG DROP - (Topic 3)
You are building a data loading pattern by using a Fabric data pipeline. The source is an Azure SQL database that contains 25 tables. The destination is a lakehouse.
In a warehouse, you create a control table named Control.Object as shown in the exhibit. (Click the Exhibit tab.)
You need to build a data pipeline that will support the dynamic ingestion of the tables listed in the control table by using a single execution.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 29
HOTSPOT - (Topic 3)
You have a Fabric workspace.
You are debugging a statement and discover the following issues: Sometimes, the statement fails to return all the expected rows.
The PurchaseDate output column is NOT in the expected format of mmm dd, yy.
You need to resolve the issues. The solution must ensure that the data types of the results are retained. The results can contain blank cells.
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

```
SELECT
  item_id as ItemId
  [ ▼ ]  as ItemName
      ,convert(varchar(20), item_name)
      ,convert(varchar(max), item_name)
      ,try_cast(item_name as varchar(20))
  ,item_description as ItemDescription
  [ ▼ ]  as PurchaseDate
      ,convert(varchar, purchase_date, 7)
      ,convert(varchar, purchase_date, 109)
      ,convert(varchar, purchase_date, 112)
FROM
  Table1
WHERE
  item_type = @itemtype_parameter
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

```
SELECT
  item_id as ItemId
  [ ▼ ]                                          as ItemName
  ,item_description as ItemDescription
  [ ▼ ]                                          as PurchaseDate
FROM
  Table1
WHERE
  item_type = @itemtype_parameter
```

- ,convert(varchar(20), item_name)
- ,convert(varchar(max), item_name)
- ,try_cast(item_name as varchar(20))

- ,convert(varchar, purchase_date, 7)
- ,convert(varchar, purchase_date, 109)
- ,convert(varchar, purchase_date, 112)

NEW QUESTION 33
- (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 Fabric eventstream that loads data into a table named Bike_Location in a KQL database. The table contains the following columns:
BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp
You need to apply transformation and filter logic to prepare the data for consumption. The
solution must return data for a neighbourhood named Sands End when No_Bikes is at least 15. The results must be ordered by No_Bikes in ascending order.
Solution: You use the following code segment:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| sort by No_Bikes
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?

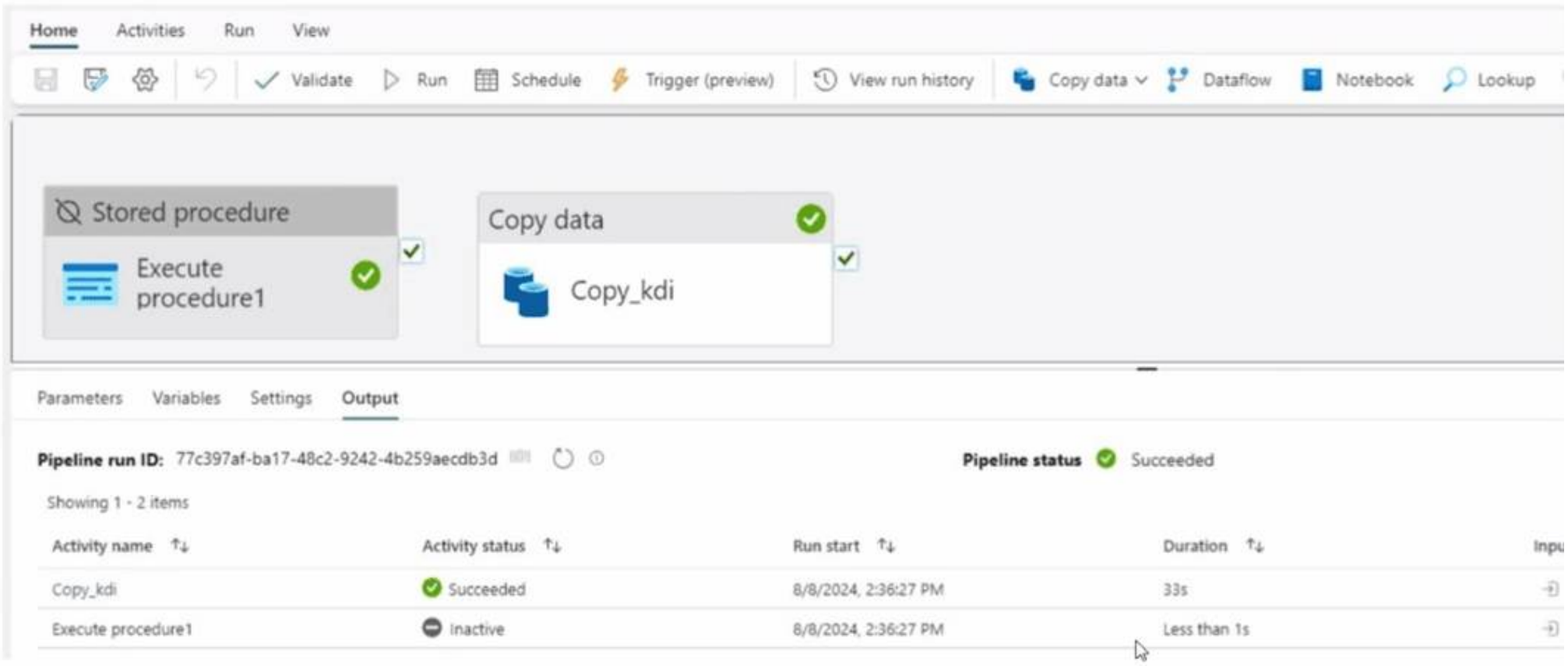
A. Yes
B. no

Answer: B

Explanation:
This code does not meet the goal because it uses sort by without specifying the order, which defaults to ascending, but explicitly mentioning asc improves clarity.
Correct code should look like:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| sort by No_Bikes asc
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

NEW QUESTION 35
- (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 37
- (Topic 3)
You have a Fabric workspace that contains an eventstream named EventStreaml. EventStreaml outputs events to a table named Tablel in a lakehouse. The streaming data is souiced from motorway sensors and represents the speed of cars.
You need to add a transformation to EventStream1 to average the car speeds. The speeds must be grouped by non-overlapping and contiguous time intervals of one minute. Each event must belong to exactly one window.
Which windowing function should you use?

A. sliding
B. hopping
C. tumbling
D. session

Answer: C

NEW QUESTION 42
- (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 Fabric eventstream that loads data into a table named Bike_Location in a KQL database. The table contains the following columns:
BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp
You need to apply transformation and filter logic to prepare the data for consumption. The solution must return data for a neighbourhood named Sands End when No_Bikes is at least 15. The results must be ordered by No_Bikes in ascending order.
Solution: You use the following code segment:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| sort by No_Bikes asc
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?

A. Yes
B. no

Answer: A

Explanation:
Filter Condition: It correctly filters rows where Neighbourhood is "Sands End" and No_Bikes is greater than or equal to 15.
Sorting: The sorting is explicitly done by No_Bikes in ascending order using sort by
No_Bikes asc.
Projection: It projects the required columns (BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp), which minimizes the data returned for consumption.

NEW QUESTION 43
- (Topic 3)
You have a Fabric workspace that contains a takehouse and a semantic model named Model1.
You use a notebook named Notebook1 to ingest and transform data from an external data source.
You need to execute Notebook1 as part of a data pipeline named Pipeline1. The process must meet the following requirements:
• Run daily at 07:00 AM UTC.
• Attempt to retry Notebook1 twice if the notebook fails.
• After Notebook1 executes successfully, refresh Model1.
Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Set the Retry setting of the Notebook activity to 2.
B. Place the Semantic model refresh activity after the Notebook activity and link the activities by using an On completion condition.
C. Place the Semantic model refresh activity after the Notebook activity and link the activities by using the On success condition.
D. From the Schedule settings of Notebook1, set the time zone to UTC.
E. From the Schedule settings of Pipeline1, set the time zone to UTC.
F. Set the Retry setting of the Semantic model refresh activity to 2.

Answer: ACE

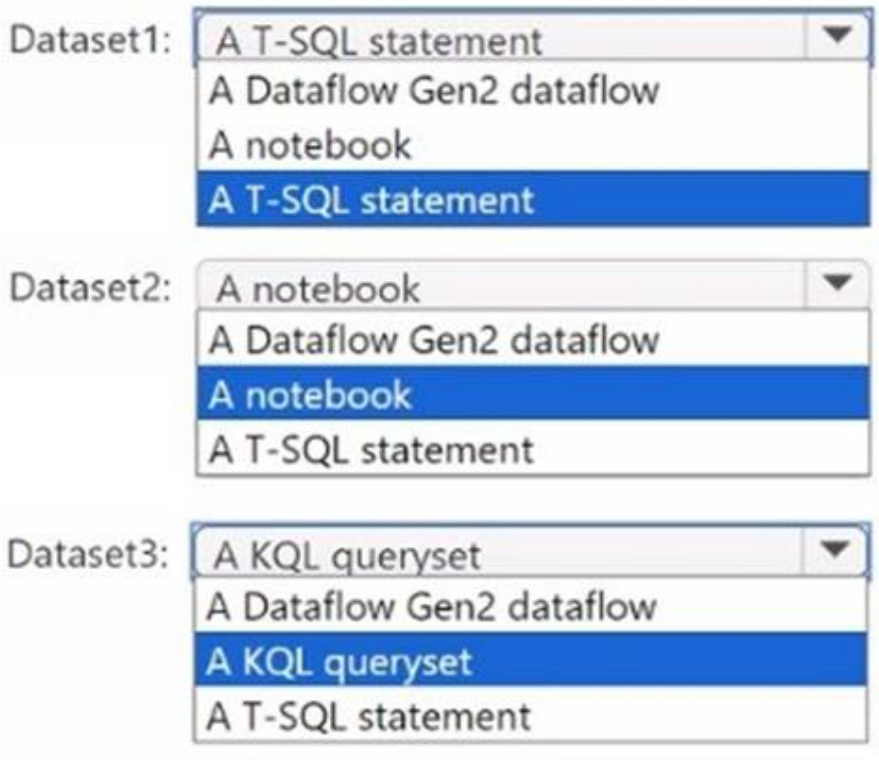
NEW QUESTION 44
- (Topic 3)
You have a Fabric workspace that contains a semantic model named Modell. You need to monitor the refresh history of Model 1 and visualize the refresh history in a chart. What should you use?

A. the refresh history from the settings of Model1.
B. a notebook
C. a Dataflow Gen2 dataflow
D. a data pipeline

Answer: A

NEW QUESTION 49
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

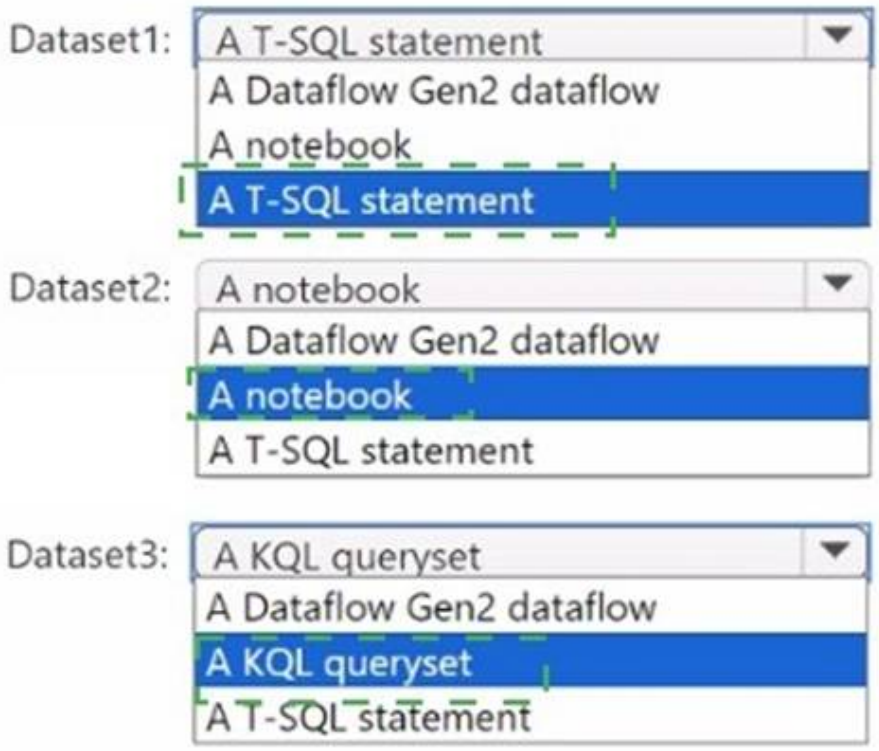


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 51
- (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

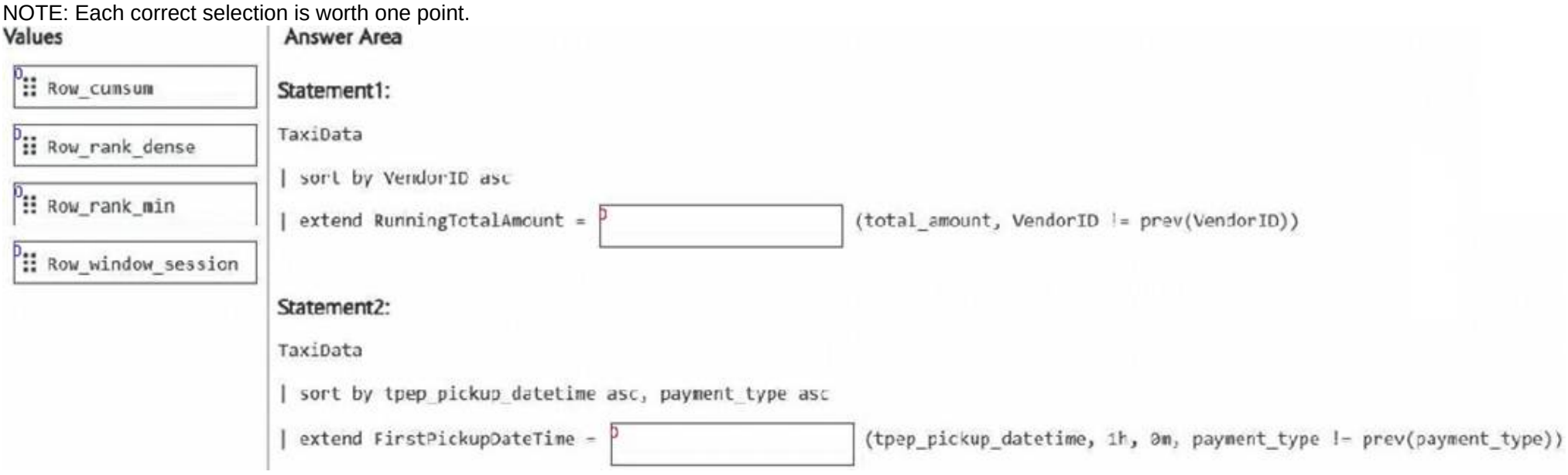
Explanation:
When integrating Azure DevOps with Fabric (Workspace1), using a service principal is the recommended authentication method. A service principal provides a way for applications (such as an Azure DevOps pipeline) to authenticate and interact with resources securely. It allows Azure DevOps 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 52
DRAG DROP - (Topic 3)

You have a Fabric eventhouse that contains a KQL database. The database contains a table named TaxiData. The following is a sample of the data in TaxiData.

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	PULocationID	DOLocationID	payment_type	total_amount
2	2022-06-06T11:08:32Z	2022-06-06T11:22:17Z	1	0.17	231	50	2	7.12
2	2022-06-06T11:12:05Z	2022-06-06T11:20:43Z	1	1.02	161	163	1	10.56
1	2022-06-06T11:15:00Z	2022-06-06T11:25:32Z	1	1.07	142	230	2	17.12
2	2022-06-06T11:29:54Z	2022-06-06T11:49:34Z	2	2.07	162	236	2	12.01
1	2022-06-06T11:50:50Z	2022-06-06T12:07:24Z	2	2.65	140	142	1	7.89

You need to build two KQL queries. The solution must meet the following requirements: One of the queries must partition RunningTotalAmount by VendorID. The other query must create a column named FirstPickupDateTime that shows the first value of each hour from tpep_pickup_datetime partitioned by payment_type.
How should you complete each query? To answer, drag the appropriate values the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Partition the RunningTotalAmount by VendorID. - Row_cumsum
The Row_cumsum function computes the cumulative sum of a column while optionally restarting the accumulation based on a condition. In this case, it calculates the cumulative sum of total_amount for each VendorID, restarting when the VendorID changes (VendorID != prev(VendorID)).

```
TaxiData
| sort by VendorID asc
| extend RunningTotalAmount = Row_cumsum(total_amount, VendorID != prev(VendorID))
```

Create a column FirstPickupDateTime that shows the first value of each hour from tpep_pickup_datetime, partitioned by payment_type - Row_window_session

```
TaxiData
| sort by tpep_pickup_datetime asc, payment_type asc
| extend FirstPickupDateTime = Row_window_session(tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))
```

NEW QUESTION 56
- (Topic 3)
You are developing a data pipeline named Pipeline1.
You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse. Which option from the Settings tab of the Copy data activity must you configure?

A. Enable logging
B. Fault tolerance
C. Enable staging
D. Degree of copy parallelism

Answer: C

NEW QUESTION 57
- (Topic 3)
You have a Fabric F32 capacity that contains a workspace. The workspace contains a warehouse named DW1 that is modelled by using MD5 hash surrogate keys.

DW1 contains a single fact table that has grown from 200 million rows to 500 million rows during the past year.
You have Microsoft Power BI reports that are based on Direct Lake. The reports show year-over-year values.
Users report that the performance of some of the reports has degraded over time and some visuals show errors.
You need to resolve the performance issues. The solution must meet the following requirements:
Provide the best query performance. Minimize operational costs.
Which should you do?

A. Change the MD5 hash to SHA256.
B. Increase the capacity.C Enable V-Order
C. Modify the surrogate keys to use a different data type.
D. Create views.

Answer: D

Explanation:
In this case, the key issue causing performance degradation likely stems from the use of MD5 hash surrogate keys. MD5 hashes are 128-bit values, which can be inefficient for large datasets like the 500 million rows in your fact table. Using a more efficient data type for surrogate keys (such as integer or bigint) would reduce the storage and processing overhead, leading to better query performance. This approach will improve performance while minimizing operational costs because it reduces the complexity of querying and indexing, as smaller data types are generally faster and more efficient to process.

NEW QUESTION 58
- (Topic 3)
You have a Fabric warehouse named DW1 that contains a Type 2 slowly changing dimension (SCD) dimension table named DimCustomer. DimCustomer contains 100 columns and 20 million rows. The columns are of various data types, including int, varchar, date, and varbinary.
You need to identify incoming changes to the table and update the records when there is a change. The solution must minimize resource consumption.
What should you use to identify changes to attributes?

A. a direct attributes comparison for the attributes in the source table.
B. a hash function to compare the attributes in the DimCustomer table.
C. a direct attributes comparison across the attributes in the DimCustomer table.
D. a hash function to compare the attributes in the source table.

Answer: D

NEW QUESTION 60
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

```
ALTER TABLE DimCustomer
ALTER COLUMN BirthDate
ADD MASKED WITH (FUNCTION = 'default()' )
```
'default()'
'partial(1900-01-01)'
'random(1900-01-01, 1900-01-01)'

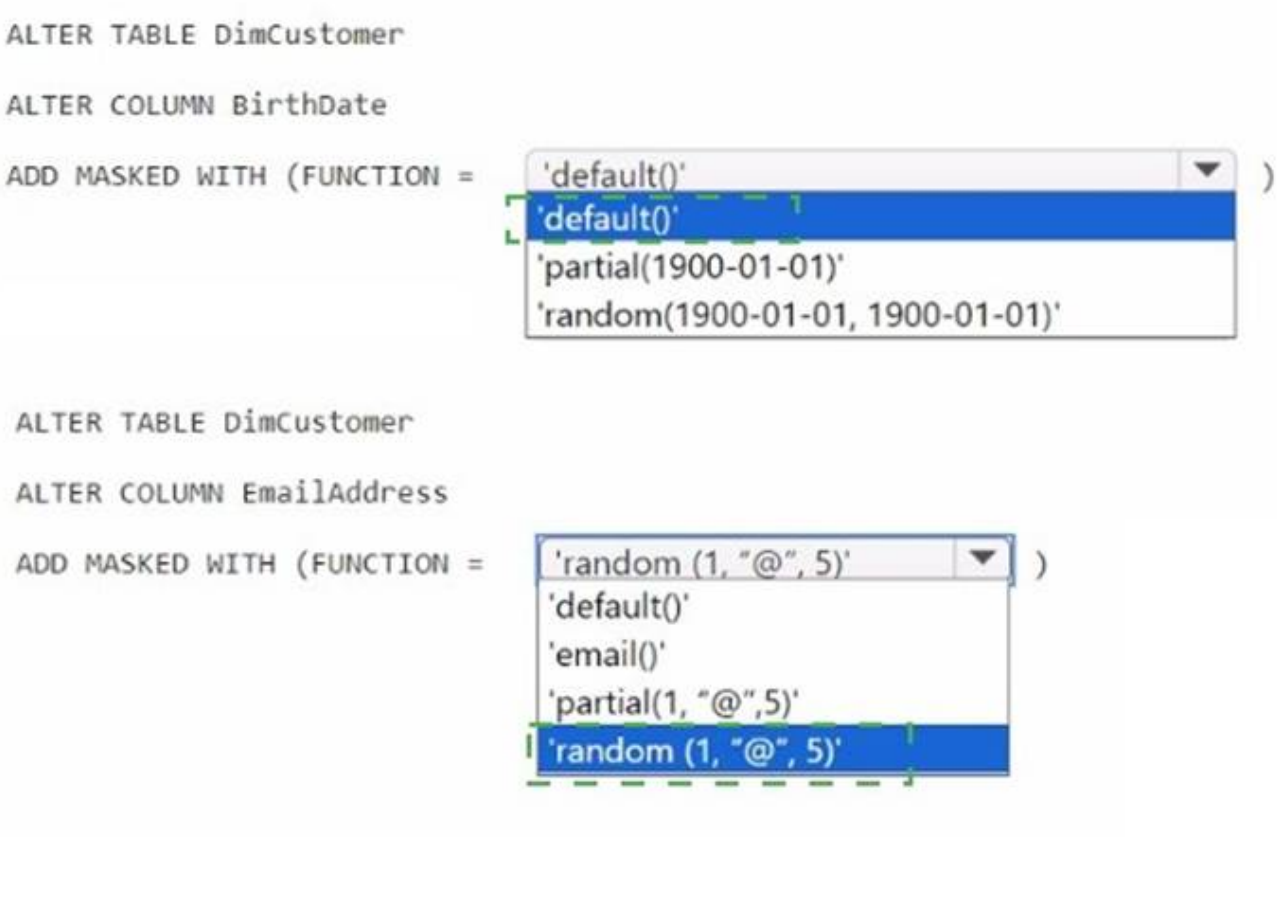
```
ALTER TABLE DimCustomer
ALTER COLUMN EmailAddress
ADD MASKED WITH (FUNCTION = 'random (1, "@", 5)' )
```
'default()'
'email()'
'partial(1, "@",5)'
'random (1, "@", 5)'

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 61
- (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 Fabric eventstream that loads data into a table named Bike_Location in a KQL database. The table contains the following columns:
BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp
You need to apply transformation and filter logic to prepare the data for consumption. The solution must return data for a neighbourhood named Sands End when No_Bikes is at least 15. The results must be ordered by No_Bikes in ascending order.
Solution: You use the following code segment:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| order by No_Bikes
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?
A. Yes
B. no

Answer: B

Explanation:
This code does not meet the goal because it uses order by, which is not valid in KQL. The correct term in KQL is sort by.
Correct code should look like:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| sort by No_Bikes asc
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

NEW QUESTION 62
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:
Orders
Customer Employee
The Employee table contains Personally Identifiable Information (PII).

A data engineer is building a workflow that requires writing data to the Customer table, however, the user does NOT have the elevated permissions required to view the contents of the Employee table.
You need to ensure that the data engineer can write data to the Customer table without reading data from the Employee table.
Which three actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Share Lakehouse1 with the data engineer.
B. Assign the data engineer the Contributor role for Workspace2.
C. Assign the data engineer the Viewer role for Workspace2.
D. Assign the data engineer the Contributor role for Workspace1.
E. Migrate the Employee table from Lakehouse1 to Lakehouse2.
F. Create a new workspace named Workspace2 that contains a new lakehouse named Lakehouse2.
G. Assign the data engineer the Viewer role for Workspace1.

Answer: ADE

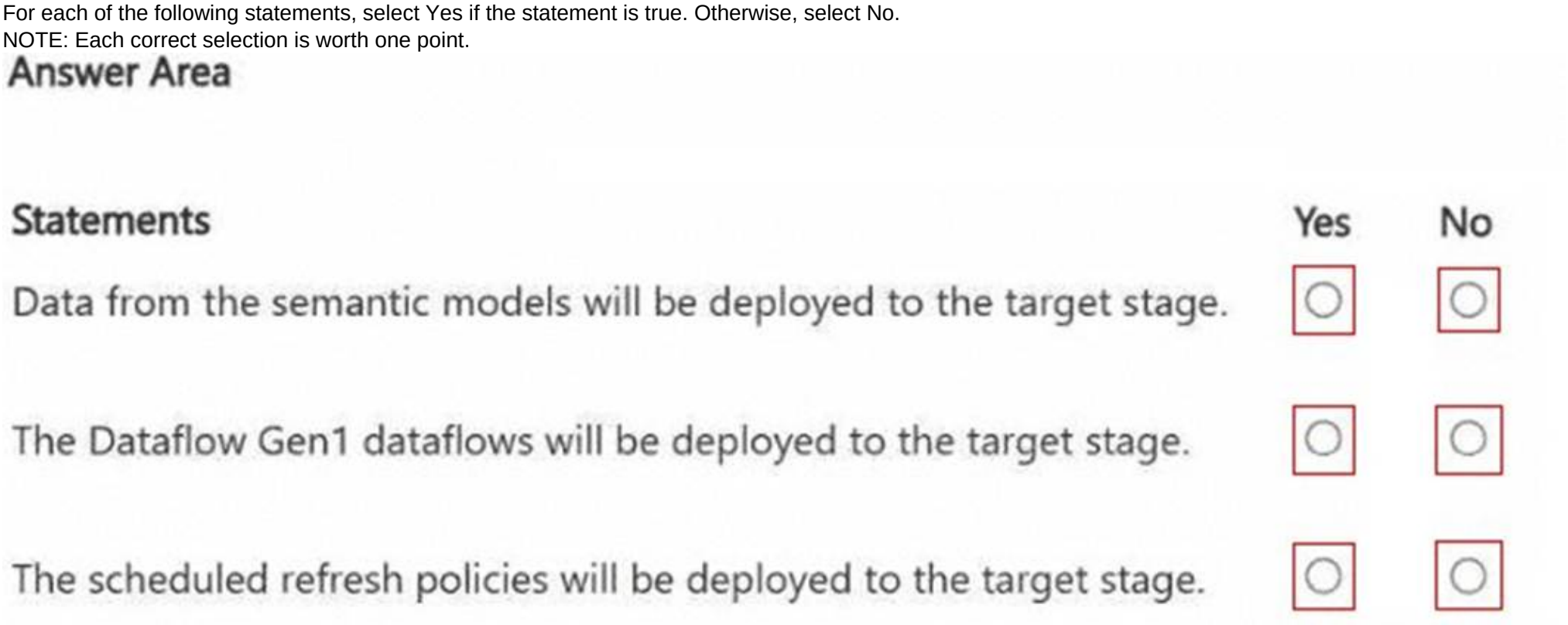
Explanation:
To meet the requirements of ensuring that the data engineer can write data to the Customer table without reading data from the Employee table (which contains Personally Identifiable Information, or PII), you can implement the following steps:
? Share Lakehouse1 with the data engineer.
By sharing Lakehouse1 with the data engineer, you provide the necessary access to the data within the lakehouse. However, this access should be controlled through roles and permissions, which will allow writing to the Customer table but prevent reading from the Employee table.
? Assign the data engineer the Contributor role for Workspace1.
Assigning the Contributor role for Workspace1 grants the data engineer the ability to perform actions such as writing to tables (e.g., the Customer table) within the workspace. This role typically allows users to modify and manage data without necessarily granting them access to view all data (e.g., PII data in the Employee table).
? Migrate the Employee table from Lakehouse1 to Lakehouse2.
To prevent the data engineer from accessing the Employee table (which contains PII), you can migrate the Employee table to a separate lakehouse (Lakehouse2) or workspace
(Workspace2). This separation of sensitive data ensures that the data engineer's access is restricted to the Customer table in Lakehouse1, while the Employee table can be managed separately and protected under different access controls.

NEW QUESTION 67
- (Topic 3)
You have a Fabric workspace named Workspace1. Your company acquires GitHub licenses.
You need to configure source control for Workpace1 to use GitHub. The solution must follow the principle of least privilege. Which permissions do you require to ensure that you can commit code to GitHub?

A. Actions (Read and write) and Contents (Read and write)
B. Actions (Read and write) only
C. Contents (Read and write) only
D. Contents (Read) and Commit statuses (Read and write)

Answer: C

NEW QUESTION 71
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



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 74
- (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 78
- (Topic 3)
You are building a Fabric notebook named MasterNotebookl in a workspace. MasterNotebookl contains the following code.

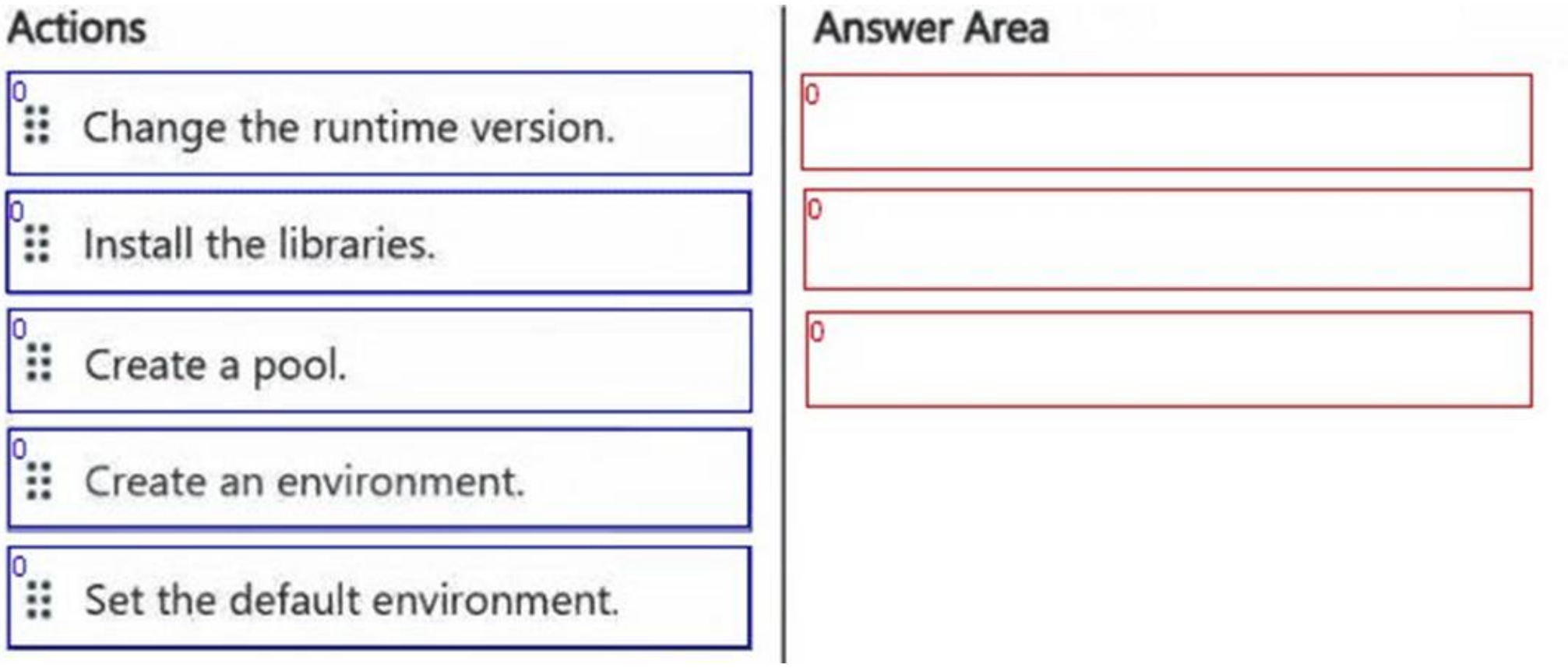
```
DAG = {
    "activities": [
        {   "name": "execute_notebook_1",
            "path": "notebook_01",
            "timeoutPerCellInSeconds": 600,
            "args": {
                        "input_value": "999"
                },
            "retry": 1,
            "retryIntervalInSeconds": 30
        },
        {   "name": "execute_notebook_2",
            "path": "notebook_02",
            "timeoutPerCellInSeconds": 400,
            "args": {
                        "input_value": "888"
                },
            "retry": 1,
            "retryIntervalInSeconds": 30
        },
        {   "name": "execute_notebook_3",
            "path": "notebook_03",
            "timeoutPerCellInSeconds": 600,
            "args": {
                        "input_value": "777"
                },
            "retry": 1,
            "retryIntervalInSeconds": 30
        }
        ],
    "timeoutInSeconds": 43200,
    "concurrency": 0
}
```

You need to ensure that the notebooks are executed in the following sequence:
* 1. Notebook_03
* 2. Notebook.Ol
* 3. Notebook_02
Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Split the Directed Acyclic Graph (DAG) definition into three separate definitions.
B. Change the concurrency to 3.
C. Move the declaration of Notebook_03 to the top of the Directed Acyclic Graph (DAG) definition.
D. Move the declaration of Notebook_02 to the bottom of the Directed Acyclic Graph (DAG) definition.
E. Add dependencies to the execution of Note boo k_02.
F. Add dependencies to the execution of Notebook_03.

Answer: CE

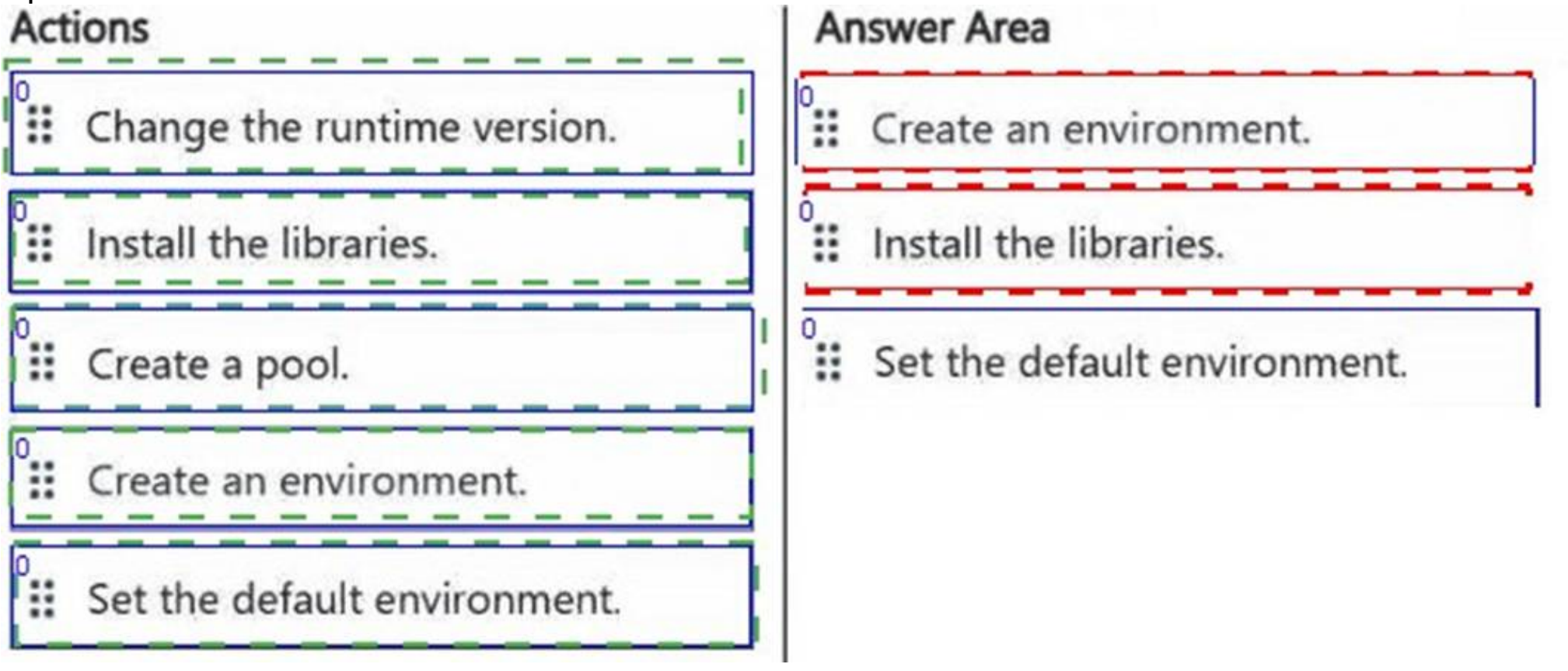
NEW QUESTION 82
DRAG DROP - (Topic 3)
Your company has a team of developers. The team creates Python libraries of reusable code that is used to transform data.
You create a Fabric workspace name Workspace1 that will be used to develop extract, transform, and load (ETL) solutions by using notebooks.
You need to ensure that the libraries are available by default to new notebooks in Workspace1.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

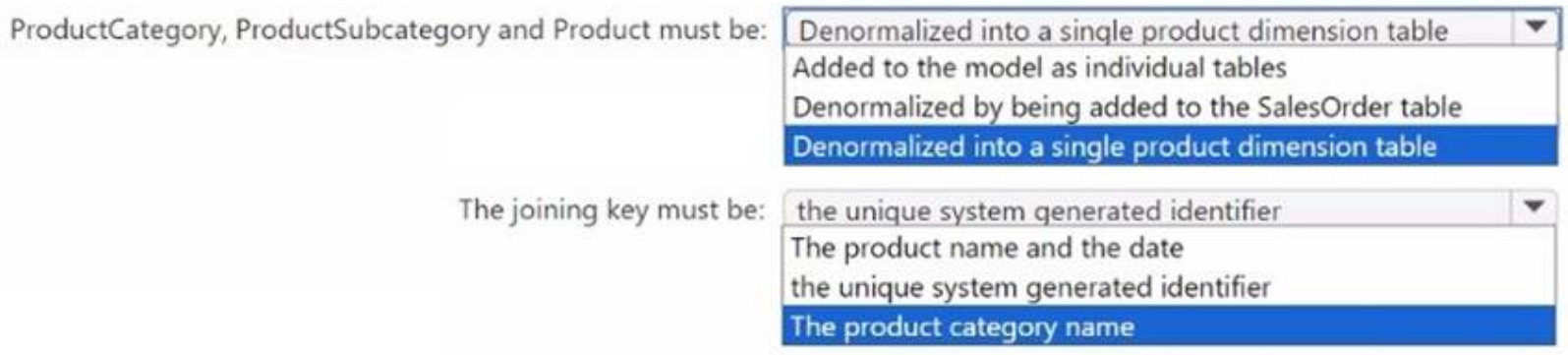
Answer: A

Explanation:



NEW QUESTION 87
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

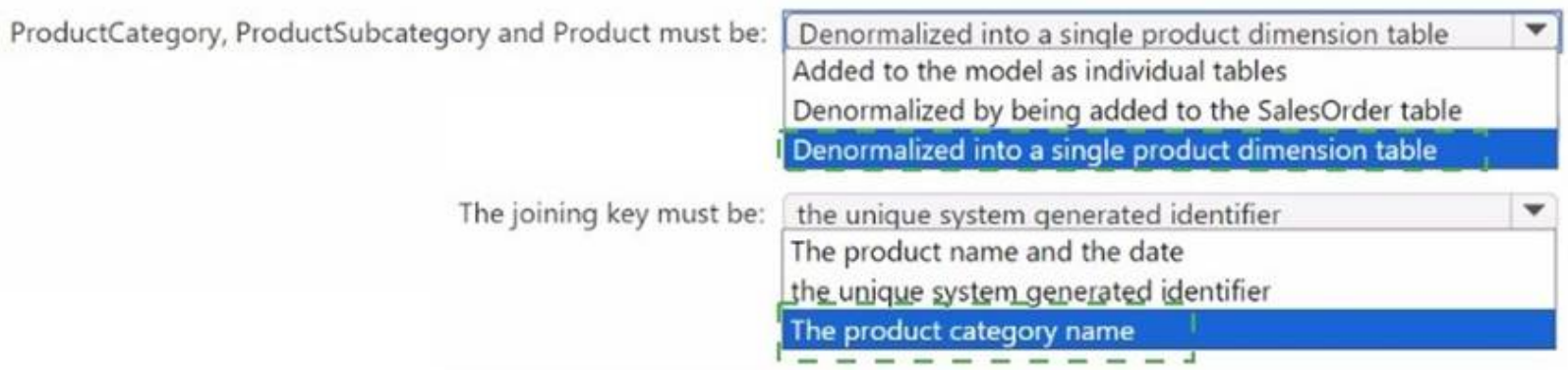


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 88
HOTSPOT - (Topic 3)
You have a table in a Fabric lakehouse that contains the following data.

SalesOrderNumber	OrderDate	CustomerName	Email
SO49172	2021-01-01	Brian Howard	brian23@adventure-works.com
SO49173	2021-01-01	Linda Alvarez	linda19@adventure-works.com
SO49174	2021-01-01	Gina Hernandez	gina4@adventure-works.com
SO49178	2021-01-01	Beth Ruiz	beth4@adventure-works.com
SO49179	2021-01-01	Evan Ward	evan13@adventure-works.com

You have a notebook that contains the following code segment.

```
01 df = df.withColumn("CustomerName", when((col("CustomerName").isNull() | (col("CustomerName")==="")),lit("Unknown")).otherwise(col("CustomerName")))
02 df = df.withColumn("Username",split(col("Email"), "@").getItem(1))
03 df = df.dropDuplicates(["OrderDate"]).select(col("OrderDate"), year("OrderDate").alias("Year"), ("CustomerName"), ("Username"))
04 display(df.head(10))
```

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
Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value.	○	○
Line 02 will extract the value before the @ character and generate a new column named Username.	○	○
Line 03 will extract the year value from the OrderDate column and keep only the first occurrence for each year.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements	Yes	No
Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value.	○	○
Line 02 will extract the value before the @ character and generate a new column named Username.	○	○
Line 03 will extract the year value from the OrderDate column and keep only the first occurrence for each year.	○	○

NEW QUESTION 91
- (Topic 3)
You have a Fabric workspace that contains a lakehouse and a notebook named Notebook1. Notebook1 reads data into a DataFrame from a table named Table1 and applies transformation logic. The data from the DataFrame is then written to a new Delta table named Table2 by using a merge operation.
You need to consolidate the underlying Parquet files in Table1. Which command should you run?

A. VACUUM
B. BROADCAST
C. OPTIMIZE
D. CACHE

Answer: C

Explanation:
To consolidate the underlying Parquet files in Table1 and improve query performance by optimizing the data layout, you should use the OPTIMIZE command in Delta Lake. The OPTIMIZE command coalesces smaller files into larger ones and reorganizes the data for more efficient reads. This is particularly useful when working with large datasets in Delta tables, as it helps reduce the number of files and improves performance for subsequent queries or operations like MERGE.

NEW QUESTION 95
- (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 100
- (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.
You need to reduce how long it takes to run the KQL queryset. Solution: You move the filter to line 02.

```
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)
```

Does this meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Moving the filter to line 02: Filtering the Stream table before performing the join operation reduces the number of rows that need to be processed during the join. This is an effective optimization technique for queries involving large datasets.

NEW QUESTION 105
- (Topic 3)
You have a Fabric workspace that contains a lakehouse named Lakehousel.
You plan to create a data pipeline named Pipeline! to ingest data into Lakehousel. You will use a parameter named paraml to pass an external value into Pipeline1!. The paraml parameter has a data type of int
You need to ensure that the pipeline expression returns param1 as an int value. How should you specify the parameter value?

A. "@pipeline(). parameter
B. paraml"
C. "@{pipeline().parameters.paraml}"
D. "@{pipeline().parameters.[paraml]}"
E. "@{pipeline().parameters.paraml}-

Answer: B

NEW QUESTION 110
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-700 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-700 Product From:

https://www.2passeasy.com/dumps/DP-700/

Money Back Guarantee

DP-700 Practice Exam Features:

* DP-700 Questions and Answers Updated Frequently

* DP-700 Practice Questions Verified by Expert Senior Certified Staff

* DP-700 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DP-700 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year