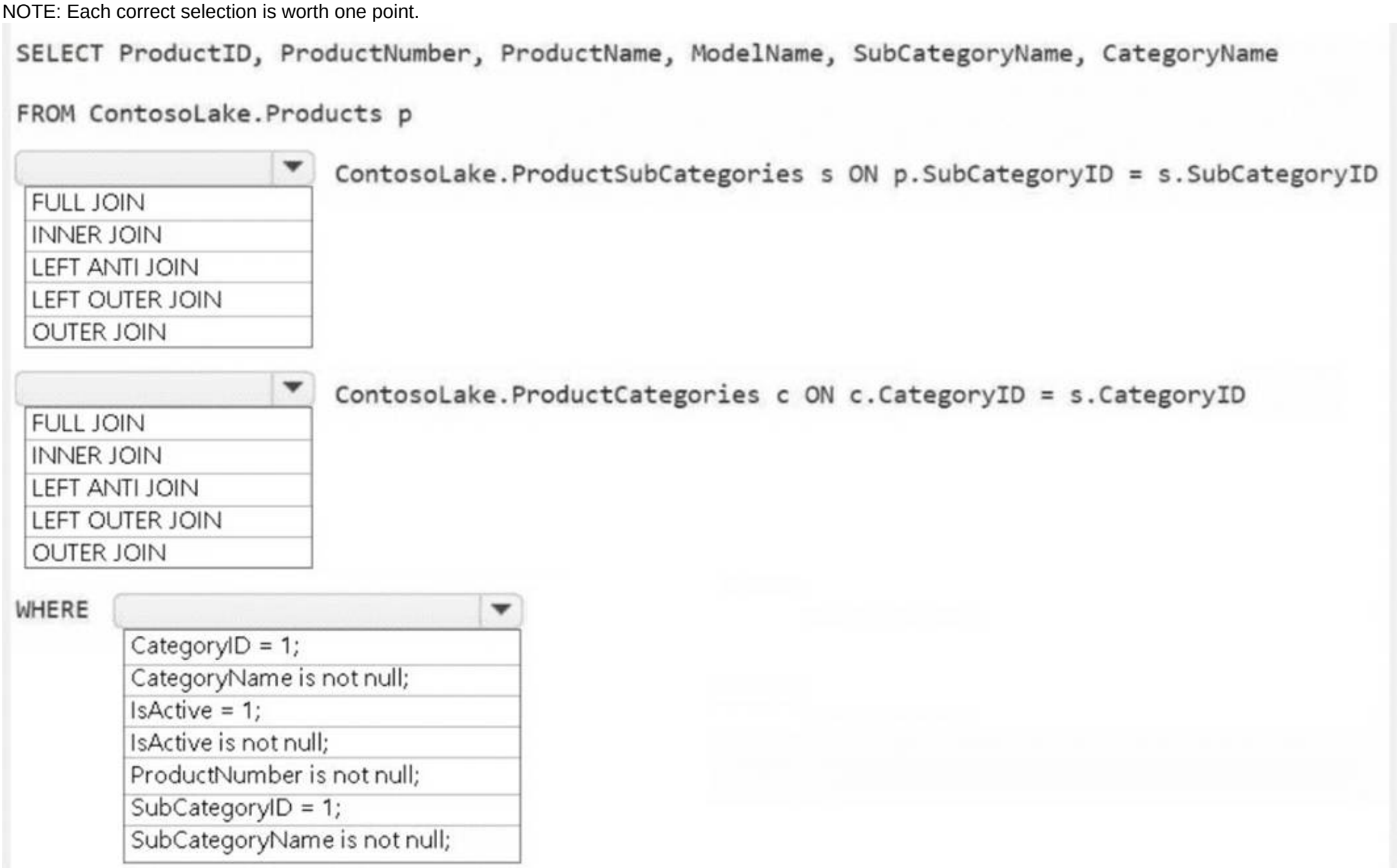
Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

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).

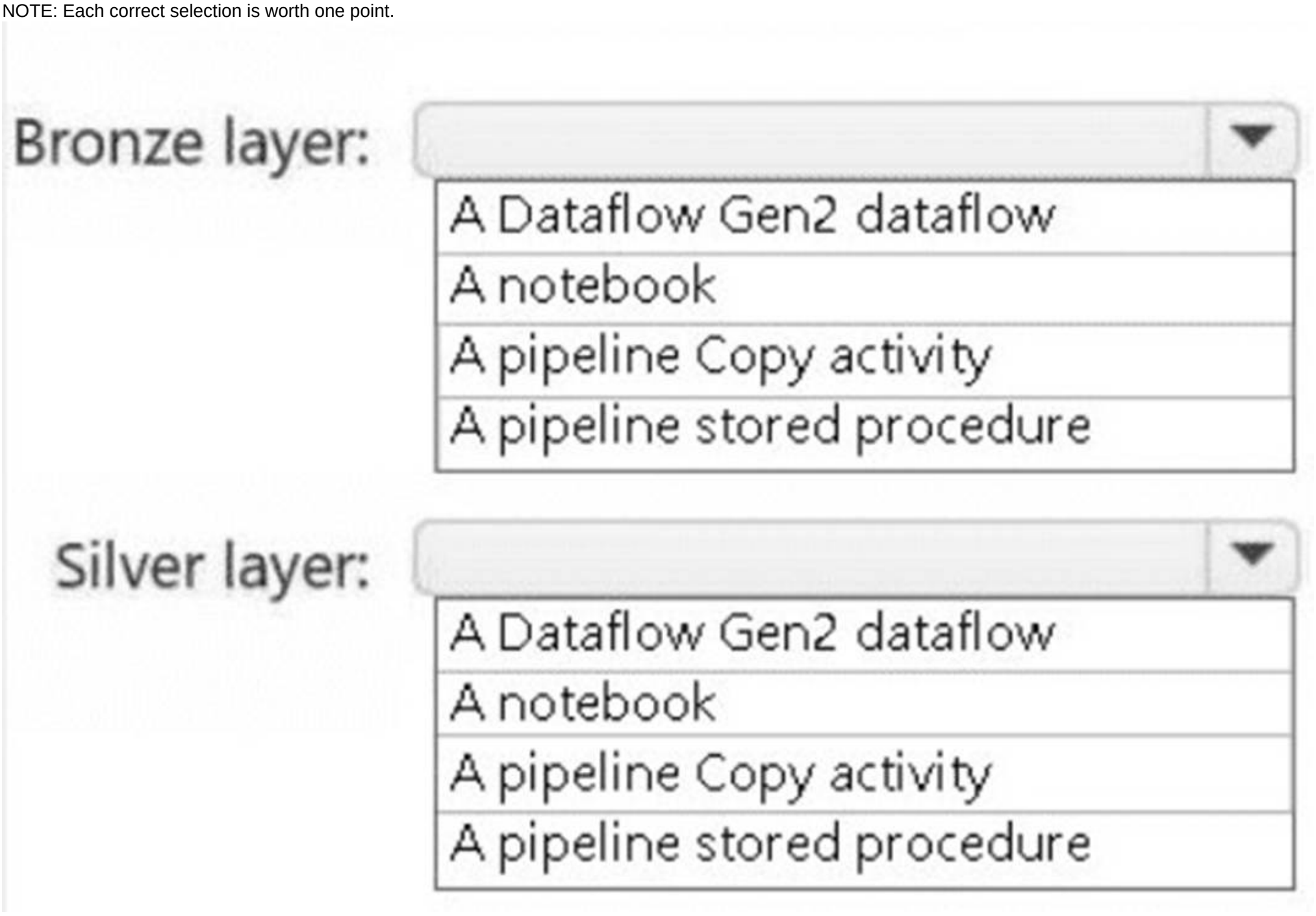
NEW QUESTION 3
- (Topic 1)
You need to ensure that the data analysts can access the gold layer lakehouse. What should you do?

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
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 5
- (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 6
- (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 7
- (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1.
While monitoring Warehouse1, you discover that query performance has degraded during the last 60 minutes.
You need to isolate all the queries that were run during the last 60 minutes. The results must include the username of the users that submitted the queries and the query statements. What should you use?

A. the Microsoft Fabric Capacity Metrics app
B. views from the queryinsights schema
C. Query activity
D. the sys.dm_exec_requests dynamic management view

Answer: B

NEW QUESTION 8
HOTSPOT - (Topic 3)
You have three users named User1, User2, and User3.
You have the Fabric workspaces shown in the following table.

Name	Workspace admin
Workspace1	User1
Workspace2	User2

You have a security group named Group1 that contains User1 and User3. The Fabric admin creates the domains shown in the following table.

Name	Domain admin
Domain1	User1
Domain2	User2

User1 creates a new workspace named Workspace3. You add Group1 to the default domain of Domain1.
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
User3 has Viewer role access to Workspace3.	○	○
User3 has Domain contributor access to Domain1.	○	○
User2 has Contributor role access to Workspace3.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area

Statements	Yes	No
User3 has Viewer role access to Workspace3.	○	○
User3 has Domain contributor access to Domain1.	○	○
User2 has Contributor role access to Workspace3.	○	○

NEW QUESTION 9
- (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 10
- (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 10
- (Topic 3)
You have a Fabric workspace that contains a Real-Time Intelligence solution and an eventhouse.
Users report that from OneLake file explorer, they cannot see the data from the eventhouse.
You enable OneLake availability for the eventhouse. What will be copied to OneLake?

A. only data added to new databases that are added to the eventhouse
B. only the existing data in the eventhouse
C. no data
D. both new data and existing data in the eventhouse
E. only new data added to the eventhouse

Answer: D

Explanation:
When you enable OneLake availability for an eventhouse, both new and existing data in the eventhouse will be copied to OneLake. This feature ensures that data, whether newly ingested or already present, becomes available for access through OneLake, making it easier for users to interact with and explore the data directly from OneLake file explorer.

NEW QUESTION 11
- (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 13
- (Topic 3)
You are implementing a medallion architecture in a Fabric lakehouse.
You plan to create a dimension table that will contain the following columns:
• ID
• CustomerCode
• CustomerName

• CustomerAddress
• CustomerLocation
• ValidFrom
• ValidTo

You need to ensure that the table supports the analysis of historical sales data by customer location at the time of each sale Which type of slowly changing dimension (SCD) should you use?

A. Type 2
B. Type 0
C. Type 1
D. Type 3

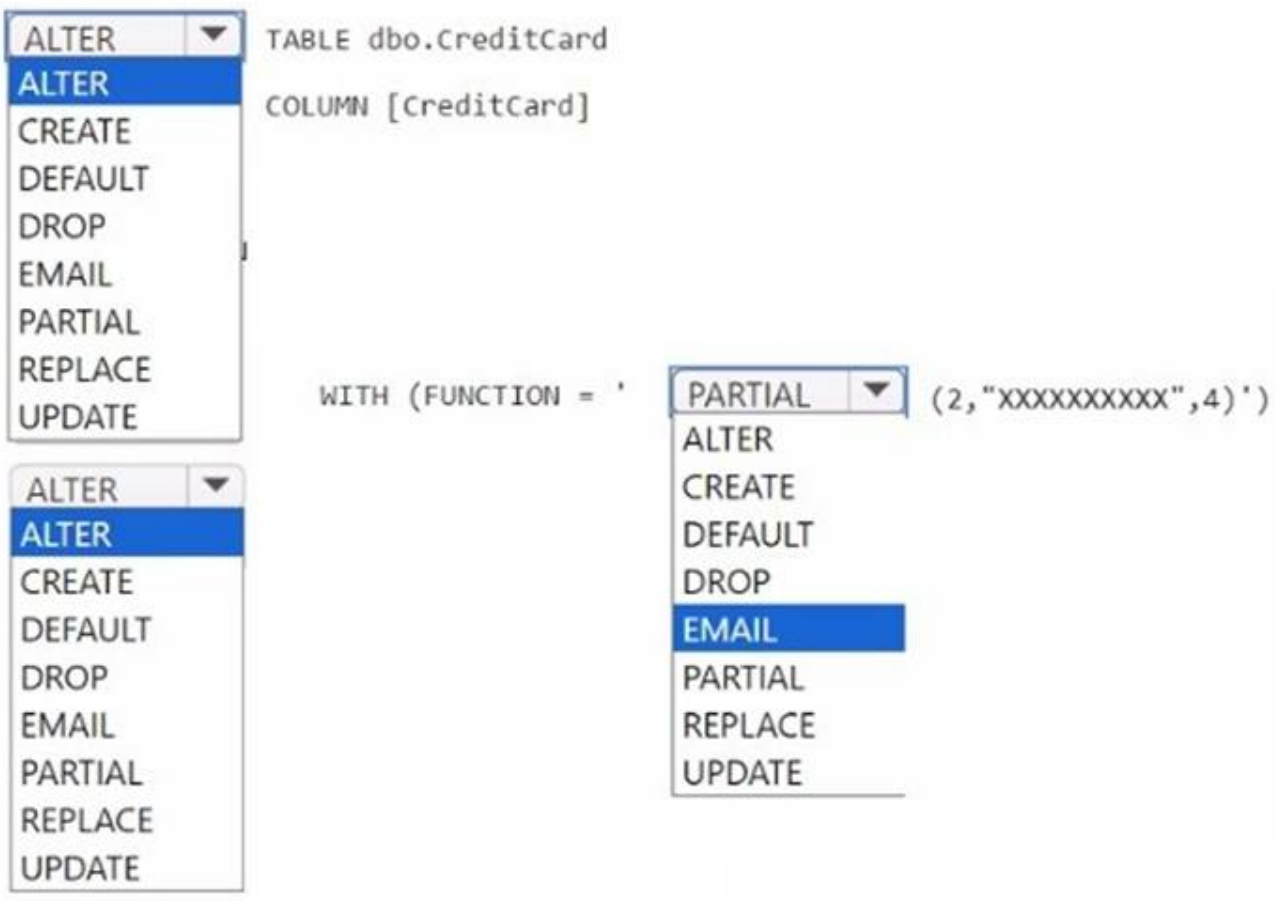
Answer: A

NEW QUESTION 17
HOTSPOT - (Topic 3)
You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse2. A team of data analysts has Viewer role access to Workspace1. You create a table by running the following statement.

```
CREATE TABLE [warehouse2].[dbo].[CreditCard]
(
 CreditCard varchar(20) NOT NULL
 ,CreditCardType varchar(10) NOT NULL)
GO
```

You need to ensure that the team can view only the first two characters and the last four characters of the Creditcard attribute.
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 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

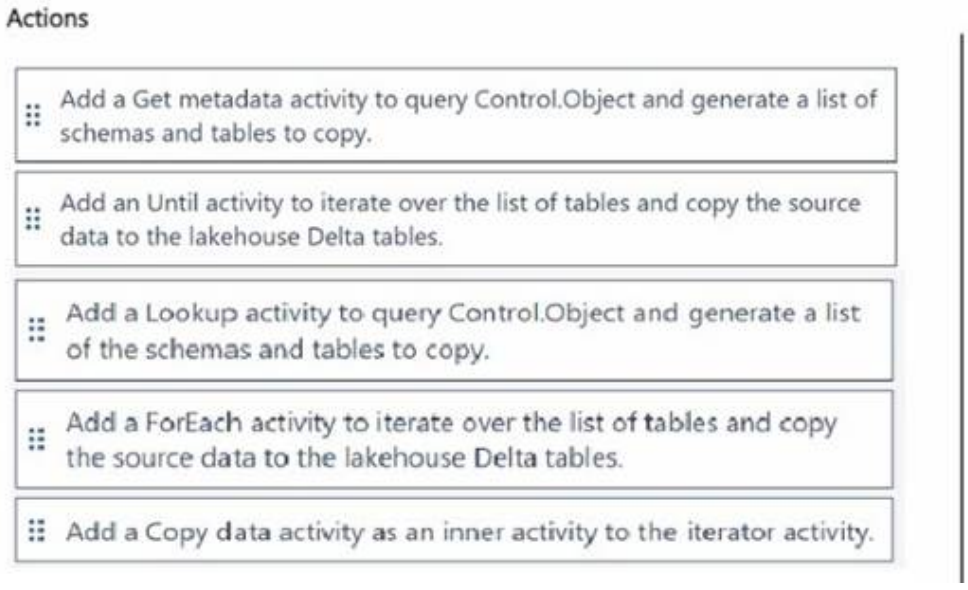
Explanation:
An outer join will include unmatched rows from both tables, increasing the dataset size and processing time. It does not improve query performance.

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

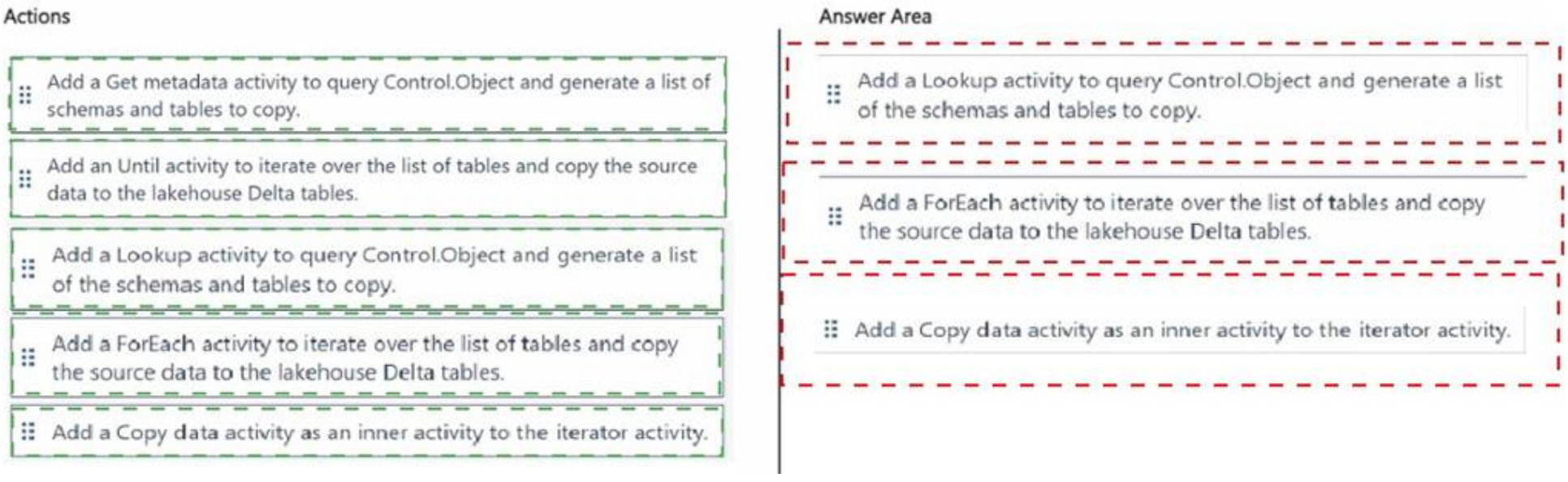
NEW QUESTION 27
DRAG DROP - (Topic 3)
You are building a data loading pattern by using a Fabric data pipeline. The source is an Azure SQL database that contains 25 tables. The destination is a lakehouse.
In a warehouse, you create a control table named Control.Object as shown in the exhibit. (Click the 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 28
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

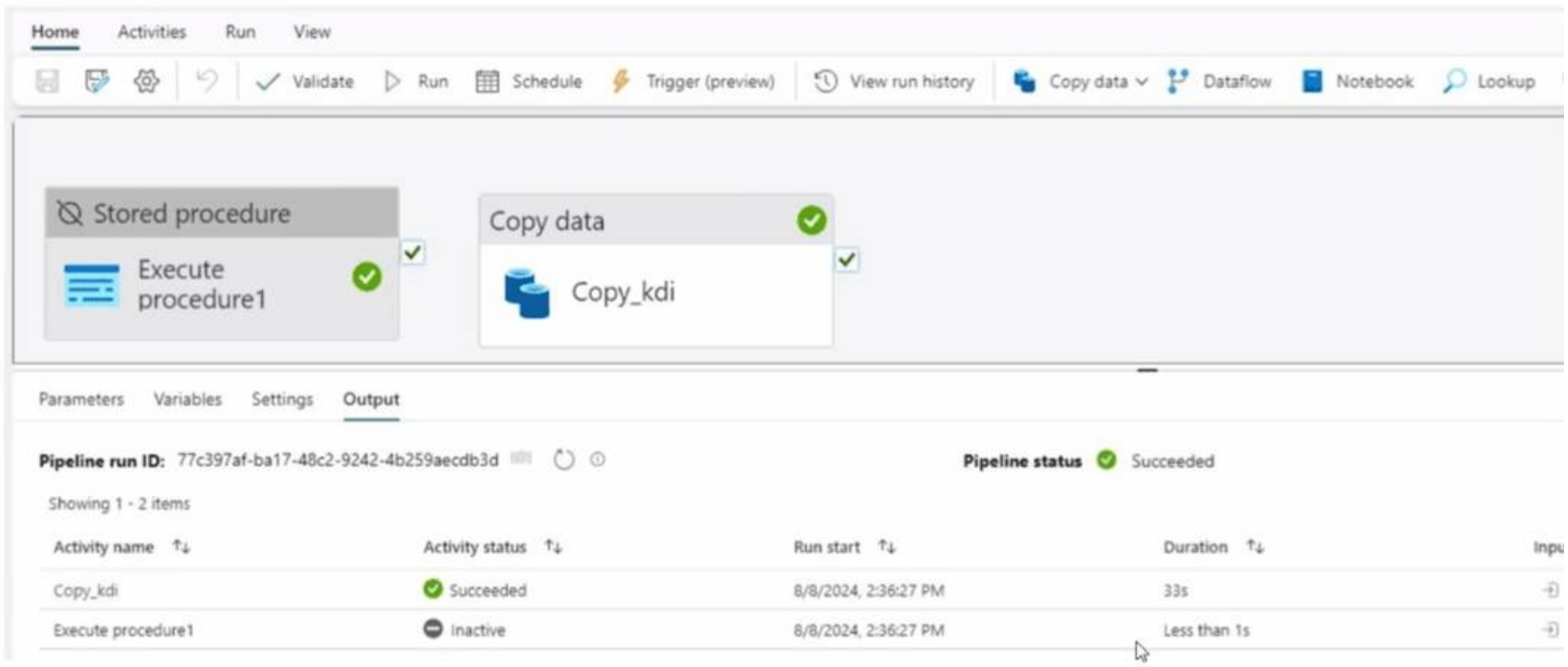
Answer: A

Explanation:

Answer Area

```
SELECT
  item_id as ItemId
  [ ▼ ]                                          as ItemName
      ,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
```

NEW QUESTION 29
- (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 32
- (Topic 3)
Your company has a sales department that uses two Fabric workspaces named Workspace1 and Workspace2.
The company decides to implement a domain strategy to organize the workspaces. You need to ensure that a user can perform the following tasks:
Create a new domain for the sales department.
Create two subdomains: one for the east region and one for the west region. Assign Workspace1 to the east region subdomain.
Assign Workspace2 to the west region subdomain. The solution must follow the principle of least privilege. Which role should you assign to the user?

A. workspace Admin
B. domain admin
C. domain contributor
D. Fabric admin

Answer: B

Explanation:
To implement a domain strategy and manage subdomains within Fabric, the domain admin role is the appropriate role for the user. A domain admin has the permissions necessary to:
? Create a new domain (for the sales department).
? Create subdomains (for the east and west regions).
? Assign workspaces (such as Workspace1 and Workspace2) to the appropriate subdomains.
The domain admin role allows for managing the structure and organization of workspaces in the context of domains and subdomains while maintaining the principle of least privilege by limiting the user's access to managing the domain structure specifically.

NEW QUESTION 36
- (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. a Dataflow Gen1 dataflow
B. a data pipeline
C. a KQL queryset
D. a notebook

Answer: B

Explanation:
To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Microsoft Fabric, the best option is to use a data pipeline. A data pipeline in Fabric allows for the orchestration of data movement, from source to destination, using connectors, transformations, and scheduled workflows. Since the data is being transferred from an on-premises database and requires the use of a data gateway, a data pipeline provides the appropriate framework to facilitate this data movement efficiently and reliably.

NEW QUESTION 39
- (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 40
- (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

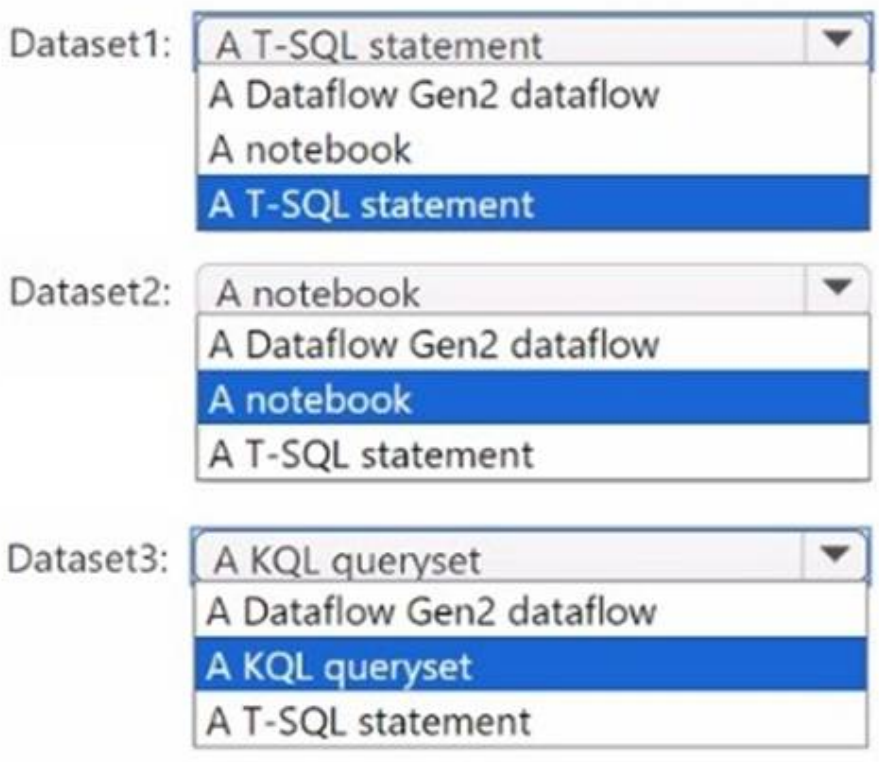
NEW QUESTION 44
- (Topic 3)
You have an Azure key vault named KeyVaultl that contains secrets.
You have a Fabric workspace named Workspace-!. Workspace! contains a notebook named Notebookl that performs the following tasks:
• Loads stage data to the target tables in a lakehouse
• Triggers the refresh 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 the key vault URL and key vault secret. Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 48
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

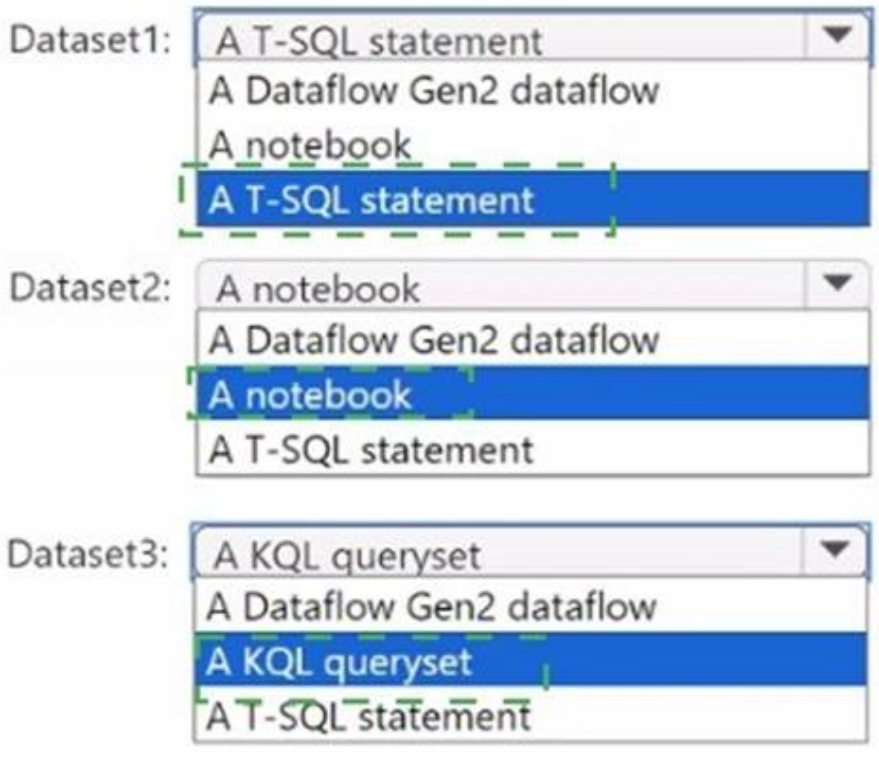


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 49
- (Topic 3)
You have a Fabric workspace that contains a lakehouse named Lakehouse1.
In an external data source, you have data files that are 500 GB each. A new file is added every day.
You need to ingest the data into Lakehouse1 without applying any transformations. The solution must meet the following requirements
Trigger the process when a new file is added. Provide the highest throughput.
Which type of item should you use to ingest the data?

A. Data pipeline
B. Environment
C. KQL queryset
D. Dataflow Gen2

Answer: A

Explanation:
To efficiently ingest large data files (500 GB each) into Lakehouse1 with high throughput and trigger the process when a new file is added, a Data pipeline is the most suitable solution. Data pipelines in Fabric are ideal for orchestrating data movement and can be configured to automatically trigger based on file arrivals or other events. This solution meets both requirements: ingesting the data without transformations (since you just need to copy the data) and triggering the process when new files are added.

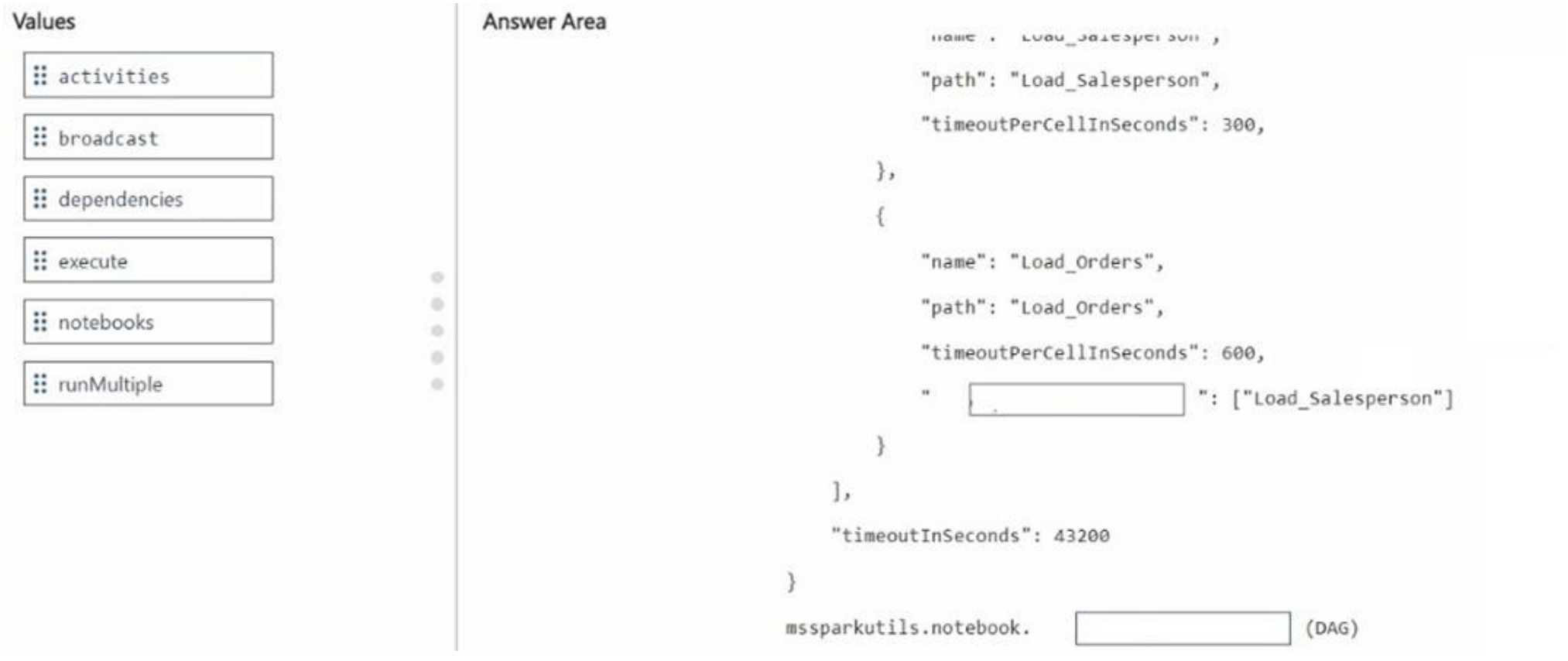
NEW QUESTION 53
- (Topic 3)
You have five Fabric workspaces.
You are monitoring the execution of items by using Monitoring hub.
You need to identify in which workspace a specific item runs. Which column should you view in Monitoring hub?

A. Start time
B. Capacity
C. Activity name
D. Submitter
E. Item type
F. Job type
G. Location

Answer: G

Explanation:
To identify in which workspace a specific item runs in Monitoring hub, you should view the Location column. This column indicates the workspace where the item is executed. Since you have multiple workspaces and need to track the execution of items across them, the Location column will show you the exact workspace associated with each item or job execution.

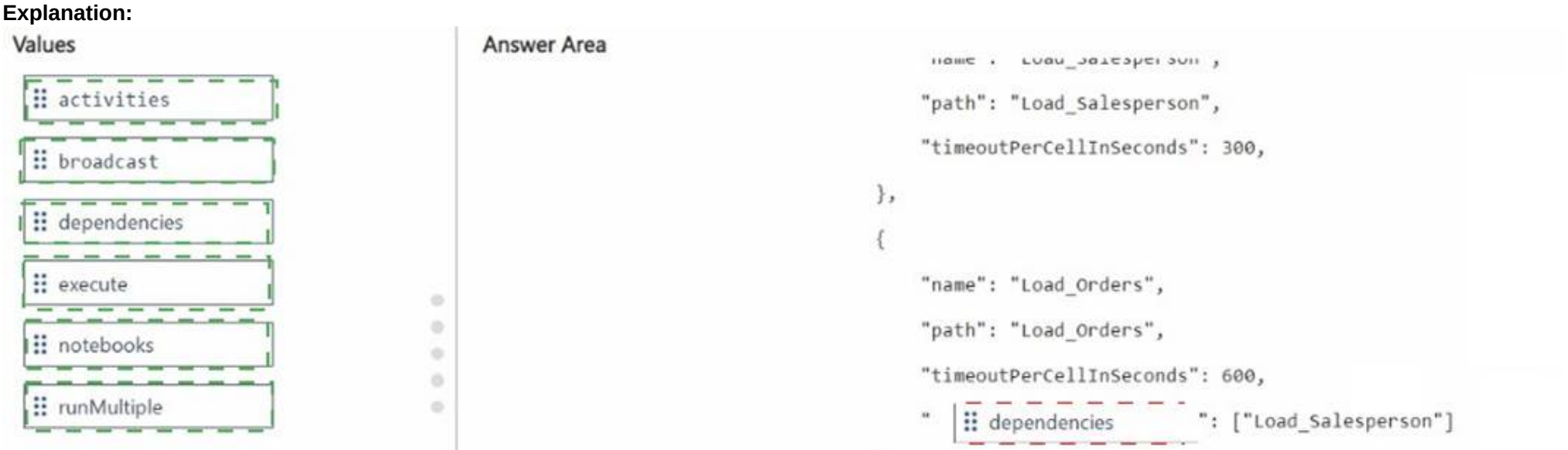
NEW QUESTION 56
DRAG DROP - (Topic 3)
You have two Fabric notebooks named Load_Salesperson and Load_Orders that read data from Parquet files in a lakehouse. Load_Salesperson writes to a Delta table named dim_salesperson. Load.Orders writes to a Delta table named fact_orders and is dependent on the successful execution of Load_Salesperson.
You need to implement a pattern to dynamically execute Load_Salesperson and Load_Orders in the appropriate order by using a notebook.
How should you complete the code? 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

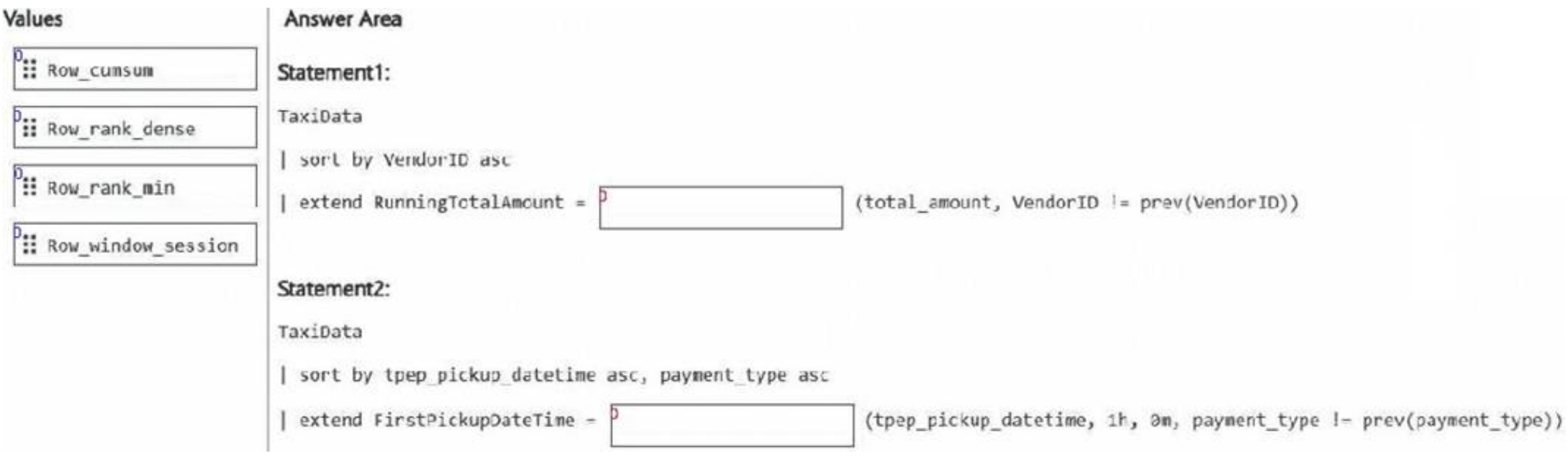
Explanation:



NEW QUESTION 59
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 64
- (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 66
- (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 71
- (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

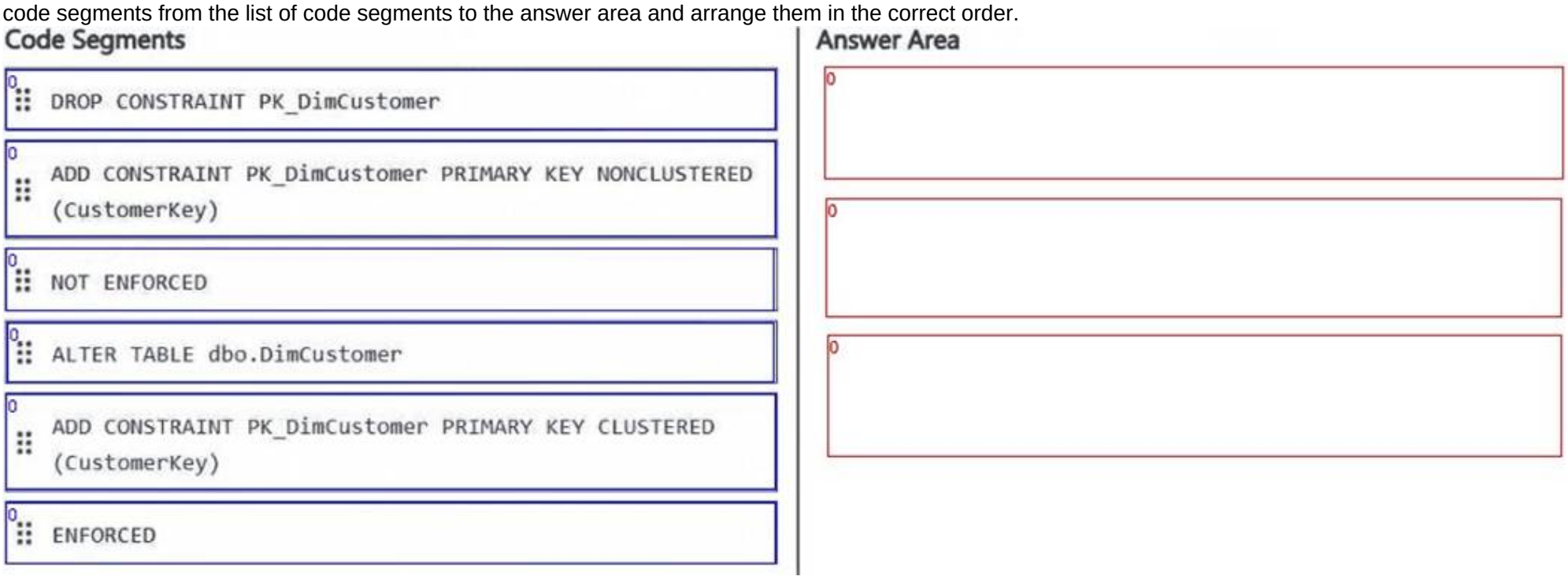
Answer: B

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 72
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
    );
```

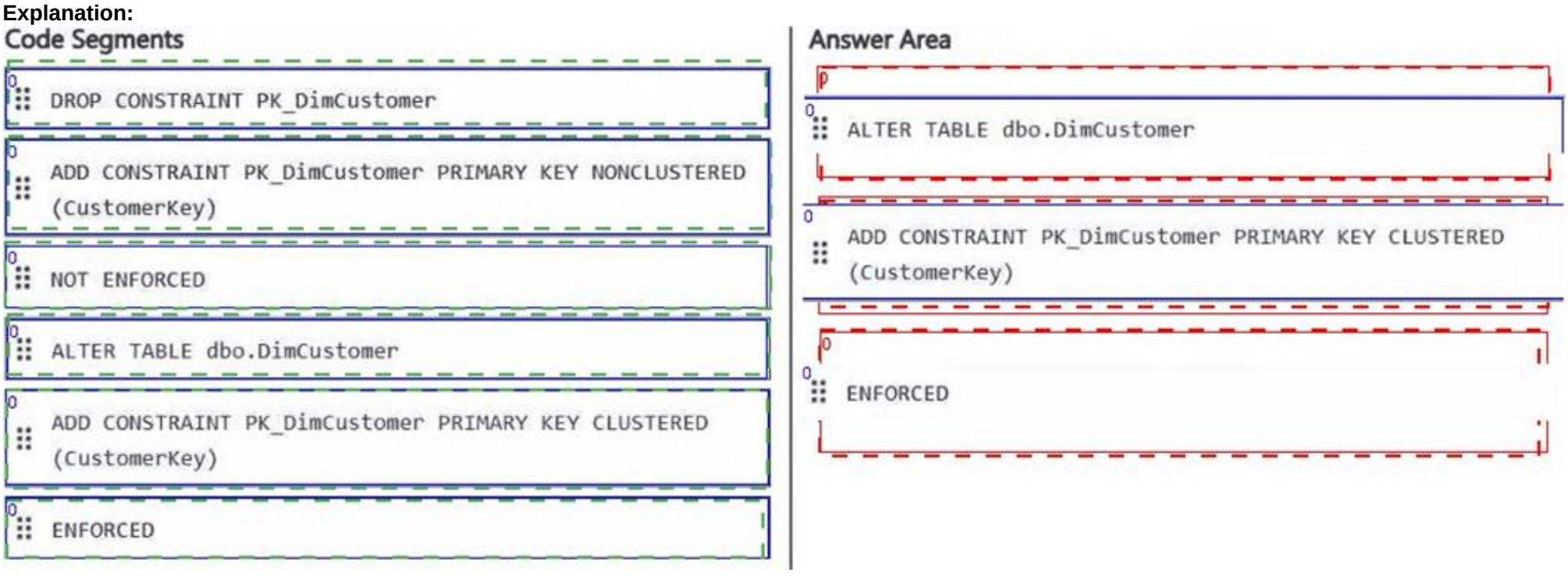
You need to set the Customerkey column as a primary key of the DimCustomer table. Which three code segments should you run in sequence? To answer, move the appropriate
code segments from the list of code segments to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

Answer: A

Explanation:

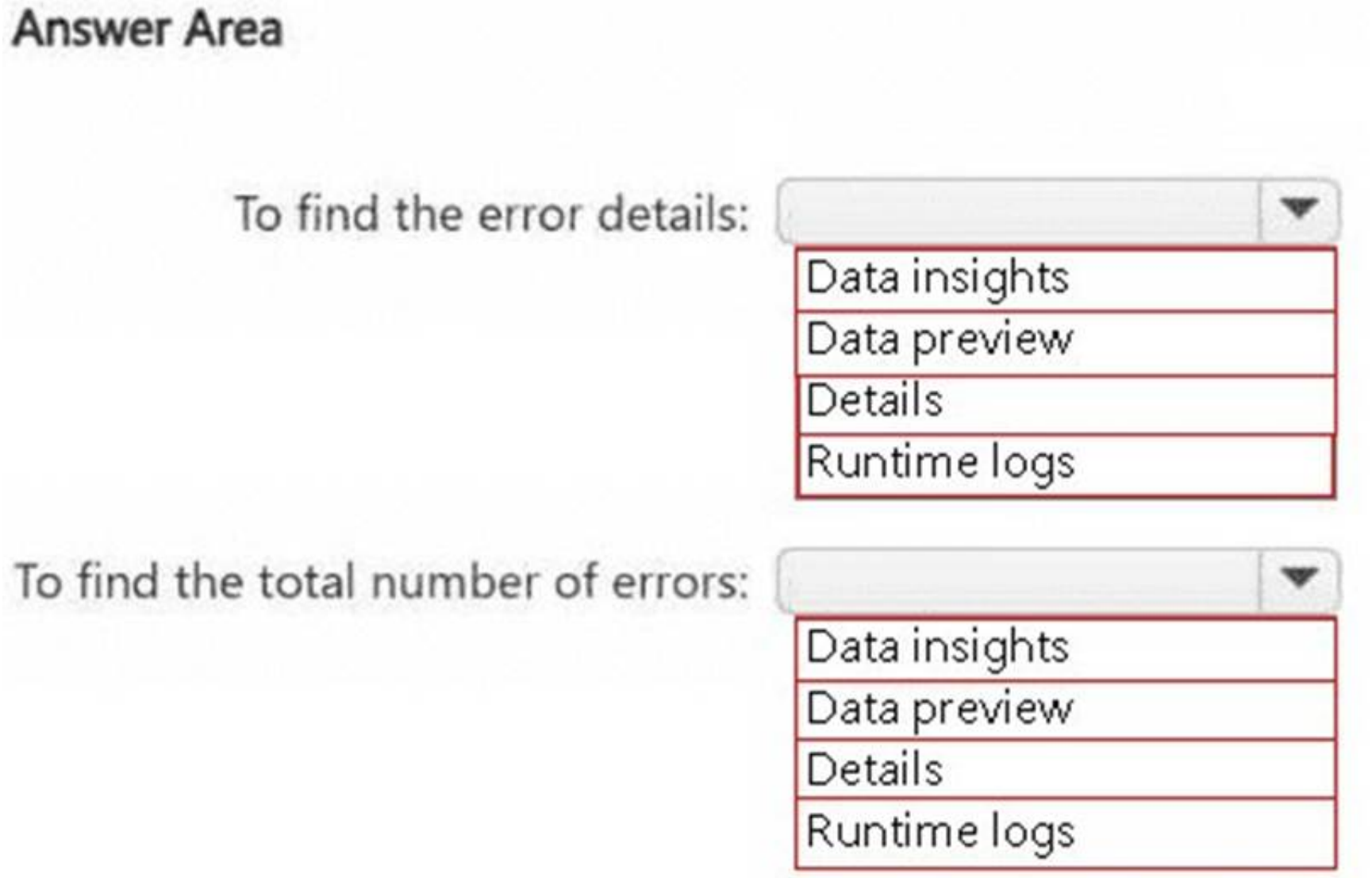


NEW QUESTION 75
- (Topic 3)
You have a Fabric warehouse named DW1 that contains a Type 2 slowly changing dimension (SCD) dimension table named DimCustomer. DimCustomer contains 100 columns and 20 million rows. The columns are of various data types, including int, varchar, date, and varbinary.
You need 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

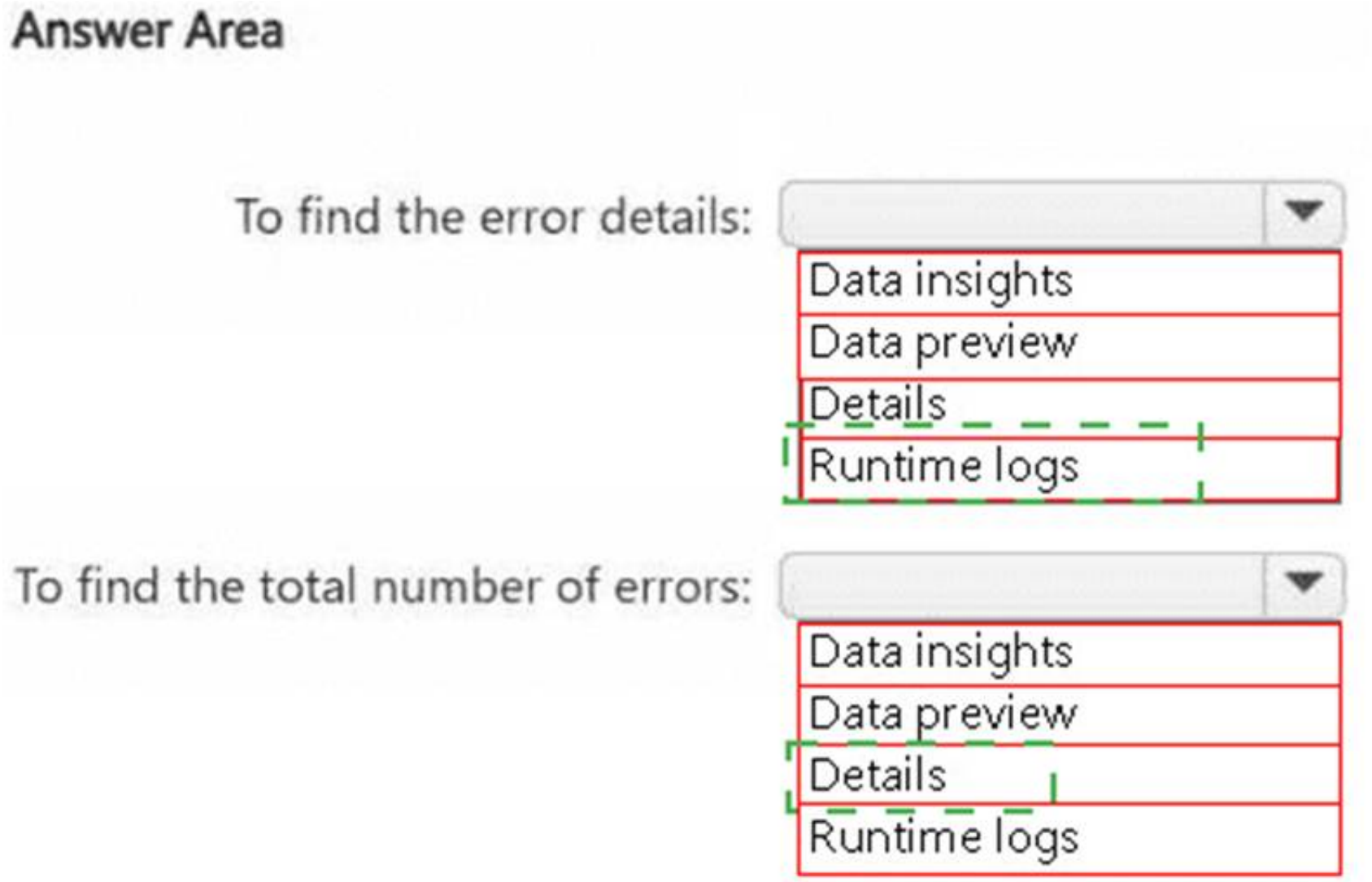
NEW QUESTION 76
HOTSPOT - (Topic 3)
You have a Fabric workspace that contains an eventstream named EventStream1. You discover that an EventStream1 transformation fails.
You need to find the following error information: The error details, including the occurrence time The total number of errors
What should 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 79
- (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.
What should you configure?

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

Answer: C

Explanation:
When using the Copy data activity in a data pipeline to move data from Snowflake to a Fabric warehouse, the process often involves intermediate staging to handle data efficiently, especially for large datasets or cross-cloud data transfers.
Staging involves temporarily storing data in an intermediate location (e.g., Blob storage or Azure Data Lake) before loading it into the target destination.
For cross-cloud data transfers (e.g., from Snowflake to Fabric), enabling staging ensures data is processed and stored temporarily in an efficient format for transfer.
Staging is especially useful when dealing with large datasets, ensuring the process is optimized and avoids memory limitations.

NEW QUESTION 84
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.	○	○

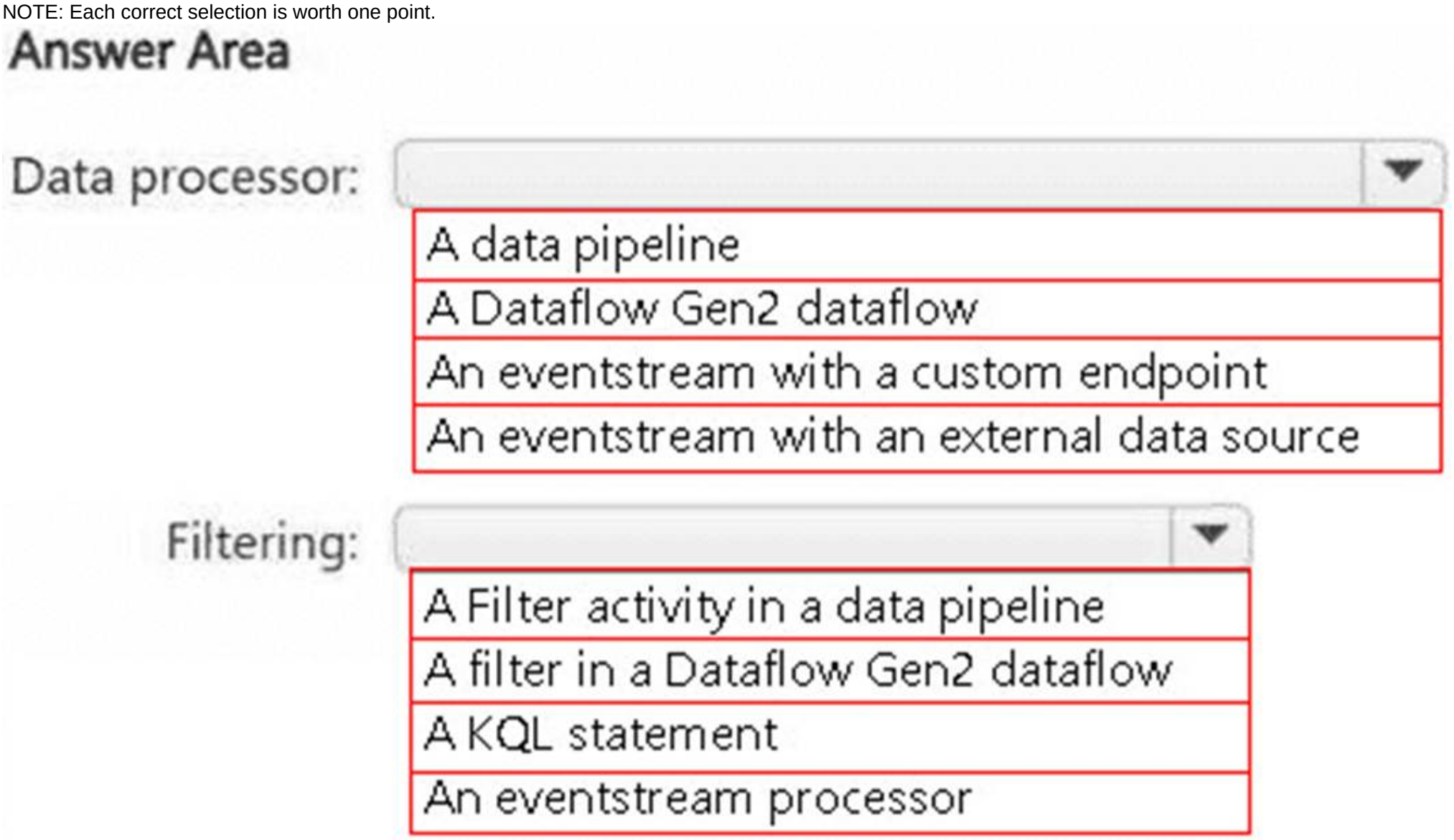
A. Mastered
B. Not Mastered

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.	○	○

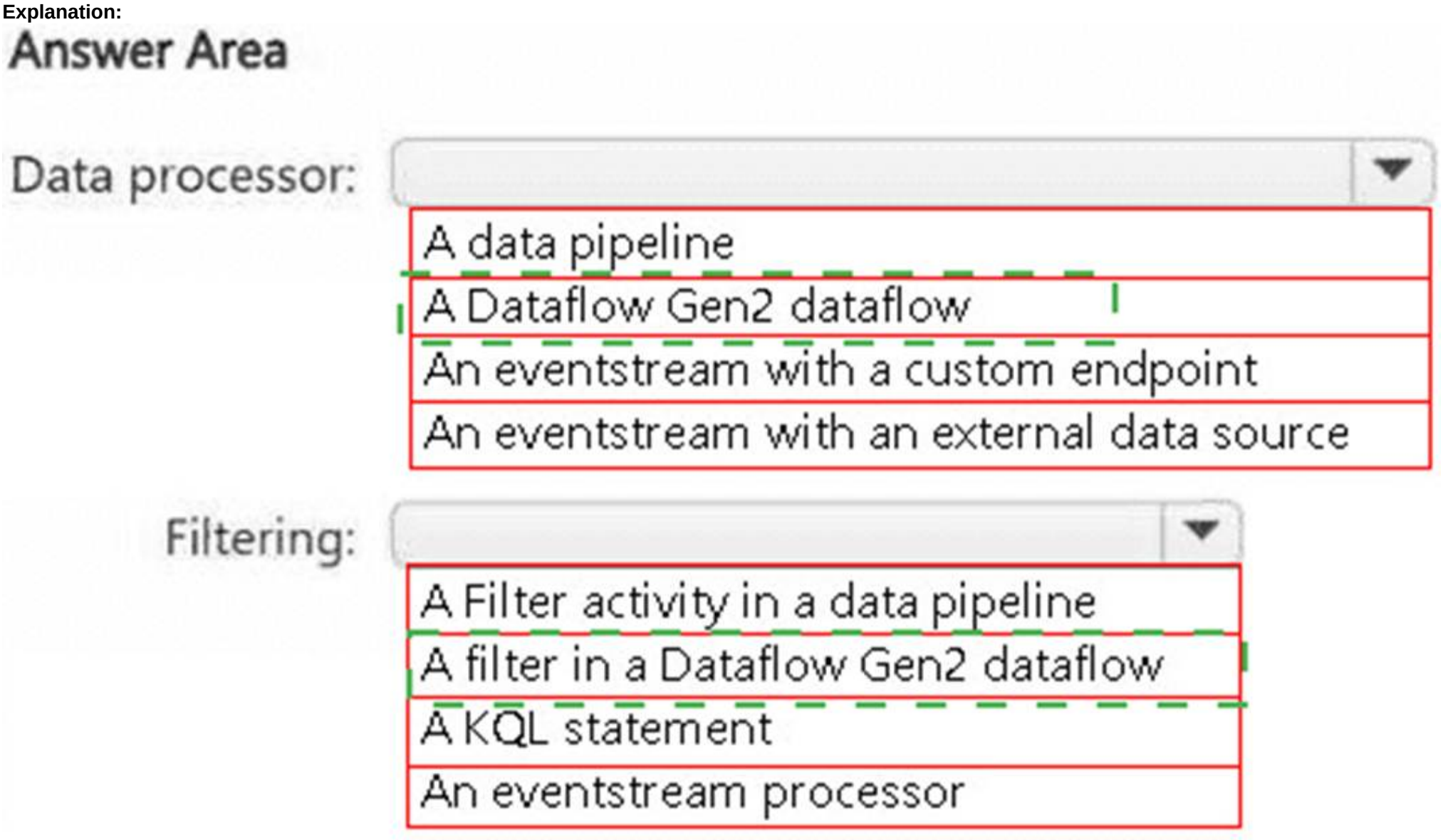
NEW QUESTION 89
HOTSPOT - (Topic 3)
You have an Azure Event Hubs data source that contains weather data.
You ingest the data from the data source by using an eventstream named Eventstream1. Eventstream1 uses a lakehouse as the destination.
You need to batch ingest only rows from the data source where 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 91
- (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 96
- (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:

```
SELECT BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
FROM bike_location
WHERE neighbourhood = 'Sands End'
AND no_bikes >= 15
ORDER BY no_bikes
```

Does this meet the goal?

A. Yes
B. no

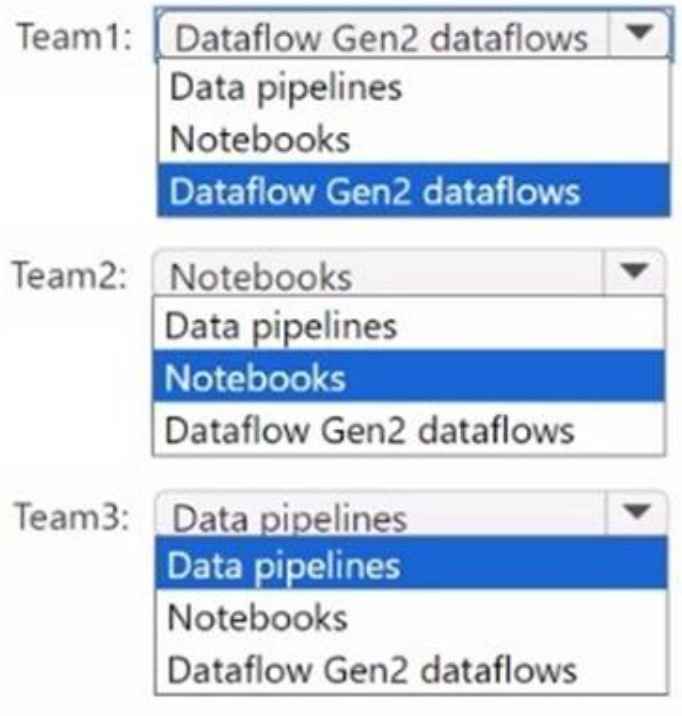
Answer: B

Explanation:
This code does not meet the goal because this is an SQL-like query and cannot be executed in KQL, which is required for the database.
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 101
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

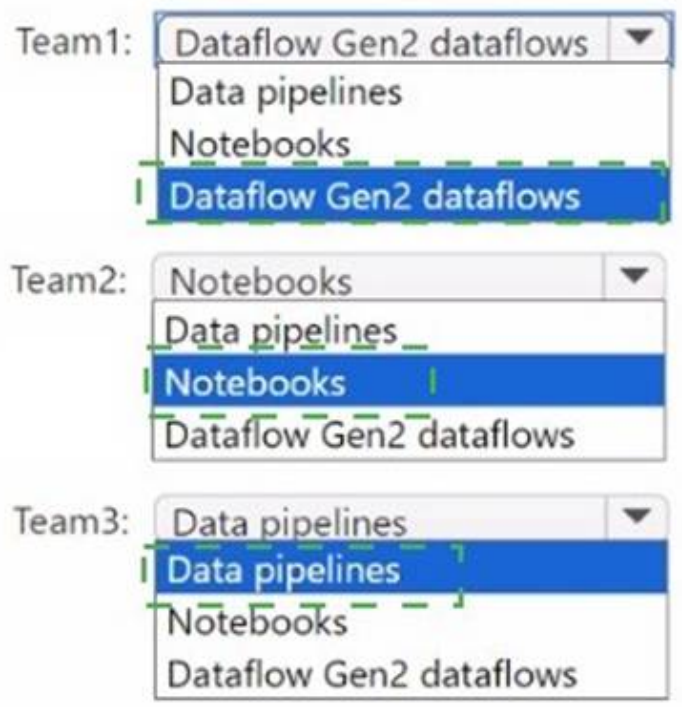


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 103
- (Topic 3)
You have a Fabric workspace named Workspace1.
You plan to configure Git integration for Workspacel by using an Azure DevOps Git repository. An Azure DevOps admin creates the required artifacts to support the integration of Workspacel Which details do you require to perform the integration?

A. the project, Git repository, branch, and Git folder
B. the organization, projec
C. Git repository, and branch
D. the Git repository URL and the Git folder
E. the personal access token (PAT) for Git authentication and the Git repository URL

Answer: B

NEW QUESTION 106
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a data pipeline named Pipeline1 and a lakehouse named Lakehouse1.

You have a deployment pipeline named deployPipeline1 that deploys Workspace1 to Workspace2.
You restructure Workspace1 by adding a folder named Folder1 and moving Pipeline1 to Folder1.
You use deployPipeline1 to deploy Workspace1 to Workspace2. What occurs to Workspace2?

A. Folder1 is created, Pipeline1 moves to Folder1, and Lakehouse1 is deployed.
B. Only Pipeline1 and Lakehouse1 are deployed.
C. Folder1 is created, and Pipeline1 and Lakehouse1 move to Folder1.
D. Only Folder1 is created and Pipeline1 moves to Folder1.

Answer: A

Explanation:
When you restructure Workspace1 by adding a new folder (Folder1) and moving Pipeline1 into it, deployPipeline1 will deploy the entire structure of Workspace1 to Workspace2, preserving the changes made in Workspace1. This includes:
Folder1 will be created in Workspace2, mirroring the structure in Workspace1.
Pipeline1 will be moved into Folder1 in Workspace2, maintaining the same folder structure. Lakehouse1 will be deployed to Workspace2 as it exists in Workspace1.

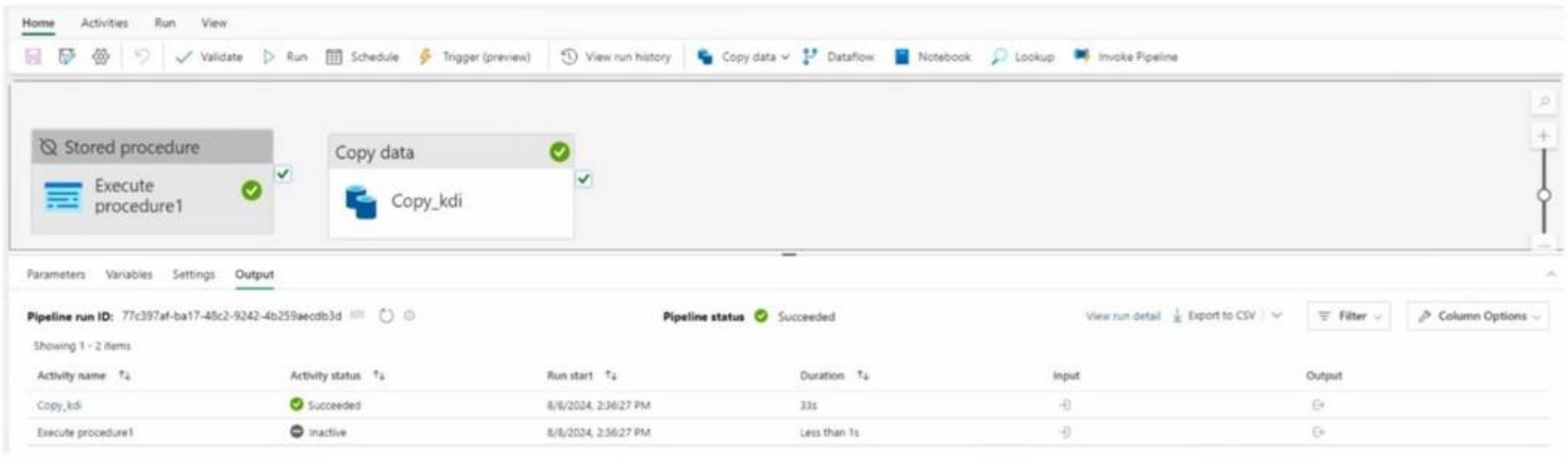
NEW QUESTION 109
- (Topic 3)
You have a Fabric workspace that contains a lakehouse and a notebook named Notebook1. Notebook1 reads data into a DataFrame from a table named Table1 and applies transformation logic. The data from the DataFrame is then written to a new Delta table named Table2 by using a 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 111
- (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 112
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 113
- (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 115
- (Topic 3)
You have a Fabric workspace that contains an eventstream named Eventstream1. Eventstream1 processes data from a thermal sensor by using event stream processing, and then stores the data in a lakehouse.
You need to modify Eventstream1 to include the standard deviation of the temperature. Which transform operator should you include in the Eventstream1 logic?

A. Expand
B. Group by
C. Union
D. Aggregate

Answer: D

Explanation:
To compute the standard deviation of the temperature from the thermal sensor data, you would use the Aggregate transform operator in Eventstream1. The Aggregate operator allows you to apply functions like sum, average, count, and statistical functions like standard deviation across a group of rows or events. This operator is ideal for operations that require summarizing or computing statistics over a dataset, such as calculating the standard deviation.

NEW QUESTION 116
.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

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