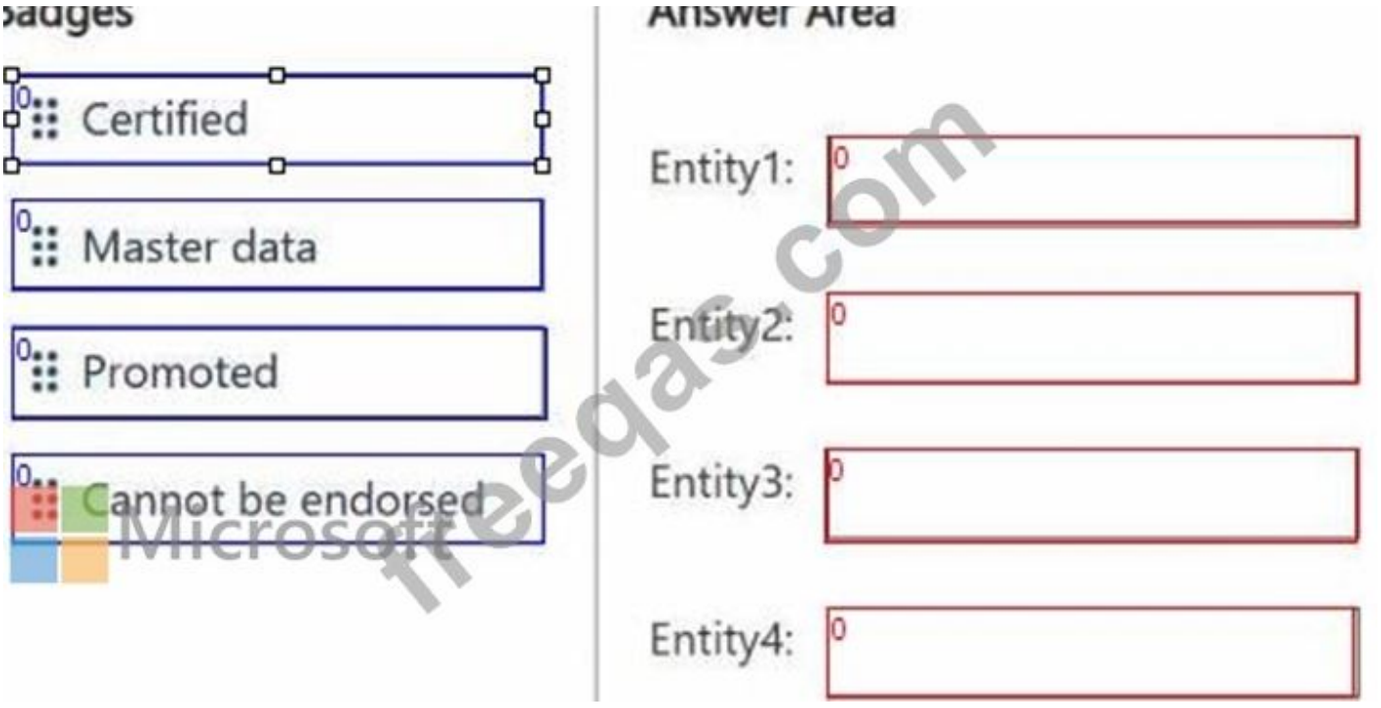
Microsoft.DP-700.v2025-08-07.q46

Exam Code:	DP-700
Exam Name:	Implementing Data Engineering Solutions Using Microsoft Fabric
Certification Provider:	Microsoft
Free Question Number:	46
Version:	v2025-08-07
# of views:	129
# of Questions views:	460
https://www.freeqas.com/qa/Microsoft/DP-700/Microsoft.DP-700.v2025-08-07.q46.html	

NEW QUESTION: 1
You are implementing the following data entities in a Fabric environment:
Entity1: Available in a lakehouse and contains data that will be used as a core organization entity
Entity2: Available in a semantic model and contains data that meets organizational standards
Entity3: Available in a Microsoft Power BI report and contains data that is ready for sharing and reuse Entity4: Available in a Power BI dashboard and contains approved data for executive-level decision making Your company requires that specific governance processes be implemented for the data.
You need to apply endorsement badges to the entities based on each entity's use case.
Which badge should you apply to each entity? To answer, drag the appropriate badges the correct entities. Each badge 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.



Answer:

NEW QUESTION: 2

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:

NEW QUESTION: 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:

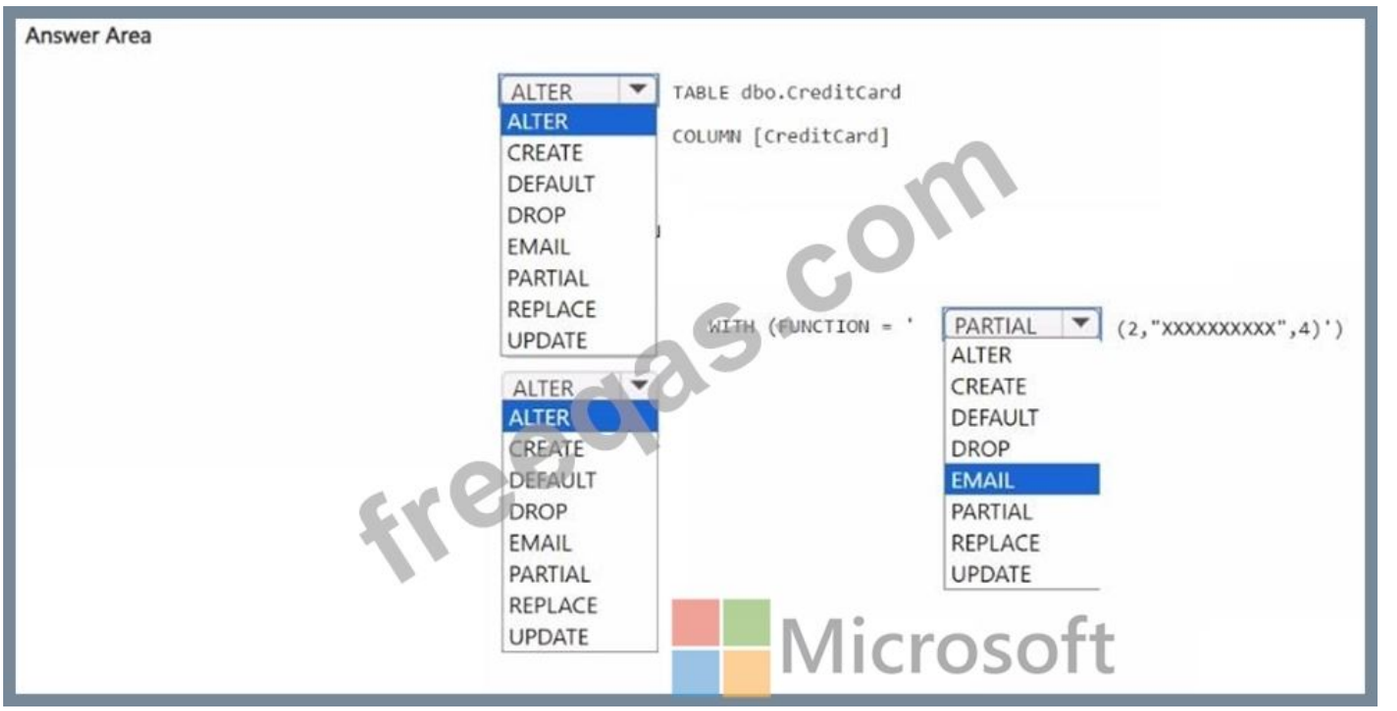
```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| sort by No_Bikes asc
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?
A. Yes
B. no
Answer: (SHOW ANSWER)
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.

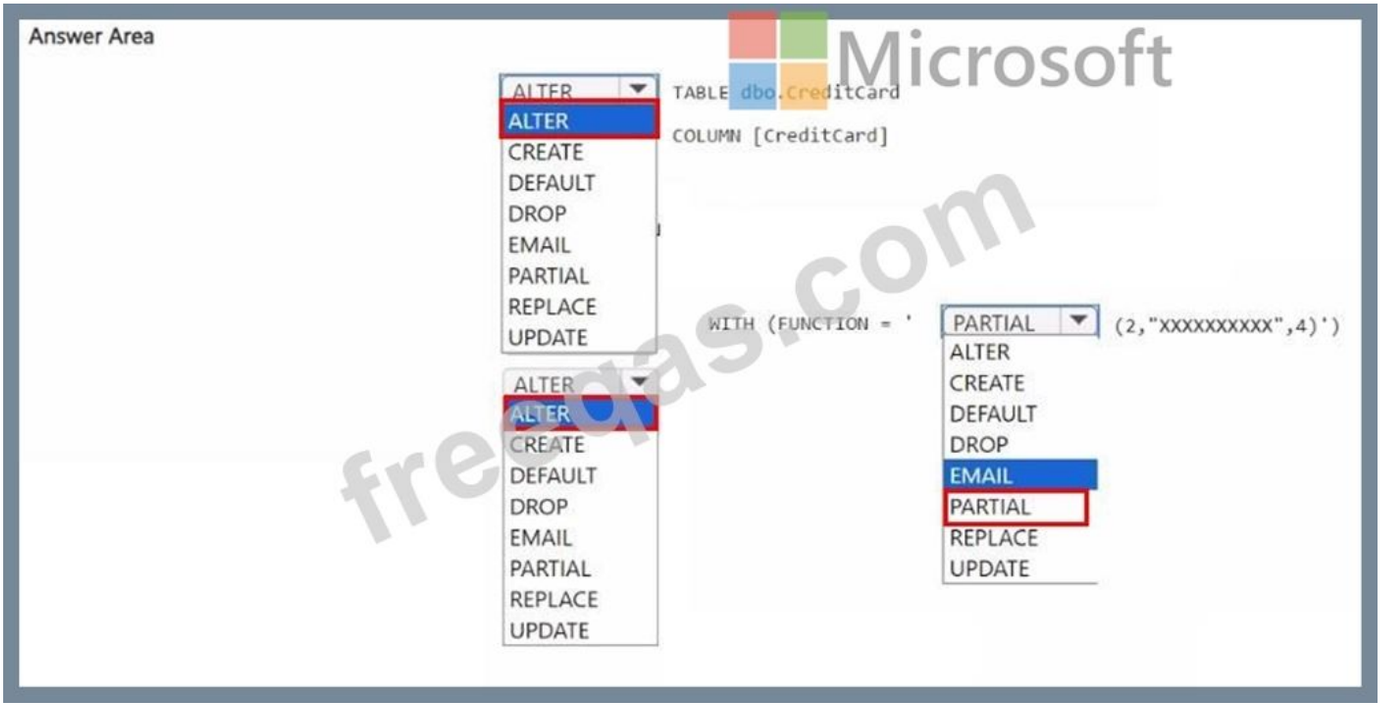
NEW QUESTION: 4
You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse2. A team of data analysts has Viewer role access to Workspace1. You create a table by running the following statement.

```
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:



NEW QUESTION: 5

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. Fault tolerance

B. Enable staging

C. Degree of copy parallelism

D. Enable logging

Answer: (SHOW ANSWER)

NEW QUESTION: 6

HOTSPOT

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 contains the following tables and columns.

Table name	Column name	Data type
Employee	EmployeeID	Int
Employee	EmployeeName	Varchar(128)
Employee	EmployeePosition	Varchar(64)
Contract	EmployeeID	Int
Contract	ContractType	Varchar(64)
Contract	StartDate	Datetime2
Contract	EndDate	Datetime2

You need to denormalize the tables and include the ContractType and StartDate columns in the Employee table. The solution must meet the following requirements:

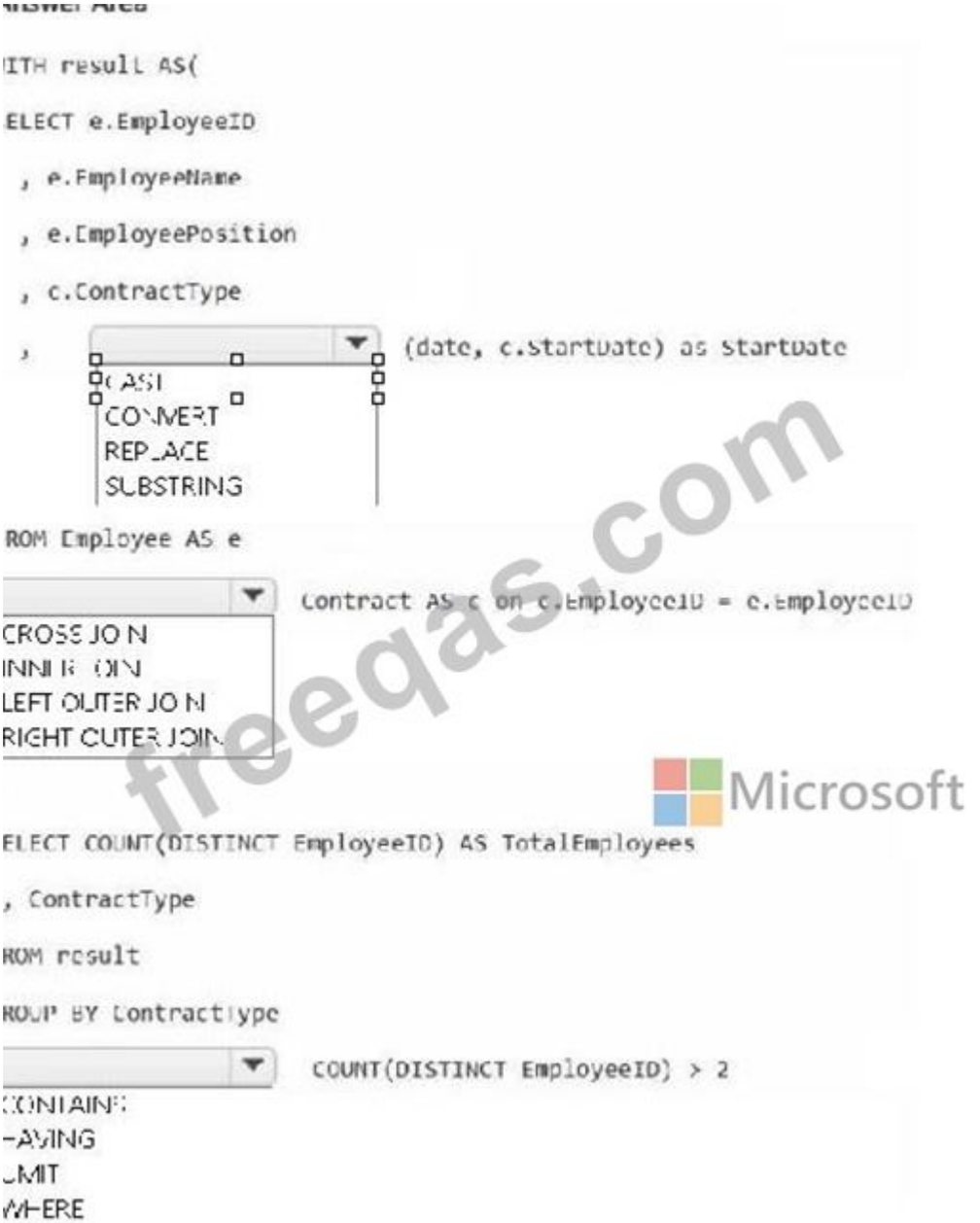
Ensure that the StartDate column is of the date data type.

Ensure that all the rows from the Employee table are preserved and include any matching rows from the Contract table.

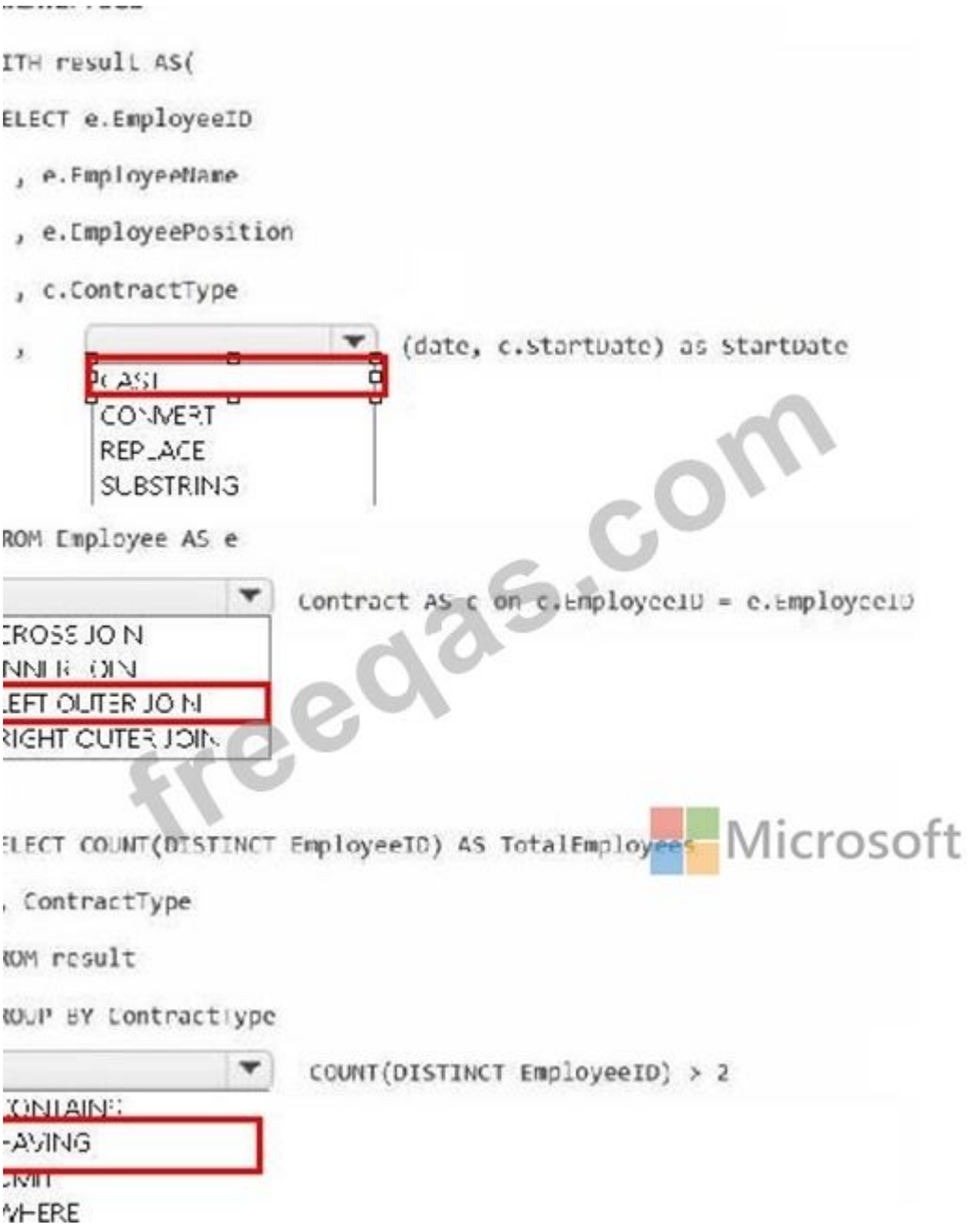
Ensure that the result set displays the total number of employees per contract type for all the contract types that have more than two employees.

How should you complete the statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.


ITH result AS(
ELECT e.EmployeeID
, e.EmployeeName
, e.EmployeePosition
, c.ContractType
, (date, c.StartDate) as StartDate
CAST
CONVERT
REPLACE
SUBSTRING
ROM Employee AS e
Contract AS c on c.EmployeeID = e.EmployeeID
CROSS JOIN
INNER JOIN
LEFT OUTER JOIN
RIGHT OUTER JOIN
ELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees
, ContractType
ROM result
ROUP BY ContractType
COUNT(DISTINCT EmployeeID) > 2
CONTAINS
HAVING
IMIT
HERE

Answer:



NEW QUESTION: 7

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 across the attributes in the DimCustomer table.

B. a direct attributes comparison for the attributes in the source table.

C. a hash function to compare the attributes in the source table.

D. a hash function to compare the attributes in the DimCustomer table.

Answer: C (LEAVE A REPLY)

NEW QUESTION: 8

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 move the filter to line 02.

Does this meet the goal?

A. Yes

B. No

Answer: A (LEAVE A REPLY)

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: 9

You have an Azure key vault named KeyVaultl that contains secrets.

You have a Fabric workspace named Workspace!. Workspace! contains a notebook named Notebookl that performs the following tasks:

* Loads stage data to the target tables in a lakehouse
* Triggers the refresh of a semantic model

You plan to add functionality to Notebookl that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVaultl to generate the authentication token. Solution: You use the following code segment:
Use notebookutils. credentials.getSecret and specify key vault URL and the name of a linked service.
Does this meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)

NEW QUESTION: 10
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:
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 (LEAVE A REPLY)
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: 11
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:

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 (LEAVE A REPLY)

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: 12

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

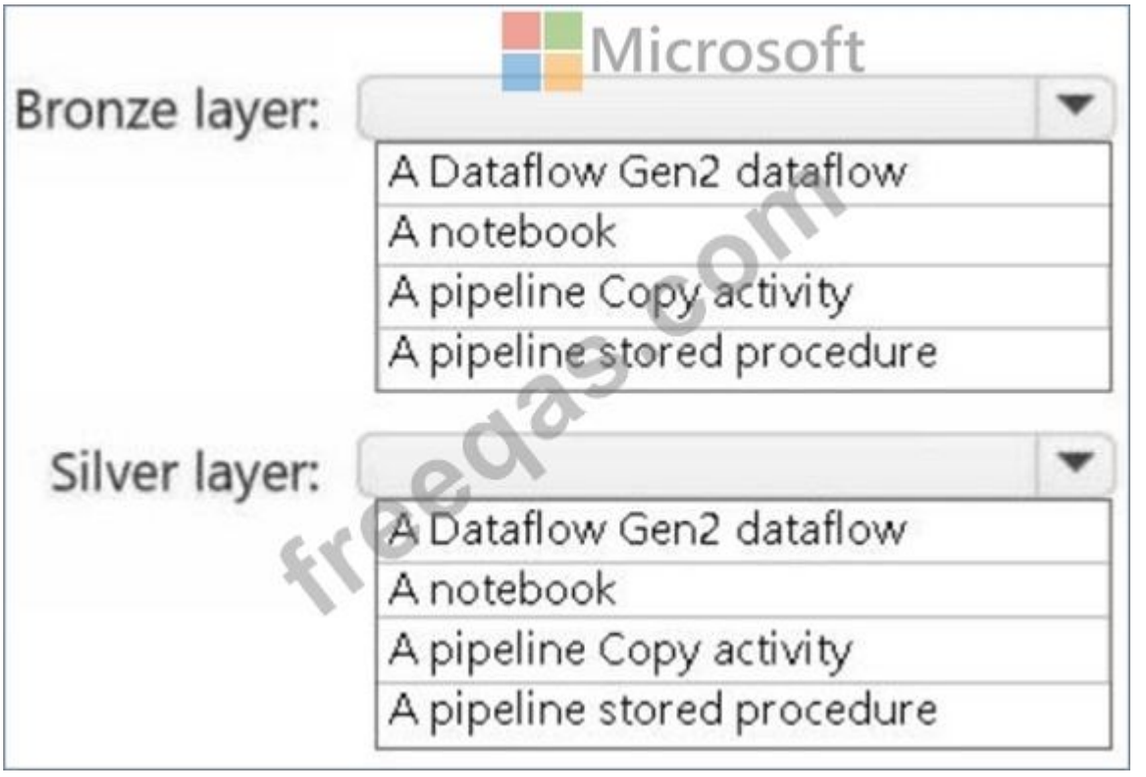
Answer: B (LEAVE A REPLY)

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

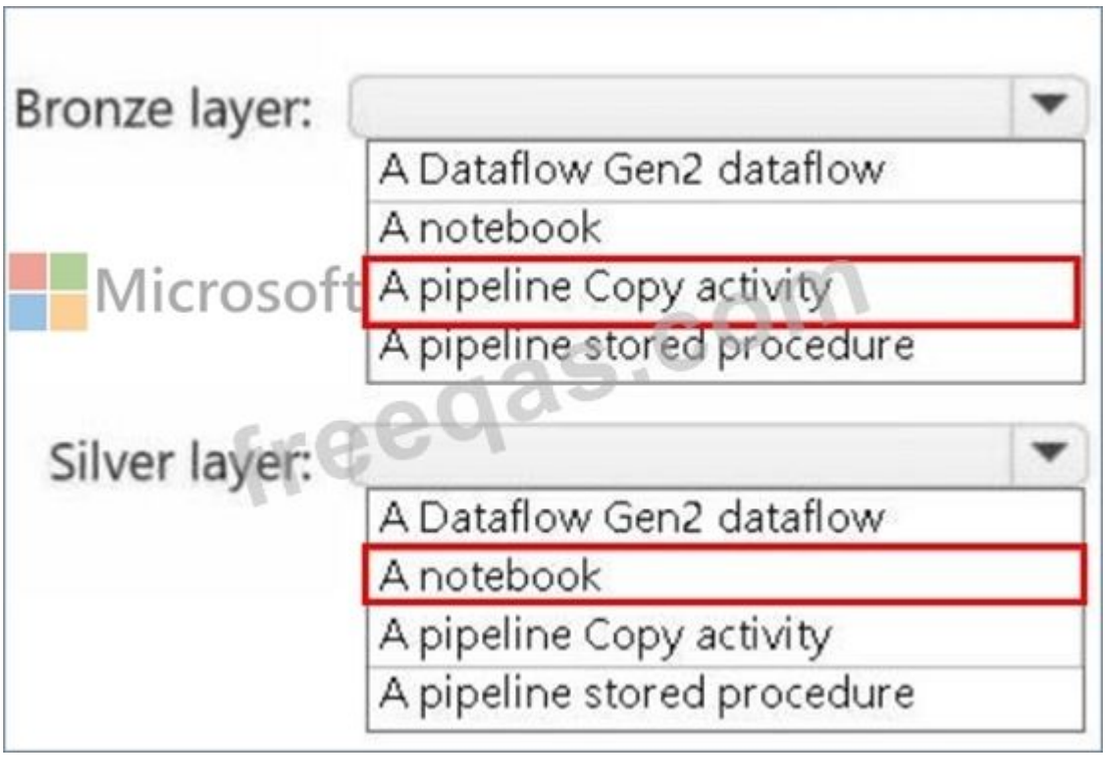
an on-premises database and requires the use of a data gateway, a data pipeline provides the appropriate framework to facilitate this data movement efficiently and reliably.

NEW QUESTION: 13
What should you do to optimize the query experience for the business users?
A. Run the VACUUM command.
B. Introduce primary keys.
C. Enable V-Order.
D. Create and update statistics.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 14
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.



Answer:



NEW QUESTION: 15

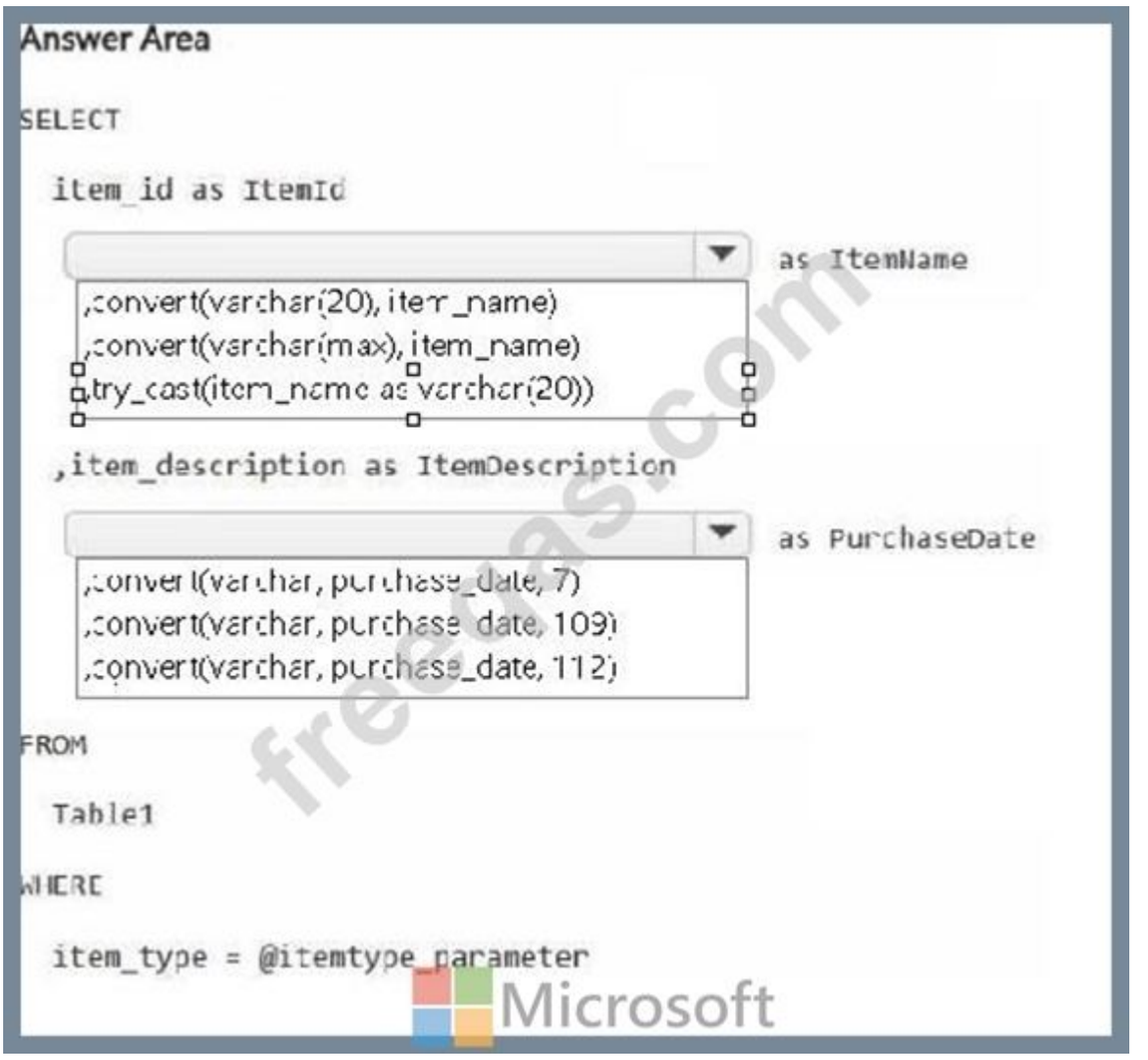
You have a Fabric workspace.

You are debugging a statement and discover the following issues:

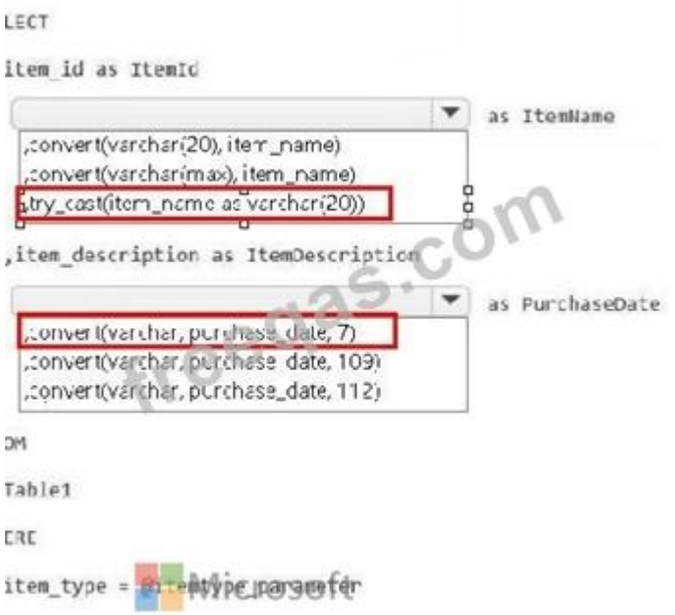
You need to resolve the issues. The solution must ensure that the data types of the results are retained. The results can contain blank cells.

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:



NEW QUESTION: 16

You have a Fabric workspace that contains a warehouse named DW1. DW1 is loaded by using a notebook named Notebook1.

You need to identify which version of Delta was used when Notebook1 was executed.
What should you use?
A. Real-Time hub
B. OneLake data hub
C. the Admin monitoring workspace
D. Fabric Monitor
E. the Microsoft Fabric Capacity Metrics app
Answer: C (LEAVE A REPLY)
To identify the version of Delta used when Notebook1 was executed, you should use the Admin monitoring workspace. The Admin monitoring workspace allows you to track and monitor detailed information about the execution of notebooks and jobs, including the underlying versions of Delta or other technologies used. It provides insights into execution details, including versions and configurations used during job runs, making it the most appropriate choice for identifying the Delta version used during the execution of Notebook1.

NEW QUESTION: 17
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. OPTIMIZE
B. CACHE
C. VACUUM
D. BROADCAST
Answer: A (LEAVE A REPLY)
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: 18

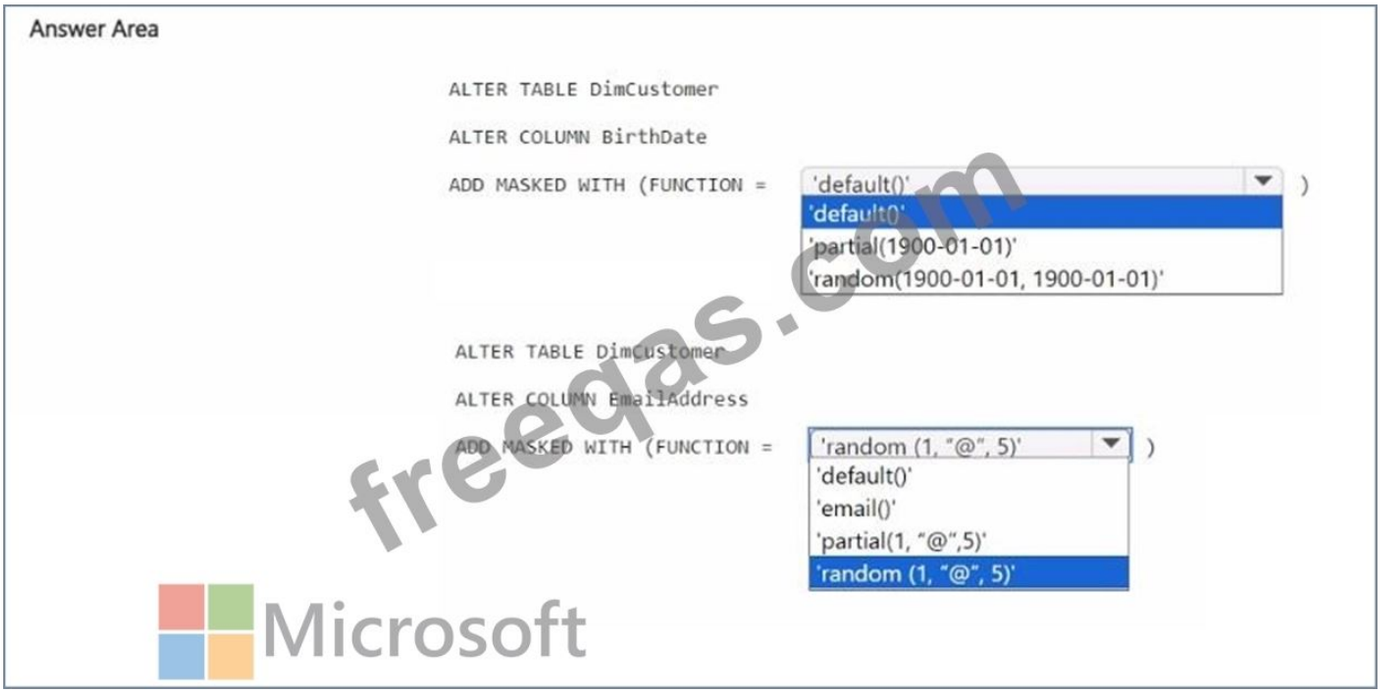
You have a Fabric workspace that contains a warehouse named Warehouse!. Warehousel contains a table named DimCustomers. DimCustomers contains the following columns:

* CustomerName
* CustomerID
* 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 are a. NOTE: Each correct selection is worth one point.



Answer:

NEW QUESTION: 19

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. the sys.dm_exec_requests dynamic management view

C. views from the queryinsights schema

D. Query activity

Answer: (SHOW ANSWER)

NEW QUESTION: 20

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: (SHOW ANSWER)

The technical requirements specify that:

Why Use a Data Pipeline That Calls Other Data Pipelines?

- 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: 21

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 (LEAVE A REPLY)
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: 22

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 (LEAVE A REPLY)
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: 23
You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse! contains a table named Customer. Customer contains the following data.

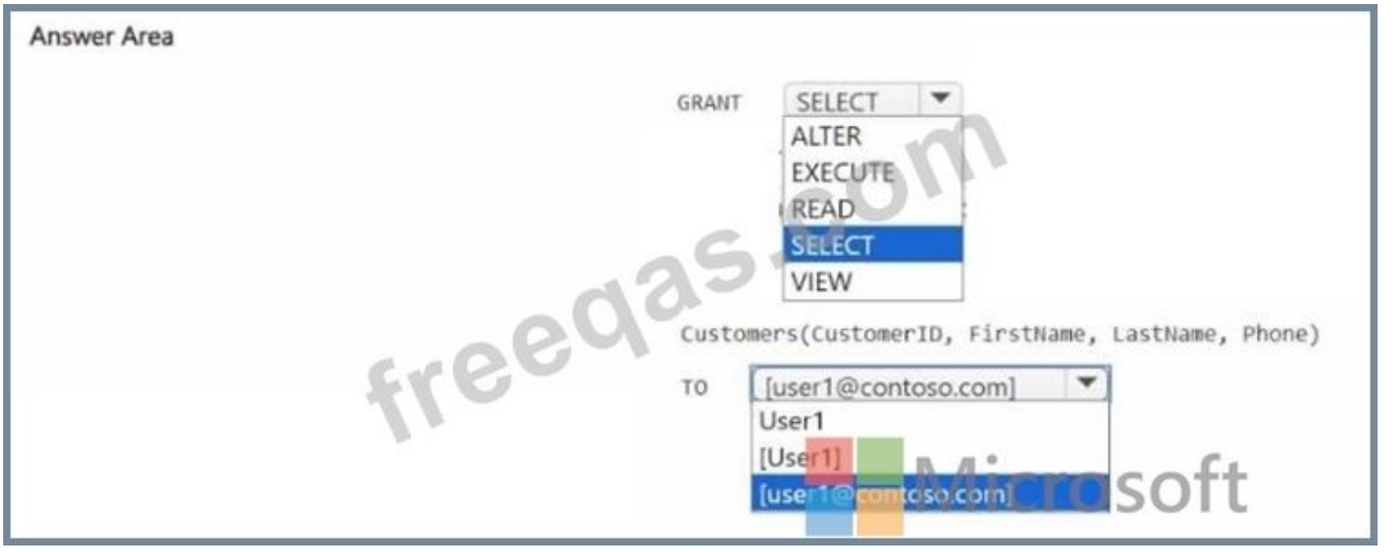
CustomerID	FirstName	LastName	Phone	CreditCard
1	John	Doe	555-123-4567	1234567812345670
2	Jane	Smith	555-987-6543	8765432187654320
3	Michael	Johnson	555-555-5555	1234987654321230
4	Emily	Davis	555-222-3333	4321123456789870
5	David	Brown	555-444-5555	5678123498761230

You have an internal Microsoft Entra user named User1 that has an email address of user1@contoso.com.

You need to provide User1 with access to the Customer table. The solution must prevent User1 from accessing the CreditCard column.

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:

NEW QUESTION: 24
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. No
B. Yes
Answer: B (LEAVE A REPLY)

NEW QUESTION: 25
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 (LEAVE A REPLY)
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: 26
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: (SHOW ANSWER)
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: 27
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 (LEAVE A REPLY)
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: 28
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 (LEAVE A REPLY)
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: 29
You have a Fabric workspace named Workspace1_DEV that contains the following items:
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:

NEW QUESTION: 30
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 (LEAVE A REPLY)
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: 31
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 (LEAVE A REPLY)
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: 32
You have a Fabric warehouse named DW1 that loads data by using a data pipeline named Pipeline1. Pipeline1 uses a Copy data activity with a dynamic SQL source. Pipeline1 is scheduled to run every 15 minutes.
You discover that Pipeline1 keeps failing.
You need to identify which SQL query was executed when the pipeline failed.
What should you do?
A. From Monitoring hub, select the latest failed run of Pipeline1, and then view the output JSON.
B. From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.
C. From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemReadFailed.
D. From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemUpdateFailed.
Answer: B (LEAVE A REPLY)
The input JSON contains the configuration details and parameters passed to the Copy data activity during execution, including the dynamically generated SQL query.
Viewing the input JSON for the failed pipeline run provides direct insight into what query was executed at the time of failure.

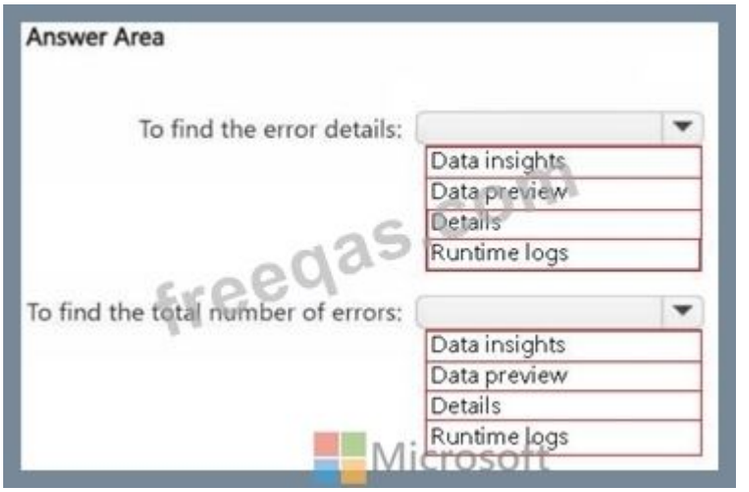
NEW QUESTION: 33
You have a Fabric workspace that contains an eventstream named Eventstream1. Eventstream1 processes data from a thermal sensor by using event stream processing, and then stores the data in a lakehouse.
You need to modify Eventstream1 to include the standard deviation of the temperature.
Which transform operator should you include in the Eventstream1 logic?
A. Expand
B. Group by
C. Union
D. Aggregate
Answer: (SHOW ANSWER)
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: 34
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.



Answer:

NEW QUESTION: 35

You have a Fabric workspace named Workspace1 that contains a warehouse named DW1 and a data pipeline named Pipeline1.

You plan to add a user named User3 to Workspace1.

You need to ensure that User3 can perform the following actions:

View all the items in Workspace1.

Update the tables in DW1.

The solution must follow the principle of least privilege.

You already assigned the appropriate object-level permissions to DW1.

Which workspace role should you assign to User3?

A. Admin
B. Member
C. Viewer
D. Contributor
Answer: D (LEAVE A REPLY)
To ensure User3 can view all items in Workspace1 and update the tables in DW1, the most appropriate workspace role to assign is the Contributor role. This role allows User3 to:
View all items in Workspace1: The Contributor role provides the ability to view all objects within the workspace, such as data pipelines, warehouses, and other resources.
Update the tables in DW1: The Contributor role allows User3 to modify or update resources within the workspace, including the tables in DW1, assuming that appropriate object-level permissions are set for the warehouse.
This role adheres to the principle of least privilege, as it provides the necessary permissions without granting broader administrative rights.

NEW QUESTION: 36
HOTSPOT
You have a Fabric workspace that contains two lakehouses named Lakehouse1 and Lakehouse2. Lakehouse1 contains staging data in a Delta table named Orderlines. Lakehouse2 contains a Type 2 slowly changing dimension (SCD) dimension table named Dim_Customer.
You need to build a query that will combine data from Orderlines and Dim_Customer to create a new fact table named Fact_Orders. The new table must meet the following requirements:
Enable the analysis of customer orders based on historical attributes.
Enable the analysis of customer orders based on the current attributes.
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
  OrderLineID order_line_id
  ,OrderDate order_date
  ,c.customer_key
  ,c.customer_id
  ,Quantity order_quantity
  ,UnitPrice unit_price
  ,TaxRate tax_rate
FROM
  Lakehouse1.orderlines o
INNER JOIN
  Lakehouse2.dim_customer c
    ON o.customerid = c.customer_id
```


AND [dropdown]

- c.is_current = 1
- o.OrderDate > c.valid_to_datetime
- o.OrderDate >= c.valid_from_datetime

AND [dropdown]

- c.is_current = 1
- o.OrderDate < c.valid_to_datetime
- o.OrderDate <= c.valid_from_datetime

Answer:

NEW QUESTION: 37
You have a Fabric warehouse named DW1. DW1 contains a table that stores sales data and is used by multiple sales representatives.
You plan to implement row-level security (RLS).
You need to ensure that the sales representatives can see only their respective data.
Which warehouse object do you require to implement RLS?
A. ISTORED PROCEDURE
B. CONSTRAINT
C. SCHEMA
D. FUNCTION
Answer: D (LEAVE A REPLY)
To implement Row-Level Security (RLS) in a Fabric warehouse, you need to use a function that defines the security logic for filtering the rows of data based on the user's identity or role. This function can be used in conjunction with a security policy to control access to specific rows in a table.
In the case of sales representatives, the function would define the filtering criteria (e.g., based on a column such as SalesRepID or SalesRepName), ensuring that each representative can only see their respective data.

NEW QUESTION: 38
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 (LEAVE A REPLY)

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: 39

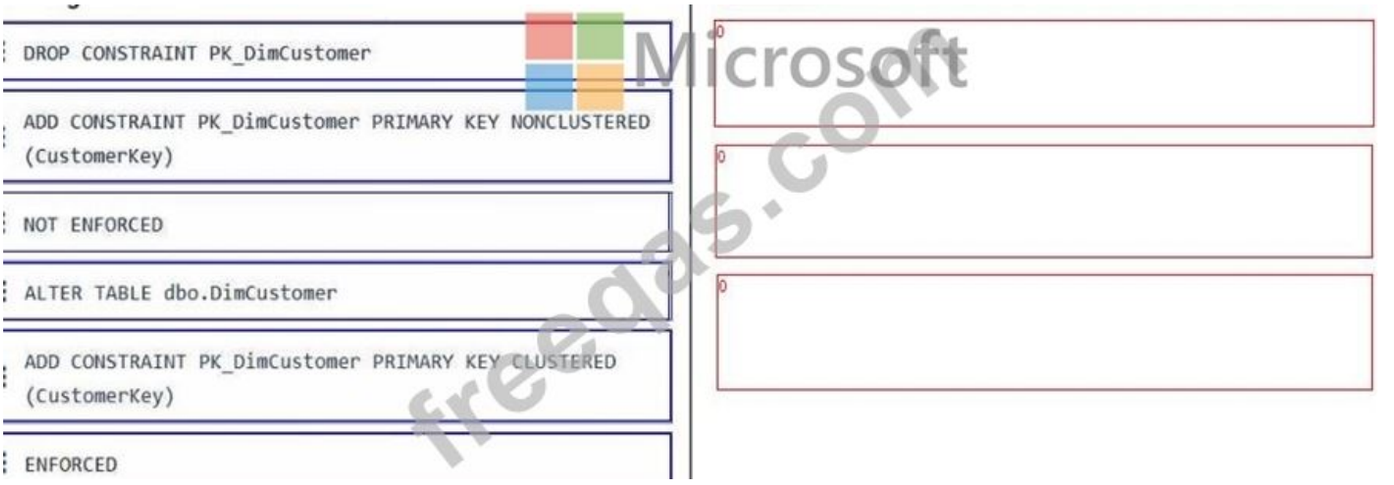
You have a Fabric workspace that contains a warehouse named Warehouse1.

In Warehouse1, you create a table named DimCustomer by running the following statement.

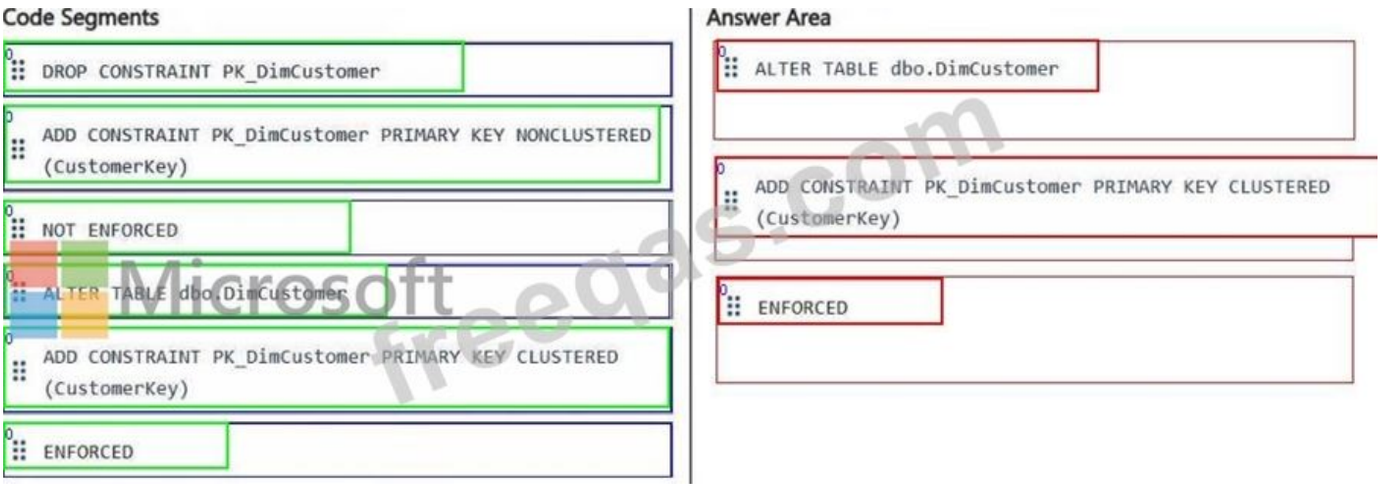
```
CREATE TABLE dbo.DimCustomer (
  CustomerKey VARCHAR(255) NOT NULL,
  Name VARCHAR(255) NOT NULL,
  Email VARCHAR(255) NOT NULL
  );
```

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.



Answer:



NEW QUESTION: 40

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 (LEAVE A REPLY)

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.

Topic 1, Litware, Inc

Overview

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment. Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

Existing Environment. Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment. Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment. Security Groups

Litware has the following security groups:

Existing Environment. Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the

reports fail to load: "The SQL query failed while running." The data engineering team wants to debug the issue and find queries that cause more than one failure.
When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.
The company's sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.
Requirements. Planned Changes
Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.
Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.
Requirements. Version Control
Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.
Requirements. Governance Requirements
To control data platform costs, the data platform must use only Fabric services and items.
Additional Azure resources must NOT be provisioned.
Requirements. Data Requirements
Litware identifies the following data requirements:

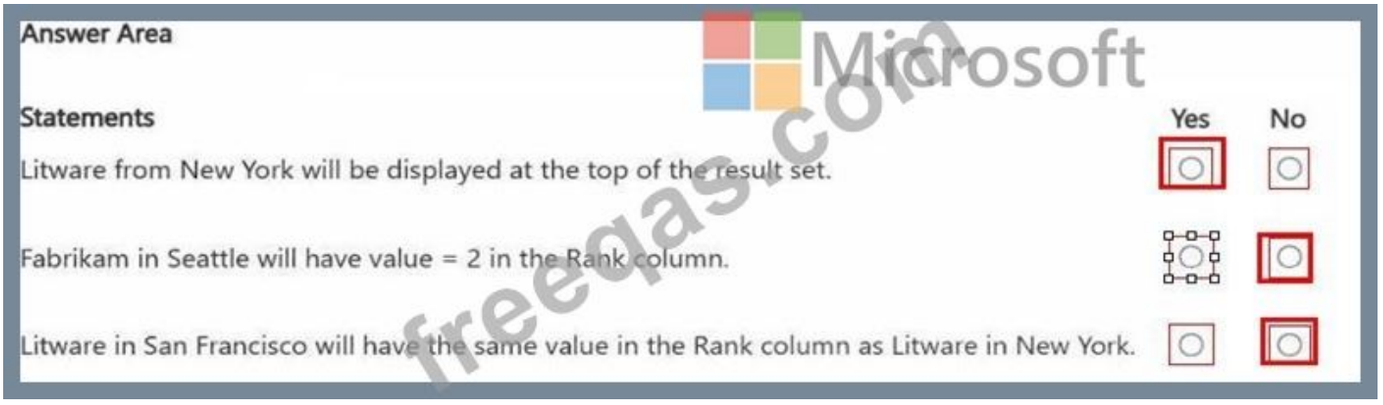
NEW QUESTION: 41
HOTSPOT
You are processing streaming data from an external data provider.
You have the following code segment.

```
"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
```

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:



NEW QUESTION: 42

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. Event stream

B. Dataflow Gen2

C. Streaming dataset

D. Data pipeline

Answer: (SHOW ANSWER)

To ingest large files (500 GB each) from an external data source into Lakehouse1 with high throughput and to trigger the process when a new file is added, an Eventstream is the best solution.

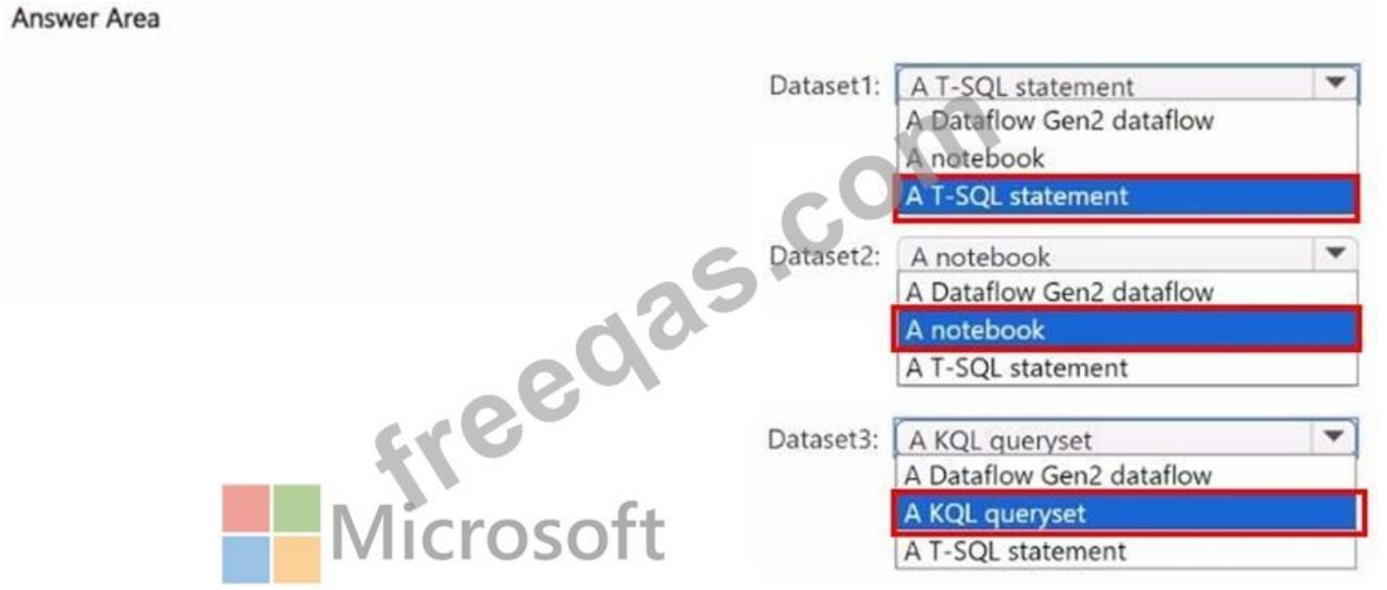
An Eventstream in Fabric is designed for handling real-time data streams and can efficiently ingest large files as soon as they are added to an external source. It is optimized for high throughput and can be configured to trigger upon detecting new files, allowing for fast and continuous ingestion of data with minimal delay.

NEW QUESTION: 43

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: (SHOW ANSWER)
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.

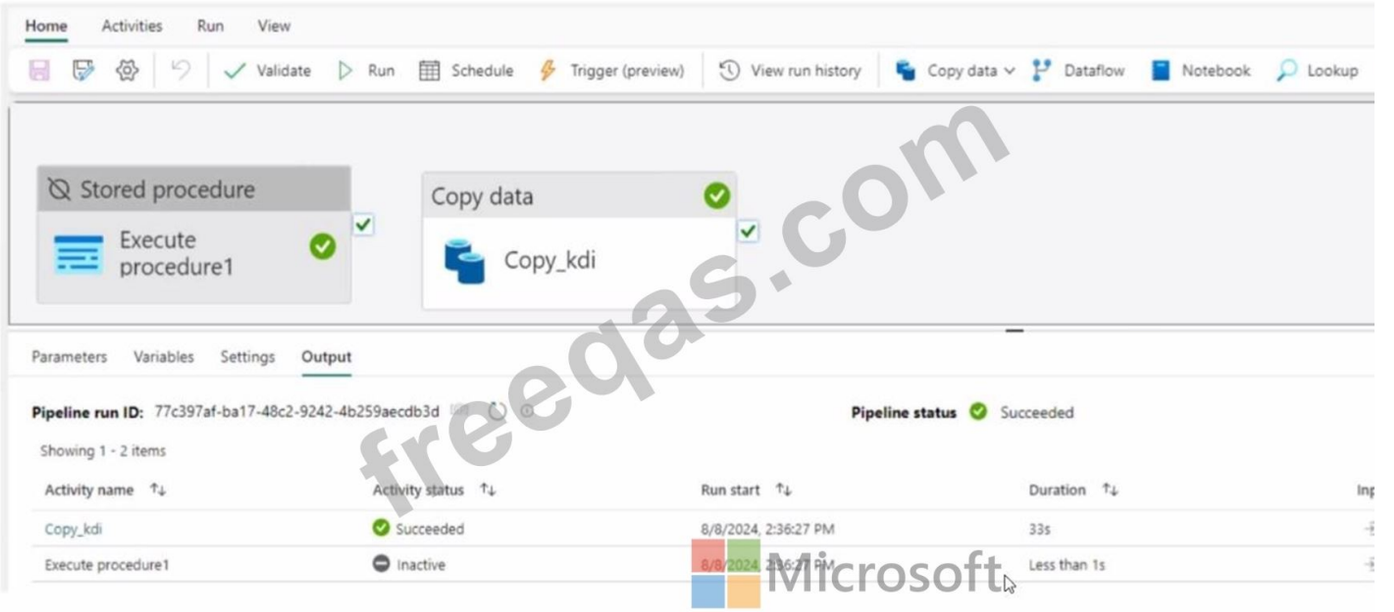
NEW QUESTION: 44
You plan to process the following three datasets by using Fabric:
* Dataset1: This dataset will be added to Fabric and will have a unique primary key between the source and the destination. The unique primary key will be an integer and will start from 1 and have an increment of 1.
* 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:



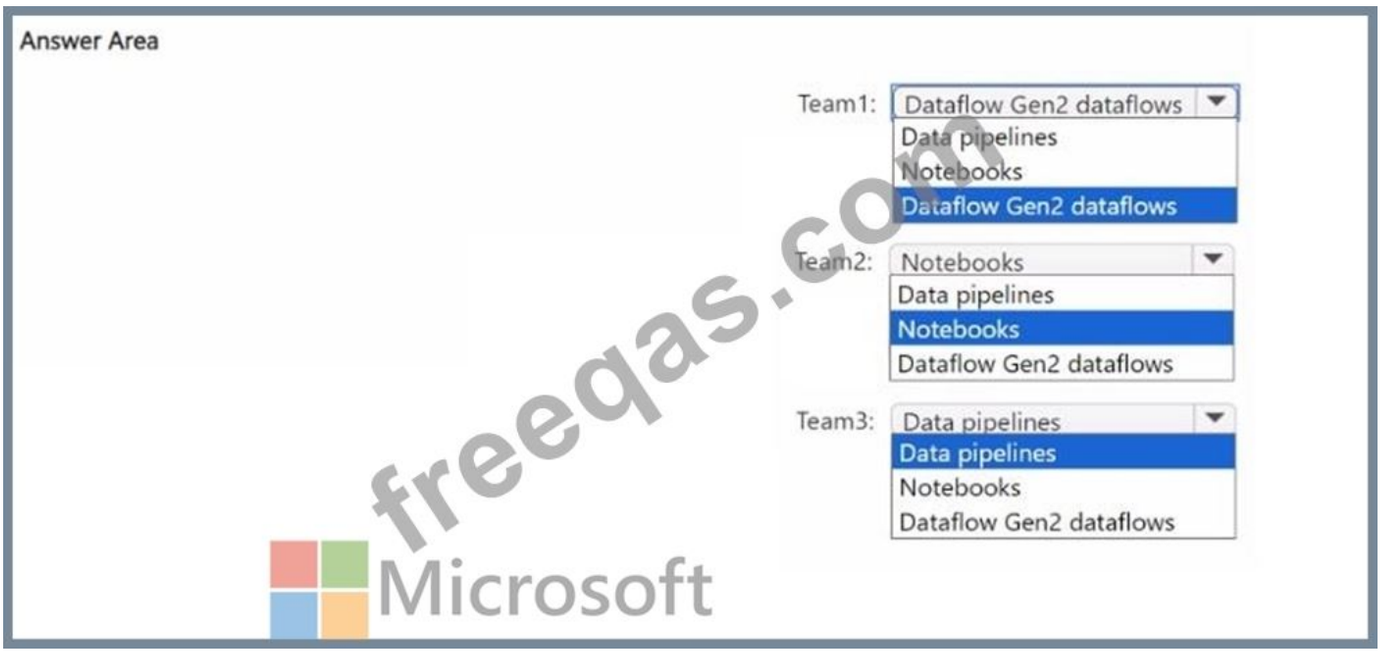
NEW QUESTION: 45

You have a Fabric workspace that contains a data pipeline named Pipeline! as shown in the exhibit.



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

NEW QUESTION: 46
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:

Valid DP-700 Dumps shared by PrepPdf.com for Helping Passing DP-700 Exam! PrepPdf.com now offer the **newest DP-700 exam dumps**, the PrepPdf.com DP-700 exam