



Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

NEW QUESTION 1

- (Topic 1)

You need to ensure that usage of the data in the Amazon S3 bucket meets the technical requirements. What should you do?

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

Answer: B

Explanation:

To ensure that the usage of the data in the Amazon S3 bucket meets the technical requirements, we must address two key points:

Minimize egress costs associated with cross-cloud data access: Using a shortcut ensures that Fabric does not replicate the data from the S3 bucket into the lakehouse but rather provides direct access to the data in its original location. This minimizes cross-cloud data transfer and avoids additional egress costs.

Prevent saving a copy of the raw data in the lakehouses: Disabling caching ensures that the raw data is not copied or persisted in the Fabric workspace. The data is accessed on-demand directly from the Amazon S3 bucket.

NEW QUESTION 2

- (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 3

- (Topic 3)

You have an Azure event hub. Each event contains the following fields: BikepointID

Street Neighbourhood

Latitude Longitude No_Bikes No_Empty_Docks

You need to ingest the events. The solution must only retain events that have a Neighbourhood value of Chelsea, and then store the retained events in a Fabric lakehouse.

What should you use?

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

Answer: B

Explanation:

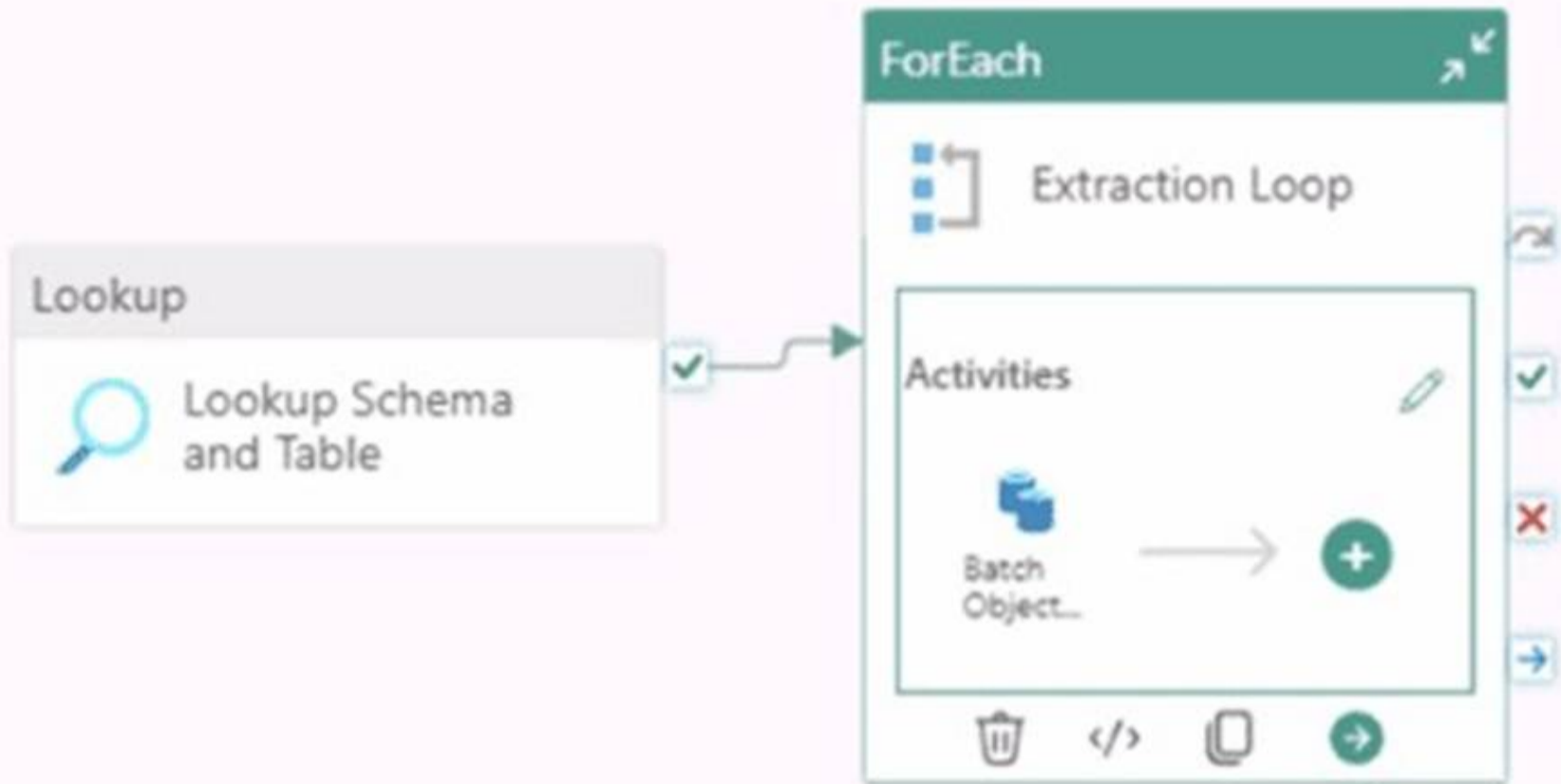
An eventstream is the best solution for ingesting data from Azure Event Hub into Fabric, while applying filtering logic such as retaining only the events that have a Neighbourhood value of "Chelsea." Eventstreams in Microsoft Fabric are designed for handling real-time data streams and can apply transformation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

Eventstreams are well-suited for stream processing, such as this case where you need to filter out only specific data (events with a Neighbourhood of "Chelsea") before storing it in the lakehouse.

NEW QUESTION 4

HOTSPOT - (Topic 3)

You are building a data orchestration pattern by using a Fabric data pipeline named Dynamic Data Copy as shown in the exhibit. (Click the Exhibit tab.)



General Settings Activities (1)

Batch count ⓘ

Items *

This property should be parameterized.

Add dynamic content [Alt+Shift+D]

Dynamic Data Copy does NOT use parametrization.

You need to configure the ForEach activity to receive the list of tables to be copied. How should you complete the pipeline expression? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

@activity('Lookup Schema and Table',
 Lookup Schema and Table
 Batch Object Copy
 Dynamic Data Copy
 Extraction Loop
 Lookup Schema and Table
).
 output.value
 output
 output.count
 output.pipelineReturnValue
 output.value

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 5

- (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 6

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. | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| User3 has Domain contributor access to Domain1. | <input type="checkbox"/> | <input type="checkbox"/> |
| User2 has Contributor role access to Workspace3. | <input type="checkbox"/> | <input type="checkbox"/> |

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Answer Area

Statements	Yes	No
User3 has Viewer role access to Workspace3.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
User3 has Domain contributor access to Domain1.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
User2 has Contributor role access to Workspace3.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

NEW QUESTION 7

- (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse1. Data is loaded daily into Warehouse1 by using data pipelines and stored procedures.

You discover that the daily data load takes longer than expected.

You need to monitor Warehouse1 to identify the names of users that are actively running queries.

Which view should you use?

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

Answer: E

Explanation:

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

NEW QUESTION 8

- (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-59feade642f6", "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 9

- (Topic 3)

You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse1.

You plan to deploy Warehouse1 to a new workspace named Workspace2.

As part of the deployment process, you need to verify whether Warehouse1 contains invalid references. The solution must minimize development effort. What should you use?

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

Answer: C

Explanation:

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

NEW QUESTION 10

- (Topic 3)

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

- A. Enable high concurrency for notebooks.
- B. Enable dynamic allocation for the Spark pool.
- C. Change the runtime version.
- D. Increase the number of executors.

Answer: A

Explanation:

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

NEW QUESTION 10

- (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 11

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

ALTER TABLE dbo.CreditCard
 COLUMN [CreditCard]
 WITH (FUNCTION = 'PARTIAL' (2, "XXXXXXXXXX",4)')

ALTER
 ALTER
 CREATE
 DEFAULT
 DROP
 EMAIL
 PARTIAL
 REPLACE
 UPDATE

ALTER
 ALTER
 CREATE
 DEFAULT
 DROP
 EMAIL
 PARTIAL
 REPLACE
 UPDATE

PARTIAL
 ALTER
 CREATE
 DEFAULT
 DROP
 EMAIL
 PARTIAL
 REPLACE
 UPDATE

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

ALTER TABLE dbo.CreditCard
 COLUMN [CreditCard]
 WITH (FUNCTION = 'PARTIAL' (2, "XXXXXXXXXX",4)')

ALTER
 ALTER
 CREATE
 DEFAULT
 DROP
 EMAIL
 PARTIAL
 REPLACE
 UPDATE

ALTER
 ALTER
 CREATE
 DEFAULT
 DROP
 EMAIL
 PARTIAL
 REPLACE
 UPDATE

PARTIAL
 ALTER
 CREATE
 DEFAULT
 DROP
 EMAIL
 PARTIAL
 REPLACE
 UPDATE

NEW QUESTION 12

HOTSPOT - (Topic 3)

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 
  o.OrderDate > c.valid_to_datetime
  o.OrderDate >= c.valid_from_datetime

AND 
  o.OrderDate < c.valid_to_datetime
  o.OrderDate <= c.valid_from_datetime
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

c.is_current = 1

o.OrderDate <= valid_c_datetime

o.OrderDate >= cval c_from_datetime

AND

c.is_current = 1

o.OrderDate <= valid_c_datetime

o.OrderDate <= cval c_from_datetime

NEW QUESTION 13

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

```

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

```

as ItemName

,item_description as ItemDescription

```

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

```

as PurchaseDate

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 18

DRAG DROP - (Topic 3)

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 to 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.

Badges

- Certified
- Master data
- Promoted
- Cannot be endorsed

Answer Area

Entity1:

Entity2:

Entity3:

Entity4:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Badges



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 Fabric eventstream that loads data into a table named Bike_Location in a KQL database. The table contains the following columns:

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp

You need to apply transformation and filter logic to prepare the data for consumption. The solution must return data for a neighbourhood named Sands End when No_Bikes is at least 15. The results must be ordered by No_Bikes in ascending order.

Solution: You use the following code segment:

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

Does this meet the goal?

- A. Yes
- B. no

Answer: A

Explanation:

Filter Condition: It correctly filters rows where Neighbourhood is "Sands End" and No_Bikes is greater than or equal to 15.

Sorting: The sorting is explicitly done by No_Bikes in ascending order using sort by No_Bikes asc.

Projection: It projects the required columns (BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp), which minimizes the data returned for consumption.

NEW QUESTION 24

- (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 25

HOTSPOT - (Topic 3)

You need to recommend a Fabric streaming solution that will use the sources shown in the following table.

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

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

Answer Area

Source1:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline
- A streaming dataflow**
- An eventstream

Source2:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline**
- A streaming dataflow

Source3:

- Apache Spark Structured Streaming
- An eventstream**
- A data pipeline
- A streaming dataflow

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Source1:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline
- A streaming dataflow**
- An eventstream

Source2:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline**
- A streaming dataflow

Source3:

- Apache Spark Structured Streaming
- An eventstream**
- A data pipeline
- A streaming dataflow

NEW QUESTION 27

- (Topic 3)

You are developing a data pipeline named Pipeline1.

You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse. Which option from the Settings tab of the Copy data activity must you configure?

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

Answer: C

NEW QUESTION 30

- (Topic 3)

You have a Fabric notebook named Notebook1 that has been executing successfully for the last week.

During the last run, Notebook1executed nine jobs. You need to view the jobs in a timeline chart. What should you use?

- A. Real-Time hub
- B. Monitoring hub
- C. the job history from the application run
- D. Spark History Server
- E. the run series from the details of the application run

Answer: E

Explanation:

The run series from the details of the application run is the most detailed and relevant feature for visualizing job execution in a timeline format, making it the correct choice for this scenario. It provides an intuitive way to analyze job execution patterns and improve the efficiency of the notebook.

NEW QUESTION 31

- (Topic 3)

You have a Fabric F32 capacity that contains a workspace. The workspace contains a warehouse named DW1 that is modelled by using MD5 hash surrogate keys.

DW1 contains a single fact table that has grown from 200 million rows to 500 million rows during the past year.

You have Microsoft Power BI reports that are based on Direct Lake. The reports show year-over-year values.

Users report that the performance of some of the reports has degraded over time and some visuals show errors.

You need to resolve the performance issues. The solution must meet the following requirements:

Provide the best query performance. Minimize operational costs.

Which should you do?

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

Answer: D

Explanation:

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

NEW QUESTION 33

- (Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a 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 35

- (Topic 3)

You have a Fabric capacity that contains a workspace named Workspace1. Workspace1 contains a lakehouse named Lakehouse1, a data pipeline, a notebook, and several Microsoft Power BI reports.

A user named User1 wants to use SQL to analyze the data in Lakehouse1. You need to configure access for User1. The solution must meet the following requirements:

Provide User1 with read access to the table data in Lakehouse1.

Prevent User1 from using Apache Spark to query the underlying files in Lakehouse1. Prevent User1 from accessing other items in Workspace1.

What should you do?

- A. Share Lakehouse1 with User1 directly and select Read all SQL endpoint data.

- B. Assign User1 the Viewer role for Workspace1. Share Lakehouse1 with User1 and select Read all SQL endpoint data.
- C. Share Lakehouse1 with User1 directly and select Build reports on the default semantic model.
- D. Assign User1 the Member role for Workspace1. Share Lakehouse1 with User1 and select Read all SQL endpoint data.

Answer: B

Explanation:

To meet the specified requirements for User1, the solution must ensure:

? Read access to the table data in Lakehouse1: User1 needs permission to access the data within Lakehouse1. By sharing Lakehouse1 with User1 and selecting the Read all SQL endpoint data option, User1 will be able to query the data via SQL endpoints.

? Prevent Apache Spark usage: By sharing the lakehouse directly and selecting the SQL endpoint data option, you specifically enable SQL-based access to the data, preventing User1 from using Apache Spark to query the data.

? Prevent access to other items in Workspace1: Assigning User1 the Viewer role for Workspace1 ensures that User1 can only view the shared items (in this case, Lakehouse1), without accessing other resources such as notebooks, pipelines, or Power BI reports within Workspace1.

This approach provides the appropriate level of access while restricting User1 to only the required resources and preventing access to other workspace assets.

NEW QUESTION 39

- (Topic 3)

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

Explanation:

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 44

- (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 49

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 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 area. NOTE: Each correct selection is worth one point.

Answer Area

ALTER TABLE DimCustomer

ALTER COLUMN BirthDate

ADD MASKED WITH (FUNCTION =

'default()'	▼)
'default()'		
'partial(1900-01-01)'		
'random(1900-01-01, 1900-01-01)'		

ALTER TABLE DimCustomer

ALTER COLUMN EmailAddress

ADD MASKED WITH (FUNCTION =

'random (1, "@", 5)'	▼)
'default()'		
'email()'		
'partial(1, "@",5)'		
'random (1, "@", 5)'		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

ALTER TABLE DimCustomer

ALTER COLUMN BirthDate

ADD MASKED WITH (FUNCTION =

'default()'	▼)
'default()'		
'partial(1900-01-01)'		
'random(1900-01-01, 1900-01-01)'		

ALTER TABLE DimCustomer

ALTER COLUMN EmailAddress

ADD MASKED WITH (FUNCTION =

'random (1, "@", 5)'	▼)
'default()'		
'email()'		
'partial(1, "@",5)'		
'random (1, "@", 5)'		

NEW QUESTION 50

- (Topic 3)

You have an Azure key vault named KeyVault1 that contains secrets.

You have a Fabric workspace named Workspace!. Workspace! contains a notebook named Notebook1 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 Notebook1 that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token. Solution: You use the following code segment:

Use notebookutils.credentials.getSecret and specify key vault URL and the name of a linked service.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 54

- (Topic 3)

You have a Fabric workspace named Workspace1 that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:

Orders

Customer Employee

The Employee table contains Personally Identifiable Information (PII).

A data engineer is building a workflow that requires writing data to the Customer table, however, the user does NOT have the elevated permissions required to view the contents of the Employee table.

You need to ensure that the data engineer can write data to the Customer table without reading data from the Employee table.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

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

Answer: ADE

Explanation:

To meet the requirements of ensuring that the data engineer can write data to the Customer table without reading data from the Employee table (which contains Personally Identifiable Information, or PII), you can implement the following steps:

? Share Lakehouse1 with the data engineer.

By sharing Lakehouse1 with the data engineer, you provide the necessary access to the data within the lakehouse. However, this access should be controlled through roles and permissions, which will allow writing to the Customer table but prevent reading from the Employee table.

? Assign the data engineer the Contributor role for Workspace1.

Assigning the Contributor role for Workspace1 grants the data engineer the ability to perform actions such as writing to tables (e.g., the Customer table) within the workspace. This role typically allows users to modify and manage data without necessarily granting them access to view all data (e.g., PII data in the Employee table).

? Migrate the Employee table from Lakehouse1 to Lakehouse2.

To prevent the data engineer from accessing the Employee table (which contains PII), you can migrate the Employee table to a separate lakehouse (Lakehouse2) or workspace

(Workspace2). This separation of sensitive data ensures that the data engineer's access is restricted to the Customer table in Lakehouse1, while the Employee table can be managed separately and protected under different access controls.

NEW QUESTION 59

- (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 61

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.	<input type="radio"/>	<input type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>

- 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.	<input type="radio"/>	<input checked="" type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input checked="" type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="radio"/>

NEW QUESTION 64

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.

Answer Area

Data processor:

- A data pipeline
- A Dataflow Gen2 dataflow
- An eventstream with a custom endpoint
- An eventstream with an external data source

Filtering:

- A Filter activity in a data pipeline
- A filter in a Dataflow Gen2 dataflow
- A KQL statement
- An eventstream processor

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Data processor:

- A data pipeline
- A Dataflow Gen2 dataflow
- An eventstream with a custom endpoint
- An eventstream with an external data source

Filtering:

- A Filter activity in a data pipeline
- A filter in a Dataflow Gen2 dataflow
- A KQL statement
- An eventstream processor

NEW QUESTION 68

- (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 72

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a warehouse named DW1. DW1 contains the following tables and columns.

Table name	Column name	Description
SalesOrderDetail	ProductID	Contains the product ID of the ordered product
SalesOrderDetail	ModifiedDate	Contains the date of an order
SalesOrderDetail	OrderQty	Contains the order quantity
Product	ProductID	Contains the unique ID of a product
Product	Name	Contains a product name

You need to create an output that presents the summarized values of all the order quantities by year and product. The results must include a summary of the order quantities at the year level for all the products.

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

NOTE: Each correct selection is worth one point.

Answer Area

```

(SO.ModifiedDate) AS OrderDate
SELECT CAST
SELECT CONVERT
SELECT YEAR
,P.Name AS ProductName
,SUM(SO.OrderQty) AS OrderQty
FROM [dbo].[SalesOrderDetail] SO
INNER JOIN [dbo].[Product] P
ON P.ProductID = SO.ProductID
GROUP BY
CLUBE(YEAR(SO.ModifiedDate), P.Name)
(ROLLUP(YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
ROLLUP(YEAR(SO.ModifiedDate), P.Name)
YEAR(SO.ModifiedDate), P.Name
ORDER BY OrderDate

```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
 Answer Area

```

(SO.ModifiedDate) AS OrderDate
SELECT CAST
SELECT CONVERT
SELECT YEAR
,P.Name AS ProductName
,SUM(SO.OrderQty) AS OrderQty
FROM [dbo].[SalesOrderDetail] SO
INNER JOIN [dbo].[Product] P
ON P.ProductID = SO.ProductID
GROUP BY
CLUBE(YEAR(SO.ModifiedDate), P.Name)
(ROLLUP(YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
ROLLUP(YEAR(SO.ModifiedDate), P.Name)
YEAR(SO.ModifiedDate), P.Name
ORDER BY OrderDate

```

NEW QUESTION 74

- (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 79

- (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 82

- (Topic 3)

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

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

Answer: B

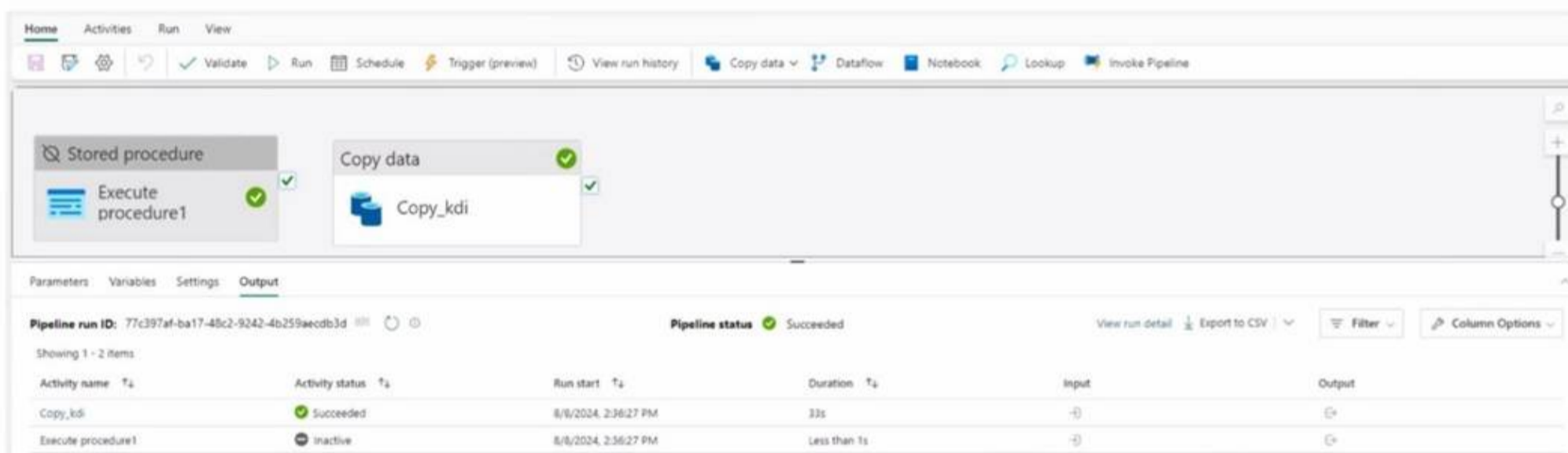
Explanation:

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

NEW QUESTION 86

- (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 91

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

Litware from New York will be displayed at the top of the result set.

Yes	No
<input type="checkbox"/>	<input type="checkbox"/>

Fabrikam in Seattle will have value = 2 in the Rank column.

<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------

Litware in San Francisco will have the same value in the Rank column as Litware in New York.

<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------

- 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 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 KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

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

You need to reduce how long it takes to run the KQL queryset. Solution: You move the filter to line 02.

```

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

Does this meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

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

NEW QUESTION 97

- (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 99

.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questions 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

100% Actual & Verified — Instant Download, Please Click
[Order The DP-700 Practice Test Here](#)