

Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)



NEW QUESTION 1

HOTSPOT - (Topic 1)

You need to create the product dimension.

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

NOTE: Each correct selection is worth one point.

```
SELECT ProductID, ProductNumber, ProductName, ModelName, SubCategoryName, CategoryName
FROM ContosoLake.Products p
    ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID
    ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID
WHERE
```

▼

FULL JOIN

INNER JOIN

LEFT ANTI JOIN

LEFT OUTER JOIN

OUTER JOIN

▼

FULL JOIN

INNER JOIN

LEFT ANTI JOIN

LEFT OUTER JOIN

OUTER JOIN

▼

CategoryID = 1;

CategoryName is not null;

IsActive = 1;

IsActive is not null;

ProductNumber is not null;

SubCategoryID = 1;

SubCategoryName is not null;

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Join between Products and ProductSubCategories: Use an INNER JOIN.

The goal is to include only products that are assigned to a subcategory. An INNER JOIN ensures that only matching records (i.e., products with a valid subcategory) are included.

Join between ProductSubCategories and ProductCategories: Use an INNER JOIN.

Similar to the above logic, we want to include only subcategories assigned to a valid product category. An INNER JOIN ensures this condition is met.

WHERE Clause Condition: IsActive = 1

Only active products (where IsActive equals 1) should be included in the gold layer. This filters out inactive products.

NEW QUESTION 2

HOTSPOT - (Topic 1)

You need to recommend a method to populate the POS1 data to the lakehouse medallion layers.

What should you recommend for each layer? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Bronze layer:

	▼
A Dataflow Gen2 dataflow	
A notebook	
A pipeline Copy activity	
A pipeline stored procedure	

Silver layer:

	▼
A Dataflow Gen2 dataflow	
A notebook	
A pipeline Copy activity	
A pipeline stored procedure	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Bronze Layer: A pipeline Copy activity

The bronze layer is used to store raw, unprocessed data. The requirements specify that no transformations should be applied before landing the data in this layer. Using a pipeline Copy activity ensures minimal development effort, built-in connectors, and the ability to ingest the data directly into the Delta format in the bronze layer.

Silver Layer: A notebook

The silver layer involves extensive data cleansing (deduplication, handling missing values, and standardizing capitalization). A notebook provides the flexibility to implement complex transformations and is well-suited for this task.

NEW QUESTION 3

DRAG DROP - (Topic 2)

You need to ensure that the authors can see only their respective sales data.

How should you complete the statement? To answer, drag the appropriate values the correct targets. Each value may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content

NOTE: Each correct selection is worth one point.

Values

- AuthorSales
- AuthorEmail
- AuthorSales.AuthorEmail
- BLOCK
- FILTER
- INLINE
- SCHEMABINDING
- USER_NAME()

Answer Area

```
CREATE FUNCTION dbo.tvf_rlspredicate(@Author AS varchar(50))
    RETURNS TABLE
WITH
AS
    RETURN SELECT 1 AS tvf_rlspredicate_result
WHERE @Author =
GO

CREATE SECURITY POLICY RLSFilter
ADD FILTER PREDICATE Security.tvf_rlspredicate(AuthorEmail)
ON
WITH (STATE = ON)
```

- No
-
-
-

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Values

- AuthorSales
- AuthorEmail
- AuthorSales.AuthorEmail
- BLOCK
- FILTER
- INLINE
- SCHEMABINDING
- USER_NAME()

Answer Area

```
CREATE FUNCTION dbo.tvf_rlspredicate(@Author AS varchar(50))
    RETURNS TABLE
WITH SCHEMABINDING
AS
    RETURN SELECT 1 AS tvf_rlspredicate_result
WHERE @Author = USER_NAME()
GO

CREATE SECURITY POLICY RLSFilter
ADD FILTER PREDICATE Security.tvf_rlspredicate(AuthorEmail)
ON AuthorSales
WITH (STATE = ON)
```

- No
-
-
-

NEW QUESTION 4

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

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

Answer: E

Explanation:

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

NEW QUESTION 5

- (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

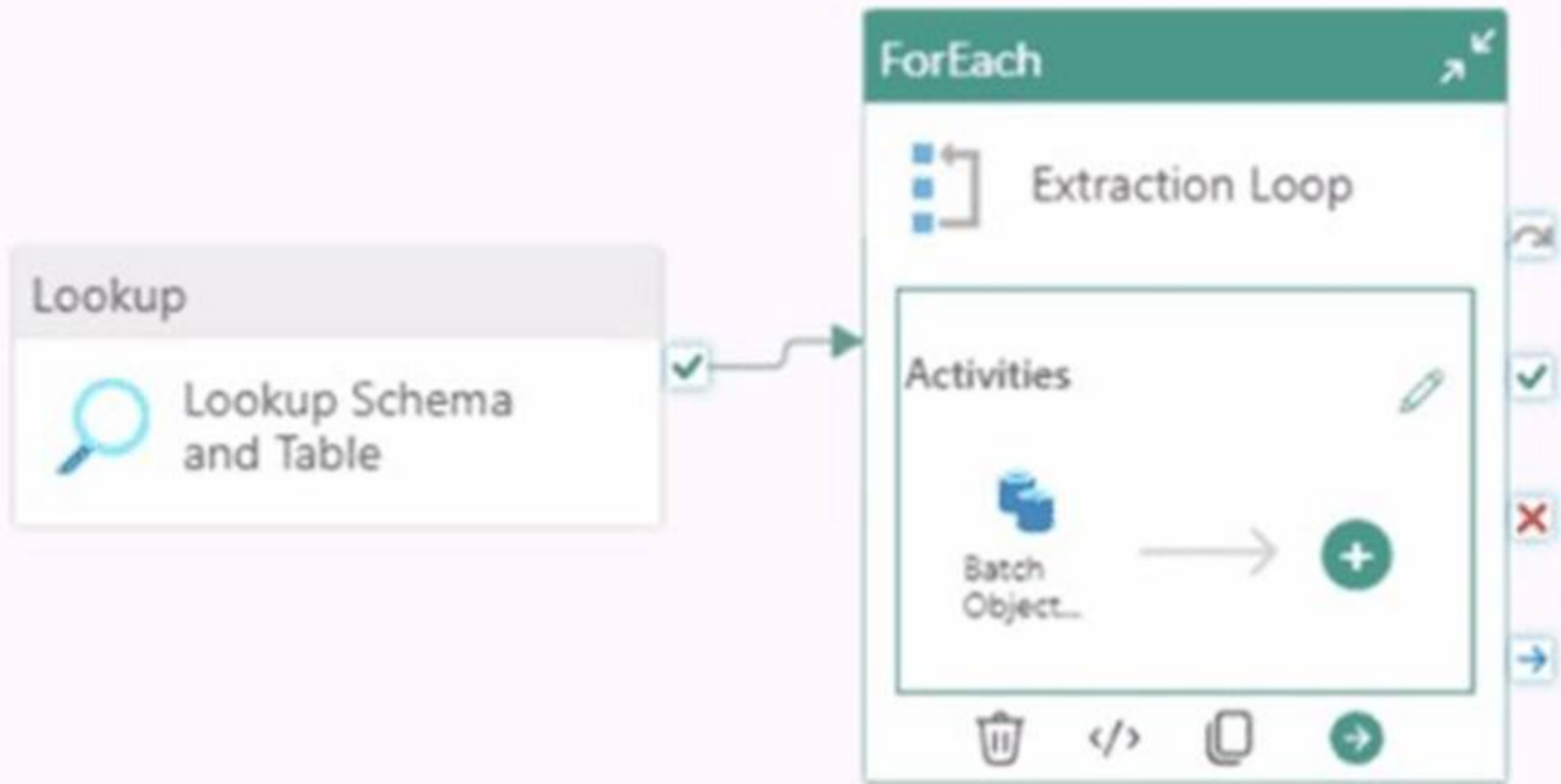
Explanation:

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 6

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').output.value

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 7

- (Topic 3)

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

While monitoring Warehouse1, you discover that query performance has degraded during the last 60 minutes.

You need to isolate all the queries that were run during the last 60 minutes. The results must include the username of the users that submitted the queries and the query statements. What should you use?

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

Answer: B

NEW QUESTION 8

HOTSPOT - (Topic 3)

You have three users named User1, User2, and User3.

You have the Fabric workspaces shown in the following table.

Name	Workspace admin
Workspace1	User1
Workspace2	User2

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

Name	Domain admin
Domain1	User1
Domain2	User2

User1 creates a new workspace named Workspace3. You add Group1 to the default domain of Domain1.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

Statements

	Yes	No
User3 has Viewer role access to Workspace3.	<input 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 9

- (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 10

- (Topic 3)

You have a Fabric workspace that contains a lakehouse named Lakehouse1.

In an external data source, you have data files that are 500 GB each. A new file is added every day.

You need to ingest the data into Lakehouse1 without applying any transformations. The solution must meet the following requirements

Trigger the process when a new file is added.

Provide the highest throughput.

Which type of item should you use to ingest the data?

- A. Event stream
- B. Dataflow Gen2
- C. Streaming dataset
- D. Data pipeline

Answer: A

Explanation:

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 10

- (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 15

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

,item_description as ItemDescription

as PurchaseDate

FROM

Table1

WHERE

item_type = @itemtype_parameter

NEW QUESTION 16

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 18

- (Topic 3)

Your company has a sales department that uses two Fabric workspaces named Workspace1 and Workspace2.

The company decides to implement a domain strategy to organize the workspaces. You need to ensure that a user can perform the following tasks:
Create a new domain for the sales department.

Create two subdomains: one for the east region and one for the west region. Assign Workspace1 to the east region subdomain.

Assign Workspace2 to the west region subdomain. The solution must follow the principle of least privilege. Which role should you assign to the user?

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

Answer: B

Explanation:

To implement a domain strategy and manage subdomains within Fabric, the domain admin role is the appropriate role for the user. A domain admin has the permissions necessary to:

- ? Create a new domain (for the sales department).
- ? Create subdomains (for the east and west regions).
- ? Assign workspaces (such as Workspace1 and Workspace2) to the appropriate subdomains.

The domain admin role allows for managing the structure and organization of workspaces in the context of domains and subdomains while maintaining the principle of least privilege by limiting the user's access to managing the domain structure specifically.

NEW QUESTION 21

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Status_Target that has the following columns:

- Key
- Status
- LastModified

The data source contains a table named Status.Source that has the same columns as Status_Target. Status.Source is used to populate Status_Target. In a notebook name Notebook1, you load Status_Source to a DataFrame named sourceDF and Status_Target to a DataFrame named targetDF. You need to implement an incremental loading pattern by using Notebook1. The solution must meet the following requirements:

- For all the matching records that have the same value of key, update the value of LastModified in Status_Target to the value of LastModified in Status_Source.
- Insert all the records that exist in Status_Source that do NOT exist in Status_Target.
- Set the value of Status in Status_Target to inactive for all the records that were last modified more than seven days ago and that do NOT exist in Status.Source.

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

```

...
(targetDF
  .merge(sourceDF, "sourceDF.Key" = "targetDF.Key")
    .whenMatchedUpdate(
    .whenMatchedInsert(
    .whenMatchedUpdate(
  ) .whenNotMatchedBySourceInsert(
  .whenNotMatchedBySourceUpdate(
  .whenNotMatchedInsert(
  .whenNotMatchedUpdate(
)
    .whenNotMatchedInsert(
    .whenMatchedInsert(
    .whenMatchedUpdate(
    .whenNotMatchedBySourceInsert(
    .whenNotMatchedBySourceUpdate(
    .whenNotMatchedInsert(
    .whenNotMatchedUpdate(
  }
)
    .whenNotMatchedBySourceUpdate(
    .whenMatchedInsert(
    .whenMatchedUpdate(
    .whenNotMatchedBySourceInsert(
    .whenNotMatchedBySourceUpdate(
    .whenNotMatchedInsert(
    .whenNotMatchedUpdate(
)
)

```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 29

- (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 33

HOTSPOT - (Topic 3)

You plan to process the following three datasets by using Fabric:

- Dataset1: This dataset will be added to Fabric and will have a unique primary key between the source and the destination. The unique primary key will be an integer and will start from 1 and have an increment of 1.
- Dataset2: This dataset contains semi-structured data that uses bulk data transfer. The dataset must be handled in one process between the source and the destination. The data transformation process will include the use of custom visuals to understand and work with the dataset in development mode.
- Dataset3: This dataset is in a takehouse. The data will be bulk loaded. The data transformation process will include row-based windowing functions during the loading process.

You need to identify which type of item to use for the datasets. The solution must minimize development effort and use built-in functionality, when possible. What should you identify for each dataset? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Dataset1:

Dataset2:

Dataset3:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Dataset1:

Dataset2:

Dataset3:

NEW QUESTION 35

DRAG DROP - (Topic 3)

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

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

You need to build two KQL queries. The solution must meet the following requirements: One of the queries must partition RunningTotalAmount by VendorID. The other query must create a column named FirstPickupDateTime that shows the first value of each hour from tpep_pickup_datetime partitioned by payment_type.

How should you complete each query? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values

-
-
-
-

Answer Area

Statement1:

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

Statement2:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Partition the RunningTotalAmount by VendorID. - Row_cumsum

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

```
TaxiData
```

```
| sort by VendorID asc
```

```
| extend RunningTotalAmount = Row_cumsum(total_amount, VendorID != prev(VendorID))
```

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

```
TaxiData
```

```
| sort by tpep_pickup_datetime asc, payment_type asc
```

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

NEW QUESTION 40

- (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 43

- (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 46

- (Topic 3)

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table in a lakehouse.

You need to remove files that are older than seven days and are no longer in use. Which command should you run?

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

Answer: A

Explanation:

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

```
VACUUM delta.`/path_to_table` RETAIN 7 HOURS;
```

NEW QUESTION 49

- (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 52

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

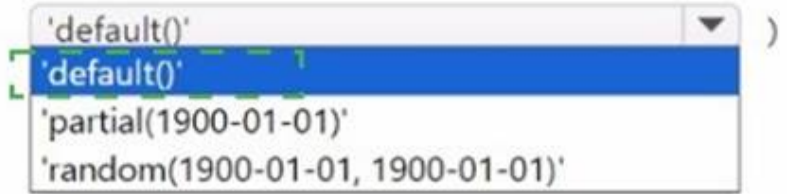
Explanation:

Answer Area

```
ALTER TABLE DimCustomer
```

```
ALTER COLUMN BirthDate
```

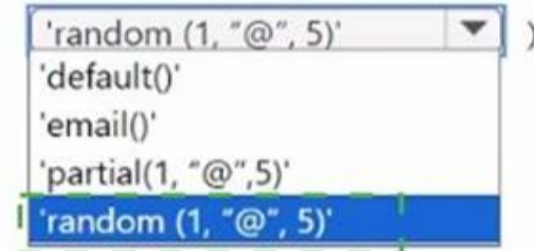
```
ADD MASKED WITH (FUNCTION =
```



```
ALTER TABLE DimCustomer
```

```
ALTER COLUMN EmailAddress
```

```
ADD MASKED WITH (FUNCTION =
```



NEW QUESTION 56

- (Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some

question sets might have more than one correct solution, while others might not have a correct solution.

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

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

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp

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

Solution: You use the following code segment:

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

Does this meet the goal?

- A. Yes
- B. no

Answer: B

Explanation:

This code does not meet the goal because it uses order by, which is not valid in KQL. The correct term in KQL is sort by.

Correct code should look like:

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

NEW QUESTION 57

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains an eventstream named EventStream1. You discover that an EventStream1 transformation fails.

You need to find the following error information: The error details, including the occurrence time The total number of errors

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

Answer Area

To find the error details:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

To find the total number of errors:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

To find the error details:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

To find the total number of errors:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

NEW QUESTION 58

- (Topic 3)

You have a Fabric workspace named Workspace1. Your company acquires GitHub licenses.

You need to configure source control for Workspace1 to use GitHub. The solution must follow the principle of least privilege. Which permissions do you require to ensure that you can commit code to GitHub?

- A. Actions (Read and write) and Contents (Read and write)

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

Answer: C

NEW QUESTION 62

- (Topic 3)

You have a Fabric workspace that contains a semantic model named Model1. You need to dynamically execute and monitor the refresh progress of Model1. What should you use?

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

Answer: D

Explanation:

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

NEW QUESTION 66

- (Topic 3)

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

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

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

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

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

You need to reduce how long it takes to run the KQL queryset. Solution: You add the make_list() function to the output columns. Does this meet the goal?

```

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)
    
```

- A. Yes
- B. No

Answer: B

Explanation:

Adding an aggregation like `make_list()` would require additional processing and memory, which could make the query slower.

NEW QUESTION 67

- (Topic 3)

You are building a Fabric notebook named MasterNotebook1 in a workspace. MasterNotebook1 contains the following code.

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

You need to ensure that the notebooks are executed in the following sequence:

- * 1. Notebook_03
- * 2. Notebook_01
- * 3. Notebook_02

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

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

Answer: CE

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

```

(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
CUBE(YEAR(SO.ModifiedDate), P.Name)
(ROLLUP(S (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
CUBE(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 77

- (Topic 3)

You have two Fabric workspaces named Workspace1 and Workspace2.

You have a Fabric deployment pipeline named deployPipeline1 that deploys items from Workspace1 to Workspace2. DeployPipeline1 contains all the items in Workspace1.

You recently modified the items in Workspace1.

The workspaces currently contain the items shown in the following table.

Workspace	Items
Workspace1	Model1 Notebook1 Report1 Lakehouse1 Pipeline1
Workspace2	Model1 Notebook2 Report1 Lakehouse2

Items in Workspace1 that have the same name as items in Workspace2 are currently paired.

You need to ensure that the items in Workspace1 overwrite the corresponding items in Workspace2. The solution must minimize effort.

What should you do?

- A. Delete all the items in Workspace2, and then run deployPipeline1.
- B. Rename each item in Workspace2 to have the same name as the items in Workspace1.
- C. Back up the items in Workspace2, and then run deployPipeline1.
- D. Run deployPipeline1 without modifying the items in Workspace2.

Answer: D

Explanation:

When running a deployment pipeline in Fabric, if the items in Workspace1 are paired with the corresponding items in Workspace2 (based on the same name), the deployment pipeline will automatically overwrite the existing items in Workspace2 with the modified items from Workspace1. There's no need to delete, rename, or back up items manually unless you need to keep versions. By simply running deployPipeline1, the pipeline will handle overwriting the existing items in Workspace2 based on the pairing, ensuring the latest version of the items is deployed with minimal effort.

NEW QUESTION 78

HOTSPOT - (Topic 3)

You have a table in a Fabric lakehouse that contains the following data.

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

You have a notebook that contains the following code segment.

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

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

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

NEW QUESTION 81

- (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

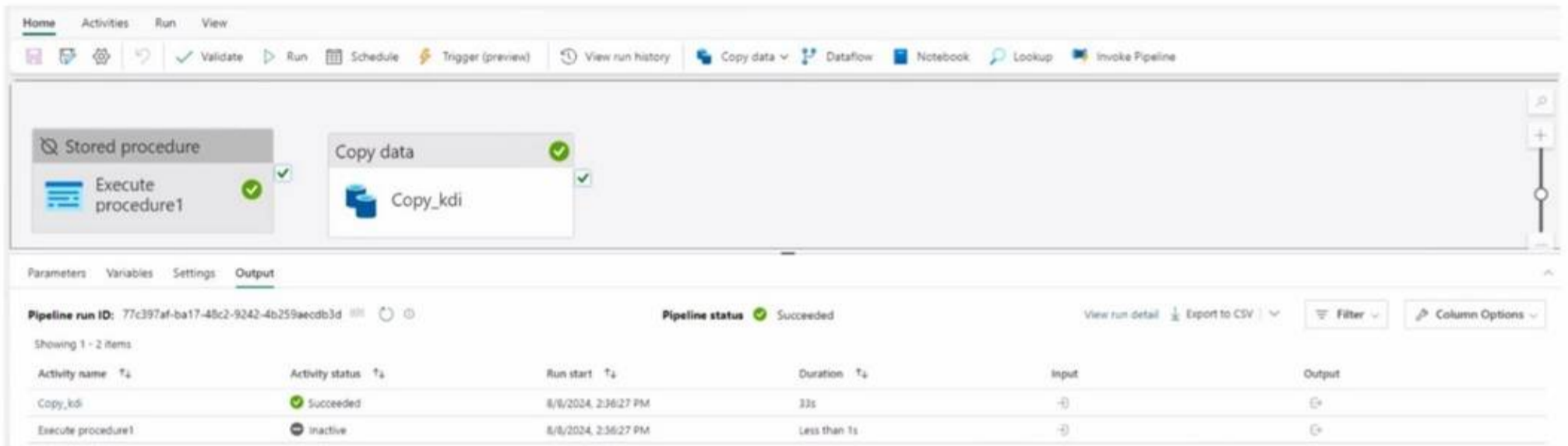
Answer: C

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 83

- (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 85

- (Topic 3)

You have a Fabric workspace that contains a lakehouse named Lakehouse1.

You plan to create a data pipeline named Pipeline1 to ingest data into Lakehouse1. You will use a parameter named param1 to pass an external value into Pipeline1. The param1 parameter has a data type of int

You need to ensure that the pipeline expression returns param1 as an int value. How should you specify the parameter value?

- A. "@pipeline(). parameter
- B. param1"
- C. "@{pipeline().parameters.param1}"
- D. "@{pipeline().parameters.[param1]}"
- E. "@{pipeline().parameters.param1}-

Answer: B

NEW QUESTION 86

- (Topic 3)

You have an Azure Data Lake Storage Gen2 account named storage1 and an Amazon S3 bucket named storage2.

You have the Delta Parquet files shown in the following table.

Name	Stored in	Size	Description
ProductFile	storage1	50 MB	Contains a list of products and their details
TripsFile	storage2	2 GB	Contains one month's worth of taxi trip data
StoreFile	storage2	25 MB	Contains a list of stores and their addresses

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

A shortcut to ProductFile aliased as Products A shortcut to StoreFile aliased as Stores

A shortcut to TripsFile aliased as Trips

The data from which shortcuts will be retrieved from the cache?

- A. Trips and Stores only
- B. Products and Store only
- C. Stores only

- D. Products only
- E. Product
- F. Stores, and Trips

Answer: B

Explanation:

When the cache for shortcuts is enabled in Fabric, the data retrieval is governed by the caching behavior, which generally retains data for a specific period after it was last accessed. The data from the shortcuts will be retrieved from the cache if the data is stored in locations that support caching. Here's a breakdown based on the data's location: Products: The ProductFile is stored in Azure Data Lake Storage Gen2 (storage1). Since Azure Data Lake is a supported storage system in Fabric and the file is relatively small (50 MB), this data is most likely cached and can be retrieved from the cache.

Stores: The StoreFile is stored in Amazon S3 (storage2), and even though it is stored in a different cloud provider, Fabric can cache data from Amazon S3 if caching is enabled. This data (25 MB) is likely cached and retrievable.

Trips: The TripsFile is stored in Amazon S3 (storage2) and is significantly larger (2 GB) compared to the other files. While Fabric can cache data from Amazon S3, the larger size of the file (2 GB) may exceed typical cache sizes or retention windows, causing this file to likely be retrieved directly from the source instead of the cache.

NEW QUESTION 90

- (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 95

.....

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](#)