



## 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
    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**

- (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 3**

- (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 4**

- (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

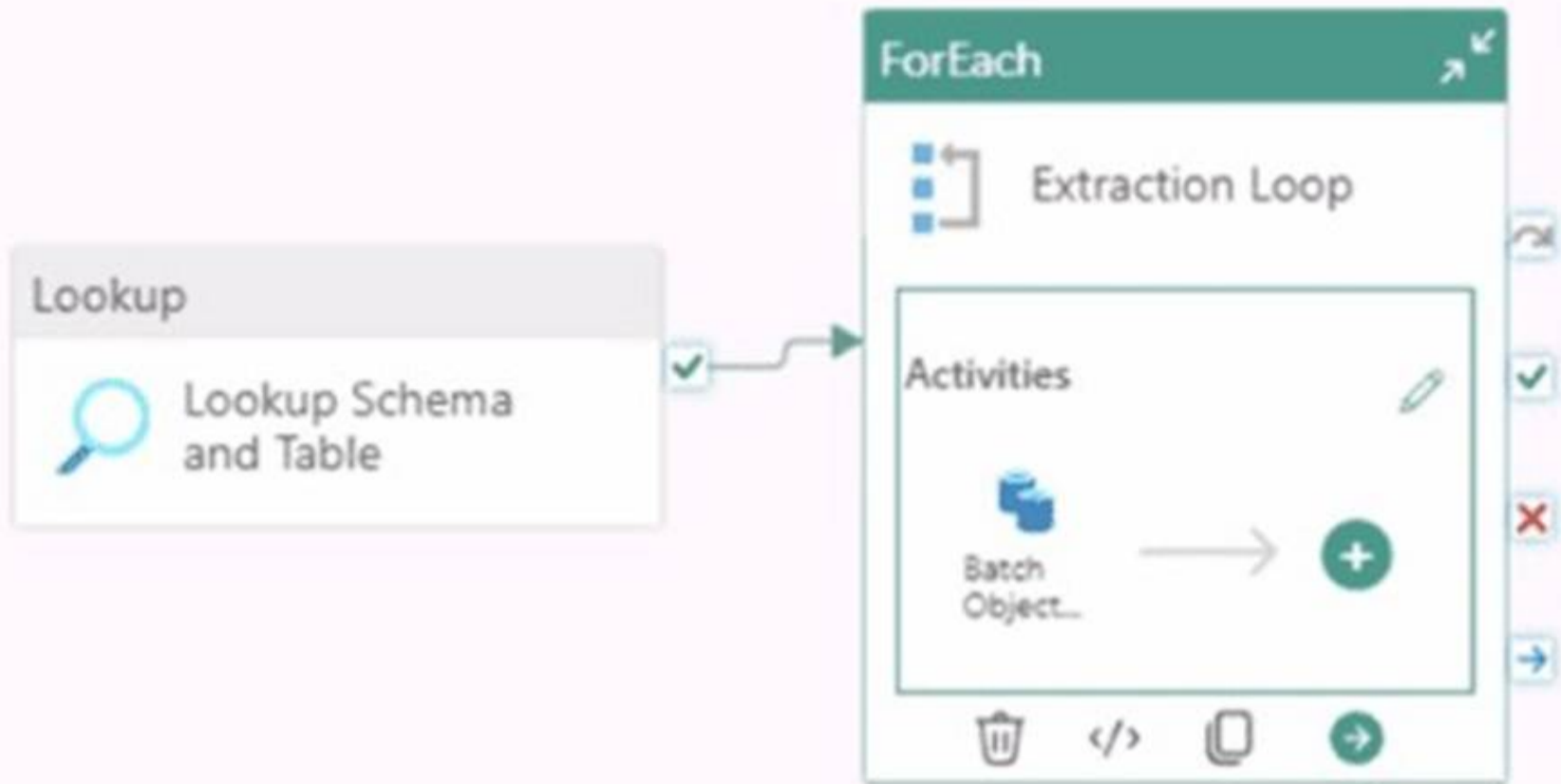
**Explanation:**

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 5**

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** <sup>1</sup> 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 6**

- (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 7**

- (Topic 3)

You have a Fabric workspace that contains a lakehouse named Lakehouse1. Data is ingested into Lakehouse1 as one flat table. The table contains the following columns.

Name	Description
TransactionID	Contains a unique ID for each transaction
Date	Contains the date of a transaction
ProductID	Contains a unique ID for each product
ProductColor	Contains a descriptive attribute that describes the color of each product
ProductName	Contains a unique name for each product
SalesAmount	Contains the sales amount of a transaction

You plan to load the data into a dimensional model and implement a star schema. From the original flat table, you create two tables named FactSales and DimProduct. You will track changes in DimProduct.

You need to prepare the data.

Which three columns should you include in the DimProduct table? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Date
- B. ProductName
- C. ProductColor
- D. TransactionID
- E. SalesAmount
- F. ProductID

**Answer: BCF**

**Explanation:**

In a star schema, the DimProduct table serves as a dimension table that contains descriptive attributes about products. It will provide context for the FactSales table, which contains transactional data. The following columns should be included in the DimProduct table:

- ? ProductName: The ProductName is an important descriptive attribute of the product, which is needed for analysis and reporting in a dimensional model.
- ? ProductColor: ProductColor is another descriptive attribute of the product. In a star schema, it makes sense to include attributes like color in the dimension table to help categorize products in the analysis.
- ? ProductID: ProductID is the primary key for the DimProduct table, which will be used to join the FactSales table to the product dimension. It's essential for uniquely identifying each product in the model.

**NEW QUESTION 8**

- (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 9**

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 contains a table named Customer. Customer contains the following data.

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

You have an internal Microsoft Entra user named User1 that has an email address of user1@contoso.com.  
 You need to provide User1 with access to the Customer table. The solution must prevent User1 from accessing the CreditCard column.  
 How should you complete the statement? To answer, select the appropriate options in the answer area.  
 NOTE: Each correct selection is worth one point.

**Answer Area**

GRANT SELECT ▼

ALTER

EXECUTE

READ

SELECT

VIEW

Customers(CustomerID, FirstName, LastName, Phone)

TO [user1@contoso.com] ▼

User1

[User1]

[user1@contoso.com]

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Answer Area



**NEW QUESTION 10**

- (Topic 3)

You have a Fabric workspace that contains an eventhouse and a KQL database named Database1. Database1 has the following:

A table named Table1 A table named Table2

An update policy named Policy1

Policy1 sends data from Table1 to Table2.

The following is a sample of the data in Table2.

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-18 12:45:17.16524	81416f30-60a2-4e75-9b19-2a84ea059735	[ { "index": 0, "eventId": "719afca0-be30-4559-bb5e-59feade642f6" } ]
2024-05-18 12:45:21.76423	bb664e1e-02aa-4e17-8c8a-116cd4458d52	[ { "index": 0, "eventId": "782222b2-fbcb-43c0-82d6-ecd49a99dbf5" } ]
2024-05-18 12:45:23.98642	717bfe7d-0e5d-498f-9f21-e60aaf258056	[ { "index": 0, "eventId": "d5730286-0da4-41f8-8e59-f75e209310a9" } ]

Recently, the following actions were performed on Table1:

An additional element named temperature was added to the StreamData column. The data type of the Timestamp column was changed to date.

The data type of the DeviceId column was changed to string. You plan to load additional records to Table2.

Which two records will load from Table1 to Table2? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-18	81416f30-60a2-4e75-9b19-2a84ea059735	[ { "index": 40, "eventId": "729afca2-be30-4559-bb5e-59feade642f3", "temperature": 32 } ]

B)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-21	81416f30	[ { "index": 0, "eventId": "719afca0-be30-4559-bb5e-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 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 14**

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 >= valid_to_datetime
o.OrderDate >= valid_from_datetime

AND 
o.OrderDate <= valid_to_datetime
o.OrderDate <= 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 <= c.valid\_to\_datetime

o.OrderDate >= c.valid\_from\_datetime

AND

c.is\_current = 1

o.OrderDate <= c.valid\_to\_datetime

o.OrderDate <= c.valid\_from\_datetime

### NEW QUESTION 17

- (Topic 3)

You have an Azure SQL database named DB1.

In a Fabric workspace, you deploy an eventstream named EventStreamDBI to stream record changes from DB1 into a lakehouse.

You discover that events are NOT being propagated to EventStreamDBI.

You need to ensure that the events are propagated to EventStreamDBI. What should you do?

- A. Create a read-only replica of DB1.
- B. Create an Azure Stream Analytics job.
- C. Enable Extended Events for DB1.
- D. Enable change data capture (CDC) for DB1.

Answer: D

### NEW QUESTION 19

- (Topic 3)

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

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

Answer: B

### NEW QUESTION 21

**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**

Badges available for selection:

- Certified
- Master data
- Promoted
- Cannot be endorsed

**Answer Area**

Entity1:

Entity2:

Entity3:

Entity4:

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

**Badges**

Badges available for selection (with dashed green boxes indicating they are the correct answer):

- Certified
- Master data
- Promoted
- Cannot be endorsed

**Answer Area**

Entity1:

Entity2:

Entity3:

Entity4:

**NEW QUESTION 26**

**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.





You need to copy data from Database1 to Warehouse1. Which item should you use?

- A. a Dataflow Gen1 dataflow
- B. a data pipeline
- C. a KQL queryset
- D. a notebook

**Answer:** B

**Explanation:**

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

**NEW QUESTION 35**

- (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 38**

HOTSPOT - (Topic 3)

You have a Fabric workspace named Workspace1 that contains the items shown in the following table.

Name	Type
Notebook1	Notebook
Notebook2	Notebook
Lakehouse1	Lakehouse
Pipeline1	Data pipeline
Model1	Semantic model

For Model1, the Keep your Direct Lake data up to date option is disabled.

You need to configure the execution of the items to meet the following requirements:

Notebook1 must execute every weekday at 8:00 AM.

Notebook2 must execute when a file is saved to an Azure Blob Storage container. Model1 must refresh when Notebook1 has executed successfully.

How should you orchestrate each item? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

**Notebook1:**  
 Add Notebook1 to an Apache Spark job definition.  
 Add Notebook1 to Pipeline1.  
 From Real-Time hub, configure the execution of Notebook1

**Notebook2:**  
 Add Notebook2 to an Apache Spark job definition.  
 Add Notebook2 to Pipeline1.  
 From Real-Time hub, configure the execution of Notebook2

**Pipeline1:**  
 Add Pipeline1 to an Apache Spark job definition.  
 Configure the execution of Pipeline1 by using a schedule  
 From Real-Time hub, configure the execution of Pipeline1.

**Model1:**  
 Add Model1 to Pipeline1  
 From Real-Time hub, configure Model1 to refresh.  
 Set Keep your Direct Lake data up to date to On.

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**  
**Answer Area**

**Notebook1:**  
 Add Notebook1 to an Apache Spark job definition.  
 Add Notebook1 to Pipeline1.  
 From Real-Time hub, configure the execution of Notebook1

**Notebook2:**  
 Add Notebook2 to an Apache Spark job definition.  
 Add Notebook2 to Pipeline1.  
 From Real-Time hub, configure the execution of Notebook2

**Pipeline1:**  
 Add Pipeline1 to an Apache Spark job definition.  
 Configure the execution of Pipeline1 by using a schedule  
 From Real-Time hub, configure the execution of Pipeline1.

**Model1:**  
 Add Model1 to Pipeline1  
 From Real-Time hub, configure Model1 to refresh.  
 Set Keep your Direct Lake data up to date to On.

**NEW QUESTION 40**

- (Topic 3)

You have a Fabric workspace that contains a semantic model named Model1. You need to monitor the refresh history of Model 1 and visualize the refresh history in a chart. What should you use?

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

**Answer:** A

**NEW QUESTION 43**

- (Topic 3)

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

You have a Fabric workspace named Workspace1. Workspace1 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 the key vault URL and key vault secret. Does this meet the goal?

- A. Yes
- B. No

Answer: A

**NEW QUESTION 45**

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 46**

- (Topic 3)

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

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

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

Trigger the process when a new file is added. Provide the highest throughput.

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

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

**Answer:** A

**Explanation:**

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

**NEW QUESTION 49**

HOTSPOT - (Topic 3)

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

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

WITH result AS(

SELECT e.EmployeeID

, e.EmployeeName

, e.EmployeePosition

, c.ContractType

, (date, c.StartDate) as StartDate

FROM Employee AS e

Contract AS c on c.EmployeeID = e.EmployeeID

)

SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

, ContractType

FROM result

GROUP BY ContractType

COUNT(DISTINCT EmployeeID) > 2

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

WITH result AS(

SELECT e.EmployeeID

, e.EmployeeName

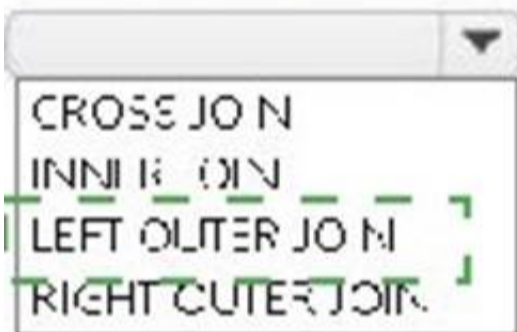
, e.EmployeePosition

, c.ContractType

, (date, c.StartDate) as StartDate



FROM Employee AS e



Contract AS c on c.EmployeeID = e.EmployeeID

)  
 SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

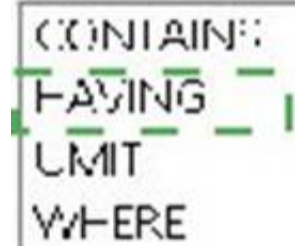
, ContractType

FROM result

GROUP BY ContractType



COUNT(DISTINCT EmployeeID) > 2



NEW QUESTION 54

- (Topic 3)

You have five Fabric workspaces.

You are monitoring the execution of items by using Monitoring hub.

You need to identify in which workspace a specific item runs. Which column should you view in Monitoring hub?

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

Answer: G

**Explanation:**

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

**NEW QUESTION 58**

- (Topic 3)

You have a Fabric workspace named Workspace1. You plan to integrate Workspace1 with Azure DevOps.

You will use a Fabric deployment pipeline named deployPipeline1 to deploy items from Workspace1 to higher environment workspaces as part of a medallion architecture. You will run deployPipeline1 by using an API call from an Azure DevOps pipeline.

You need to configure API authentication between Azure DevOps and Fabric. Which type of authentication should you use?

- A. service principal
- B. Microsoft Entra username and password
- C. managed private endpoint
- D. workspace identity

**Answer:** A

**Explanation:**

When integrating Azure DevOps with Fabric (Workspace1), using a service principal is the recommended authentication method. A service principal provides a way for applications (such as an Azure DevOps pipeline) to authenticate and interact with resources securely. It allows Azure DevOps to authenticate API calls to Fabric without requiring direct user credentials. This method is ideal for automating tasks such as deploying items through a Fabric deployment pipeline.

**NEW QUESTION 63**

- (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 64**

- (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 69**

- (Topic 3)

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

You have an on-premises Microsoft SQL Server database named Database1 that is accessed by using an on-premises data gateway.

You need to copy data from Database1 to Warehouse1. Which item should you use?

- A. an Apache Spark job definition
- B. a data pipeline
- C. a Dataflow Gen1 dataflow
- D. an eventstream

**Answer: B**

**Explanation:**

To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Fabric, a data pipeline is the most appropriate tool. A data pipeline in Fabric is designed to move data between various data sources and destinations, including on-premises databases like SQL Server, and cloud-based storage like Fabric warehouses. The data pipeline can handle the connection through an on-premises data gateway, which is required to access on-premises data. This solution facilitates the orchestration of data movement and transformations if needed.

**NEW QUESTION 70**

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 71

- (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 73**

- (Topic 3)

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

You have a Fabric workspace named Workspace1. Workspace1 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 76**

- (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 80**

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 83**

- (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 84**

- (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 89**

HOTSPOT - (Topic 3)

You have a Fabric warehouse named DW1 that contains four staging tables named ProductCategory, ProductSubcategory, Product, and SalesOrder.

ProductCategory, ProductSubcategory, and Product are used often in analytical queries.

You need to implement a star schema for DW1. The solution must minimize development effort.

Which design approach should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

ProductCategory, ProductSubcategory and Product must be:

- Denormalized into a single product dimension table
- Added to the model as individual tables
- Denormalized by being added to the SalesOrder table
- Denormalized into a single product dimension table**

The joining key must be:

- the unique system generated identifier
- The product name and the date
- the unique system generated identifier
- The product category name**

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

ProductCategory, ProductSubcategory and Product must be:

- Denormalized into a single product dimension table
- Added to the model as individual tables
- Denormalized by being added to the SalesOrder table
- Denormalized into a single product dimension table**

The joining key must be:

- the unique system generated identifier
- The product name and the date
- the unique system generated identifier
- The product category name**

### NEW QUESTION 92

- (Topic 3)

You have a Fabric workspace named Workspace1.

You plan to configure Git integration for Workspace1 by using an Azure DevOps Git repository. An Azure DevOps admin creates the required artifacts to support the integration of Workspace1. Which details do you require to perform the integration?

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

Answer: B

### NEW QUESTION 95

- (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 100

- (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

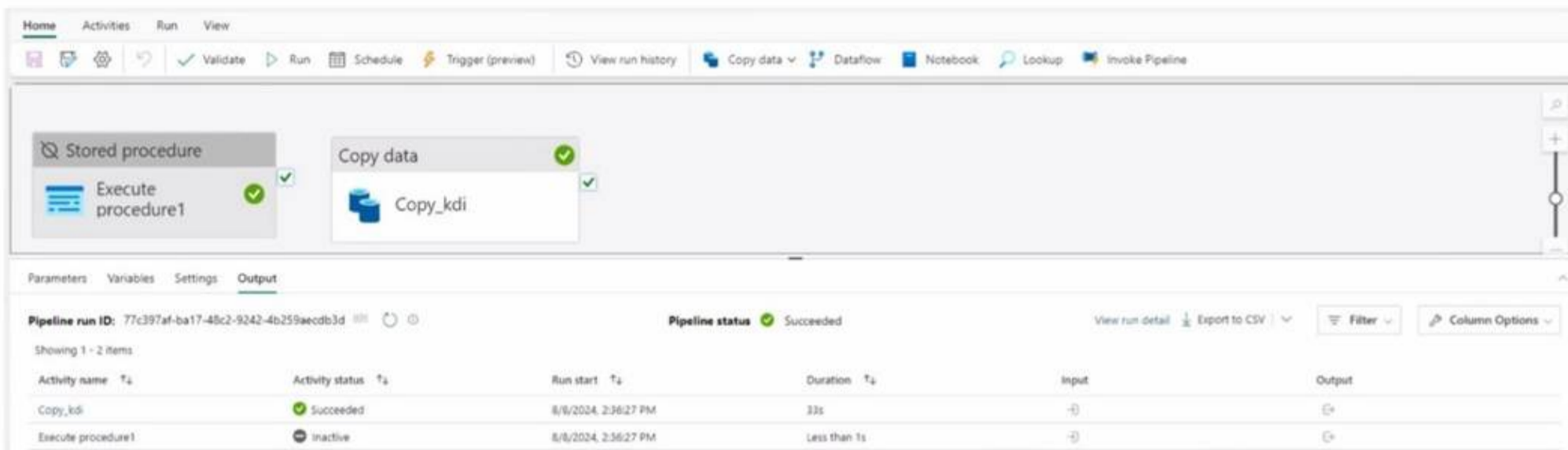
Answer: B

**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 103**

- (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 104**

- (Topic 3)

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

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

Both tables contain millions of rows. You have the following KQL queryset. You need to reduce how long it takes to run the KQL queryset. Solution: You move the filter to line 02.

```

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

```

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:**

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

**NEW QUESTION 105**

- (Topic 3)

You have 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 107**

- (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 112**

.....

## 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](#)