

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

FULL JOIN

INNER JOIN

LEFT ANTI JOIN

LEFT OUTER JOIN

OUTER JOIN

ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID

FULL JOIN

INNER JOIN

LEFT ANTI JOIN

LEFT OUTER JOIN

OUTER JOIN

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

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.

Passing Certification Exams Made Easy

visit - <https://www.surepassexam.com>

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 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 2)

You need to implement the solution for the book reviews.
Which should you do?

- A. Create a Dataflow Gen2 dataflow.
- B. Create a shortcut.
- C. Enable external data sharing.
- D. Create a data pipeline.

Answer: B

Explanation:

The requirement specifies that Litware plans to make the book reviews available in the lakehouse without making a copy of the data. In this case, creating a shortcut in Fabric is the most appropriate solution. A shortcut is a reference to the external data, and it allows Litware to access the book reviews stored in Amazon S3 without duplicating the data into the lakehouse.

NEW QUESTION 6

- (Topic 2)

What should you do to optimize the query experience for the business users?

- A. Enable V-Order.
- B. Create and update statistics.
- C. Run the VACUUM command.
- D. Introduce primary keys.

Answer: B

NEW QUESTION 7

- (Topic 3)

You have 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 8

- (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 9

- (Topic 3)

You have a Fabric warehouse named DW1. DW1 contains a table that stores sales data and is used by multiple sales representatives.

You plan to implement row-level security (RLS).

You need to ensure that the sales representatives can see only their respective data. Which warehouse object do you require to implement RLS?

- A. ISTORED PROCEDURE
- B. CONSTRAINT
- C. SCHEMA

D. FUNCTION

Answer: D

Explanation:

To implement Row-Level Security (RLS) in a Fabric warehouse, you need to use a function that defines the security logic for filtering the rows of data based on the user's identity or role. This function can be used in conjunction with a security policy to control access to specific rows in a table.
In the case of sales representatives, the function would define the filtering criteria (e.g., based on a column such as SalesRepID or SalesRepName), ensuring that each representative can only see their respective data.

NEW QUESTION 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- fbc0-43c0-82d6-ecd49a99dbf5" }]
2024-05-18 12:45:23.98642	717bfe7d- 0e5d-498f- 9f21- e60aaf258056	[{ "index": 0, "eventid": "d5730286- 0da4-41f8-8e59-f75e209310a9" }]

Recently, the following actions were performed on Table1:

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

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

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

NOTE: Each correct selection is worth one point.

A)

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

B)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-21	81416f30	[{ "index": 0, "eventId": "719afca0-be30-4559-bb5e-5werade642f6", "temperature": 27 }]

C)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-23	81416f3060a24e759b192a84ea05973532dhdyte3	[{ "index": 0, "eventId": "719afca0-be30-4559-bb5e-59feade642f6" }]

D)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-24	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 0, "eventId": "719afca0-be30-4559-bb5e-59feade642f6" }]

- A. Option A
- B. Option B
- C. Option c
- D. Option D

Answer: BD

Explanation:

Changes to Table1 Structure:
StreamData column: An additional temperature element was added. Timestamp column: Data type changed from datetime to date. DeviceId column: Data type changed from guid to string.
Impact of Changes:
Only records that comply with Table2??s structure will load.
Records that deviate from Table2??s column data types or structure will be rejected.
Record B:
Timestamp: Matches Table2 (datetime format). DeviceId: Matches Table2 (guid format).
StreamData: Contains only the index and eventId, which matches Table2. Accepted because it fully matches Table2??s structure and data types.
Record D:
Timestamp: Matches Table2 (datetime format). DeviceId: Matches Table2 (guid format). StreamData: Matches Table2??s structure.
Accepted because it fully matches Table2??s structure and data types.

NEW QUESTION 10

- (Topic 3)
You have a Fabric workspace 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 11

- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a notebook named Notebook1.
In Workspace1, you create a new notebook named Notebook2.
You need to ensure that you can attach Notebook2 to the same Apache Spark session as Notebook1.
What should you do?

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

Answer: A

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

NEW QUESTION 12

- (Topic 3)
You have a Fabric workspace. You have semi-structured data.
You need to read the data by using T-SQL, KQL, and Apache Spark. The data will only be written by using Spark.
What should you use to store the data?

- A. a lakehouse
- B. an eventhouse
- C. a datamart
- D. a warehouse

Answer: A

Explanation:
A lakehouse is the best option for storing semi-structured data when you need to read it using T-SQL, KQL, and Apache Spark. A lakehouse combines the flexibility of a data lake (which can handle semi-structured and unstructured data) with the performance features of a data warehouse. It allows data to be written using Apache Spark and can be queried using different technologies such as T-SQL (for SQL-based querying), KQL (Kusto Query Language for querying), and Apache Spark (for distributed processing). This solution is ideal when dealing with semi-structured data and requiring a versatile querying approach.

NEW QUESTION 13

- (Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

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

01 Stream

02 | extend lat = todecimal(GeoLocation.Latitude), long = todecimal(GeoLocation.Longitude)

03 | join kind=inner Reference on DeviceId

04 | project Timestamp, lat, long, Temperature, DeviceName

05 | filter Temperature >= 10

06 | render scatterchart with (kind = map)

You need to reduce how long it takes to run the KQL queryset. Solution: You change the join type to kind=outer. Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 14

DRAG DROP - (Topic 3)

You are building a data loading pattern by using a Fabric data pipeline. The source is an Azure SQL database that contains 25 tables. The destination is a lakehouse.

In a warehouse, you create a control table named Control.Object as shown in the exhibit. (Click the Exhibit tab.)

You need to build a data pipeline that will support the dynamic ingestion of the tables listed in the control table by using a single execution.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

- ⋮ Add a Get metadata activity to query Control.Object and generate a list of schemas and tables to copy.
- ⋮ Add an Until activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
- ⋮ Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.
- ⋮ Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
- ⋮ Add a Copy data activity as an inner activity to the iterator activity.

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions

- ⋮ Add a Get metadata activity to query Control.Object and generate a list of schemas and tables to copy.
- ⋮ Add an Until activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
- ⋮ Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.
- ⋮ Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
- ⋮ Add a Copy data activity as an inner activity to the iterator activity.

Answer Area

- ⋮ Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.
- ⋮ Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
- ⋮ Add a Copy data activity as an inner activity to the iterator activity.

NEW QUESTION 15

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

Certified

Master data

Promoted

Cannot be endorsed

Answer Area

Entity1:

Master data

Entity2:

Certified

Entity3:

Promoted

Entity4:

Certified

NEW QUESTION 19

- (Topic 3)

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

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

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

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp

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

Solution: You use the following code segment:

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

Does this meet the goal?

- A. Yes
- B. no

Answer: B

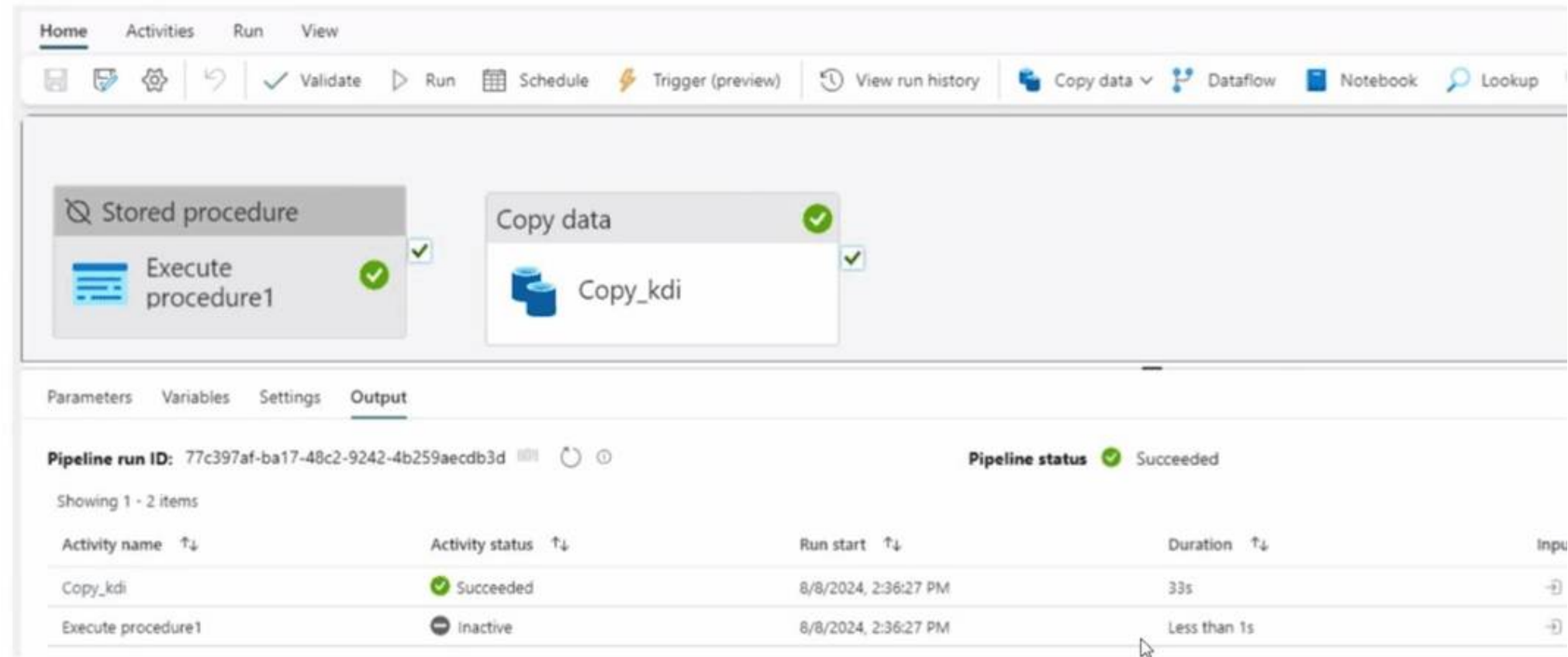
Explanation:

This code does not meet the goal because it uses sort by without specifying the order, which defaults to ascending, but explicitly mentioning asc improves clarity. 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 24

- (Topic 3)
You have a Fabric workspace that contains a data pipeline named Pipeline1 as shown in the exhibit.



What will occur the next time Pipeline1 runs?

- A. Both activities will run simultaneously.
- B. Both activities will be skipped.
- C. Execute procedure1 will run and Copy_kdi will be skipped.
- D. Copy_kdi will run and Execute procedure1 will be skipped.
- E. Execute procedure1 will run first, and then Copy_kdi will run.
- F. Copy_kdi will run first, and then Execute procedure1 will run.

Answer: A

NEW QUESTION 26

- (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 31

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

- NOTE: Each correct selection is worth one point.

• • •

)

}

)

```
ent_date() - INTERVAL '7' DAY)",
```

)

- visit - <https://www.surepassexam.com>

Answer Area

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

NEW QUESTION 33

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:	<input type="checkbox"/> Add Notebook1 to an Apache Spark job definition. <input type="checkbox"/> Add Notebook1 to Pipeline1. <input type="checkbox"/> From Real-Time hub, configure the execution of Notebook1
Notebook2:	<input type="checkbox"/> Add Notebook2 to an Apache Spark job definition. <input type="checkbox"/> Add Notebook2 to Pipeline1. <input type="checkbox"/> From Real-Time hub, configure the execution of Notebook2
Pipeline1:	<input type="checkbox"/> Add Pipeline1 to an Apache Spark job definition. <input type="checkbox"/> Configure the execution of Pipeline1 by using a schedule <input type="checkbox"/> From Real-Time hub, configure the execution of Pipeline1.
Model1:	<input type="checkbox"/> Add Model1 to Pipeline1 <input type="checkbox"/> From Real-Time hub, configure Model1 to refresh. <input type="checkbox"/> Set Keep your Direct Lake data up to date to On.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Notebook1:	<input checked="" type="checkbox"/> Add Notebook1 to an Apache Spark job definition. <input checked="" type="checkbox"/> Add Notebook1 to Pipeline1. <input checked="" type="checkbox"/> From Real-Time hub, configure the execution of Notebook1
Notebook2:	<input type="checkbox"/> Add Notebook2 to an Apache Spark job definition. <input type="checkbox"/> Add Notebook2 to Pipeline1. <input checked="" type="checkbox"/> From Real-Time hub, configure the execution of Notebook2
Pipeline1:	<input type="checkbox"/> Add Pipeline1 to an Apache Spark job definition. <input checked="" type="checkbox"/> Configure the execution of Pipeline1 by using a schedule <input checked="" type="checkbox"/> From Real-Time hub, configure the execution of Pipeline1.
Model1:	<input checked="" type="checkbox"/> Add Model1 to Pipeline1 <input type="checkbox"/> From Real-Time hub, configure Model1 to refresh. <input type="checkbox"/> Set Keep your Direct Lake data up to date to On.

NEW QUESTION 34

- (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 38

- (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 NotebookI that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVaultI 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 41

- (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 43

- 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

CASE
CONVERT
REPLACE
SUBSTRING

FROM Employee AS e

Contract AS c on c.EmployeeID = e.EmployeeID

CROSS JOIN
INNER JOIN
LEFT OUTER JOIN
RIGHT OUTER JOIN

)

SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

, ContractType

FROM result

GROUP BY ContractType

COUNT(DISTINCT EmployeeID) > 2

CONTAINS
HAVING
LIMIT
WHERE

- 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



)

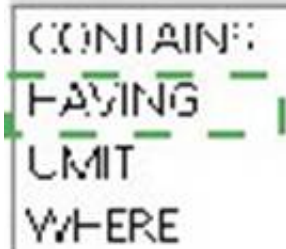
SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

, ContractType

FROM result

GROUP BY ContractType

COUNT(DISTINCT EmployeeID) > 2



NEW QUESTION 46

- (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 49

DRAG DROP - (Topic 3)

You have two Fabric notebooks named Load_Salesperson and Load_Orders that read data from Parquet files in a lakehouse. Load_Salesperson writes to a Delta table named dim_salesperson. Load.Orders writes to a Delta table named fact_orders and is dependent on the successful execution of Load_Salesperson. You need to implement a pattern to dynamically execute Load_Salesperson and Load_Orders in the appropriate order by using a notebook. How should you complete the code? To answer, drag the appropriate values the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.

Values

activities

broadcast

dependencies

execute

notebooks

runMultiple

Answer Area

```
name : Load_Salesperson ,
"path": "Load_Salesperson",
"timeoutPerCellInSeconds": 300,
},
{
"name": "Load_Orders",
"path": "Load_Orders",
"timeoutPerCellInSeconds": 600,
"  ": ["Load_Salesperson"]
}
},
"timeoutInSeconds": 43200
}
mssparkutils.notebook.  (DAG)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Values

activities

broadcast

dependencies

execute

notebooks

runMultiple

Answer Area

```
name : Load_Salesperson ,
"path": "Load_Salesperson",
"timeoutPerCellInSeconds": 300,
},
{
"name": "Load_Orders",
"path": "Load_Orders",
"timeoutPerCellInSeconds": 600,
"  dependencies  ": ["Load_Salesperson"]
}
},
"timeoutInSeconds": 43200
}
mssparkutils.notebook.  runMultiple  (DAG)
```

NEW QUESTION 54

DRAG DROP - (Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

Values

Row_cumsum

Row_rank_dense

Row_rank_min

Row_window_session

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 59

- (Topic 3)

You are developing a data pipeline named Pipeline1.

You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse. 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 63

- (Topic 3)

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

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

requirements:

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

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

What should you do?

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

Answer: B

Explanation:

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

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

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

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

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

NEW QUESTION 66

- (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 71

DRAG DROP - (Topic 3)

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

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

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

You need to set the Customerkey column as a primary key of the DimCustomer table. Which three code segments should you run in sequence? To answer, move the appropriate

code segments from the list of code segments to the answer area and arrange them in the correct order.

Code Segments

0

⋮

DROP CONSTRAINT PK_DimCustomer

0

⋮

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)

0

⋮

NOT ENFORCED

0

⋮

ALTER TABLE dbo.DimCustomer

0

⋮

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)

0

⋮

ENFORCED

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Code Segments

0

⋮

DROP CONSTRAINT PK_DimCustomer

0

⋮

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)

0

⋮

NOT ENFORCED

0

⋮

ALTER TABLE dbo.DimCustomer

0

⋮

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)

0

⋮

ENFORCED

NEW QUESTION 74

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 75

- (Topic 3)
You are developing a data pipeline named Pipeline1.
You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse.
What should you configure?

- A. Degree of copy parallelism

- B. Fault tolerance
- C. Enable staging
- D. Enable logging

Answer: C

Explanation:

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

NEW QUESTION 76

HOTSPOT - (Topic 3)

You have a Fabric workspace named Workspace1_DEV that contains the following items: 10 reports
Four notebooks Three lakehouses Two data pipelines
Two Dataflow Gen1 dataflows Three Dataflow Gen2 dataflows
Five semantic models that each has a scheduled refresh policy
You create a deployment pipeline named Pipeline1 to move items from Workspace1_DEV to a new workspace named Workspace1_TEST.
You deploy all the items from Workspace1_DEV to Workspace1_TEST.
For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input checked="" type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="radio"/>

NEW QUESTION 81

- (Topic 3)

You have a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Table1.
You analyze Table1 and discover that Table1 contains 2,000 Parquet files of 1 MB each. You need to minimize how long it takes to query Table1.

What should you do?

- A. Disable V-Order and run the OPTIMIZE command.
- B. Disable V-Order and run the VACUUM command.
- C. Run the OPTIMIZE and VACUUM commands.

Answer: C

Explanation:

Problem Overview:

Table1 has 2,000 small Parquet files (1 MB each).

Query performance suffers when the table contains numerous small files because the query engine must process each file individually, leading to significant overhead.

Solution:

To improve performance, file compaction is necessary to reduce the number of small files and create larger, optimized files.

Commands and Their Roles: OPTIMIZE Command:

- Compacts small Parquet files into larger files to improve query performance.
 - It supports optional features like V-Order, which organizes data for efficient scanning.
- VACUUM Command:
- Removes old, unreferenced data files and metadata from the Delta table.
 - Running VACUUM after OPTIMIZE ensures unnecessary files are cleaned up, reducing storage overhead and improving performance.

NEW QUESTION 85

DRAG DROP - (Topic 3)

Your company has a team of developers. The team creates Python libraries of reusable code that is used to transform data.

You create a Fabric workspace name Workspace1 that will be used to develop extract, transform, and load (ETL) solutions by using notebooks.

You need to ensure that the libraries are available by default to new notebooks in Workspace1.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

- 0

⋮

Change the runtime version.
- 0

⋮

Install the libraries.
- 0

⋮

Create a pool.
- 0

⋮

Create an environment.
- 0

⋮

Set the default environment.

Answer Area

- 0
- 0
- 0

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions

- 0 ☐ Change the runtime version.
- 0 ☐ Install the libraries.
- 0 ☐ Create a pool.
- 0 ☐ Create an environment.
- 0 ☐ Set the default environment.

Answer Area

- 0 ☐ Create an environment.
- 0 ☐ Install the libraries.
- 0 ☐ Set the default environment.

NEW QUESTION 88

- (Topic 3)

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

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

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

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks

Timestamp

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

Solution: You use the following code segment:

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

Does this meet the goal?

- A. Yes
- B. no

Answer: B

Explanation:

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

Correct code should look like:

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

NEW QUESTION 90

HOTSPOT - (Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

(SO.ModifiedDate) AS OrderDate

SELECT CAST

SELECT CONVERT

SELECT YEAR

,P.Name AS ProductName

,SUM(SO.OrderQty) AS OrderQty

FROM [dbo].[SalesOrderDetail] SO

INNER JOIN [dbo].[Product] P

ON P.ProductID = SO.ProductID

GROUP BY

CUBE(YEAR(SO.ModifiedDate), P.Name)

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

ROLLUP(YEAR(SO.ModifiedDate), P.Name)

YEAR(SO.ModifiedDate), P.Name

ORDER BY OrderDate

NEW QUESTION 95
HOTSPOT - (Topic 3)

Your company has three newly created data engineering teams named Team1, Team2, and Team3 that plan to use Fabric. The teams have the following personas:

- Team1 consists of members who currently use Microsoft Power BI. The team wants to transform data by using by a low-code approach.
- Team2 consists of members that have a background in Python programming. The team wants to use PySpark code to transform data.
- Team3 consists of members who currently use Azure Data Factory. The team wants to move data between source and sink environments by using the least amount of effort.

You need to recommend tools for the teams based on their current personas.

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

NOTE: Each correct selection is worth one point.

Answer Area

Team1:

Dataflow Gen2 dataflows

Data pipelines

Notebooks

Dataflow Gen2 dataflows

Team2:

Notebooks

Data pipelines

Notebooks

Dataflow Gen2 dataflows

Team3:

Data pipelines

Data pipelines

Notebooks

Dataflow Gen2 dataflows

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Team1:

Dataflow Gen2 dataflows

Data pipelines

Notebooks

Dataflow Gen2 dataflows

Team2:

Notebooks

Data pipelines

Notebooks

Dataflow Gen2 dataflows

Team3:

Data pipelines

Data pipelines

Notebooks

Dataflow Gen2 dataflows

NEW QUESTION 100

- (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, and Git repository
- 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 102

- (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 103

- (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 108

- (Topic 3)

You have a Fabric workspace named Workspace1 that contains an Apache Spark job definition named Job1.

You have an Azure SQL database named Source1 that has public internet access disabled.

You need to ensure that Job1 can access the data in Source1. What should you create?

- A. an on-premises data gateway
- B. a managed private endpoint
- C. an integration runtime
- D. a data management gateway

Answer: B

Explanation:

To allow Job1 in Workspace1 to access an Azure SQL database (Source1) with public internet access disabled, you need to create a managed private endpoint. A managed private endpoint is a secure, private connection that enables services like Fabric (or other Azure services) to access resources such as databases, storage accounts, or other services within a virtual network (VNet) without requiring public internet access. This approach maintains the security and integrity of your data while enabling access to the Azure SQL database.

NEW QUESTION 109

HOTSPOT - (Topic 3)

You are processing streaming data from an external data provider. You have the following code segment.

```
datatable (Location:string, Company:string, UnitsSold:long)
[
  "New York", "Contoso", 300,
  "New York", "Litware", 1000,
  "New York", "Relecloud", 300,
  "New York", "Fabrikam", 200,
  "Seattle", "Contoso", 300,
  "Seattle", "Litware", 100,
  "Seattle", "Fabrikam", 100,
  "San Francisco", "Relecloud", 500,
  "San Francisco", "Litware", 500,
  "Washington DC", "Litware", 300,
  "Washington DC", "Contoso", 400
]
| sort by Location desc, UnitsSold desc
| extend Rank=row_rank_dense(UnitsSold, prev(Location) != Location)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
Litware from New York will be displayed at the top of the result set.	<input type="radio"/>	<input type="radio"/>
Fabrikam in Seattle will have value = 2 in the Rank column.	<input checked="" type="radio"/>	<input type="radio"/>
Litware in San Francisco will have the same value in the Rank column as Litware in New York.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Litware from New York will be displayed at the top of the result set – Yes
The data is sorted first by Location in descending order and then by UnitsSold in descending order. Since "New York" is alphabetically the last Location, it will appear first in the result set. Within "New York", Litware has the highest UnitsSold (1000), so it will be displayed at the top.
Fabrikam in Seattle will have value = 2 in the Rank column – No
The row_rank_dense function assigns dense ranks based on UnitsSold within each location. In "Seattle":
Contoso has UnitsSold = 300 Rank 1 Litware has UnitsSold = 100 Rank 2
Fabrikam also has UnitsSold = 100, so it shares the same rank (2) as Litware.
Litware in San Francisco will have the same value in the Rank column as Litware in New York – No
The rank is calculated separately for each location. In "San Francisco":
Both Relecloud and Litware have UnitsSold = 500, so they share the same rank (1). In "New York", Litware has the highest UnitsSold = 1000 Rank 1.
Since ranks are calculated independently for each location, Litware in San Francisco does not share the same rank as Litware in New York.

NEW QUESTION 113

- (Topic 3)
You have a Fabric workspace that contains a lakehouse named 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. paraml"
- C. "@{pipeline().parameters.paraml}"
- D. "@{pipeline().parameters.[paraml]}"
- E. "@{pipeline().parameters.paraml}-

Answer: B

NEW QUESTION 117

.....

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](#)