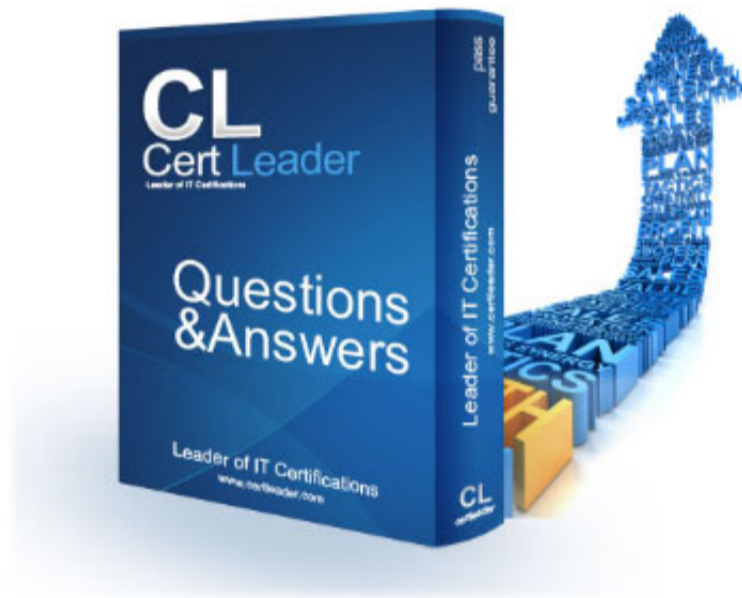


DP-700 Dumps

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

<https://www.certleader.com/DP-700-dumps.html>



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

HOTSPOT - (Topic 1)

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

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

NOTE: Each correct selection is worth one point.

Bronze layer:

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

Silver layer:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Bronze Layer: A pipeline Copy activity

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

Silver Layer: A notebook

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

NEW QUESTION 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

HOTSPOT - (Topic 2)

You need to troubleshoot the ad-hoc query issue.

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

NOTE: Each correct selection is worth one point.

```
SELECT last_run_start_time, last_run_command
```

FROM

queryinsights.exec_requests_history
queryinsights.exec_sessions_history
queryinsights.frequently_run_queries
queryinsights.long_running_queries

```
WHERE last_run_total_elapsed_time_ms > 7200000
```

AND

max_run_total_elapsed_time_ms > 7200000
median_total_elapsed_time_ms > 7200000
number_of_canceled_runs > 1
number_of_failed_runs > 1
number_of_runs > 1

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

SELECT last_run_start_time, last_run_command: These fields will help identify the execution details of the long-running queries.

FROM queryinsights.long_running_queries: The correct solution is to check the long- running queries using the queryinsights.long_running_queries view, which provides insights into queries that take longer than expected to execute.

WHERE last_run_total_elapsed_time_ms > 7200000: This condition filters queries that took more than 2 hours to complete (7200000 milliseconds), which is relevant to the issue described.

AND number_of_failed_runs > 1: This condition is key for identifying queries that have failed more than once, helping to isolate the problematic queries that cause failures and need attention.

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

HOTSPOT - (Topic 3)

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

You have the Fabric workspaces shown in the following table.

Name	Workspace admin
Workspace1	User1
Workspace2	User2

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

Name	Domain admin
Domain1	User1
Domain2	User2

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

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

NOTE: Each correct selection is worth one point.

Answer Area

Statements

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements

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

NEW QUESTION 9

- (Topic 3)

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

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

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

Which view should you use?

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

Answer: E

Explanation:

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

NEW QUESTION 10

- (Topic 3)

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

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

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

Trigger the process when a new file is added.

Provide the highest throughput.

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

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

Answer: A

Explanation:

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

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

NEW QUESTION 10

- (Topic 3)

You have a Fabric workspace 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 15

- (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 16

- (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 21

- (Topic 3)

You are implementing a medallion architecture in a Fabric lakehouse.

You plan to create a dimension table that will contain the following columns:

- ID
- CustomerCode
- CustomerName
- CustomerAddress
- CustomerLocation
- ValidFrom
- ValidTo

You need to ensure that the table supports the analysis of historical sales data by customer location at the time of each sale. Which type of slowly changing dimension (SCD) should you use?

- A. Type 2
- B. Type 0
- C. Type 1
- D. Type 3

Answer: A

NEW QUESTION 23

- (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 26

- (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 31

- (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 35

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

Answer Area

- ☰ 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.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions	Answer Area
<ul style="list-style-type: none"> Add a Get metadata activity to query Control.Object and generate a list of schemas and tables to copy. 	<ul style="list-style-type: none"> Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.
<ul style="list-style-type: none"> Add an Until activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables. 	<ul style="list-style-type: none"> Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
<ul style="list-style-type: none"> Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy. 	
<ul style="list-style-type: none"> Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables. 	<ul style="list-style-type: none"> Add a Copy data activity as an inner activity to the iterator activity.
<ul style="list-style-type: none"> Add a Copy data activity as an inner activity to the iterator activity. 	

NEW QUESTION 40

HOTSPOT - (Topic 3)

You have a Fabric workspace.

You are debugging a statement and discover the following issues: Sometimes, the statement fails to return all the expected rows.

The PurchaseDate output column is NOT in the expected format of mmm dd, yy.

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

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

NOTE: Each correct selection is worth one point.

Answer Area

SELECT

item_id as ItemId

▼ as ItemName

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

,item_description as ItemDescription

▼ as PurchaseDate

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

FROM

Table1

WHERE

item_type = @itemtype_parameter

- A. Mastered
- B. Not Mastered

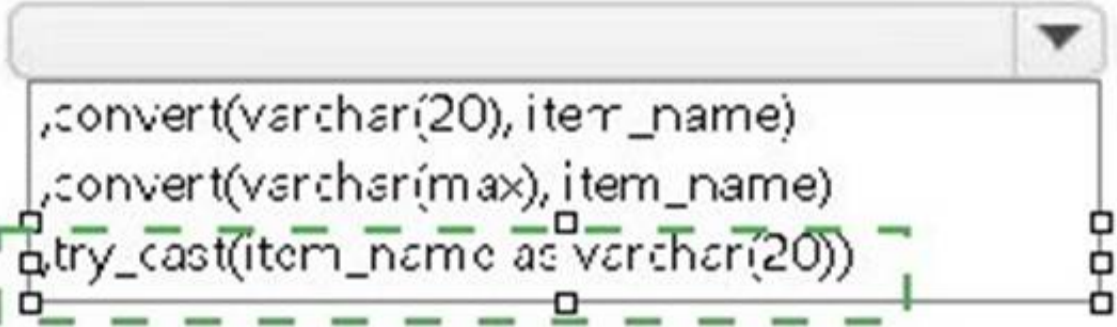
Answer: A

Explanation:

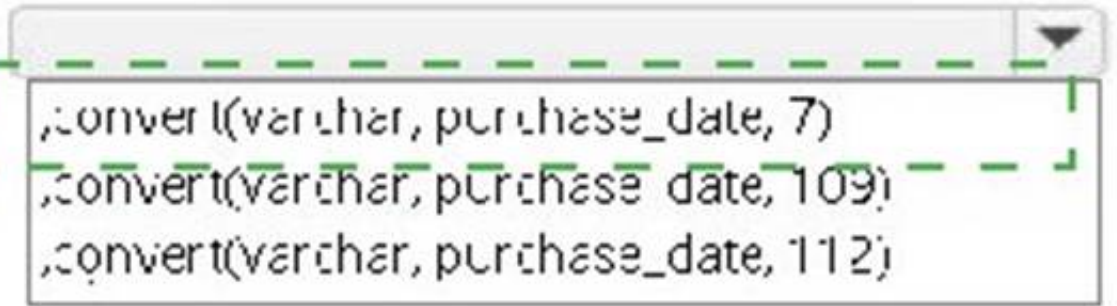
Answer Area

SELECT

item_id as ItemId

 as ItemName

,item_description as ItemDescription

 as PurchaseDate

FROM

Table1

WHERE

item_type = @itemtype_parameter

NEW QUESTION 45

- (Topic 3)

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table named Table1 in a lakehouse. The streaming data is sourced from motorway sensors and represents the speed of cars.

You need to add a transformation to EventStream1 to average the car speeds. The speeds must be grouped by non-overlapping and contiguous time intervals of one minute. Each event must belong to exactly one window.

Which windowing function should you use?

- A. sliding
- B. hopping
- C. tumbling
- D. session

Answer: C

NEW QUESTION 50

- (Topic 3)

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

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

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

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp

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

Solution: You use the following code segment:

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

Does this meet the goal?

- A. Yes
- B. no

Answer: A

Explanation:

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

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

No_Bikes asc.

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

NEW QUESTION 53

- (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 56

HOTSPOT - (Topic 3)

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

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

The solution must minimize development effort.

What should you include in the recommendation for each source? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Source1:

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

Source2:

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

Source3:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Source1:

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

Source2:

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

Source3:

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

NEW QUESTION 58

- (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 62

- (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 64

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	Answer Area
activities	<pre> name : Load_Salesperson , "path": "Load_Salesperson", "timeoutPerCellInSeconds": 300, }, { "name": "Load_Orders", "path": "Load_Orders", "timeoutPerCellInSeconds": 600, " [] ": ["Load_Salesperson"] } }, "timeoutInSeconds": 43200 } mssparkutils.notebook. [] (DAG) </pre>
broadcast	
dependencies	
execute	
notebooks	
runMultiple	

- 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 69

- (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 70

- (Topic 3)

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

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

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

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

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

Provide the best query performance. Minimize operational costs.

Which should you do?

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

Answer: D

Explanation:

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

NEW QUESTION 72

- (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 73

- (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 74

- (Topic 3)

You have a Fabric warehouse named DW1 that contains a Type 2 slowly changing dimension (SCD) dimension table named DimCustomer. DimCustomer contains 100 columns and 20 million rows. The columns are of various data types, including int, varchar, date, and varbinary. You need to identify incoming changes to the table and update the records when there is a change. The solution must minimize resource consumption. What should you use to identify changes to attributes?

- A. a direct attributes comparison for the attributes in the source table.
- B. a hash function to compare the attributes in the DimCustomer table.
- C. a direct attributes comparison across the attributes in the DimCustomer table.
- D. a hash function to compare the attributes in the source table.

Answer: D

NEW QUESTION 75

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse!. Warehouse! 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()' )
```



```
ALTER TABLE DimCustomer
ALTER COLUMN EmailAddress
ADD MASKED WITH (FUNCTION = 'random (1, "@", 5)' )
```

NEW QUESTION 76

- (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 77

- (Topic 3)

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

You have a Fabric workspace named Workspace!. Workspace! contains a notebook named Notebook1 that performs the following tasks:

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

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token. Solution: You use the following code segment:
Use notebookutils.credentials.getSecret and specify key vault URL and the name of a linked service.
Does this meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 78

- (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 81

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 82

- (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 85

- (Topic 3)

You have a Google Cloud Storage (GCS) container named storage1 that contains the files shown in the following table.

Name	Size
ProductFile.parquet	8 MB
StoreFile.json	500 MB
TripsFile.csv	99 MB

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

Name	Source	Last accessed
Products	ProductFile	12 hours ago
Stores	StoreFile	4 hours ago
Trips	TripsFile	48 hours ago

You need to read data from all the shortcuts. Which shortcuts will retrieve data from the cache?

- A. Stores only
- B. Products only
- C. Stores and Products only
- D. Products, Stores, and Trips
- E. Trips only
- F. Products and Trips only

Answer: C

Explanation:

When reading data from shortcuts in Fabric (in this case, from a lakehouse like Lakehouse1), the cache for shortcuts helps by storing the data locally for quick access. The last accessed timestamp and the cache expiration rules determine whether data is fetched from the cache or from the source (Google Cloud Storage, in this case).

Products: The ProductFile.parquet was last accessed 12 hours ago. Since the cache has data available for up to 12 hours, it is likely that this data will be retrieved from the cache, as it hasn't been too long since it was last accessed.

Stores: The StoreFile.json was last accessed 4 hours ago, which is within the cache retention period. Therefore, this data will also be retrieved from the cache.

Trips: The TripsFile.csv was last accessed 48 hours ago. Given that it's outside the typical caching window (assuming the cache has a maximum retention period of around 24 hours), it would not be retrieved from the cache. Instead, it will likely require a fresh read from the source.

NEW QUESTION 90

- (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 95

HOTSPOT - (Topic 3)

You have an Azure Event Hubs data source that contains weather data.

You ingest the data from the data source by using an eventstream named Eventstream1. Eventstream1 uses a lakehouse as the destination.

You need to batch ingest only rows from the data source where the City attribute has a value of Kansas. The filter must be added before the destination. The solution must minimize development effort.

What should you use for the data processor and filtering? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Data processor:

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

Filtering:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Data processor:

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

Filtering:

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

NEW QUESTION 97

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 101

- (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 104

- (Topic 3)

You have two Fabric workspaces named Workspace1 and Workspace2.

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

You recently modified the items in Workspaces1.

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

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

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

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

What should you do?

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

Answer: D

Explanation:

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

NEW QUESTION 105

- (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
- 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 109

- (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 114

HOTSPOT - (Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

Statements

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

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

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

<input type="checkbox"/>	<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 117

- (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 119

- (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 121

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-700 Exam with Our Prep Materials Via below:

<https://www.certleader.com/DP-700-dumps.html>