

Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)



NEW QUESTION 1

- (Topic 1)

You need to ensure that the data analysts can access the gold layer lakehouse. What should you do?

- A. Add the DataAnalyst group to the Viewer role for WorkspaceA.
- B. Share the lakehouse with the DataAnalysts group and grant the Build reports on the default semantic model permission.
- C. Share the lakehouse with the DataAnalysts group and grant the Read all SQL Endpoint data permission.
- D. Share the lakehouse with the DataAnalysts group and grant the Read all Apache Spark permission.

Answer: C

Explanation:

Data Analysts' Access Requirements must only have read access to the Delta tables in the gold layer and not have access to the bronze and silver layers. The gold layer data is typically queried via SQL Endpoints. Granting the Read all SQL Endpoint data permission allows data analysts to query the data using familiar SQL-based tools while restricting access to the underlying files.

NEW QUESTION 2

- (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 3

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 4

- (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 5

- (Topic 3)

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

Street Neighbourhood

Latitude Longitude No_Bikes No_Empty_Docks

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

What should you use?

- A. a KQL queryset

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

Answer: B

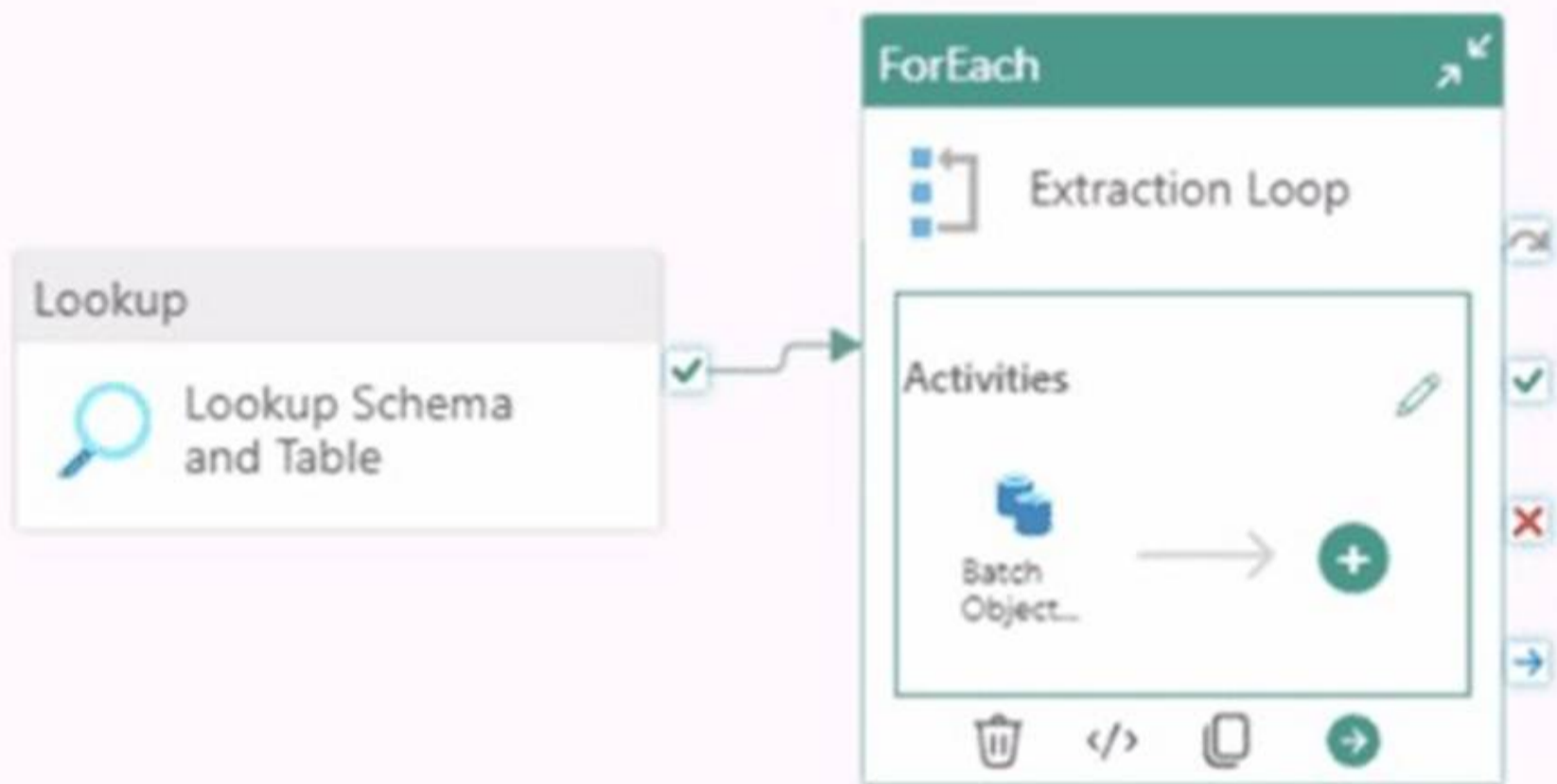
Explanation:

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

NEW QUESTION 6

HOTSPOT - (Topic 3)

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



General **Settings** Activities (1)

Batch count ⓘ

Items *

This property should be parameterized.

Add dynamic content [Alt+Shift+D]

Dynamic Data Copy does NOT use parametrization.

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

NOTE: Each correct selection is worth one point.

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 7

- (Topic 3)

You have a Fabric warehouse named DW1 that loads data by using a data pipeline named Pipeline1. Pipeline1 uses a Copy data activity with a dynamic SQL source. Pipeline1 is scheduled to run every 15 minutes.

You discover that Pipeline1 keeps failing.

You need to identify which SQL query was executed when the pipeline failed. What should you do?

- A. From Monitoring hub, select the latest failed run of Pipeline1, and then view the output JSON.
- B. From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.
- C. From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemReadFailed.
- D. From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemUpdateFailed.
- E. From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemReadFailed.

Answer: B

Explanation:

The input JSON contains the configuration details and parameters passed to the Copy data activity during execution, including the dynamically generated SQL query.

Viewing the input JSON for the failed pipeline run provides direct insight into what query was executed at the time of failure.

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 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 named Workspacel that contains the following items:

- A Microsoft Power BI report named Reportl
- A Power BI dashboard named Dashboardl
- A semantic model named Modell
- A lakehouse name Lakehouse1

Your company requires that specific governance processes be implemented for the items. Which items can you endorse in Fabric?

- A. Lakehouse1, Modell, and Dashboard1 only
- B. Lakehouse1, Modell, Report1 and Dashboard1
- C. Report1 and Dashboard1 only
- D. Model1, Report1, and Dashboard1 only
- E. Lakehouse1, Model1, and Report1 only

Answer: B

NEW QUESTION 10

HOTSPOT - (Topic 3)

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

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

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

Answer Area

GRANT

- ALTER
- EXECUTE
- READ
- SELECT
- VIEW

Customers(CustomerID, FirstName, LastName, Phone)

TO

- User1
- [User1]
- [user1@contoso.com]

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 15

- (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 17

- (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 20

- (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

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains two lakehouses named Lakehouse1 and Lakehouse2. Lakehouse1 contains staging data in a Delta table named Orderlines. Lakehouse2 contains a Type 2 slowly changing dimension (SCD) dimension table named Dim_Customer.

You need to build a query that will combine data from Orderlines and Dim_Customer to create a new fact table named Fact_Orders. The new table must meet the following requirements:

Enable the analysis of customer orders based on historical attributes. Enable the analysis of customer orders based on the current attributes.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
SELECT
    OrderLineID order_line_id
    ,OrderDate order_date
    ,c.customer_key
    ,c.customer_id
    ,Quantity order_quantity
    ,unitPrice unit_price
    ,taxRate tax_rate
FROM
    Lakehouse1.orderlines o
INNER JOIN
    Lakehouse2.dim_customer c
    ON o.customerid = c.customer_id

AND 
    o.OrderDate >= valid_to_datetime
    o.OrderDate >= valid_from_datetime

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

SELECT

```
OrderLineID order_line_id
,OrderDate order_date
,c.customer_key
,c.customer_id
,Quantity order_quantity
,unitPrice unit_price
,taxRate tax_rate
```

FROM

```
Lakehouse1.orderlines o
```

INNER JOIN

```
Lakehouse2.dim_customer c
ON o.customerid = c.customer_id
```

AND

```
c.is_current = 1
o.OrderDate <= c.valid_to_datetime
o.OrderDate >= c.valid_from_datetime
```

AND

```
c.is_current = 1
o.OrderDate <= c.valid_to_datetime
o.OrderDate <= c.valid_from_datetime
```

NEW QUESTION 28

- (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 29

HOTSPOT - (Topic 3)

You have a Fabric workspace.

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

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

SELECT

item_id as ItemId

▼

```

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

as ItemName

,item_description as ItemDescription

▼

```

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

as PurchaseDate

FROM

Table1

WHERE

item_type = @itemtype_parameter

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

SELECT

item_id as ItemId

as ItemName
 ,convert(varchar(20), item_name)
 ,convert(varchar(max), item_name)
 ,try_cast(item_name as varchar(20))
 ,item_description as ItemDescription

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

FROM

Table1

WHERE

item_type = @itemtype_parameter

NEW QUESTION 30

DRAG DROP - (Topic 3)

You are implementing the following data entities in a Fabric environment:

Entity1: Available in a lakehouse and contains data that will be used as a core organization entity

Entity2: Available in a semantic model and contains data that meets organizational standards

Entity3: Available in a Microsoft Power BI report and contains data that is ready for sharing and reuse

Entity4: Available in a Power BI dashboard and contains approved data for executive-level decision making

Your company requires that specific governance processes be implemented for the data. You need to apply endorsement badges to the entities based on each entity's use case.

Which badge should you apply to each entity? To answer, drag the appropriate badges to the correct entities. Each badge may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Badges

- Certified
- Master data
- Promoted
- Cannot be endorsed

Answer Area

- Entity1:
- Entity2:
- Entity3:
- Entity4:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Badges



Answer Area



NEW QUESTION 35

- (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 36

HOTSPOT - (Topic 3)

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

- Key
- Status
- LastModified

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

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

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

NOTE: Each correct selection is worth one point.

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

NEW QUESTION 43

- (Topic 3)

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

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

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

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

Answer: B

Explanation:

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

NEW QUESTION 46

- (Topic 3)

You have a Fabric workspace that contains a semantic model named Modell. 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 48

- (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 50

HOTSPOT - (Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

Dataset1:

Dataset2:

Dataset3:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Dataset1:

Dataset2:

Dataset3:

NEW QUESTION 52

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,
"dependencies": ["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 56

- (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 61

- (Topic 3)

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

During the last run, Notebook1 executed nine jobs. You need to view the jobs in a timeline chart. What should you use?

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

Answer: E

Explanation:

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

NEW QUESTION 66

- (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 69

- (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 70

- (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 72

- (Topic 3)

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

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

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

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

Answer: B

Explanation:

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

NEW QUESTION 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 79

- (Topic 3)

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

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

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

Answer: C

NEW QUESTION 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 84

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 88

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

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

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions

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

Answer Area

Create an environment.

Install the libraries.

Set the default environment.

NEW QUESTION 89

HOTSPOT - (Topic 3)

You have a Fabric warehouse named DW1 that contains four staging tables named ProductCategory, ProductSubcategory, Product, and SalesOrder. ProductCategory, ProductSubcategory, and Product are used often in analytical queries.

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

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

NOTE: Each correct selection is worth one point.

Answer Area

ProductCategory, ProductSubcategory and Product must be:

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

The joining key must be:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

ProductCategory, ProductSubcategory and Product must be:

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

The joining key must be:

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

NEW QUESTION 93

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(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(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

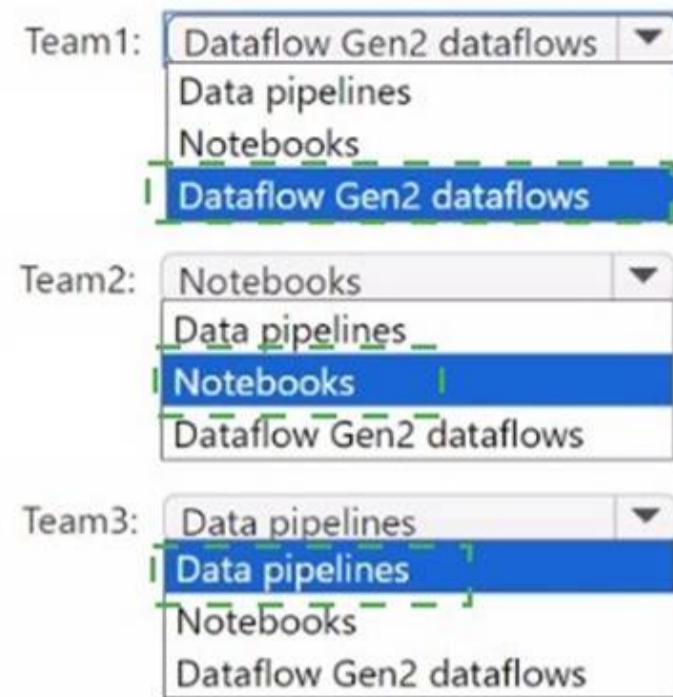
Team3: Data pipelines Data pipelines Notebooks Dataflow Gen2 dataflows

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 96

- (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 folder
- 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 98

- (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 100

- (Topic 3)

You have a Fabric deployment pipeline that uses three workspaces named Dev, Test, and Prod.

You need to deploy an eventhouse as part of the deployment process. What should you use to add the eventhouse to the deployment process?

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

Answer: B

Explanation:

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

NEW QUESTION 101

- (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 104

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 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 106

- (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 111

- (Topic 3)

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

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

Pipeline1!. The param1 parameter has a data type of int

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

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

Answer: B

NEW QUESTION 114

- (Topic 3)

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

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

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

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

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

A shortcut to TripsFile aliased as Trips

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

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

Answer: B

Explanation:

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

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

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

NEW QUESTION 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

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](#)