Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

NEW QUESTION 1
HOTSPOT - (Topic 1)
You need to create the product dimension.
How should you complete the Apache Spark SQL code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
SELECT ProductID, ProductNumber, ProductName, ModelName, SubCategoryName, CategoryName

FROM ContosoLake.Products p
```

▼	ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID
FULL JOIN	
INNER JOIN	
LEFT ANTI JOIN	
LEFT OUTER JOIN	
OUTER JOIN	

▼	ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID
FULL JOIN	
INNER JOIN	
LEFT ANTI JOIN	
LEFT OUTER JOIN	
OUTER JOIN	

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 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 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 5
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.	⊙	○
User3 has Domain contributor access to Domain1.	○	○
User2 has Contributor role access to Workspace3.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area

Statements	Yes	No
User3 has Viewer role access to Workspace3.	⊙	○
User3 has Domain contributor access to Domain1.	○	⊙
User2 has Contributor role access to Workspace3.	○	⊙

NEW QUESTION 6
- (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-5werade642f6",`
		`    "temperature": 27`
		`  }`
		`]`

C)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)

2024-05-23	81416f3060a24e759b192a84ea05973532dhdyte3	`[`
		`  {`
		`    "index": 0,`
		`    "eventid": "719afca0-be30-4559-bb5e-59feade642f6"`
		`  }`
		`]`

D)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)

2024-05-24	81416f30-60a2-4e75-9b19-2a84ea059735	<pre>[
		  {
		    "index": 0,
		    "eventid": "719afca0-be30-4559-bb5e-59feade642f6"
		  }
		]</pre>

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 7
- (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 8
- (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 9
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a warehouse named DW1 and a data pipeline named Pipeline1.
You plan to add a user named User3 to Workspace1.
You need to ensure that User3 can perform the following actions: View all the items in Workspace1.

Update the tables in DW1.
The solution must follow the principle of least privilege.
You already assigned the appropriate object-level permissions to DW1. Which workspace role should you assign to User3?

A. Admin
B. Member
C. Viewer
D. Contributor

Answer: D

Explanation:
To ensure User3 can view all items in Workspace1 and update the tables in DW1, the most appropriate workspace role to assign is the Contributor role. This role allows User3 to: View all items in Workspace1: The Contributor role provides the ability to view all objects within the workspace, such as data pipelines, warehouses, and other resources.
Update the tables in DW1: The Contributor role allows User3 to modify or update resources within the workspace, including the tables in DW1, assuming that appropriate object-level permissions are set for the warehouse.
This role adheres to the principle of least privilege, as it provides the necessary permissions without granting broader administrative rights.

NEW QUESTION 10
- (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 10
HOTSPOT - (Topic 3)
You have a Fabric workspace named Workspace1 that contains the items shown in the following table.

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

For Model1, the Keep your Direct Lake data up to date option is disabled.
You need to configure the execution of the items to meet the following requirements:
Notebook1 must execute every weekday at 8:00 AM.
Notebook2 must execute when a file is saved to an Azure Blob Storage container. Model1 must refresh when Notebook1 has executed successfully.
How should you orchestrate each item? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

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

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

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

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

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

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

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

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

NEW QUESTION 13
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: A streaming dataflow ▼
  Apache Spark Structured Streaming
  An eventstream
  A data pipeline
  A streaming dataflow
  An eventstream ▼

Source2: A data pipeline ▼
  Apache Spark Structured Streaming
  An eventstream
  A data pipeline
  A streaming dataflow

Source3: An eventstream ▼
  Apache Spark Structured Streaming
  An eventstream
  A data pipeline
  A streaming dataflow

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Source1:	A streaming dataflow ▼
	Apache Spark Structured Streaming
	An eventstream
	A data pipeline
	A streaming dataflow
	An eventstream ▼

Source2:	A data pipeline ▼
	Apache Spark Structured Streaming
	An eventstream
	A data pipeline
	A streaming dataflow

Source3:	An eventstream ▼
	Apache Spark Structured Streaming
	An eventstream
	A data pipeline
	A streaming dataflow

NEW QUESTION 18
HOTSPOT - (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 contains the following tables and columns.

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

You need to denormalize the tables and include the ContractType and StartDate columns
in the Employee table. The solution must meet the following requirements:
Ensure that the StartDate column is of the date data type.
Ensure that all the rows from the Employee table are preserved and include any matching rows from the Contract table.
Ensure that the result set displays the total number of employees per contract type for all the contract types that have more than two employees.
How should you complete the statement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
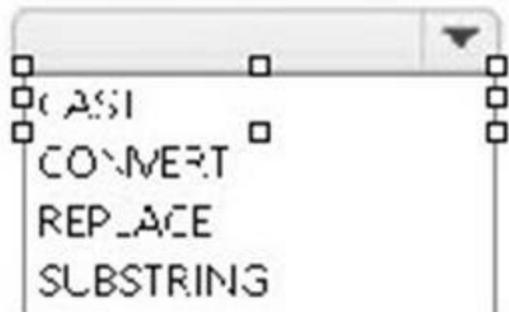
Answer Area

```
WITH result AS(

SELECT e.EmployeeID

    , e.EmployeeName

    , e.EmployeePosition

    , c.ContractType

    ,  [ ▼ ]        (date, c.StartDate) as StartDate
       ┌──────────┐
       │ CASE     │
       │ CONVERT  │
       │ REPLACE  │
       │ SUBSTRING│
       └──────────┘

FROM Employee AS e

[ ▼ ]            Contract AS c on c.EmployeeID = e.EmployeeID
┌──────────────────┐
│ CROSS JOIN       │
│ INNER JOIN       │
│ LEFT OUTER JOIN  │
│ RIGHT OUTER JOIN │
└──────────────────┘
)

SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

   , ContractType

FROM result

GROUP BY ContractType

[ ▼ ]            COUNT(DISTINCT EmployeeID) > 2
┌──────────┐
│ CONTAINS │
│ HAVING   │
│ LIMIT    │
│ WHERE    │
└──────────┘
```
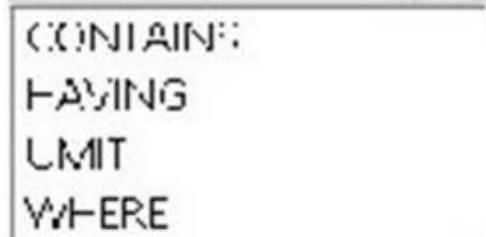
A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

```
WITH result AS(

SELECT e.EmployeeID

  , e.EmployeeName

  , e.EmployeePosition

  , c.ContractType
```

Dropdown options:
- CAST
- CONVERT
- REPLACE
- SUBSTRING

```
  ,                    (date, c.StartDate) as StartDate

FROM Employee AS e
```

Dropdown options:
- CROSS JOIN
- INNER JOIN
- LEFT OUTER JOIN
- RIGHT OUTER JOIN
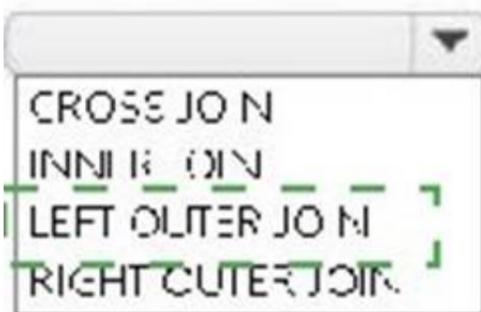
```
         Contract AS c on c.EmployeeID = e.EmployeeID

)

SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

  , ContractType

FROM result

GROUP BY ContractType
```

Dropdown options:
- CONTAINS
- HAVING
- LIMIT
- WHERE

```
                    COUNT(DISTINCT EmployeeID) > 2
```

NEW QUESTION 20
- (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 22

- (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 24
- (Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

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

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

You need to reduce how long it takes to run the KQL queryset. Solution: You change project to extend.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Using extend retains all columns in the table, potentially increasing the size of the output unnecessarily. project is more efficient because it selects only the required columns.

NEW QUESTION 28
- (Topic 3)
You have a Fabric capacity that contains a workspace named Workspace1. Workspace1 contains a lakehouse named Lakehouse1, a data pipeline, a notebook, and several Microsoft Power BI reports.
A user named User1 wants to use SQL to analyze the data in Lakehouse1. You need to configure access for User1. The solution must meet the following requirements:
Provide User1 with read access to the table data in Lakehouse1.
Prevent User1 from using Apache Spark to query the underlying files in Lakehouse1. Prevent User1 from accessing other items in Workspace1.
What should you do?

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

Answer: B

Explanation:
To meet the specified requirements for User1, the solution must ensure:
? Read access to the table data in Lakehouse1: User1 needs permission to access the data within Lakehouse1. By sharing Lakehouse1 with User1 and selecting the Read all SQL endpoint data option, User1 will be able to query the data via SQL endpoints.
? Prevent Apache Spark usage: By sharing the lakehouse directly and selecting the SQL endpoint data option, you specifically enable SQL-based access to the data, preventing User1 from using Apache Spark to query the data.
? Prevent access to other items in Workspace1: Assigning User1 the Viewer role for Workspace1 ensures that User1 can only view the shared items (in this case, Lakehouse1), without accessing other resources such as notebooks, pipelines, or Power BI reports within Workspace1.
This approach provides the appropriate level of access while restricting User1 to only the required resources and preventing access to other workspace assets.

NEW QUESTION 31
- (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 34
HOTSPOT - (Topic 3)
You are building a data loading pattern for Fabric notebook workloads. You have the following code segment:

```
def loading_pattern_sample(df_source):
    try:
        deltaTable = DeltaTable.forName(spark, target_table)
    except Exception:
        try:
            df_source.write.format('delta').mode('overwrite').saveAsTable(f"{target_table}")
        except Exception as e:
            print(f':Load for table {target_table} failed with error: {str(e)}')
            raise
        return

    try:
        change_detection_columns = [col for col in df_source.columns if col not in candidate_key]

        match_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate_key])
        update_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change_detection_columns])

        update_expr = {col: f'source.{col}' for col in df_source.columns}

        merge_operation = deltaTable.alias('target').merge(
            source=df_source.alias('source'),
            condition=match_condition
        ).whenMatchedUpdate(
            condition=update_condition,
            set=update_expr
        ).whenNotMatchedInsertAll()

        merge_operation.execute()
    except Exception as e:
        print(f'Insert operation for table {target_table} failed with error: {str(e)}')
    return
```

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
The target table will always be overwritten.	O	O
The merge operation will always run.	O	O
The loading pattern supports both full and incremental loading requirements.	O	O

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements	Yes	No
The target table will always be overwritten.	○	○
The merge operation will always run.	○	○
The loading pattern supports both full and incremental loading requirements.	○	○

NEW QUESTION 36
HOTSPOT - (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse!. Warehousel contains a table named DimCustomers. DimCustomers contains the following columns:
• CustomerName
• CustomerID
• BirthDate
• Email
You need to configure security to meet the following requirements:
• BirthDate in DimCustomer must be masked and display 1900-01-01.
• Email in DimCustomer must be masked and display only the first leading character and the last five characters.
How should you complete the statement? To answer, select the appropriate options in the
answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
ALTER TABLE DimCustomer

ALTER COLUMN BirthDate

ADD MASKED WITH (FUNCTION =   'default()'          ▼ )
                              'default()'
                              'partial(1900-01-01)'
                              'random(1900-01-01, 1900-01-01)'


ALTER TABLE DimCustomer

ALTER COLUMN EmailAddress

ADD MASKED WITH (FUNCTION =   'random (1, "@", 5)'  ▼ )
                              'default()'
                              'email()'
                              'partial(1, "@",5)'
                              'random (1, "@", 5)'
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

```
ALTER TABLE DimCustomer

ALTER COLUMN BirthDate

ADD MASKED WITH (FUNCTION =     'default()'            ▼  )
                                'default()'
                                'partial(1900-01-01)'
                                'random(1900-01-01, 1900-01-01)'


ALTER TABLE DimCustomer

ALTER COLUMN EmailAddress

ADD MASKED WITH (FUNCTION =     'random (1, "@", 5)'   ▼  )
                                'default()'
                                'email()'
                                'partial(1, "@",5)'
                                'random (1, "@", 5)'
```
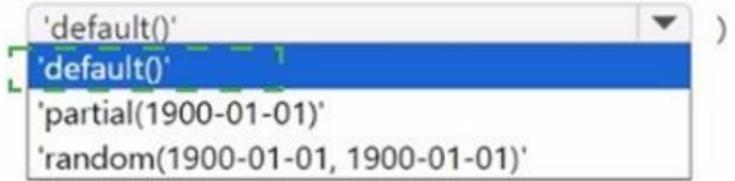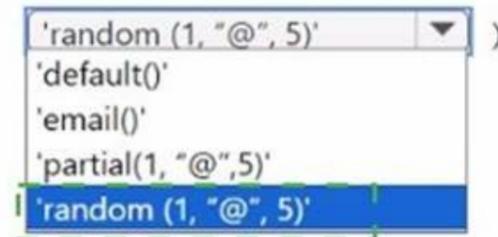
NEW QUESTION 37
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:
Orders
Customer Employee
The Employee table contains Personally Identifiable Information (PII).
A data engineer is building a workflow that requires 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 39
- (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 42
- (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 47
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

(empty)

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 50
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)
      GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
      ROLLUP(YEAR(SO.ModifiedDate), P.Name)
      YEAR(SO.ModifiedDate), P.Name

    ORDER BY OrderDate
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area

```
[_____▼] (SO.ModifiedDate) AS OrderDate
  SELECT CAST
  SELECT CONVERT
  [SELECT YEAR]

        ,P.Name AS ProductName

        ,SUM(SO.OrderQty) AS OrderQty

    FROM [dbo].[SalesOrderDetail] SO

    INNER JOIN [dbo].[Product] P

      ON P.ProductID = SO.ProductID

    GROUP BY

    [_____▼]
      CUBE(YEAR(SO.ModifiedDate), P.Name)
      GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
      [ROLLUP(YEAR(SO.ModifiedDate), P.Name)]
      YEAR(SO.ModifiedDate), P.Name

    ORDER BY OrderDate
```

NEW QUESTION 51
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons 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