



Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

NEW QUESTION 1

- (Topic 1)

You need to populate the MAR1 data in the bronze layer.

Which two types of activities should you include in the pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. ForEach
- B. Copy data
- C. WebHook
- D. Stored procedure

Answer: AB

Explanation:

MAR1 has seven entities, each accessible via a different API endpoint. A ForEach activity is required to iterate over these endpoints to fetch data from each one. It enables dynamic execution of API calls for each entity.

The Copy data activity is the primary mechanism to extract data from REST APIs and load it into the bronze layer in Delta format. It supports native connectors for REST APIs and Delta, minimizing development effort.

You need to schedule the population of the medallion layers to meet the technical requirements.

What should you do?

- * A. Schedule a data pipeline that calls other data pipelines.
- * B. Schedule a notebook.
- * C. Schedule an Apache Spark job.
- * D. Schedule multiple data pipelines.

* Answer: A

The technical requirements specify that:

Medallion layers must be fully populated sequentially (bronze silver gold). Each layer must be populated before the next.

If any step fails, the process must notify the data engineers. Data imports should run simultaneously when possible.

Why Use a Data Pipeline That Calls Other Data Pipelines?

A data pipeline provides a modular and reusable approach to orchestrating the sequential population of medallion layers.

By calling other pipelines, each pipeline can focus on populating a specific layer (bronze, silver, or gold), simplifying development and maintenance.

A parent pipeline can handle:

- Sequential execution of child pipelines.
- Error handling to send email notifications upon failures.
- Parallel execution of tasks where possible (e.g., simultaneous imports into the bronze layer).

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

DRAG DROP - (Topic 2)

You need to ensure that the authors can see only their respective sales data.

How should you complete the statement? To answer, drag the appropriate values to 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

- AuthorSales
- AuthorEmail
- AuthorSales.AuthorEmail
- BLOCK
- FILTER
- INLINE
- SCHEMABINDING
- USER_NAME()

Answer Area

```
CREATE FUNCTION dbo.tvf_riSPredicate(@Author AS varchar(50))
    RETURNS TABLE
WITH
AS
    RETURN SELECT 1 AS tvf_riSPredicate_result
WHERE @Author =
GO

CREATE SECURITY POLICY RLSFilter
ADD FILTER PREDICATE Security.tvf_riSPredicate(AuthorEmail)
ON
WITH (STATE = ON)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Values

- AuthorSales
- AuthorEmail
- AuthorSales.AuthorEmail
- BLOCK
- FILTER
- INLINE
- SCHEMABINDING
- USER_NAME()

Answer Area

```
CREATE FUNCTION dbo.tvf_riSPredicate(@Author AS varchar(50))
    RETURNS TABLE
WITH
AS
    RETURN SELECT 1 AS tvf_riSPredicate_result
WHERE @Author =
GO

CREATE SECURITY POLICY RLSFilter
ADD FILTER PREDICATE Security.tvf_riSPredicate(AuthorEmail)
ON
WITH (STATE = ON)
```

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

- (Topic 3)

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

columns.

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

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

You need to prepare the data.

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

NOTE: Each correct selection is worth one point.

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

Answer: BCF

Explanation:

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

? ProductName: The ProductName is an important descriptive attribute of the product, which is needed for analysis and reporting in a dimensional model.

? ProductColor: ProductColor is another descriptive attribute of the product. In a star schema, it makes sense to include attributes like color in the dimension table to help categorize products in the analysis.

? ProductID: ProductID is the primary key for the DimProduct table, which will be used to join the FactSales table to the product dimension. It's essential for uniquely identifying each product in the model.

NEW QUESTION 6

- (Topic 3)

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

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

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

Trigger the process when a new file is added.

Provide the highest throughput.

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

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

Answer: A

Explanation:

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

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

NEW QUESTION 7

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

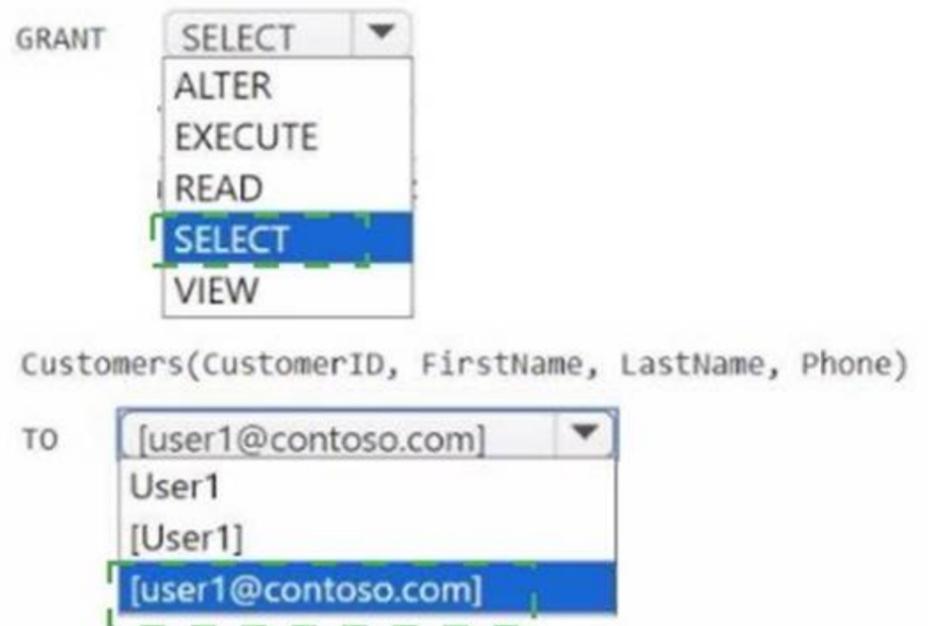
- User1
- [User1]
- [user1@contoso.com]**

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 8

- (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 9

- (Topic 3)

You have an Azure SQL database named DB1.

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

You discover that events are NOT being propagated to EventStreamDBI.

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

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

Answer: D

NEW QUESTION 10

- (Topic 3)

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

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

You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

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

```

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

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

- A. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 10

DRAG DROP - (Topic 3)

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

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

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

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

Actions

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

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions

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

Answer Area

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

NEW QUESTION 15

DRAG DROP - (Topic 3)

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

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

Badges

- ☰ Certified
- ☰ Master data
- ☰ Promoted
- ☰ Cannot be endorsed

Answer Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Badges

- ☰ Certified
- ☰ Master data
- ☰ Promoted
- ☰ Cannot be endorsed

Answer Area

- Entity1: ☰ Master data
- Entity2: ☰ Certified
- Entity3: ☰ Promoted
- Entity4: ☰ Certified

NEW QUESTION 17

- (Topic 3)

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

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

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

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp

You need to apply transformation and filter logic to prepare the data for consumption. The

solution must return data for a neighbourhood named Sands End when No_Bikes is at least 15. The results must be ordered by No_Bikes in ascending order.

Solution: You use the following code segment:

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

Does this meet the goal?

- A. Yes
- B. no

Answer: B

Explanation:

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

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

NEW QUESTION 21

- (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 22

- (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 26

HOTSPOT - (Topic 3)

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

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

The solution must minimize development effort.
 What should you include in the recommendation for each source? To answer, select the appropriate options in the answer area.
 NOTE: Each correct selection is worth one point.

Answer Area

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

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

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Source1:

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

Source2:

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

Source3:

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

NEW QUESTION 31

HOTSPOT - (Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

Dataset1:

- A Dataflow Gen2 dataflow
- A notebook
- A T-SQL statement**

Dataset2:

- A Dataflow Gen2 dataflow
- A notebook**
- A T-SQL statement

Dataset3:

- A Dataflow Gen2 dataflow
- A KQL queryset**
- A T-SQL statement

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Dataset1:

Dataset2:

Dataset3:

NEW QUESTION 34

DRAG DROP - (Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

Values

-
-
-
-

Answer Area

Statement1:

```
TaxiData
| sort by VendorID asc
| extend RunningTotalAmount =  (total_amount, VendorID != prev(VendorID))
```

Statement2:

```
TaxiData
| sort by tpep_pickup_datetime asc, payment_type asc
| extend FirstPickupDateTime =  (tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Partition the RunningTotalAmount by VendorID. - Row_cumsum

The Row_cumsum function computes the cumulative sum of a column while optionally restarting the accumulation based on a condition. In this case, it calculates the cumulative sum of total_amount for each VendorID, restarting when the VendorID changes (VendorID != prev(VendorID)).

```
TaxiData
```

```
| sort by VendorID asc
```

```
| extend RunningTotalAmount = Row_cumsum(total_amount, VendorID != prev(VendorID))
```

Create a column FirstPickupDateTime that shows the first value of each hour from tpep_pickup_datetime, partitioned by payment_type - Row_window_session

```
TaxiData
```

```
| sort by tpep_pickup_datetime asc, payment_type asc
```

```
| extend FirstPickupDateTime = Row_window_session(tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))
```

NEW QUESTION 39

- (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 42

- (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 46

- (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 47

DRAG DROP - (Topic 3)

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

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

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

You need to set the Customerkey column as a primary key of the DimCustomer table. Which three code segments should you run in sequence? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.

Code Segments

- 0 DROP CONSTRAINT PK_DimCustomer
- 0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)
- 0 NOT ENFORCED
- 0 ALTER TABLE dbo.DimCustomer
- 0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)
- 0 ENFORCED

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Code Segments

- 0 DROP CONSTRAINT PK_DimCustomer
- 0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)
- 0 NOT ENFORCED
- 0 ALTER TABLE dbo.DimCustomer
- 0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)
- 0 ENFORCED

Answer Area

0 ALTER TABLE dbo.DimCustomer

0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)

0 ENFORCED

NEW QUESTION 50

- (Topic 3)

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

You have a Fabric workspace named Workspace1. Workspace1 contains a notebook named Notebook1 that performs the following tasks:

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

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token. Solution: You use the following code segment:

Use notebookutils.credentials.getSecret and specify key vault URL and the name of a linked service.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 51

- (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 53

- (Topic 3)

You have two Fabric workspaces named Workspace1 and Workspace2. You have a Fabric deployment pipeline named deployPipeline1 that deploys items from Workspace1 to Workspace2. DeployPipeline1 contains all the items in Workspace1. You recently modified the items in Workspaces1. The workspaces currently contain the items shown in the following table.

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

Items in Workspace1 that have the same name as items in Workspace2 are currently paired. You need to ensure that the items in Workspace1 overwrite the corresponding items in Workspace2. The solution must minimize effort. What should you do?

- A. Delete all the items in Workspace2, and then run deployPipeline1.
- B. Rename each item in Workspace2 to have the same name as the items in Workspace1.

- C. Back up the items in Workspace2, and then run deployPipeline1.
- D. Run deployPipeline1 without modifying the items in Workspace2.

Answer: D

Explanation:

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

NEW QUESTION 55

- (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 60

- (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 65

- (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 66

.....

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](#)