

100% Money Back
Guarantee

Vendor:Microsoft

Exam Code:DP-600

Exam Name:Implementing Analytics Solutions Using
Microsoft Fabric

Version:Demo

QUESTION 1

What should you recommend using to ingest the customer data into the data store in the AnalyticsPOC workspace?

- A. a stored procedure
- B. a pipeline that contains a KQL activity
- C. a Spark notebook
- D. a dataflow

Correct Answer: D

Explanation: For ingesting customer data into the data store in the AnalyticsPOC workspace, a dataflow (D) should be recommended. Dataflows are designed within the Power BI service to ingest, cleanse, transform, and load data into the Power BI environment. They allow for the low-code ingestion and transformation of data as needed by Litware's technical requirements. References = You can learn more about dataflows and their use in Power BI environments in Microsoft's Power BI documentation.

QUESTION 2

You need to create a DAX measure to calculate the average overall satisfaction score.

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

NOTE: Each correct selection is worth one point.

Hot Area:

```

Rolling 12 Overall Satisfaction =
VAR NumberOfMonths = 12
VAR LastCurrentDate = MAX ( 'Date'[Date] )
VAR Period = DATESINPERIOD ( 'Date'[Date], LastCurrentDate, - NumberOfMonths, MONTH )
VAR Result =
    CALCULATE (
        AVERAGE('Survey'[Response Value]),
        AVERAGE('Survey'[Response Value]),
        AVERAGEA('Question'[Question Text]),
        AVERAGEX(VALUES('Survey'[Customer Key]),
            NumberOfMonths,
            LastCurrentDate,
            NumberOfMonths,
            Period,
        'Survey Question'[Question Title] = "Overall Satisfaction"
    )
RETURN
    Result

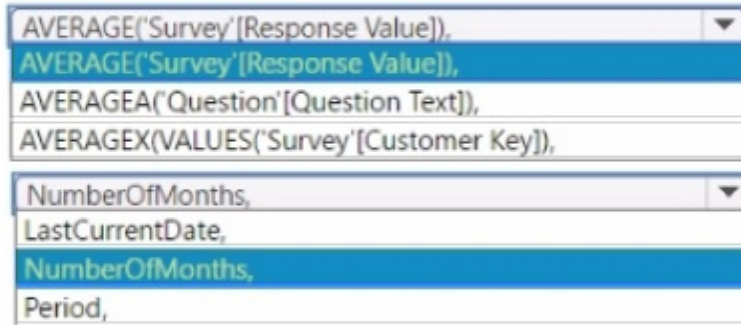
```

Correct Answer:

```

Rolling 12 Overall Satisfaction =
VAR NumberOfMonths = 12
VAR LastCurrentDate = MAX ( 'Date'[Date] )
VAR Period = DATESINPERIOD ( 'Date'[Date], LastCurrentDate, - NumberOfMonths, MONTH )
VAR Result =
    CALCULATE (
        AVERAGE('Survey'[Response Value]),
        AVERAGE('Survey'[Response Value]),
        AVERAGEA('Question'[Question Text]),
        AVERAGEX(VALUES('Survey'[Customer Key]),
            NumberOfMonths,
            LastCurrentDate,
            NumberOfMonths,
            Period,
        'Survey Question'[Question Title] = "Overall Satisfaction"
    )
RETURN
    Result

```



The measure should use the AVERAGE function to calculate the average value.

It should reference the Response Value column from the 'Survey' table.

The 'Number of months' should be used to define the period for the average calculation.

To calculate the average overall satisfaction score using DAX, you would need to use the AVERAGE function on the response values related to satisfaction questions. The DATESINPERIOD function will help in calculating the rolling average

over the last 12 months.

QUESTION 3

You are analyzing customer purchases in a Fabric notebook by using PySpark. You have the following DataFrames:

- transactions: Contains five columns named transaction_id, customer_id, product_id, amount, and date and has 10 million rows, with each row representing a transaction
- customers: Contains customer details in 1,000 rows and three columns named customer_id, name, and country

You need to join the DataFrames on the customer_id column. The solution must minimize data shuffling. You write the following code.

```
from pyspark.sql import functions as F
```

```
results =
```

Which code should you run to populate the results DataFrame? A)

```
transactions.join(F.broadcast(customers), transactions.customer_id == customers.customer_id)
```

B)

```
transactions.join(customers, transactions.customer_id == customers.customer_id).distinct()
```

C)

```
transactions.join(customers, transactions.customer_id == customers.customer_id)
```

D)

```
transactions.crossJoin(customers).where(transactions.customer_id == customers.customer_id)
```

A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: A

Explanation: The correct code to populate the results DataFrame with minimal data shuffling is Option A. Using the broadcast function in PySpark is a way to minimize data movement by broadcasting the smaller DataFrame (customers) to each node in the cluster. This is ideal when one DataFrame is much smaller than the other, as in this case with customers. References = You can refer to the official Apache Spark documentation for more details on joins and the broadcast hint.

QUESTION 4

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a subfolder named Subfolder1 that contains CSV files. You need to convert the CSV files into the delta format that has V-Order optimization enabled. What should you do from Lakehouse explorer?

A. Use the Load to Tables feature.

B. Create a new shortcut in the Files section.

C. Create a new shortcut in the Tables section.

D. Use the Optimize feature.

Correct Answer: D

Explanation: To convert CSV files into the delta format with Z-Order optimization enabled, you should use the Optimize

feature (D) from Lakehouse Explorer. This will allow you to optimize the file organization for the most efficient querying. References = The process for converting and optimizing file formats within a lakehouse is discussed in the lakehouse management documentation.

QUESTION 5

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Nyctaxi_raw. Nyctaxi_raw contains the following columns.

Name	Data type
pickupDateTime	Timestamp
passengerCount	Integer
fareAmount	Double
paymentType	String
tipAmount	Double

You create a Fabric notebook and attach it to lakehouse1.

You need to use PySpark code to transform the data. The solution must meet the following requirements:

Correct Answer:

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")
df2 = df.withColumn("pickupDate",df['tpepPickupDateTime'].cast('date')).filter("fareAmount > 0 AND fareAmount < 100")
```

Add the pickupDate column: `.withColumn("pickupDate", df["pickupDateTime"].cast("date"))`
 Filter the DataFrame: `.filter("fareAmount > 0 AND fareAmount`

In PySpark, you can add a new column to a DataFrame using the `.withColumn` method, where the first argument is the new column name and the second argument is the expression to generate the content of the new column. Here, we use

the `.cast("date")` function to extract only the date part from a timestamp. To filter the DataFrame, you use the `.filter` method with a condition that selects rows where `fareAmount` is greater than 0 and less than 100, thus ensuring only positive

values less than 100 are included.

QUESTION 6

You are analyzing the data in a Fabric notebook.

You have a Spark DataFrame assigned to a variable named `df`.

You need to use the Chart view in the notebook to explore the data manually.

Which function should you run to make the data available in the Chart view?

A. `displayHTML`

B. `show`

C. `write`

D. `display`

Correct Answer: D

Explanation: The `display` function is the correct choice to make the data available in the Chart view within a Fabric notebook. This function is used to visualize Spark DataFrames in various formats including charts and graphs directly within the notebook environment. References = Further explanation of the `display` function can be found in the official documentation on Azure Synapse Analytics notebooks.

QUESTION 7

You have a Fabric workspace named `Workspace1` and an Azure Data Lake Storage Gen2 account named `storage!`. `Workspace1` contains a lakehouse named `Lakehouse1`.

You need to create a shortcut to `storage!` in `Lakehouse1`.

Which connection and endpoint should you specify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Connection:

abfss	▼
abfs	
abfss	
https	

Endpoint:

dfs	▼
blob	
dfs	
file	

Correct Answer:

Connection:

abfss	▼
abfs	
abfss	
https	

Endpoint:

dfs	▼
blob	
dfs	
file	

When creating a shortcut to an Azure Data Lake Storage Gen2 account in a lakehouse, you should use the abfss (Azure Blob File System Secure) connection string and the dfs (Data Lake File System) endpoint. The abfss is used for secure

access to Azure Data Lake Storage, and the dfs endpoint indicates that the Data Lake Storage Gen2 capabilities are to be used.

QUESTION 8

You need to assign permissions for the data store in the AnalyticsPOC workspace. The solution must meet the security requirements.

Which additional permissions should you assign when you share the data store? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

DataEngineers:

Build Reports on the default dataset	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataAnalysts:

Read All Apache Spark	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataScientists:

Read All SQL analytics endpoint data	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

Correct Answer:

DataEngineers:

Build Reports on the default dataset	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataAnalysts:

Read All Apache Spark	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataScientists:

Read All SQL analytics endpoint data	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

Data Engineers: Read All SQL analytics endpoint data

Data Analysts: Read All Apache Spark

Data Scientists: Read All SQL analytics endpoint data

The permissions for the data store in the AnalyticsPOC workspace should align with the principle of least privilege:

Data Engineers need read and write access but not to datasets or reports.

Data Analysts require read access specifically to the dimensional model objects and the ability to create Power BI reports.

Data Scientists need read access via Spark notebooks. These settings ensure each role has the necessary permissions to fulfill their responsibilities without exceeding their required access level.

QUESTION 9

You are creating a dataflow in Fabric to ingest data from an Azure SQL database by using a T-SQL statement.

You need to ensure that any foldable Power Query transformation steps are processed by the Microsoft SQL Server engine.

How should you complete the code? 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.

Select and Place:

Values

- EnableFolding
- NativeQuery
- Optimize
- Record
- StopFolding
- Table
- Value

Answer Area

```
let
    Source = Sql.Databases(
        "server.database.windows.net"
    ),
    Database = Source[{Name = "db"}][Data],
    Query = [ ] - [ ] (
        Database,
        " SELECT * FROM customer WHERE country IN ('USA', 'UK')",
        null,
        [ [ ] = true]
    )
in
    Query
```

Correct Answer:

Values

Answer Area

```
let
    Source = Sql.Databases(
        "server.database.windows.net"
    ),
    Database = Source{[Name = "db"]}[Data],
    Query = [Table] - [NativeQuery] (
        Database,
        " SELECT * FROM customer WHERE country IN ('USA', 'UK')",
        null,
        [ EnableFolding ] = true
    )
in
    Query
```

You should complete the code as follows: Table NativeQuery EnableFolding

In Power Query, using Table before the SQL statement ensures that the result of the SQL query is treated as a table. NativeQuery allows a native database query to be passed through from Power Query to the source database. The EnableFolding option ensures that any subsequent transformations that can be folded will be sent back and executed at the source database (Microsoft SQL Server engine in this case).

QUESTION 10

You have a Fabric tenant tha1 contains a takehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer.

When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table.

You need to identify whether maintenance tasks were performed on Customer.

Solution: You run the following Spark SQL statement:

```
EXPLAIN TABLE customer
```

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Explanation: No, the EXPLAIN TABLE statement does not identify whether maintenance tasks were performed on a table. It shows the execution plan for a query. References = The usage and output of the EXPLAIN command can be

found in the Spark SQL documentation.

QUESTION 11

You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains a fact table named FactSales that has one billion rows. You run the following TSQL statement.

```
CREATE TABLE test.FactSales AS CLONE OF Dbo.FactSales;
```

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

Hot Area:

Statements	Yes	No
A replica of <code>dbo.Sales</code> is created in the test schema by copying the metadata only.	<input type="radio"/>	<input type="radio"/>
Additional schema changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input type="radio"/>
Additional data changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input type="radio"/>

Correct Answer:

Statements	Yes	No
A replica of <code>dbo.Sales</code> is created in the test schema by copying the metadata only.	<input type="radio"/>	<input checked="" type="radio"/>
Additional schema changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input checked="" type="radio"/>
Additional data changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input checked="" type="radio"/>	<input type="radio"/>

A replica of `dbo.Sales` is created in the test schema by copying the metadata only. - No Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`. - No Additional data changes to `dbo.FactSales` will also apply to `test.FactSales`. - Yes The CREATE TABLE AS CLONE statement creates a copy of an existing table, including its data and any associated data structures, like indexes. Therefore, the statement does not merely copy metadata; it also copies the data. However, subsequent schema changes to the original table do not automatically propagate to the cloned table. Any data changes in the original table after the clone operation will not be reflected in the clone unless explicitly updated. References = CREATE TABLE AS SELECT (CTAS) in SQL Data Warehouse

QUESTION 12

You have a Fabric tenant that contains a takehouse named lakehouse1. Lakehouse1 contains a Delta table named Customer.

When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table.

You need to identify whether maintenance tasks were performed on Customer.

Solution: You run the following Spark SQL statement:

```
DESCRIBE HISTORY customer
```

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

Explanation: Yes, the DESCRIBE HISTORY statement does meet the goal. It provides information on the history of operations, including maintenance tasks, performed on a Delta table. References = The functionality of the DESCRIBE HISTORY statement can be verified in the Delta Lake documentation.