DP-600$^{Q\&As}$

Implementing Analytics Solutions Using Microsoft Fabric

Pass Microsoft DP-600 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/dp-600.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

You are analyzing customer purchases in a Fabric notebook by using PySpanc 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 2

You are the administrator of a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:

Table1: A Delta table created by using a shortcut Table2: An external table created by using Spark Table3: A managed table

You plan to connect to Lakehouse1 by using its SQL endpoint. What will you be able to do after connecting to Lakehouse1?

A. ReadTable3.

B. Update the data Table3.

C. ReadTable2.

D. Update the data in Table1.

Correct Answer: D

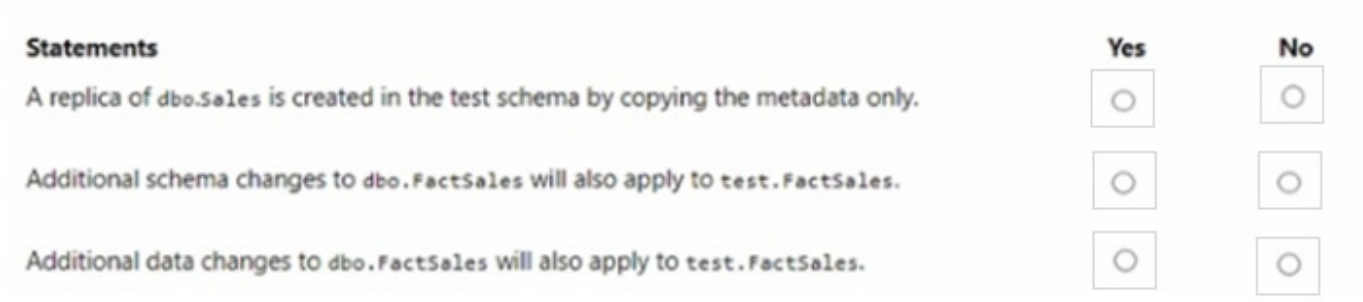
QUESTION 3

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.

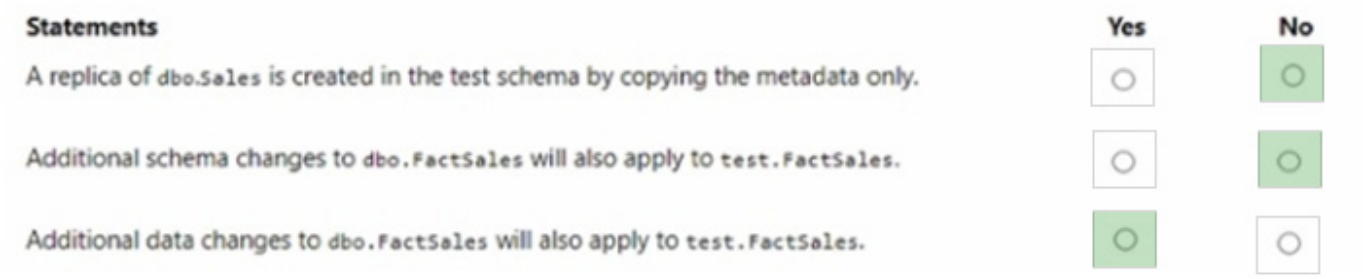
CREATE TABLE test.FactSales AS CLONE OF Dbo.FactSales;

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:



Correct Answer:



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 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 4

You have a Fabric workspace named Workspace1 that contains a data flow named Dataflow1. Dataflow1 contains a query that returns the data shown in the following exhibit.

You need to transform the date columns into attribute-value pairs, where columns become rows.

You select the VendorID column.

Which transformation should you select from the context menu of the VendorID column?

A. Group by

B. Unpivot columns

C. Unpivot other columns

D. Split column

E. Remove other columns

Correct Answer: B

Explanation: The transformation you should select from the context menu of the VendorID column to transform the date columns into attribute-value pairs, where columns become rows, is Unpivot columns (B). This transformation will turn the selected columns into rows with two new columns, one for the attribute (the original column names) and one for the value (the data from the cells). References = Techniques for unpivoting columns are covered in the Power Query documentation, which explains how to use the transformation in data modeling.

QUESTION 5

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.