Microsoft.DP-600.by.Rian.30q

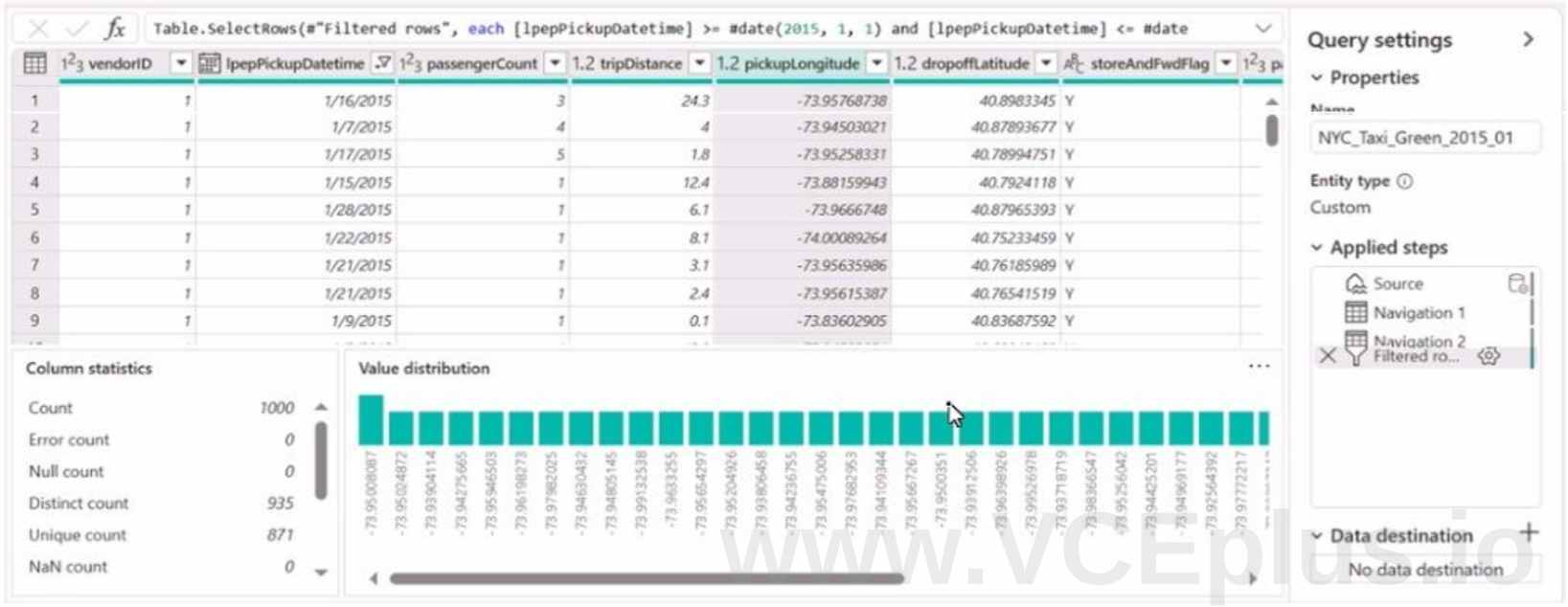
Number: DP-600
Passing Score: 800
Time Limit: 120
File Version: 4.0

DP-600

Exam Name: Implementing Analytics Solutions Using Microsoft Fabric

Exam A

QUESTION 1
You have a Fabric workspace named Workspace 1 that contains a dataflow named Dataflow1. Dataflow! has a query that returns 2.000 rows. You view the query in Power Query as shown in the following exhibit.



What can you identify about the pickupLongitude column?

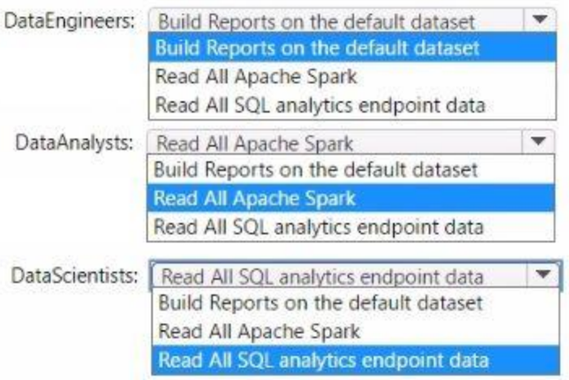
A. The column has duplicate values.
B. All the table rows are profiled.
C. The column has missing values.
D. There are 935 values that occur only once.

Correct Answer: B
Section:

QUESTION 2
HOTSPOT
You to need 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.

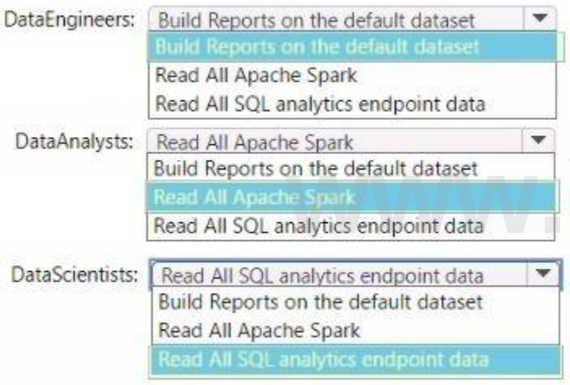
Hot Area:

Answer Area



Answer Area:

Answer Area



Section:
Explanation:

QUESTION 3
HOTSPOT
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.

Answer Area

```
Rolling 12 Overall Satisfaction =
VAR NumberOfMonths = 12
VAR LastCurrentDate = MAX ( 'Date'[Date] )
VAR Period = DATESINPERIOD ( 'Date'[Date], LastCurrentDate, - NumberOfMonths, MONTH )
VAR Result =
  CALCULATE (
    [                                                    ▼]
    [                                                    ▼]
    'Survey Question'[Question Title] = "Overall Satisfaction"
  )
RETURN
  Result
```

Hot Area:

Answer 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]),                   ▼
    NumberOfMonths,                                      ▼
    'Survey Question'[Question Title] = "Overall Satisfaction"
  )
RETURN
  Result
```

- AVERAGE('Survey'[Response Value]),
- AVERAGEA('Question'[Question Text]),
- AVERAGEX(VALUES('Survey'[Customer Key]),

- LastCurrentDate,
- NumberOfMonths,
- Period,

Answer Area:

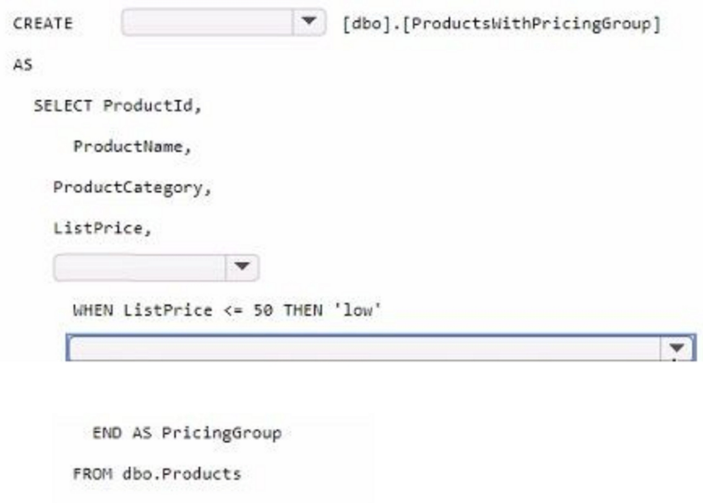
Answer Area

```
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
```

Section:
Explanation:

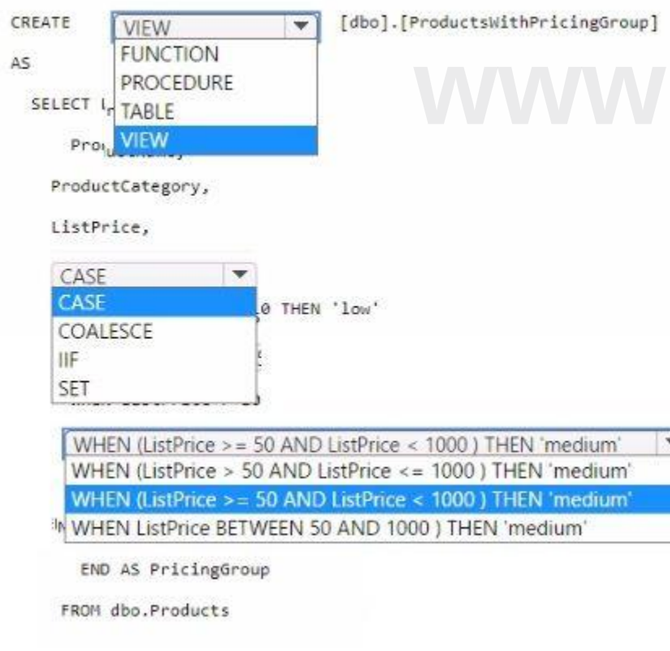
QUESTION 4
HOTSPOT
You need to resolve the issue with the pricing group classification.
How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area



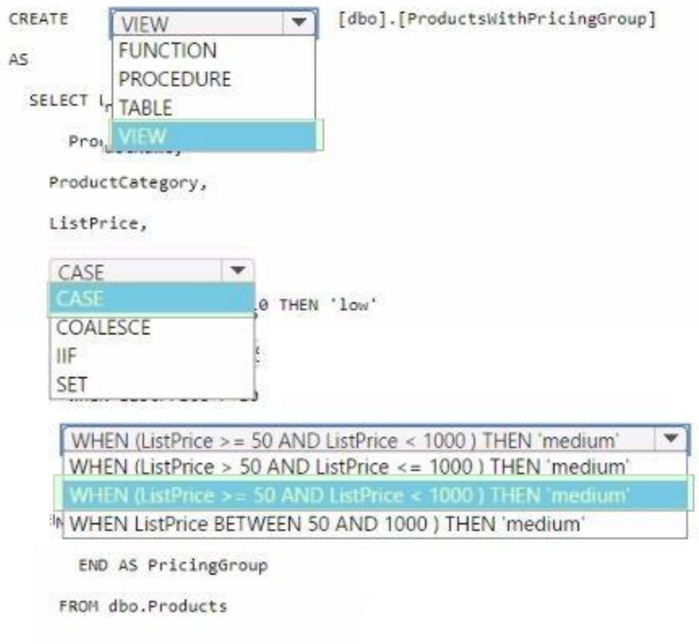
Hot Area:

Answer Area



Answer Area:

Answer Area



Section:
Explanation:

QUESTION 5
What should you recommend using to ingest the customer data into the data store in the AnatyticsPOC workspace?

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

Correct Answer: D
Section:

QUESTION 6
Which type of data store should you recommend in the AnalyticsPOC workspace?

A. a data lake
B. a warehouse
C. a lakehouse
D. an external Hive metaStore

Correct Answer: C
Section:

QUESTION 7

You have a Fabric warehouse that contains a table named Staging.Sales. Staging.Sales contains the following columns.

Name	Data type	Nullable
ProductID	Integer	No
ProductName	Varchar(30)	No
SalesDate	Datetime2(6)	No
WholesalePrice	Decimal(18, 2)	Yes
Amount	Decimal(18, 2)	Yes

You need to write a T-SQL query that will return data for the year 2023 that displays ProductID and ProductName arxl has a summarized Amount that is higher than 10,000. Which query should you use?

A)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000
```

B)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
GROUP BY ProductID, ProductName
HAVING DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000
```

C)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000
```

D)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING TotalAmount > 10000
```

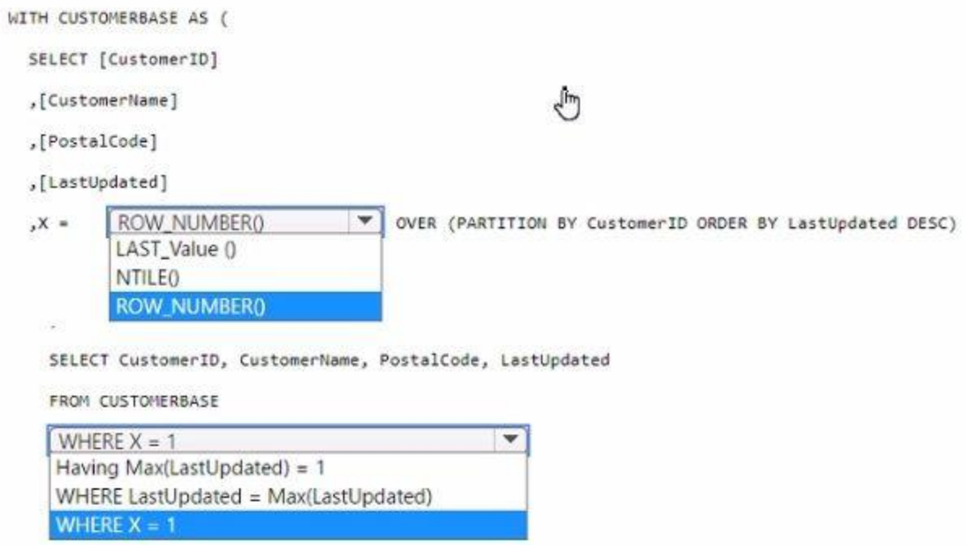

A. Option A
B. Option B
C. Option C
D. Option D

Correct Answer: A
Section:

QUESTION 8
HOTSPOT
You have a data warehouse that contains a table named Stage. Customers. Stage-Customers contains all the customer record updates from a customer relationship management (CRM) system. There can be multiple updates per customer
You need to write a T-SQL query that will return the customer ID, name, postal code, and the last updated time of the most recent row for each customer ID.
How should you complete the code? To answer, select the appropriate options in the answer area,
NOTE Each correct selection is worth one point.

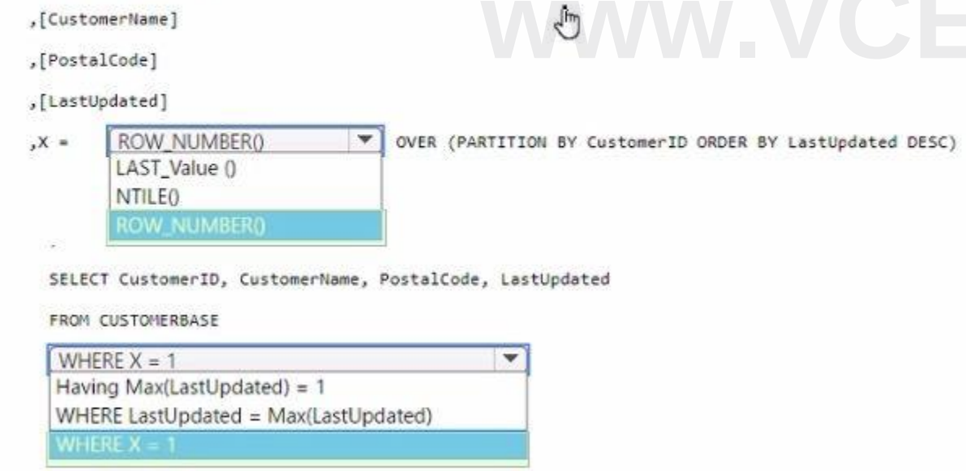
Hot Area:

Answer Area



Answer Area:

Answer Area



Section:
Explanation:

QUESTION 9
You have a Fabric workspace that contains a DirectQuery semantic model. The model queries a data source that has 500 million rows.
You have a Microsoft Power Bl report named Report1 that uses the model. Report! contains visuals on multiple pages.
You need to reduce the query execution time for the visuals on all the pages.
What are two features that you can use? Each correct answer presents a complete solution.
NOTE: Each correct answer is worth one point.

A. user-defined aggregations
B. automatic aggregation
C. query caching
D. OneLake integration

Correct Answer: A, B
Section:

QUESTION 10
You have a Fabric tenant that contains 30 CSV files in OneLake. The files are updated daily.
You create a Microsoft Power Bl semantic model named Modell that uses the CSV files as a data source. You configure incremental refresh for Model 1 and publish the model to a Premium capacity in the Fabric tenant.
When you initiate a refresh of Model1, the refresh fails after running out of resources.
What is a possible cause of the failure?

A. Query folding is occurring.
B. Only refresh complete days is selected.
C. XMLA Endpoint is set to Read Only.
D. Query folding is NOT occurring.
E. The data type of the column used to partition the data has changed.

Correct Answer: E
Section:

QUESTION 11
You have a Fabric tenant that uses a Microsoft tower Bl Premium capacity. You need to enable scale-out for a semantic model. What should you do first?

A. At the semantic model level, set Large dataset storage format to Off.
B. At the tenant level, set Create and use Metrics to Enabled.
C. At the semantic model level, set Large dataset storage format to On.
D. At the tenant level, set Data Activator to Enabled.

Correct Answer: C
Section:

QUESTION 12
You have a Fabric tenant that contains a warehouse. The warehouse uses row-level security (RLS). You create a Direct Lake semantic model that uses the Delta tables and RLS of the warehouse. When users interact with a report built from the model, which mode will be used by the DAX queries?

A. DirectQuery
B. Dual
C. Direct Lake
D. Import

Correct Answer: C
Section:

QUESTION 13

You have a Fabric tenant that contains a complex semantic model. The model is based on a star schema and contains many tables, including a fact table named Sales. You need to create a diagram of the model. The diagram must contain only the Sales table and related tables. What should you use from Microsoft Power Bl Desktop?

A. data categories
B. Data view
C. Model view
D. DAX query view

Correct Answer: C
Section:

QUESTION 14
You have a Fabric tenant that contains a semantic model. The model uses Direct Lake mode.
You suspect that some DAX queries load unnecessary columns into memory.
You need to identify the frequently used columns that are loaded into memory.
What are two ways to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct answer is worth one point.

A. Use the Analyze in Excel feature.
B. Use the Vertipaq Analyzer tool.
C. Query the $system.discovered_STORAGE_TABLE_COLUMN-iN_SEGMeNTS dynamic management view (DMV).
D. Query the discover_hehory6Rant dynamic management view (DMV).

Correct Answer: B, C
Section:

QUESTION 15
You have a Fabric tenant that contains a warehouse.
Several times a day. the performance of all warehouse queries degrades. You suspect that Fabric is throttling the compute used by the warehouse.
What should you use to identify whether throttling is occurring?

A. the Capacity settings
B. the Monitoring hub
C. dynamic management views (DMVs)
D. the Microsoft Fabric Capacity Metrics app

Correct Answer: B
Section:

QUESTION 16
HOTSPOT
You have a Fabric workspace that uses the default Spark starter pool and runtime version 1,2.
You plan to read a CSV file named Sales.raw.csv in a lakehouse, select columns, and save the data as a Delta table to the managed area of the lakehouse. Sales_raw.csv contains 12 columns.
You have the following code.

```
from pyspark.sql.functions import year

(spark
  .read
  .format("csv")
  .option("heade" ,'true')
  .load("Files/sales_raw.csv")
  .select('SalesOrderNumber', 'OrderDate','CustomerName', 'UnitPrice')
  .withColumn("Year",year("OrderDate"))
  .write
  .partitionBy('Year')
  .saveAsTable("sales")
)
```

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:

Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate','CustomerName', 'UnitPrice' columns from Sales_raw.csv.	○	○
Removing the partition will reduce the execution time of the query.	○	○
Adding `inferSchema='true'` to the options will increase the execution time of the query.	○	○

Answer Area:

Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate','CustomerName', 'UnitPrice' columns from Sales_raw.csv.	○	○
Removing the partition will reduce the execution time of the query.	○	○
Adding `inferSchema='true'` to the options will increase the execution time of the query.	○	○

Section:
Explanation:

QUESTION 17
You have a Fabric tenant that contains a warehouse.
A user discovers that a report that usually takes two minutes to render has been running for 45 minutes and has still not rendered.
You need to identify what is preventing the report query from completing.
Which dynamic management view (DMV) should you use?

A. sys.dm-exec_requests
B. sys.dn_.exec._sessions
C. sys.dm._exec._connections
D. sys.dm_pdw_exec_requests

Correct Answer: A
Section:

QUESTION 18
DRAG DROP
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

- Optimize
- Record
- StopFolding
- Table

Answer Area

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

Section:
Explanation:

QUESTION 19
DRAG DROP
You have a Fabric tenant that contains a lakehouse named Lakehouse1
Readings from 100 IoT devices are appended to a Delta table in Lakehouse1. Each set of readings is approximately 25 KB. Approximately 10 GB of data is received daily.
All the table and SparkSession settings are set to the default.
You discover that queries are slow to execute. In addition, the lakehouse storage contains data and log files that are no longer used.
You need to remove the files that are no longer used and combine small files into larger files with a target size of 1 GB per file.
What should you do? To answer, drag the appropriate actions to the correct requirements. Each action 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:

Actions

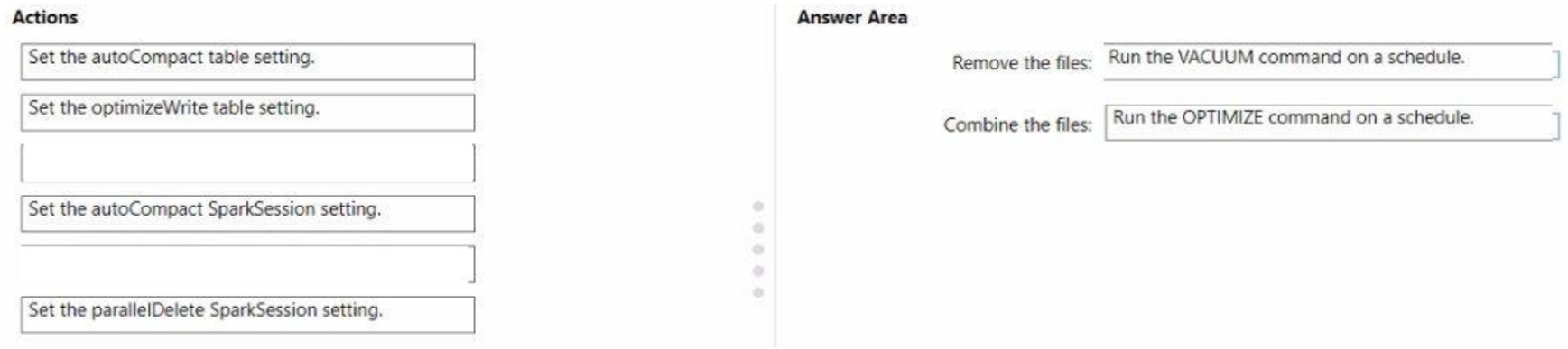
- Set the autoCompact table setting.
- Set the optimizeWrite table setting.
- Run the VACUUM command on a schedule.
- Set the autoCompact SparkSession setting.
- Run the OPTIMIZE command on a schedule.
- Set the parallelDelete SparkSession setting.

Answer Area

Remove the files: [ ]

Combine the files: [ ]

Correct Answer:



Section:
Explanation:

QUESTION 20
You need to create a data loading pattern for a Type 1 slowly changing dimension (SCD).
Which two actions should you include in the process? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. Update rows when the non-key attributes have changed.
B. Insert new rows when the natural key exists in the dimension table, and the non-key attribute values have changed.
C. Update the effective end date of rows when the non-key attribute values have changed.
D. Insert new records when the natural key is a new value in the table.

Correct Answer: A, D
Section:

QUESTION 21
HOTSPOT
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:

Answer Area:

Section:
Explanation:

QUESTION 22
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)
```

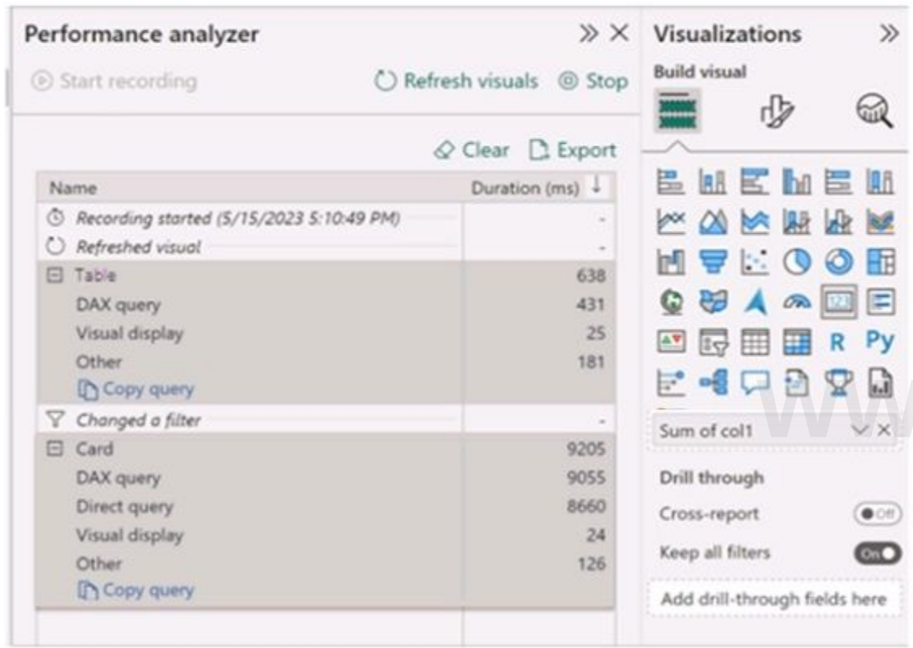
D)

```
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
Section:

QUESTION 23
HOTSPOT
You have a Microsoft Power B1 report and a semantic model that uses Direct Lake mode. From Power Si Desktop, you open Performance analyzer as shown in the following exhibit.



Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic. NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area:

Section:
Explanation:

QUESTION 24
You have a Fabric tenant that contains a lakehouse named lakehouse1. Lakehouse1 contains an unpartitioned table named Table1.
You plan to copy data to Table1 and partition the table based on a date column in the source data.
You create a Copy activity to copy the data to Table1.
You need to specify the partition column in the Destination settings of the Copy activity.
What should you do first?

A. From the Destination tab, set Mode to Append.
B. From the Destination tab, select the partition column,
C. From the Source tab, select Enable partition discovery
D. From the Destination tab, set Mode to Overwrite.

Correct Answer: B
Section:

QUESTION 25
HOTSPOT

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 T-SQL 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:

Answer Area

Statements	Yes	No
A replica of `dbo.Sales` is created in the test schema by copying the metadata only.	○	○
Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	○
Additional data changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	○

Answer Area:

Answer Area

Statements	Yes	No
A replica of `dbo.Sales` is created in the test schema by copying the metadata only.	○	○
Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	○
Additional data changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	○

Section:
Explanation:

QUESTION 26
You have source data in a folder on a local computer.
You need to create a solution that will use Fabric to populate a data store. The solution must meet the following requirements:
* Support the use of dataflows to load and append data to the data store.
* Ensure that Delta tables are V-Order optimized and compacted automatically.
Which type of data store should you use?

A. a lakehouse
B. an Azure SQL database
C. a warehouse
D. a KQL database

Correct Answer: A
Section:

QUESTION 27
HOTSPOT
You have a Fabric tenant that contains a lakehouse.
You are using a Fabric notebook to save a large DataFrame by using the following code.

```
df.write.partitionBy("year","month","day").mode("overwrite").parquet("Files/SalesOrder")
```

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:

Answer Area

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	○	○
The resulting file partitions can be read in parallel across multiple nodes.	○	○
The resulting file partitions will use file compression.	○	○

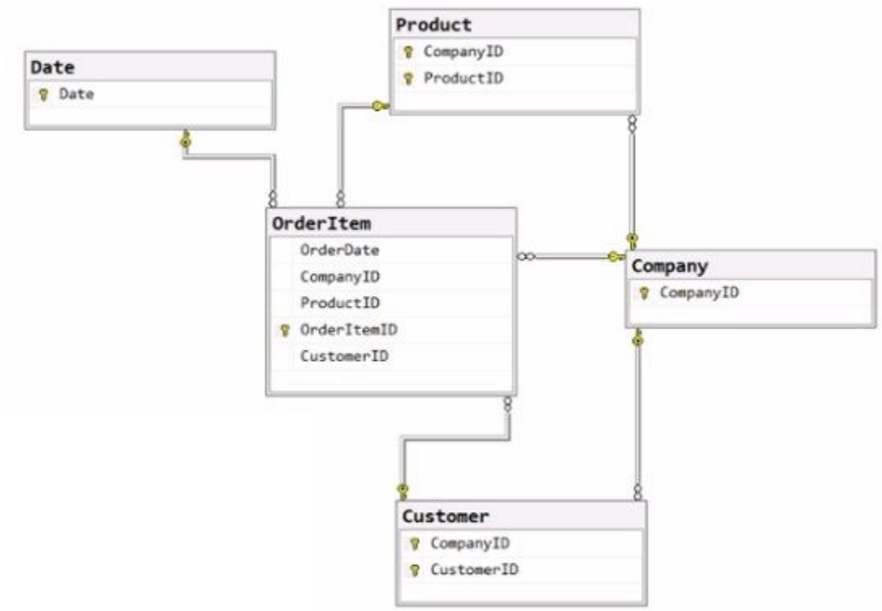
Answer Area:

Answer Area

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	○	○
The resulting file partitions can be read in parallel across multiple nodes.	○	○
The resulting file partitions will use file compression.	○	○

Section:
Explanation:

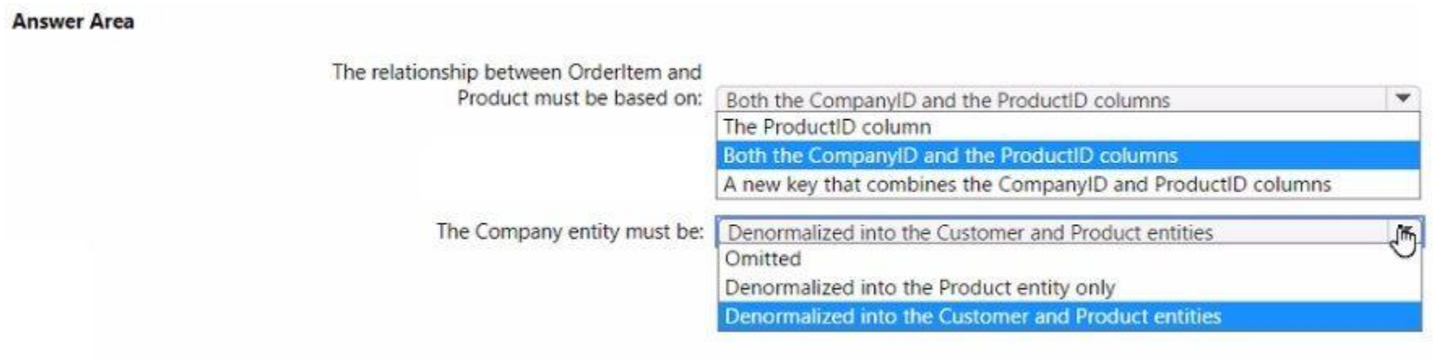
QUESTION 28
HOTSPOT
You have the source data model shown in the following exhibit.



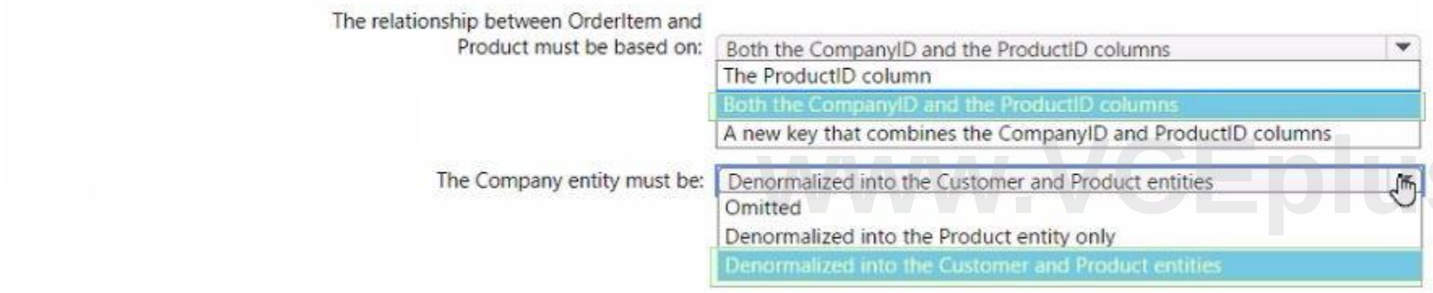
The primary keys of the tables are indicated by a key symbol beside the columns involved in each key.

You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer.
What should you include in the solution? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area:



Section:
Explanation:

QUESTION 29
You have a Fabric tenant that contains a semantic model named Model1. Model1 uses Import mode. Model1 contains a table named Orders. Orders has 100 million rows and the following fields.

Name	Data type	Description
OrderId	Integer	Column imported from the source
OrderDateTime	Date/time	Column imported from the source
Quantity	Integer	Column imported from the source
Price	Decimal	Column imported from the source
TotalSalesAmount	Decimal	Calculated column that multiplies Quantity and Price
TotalQuantity	Integer	Measure

You need to reduce the memory used by Model! and the time it takes to refresh the model. Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct answer is worth one point.

A. Split OrderDateTime into separate date and time columns.
B. Replace TotalQuantity with a calculated column.
C. Convert Quantity into the Text data type.
D. Replace TotalSalesAmount with a measure.

Correct Answer: A, D
Section:

QUESTION 30
You have a Fabric tenant that contains a semantic model.
You need to prevent report creators from populating visuals by using implicit measures.
What are two tools that you can use to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct answer is worth one point.

A. Microsoft Power BI Desktop
B. Tabular Editor
C. Microsoft SQL Server Management Studio (SSMS)
D. DAX Studio

Correct Answer: A, B
Section: