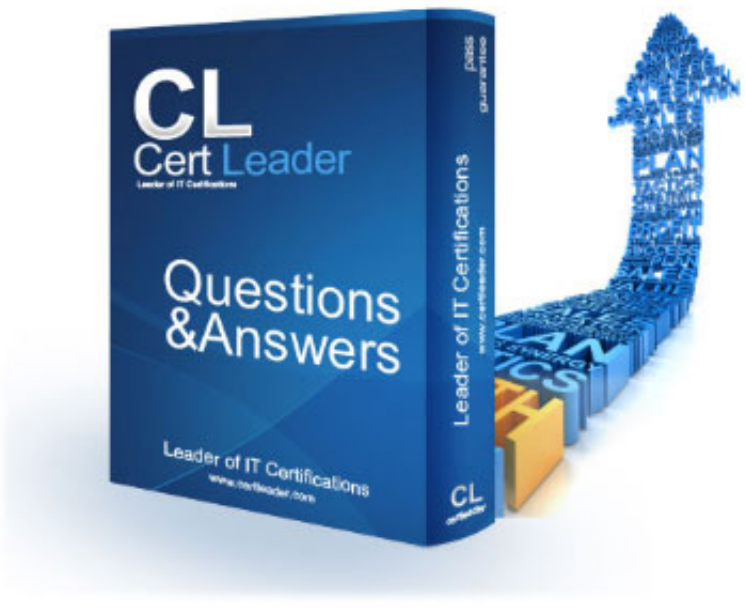
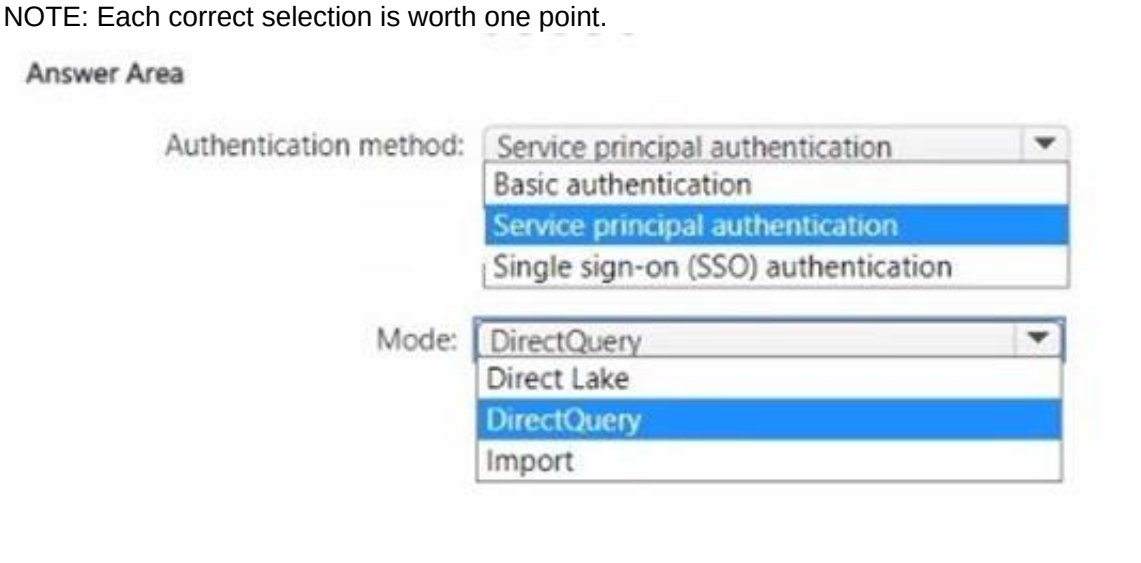
DP-600 Dumps

Implementing Analytics Solutions Using Microsoft Fabric

https://www.certleader.com/DP-600-dumps.html



NEW QUESTION 1
HOTSPOT - (Topic 1)
You need to design a semantic model for the customer satisfaction report.
Which data source authentication method and mode should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
For the semantic model design required for the customer satisfaction report, the choices for data source authentication method and mode should be made based on security and performance considerations as per the case study provided.
Authentication method: The data should be accessed securely, and given that row-level security (RLS) is required for users executing T-SQL queries, you should use an authentication method that supports RLS. Service principal authentication is suitable for automated and secure access to the data, especially when the access needs to be controlled programmatically and is not tied to a specific user's credentials.
Mode: The report needs to show data as soon as it is updated in the data store, and it should only contain data from the current and previous year. DirectQuery mode allows for real-time reporting without importing data into the model, thus meeting the need for up-to- date data. It also allows for RLS to be implemented and enforced at the data source level, providing the necessary security measures.
Based on these considerations, the selections should be:
? Authentication method: Service principal authentication
? Mode: DirectQuery

NEW QUESTION 2
- (Topic 1)
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

Answer: C

Explanation:
A lakehouse (C) should be recommended for the AnalyticsPOC workspace. It combines the capabilities of a data warehouse with the flexibility of a data lake. A lakehouse supports semi-structured and unstructured data and allows for T-SQL and Python read access, fulfilling the technical requirements outlined for Litware.
References = For further understanding, Microsoft's documentation on the lakehouse architecture provides insights into how it supports various data types and analytical operations.

NEW QUESTION 3
- (Topic 1)
You need to implement the date dimension in the data store. The solution must meet the technical requirements.
What are two ways to achieve the goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. Populate the date dimension table by using a dataflow.
B. Populate the date dimension table by using a Stored procedure activity in a pipeline.
C. Populate the date dimension view by using T-SQL.
D. Populate the date dimension table by using a Copy activity in a pipeline.

Answer: AB

Explanation:
Both a dataflow (A) and a Stored procedure activity in a pipeline (B) are capable of creating and populating a date dimension table. A dataflow can perform the transformation needed to create the date dimension, and it aligns with the preference for using low-code tools for data ingestion when possible. A Stored procedure could be written to generate the necessary date dimension data and executed within a pipeline, which also adheres to the technical requirements for the PoC.

NEW QUESTION 4
- (Topic 2)

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

Answer: A

Explanation:
When users interact with a report built from a Direct Lake semantic model that uses row-level security (RLS), the DAX queries will operate in DirectQuery mode (A). This is because the model directly queries the underlying data source without importing data into Power BI. References = The Power BI documentation on DirectQuery provides detailed explanations of how RLS and DAX queries function in this mode.

NEW QUESTION 5
- (Topic 2)
You have a semantic model named Model 1. Model 1 contains five tables that all use Import mode. Model1 contains a dynamic row-level security (RLS) role named HR. The HR role filters employee data so that HR managers only see the data of the department to which they are assigned.
You publish Model1 to a Fabric tenant and configure RLS role membership. You share the model and related reports to users.
An HR manager reports that the data they see in a report is incomplete. What should you do to validate the data seen by the HR Manager?

A. Ask the HR manager to open the report in Microsoft Power Bl Desktop.
B. Select Test as role to view the data as the HR role.
C. Select Test as role to view the report as the HR manager,
D. Filter the data in the report to match the intended logic of the filter for the HR department.

Answer: B

Explanation:
To validate the data seen by the HR manager, you should use the 'Test as role' feature in Power BI service. This allows you to see the data exactly as it would appear for the HR role, considering the dynamic RLS setup. Here is how you would proceed:
? Navigate to the Power BI service and locate Model1.
? Access the dataset settings for Model1.
? Find the security/RLS settings where you configured the roles.
? Use the 'Test as role' feature to simulate the report viewing experience as the HR role.
? Review the data and the filters applied to ensure that the RLS is functioning
correctly.
? If discrepancies are found, adjust the RLS expressions or the role membership as needed.
References: The 'Test as role' feature and its use for validating RLS in Power BI is covered in the Power BI documentation available on Microsoft's official documentation.

NEW QUESTION 6
- (Topic 2)
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.

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.

NEW QUESTION 7
- (Topic 2)
You have a Fabric tenant that contains a Microsoft Power Bl report. You are exploring a new semantic model.
You need to display the following column statistics:
• Count
• Average
• Null count
• Distinct count
• Standard deviation
Which Power Query function should you run?

A. Tabl
B. FuzzyGroup
C. Table.Profile
D. Table.View
E. Table.Schema

Answer: B

Explanation:
The Table.Profile function in Power Query is used to generate column statistics such as count, average, null count, distinct count, and standard deviation. You can

use this function as follows:
? Invoke the Power Query Editor.
? Apply the Table.Profile function to your table.
? The result will be a table where each row represents a column from the original table, and each column in the result represents a different statistic such as those listed in the requirement.
References: The use of Table.Profile is part of Power Query M function documentation where it explains how to gather column statistics for a given table.

NEW QUESTION 8
- (Topic 2)
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

Answer: C

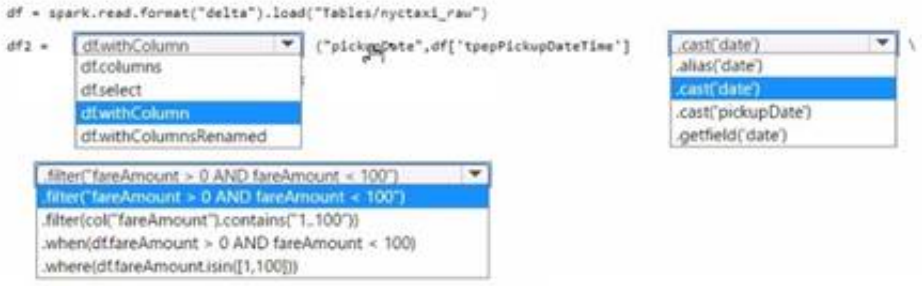
Explanation:
To create a diagram that contains only the Sales table and related tables, you should use the Model view (C) in Microsoft Power BI Desktop. This view allows you to visualize and manage the relationships between tables within your semantic model. References = Microsoft Power BI Desktop documentation outlines the functionalities available in Model view for managing semantic models.

NEW QUESTION 9
HOTSPOT - (Topic 2)
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:
• Add a column named pickupDate that will contain only the date portion of pickupDateTime.
• Filter the DataFrame to include only rows where fareAmount is a positive number that is less than 100.
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



A. Mastered
B. Not Mastered

Answer: A

Explanation:
? Add the pickupDate column: .withColumn("pickupDate", df["pickupDateTime"].cast("date"))
? Filter the DataFrame: .filter("fareAmount > 0 AND fareAmount < 100")
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.

NEW QUESTION 10
- (Topic 2)
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).

Answer: BC

Explanation:
The Vertipaq Analyzer tool (B) and querying the $system.discovered_STORAGE_TABLE_COLUMNS_IN_SEGMENTS dynamic management view (DMV) (C) can help identify which columns are frequently loaded into memory. Both methods provide insights into the storage and retrieval aspects of the semantic model. References = The Power BI documentation on Vertipaq Analyzer and DMV queries offers detailed guidance on how to use these tools for performance analysis.

NEW QUESTION 10
- (Topic 2)
You have a Microsoft Power Bl report named Report1 that uses a Fabric semantic model. Users discover that Report1 renders slowly.
You open Performance analyzer and identify that a visual named Orders By Date is the slowest to render. The duration breakdown for Orders By Date is shown in the following table.

Name	Duration (ms)
DAX query	27
Visual display	39
Other	1047

What will provide the greatest reduction in the rendering duration of Report1?

A. Change the visual type of Orders By Dale.
B. Enable automatic page refresh.
C. Optimize the DAX query of Orders By Date by using DAX Studio.
D. Reduce the number of visuals in Report1.

Answer: C

Explanation:
Based on the duration breakdown provided, the major contributor to the rendering duration is categorized as "Other," which is significantly higher than DAX Query and Visual display times. This suggests that the issue is less likely with the DAX calculation or visual rendering times and more likely related to model performance or the complexity of the visual. However, of the options provided, optimizing the DAX query can be a crucial step, even if "Other" factors are dominant. Using DAX Studio, you can analyze and optimize the DAX queries that power your visuals for performance improvements. Here's how you might proceed:
? Open DAX Studio and connect it to your Power BI report.
? Capture the DAX query generated by the Orders By Date visual.
? Use the Performance Analyzer feature within DAX Studio to analyze the query.
? Look for inefficiencies or long-running operations.
? Optimize the DAX query by simplifying measures, removing unnecessary calculations, or improving iterator functions.
? Test the optimized query to ensure it reduces the overall duration.
References: The use of DAX Studio for query optimization is a common best practice for improving Power BI report performance as outlined in the Power BI documentation.

NEW QUESTION 14
- (Topic 2)
You have a Fabric tenant that contains a new semantic model in OneLake. You use a Fabric notebook to read the data into a Spark DataFrame.
You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.
Solution: You use the following PySpark expression: df .sumary ()
Does this meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Yes, the df.summary() method does meet the goal. This method is used to compute specified statistics for numeric and string columns. By default, it provides statistics such as count, mean, stddev, min, and max. References = The PySpark API documentation details the summary() function and the statistics it provides.

NEW QUESTION 16
- (Topic 2)
You have a Fabric tenant that contains a lakehouse.
You plan to query sales data files by using the SQL endpoint. The files will be in an Amazon Simple Storage Service (Amazon S3) storage bucket.
You need to recommend which file format to use and where to create a shortcut. Which two actions should you include in the recommendation? Each correct answer
presents part of the solution.
NOTE: Each correct answer is worth one point.

A. Create a shortcut in the Files section.
B. Use the Parquet format
C. Use the CSV format.
D. Create a shortcut in the Tables section.
E. Use the delta format.

Answer: BD

Explanation:
You should use the Parquet format (B) for the sales data files because it is optimized for performance with large datasets in analytical processing and create a shortcut in the Tables section (D) to facilitate SQL queries through the lakehouse's SQL endpoint. References = The best practices for working with file formats and shortcuts in a lakehouse environment are covered in the lakehouse and SQL endpoint documentation provided by the cloud data platform services.

NEW QUESTION 20
HOTSPOT - (Topic 2)
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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
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.

NEW QUESTION 21
- (Topic 2)
You have a Fabric tenant that contains a workspace named Workspace^ Workspacel is assigned to a Fabric capacity.
You need to recommend a solution to provide users with the ability to create and publish custom Direct Lake semantic models by using external tools. The solution must follow the principle of least privilege.
Which three actions in the Fabric Admin portal should you include in the recommendation? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. From the Tenant settings, set Allow XMLA Endpoints and Analyze in Excel with on- premises datasets to Enabled
B. From the Tenant settings, set Allow Azure Active Directory guest users to access Microsoft Fabric to Enabled
C. From the Tenant settings, select Users can edit data models in the Power Bl service.
D. From the Capacity settings, set XMLA Endpoint to Read Write
E. From the Tenant settings, set Users can create Fabric items to Enabled
F. From the Tenant settings, enable Publish to Web

Answer: ACD

Explanation:
For users to create and publish custom Direct Lake semantic models using external tools, following the principle of least privilege, the actions to be included are enabling XMLA Endpoints (A), editing data models in Power BI service (C), and setting XMLA Endpoint to Read-Write in the capacity settings (D). References = More information can be found in the Admin portal of the Power BI service documentation, detailing tenant and capacity settings.

NEW QUESTION 25
- (Topic 2)
You have a Fabric tenant that contains a warehouse.
You are designing a star schema model that will contain a customer dimension. The customer dimension table will be a Type 2 slowly changing dimension (SCD).
You need to recommend which columns to add to the table. The columns must NOT already exist in the source.
Which three types of columns should you recommend? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. an effective end date and time
B. a foreign key
C. a surrogate key
D. a natural key
E. an effective start date and time

Answer: ACE

Explanation:
For a Type 2 slowly changing dimension (SCD), you typically need to add the following types of columns that do not exist in the source system:
? An effective start date and time (E): This column records the date and time from which the data in the row is effective.
? An effective end date and time (A): This column indicates until when the data in the row was effective. It allows you to keep historical records for changes over time.
? A surrogate key (C): A surrogate key is a unique identifier for each row in a table, which is necessary for Type 2 SCDs to differentiate between historical and current records.
References: Best practices for designing slowly changing dimensions in data warehousing solutions, which include Type 2 SCDs, are commonly discussed in data warehousing and business intelligence literature and would be part of the modeling guidance in a Fabric tenant's documentation.

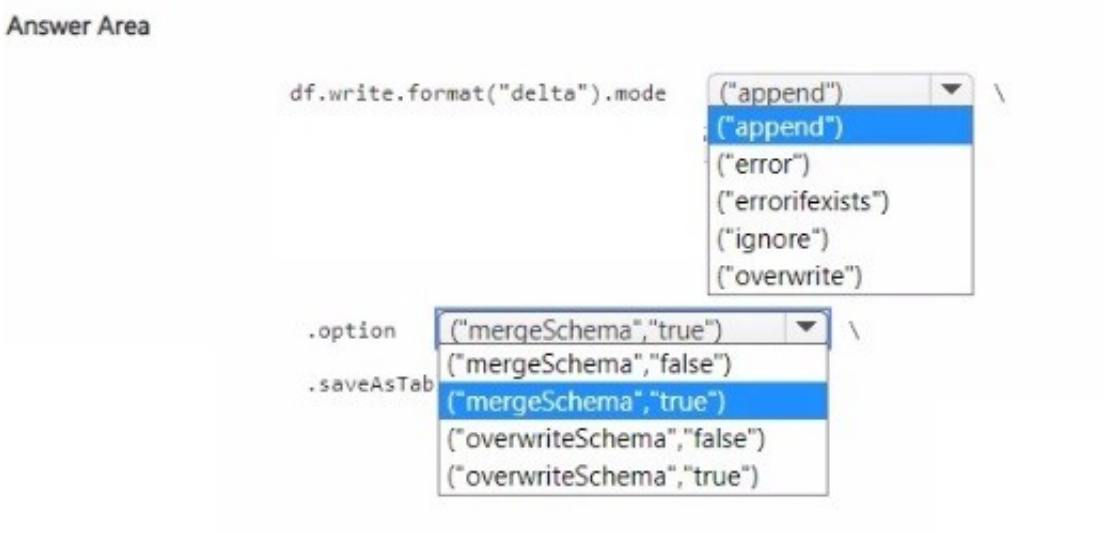
NEW QUESTION 27
- (Topic 2)
You have a Fabric tenant that contains a lakehouse named Lakehouse1.
You need to prevent new tables added to Lakehouse1 from being added automatically to the default semantic model of the lakehouse.
What should you configure? (5)

A. the semantic model settings
B. the Lakehouse1 settings
C. the workspace settings
D. the SQL analytics endpoint settings

Answer: A

Explanation:
To prevent new tables added to Lakehouse1 from being automatically added to the default semantic model, you should configure the semantic model settings. There should be an option within the settings of the semantic model to include or exclude new tables by default. By adjusting these settings, you can control the automatic inclusion of new tables.
References: The management of semantic models and their settings would be covered under the documentation for the semantic layer or modeling features of the Fabric tenant's lakehouse solution.

NEW QUESTION 32
HOTSPOT - (Topic 2)
You have a Fabric tenant that contains lakehouse named Lakehousel. Lakehousel contains a Delta table with eight columns. You receive new data that contains the same eight columns and two additional columns.
You create a Spark DataFrame and assign the DataFrame to a variable named df. The DataFrame contains the new data. You need to add the new data to the Delta table to meet the following requirements:
• Keep all the existing rows.
• Ensure that all the new data is added to the table.
How should you complete the code? To answer, select the appropriate options in the answer area.



A. Mastered
B. Not Mastered

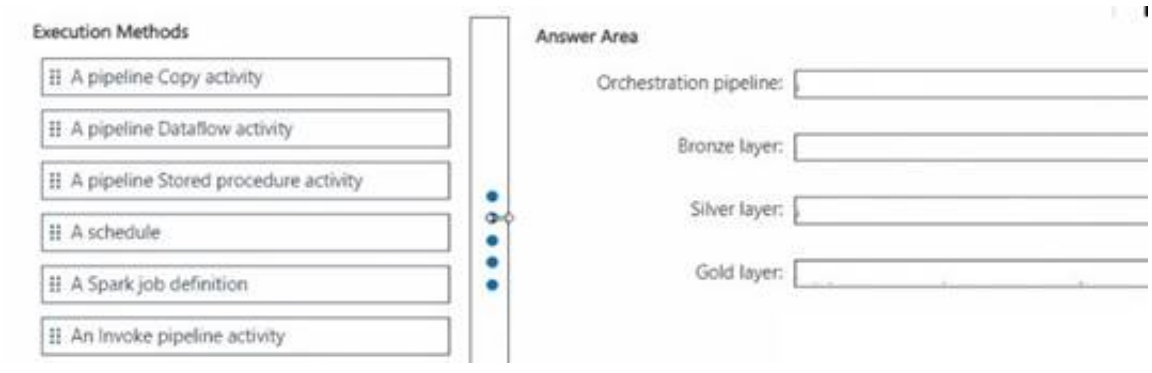
Answer: A

Explanation:
o add new data to the Delta table while meeting the specified requirements:
? You should use the append mode to ensure that all new data is added to the table without affecting the existing rows.
? You should set the mergeSchema option to true to allow the schema of the Delta table to be updated with the new columns found in the DataFrame.
The completed code would look like this:
df.write.format("delta").mode("append") option("mergeSchema", "true") saveAsTable("Lakehouse1.TableName")

NEW QUESTION 36
DRAG DROP - (Topic 2)
You are implementing a medallion architecture in a single Fabric workspace.
You have a lakehouse that contains the 8ronze and Silver layers and a warehouse that contains the Gold layer.
You create the items required to populate the layers as shown in the following table.

Layer	Data integration tool
Bronze	Pipelines with Copy activities
Silver	Dataflows
Gold	Stored procedures

You need to ensure that the layers are populated daily in sequential order such that Silver is populated only after Bronze is complete, and Gold is populated only after Silver is complete. The solution must minimize development effort and complexity.
What should you use to execute each set of items? To answer, drag the appropriate options to the correct items. Each option 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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
To execute each set of items in sequential order with minimized development effort and complexity, you should use the following options:
? Orchestration pipeline: Use a pipeline with an Invoke pipeline activity. This allows for orchestrating and scheduling the execution of other pipelines, ensuring they run in the correct sequence.
? Bronze layer: Implement a pipeline Copy activity. This aligns with the table indicating that the Bronze layer uses pipelines with Copy activities for data integration.
? Silver layer: Implement a pipeline Dataflow activity. The table specifies that Dataflows are used for the Silver layer.
? Gold layer: Implement a pipeline Stored procedure activity. Stored procedures are specified for the Gold layer according to the table.

NEW QUESTION 41
HOTSPOT - (Topic 2)
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.
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.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
? The results will form a hierarchy of folders for each partition key. - Yes
? The resulting file partitions can be read in parallel across multiple nodes. - Yes
? The resulting file partitions will use file compression. - No
Partitioning data by columns such as year, month, and day, as shown in the DataFrame write operation, organizes the output into a directory hierarchy that reflects the partitioning structure. This organization can improve the performance of read operations, as queries that filter by the partitioned columns can scan only the relevant directories. Moreover, partitioning facilitates parallelism because each partition can be processed independently across different nodes in a distributed system like Spark. However, the code snippet provided does not explicitly specify that file compression should be used, so we cannot assume that the output will be compressed without additional context.
References =
? DataFrame write partitionBy
? Apache Spark optimization with partitioning

NEW QUESTION 45
- (Topic 2)
You have a Fabric tenant that contains a warehouse.
You use a dataflow to load a new dataset from OneLake to the warehouse.
You need to add a Power Query step to identify the maximum values for the numeric columns.
Which function should you include in the step?

A. Tabl
B. MaxN
C. Table.Max
D. Table.Range
E. Table.Profile

Answer: B

Explanation:
The Table.Max function should be used in a Power Query step to identify the maximum values for the numeric columns. This function is designed to calculate the maximum value across each column in a table, which suits the requirement of finding maximum values for numeric columns. References = For detailed information on Power Query functions, including Table.Max, please refer to Power Query M function reference.

NEW QUESTION 48
HOTSPOT - (Topic 2)
You have a Fabric tenant.
You plan to create a Fabric notebook that will use Spark DataFrames to generate Microsoft Power BI visuals.
You run the following code.

```
from powerbiclient import QuickVisualize, get_dataset_config, Report

PBI_visualize = QuickVisualize(get_dataset_config(df))
PBI_visualize
```

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 code embeds an existing Power BI report.	○	○
The code creates a Power BI report.	○	○
The code displays a summary of the DataFrame.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
? The code embeds an existing Power BI report. - No
? The code creates a Power BI report. - No
? The code displays a summary of the DataFrame. - Yes
The code provided seems to be a snippet from a SQL query or script which is neither creating nor embedding a Power BI report directly. It appears to be setting up a DataFrame for use within a larger context, potentially for visualization in Power BI, but the code itself does not perform the creation or embedding of a report. Instead, it's likely part of a data processing step that summarizes data.
References =
? Introduction to DataFrames - Spark SQL
? Power BI and Azure Databricks

NEW QUESTION 49
- (Topic 2)
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

Answer: AB

Explanation:
User-defined aggregations (A) and query caching (C) are two features that can help reduce query execution time. User-defined aggregations allow precalculation of large datasets, and query caching stores the results of queries temporarily to speed up future queries. References = Microsoft Power BI documentation on performance optimization offers in-depth knowledge on these features.

NEW QUESTION 54
.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-600 Exam with Our Prep Materials Via below:

https://www.certleader.com/DP-600-dumps.html