Microsoft

Exam Questions DP-600

Implementing Analytics Solutions Using Microsoft Fabric

About Exambible

Your Partner of IT Exam

Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

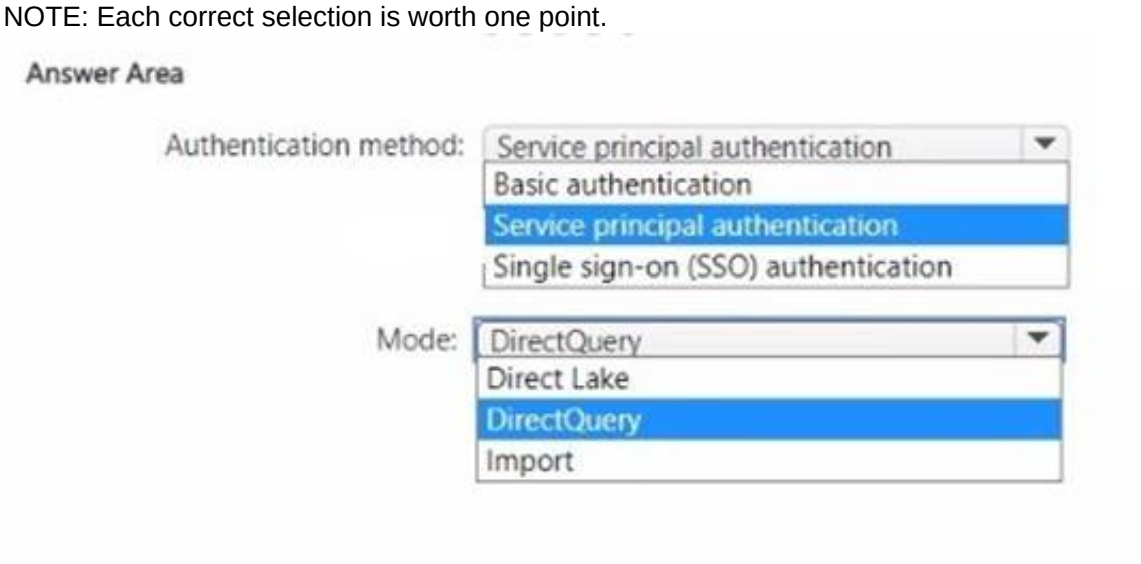
* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

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.

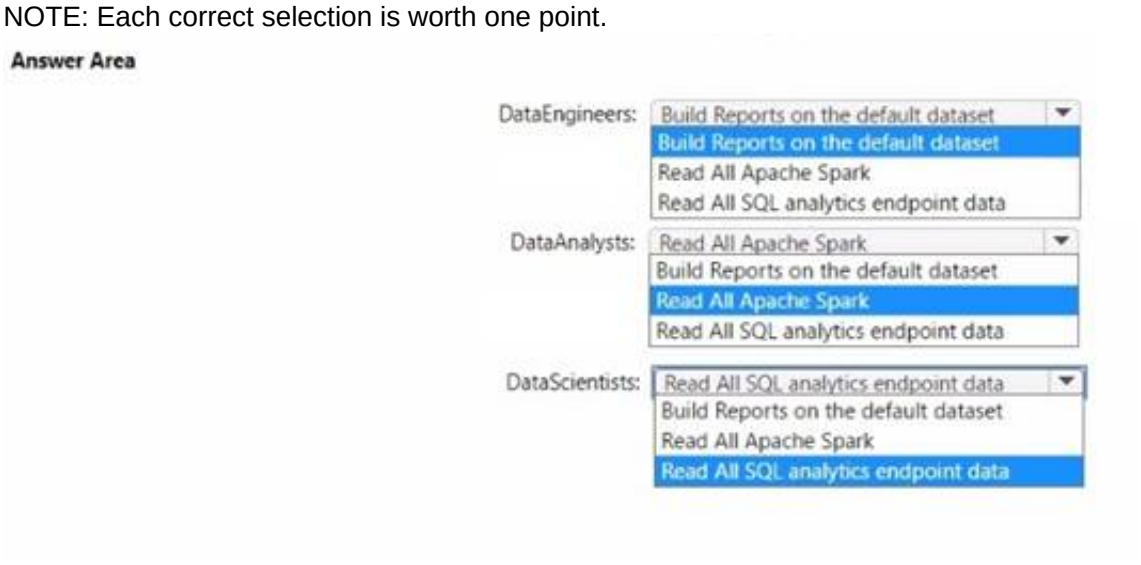


A. Mastered
B. Not Mastered

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
HOTSPOT - (Topic 1)
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
? 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.

NEW QUESTION 3

HOTSPOT - (Topic 1)
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

Answer Area

A. Mastered
B. Not Mastered

Answer: A

Explanation:
C:\Users\Waqas Shahid\Desktop\Mudassir\Untitled.jpg
? You should use CREATE VIEW to make the pricing group logic available for T- SQL queries.
? The CASE statement should be used to determine the pricing group based on the
list price.
The T-SQL statement should create a view that classifies products into pricing groups based on the list price. The CASE statement is the correct conditional logic to assign each product to the appropriate pricing group. This view will standardize the pricing group logic across different databases and semantic models.

NEW QUESTION 4
- (Topic 1)
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

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.

NEW QUESTION 5
- (Topic 2)

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.

Answer: E

Explanation:
A possible cause for the failure is that query folding is NOT occurring (D). Query folding helps optimize refresh by pushing down the query logic to the source system, reducing the amount of data processed and transferred, hence conserving resources. References = The Power BI documentation on incremental refresh and query folding provides detailed information on this topic.

NEW QUESTION 6
- (Topic 2)
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

Answer: A

Explanation:
A lakehouse (A) is the type of data store you should use. It supports dataflows to load and append data and ensures that Delta tables are Z-Order optimized and compacted automatically. References = The capabilities of a lakehouse and its support for Delta tables are described in the lakehouse and Delta table documentation.

NEW QUESTION 7
- (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 8
- (Topic 2)
You have a Fabric tenant that contains a lakehouse. You plan to use a visual query to merge two tables.
You need to ensure that the query returns all the rows that are present in both tables. Which type of join should you use?

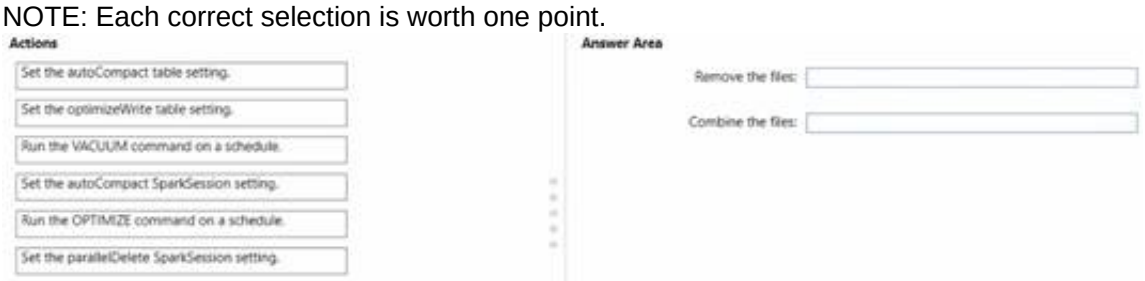
A. left outer
B. right anti
C. full outer
D. left anti
E. right outer
F. inner

Answer: C

Explanation:
When you need to return all rows that are present in both tables, you use a full outer join. This type of join combines the results of both left and right outer joins and returns all rows from both tables, with matching rows from both sides where available. If there is no match, the result is NULL on the side of the join where there is no match. References: Information about joins and their use in querying data in a lakehouse can be typically found in the SQL and data processing documentation of the Fabric tenant or lakehouse solutions.

NEW QUESTION 9
DRAG DROP - (Topic 2)
You have a Fabric tenant that contains a lakehouse named Lakehouse1
Readings from 100 loT 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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
? Remove the files: Run the VACUUM command on a schedule.
? Combine the files: Set the optimizeWrite table setting. or Run the OPTIMIZE command on a schedule.
To remove files that are no longer used, the VACUUM command is used in Delta Lake to clean up invalid files from a table. To combine smaller files into larger ones, you can either set the optimizeWrite setting to combine files during write operations or use the OPTIMIZE command, which is a Delta Lake operation used to compact small files into larger ones.

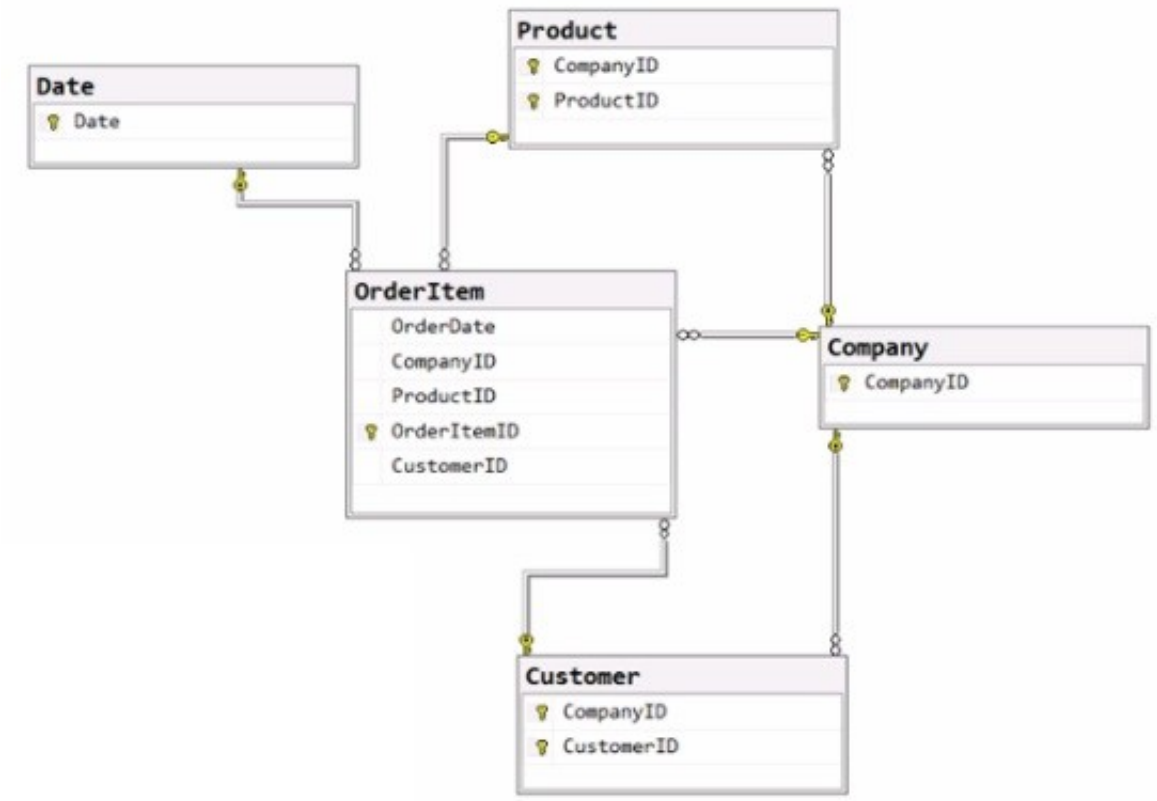
NEW QUESTION 10
- (Topic 2)
You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a subfolder named Subfolder1 that contains CSV files. You need to convert the CSV files into the delta format that has V-Order optimization enabled. What should you do from Lakehouse explorer?

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.

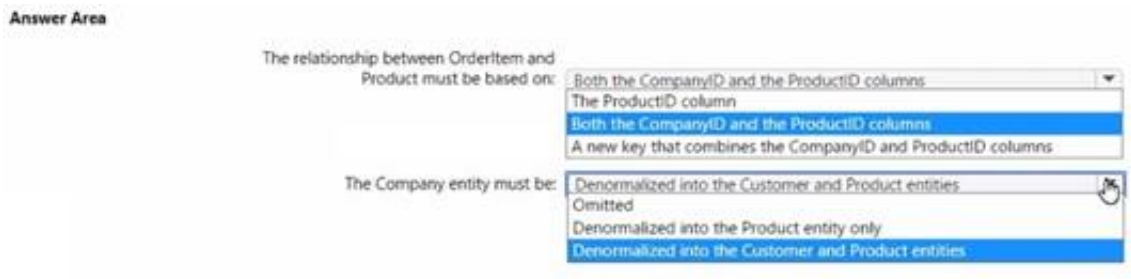
Answer: D

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 10
HOTSPOT - (Topic 2)
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
? The relationship between OrderItem and Product must be based on: Both the CompanyID and the ProductID columns
? The Company entity must be: Denormalized into the Customer and Product
entities
In a dimensional model, the relationships are typically based on foreign key constraints between the fact table (OrderItem) and dimension tables (Product, Customer, Date). Since CompanyID is present in both the OrderItem and Product tables, it acts as a foreign key in the relationship. Similarly, ProductID is a foreign key that relates these two tables. To enable analysis by date, product, and customer, the Company entity would need to be denormalized into the Customer and Product entities to ensure that the relevant company information is available within those dimensions for querying and reporting purposes. References =
? Dimensional modeling
? Star schema design

NEW QUESTION 11
- (Topic 2)
You have a Fabric tenant named Tenant1 that contains a workspace named WS1. WS1 uses a capacity named C1 and contains a dawset named DS1. You need to ensure read- write access to DS1 is available by using the XMLA endpoint. What should be modified first?

A. the DS1 settings
B. the WS1 settings
C. the C1 settings
D. the Tenant1 settings

Answer: C

Explanation:
To ensure read-write access to DS1 is available by using the XMLA endpoint, the C1 settings (which refer to the capacity settings) should be modified first. XMLA endpoint configuration is a capacity feature, not specific to individual datasets or workspaces. References = The configuration of XMLA endpoints in Power BI capacities is detailed in the Power BI documentation on dataset management.

NEW QUESTION 13
- (Topic 2)
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

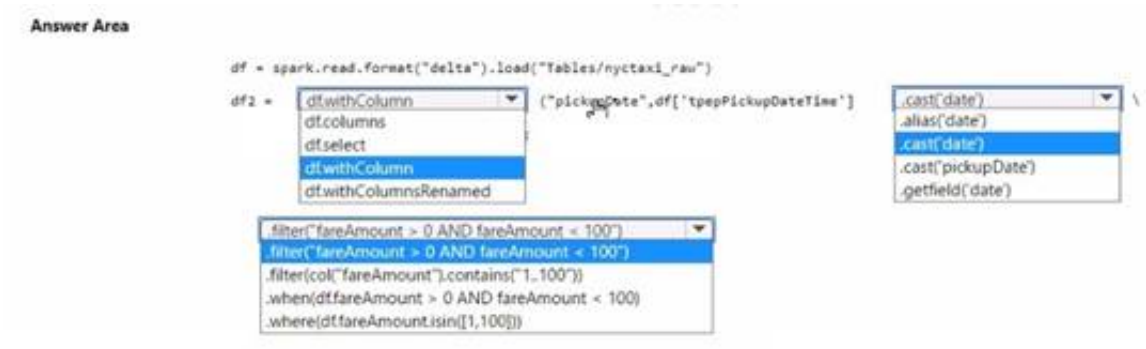
Answer: B

Explanation:
To identify whether throttling is occurring, you should use the Monitoring hub (B). This provides a centralized place where you can monitor and manage the health, performance, and reliability of your data estate, and see if the compute resources are being throttled. References = The use of the Monitoring hub for performance management and troubleshooting is detailed in the Azure Synapse Analytics documentation.

NEW QUESTION 15
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.



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 17
- (Topic 2)
You have a Fabric tenant that contains a semantic model. The model contains 15 tables.
You need to programmatically change each column that ends in the word Key to meet the following requirements:
• Hide the column.
• Set Nullable to False.
• Set Summarize By to None
• Set Available in MDX to False.
• Mark the column as a key column. What should you use?

A. Microsoft Power Bl Desktop
B. Tabular Editor
C. ALM Toolkit
D. DAX Studio

Answer: B

Explanation:
Tabular Editor is an advanced tool for editing Tabular models outside of Power BI Desktop that allows you to script out changes and apply them across multiple columns or tables. To accomplish the task programmatically, you would:
? Open the model in Tabular Editor.
? Create an Advanced Script using C# to iterate over all tables and their respective columns.
? Within the script, check if the column name ends with 'Key'.
? For columns that meet the condition, set the properties accordingly: IsHidden = true, IsNullable = false, SummarizeBy = None, IsAvailableInMDX = false.
? Additionally, mark the column as a key column.
? Save the changes and deploy them back to the Fabric tenant.
References: The ability to batch-edit properties using scripts in Tabular Editor is well- documented in the tool's official documentation and user community resources.

NEW QUESTION 18
- (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?

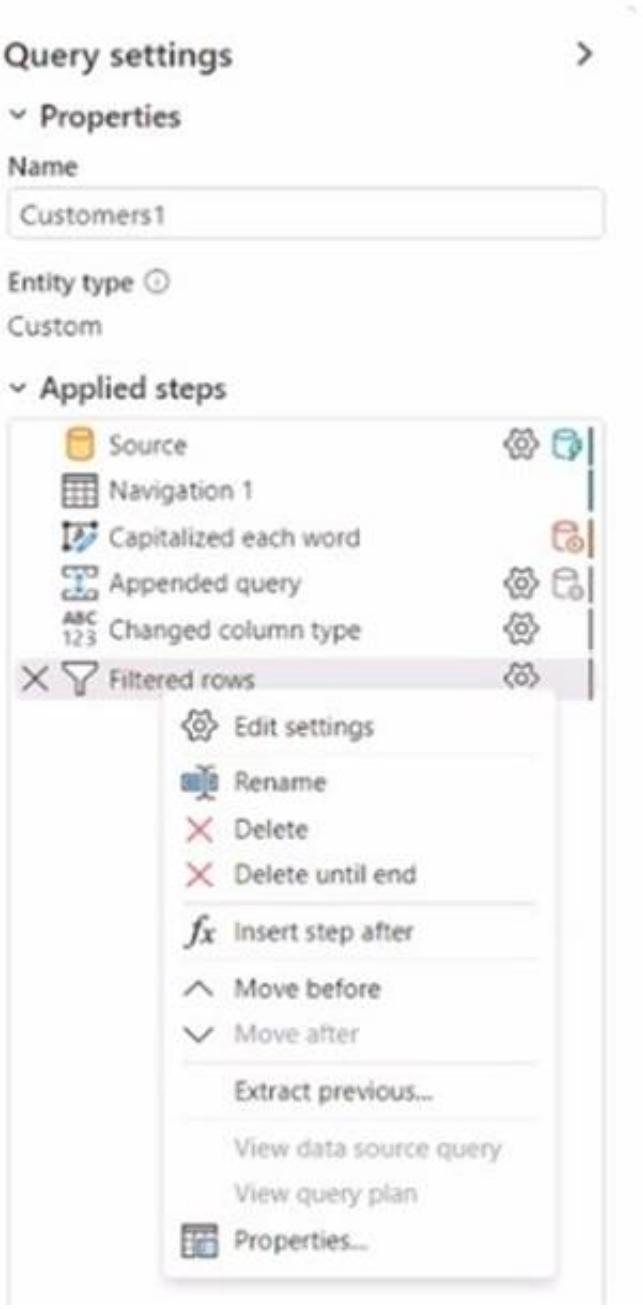
A. Change the visual type of Orders By Dale.
B. Enable automatic page refresh.
C. Optimize the DAX query of Orders By Date by using DAX Studio.
D. Reduce the number of visuals in Report1.

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 21
HOTSPOT - (Topic 2)
You have a Fabric tenant that contains two lakehouses.
You are building a dataflow that will combine data from the lakehouses. The applied steps from one of the queries in the dataflow is 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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Folding in Power Query refers to operations that can be translated into source queries. In this case, "some" of the steps can be folded, which means that some transformations will be executed at the data source level. The steps that cannot be folded will be executed within the Power Query engine. Custom steps, especially those that are not standard query operations, are usually executed within Power Query engine rather than being pushed down to the source system.
References =
? Query folding in Power Query
? Power Query M formula language

NEW QUESTION 25
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 30
- (Topic 2)
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.

Answer: AD

Explanation:
To reduce memory usage and refresh time, splitting the OrderDateTime into separate date and time columns (A) can help optimize the model because date/time data types can be more memory-intensive than separate date and time columns. Moreover, replacing TotalSalesAmount with a measure (D) instead of a calculated column ensures that the calculation is performed at query time, which can reduce the size of the model as the value is not stored but calculated on the fly. References = The best practices for optimizing Power BI models are detailed in the Power BI documentation, which recommends using measures for calculations that don't need to be stored and adjusting data types to improve performance.

NEW QUESTION 31
- (Topic 2)
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

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.

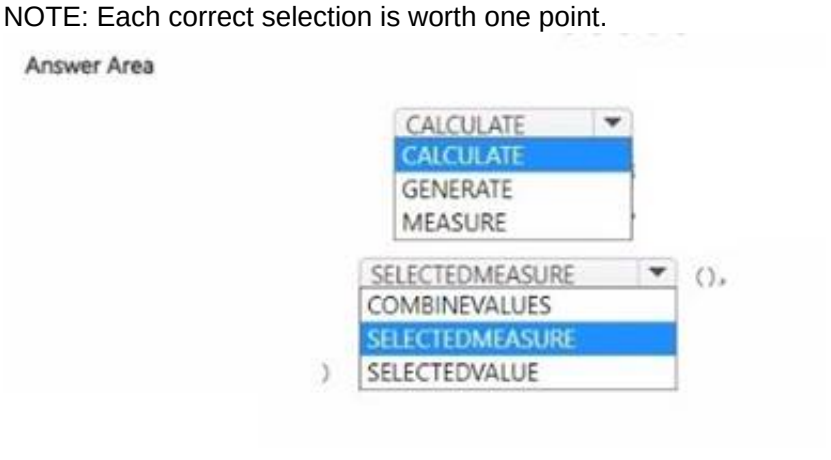
NEW QUESTION 34
- (Topic 2)
You have a Microsoft Power Bl semantic model that contains measures. The measures use multiple calculate functions and a filter function.
You are evaluating the performance of the measures.
In which use case will replacing the filter function with the keepfilters function reduce execution time?

A. when the filter function uses a nested calculate function
B. when the filter function references a column from a single table that uses Import mode
C. when the filter function references columns from multiple tables
D. when the filter function references a measure

Answer: A

Explanation:
The KEEPFILTERS function modifies the way filters are applied in calculations done through the CALCULATE function. It can be particularly beneficial to replace the FILTER function with KEEPFILTERS when the filter context is being overridden by nested CALCULATE functions, which may remove filters that are being applied on a column. This can potentially reduce execution time because KEEPFILTERS maintains the existing filter context and allows the nested CALCULATE functions to be evaluated more efficiently. References: This information is based on the DAX reference and performance optimization guidelines in the Microsoft Power BI documentation.

NEW QUESTION 35
HOTSPOT - (Topic 2)
You have a Microsoft Power Bl semantic model. You plan to implement calculation groups.
You need to create a calculation item that will change the context from the selected date to month-to-date (MTD).
How should you complete the DAX expression? 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:
To create a calculation item that changes the context from the selected date to month-to- date (MTD), the appropriate DAX expression involves using the CALCULATE function to alter the filter context and the DATESMTD function to specify the month-to-date context. The correct completion for the DAX expression would be:
? In the first dropdown, select CALCULATE.
? In the second dropdown, select SELECTEDMEASURE. This would create a DAX expression in the form:
CALCULATE(SELECTEDMEASURE(),
DATESMTD('Date'[DateColumn])
)

NEW QUESTION 39
- (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)

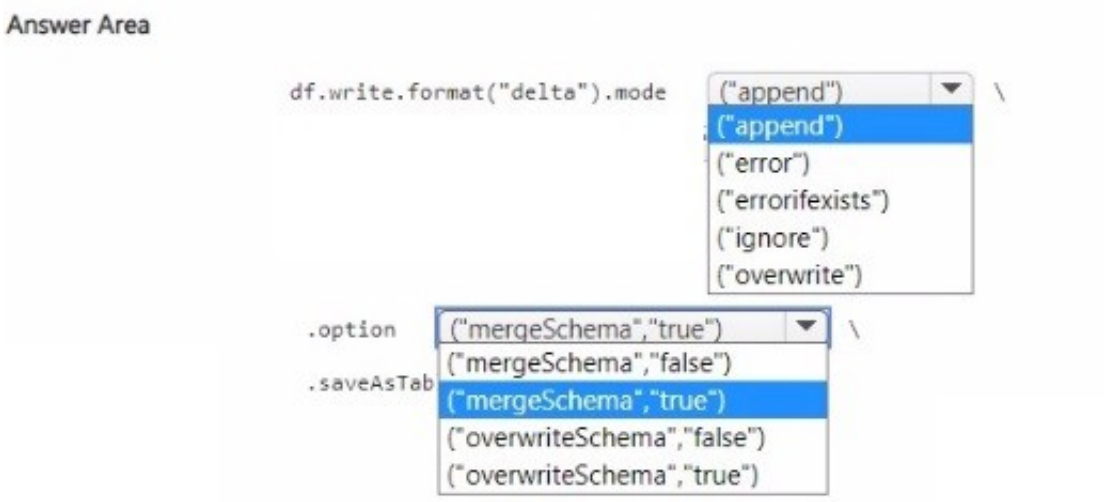
A. the semantic model settings
B. the Lakehouse1 settings
C. the workspace settings
D. the SQL analytics endpoint settings

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 43
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

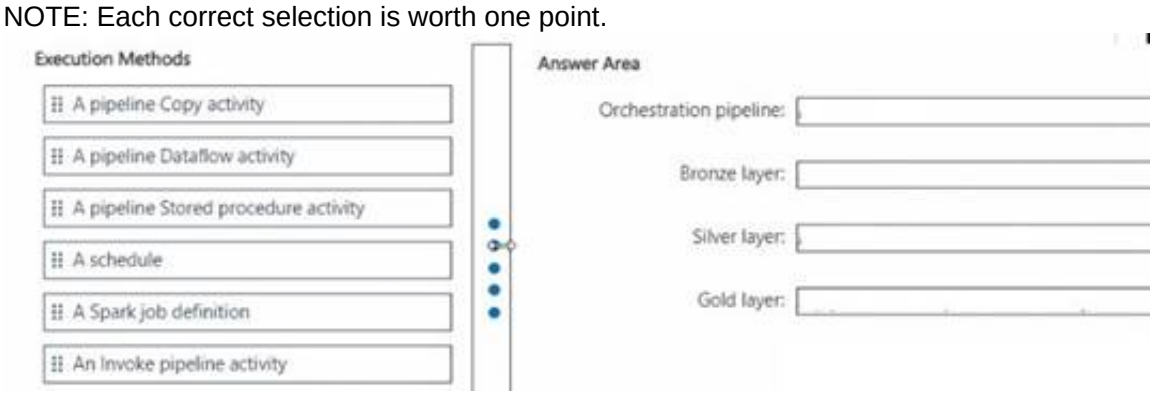
Answer: A

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 44
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 48
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 53
- (Topic 2)
You have a Fabric tenant.
You are creating a Fabric Data Factory pipeline.
You have a stored procedure that returns the number of active customers and their average sales for the current month.
You need to add an activity that will execute the stored procedure in a warehouse. The returned values must be available to the downstream activities of the pipeline.
Which type of activity should you add?

A. Stored procedure
B. Get metadata
C. Lookup
D. Copy data

Answer: C

Explanation:
In a Fabric Data Factory pipeline, to execute a stored procedure and make the returned values available for downstream activities, the Lookup activity is used. This activity can retrieve a dataset from a data store and pass it on for further processing. Here's how you would use the Lookup activity in this context:
? Add a Lookup activity to your pipeline.
? Configure the Lookup activity to use the stored procedure by providing the necessary SQL statement or stored procedure name.
? In the settings, specify that the activity should use the stored procedure mode.
? Once the stored procedure executes, the Lookup activity will capture the results and make them available in the pipeline's memory.
? Downstream activities can then reference the output of the Lookup activity. References: The functionality and use of Lookup activity within Azure Data Factory is documented in Microsoft's official documentation for Azure Data Factory, under the section
for pipeline activities.

NEW QUESTION 56
- (Topic 2)
You have a Fabric tenant that contains a lakehouse named lakehouse1. Lakehouse1 contains a table named Table1.
You are creating a new data pipeline.
You plan to copy external data to Table1. The schema of the external data changes regularly.
You need the copy operation to meet the following requirements:
• Replace Table1 with the schema of the external data.
• Replace all the data in Table1 with the rows in the external data.
You add a Copy data activity to the pipeline. What should you do for the Copy data activity?

A. From the Source tab, add additional columns.
B. From the Destination tab, set Table action to Overwrite.
C. From the Settings tab, select Enable staging
D. From the Source tab, select Enable partition discovery

E. From the Source tab, select Recursively

Answer: B

Explanation:
For the Copy data activity, from the Destination tab, setting Table action to Overwrite (B) will ensure that Table1 is replaced with the schema and rows of the external data, meeting the requirements of replacing both the schema and data of the destination table. References = Information about Copy data activity and table actions in Azure Data Factory, which can be applied to data pipelines in Fabric, is available in the Azure Data Factory documentation.

NEW QUESTION 59
- (Topic 2)
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)

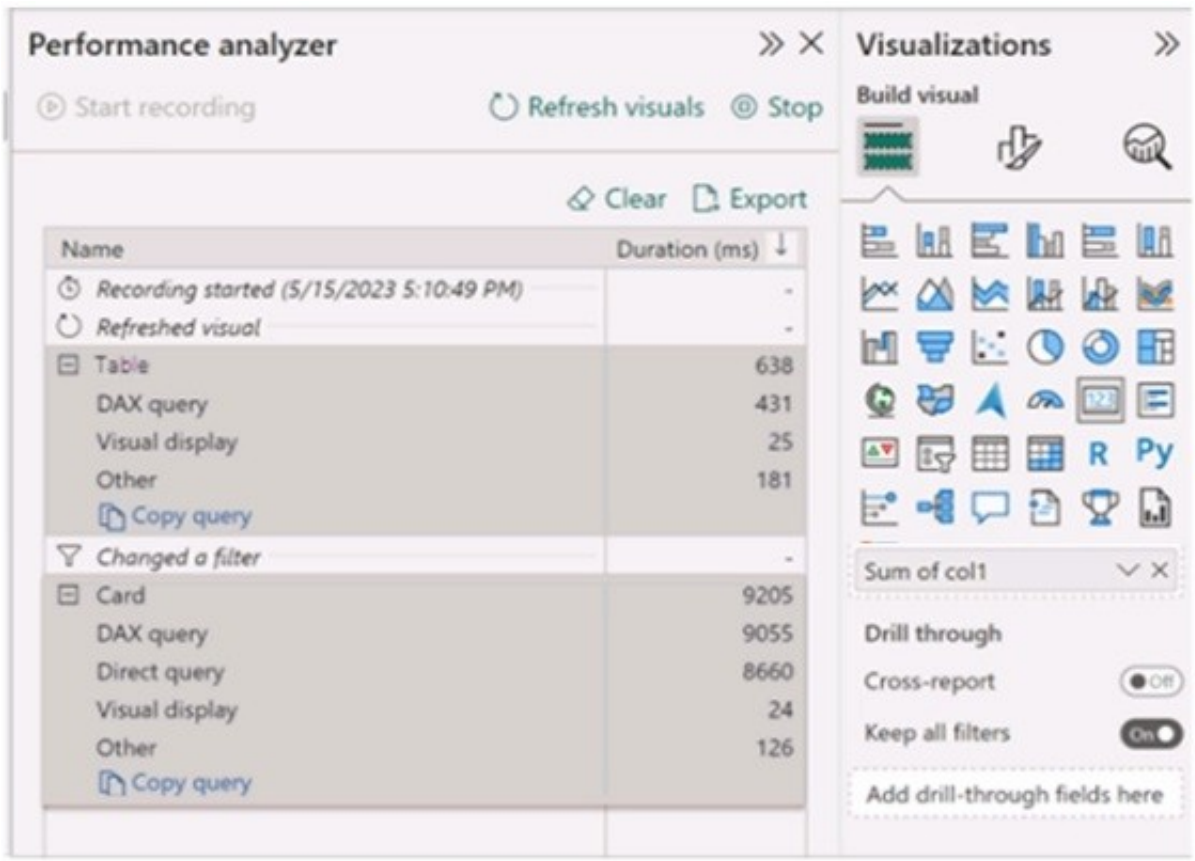
```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING TotalAmount > 10000
```

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

Answer: B

Explanation:
The correct query to use in order to return data for the year 2023 that displays ProductID, ProductName, and has a summarized Amount greater than 10,000 is Option B. The reason is that it uses the GROUP BY clause to organize the data by ProductID and ProductName and then filters the result using the HAVING clause to only include groups where the sum of Amount is greater than 10,000. Additionally, the DATEPART(YEAR, SaleDate) = '2023' part of the HAVING clause ensures that only records from the year 2023 are included. References = For more information, please visit the official documentation on T-SQL queries and the GROUP BY clause at T-SQL GROUP BY.

NEW QUESTION 62
HOTSPOT - (Topic 2)
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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
? The Direct Lake fallback behavior is set to: DirectQueryOnly
? The query for the table visual is executed by using: DirectQuery
In the context of Microsoft Power BI, when using DirectQuery in Direct Lake mode, there is no caching of data and all queries are sent directly to the underlying data source. The Performance Analyzer tool shows the time taken for different operations, and from the options provided, it indicates that DirectQuery mode is being used for the visuals, which is consistent with the Direct Lake setting. DirectQueryOnly as the fallback behavior ensures that only DirectQuery will be used without reverting to import mode.

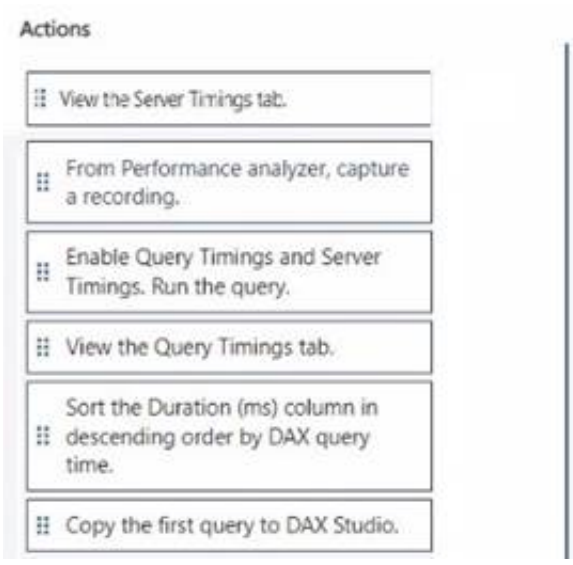
NEW QUESTION 67
- (Topic 2)
You have a Fabric workspace that contains a DirectQuery semantic model. The model queries a data source that has 500 million rows.
You have a Microsoft Power Bl report named Report1 that uses the model. Report! contains visuals on multiple pages.
You need to reduce the query 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 71
DRAG DROP - (Topic 2)
You have a Fabric tenant that contains a Microsoft Power Bl report named Report 1. Report1 is slow to render. You suspect that an inefficient DAX query is being executed.
You need to identify the slowest DAX query, and then review how long the query spends in the formula engine as compared to the storage engine.
Which five 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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
To identify the slowest DAX query and analyze the time it spends in the formula engine compared to the storage engine, you should perform the following actions in sequence:
? From Performance analyzer, capture a recording.
? View the Server Timings tab.
? Enable Query Timings and Server Timings. Run the query.
? View the Query Timings tab.
? Sort the Duration (ms) column in descending order by DAX query time.

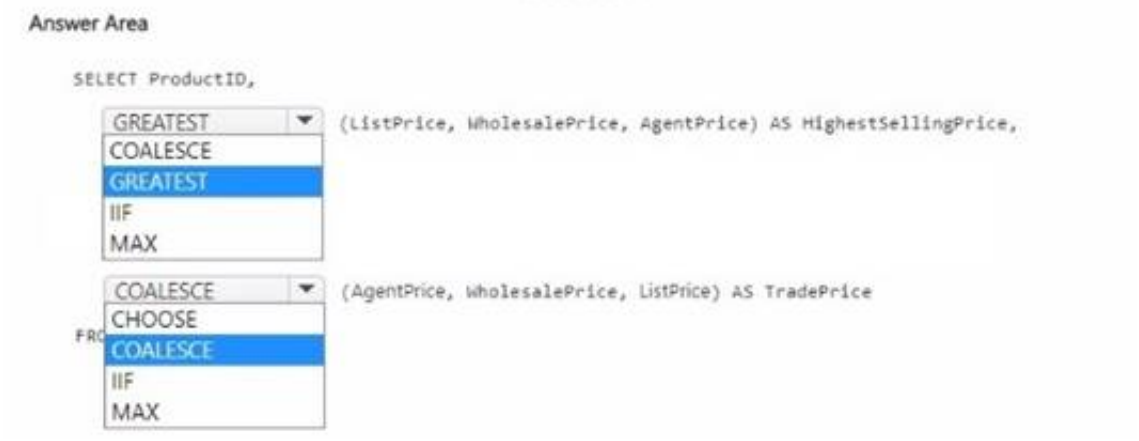
NEW QUESTION 75
HOTSPOT - (Topic 2)
You have a Fabric warehouse that contains a table named Sales.Products. Sales.Products contains the following columns.

Name	Data type	Nullable
ProductID	Integer	No
ProductName	Varchar(30)	No
ListPrice	Decimal(18, 2)	No
WholesalePrice	Decimal(18, 2)	Yes
AgentPrice	Decimal(18, 2)	Yes

You need to write a T-SQL query that will return the following columns.

Name	Description
ProductID	Return the ProductID value
HighestSellingPrice	Returns the highest value from ListPrice, WholesalePrice, and AgentPrice
TradePrice	Returns the AgentPrice value if present, otherwise returns the WholesalePrice value if present, otherwise returns the ListPrice value

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



A. Mastered
B. Not Mastered

Answer: A

Explanation:
? For the HighestSellingPrice, you should use the GREATEST function to find the highest value from the given price columns. However, T-SQL does not have a GREATEST function as found in some other SQL dialects, so you would typically use a CASE statement or an IIF statement with nested MAX functions. Since neither of those are provided in the options, you should select MAX as a placeholder to indicate the function that would be used to find the highest value if combining multiple MAX functions or a similar logic was available.

? For the TradePrice, you should use the COALESCE function, which returns the first non-null value in a list. The COALESCE function is the correct choice as it will return AgentPrice if it's not null; if AgentPrice is null, it will check WholesalePrice, and if that is also null, it will return ListPrice.
The complete code with the correct SQL functions would look like this:
SELECT ProductID,
MAX(ListPrice, WholesalePrice, AgentPrice) AS HighestSellingPrice, -- MAX is used as a placeholder
COALESCE(AgentPrice, WholesalePrice, ListPrice) AS TradePrice FROM Sales.Products
Select MAX for HighestSellingPrice and COALESCE for TradePrice in the answer area.

NEW QUESTION 76
.....

Relate Links

100% Pass Your DP-600 Exam with Exambible Prep Materials

https://www.exambible.com/DP-600-exam/

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - https://www.exambible.com/