



Product Questions: 80

Topic 1, Litware. Inc. Case Study

Overview

Litware. Inc. is a manufacturing company that has offices throughout North America.

a. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

Existing Environment

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

Table 1: Data sources

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

- Survey
- Question
- Response

For each survey submitted, the following occurs:

- One row is added to the Survey table.
- One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score. Customers can submit a survey after each purchase.

User Problems

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

Planned Changes

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Litware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity.

The following three workspaces will be created:

- AnalyticsPOC: Will contain the data store, semantic models, reports, pipelines, dataflows, and notebooks used to populate the data store
- DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate **Onelake**
- DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

- A data store (type to be decided)
- A custom semantic model
- A default semantic model
- Interactive reports

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

Technical Requirements

The data store must support the following:

- Read access by using T-SQL or Python
- Semi-structured and unstructured data
- Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model.

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model.

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location.

The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

- List prices that are less than or equal to 50 are in the low pricing group.
- List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
- List prices that are greater than 1,000 are in the high pricing group.

Security Requirements

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC. Litware identifies the following security requirements for the Fabric items in the

AnalyticsPOC workspace:

- Fabric administrators will be the workspace administrators.
- The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
- The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
- The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook.
- The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
- The date dimension must be available to all users of the data store.
- The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

- FabricAdmins: Fabric administrators
- AnalyticsTeam: All the members of the analytics team
- DataAnalysts: The data analysts on the analytics team
- DataScientists: The data scientists on the analytics team
- Data Engineers: The data engineers on the analytics team
- Analytics Engineers: The analytics engineers on the analytics team

Report Requirements

The data analysis must create a customer satisfaction report that meets the following requirements:

- Enables a user to select a product to filter customer survey responses to only those who have purchased that product
- Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date
- Shows data as soon as the data is updated in the data store
- Ensures that the report and the semantic model only contain data from the current and previous year
- Ensures that the report respects any table-level security specified in the source data store
- Minimizes the execution time of report queries

Question: 1

HOTSPOT

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

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

NOTE: Each correct selection is worth one point.

Answer Area

DataEngineers:	<div>Build Reports on the default dataset ▼ Build Reports on the default dataset Read All Apache Spark Read All SQL analytics endpoint data</div>
DataAnalysts:	<div>Read All Apache Spark ▼ Build Reports on the default dataset Read All Apache Spark Read All SQL analytics endpoint data</div>
DataScientists:	<div>Read All SQL analytics endpoint data ▼ Build Reports on the default dataset Read All Apache Spark Read All SQL analytics endpoint data</div>

Answer:

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.

Question: 2

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

```

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

```

Answer:

Explanation:

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

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

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

To calculate the average overall satisfaction score using DAX, you would need to use the AVERAGE function on the response values related to satisfaction questions. The DATESINPERIOD function will help in calculating the rolling average over the last 12 months.

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
```

Question: 3

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

```

CREATE [ ] [dbo].[ProductsWithPricingGroup]
AS
SELECT ProductId,
        ProductName,
        ProductCategory,
        ListPrice,
        [ ]
        WHEN ListPrice <= 50 THEN 'low'
        [ ]

END AS PricingGroup
FROM dbo.Products

```

Answer Area

```

CREATE [VIEW] [dbo].[ProductsWithPricingGroup]
AS
SELECT ProductId,
        ProductName,
        ProductCategory,
        ListPrice,
        CASE
        CASE THEN 'low'
        COALESCE
        IIF
        SET

        WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
        WHEN (ListPrice > 50 AND ListPrice <= 1000 ) THEN 'medium'
        WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
        WHEN ListPrice BETWEEN 50 AND 1000 ) THEN 'medium'

END AS PricingGroup
FROM dbo.Products

```

Answer:

Explanation:

Answer Area

```

CREATE VIEW [dbo].[ProductswithPricingGroup]
AS
SELECT ProductID,
ProductCategory,
ListPrice,
CASE
WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
WHEN (ListPrice > 50 AND ListPrice <= 1000 ) THEN 'medium'
WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
WHEN ListPrice BETWEEN 50 AND 1000 ) THEN 'medium'
END AS PricingGroup
FROM dbo.Products

```

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.

Question: 4

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

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

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. Reference = You can learn

more about dataflows and their use in Power BI environments in Microsoft's Power BI documentation.

Question: 5

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. Reference = For further understanding, Microsoft's documentation on the lakehouse architecture provides insights into how it supports various data types and analytical operations.

Question: 6

You need to recommend a solution to prepare the tenant for the PoC.

Which two actions should you recommend performing from the Fabric Admin portal? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Enable the Users can try Microsoft Fabric paid features option for specific security groups.
- B. Enable the Allow Azure Active Directory guest users to access Microsoft Fabric option for specific security groups.
- C. Enable the Users can create Fabric items option and exclude specific security groups.
- D. Enable the Users can try Microsoft Fabric paid features option for the entire organization.
- E. Enable the Users can create Fabric items option for specific security groups.

Answer: A, E

Explanation:

The PoC is planned to be completed using a Fabric trial capacity, which implies that users involved in the PoC should be able to try paid features. However, this should be limited to specific security groups involved in the PoC to prevent the entire organization from accessing these features before the trial is proven successful (A). The ability for users to create Fabric items should also be enabled for specific security groups to ensure that only the relevant team members participating in the PoC can create items in the Fabric environment (E).

Question: 7

You need to ensure the data loading activities in the AnalyticsPOC workspace are executed in the appropriate sequence. The solution must meet the technical requirements.
What should you do?

- A. Create a pipeline that has dependencies between activities and schedule the pipeline.
- B. Create and schedule a Spark job definition.
- C. Create a dataflow that has multiple steps and schedule the dataflow.
- D. Create and schedule a Spark notebook.

Answer: A

Explanation:

To meet the technical requirement that data loading activities must ensure the raw and cleansed data is updated completely before populating the dimensional model, you would need a mechanism that allows for ordered execution. A pipeline in Microsoft Fabric with dependencies set between activities can ensure that activities are executed in a specific sequence. Once set up, the pipeline can be scheduled to run at the required intervals (hourly or daily depending on the data source).

Question: 8

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: A, B

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.

Question: 9

HOTSPOT

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.

Answer Area

Authentication method: Service principal authentication
Basic authentication
Service principal authentication
Single sign-on (SSO) authentication

Mode: DirectQuery
Direct Lake
DirectQuery
Import

Answer:

Explanation:

Answer Area

Authentication method: Service principal authentication

Mode: DirectQuery

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

Topic 2, Misc. Questions

Question: 10

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 and 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. Reference = For more information, please visit the official documentation on T-SQL queries and the GROUP BY clause at T-SQL GROUP BY.

Question: 11

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.

Answer Area

```
WITH CUSTOMERBASE AS (
    SELECT [CustomerID]
    , [CustomerName]
    , [PostalCode]
    , [LastUpdated]
    , X = ROW_NUMBER() OVER (PARTITION BY CustomerID ORDER BY LastUpdated DESC)
    FROM [LakehousePOC].[dbo].[CustomerChanges]
)
SELECT CustomerID, CustomerName, PostalCode, LastUpdated
FROM CUSTOMERBASE
WHERE X = 1
```

Answer:

Explanation:

Answer Area

```
WITH CUSTOMERBASE AS (
    SELECT [CustomerID]
    , [CustomerName]
    , [PostalCode]
    , [LastUpdated]
    , X = ROW_NUMBER() OVER (PARTITION BY CustomerID ORDER BY LastUpdated DESC)
    FROM [LakehousePOC].[dbo].[CustomerChanges]
)
SELECT CustomerID, CustomerName, PostalCode, LastUpdated
FROM CUSTOMERBASE
WHERE X = 1
```

In the ROW_NUMBER() function, choose OVER (PARTITION BY CustomerID ORDER BY LastUpdated DESC).

In the WHERE clause, choose WHERE X = 1.

To select the most recent row for each customer ID, you use the ROW_NUMBER() window function partitioned by CustomerID and ordered by LastUpdated in descending order. This will assign a row number of 1 to the most recent update for each customer. By selecting rows where the row number

(X) is 1, you get the latest update per customer.

Reference =

Use the OVER clause to aggregate data per partition

Use window functions

Question: 12

HOTSPOT

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.	<input type="radio"/>	<input type="radio"/>
The code creates a Power BI report.	<input type="radio"/>	<input type="radio"/>
The code displays a summary of the DataFrame.	<input type="radio"/>	<input type="radio"/>

Answer:

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.

Reference =

Introduction to DataFrames - Spark SQL

Power BI and Azure Databricks

Question: 13

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

Lakehouse1 contains the following tables:

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

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

- A. ReadTable3.
- B. Update the data Table3.
- C. ReadTable2.
- D. Update the data in Table1.

Answer: D

Explanation:

Question: 14

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. Table.MaxN
- B. Table.Max
- C. Table.Range
- D. 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. Reference = For detailed information on Power Query functions, including Table.Max, please refer to Power Query M function reference.

Question: 15

You have a Fabric tenant that contains a machine learning model registered in a Fabric workspace.
You need to use the model to generate predictions by using the predict function in a fabric notebook.
Which two languages can you use to perform model scoring? Each correct answer presents a complete solution. NOTE: Each correct answer is worth one point.

- A. T-SQL
- B. DAX EC.
- C. Spark SQL
- D. PySpark

Answer: C, D

Explanation:

The two languages you can use to perform model scoring in a Fabric notebook using the predict function are Spark SQL (option C) and PySpark (option D). These are both part of the Apache Spark

ecosystem and are supported for machine learning tasks in a Fabric environment. Reference = You can find more information about model scoring and supported languages in the context of Fabric notebooks in the official documentation on Azure Synapse Analytics.

Question: 16

You are analyzing the data in a Fabric notebook.
You have a Spark DataFrame assigned to a variable named df.
You need to use the Chart view in the notebook to explore the data manually.
Which function should you run to make the data available in the Chart view?

- A. displayMTML
- B. show
- C. write
- D. display

Answer: D

Explanation:

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

Question: 17

You have a Fabric tenant that contains a Microsoft Power BI report named Report 1. Report1 includes a Python visual. Data displayed by the visual is grouped automatically and duplicate rows are NOT displayed. You need all rows to appear in the visual. What should you do?

- A. Reference the columns in the Python code by index.
- B. Modify the Sort Column By property for all columns.
- C. Add a unique field to each row.
- D. Modify the Summarize By property for all columns.

Answer: C

Explanation:

To ensure all rows appear in the Python visual within a Power BI report, option C, adding a unique field to each row, is the correct solution. This will prevent automatic grouping by unique values and allow for all instances of data to be represented in the visual. Reference = For more on Power BI Python visuals and how they handle data, please refer to the Power BI documentation.

Question: 18

DRAG DROP

You have a Fabric tenant that contains a semantic model. The model contains data about retail

stores.

You need to write a DAX query that will be executed by using the XMLA endpoint. The query must return a table of stores that have opened since December 1, 2023.

How should you complete the DAX expression? 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.

The screenshot shows a DAX query editor interface. On the left, under the 'Values' pane, there are five buttons: DEFINE, EVALUATE, FILTER, SUMMARIZE, and TABLE. On the right, under the 'Answer Area', the following DAX query is partially entered:

```
VAR _SalesSince =
    DATE ( 2023, 12, 01 )
    FILTER (
        [ ] ( Store, Store[Name], Store[OpenDate] ),
        Store[OpenDate] >= _SalesSince
    )
```

The [] in the FILTER function represents a blank space where a table or expression should be placed.

Answer:

Explanation:

The correct order for the DAX expression would be:

```
DEFINE VAR _SalesSince = DATE ( 2023, 12, 01 )
```

```
EVALUATE
```

```
FILTER (
```

```
SUMMARIZE ( Store, Store[Name], Store[OpenDate] ),
```

```
Store[OpenDate] >= _SalesSince )
```

In this DAX query, you're defining a variable `_SalesSince` to hold the date from which you want to filter the stores. `EVALUATE` starts the definition of the query. The `FILTER` function is used to return a table that filters another table or expression. `SUMMARIZE` creates a summary table for the stores, including the `Store[Name]` and `Store[OpenDate]` columns, and the filter expression `Store[OpenDate] >= _SalesSince` ensures only stores opened on or after December 1, 2023, are included in the results.

Reference =

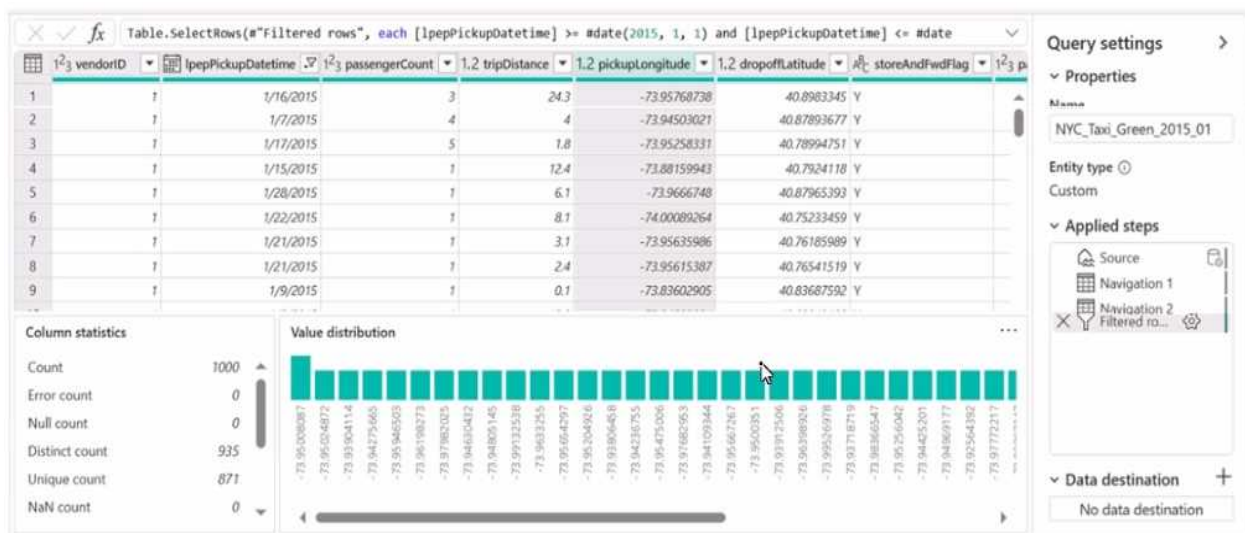
DAX FILTER Function

DAX SUMMARIZE Function

Question: 19

You have a Fabric workspace named Workspace 1 that contains a dataflow named Dataflow1.

Dataflow1 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.

Answer: B

Explanation:

The pickupLongitude column has duplicate values. This can be inferred because the 'Distinct count' is 935 while the 'Count' is 1000, indicating that there are repeated values within the column. Reference = Microsoft Power BI documentation on data profiling could provide further insights into understanding and interpreting column statistics like these.

Question: 20

You have a Fabric tenant named Tenant1 that contains a workspace named WS1. WS1 uses a capacity named C1 and contains a dataset 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. Reference = The configuration of XMLA endpoints in Power BI capacities is detailed in the Power BI documentation on dataset management.

Question: 21

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 BI 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). Reference = More information can be found in the Admin portal of the Power BI service documentation, detailing tenant and capacity settings.

Question: 22

You are creating a semantic model in Microsoft Power BI Desktop.

You plan to make bulk changes to the model by using the Tabular Model Definition Language (TMDL) extension for Microsoft Visual Studio Code.

You need to save the semantic model to a file.

Which file format should you use?

- A. PBIP
- B. PBIX
- C. PBIT
- D. PBIDS

Answer: B

Explanation:

When saving a semantic model to a file that can be edited using the Tabular Model Scripting Language (TMSL) extension for Visual Studio Code, the PBIX (Power BI Desktop) file format is the

correct choice. The PBIX format contains the report, data model, and queries, and is the primary file format for editing in Power BI Desktop. Reference = Microsoft's documentation on Power BI file formats and Visual Studio Code provides further clarification on the usage of PBIX files.

Question: 23

HOTSPOT

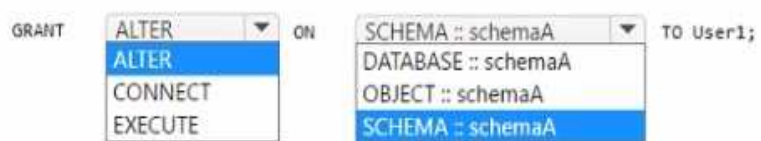
You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains three schemas named schemaA, schemaB, and schemaC

You need to ensure that a user named User1 can truncate tables in schemaA only.

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:

Explanation:

Answer Area



GRANT ALTER ON SCHEMA::schemaA TO User1;

The ALTER permission allows a user to modify the schema of an object, and granting ALTER on a schema will allow the user to perform operations like TRUNCATE TABLE on any object within that schema

a. It is the correct permission to grant to User1 for truncating tables in schemaA.

Reference =

GRANT Schema Permissions

Permissions That Can Be Granted on a Schema

Question: 24

You need to provide Power BI developers with access to the pipeline. The solution must meet the following requirements:

- Ensure that the developers can deploy items to the workspaces for Development and Test.
- Prevent the developers from deploying items to the workspace for Production.
- Follow the principle of least privilege.

Which three levels of access should you assign to the developers? Each correct answer presents part of the solution. NOTE: Each correct answer is worth one point.

- A. Build permission to the production semantic models
- B. Admin access to the deployment pipeline
- C. Viewer access to the Development and Test workspaces
- D. Viewer access to the Production workspace

- E. Contributor access to the Development and Test workspaces
- F. Contributor access to the Production workspace

Answer: B, D, E

Explanation:

To meet the requirements, developers should have Admin access to the deployment pipeline (B), Contributor access to the Development and Test workspaces (E), and Viewer access to the Production workspace (D). This setup ensures they can perform necessary actions in development and test environments without having the ability to affect production. Reference = The Power BI documentation on workspace access levels and deployment pipelines provides guidelines on assigning appropriate permissions.

Question: 25

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 BI report named Report1 that uses the model. Report1 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: A, B

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. Reference = Microsoft Power BI documentation on performance optimization offers in-depth knowledge on these features.

Question: 26

You have a Fabric tenant that contains 30 CSV files in OneLake. The files are updated daily.

You create a Microsoft Power BI semantic model named Model1 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. Reference = The Power BI documentation on incremental refresh and query folding provides detailed information on this topic.

Question: 27

You have a Fabric tenant that uses a Microsoft tower BI 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.

Answer: C

Explanation:

To enable scale-out for a semantic model, you should first set Large dataset storage format to On (C) at the semantic model level. This configuration is necessary to handle larger datasets effectively in a scaled-out environment. Reference = Guidance on configuring large dataset storage formats for scale-out is available in the Power BI documentation.

Question: 28

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. Reference = The

Power BI documentation on DirectQuery provides detailed explanations of how RLS and DAX queries function in this mode.

Question: 29

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 BI 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. Reference = Microsoft Power BI Desktop documentation outlines the functionalities available in Model view for managing semantic models.

Question: 30

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_COLUMNS_IN_SEGMENTS dynamic management view (DMV).
- D. Query the discover_hehory6Rant dynamic management view (DMV).

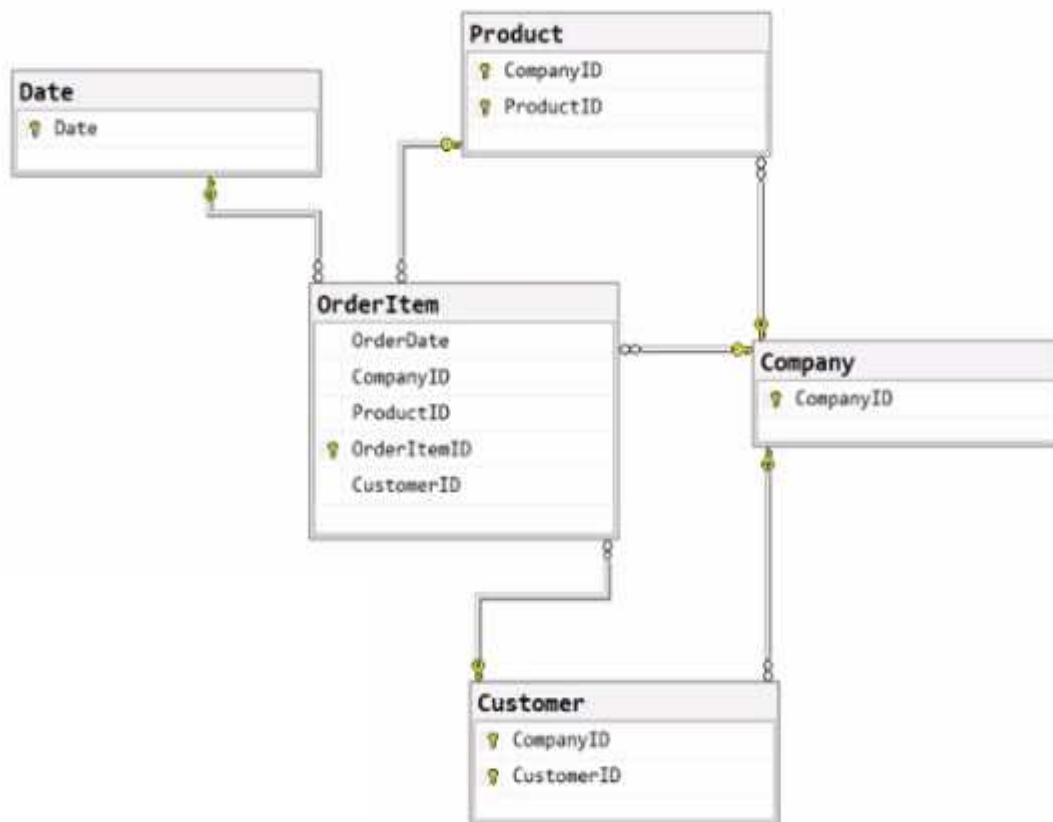
Answer: B, C

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. Reference = The Power BI documentation on Vertipaq Analyzer and DMV queries offers detailed guidance on how to use these tools for performance analysis.

Question: 31**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.

Answer Area

The relationship between OrderItem and Product must be based on:

- Both the CompanyID and the ProductID columns
- The ProductID column
- Both the CompanyID and the ProductID columns
- A new key that combines the CompanyID and ProductID columns

The Company entity must be:

- Denormalized into the Customer and Product entities
- Omitted
- Denormalized into the Product entity only
- Denormalized into the Customer and Product entities

Answer:

Explanation:

Answer Area

The relationship between OrderItem and Product must be based on:

The Company entity must be:

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.

Reference =

Dimensional modeling

Star schema design

Question: 32

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 Model1 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: A, D

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. Reference = 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.

Question: 33

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

Answer: A, B

Explanation:

Microsoft Power BI Desktop (A) and Tabular Editor (B) are the tools you can use to prevent report creators from using implicit measures. In Power BI Desktop, you can define explicit measures which can be used in visuals. Tabular Editor allows for advanced model editing, where you can enforce the use of explicit measures. Reference = Guidance on using explicit measures and preventing implicit measures in reports can be found in the Power BI and Tabular Editor official documentation.

Question: 34

HOTSPOT

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.

Answer Area

[Answer choice] of the transformation steps in the query will fold. Some

The Added custom step will be performed in [answer choice]. the Microsoft Power Query engine

Answer:

Explanation:

Answer Area

[Answer choice] of the transformation steps in the query will fold. Some

The Added custom step will be performed in [answer choice]. the Microsoft Power Query engine

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.

Reference =

Query folding in Power Query

Power Query M formula language

Question: 35

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. Reference = 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.

Question: 36

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. Reference = 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.

Question: 37

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. Reference = The process for converting and optimizing file formats within a lakehouse is discussed in the lakehouse management documentation.

Question: 38

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.

Answer: B

Explanation:

Before specifying the partition column in the Destination settings of the Copy activity, you should set Mode to Append (A). This will allow the Copy activity to add data to the table while taking the partition column into account. Reference = The configuration options for Copy activities and partitioning in Azure Data Factory, which are applicable to Fabric dataflows, are outlined in the official Azure Data Factory documentation.

Question: 39

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.

Answer Area

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

Answer:

Explanation:

A replica of dbo.Sales is created in the test schema by copying the metadata only. - No

Additional schema changes to dbo.FactSales will also apply to test.FactSales. - No

Additional data changes to dbo.FactSales will also apply to test.FactSales. - Yes

The CREATE TABLE AS CLONE statement creates a copy of an existing table, including its data and any associated data structures, like indexes. Therefore, the statement does not merely copy metadata; it also copies the data.

a. However, subsequent schema changes to the original table do not automatically propagate to the cloned table. Any data changes in the original table after the clone operation will not be reflected in the clone unless explicitly updated.

Reference =

CREATE TABLE AS SELECT (CTAS) in SQL Data Warehouse

Question: 40

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. Reference = The capabilities of a lakehouse and its support for Delta tables are described in the lakehouse and Delta table documentation.

Question: 41

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.

Answer Area

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	<input type="radio"/>	<input type="radio"/>
The resulting file partitions can be read in parallel across multiple nodes.	<input type="radio"/>	<input type="radio"/>
The resulting file partitions will use file compression.	<input type="radio"/>	<input type="radio"/>

Answer:

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.

Reference =

DataFrame write partitionBy

Apache Spark optimization with partitioning

Question: 42

You have a Fabric workspace named Workspace1 that contains a data flow named Dataflow1.

Dataflow1 contains a query that returns the data shown in the following exhibit.

	VendorID	2015-01-01	2015-01-02	2015-01-03	2015-01-04	2015-01-05	2015-01-06	2015-01-07	2015-01-08	2015-01-09
1	1	16	15	0	9	12	9	9		
2	2	20	15	7	17	20	6	6		

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

You select the VendorID column.

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

- A. Group by
- B. Unpivot columns
- C. Unpivot other columns
- D. Split column
- E. Remove other columns

Answer: B

Explanation:

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

Question: 43

You have a Fabric tenant that contains a data pipeline.

You need to ensure that the pipeline runs every four hours on Mondays and Fridays.

To what should you set Repeat for the schedule?

- A. Daily
- B. By the minute
- C. Weekly
- D. Hourly

Answer: C

Explanation:

You should set Repeat for the schedule to Weekly (C). This allows you to specify the pipeline to run on specific days of the week, in this case, every four hours on Mondays and Fridays. Reference = Scheduling options for data pipelines are available in the Azure Data Factory documentation, which includes details on configuring recurring triggers.

Question: 44

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. Reference = The use of the Monitoring hub for performance management and troubleshooting is detailed in the Azure Synapse Analytics documentation.

Question: 45

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("header", '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.

Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate','CustomerName', 'UnitPrice' columns from Sales_raw.csv.	<input type="radio"/>	<input type="radio"/>
Removing the partition will reduce the execution time of the query.	<input type="radio"/>	<input type="radio"/>
Adding inferSchema='true' to the options will increase the execution time of the query.	<input type="radio"/>	<input type="radio"/>

Answer:

Explanation:

Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate','CustomerName', 'UnitPrice' columns from Sales_raw.csv.	<input checked="" type="radio"/>	<input type="radio"/>
Removing the partition will reduce the execution time of the query.	<input type="radio"/>	<input checked="" type="radio"/>
Adding inferSchema='true' to the options will increase the execution time of the query.	<input checked="" type="radio"/>	<input type="radio"/>

The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv. - Yes

Removing the partition will reduce the execution time of the query. - No

Adding inferSchema='true' to the options will increase the execution time of the query. - Yes

The code specifies the selection of certain columns, which means only those columns will be read into the DataFrame. Partitions in Spark are a way to optimize the execution of queries by organizing the data into parts that can be processed in parallel. Removing the partition could potentially increase the execution time because Spark would no longer be able to process the data in parallel efficiently. The inferSchema option allows Spark to automatically detect the column data types, which can increase the execution time of the initial read operation because it requires Spark to read through the data to infer the schema.

Question: 46

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.dm_exec_sessions
- C. sys.dm_exec_connections
- D. sys.dm_pdw_exec_requests

Answer: D

Explanation:

The correct DMV to identify what is preventing the report query from completing is sys.dm_pdw_exec_requests (D). This DMV is specific to Microsoft Analytics Platform System (previously known as SQL Data Warehouse), which is the environment assumed to be used here. It provides information about all queries and load commands currently running or that have recently run. Reference = You can find more about DMVs in the Microsoft documentation for Analytics Platform System.

Question: 47

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.

Values	Answer Area
EnableFolding	<pre> 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 </pre>
NativeQuery	
Optimize	
Record	
StopFolding	
Table	
Value	

Answer:

Explanation:

You should complete the code as follows:

Table

NativeQuery

EnableFolding

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

Question: 48

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.

Actions	Answer Area
<input type="checkbox"/> Set the autoCompact table setting.	Remove the files: <input type="text"/>
<input type="checkbox"/> Set the optimizeWrite table setting.	Combine the files: <input type="text"/>
<input type="checkbox"/> Run the VACUUM command on a schedule.	
<input type="checkbox"/> Set the autoCompact SparkSession setting.	
<input type="checkbox"/> Run the OPTIMIZE command on a schedule.	
<input type="checkbox"/> Set the parallelDelete SparkSession setting.	

Answer:

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.

Question: 49

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.

Answer: A, D

Explanation:

For a Type 1 SCD, you should include actions that update rows when non-key attributes have changed (A), and insert new records when the natural key is a new value in the table (D). A Type 1 SCD does not track historical data, so you always overwrite the old data with the new data for a given

key. Reference = Details on Type 1 slowly changing dimension patterns can be found in data warehousing literature and Microsoft's official documentation.

Question: 50

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.

Answer Area

Connection:

Endpoint:

Answer:

Explanation:

Answer Area

Connection:

Endpoint:

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

Question: 51

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

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

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

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

```
results =
```

Which code should you run to populate the results DataFrame?

A)

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

B)

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

C)

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

D)

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

A. Option A

B. Option B

C. Option C

D. Option D

Answer: A

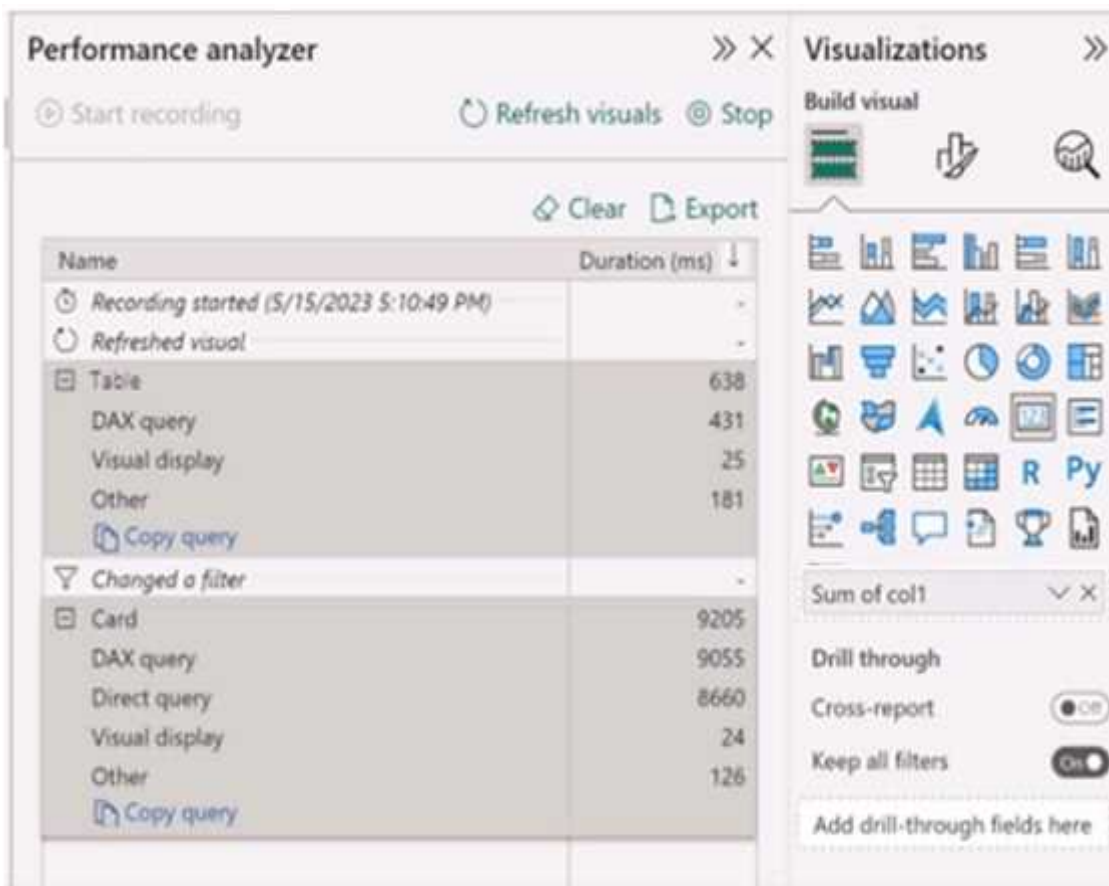
Explanation:

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. Reference = You can refer to the official Apache Spark documentation for more details on joins and the broadcast hint.

Question: 52

HOTSPOT

You have a Microsoft Power BI report and a semantic model that uses Direct Lake mode. From Power BI 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.

Answer Area

The Direct Lake fallback behavior is set to [answer choice].

the query for the table visual is executed by using [answer choice].

Answer:

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.

Question: 53

HOTSPOT

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
<code>pickupDateTime</code>	Timestamp
<code>passengerCount</code>	Integer
<code>fareAmount</code>	Double
<code>paymentType</code>	String
<code>tipAmount</code>	Double

You create a Fabric notebook and attach it to lakehouse1.

You need to use PySpark code to transform the data.

a. 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

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")
```

df2 =

df.withColumn
df.columns
df.select
df.withColumn
df.withColumnsRenamed

("pickupDate", df['tpepPickupDateTime']

.cast('date')
.alias('date')
.cast('date')
.cast('pickupDate')
.getfield('date')

.filter("fareAmount > 0 AND fareAmount < 100")
filter("fareAmount > 0 AND fareAmount < 100")
.filter(col("fareAmount").contains("1..100"))
.when(df.fareAmount > 0 AND fareAmount < 100)
.where(df.fareAmount.isin([1,100]))

Answer:

Explanation:

Answer Area

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")
```

df2 =

df.withColumn
("pickupDate", df['tpepPickupDateTime']
.cast('date')

.filter("fareAmount > 0 AND fareAmount < 100")

Add the `pickupDate` column: `.withColumn("pickupDate", df["pickupDateTime"].cast("date"))`

Filter the DataFrame: `.filter("fareAmount > 0 AND fareAmount < 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.

Question: 54

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.explain()
```

Does this meet the goal?

A. Yes

B. No

Answer: B

Explanation:

The `df.explain()` method does not meet the goal of evaluating data to calculate statistical functions. It is used to display the physical plan that Spark will execute. Reference = The correct usage of the `explain()` function can be found in the PySpark documentation.

Question: 55

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.show()
```

Does this meet the goal?

A. Yes

B. No

Answer: B

Explanation:

The `df.show()` method also does not meet the goal. It is used to show the contents of the DataFrame, not to compute statistical functions. Reference = The usage of the `show()` function is documented in the PySpark API documentation.

Question: 56

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.summary()
```

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. Reference = The PySpark API documentation details the `summary()` function and the statistics it provides.

Question: 57

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

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

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

Solution: You run the following Spark SQL statement:

```
DESCRIBE HISTORY customer
```

Does this meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

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

Question: 58

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

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

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

Solution: You run the following Spark SQL statement:

```
REFRESH TABLE customer
```

Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

No, the REFRESH TABLE statement does not provide information on whether maintenance tasks were performed. It only updates the metadata of a table to reflect any changes on the data files. Reference = The use and effects of the REFRESH TABLE command are explained in the Spark SQL documentation.

Question: 59

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

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

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

Solution: You run the following Spark SQL statement:

```
EXPLAIN TABLE customer
```

Does this meet the goal?

A. Yes

B. No

Answer: B

Explanation:

No, the EXPLAIN TABLE statement does not identify whether maintenance tasks were performed on a table. It shows the execution plan for a query. Reference = The usage and output of the EXPLAIN command can be found in the Spark SQL documentation.

Question: 60

You have a Microsoft Power BI semantic model.

You need to identify any surrogate key columns in the model that have the Summarize By property set to a value other than to None. The solution must minimize effort.

What should you use?

A. DAX Formatter in DAX Studio

B. Model view in Microsoft Power BI Desktop

C. Model explorer in Microsoft Power BI Desktop

D. Best Practice Analyzer in Tabular Editor

Answer: D

Explanation:

To identify surrogate key columns with the "Summarize By" property set to a value other than "None," the Best Practice Analyzer in Tabular Editor is the most efficient tool. The Best Practice Analyzer can analyze the entire model and provide a report on all columns that do not meet a

specified best practice, such as having the "Summarize By" property set correctly for surrogate key columns. Here's how you would proceed:

Open your Power BI model in Tabular Editor.

Go to the Advanced Scripting window.

Write or use an existing script that checks the "Summarize By" property of each column.

Execute the script to get a report on the surrogate key columns that do not have their "Summarize By" property set to "None".

You can then review and adjust the properties of the columns directly within the Tabular Editor.

Reference: The functionality of the Best Practice Analyzer in Tabular Editor is documented in the community and learning resources for Power BI.

Question: 61

DRAG DROP

You create a semantic model by using Microsoft Power BI Desktop. The model contains one security role named SalesRegionManager and the following tables:

- Sales
- SalesRegion
- Sales Address

You need to modify the model to ensure that users assigned the SalesRegionManager role cannot see a column named Address in Sales Address.

Which three 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.

Actions

⋮ Open the model in Power BI Desktop.

⋮ Set Object Level Security to **Default** for SalesRegionManager.

⋮ Set the Hidden property to **True**.

⋮ Open the model in Tabular Editor.

⋮ Select the **Address** column in SalesAddress.

⋮ Set Object Level Security to **None** for SalesRegionManager.

Answer Area

Answer:

Explanation:

Actions

⋮ Open the model in Power BI Desktop.

⋮ Set Object Level Security to **Default** for SalesRegionManager.

⋮ Set the Hidden property to **True**.

Answer Area

1 ⋮ Open the model in Tabular Editor.

2 ⋮ Select the **Address** column in SalesAddress.

3 ⋮ Set Object Level Security to **None** for SalesRegionManager.

To ensure that users assigned the SalesRegionManager role cannot see the Address column in the SalesAddress table, follow these steps in sequence:

Open the model in Tabular Editor.

Select the Address column in SalesAddress.

Set Object Level Security to None for SalesRegionManager.

Question: 62

You have a Microsoft Power BI 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.

Reference: This information is based on the DAX reference and performance optimization guidelines in the Microsoft Power BI documentation.

Question: 63

You have a Microsoft Power BI 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.

Reference: 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.

Question: 64

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 BI 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.

Reference: 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.

Question: 65

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 BI 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.

Reference: The ability to batch-edit properties using scripts in Tabular Editor is well-documented in the tool's official documentation and user community resources.

Question: 66

DRAG DROP

You have a Fabric tenant that contains a Microsoft Power BI 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.

Actions	Answer Area
<div>⋮ View the Server Timings tab.</div>	
<div>⋮ From Performance analyzer, capture a recording.</div>	
<div>⋮ Enable Query Timings and Server Timings. Run the query.</div>	
<div>⋮ View the Query Timings tab.</div>	
<div>⋮ Sort the Duration (ms) column in descending order by DAX query time.</div>	
<div>⋮ Copy the first query to DAX Studio.</div>	

Answer:

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.

Question: 67

HOTSPOT

You have a Microsoft Power BI semantic model.

You plan to implement calculation groups.

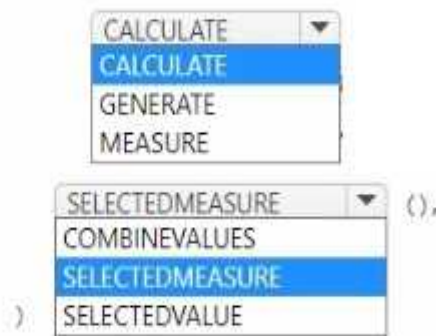
You need to create a calculation item that will change the context from the selected date to month-to-date (MTD).

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.

Answer Area



Answer:

Explanation:

Answer Area



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])
)
```

Question: 68

DRAG DROP

You are implementing two dimension tables named Customers and Products in a Fabric warehouse.

You need to use slowly changing dimension (SCD) to manage the versioning of data.

a. The solution must meet the requirements shown in the following table.

Table	Change action
Customers	Create a new version of the row.
Products	Overwrite the existing value in the latest row.

Which type of SCD should you use for each table? To answer, drag the appropriate SCD types to the correct tables. Each SCD type 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.

SCD Types	Answer Area
<div>Type 0</div> <div>Type 1</div>	Customers: <input type="text"/>
<div>Type 2</div> <div>Type 3</div>	Products: <input type="text"/>

Answer:

Explanation:

SCD Types	Answer Area
<div>Type 0</div> <div>Type 1</div>	Customers: <input type="text" value="Type 2"/>
<div>Type 2</div> <div>Type 3</div>	Products: <input type="text" value="Type 1"/>

For the Customers table, where the requirement is to create a new version of the row, you would use:

Type 2 SCD: This type allows for the creation of a new record each time a change occurs, preserving the history of changes over time.

For the Products table, where the requirement is to overwrite the existing value in the latest row, you would use:

Type 1 SCD: This type updates the record directly, without preserving historical data.

Question: 69

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.

Reference: 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.

Question: 70

HOTSPOT

You have a Fabric tenant that contains lakehouse named Lakehouse1. Lakehouse1 contains a Delta table with eight columns. You receive new data that contains the same eight columns and two additional columns.

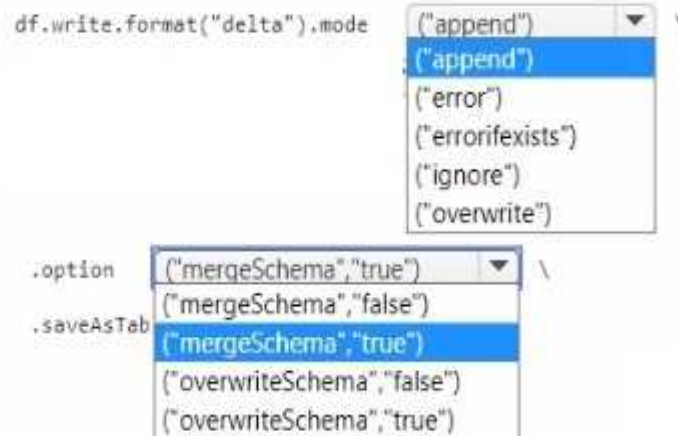
You create a Spark DataFrame and assign the DataFrame to a variable named df. The DataFrame contains the new data

a. 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.

Answer Area



Answer:

Explanation:

Answer Area

```
df.write.format("delta").mode ("append") \
.option ("mergeSchema","true") \
.saveAsTable("table")
```

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")
```

Question: 71

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.

Reference: 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.

Question: 72

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: A, C, E

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.

Reference: 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.

Question: 73

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.

Reference: 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.

Question: 74

DRAG DROP

You are implementing a medallion architecture in a single Fabric workspace.

You have a lakehouse that contains the Bronze 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.

Execution Methods

A pipeline Copy activity

A pipeline Dataflow activity

A pipeline Stored procedure activity

A schedule

A Spark job definition

An Invoke pipeline activity

Answer Area

Orchestration pipeline:

Bronze layer:

Silver layer:

Gold layer:

Answer:

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.

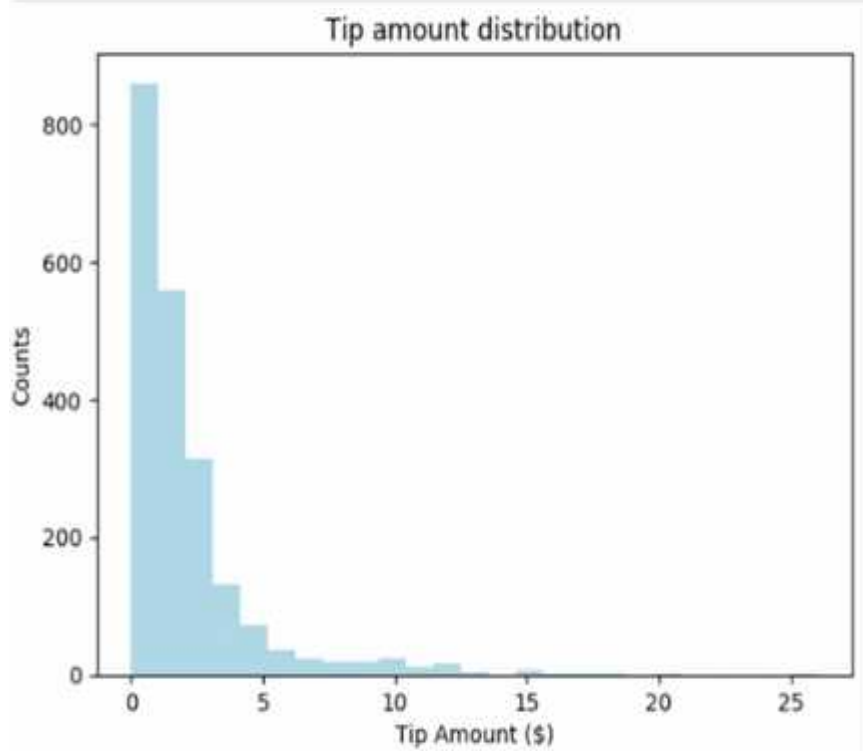
Gold layer: Implement a pipeline Stored procedure activity. Stored procedures are specified for the Gold layer according to the table.

Question: 75

You have a Fabric notebook that has the Python code and output shown in the following exhibit.

```
# Look at a histogram of tips by count by using Matplotlib
```

```
ax1 = sampled_taxi_pd_df['tipAmount'].plot(kind='hist', bins=25, facecolor='lightblue')
ax1.set_title('Tip amount distribution')
ax1.set_xlabel('Tip Amount ($)')
ax1.set_ylabel('Counts')
plt.suptitle('')
plt.show()
```



Which type of analytics are you performing?

- A. predictive
- B. descriptive
- C. prescriptive
- D. diagnostic

Answer: B

Explanation:

The Python code and output shown in the exhibit display a histogram, which is a representation of

the distribution of data. This kind of analysis is descriptive analytics, which is used to describe or summarize the features of a dataset. Descriptive analytics answers the question of "what has happened" by providing insight into past data through tools such as mean, median, mode, standard deviation, and graphical representations like histograms.

Reference: Descriptive analytics and the use of histograms as a way to visualize data distribution are basic concepts in data analysis, often covered in introductory analytics and Python programming resources.

Question: 76

You have an Azure Repos Git repository named Repo1 and a Fabric-enabled Microsoft Power BI Premium capacity. The capacity contains two workspaces named Workspace1 and Workspace2. Git integration is enabled at the workspace level.

You plan to use Microsoft Power BI Desktop and Workspace1 to make version-controlled changes to a semantic model stored in Repo1. The changes will be built and deployed to Workspace2 by using Azure Pipelines.

You need to ensure that report and semantic model definitions are saved as individual text files in a folder hierarchy. The solution must minimize development and maintenance effort.

In which file format should you save the changes?

- A. PBIP
- B. PBIT
- C. PBIX
- D. PBIDS

Answer: C

Explanation:

When working with Power BI Desktop and Git integration for version control, report and semantic model definitions should be saved in the PBIX format. PBIX is the Power BI Desktop file format that contains definitions for reports, data models, and queries, and it can be easily saved and tracked in a version-controlled environment. The solution should minimize development and maintenance effort, and saving in PBIX format allows for the easiest transition from development to deployment, especially when using Azure Pipelines for CI/CD (continuous integration/continuous deployment) practices.

Reference: The use of PBIX files with Power BI Desktop and Azure Repos for version control is discussed in Microsoft's official Power BI documentation, particularly in the sections covering Power BI Desktop files and Azure DevOps integration.

Question: 77

You have a Microsoft Fabric tenant that contains a dataflow.

You are exploring a new semantic model.

From Power Query, you need to view column information as shown in the following exhibit.



Which three Data view options should you select? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Enable column profile
- B. Show column quality details
- C. Show column profile in details pane
- D. Enable details pane
- E. Show column value distribution

Answer: A, B, E

Explanation:

To view column information like the one shown in the exhibit in Power Query, you need to select the options that enable profiling and display quality and distribution details. These are: A. Enable column profile - This option turns on profiling for each column, showing statistics such as distinct and unique values. B. Show column quality details - It displays the column quality bar on top of each column showing the percentage of valid, error, and empty values. E. Show column value distribution - It enables the histogram display of value distribution for each column, which visualizes how often each value occurs.

Reference: These features and their descriptions are typically found in the Power Query documentation, under the section for data profiling and quality features.

Question: 78

HOTSPOT

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.

Answer Area

SELECT ProductID,

GREATEST
 COALESCE
 GREATEST
 IIF
 MAX

(ListPrice, WholesalePrice, AgentPrice) AS HighestSellingPrice,

COALESCE
 CHOOSE
 COALESCE
 IIF
 MAX

(AgentPrice, WholesalePrice, ListPrice) AS TradePrice

FROM

Answer:

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.

Question: 79

You have a Fabric tenant that contains a Microsoft Power BI 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. Table.FuzzyGroup
- B. Table.Profile
- C. Table.View
- D. 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.

Reference: 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.

Question: 80

HOTSPOT

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

Name	Data type	Nullable
OrderID	Integer	No
CustomerID	Integer	No
OrderDate	Date	No
Quantity	Integer	Yes
Weight	Decimal(18, 3)	Yes
ListPrice	Decimal(18, 2)	No
SalePrice	Decimal(18, 2)	Yes

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

Name	Description
OrderID	Returns OrderID
CustomerID	Returns CustomerID
PeriodDate	Returns a date representing the first day of the month for OrderDate
DayName	Returns the name of the day for OrderDate, such as Wednesday

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

SELECT OrderID, CustomerID,

DATEFROMPARTS
DATE_BUCKET
DATEFROMPARTS
DATEPART
DATETRUNC

DATEPART, OrderDate) AS DayName

FROM Sales.Orders

weekday
day
dayofyear
weekday

Answer:

Explanation:

Answer Area

SELECT OrderID, CustomerID,

DATEFROMPARTS (month, OrderDate) AS PeriodDate,

DATENAME(weekday, OrderDate) AS DayName

FROM Sales.Orders;

For the PeriodDate that returns the first day of the month for OrderDate, you should use DATEFROMPARTS as it allows you to construct a date from its individual components (year, month, day).

For the DayName that returns the name of the day for OrderDate, you should use DATENAME with the weekday date part to get the full name of the weekday.

The complete SQL query should look like this:

```
SELECT OrderID, CustomerID,
       DATEFROMPARTS(YEAR(OrderDate), MONTH(OrderDate), 1) AS PeriodDate,
       DATENAME(weekday, OrderDate) AS DayName
FROM Sales.Orders
```

Select DATEFROMPARTS for the PeriodDate and weekday for the DayName in the answer area.

Thank you for your visit.