

Microsoft

Exam Questions DP-600

Implementing Analytics Solutions Using Microsoft Fabric



NEW QUESTION 1

- (Topic 1)

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: AE

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).

NEW QUESTION 2

- (Topic 1)

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

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

Answer: C

Explanation:

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

NEW QUESTION 3

HOTSPOT - (Topic 1)

You need to resolve the issue with the pricing group classification.

How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

• • • • •

Answer Area

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

Answer Area

CREATE

AS

SELECT

ProductCategory,

ListPrice,

CASE

END AS PricingGroup

FROM dbo.Products

[dbo].[ProductsWithPricingGroup]

VIEW

FUNCTION

PROCEDURE

TABLE

VIEW

CASE

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'

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

C:\Users\Waqas Shahid\Desktop\Mudassir\Untitled.jpg
? You should use CREATE VIEW to make the pricing group logic available for T- SQL queries.
? The CASE statement should be used to determine the pricing group based on the list price.

The T-SQL statement should create a view that classifies products into pricing groups based on the list price. The CASE statement is the correct conditional logic to assign each product to the appropriate pricing group. This view will standardize the pricing group logic across different databases and semantic models.

NEW QUESTION 4

HOTSPOT - (Topic 1)
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
```

- A. Mastered
- B. Not Mastered

Answer: A

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.

NEW QUESTION 5

- (Topic 1)

You need to implement the date dimension in the data store. The solution must meet the technical requirements.

What are two ways to achieve the goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

Answer: AB

Explanation:

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

NEW QUESTION 6

- (Topic 1)

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. References = You can learn more about dataflows and their use in Power BI environments in Microsoft's Power BI documentation.

NEW QUESTION 7

HOTSPOT - (Topic 2)

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



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

References =

? GRANT Schema Permissions

? Permissions That Can Be Granted on a Schema

NEW QUESTION 8

- (Topic 2)

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 Modell that uses the CSV files as a data source. You configure incremental refresh for Model 1 and publish the model to a Premium capacity in the Fabric tenant.

When you initiate a refresh of Model1, the refresh fails after running out of resources. What is a possible cause of the failure?

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

Answer: E

Explanation:

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

NEW QUESTION 9

- (Topic 2)

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

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

Answer: A

Explanation:

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

NEW QUESTION 10

HOTSPOT - (Topic 2)

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"/>

- A. Mastered
- B. Not Mastered

Answer: A

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. 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.

References =

? CREATE TABLE AS SELECT (CTAS) in SQL Data Warehouse

NEW QUESTION 10

- (Topic 2)

You have a Fabric tenant that contains a new semantic model in OneLake. You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression: df.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.

References = The correct usage of the explain() function can be found in the PySpark documentation.

NEW QUESTION 15

- (Topic 2)

You have a semantic model named Model 1. Model 1 contains five tables that all use Import mode. Model1 contains a dynamic row-level security (RLS) role named HR. The HR role filters employee data so that HR managers only see the data of the department to which they are assigned.

You publish Model1 to a Fabric tenant and configure RLS role membership. You share the model and related reports to users.

An HR manager reports that the data they see in a report is incomplete. What should you do to validate the data seen by the HR Manager?

A. Ask the HR manager to open the report in Microsoft Power 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.

References: The 'Test as role' feature and its use for validating RLS in Power BI is covered in the Power BI documentation available on Microsoft's official documentation.

NEW QUESTION 20

HOTSPOT - (Topic 2)

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)  
    , NTILE(1) OVER (PARTITION BY CustomerID ORDER BY LastUpdated DESC) AS NTILE  
    FROM CUSTOMERBASE  
    WHERE X = 1  
    HAVING Max(LastUpdated) = 1  
    WHERE LastUpdated = Max(LastUpdated)  
    WHERE X = 1  
)
```

A. Mastered

B. Not Mastered

Answer: A

Explanation:

? 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. References =

? Use the OVER clause to aggregate data per partition

? Use window functions

NEW QUESTION 25

- (Topic 2)

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

A. Use the Load to Tables feature.

B. Create a new shortcut in the Files section.

C. Create a new shortcut in the Tables section.

D. Use the Optimize feature.

Answer: D

Explanation:

To convert CSV files into the delta format with Z-Order optimization enabled, you should use the Optimize feature (D) from Lakehouse Explorer. This will allow you to optimize the file organization for the most efficient querying. References = The process for converting and optimizing file formats within a lakehouse is discussed in the lakehouse management documentation.

NEW QUESTION 29

HOTSPOT - (Topic 2)

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
 FR DATEPART
 DATETRUNC
 DATENAME(
 FROM Sales.Ord
 , OrderDate) AS DayName

A. Mastered

B. Not Mastered

Answer: A

Explanation:

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.

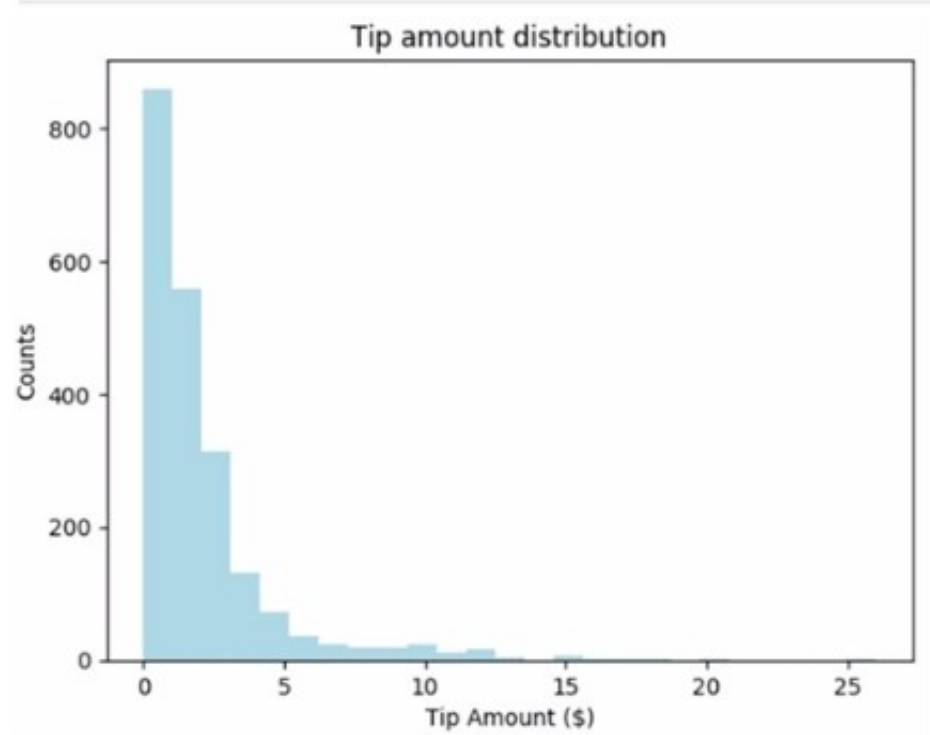
NEW QUESTION 32

- (Topic 2)

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:

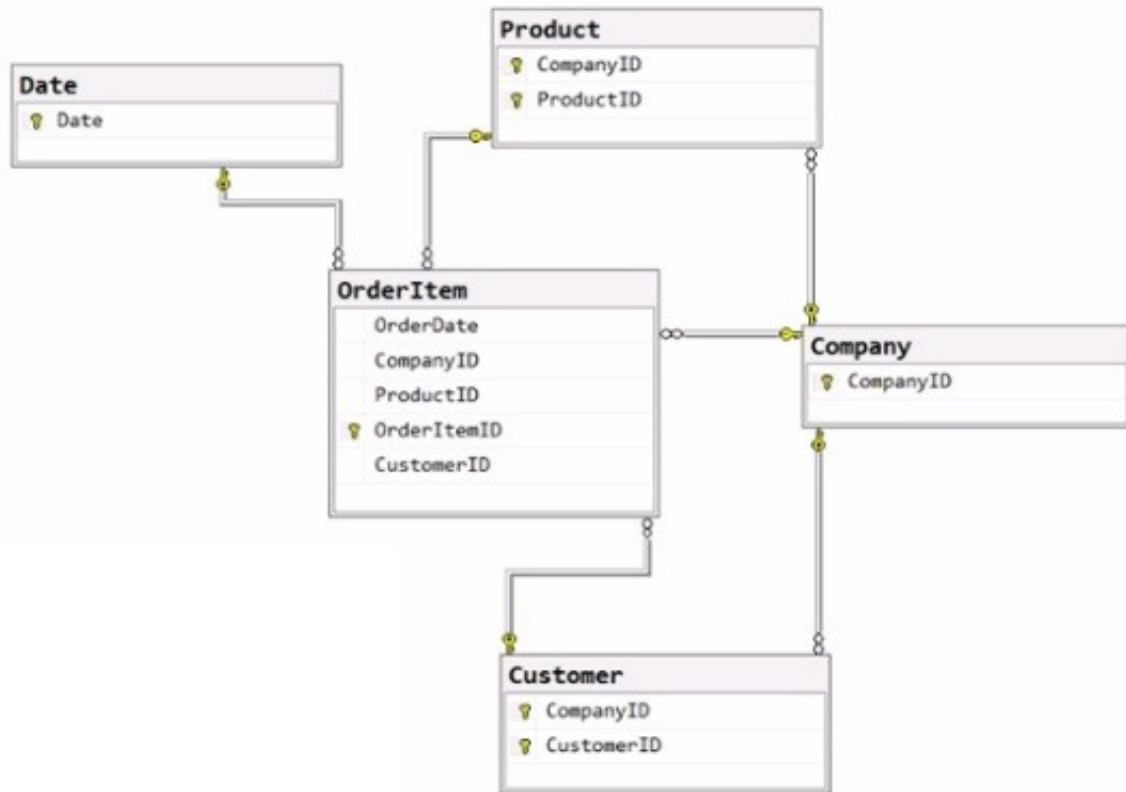
The Python code and output shown in the exhibit display a histogram, which is a representation of the distribution of data. This kind of analysis is descriptive analytics, which is used to describe or summarize the features of a dataset. Descriptive analytics answers the question of "what has happened"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

References: 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.

NEW QUESTION 35

HOTSPOT - (Topic 2)

You have the source data model shown in the following exhibit.



The primary keys of the tables are indicated by a key symbol beside the columns involved in each key. You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer. What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? The relationship between OrderItem and Product must be based on: Both the CompanyID and the ProductID columns

? The Company entity must be: Denormalized into the Customer and Product entities

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

? Dimensional modeling

? Star schema design

NEW QUESTION 36

- (Topic 2)

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. References = The configuration of XMLA endpoints in Power BI capacities is detailed in the Power BI documentation on dataset management.

NEW QUESTION 41

- (Topic 2)

You have a Fabric tenant that contains a warehouse.

Several times a day, the performance of all warehouse queries degrades. You suspect that Fabric is throttling the compute used by the warehouse.

What should you use to identify whether throttling is occurring?

- A. the Capacity settings
- B. the Monitoring hub
- C. dynamic management views (DMVs)

D. the Microsoft Fabric Capacity Metrics app

Answer: B

Explanation:

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

NEW QUESTION 45

- (Topic 2)

You have a Fabric tenant that contains a semantic model. The model contains 15 tables.

You need to programmatically change each column that ends in the word Key to meet the following requirements:

- Hide the column.
- Set Nullable to False.
- Set Summarize By to None
- Set Available in MDX to False.
- Mark the column as a key column. What should you use?

- A. Microsoft Power 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.

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

NEW QUESTION 48

- (Topic 2)

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.

References: The use of DAX Studio for query optimization is a common best practice for improving Power BI report performance as outlined in the Power BI documentation.

NEW QUESTION 52

- (Topic 2)

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: CD

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. References = 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.

NEW QUESTION 56

HOTSPOT - (Topic 2)

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.

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"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? 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.

NEW QUESTION 61

- (Topic 2)

You have a Fabric tenant that contains a new semantic model in OneLake. You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression: df.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. References = The PySpark API documentation details the summary() function and the statistics it provides.

NEW QUESTION 64

DRAG DROP - (Topic 2)

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 Ad dress

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

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

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.

NEW QUESTION 68

- (Topic 2)

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



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

You select the VendorID column.

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

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

Answer: B

Explanation:

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

NEW QUESTION 70

- (Topic 2)

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

NEW QUESTION 71

- (Topic 2)

You have a Fabric tenant that contains a warehouse.

You are designing a star schema model that will contain a customer dimension. The customer dimension table will be a Type 2 slowly changing dimension (SCD).

You need to recommend which columns to add to the table. The columns must NOT already exist in the source.

Which three types of columns should you recommend? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

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

Answer: ACE

Explanation:

For a Type 2 slowly changing dimension (SCD), you typically need to add the following types of columns that do not exist in the source system:

? An effective start date and time (E): This column records the date and time from which the data in the row is effective.

? An effective end date and time (A): This column indicates until when the data in the row was effective. It allows you to keep historical records for changes over time.

? A surrogate key (C): A surrogate key is a unique identifier for each row in a table, which is necessary for Type 2 SCDs to differentiate between historical and current records.

References: Best practices for designing slowly changing dimensions in data warehousing solutions, which include Type 2 SCDs, are commonly discussed in data warehousing and business intelligence literature and would be part of the modeling guidance in a Fabric tenant's documentation.

NEW QUESTION 76

- (Topic 2)

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.

References: 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.

NEW QUESTION 77

- (Topic 2)

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. References = Microsoft's documentation on Power BI file formats and Visual Studio Code provides further clarification on the usage of PBIX files.

NEW QUESTION 82

HOTSPOT - (Topic 2)

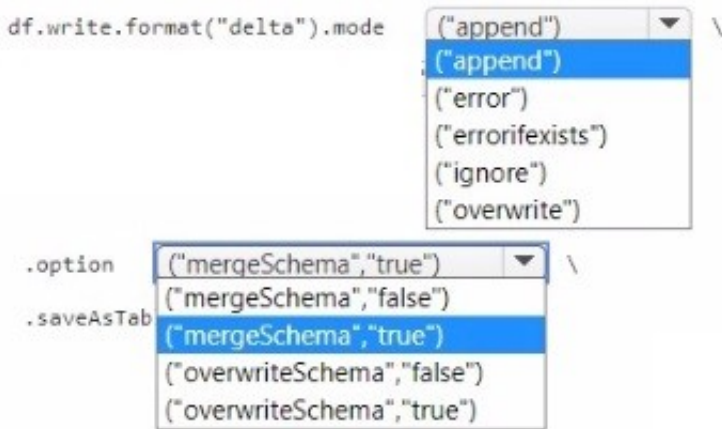
You have a Fabric tenant that contains lakehouse named Lakehouse1. Lakehouse1 contains a Delta table with eight columns. You receive new data that contains the same eight columns and two additional columns.

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



- A. Mastered
B. Not Mastered

Answer: A

Explanation:

o add new data to the Delta table while meeting the specified requirements:

? You should use the append mode to ensure that all new data is added to the table without affecting the existing rows.

? You should set the mergeSchema option to true to allow the schema of the Delta table to be updated with the new columns found in the DataFrame.

The completed code would look like this:

```
df.write.format("delta").mode("append") option("mergeSchema", "true") saveAsTable("Lakehouse1.TableName")
```

NEW QUESTION 84

DRAG DROP - (Topic 2)

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.

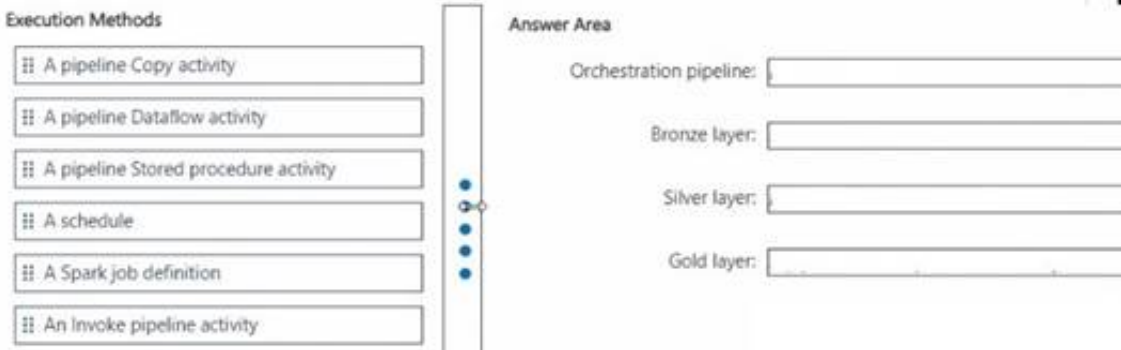
You create the items required to populate the layers as shown in the following table.

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.



- A. Mastered
B. Not Mastered

Answer: A

Explanation:

To execute each set of items in sequential order with minimized development effort and complexity, you should use the following options:

? Orchestration pipeline: Use a pipeline with an Invoke pipeline activity. This allows for orchestrating and scheduling the execution of other pipelines, ensuring they run in the correct sequence.

? Bronze layer: Implement a pipeline Copy activity. This aligns with the table indicating that the Bronze layer uses pipelines with Copy activities for data integration.

? Silver layer: Implement a pipeline Dataflow activity. The table specifies that Dataflows are used for the Silver layer.

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

NEW QUESTION 89

- (Topic 2)

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.

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

NEW QUESTION 90

- (Topic 2)

You have a Fabric tenant.

You are creating a Fabric Data Factory pipeline.

You have a stored procedure that returns the number of active customers and their average sales for the current month.

You need to add an activity that will execute the stored procedure in a warehouse. The returned values must be available to the downstream activities of the pipeline.

Which type of activity should you add?

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

Answer: C

Explanation:

In a Fabric Data Factory pipeline, to execute a stored procedure and make the returned values available for downstream activities, the Lookup activity is used. This activity can retrieve a dataset from a data store and pass it on for further processing. Here's how you would use the Lookup activity in this context:

? Add a Lookup activity to your pipeline.

? Configure the Lookup activity to use the stored procedure by providing the necessary SQL statement or stored procedure name.

? In the settings, specify that the activity should use the stored procedure mode.

? Once the stored procedure executes, the Lookup activity will capture the results and make them available in the pipeline's memory.

? Downstream activities can then reference the output of the Lookup activity. References: The functionality and use of Lookup activity within Azure Data Factory is documented in Microsoft's official documentation for Azure Data Factory, under the section for pipeline activities.

NEW QUESTION 94

- (Topic 2)

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:

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. References = The usage and output of the EXPLAIN command can be found in the Spark SQL documentation.

NEW QUESTION 98

- (Topic 2)

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. References = For more on Power BI Python visuals and how they handle data, please refer to the Power BI documentation.

NEW QUESTION 99

- (Topic 2)

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

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

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

A)

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

B)

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

C)

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

D)

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

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

Answer: B

Explanation:

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

NEW QUESTION 100

HOTSPOT - (Topic 2)

You have a Fabric tenant.

You plan to create a Fabric notebook that will use Spark DataFrames to generate Microsoft Power BI visuals.

You run the following code.

```
from powerbiclient import QuickVisualize, get_dataset_config, Report

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

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
The code embeds an existing Power BI report.	<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"/>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

? The code embeds an existing Power BI report. - No

? The code creates a Power BI report. - No

? The code displays a summary of the DataFrame. - Yes

The code provided seems to be a snippet from a SQL query or script which is neither creating nor embedding a Power BI report directly. It appears to be setting up a DataFrame for use within a larger context, potentially for visualization in Power BI, but the code itself does not perform the creation or embedding of a report.

Instead, it's likely part of a data processing step that summarizes data.

References =

? Introduction to DataFrames - Spark SQL

? Power BI and Azure Databricks

NEW QUESTION 105

- (Topic 2)

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: BDE

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. References = The Power BI documentation on workspace access levels and deployment pipelines provides guidelines on assigning appropriate permissions.

NEW QUESTION 106

HOTSPOT - (Topic 2)

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

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

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

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

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

Answer Area

SELECT ProductID,

GREATEST

COALESCE

GREATEST

IIF

MAX

(ListPrice, WholesalePrice, AgentPrice) AS HighestSellingPrice,

FROM

COALESCE

CHOOSE

COALESCE

IIF

MAX

(AgentPrice, WholesalePrice, ListPrice) AS TradePrice

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

? For the TradePrice, you should use the COALESCE function, which returns the first non-null value in a list. The COALESCE function is the correct choice as it will return AgentPrice if it's not null; if AgentPrice is null, it will check WholesalePrice, and if that is also null, it will return ListPrice.

The complete code with the correct SQL functions would look like this:

```
SELECT ProductID,
MAX(ListPrice, WholesalePrice, AgentPrice) AS HighestSellingPrice, -- MAX is used as a placeholder
COALESCE(AgentPrice, WholesalePrice, ListPrice) AS TradePrice FROM Sales.Products
Select MAX for HighestSellingPrice and COALESCE for TradePrice in the answer area.
```

NEW QUESTION 110

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DP-600 Practice Exam Features:

- * DP-600 Questions and Answers Updated Frequently
- * DP-600 Practice Questions Verified by Expert Senior Certified Staff
- * DP-600 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * DP-600 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DP-600 Practice Test Here](#)