Exam Code: DP-500
Exam Name: Designing and Implementing Enterprise-Scale Analytics Solutions Using Microsoft Azure and Microsoft Power BI

Case Study 01

Topic 1, Litware, Inc.

Overview

Litware, Inc. is a retail company that sells outdoor recreational goods and accessories. The company sells goods both online and at its stores located in six countries.

Azure Resources

Litware has the following Azure resources:

• An Azure Synapse Analytics workspace named synapseworkspace1
• An Azure Data Lake Storage Gen2 account named datalake1 that is associated with synapseworkspace1
• A Synapse Analytics dedicated SQL pool named SQLDW

Dedicated SQL Pool

SQLDW contains a dimensional model that contains the following table.

Name	Relevant column	Description
dbo.Customer	CustomerKey, CustomerID, CustomerEmail	The table currently contains 250,000 rows. Each row identifies a unique customer.
dbo.Product	ProductKey, ProductID, ProductName, ProductCategory, IsActive	The table currently contains 2,500 rows. Each row identifies a unique product.
dbo.Date	Date, Month, Year	The table currently contains 3,653 rows. Each row identifies a unique date.
dbo.SalesTransactions	CustomerKey, ProductKey, SalesDate, SalesChannelKey, SalesAmount, QuantitySold	The table currently contains 75 million rows. Each row identifies the purchase of a single product in a sales transaction.
dbo.SalesChannel	SalesChannelKey, SalesChannel	The table currently contains two rows to identify whether a sale occurred online or in a store.

SQLDW contains the following additional tables.

Name	Relevant column	Description
MLModel	Model, Model_Name	The table contains a machine learning model named PredictPurchase that predicts the likelihood of customers purchasing a specific product based on their past purchases.
CustomersWithProductScore	CustomerID, CustomerEmail, ProductID, ProductName, Score	The Score column contains the results from calling the predictive model.

SQLDW contains a view named dbo.CustomerPurchases that creates a distinct list of values from dbo.Customer [customeriD], dbo.Customer [CustomerEmail], dbo.ProductfProductID] and dbo.Product[ProductName].

The sales data in SQLDW is updated every 30 minutes. Records in dbo.SalesTransactions are updated in SQLDW up to three days after being created. The records do NOT change after three days.

Power BI

Litware has a new Power Bl tenant that contains an empty workspace named Sales Analytics.

All users have Power B1 Premium per user licenses.

IT data analysts are workspace administrators. The IT data analysts will create datasets and reports.

A single imported dataset will be created to support the company's sales analytics goals. The dataset will be refreshed every 30 minutes.

Analytics Goals

Litware identifies the following analytics goals:

• Provide historical reporting of sales by product and channel over time.
• Allow sales managers to perform ad hoc sales reporting with minimal effort.
• Perform market basket analysis to understand which products are commonly purchased in the same transaction.
• Identify which customers should receive promotional emails based on their likelihood of purchasing promoted products.

Litware plans to monitor the adoption of Power Bl reports over time. The company wants custom Power Bl usage reporting that includes the percent change of users that view reports in the Sales Analytics workspace each month.

Security Requirements

Litware identifies the following security requirements for the analytics environment:

• All the users in the sales department and the marketing department must be able to see Power B1 reports that contain market basket analysis and data about which customers are likely to purchase a product.
• Customer contact data in SQLDW and the Power B1 dataset must be labeled as Sensitive. Records must be kept of any users that use the sensitive data.
• Sales associates must be prevented from seeing the CustomerEmail column in Power B1 reports.
• Sales managers must be prevented from modifying reports created by other users.

Development Process Requirements

Litware identifies the following development process requirements:

• SQLDW and datalake1 will act as the development environment. Once feature development is complete, all entities in synapseworkspace1 will be promoted to a test workspace, and then to a production workspace.
• Power Bl content must be deployed to test and production by using deployment pipelines.

• All SQL scripts must be stored in Azure Repos.
The IT data analysts prefer to build Power BI reports in Synapse Studio.

QUESTION 1
DRAG DROP
You need to implement object-level security (OLS) in the Power BI dataset for the sales associates.
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.

Select and Place:

Actions	Answer Area
From Power BI Desktop, add a table filter to the role.	
From Power BI Desktop, create a role for the sales associates.	
From Tabular Editor, set Object Level Security to **None** for the Customer[Email] column and save the changes.	
From Power BI Desktop, publish the dataset to the Sales Analytics workspace.	
From Tabular Editor, set Object Level Security to **None** for the Customer table and save the changes.	

Correct Answer:

Actions	Answer Area
From Power BI Desktop, add a table filter to the role.	From Power BI Desktop, create a role for the sales associates.
	From Tabular Editor, set Object Level Security to **None** for the Customer[Email] column and save the changes.
	From Power BI Desktop, publish the dataset to the Sales Analytics workspace.
From Tabular Editor, set Object Level Security to **None** for the Customer table and save the changes.	

Section:
Explanation:

QUESTION 2
You need to configure the Sales Analytics workspace to meet the ad hoc reporting requirements.
What should you do?

A. Grant the sales managers the Build permission for the existing Power BI datasets.

B. Grant the sales managers admin access to the existing Power BI workspace.

C. Create a deployment pipeline and grant the sales managers access to the pipeline.

D. Create a PBIT file and distribute the file to the sales managers.

Correct Answer: D
Section:
Explanation:
Allow sales managers to perform ad hoc sales reporting with minimal effort Power BI report templates contain the following information from the report from which they were generated:
Report pages, visuals, and other visual elements
The data model definition, including the schema, relationships, measures, and other model definition items
All query definitions, such as queries, Query Parameters, and other query elements
What is not included in templates is the report's data.
Report templates use the file extension .PBIT (compare to Power BI Desktop reports, which use the .PBIX extension).
Note: With Power BI Desktop, you can create compelling reports that share insights across your entire organization. With Power BI Desktop templates, you can streamline your work by creating a report template, based on an existing template, which you or other users in your organization can use as a starting point for a new report's layout, data model, and queries. Templates in Power BI
Desktop help you jump-start and standardize report creation.
Reference: https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-templates

QUESTION 3
What should you configure in the deployment pipeline?

A. a backward deployment

B. a selective deployment

C. auto-binding

D. a data source rule

Correct Answer: D
Section:
Explanation:
Development Process Requirements
Litware identifies the following development process requirements:
SQLDW and datalake1 will act as the development environment. Once feature development is complete, all entities in synapseworkspace1 will be promoted to a test workspace, and then to a production workspace.
Power BI content must be deployed to test and production by using deployment pipelines.
Create deployment rules When working in a deployment pipeline, different stages may have different configurations. For example, each stage can have different databases or different query parameters. The development stage might query sample data from the database, while the test and production stages query the entire database.
When you deploy content between pipeline stages, configuring deployment rules enables you to allow changes to content, while keeping some settings intact. For example, if you want a dataset in a production stage to point to a production database, you can define a rule for this. The rule is defined in the production stage, under the appropriate dataset. Once the rule is defined, content deployed from test to production, will inherit the value as defined in the deployment rule, and will always apply as long as the rule is unchanged and valid.
You can configure data source rules and parameter rules.
Incorrect:
Not B: if you already have a steady production environment, you can deploy it backward (to Test or
Dev, based on your need) and set up the pipeline. The feature is not limited to any sequential orders.
Reference: https://docs.microsoft.com/en-us/power-bi/create-reports/deployment-pipelines-getstarted#step-4---create-deployment-rules

QUESTION 4
How should you configure the Power BI dataset refresh for the dbo.SalesTransactions table?

A. an incremental refresh of Product where the ModifiedDate value is during the last three days.

B. an incremental refresh of dbo.SalesTransactions where the SalesDate value is during the last three days.

C. a full refresh of all the tables

D. an incremental refresh of dbo.SalesTransactions where the SalesDate value is during the last hour.

Correct Answer: B
Section:
Explanation:
The sales data in SQLDW is updated every 30 minutes. Records in dbo.SalesTransactions are updated in SQLDW up to three days after being created. The records do NOT change after three days.

QUESTION 5
DRAG DROP
You need to create the customized Power BI usage reporting. The Usage Metrics Report dataset has already been created. The solution must minimize development and administrative effort.
Which four 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.

Select and Place:

Correct Answer:

Actions

From powerbi.com, create a new report from the Usage Metrics Report dataset in the Sales Analytics workspace.
Request access to the Power BI audit logs.

Answer Area

From Power BI Desktop, open the **Usage Metrics Report** dataset in the Sales Analytics workspace.
Add a report measure.
Add visuals to the report.
Publish the report to the Sales Analytics workspace.

Section:
Explanation:
Step 1: From powerbi.com, create a new report..
The company wants custom Power BI usage reporting that includes the percent change of users that view reports in the Sales Analytics workspace each month.
Step 2: Add a report measure
Measures are used in some of the most common data analyses. Simple summarizations such as sums, averages, minimum, maximum and counts can be set through the Fields well. The calculated results of measures are always changing in response to your interaction with your reports, allowing for fast and dynamic ad-hoc data exploration.
Step 3: Add visuals to the report
Step 4: Publish the report to the Sales Analytics workspace
Reference: https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-measures

QUESTION 6
HOTSPOT
You need to create a measure to count orders for the market basket analysis.
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.

Hot Area:

Answer Area

BasketsWithProducts =
VAR CountBaskets =
     SUMX (SUMMARIZE (Sales, [▼]), 1)

Customer[CustomerKey]
SalesTransactions[CustomerKey]
SalesTransactions[Order Number]
SalesTransactions[ProductKey]

VAR BasketsWithAndProducts =

[▼]
CALCULATE(
CALCULATABLE(
INTERSECT(
KEEPFILTERS(

    SUMMARIZE (SalesTransactions, SalesTransactions[OrderNumber]),
    REMOVEFILTERS ('Product'),
    REMOVEFILTERS (SalesTransactions[ProductKey]),
    USERELATIONSHIP (SalesTransactionsp[ProductKey], 'And Product' [And ProductKey])
)
VAR BasketsWithBothProducts =
    CALCULATE ([CountBaskets], KEEPFILTERS (BasketsWithAndProducts))
VAR RemoveEmpty =
    IF (
     ISEMPTY (

[▼]
CALCULATETABLE(
CROSSJOIN(
INTERSECT(
KEEPFILTERS(

     DISTINCT ('Product' [ProductKey]),
     DISTINCT ('And Product' [And ProductKey])
)
    ),
    [BasketsWithBothProducts]
)
  RETURN
   RemoveEmpty

Answer Area:

Answer Area

BasketsWithProducts =
VAR CountBaskets =
     SUMX (SUMMARIZE (Sales, [▼]), 1)

Customer[CustomerKey]
SalesTransactions[CustomerKey]
SalesTransactions[Order Number]
SalesTransactions[ProductKey]

VAR BasketsWithAndProducts =

[▼]
CALCULATE(
CALCULATABLE(
INTERSECT(
KEEPFILTERS(

     SUMMARIZE (SalesTransactions, SalesTransactions[OrderNumber]),
     REMOVEFILTERS ('Product'),
     REMOVEFILTERS (SalesTransactions[ProductKey]),
     USERELATIONSHIP (SalesTransactionsp[ProductKey], 'And Product' [And ProductKey])
)
VAR BasketsWithBothProducts =
     CALCULATE ([CountBaskets], KEEPFILTERS (BasketsWithAndProducts))
VAR RemoveEmpty =
     IF (
      ISEMPTY (

[▼]
CALCULATETABLE(
CROSSJOIN(
INTERSECT(
KEEPFILTERS(

       DISTINCT ('Product' [ProductKey]),
       DISTINCT ('And Product' [And ProductKey])
)
    ),
    [BasketsWithBothProducts]
)
  RETURN
   RemoveEmpty

Section:
Explanation:

Exam A

QUESTION 1

You have a 2-GB Power BI dataset.
You need to ensure that you can redeploy the dataset by using Tabular Editor. The solution must minimize how long it will take to apply changes to the dataset from powerbi.com.
Which two actions should you perform in powerbi.com? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point

A. Enable service principal authentication for read-only admin APIs.

B. Turn on Large dataset storage format.

C. Connect the target workspace to an Azure Data Lake Storage Gen2 account.

D. Enable XMLA read-write.

Correct Answer: B, D
Section:
Explanation:
Optimize datasets for write operations by enabling large models
When using the XMLA endpoint for dataset management with write operations, it's recommended you enable the dataset for large models. This reduces the overhead of write operations, which can make them considerably faster. For datasets over 1 GB in size (after compression), the difference can be significant.
Tabular Editor supports Azure Analysis Services and Power BI Premium Datasets through XMLA read/write.
Note: Tabular Editor - An open-source tool for creating, maintaining, and managing tabular models using an intuitive, lightweight editor. A hierarchical view shows all objects in your tabular model.
Objects are organized by display folders with support for multi-select property editing and DAX syntax highlighting. XMLA read-only is required for query operations. Read-write is required for metadata operations.
Reference: https://docs.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools
https://tabulareditor.github.io/

QUESTION 2
You have five Power BI reports that contain R script data sources and R visuals.
You need to publish the reports to the Power BI service and configure a daily refresh of datasets.
What should you include in the solution?

A. a Power BI Embedded capacity

B. an on-premises data gateway (standard mode)

C. a workspace that connects to an Azure Data Lake Storage Gen2 account

D. an on-premises data gateway (personal mode)

Correct Answer: B
Section:
Explanation:

QUESTION 3
You have new security and governance protocols for Power BI reports and datasets. The new protocols must meet the following requirements.
• New reports can be embedded only in locations that require authentication.
• Live connections are permitted only for workspaces that use Premium capacity datasets.
Which three actions should you recommend performing in the Power BI Admin portal? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. From Tenant settings, disable Allow XMLA endpoints and Analyze in Excel with on-premises datasets.

B. From the Premium per user settings, set XMLA Endpoint to Off.

C. From Embed Codes, delete all the codes.

D. From Capacity settings, set XMLA Endpoint to Read Write.

E. From Tenant settings, set Publish to web to Disable.

Correct Answer: A, D, E
Section:
Explanation:
Reference: https://docs.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools
https://powerbi.microsoft.com/en-us/blog/power-bi-february-service-update

QUESTION 4
You have an Azure Synapse Analytics dataset that contains data about jet engine performance. You need to score the dataset to identify the likelihood of an engine failure. Which function should you use in the query?

A. PIVOT
B. GROUPING
C. PREDICT
D. CAST

Correct Answer: A
Section:

QUESTION 5
You are optimizing a dataflow in a Power BI Premium capacity. The dataflow performs multiple joins.
You need to reduce the load time of the dataflow.
Which two actions should you perform? Each correct answer presents part of the solution. NOTE:
Each correct selection is worth one point.

A. Reduce the memory assigned to the dataflows.
B. Execute non-foldable operations before foldable operations.
C. Execute foldable operations before non-foldable operations.
D. Place the ingestion operations and transformation operations in a single dataflow.
E. Place the ingestion operations and transformation operations in separate dataflows.

Correct Answer: C, E
Section:
Explanation:
Using the compute engine to improve performance
Take the following steps to enable workloads trigger the compute engine, and always improve performance:
For computed and linked entities in the same workspace:
Ensure you perform the operations that fold, such as merges, joins, conversion, and others.
For ingestion focus on getting the data into the storage as fast as possible, using filters only if they reduce the overall dataset size. It's best practice to keep your transformation logic separate from this step, and allow the engine to focus on the initial gathering of ingredients. Next, separate your transformation and business logic into a separate dataflow in the same workspace, using linked or computed entities; doing so allows for the engine to activate and accelerate your computations. In our analogy, it's like food preparation in the kitchen: food preparation is typically a separate and distinct step from gathering your raw ingredients, and a pre-requisite for putting the food in the oven. Similarly, your logic needs to be prepared separately before it can take advantage of the compute engine.
Reference: https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflowspremium-workload-configuration

QUESTION 6
Note: This question is part of a scries of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the Power BI data model shown in the exhibit. (Click the Exhibit tab.)
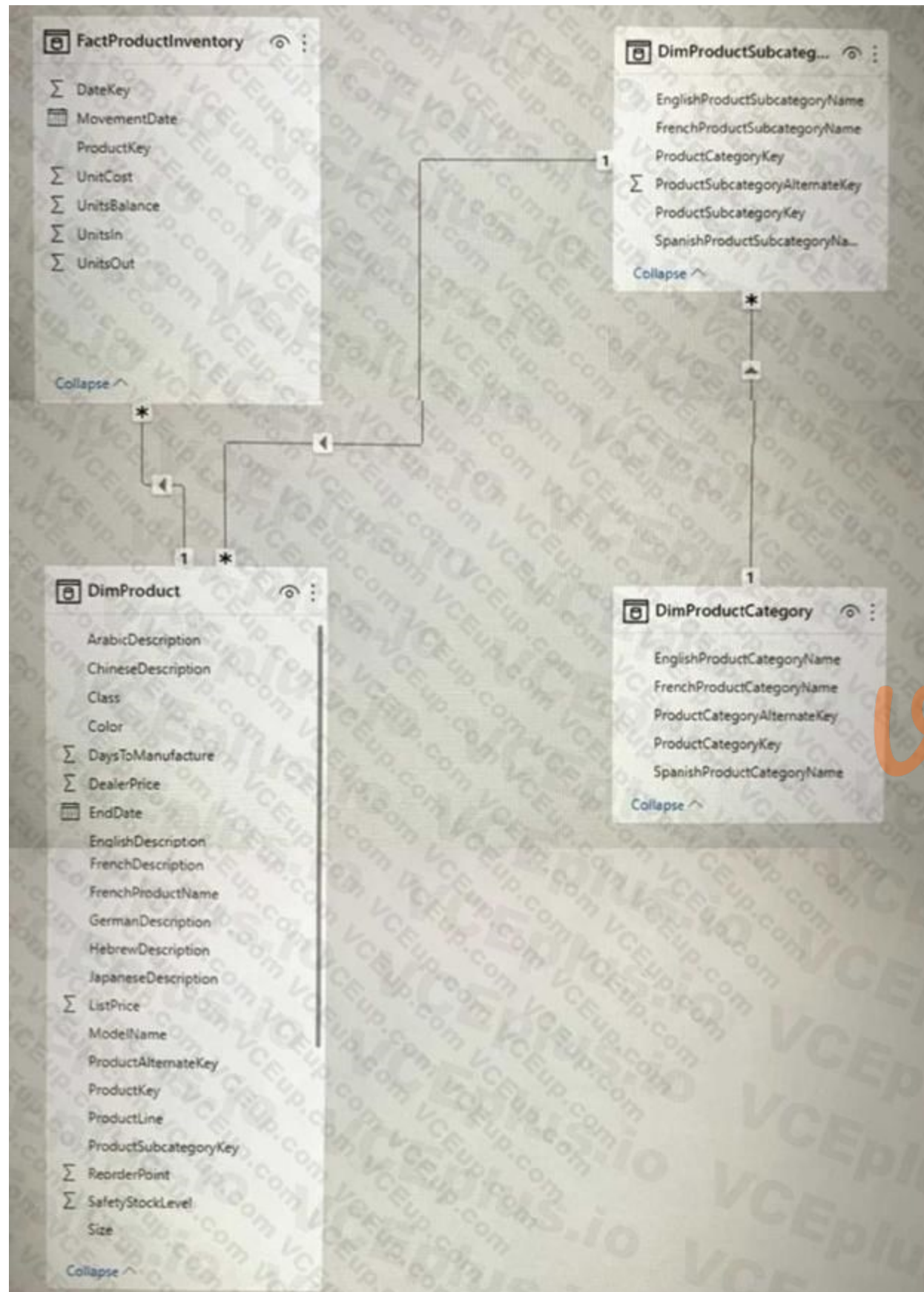
Users indicate that when they build reports from the data model, the reports take a long time to load.

You need to recommend a solution to reduce the load times of the reports.

Solution: You recommend creating a perspective that contains the commonly used fields.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B
Section:
Explanation:
Instead denormalize For Performance.
Even though it might mean storing a bit of redundant data, schema denormalization can sometimes provide better query performance. The only question then becomes is the extra space used worth the performance benefit.
Reference: https://www.mssqltips.com/sqlservertutorial/3211/denormalize-for-performance/

QUESTION 7
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Power Bl dataset named Dataset1.
In Dataset1, you currently have 50 measures that use the same time intelligence logic.
You need to reduce the number of measures, while maintaining the current functionality.
Solution: From Power Bl Desktop, you create a hierarchy.
Does this meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
Instead use the solution: From DAX Studio, you write a query that uses grouping sets.
A grouping is a set of discrete values that are used to group measure fields.
Note: A hierarchy is an ordered set of values that are linked to the level above. An example of a hierarchy could be Country, State, and City. Cities are in a State, and States make up a Country. In Power BI visuals can handle hierarchy data and provide controls for the user to navigate up and down the hierarchy.
Reference: https://docs.microsoft.com/en-us/power-bi/developer/visuals/capabilities
https://powerbi.tips/2018/09/how-to-navigate-hierarchies/

QUESTION 8
DRAG DROP
You have a Power Bl dataset that contains the following measures:
• Budget
• Actuals
• Forecast
You create a report that contains 10 visuals.
You need provide users with the ability to use a slicer to switch between the measures in two visuals only.
You create a dedicated measure named cg Measure switch.
How should you complete the DAX expression for the Actuals measure? To answer, drag the appropriate values to the targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Select and Place:

Values

| ISSELECTEDMEASURE() |
| SELECTEDMEASURENAME() |
| SELECTEDMEASURE() |
| SELECTEDVALUE() |

Answer Area

IF([Value] = "CG Measure Switch", [Actuals], [Value])

Correct Answer:

Values

| ISSELECTEDMEASURE() |
| |
| SELECTEDMEASURE() |
| |

Answer Area

IF(SELECTEDMEASURENAME() = "CG Measure Switch", [Actuals], SELECTEDVALUE())
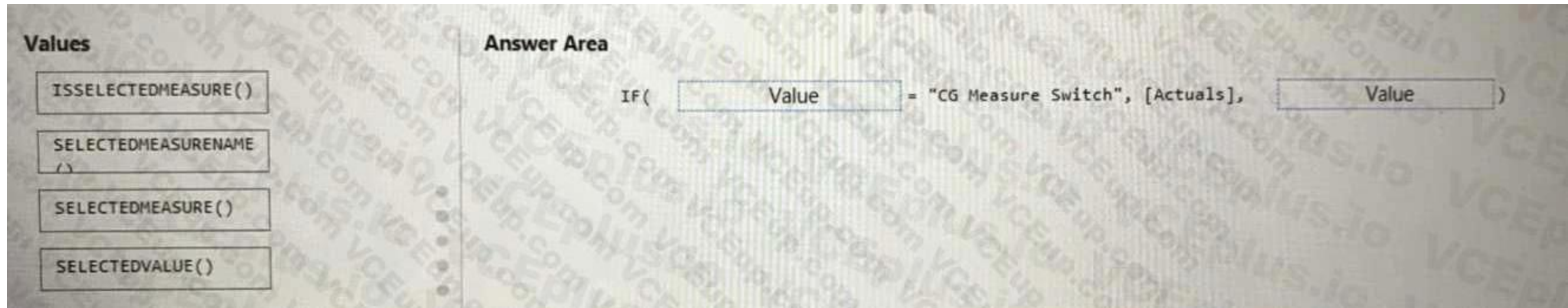
Section:
Explanation:
Box 1: SELECTEDMEASURENAME()
SELECTEDMEASURENAME is used by expressions for calculation items to determine the measure that is in context by name.
Syntax: SELECTEDMEASURENAME()
No parameters.
Example:
The following calculation item expression checks if the current measure is Expense Ratio and conditionally applies calculation logic. Since the check is based on a string comparison, it is not subject to formula fixup and will not benefit from object renaming being automatically reflected. For a similar comparison that would benefit from formula fixup, please see the ISSLECTEDMEASURE function instead.
IF (
SELECTEDMEASURENAME = "Expense Ratio",
SELECTEDMEASURE (),
DIVIDE (SELECTEDMEASURE (), COUNTROWS (DimDate))
)
Box 2: SELECTEDVALUE()
SELECTEDVALUE returns the value when the context for columnName has been filtered down to one distinct value only. Otherwise returns alternateResult.
Syntax:
SELECTEDVALUE(<columnName>[, <alternateResult>])
M1, M2, ... - A list of measures.
Reference: https://docs.microsoft.com/en-us/dax/selectedmeasurename-function-dax
https://docs.microsoft.com/en-us/dax/selectedvalue-function

QUESTION 9
DRAG DROP
You have a Power BI dataset that contains two tables named Table1 and Table2. The dataset is used by one report.
You need to prevent project managers from accessing the data in two columns in Table1 named Budget and Forecast.
Which four 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.

Select and Place:

Actions		Answer Area
For Table1, set the permissions for the Project Manager role to **None**.		
From Power BI Desktop, create a role named Project Managers.		
For Table1, set the permissions for the Project Manager role to **Read**.	>	
Open **DAX Studio**.	<	
From Power BI Desktop, add a DAX filter to the Project Managers role.		
For the Budget and Forecast columns, set the permissions to **None**.		
Open **Tabular Editor**.		

Correct Answer:

Actions		Answer Area
For Table1, set the permissions for the Project Manager role to **None**.		From Power BI Desktop, create a role named Project Managers.
		Open **Tabular Editor**.
For Table1, set the permissions for the Project Manager role to **Read**.	>	From Power BI Desktop, add a DAX filter to the Project Managers role.
Open **DAX Studio**.	<	For the Budget and Forecast columns, set the permissions to **None**.

Section:
Explanation:
Step 1: From Power BI Desktop, create a role named Project Managers.
Create roles

You can define roles within Power BI Desktop.
Step 2: Open Tabular Editor
Under Tables, select the table to which you want to apply a DAX rule.
In the Table filter DAX expression box, enter the DAX expressions. This expression returns a value of true or false. For example: [Entity ID] = "Value".
Step 3: From Power BI Desktop, add a DAX filter to the Project Managers role.
Step 4: For Table1, the Budget and Forecast columns, set the permissions to None.
Reference: https://docs.microsoft.com/en-us/power-bi/guidance/rls-guidance

QUESTION 10
You have an Azure Synapse Analytics serverless SQL pool.
You need to catalog the serverless SQL pool by using Azure Purview.
Which three actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Create a managed identity in Azure Active Directory (Azure AD).
B. Assign the Storage Blob Data Reader role to the Azure Purview managed service identity (MSI) for the storage account associated to the Synapse Analytics workspace.
C. Assign the Owner role to the Azure Purview managed service identity (MSI) for the Azure Purview resource group.
D. Register a data source.
E. Assign the Reader role to the Azure Purview managed service identity (MSI) for the Synapse Analytics workspace.

Correct Answer: A, B, E
Section:
Explanation:
Authentication for enumerating serverless SQL database resources
There are three places you'll need to set authentication to allow Microsoft Purview to enumerate your serverless SQL database resources:
The Azure Synapse workspace
The associated storage
The Azure Synapse serverless databases
The steps below will set permissions for all three.
Azure Synapse workspace
In the Azure portal, go to the Azure Synapse workspace resource.
On the left pane, select Access Control (IAM).
Select the Add button.
Set the Reader role and enter your Microsoft Purview account name, which represents its managed service identity (MSI).
Select Save to finish assigning the role
Azure Synapse Analytics serverless SQL pool catalog Purview Azure Purview managed service identity
Storage account
In the Azure portal, go to the Resource group or Subscription that the storage account associated with the Azure Synapse workspace is in.
On the left pane, select Access Control (IAM).
Select the Add button.
Set the Storage blob data reader role and enter your Microsoft Purview account name (which represents its MSI) in the Select box.
Select Save to finish assigning the role.
Azure Synapse serverless database
Go to your Azure Synapse workspace and open the Synapse Studio.
Select the Data tab on the left menu.
Select the ellipsis (...) next to one of your databases, and then start a new SQL script.
Add the Microsoft Purview account MSI (represented by the account name) on the serverless SQL databases. You do so by running the following command in your SQL script:
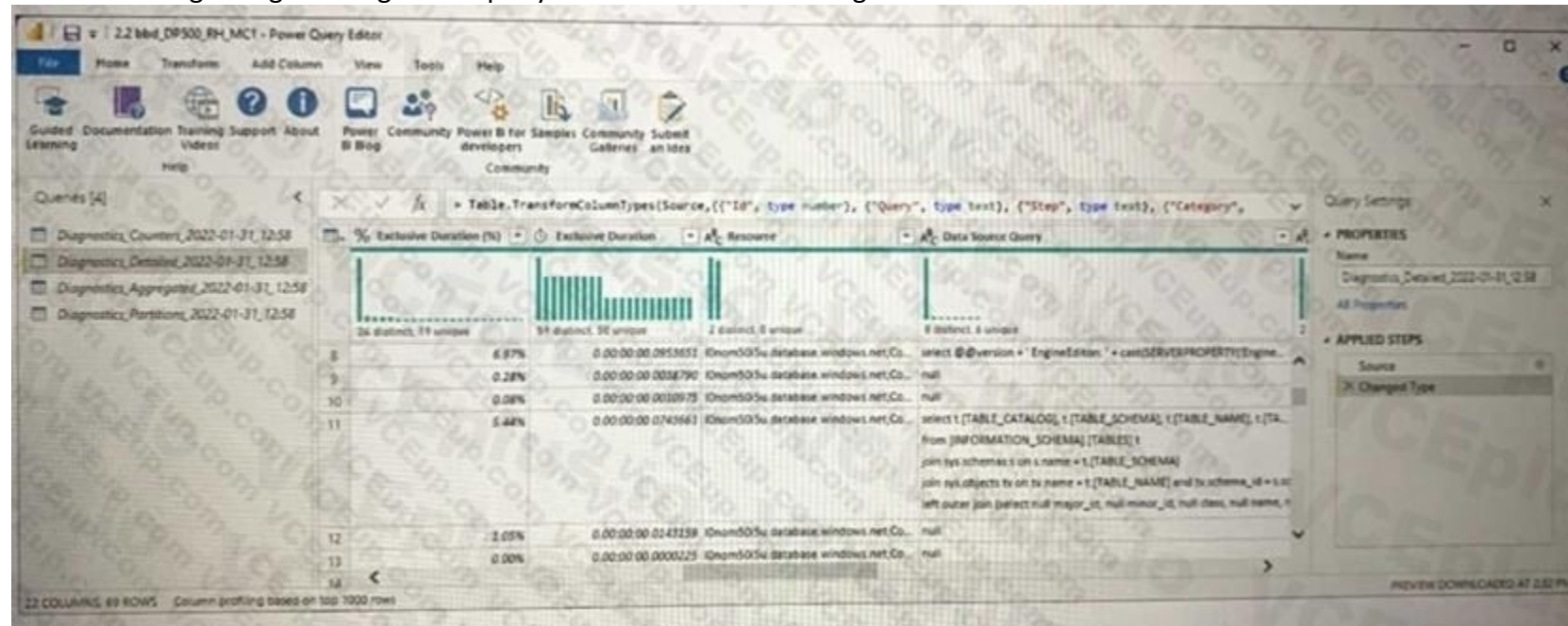SQL
CREATE LOGIN [PurviewAccountName] FROM EXTERNAL PROVIDER;

Apply permissions to scan the contents of the workspace
You can set up authentication for an Azure Synapse source in either of two ways. Select your scenario below for steps to apply permissions.

QUESTION 11
You are running a diagnostic against a query as shown in the following exhibit.

What can you identify from the diagnostics query?

A. All the query steps are folding.

B. Elevated permissions are being used to query records.

C. The query is timing out.

D. Some query steps are folding.

Correct Answer: A
Section:
Explanation:
Understanding folding with Query Diagnostics
One of the most common reasons to use Query Diagnostics is to have a better understanding of what operations were 'pushed down' by Power Query to be performed by the back-end data source, which is also known as 'folding'. If we want to see what folded, we can look at what is the 'most specific' query, or queries, that get sent to the back-end data source. We can look at this for both ODATA and SQL.
Reference: https://docs.microsoft.com/en-us/power-query/querydiagnosticsfolding

QUESTION 12
You are creating an external table by using an Apache Spark pool in Azure Synapse Analytics. The table will contain more than 20 million rows partitioned by date. The table will be shared with the SQL engines.
You need to minimize how long it takes for a serverless SQL pool to execute a query data against the table.
In which file format should you recommend storing the table data?

A. JSON

B. Apache Parquet

C. CSV

D. Delta

Correct Answer: B

Explanation:
Prepare files for querying
If possible, you can prepare files for better performance:
* Convert large CSV and JSON files to Parquet. Parquet is a columnar format. Because it's compressed, its file sizes are smaller than CSV or JSON files that contain the same data. Serverless
SQL pool skips the columns and rows that aren't needed in a query if you're reading Parquet files.
Serverless SQL pool needs less time and fewer storage requests to read it.
Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-serverlesssql-pool
https://stackoverflow.com/questions/65320949/parquet-vs-delta-format-in-azure-data-lake-gen-2store

QUESTION 13
You are configuring a Power BI report for accessibility as shown in the following table.

You need to change the default colors of all three visuals to make the report more accessible to users who have color vision deficiency. Which two settings should you configure in the Customize theme window? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Theme colors

B. Sentiment colors

C. Divergent colors

D. First-level elements colors

Correct Answer: A, B
Explanation:
Reference: https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-report-themes

QUESTION 14
You are creating a Python visual in Power BI Desktop.
You need to retrieve the value of a column named Unit Price from a DataFrame.
How should you reference the Unit Price column in the Python code?

A. pandas.DataFrame('Unit Price')

B. dataset['Unit Price']

C. data = [Unit Price]

D. ('Unit Price')

Correct Answer: A
Section:
Explanation:
You can retrieve a column in a pandas DataFrame object by using the DataFrame object name, followed by the label of the column name in brackets.

So if the DataFrame object name is dataframe1 and the column we are trying to retrieve the 'X' column, then we retrieve the column using the statement, dataframe1['X'].

Here's a simple Python script that imports pandas and uses a data frame:

import pandas as pd data = [['Alex',10],['Bob',12],['Clarke',13]] df = pd.DataFrame(data,columns=['Name','Age'],dtype=float) print (df)

When run, this script returns:

Name Age

0 Alex 10.0

1 Bob 12.0

2 Clarke 13.0

Reference: http://www.learningaboutelectronics.com/Articles/How-to-retrieve-a-column-from-a-

QUESTION 15
You have a Power BI report that contains one visual.

You need to provide users with the ability to change the visual type without affecting the view for other users.

What should you do?

A. From Report setting, select Personalize visuals.

B. From Tabular Editor, create a new perspective.

C. From the Bookmarks pane, select Focus mode, and then select Add.

D. From Visual options in Report settings, select Use the modern visual header with updated styling options.

Correct Answer: A
Section:
Explanation:
Enable personalization in a report

You can enable the feature either in Power BI Desktop or the Power BI service. You can also enable it in embedded reports.

To enable the feature in Power BI Desktop, go to File > Options and settings > Options > Current file > Report settings. Make sure Personalize visuals is turned on.

QUESTION 16
You have a Power BI report that contains the visual shown in the following exhibit.

Product	Sales
Amarilla	17,747,116.06
Carretera	13,815,307.89
Montana	15,390,801.88
Paseo	33,011,143.95
Velo	18,250,059.47
VTT	20,511,921.02
Total	118,726,350.26

You need to make the visual more accessible to users who have color vision deficiency. What should you do?

A. Change the font color of values in the Sales column to white.

B. Change the red background color to orange.

C. Add icons to represent the sales status of each product.

D. Add additional measures to the table values.

Correct Answer: B
Section:
Explanation:
Themes, contrast and colorblind-friendly colors
You should ensure that your reports have enough contrast between text and any background colors.
Certain color combinations are particularly difficult for users with color vision deficiencies to distinguish. These include the following combinations:
**---> green and black
green and red
green and brown
blue and purple
green and blue
light green and yellow
blue and grey
green and grey
Avoid using these colors together in a chart, or on the same report page.
Reference: https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-accessibility-creating-reports

QUESTION 17
You have the following Python code in an Apache Spark notebook.

```
import matplotlib.pyplot as plt
import numpy as np
ys = 300 + np.random.randn(100)
x = [x for x in range(len(ys))]
plt.plot(x, ys, '-')
plt.fill_between(x, ys, 395, where=(ys > 395), facecolor='g', alpha=0.5)
plt.title("Chart Sample")
plt.show()
```

Which type of chart will the code produce?

A. a stacked bar chart

B. a pie chart

C. a bar chart

D. an area chart

Correct Answer: D
Section:
Explanation:
The matplotlib.pyplot.fill_between function fills the area between two horizontal curves.
The curves are defined by the points (x, y1) and (x, y2). This creates one or multiple polygons describing the filled area.
Reference: https://matplotlib.org/3.5.0/api/_as_gen/matplotlib.pyplot.fill_between.html

QUESTION 18
You use Azure Synapse Analytics and Apache Spark notebooks to You need to use PySpark to gain access to the visual libraries. Which Python libraries should you use?

A. Seaborn only

B. Matplotlib and Seaborn

C. Matplotlib only

D. Matplotlib and TensorFlow

E. TensorFlow only

F. Seaborn and TensorFlow

Correct Answer: B
Section:
Explanation:
Matplotlib
You can render standard plotting libraries, like Matplotlib, using the built-in rendering functions for each library.
Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.
Additional libraries
Beyond these libraries, the Azure Synapse Analytics Runtime also includes the following set of libraries that are often used for data visualization:
Seaborn Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-datavisualization
https://seaborn.pydata.org/

QUESTION 19
You are using a Python notebook in an Apache Spark pool in Azure Synapse Analytics. You need to present the data distribution statistics from a DataFrame in a tabular view. Which method should you invoke on the DataFrame?

A. freqItems

B. explain

C. rollup

D. describe

Correct Answer: D
Section:
Explanation:
The aggregating statistic can be calculated for multiple columns at the same time with the describe function.
Example:

titanic[["Age", "Fare"]].describe()
Out[6]:
Age Fare
count 714.000000 891.000000
mean 29.699118 32.204208
std 14.526497 49.693429
min 0.420000 0.000000
25% 20.125000 7.910400
50% 28.000000 14.454200
75% 38.000000 31.000000
max 80.000000 512.329200
Reference:
https://pandas.pydata.org/docs/getting_started/intro_tutorials/06_calculate_statistics.html

QUESTION 20
You have a kiosk that displays a Power BI report page. The report uses a dataset that uses Import storage mode. You need to ensure that the report page updates all the visuals every 30 minutes.
Which two actions should you perform? Each correct answer presents part of the solution. NOTE:
Each correct selection is worth one point.

A. Enable Power BI embedded.
B. Configure the data sources to use DirectQuery.
C. Configure the data sources to use a streaming dataset
D. Select Auto page refresh.
E. Enable the XMIA endpoint.
F. Add a Microsoft Power Automate visual to the report page.

Correct Answer: B, D
Section:
Explanation:
Automatic page refresh in Power BI enables your active report page to query for new data, at a predefined cadence, for DirectQuery sources.
Automatic page refresh is available for DirectQuery sources and some LiveConnect scenarios, so it will only be available when you are connected to a supported data source. This restriction applies to both automatic page refresh types.
Reference: https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-automatic-pagerefresh

QUESTION 21
You have an Azure Synapse Analytics dedicated SQL pool.
You need to ensure that the SQL pool is scanned by Azure Purview.
What should you do first?

A. Register a data source.
B. Search the data catalog.
C. Create a data share connection.
D. Create a data policy.

Correct Answer: B
Section:

QUESTION 22

You have a Power BI workspace that contains one dataset and four reports that connect to the dataset. The dataset uses Import storage mode and contains the following data sources:
• A CSV file in an Azure Storage account
• An Azure Database for PostgreSQL database
You plan to use deployment pipelines to promote the content from development to test to production. There will be different data source locations for each stage. What should you include in the deployment pipeline to ensure that the appropriate data source locations are used during each stage?

A. parameter rules

B. selective deployment

C. auto-binding across pipelines

D. data source rules

Correct Answer: A
Section:
Explanation:
Note: Create deployment rules
When working in a deployment pipeline, different stages may have different configurations. For example, each stage can have different databases or different query parameters. The development stage might query sample data from the database, while the test and production stages query the entire database.
When you deploy content between pipeline stages, configuring deployment rules enables you to allow changes to content, while keeping some settings intact. For example, if you want a dataset in a production stage to point to a production database, you can define a rule for this. The rule is defined in the production stage, under the appropriate dataset. Once the rule is defined, content deployed from test to production, will inherit the value as defined in the deployment rule, and will always apply as long as the rule is unchanged and valid.

QUESTION 23
You need to provide users with a reproducible method to connect to a data source and transform the data by using an AI function. The solution must meet the following requirement
• Minimize development effort.
• Avoid including data in the file.
Which type of file should you create?

A. PBIDS

B. PBIX

C. PBIT

Correct Answer: C
Section:
Explanation:

QUESTION 24
You are planning a Power BI solution for a customer.
The customer will have 200 Power BI users. The customer identifies the following requirements:
• Ensure that all the users can create paginated reports.
• Ensure that the users can create reports containing AI visuals.
• Provide autoscaling of the CPU resources during heavy usage spikes.
You need to recommend a Power BI solution for the customer. The solution must minimize costs.
What should you recommend?

A. Power BI Premium per user

B. a Power BI Premium per capacity

C. Power BI Pro per user

D. Power Bl Report Server

Correct Answer: A
Section:
Explanation:
Announcing Power BI Premium Per User general availability and autoscale preview for Gen2.
Power BI Premium per user features and capabilities
* Pixel perfect paginated reports are available for operational reporting capabilities based on SSRS technology. Users can create highly formatted reports in various formats such as PDF and PPT, which are embeddable in applications and are designed to be printed or shared.
* Automated machine learning (AutoML) in Power BI enables business users to build ML models to predict outcomes without having to write any code.
* Etc.
Note:
Power BI empowers every business user and business analyst to get amazing insights with AI infused experiences. With Power BI Premium, we enable business analysts to not only analyze and visualize their data, but to also build an end-to-end data platform through drag and drop experiences.
Everything from ingesting and transforming data at scale, to building automated machine learning models, and analyzing massive volumes of data is now possible for our millions of business analysts.
Reference: https://powerbi.microsoft.com/nl-be/blog/announcing-power-bi-premium-per-usergeneral-availability-and-autoscale-preview-for-gen2/

QUESTION 25
You have a Power Bl dataset named Dataset1 that uses DirectQuery against an Azure SQL database named DB1. DB1 is a transactional database in the third normal form.
You need to recommend a solution to minimize how long it takes to execute the query. The solution must maintain the current functionality. What should you include in the recommendation?

A. Create calculated columns in Dataset1.

B. Remove the relationships from Dataset1.

C. Normalize the tables in DB1.

D. Denormalize the tables in DB1.

Correct Answer: D
Section:
Explanation:
Denormalize to improve query performance.
Note: Normalization prevents data duplications, preserves disk space, and improves the performance of the disk I/O operations. The downside of the normalization is that the queries based on these normalized tables require more table joins.
Schema denormalization (i.e. consolidation of some dimension tables) for such databases can significantly reduce costs of the analytical queries and improve the performance. Reference:
https://www.mssqltips.com/sqlservertip/7114/denormalization-dimensions-synapsemapping-data-flow/

QUESTION 26
You are building a Power Bl dataset that will use two data sources.
The dataset has a query that uses a web data source. The web data source uses anonymous authentication.
You need to ensure that the query can be used by all the other queries in the dataset.
Which privacy level should you select for the data source?

A. Public

B. Organizational

C. Private

D. None

Correct Answer: A
Section:
Explanation:

A Public data source gives everyone visibility to the data contained in the data source. Only files, internet data sources, or workbook data can be marked Public. Data from a Public data source may be freely folded to other sources.

Reference: https://docs.microsoft.com/en-us/power-bi/enterprise/desktop-privacy-levels

QUESTION 27

You have a Power BI report that contains the table shown in the following exhibit.

Store ID	Store	Returns	
6	Leo	$6,108	🔴
5	Fama	$6,097	🔴
13	Contoso	$5,214	🔴
11	Pomum	$4,968	🟡
7	VanArsdel	$4,964	🟡
10	Pirum	$4,644	🟡
2	Aliqui	$4,479	🟡
1	Abbas	$4,070	🟡
8	Natura	$3,376	🟡
14	Victoria	$2,317	🟢
4	Salvus	$2,296	🟢
12	Quibus	$2,208	🟢
3	Barba	$1,601	🟢
Total		**$52,342**	

The table contains conditional formatting that shows which stores are above, near, or below the monthly quota for returns. You need to ensure that the table is accessible to consumers of reports who have color vision deficiency. What should you do?

A. Add alt text to explain the information that each color conveys.

B. Move the conditional formatting icons to a tooltip report.

C. Change the icons to use a different shape for each color.

D. Remove the icons and use red, yellow, and green background colors instead.

Correct Answer: A
Section:
Explanation:
Report accessibility checklist, All Visuals.
* Ensure alt text is added to all non-decorative visuals on the page.
* Avoid using color as the only means of conveying information. Use text or icons to supplement or replace the color.
* Check that your report page works for users with color vision deficiency.
* Etc.
Reference: https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-accessibilitycreating-reports

QUESTION 28
You are using an Azure Synapse Analytics serverless SQL pool to query network traffic logs in the Apache Parquet format. A sample of the data is shown in the following table.

source		destination	
name	ip	name	ip
Network01	192.168.0.1	Internet	0.0.0.0

You need to create a Transact-SQL query that will return the source IP address.
Which function should you use in the select statement to retrieve the source IP address?

A. JSON_VALUE

B. FOR.JSON

C. CONVERT

D. FIRST VALUE

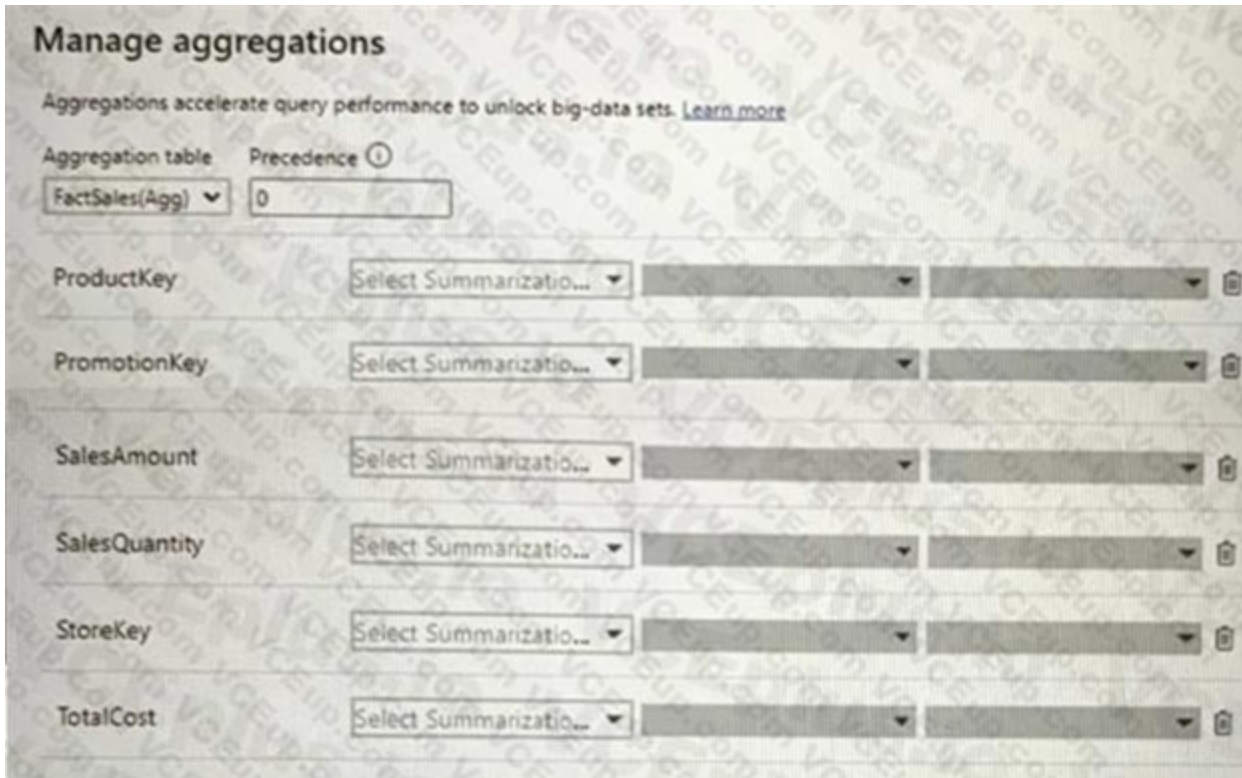**Correct Answer: A**
Section:

QUESTION 29
HOTSPOT
You are configuring an aggregation table as shown in the following exhibit.

Manage aggregations

Aggregations accelerate query performance to unlock big-data sets. Learn more

Aggregation table   Precedence ⓘ
FactSales(Agg) ▾   0

ProductKey   Select Summarizatio... ▾
PromotionKey   Select Summarizatio... ▾
SalesAmount   Select Summarizatio... ▾
SalesQuantity   Select Summarizatio... ▾
StoreKey   Select Summarizatio... ▾
TotalCost   Select Summarizatio... ▾

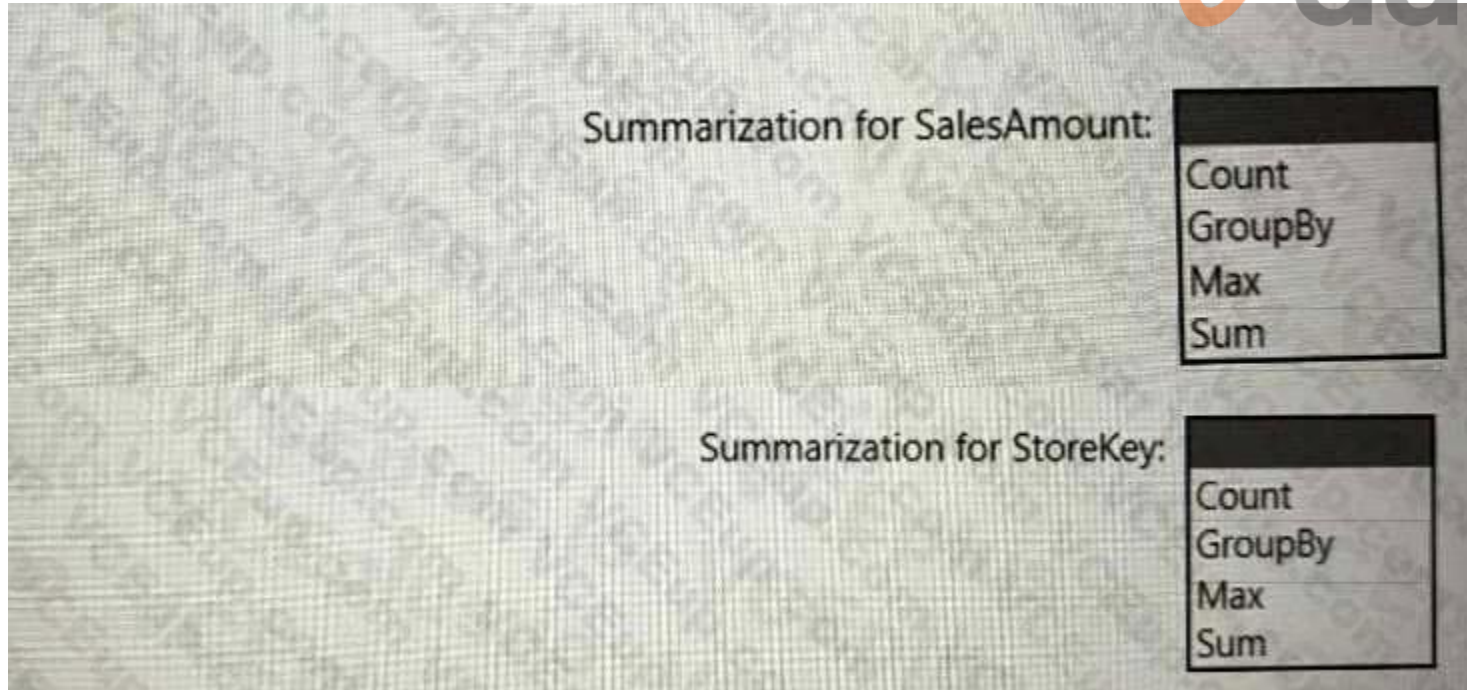The detail table is named FactSales and the aggregation table is named FactSales(Agg).

You need to aggregate SalesAmount for each store.

Which type of summarization should you use for SalesAmount and StoreKey? To answer, select the appropriate options in the answer area,
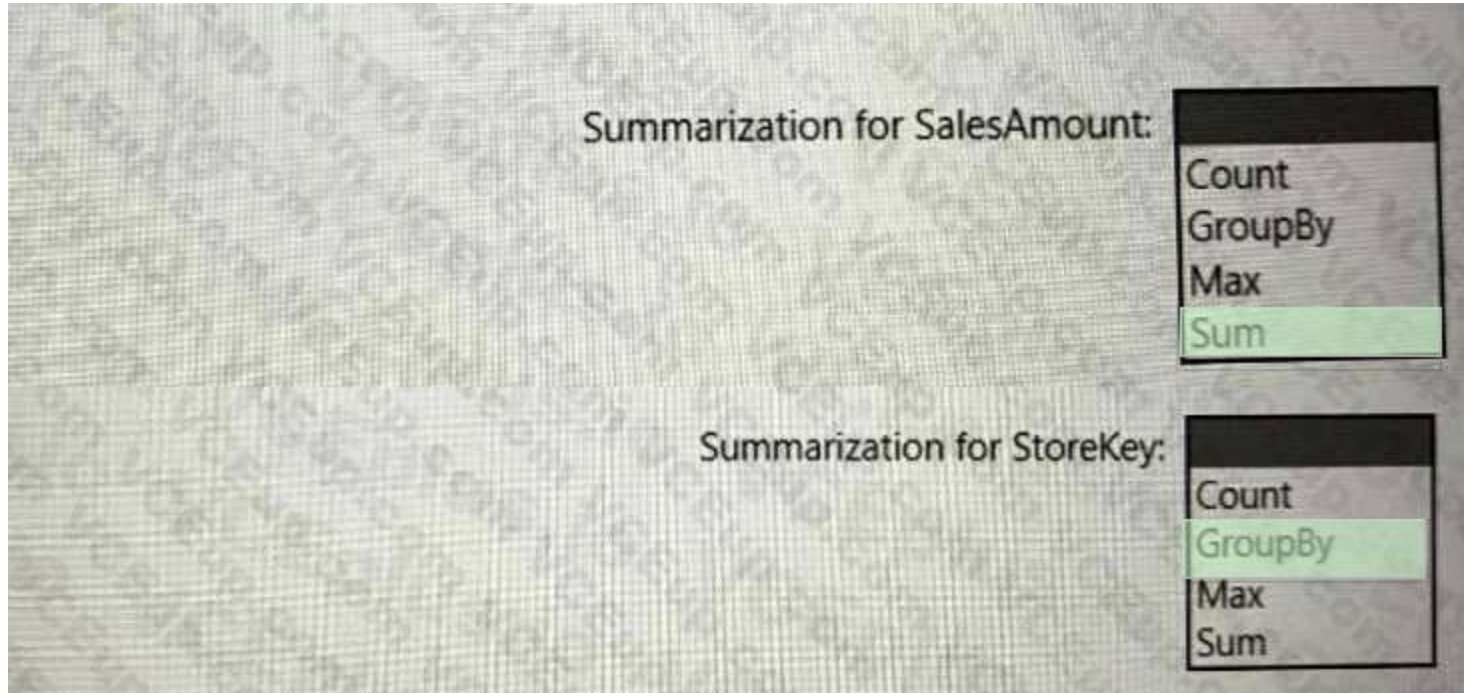
NOTE: Each correct selection is worth one point.

Hot Area:

Summarization for SalesAmount:

| Count |
| GroupBy |
| Max |
| Sum |

Summarization for StoreKey:

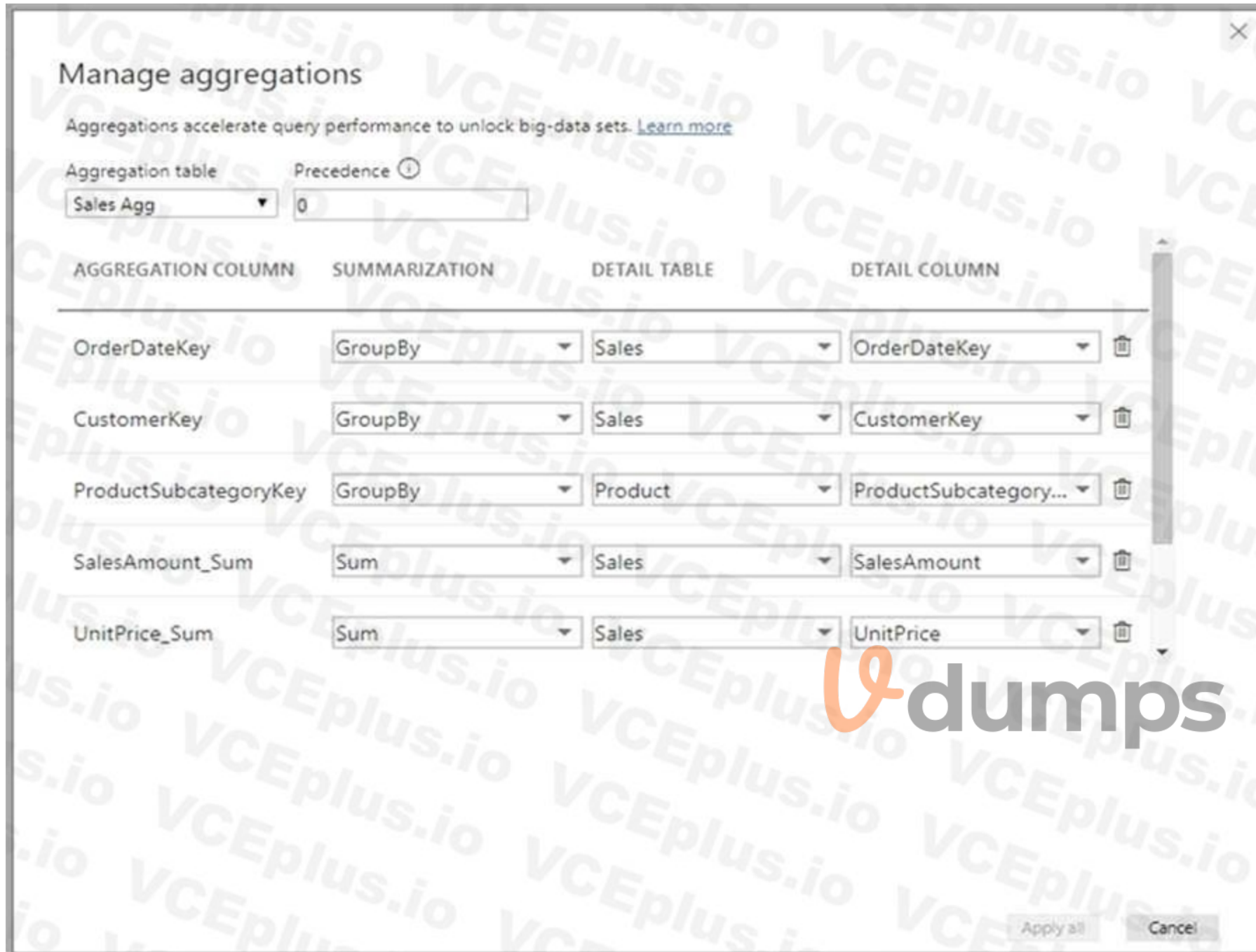| Count |
| GroupBy |
| Max |
| Sum |

Answer Area:

Section:
Explanation:
Box 1: Sum
The Manage aggregations dialog shows a row for each column in the table, where you can specify the aggregation behavior. In the following example, queries to the Sales detail table are internally redirected to the Sales Agg aggregation table.

Box 2: GroupBy
Reference: https://docs.microsoft.com/en-us/power-bi/transform-model/aggregations-advanced

QUESTION 30
DRAG DROP
You have a Power BI dataset. The dataset contains data that is updated frequently.
You need to improve the performance of the dataset by using incremental refreshes.
Which four actions should you perform in sequence to enable the incremental refreshes? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

Actions

Define the incremental refresh policy for the table.

Enable query caching.

Publish the model to the Power BI service.

Create `RangeStart` and `RangeEnd` parameters.

Use the Power BI REST API to post a message to `/refreshes`.

Apply a custom Date/Time filter to the data.

Answer Area

Correct Answer:

Actions

Enable query caching.

Use the Power BI REST API to post a message to `/refreshes`.

Answer Area

Create `RangeStart` and `RangeEnd` parameters.

Apply a custom Date/Time filter to the data.

Define the incremental refresh policy for the table.

Publish the model to the Power BI service.

Section:

Explanation:

Step 1: Create RangeStart and RangeEnd parameters.

Create parameters

In this task, use Power Query Editor to create RangeStart and RangeEnd parameters with default values. The default values apply only when filtering the data to be loaded into the model in Power BI Desktop. The values you enter should include only a small amount of the most recent data from your data source. When published to the service, these values are overridden by the incremental refresh policy.

Step 2: Apply a custom Date/Time filter to the data.
Filter data
With RangeStart and RangeEnd parameters defined, apply a filter based on conditions in the RangeStart and RangeEnd parameters.
Before continuing with this task, verify your source table has a date column of Date/Time data type.
Step 3: Define the incremental refresh policy for the table.
Define policy
After you've defined RangeStart and RangeEnd parameters, and filtered data based on those parameters, you define an incremental refresh policy. The policy is applied only after the model is published to the service and a manual or scheduled refresh operation is performed.
Step 4: Publish the model to the Power BI service.
Save and publish to the service When your RangeStart and RangeEnd parameters, filtering, and refresh policy settings are complete, be sure to save your model, and then publish to the service.
Reference: https://docs.microsoft.com/en-us/power-bi/connect-data/incremental-refresh-configure

QUESTION 31
HOTSPOT
You have a Power BI dataset that has the query dependencies shown in the following exhibit.
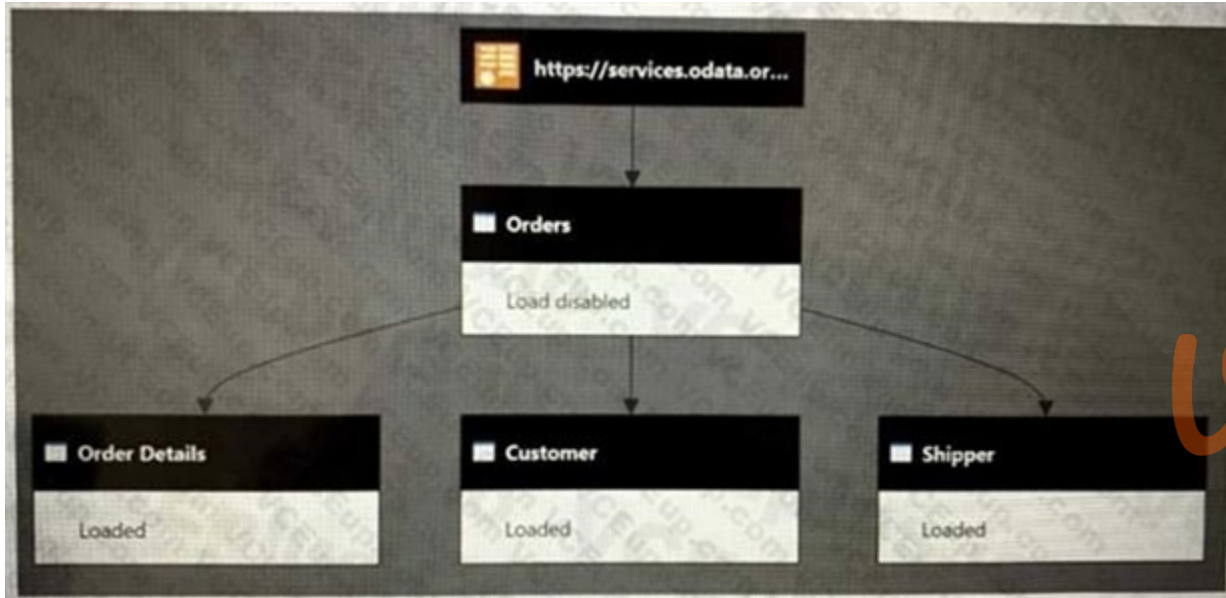
Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area:

Answer Area

When the dataset refreshes, the `Orders` query will be executed [answer choice] times.

0
1
3

[Answer choice] will reduce data refresh times for this model.

Duplicating the Orders query instead of referencing the query
Replacing the Orders query with a dataflow
Using Table.Buffer in the Orders query

Section:
Explanation:
Box 1: 3
Power Query doesn't start at the first query and work down, it starts at the bottom (last) query and works backwards, so 3 tables from 1 will cause it to process that first source table 3 times.
Box 2: Using Table.Buffer in the Orders query
Table.Buffer buffers a table in memory, isolating it from external changes during evaluation.
Buffering is shallow. It forces the evaluation of any scalar cell values, but leaves non-scalar values (records, lists, tables, and so on) as-is.
Note that using this function might or might not make your queries run faster. In some cases, it can make your queries run more slowly due to the added cost of reading all the data and storing it in memory, as well as the fact that buffering prevents downstream folding.
Example 1
Load all the rows of a SQL table into memory, so that any downstream operations will no longer be able to query the SQL server.
Usage
let
Source = Sql.Database("SomeSQLServer", "MyDb"),
MyTable = Source{[Item="MyTable"]}[Data],
BufferMyTable = Table.Buffer(dbo_MyTable)
in
BufferMyTable
Output
table
Reference: https://radacad.com/performance-tip-for-power-bi-enable-load-sucks-memory-up
https://docs.microsoft.com/en-us/powerquery-m/table-buffer

QUESTION 32
DRAG DROP
You are configuring Azure Synapse Analytics pools to support the Azure Active Directory groups shown in the following table.

Name	Requirement
Group1	Analyze data to create and train machine learning models in Synapse Analytics.
Group2	Execute complex queries with multiple joins against relational data. Results will be exported by using PolyBase.
Group3	Query and load data from Apache Parquet files stored in Azure Data Lake Storage Gen2. Costs must be based on the amount of data processed.

Which type of pool should each group use? To answer, drag the appropriate pool types to the groups.

Each pool 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.

Select and Place:

Pool Types	Answer Area	
Apache Spark pool	Group1:	Pool Type
Dedicated SQL pool	Group2:	Pool Type
Serverless SQL pool	Group3:	Pool Type

Correct Answer:

Pool Types	Answer Area	
	Group1:	Apache Spark pool
	Group2:	Dedicated SQL pool
	Group3:	Serverless SQL pool

Section:
Explanation:
Box 1: Apache Spark pool
An Apache Spark pool provides open-source big data compute capabilities. After you've created an
Apache Spark pool in your Synapse workspace, data can be loaded, modeled, processed, and distributed for faster analytic insight.
Box 2: Dedicated SQL Pool
Dedicated SQL Pool - Data is stored in relational tables
Box 3: Serverless SQL pool
Serverless SQL pool - Cost is incurred for the data processed per query
Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-create-apachespark-pool-portal
https://www.royalcyber.com/blog/data-services/dedicated-sql-pool-vs-serverless-sql/

QUESTION 33
HOTSPOT
You use Advanced Editor in Power Query Editor to edit a query that references two tables named Sales and Commission. A sample of the data in the Sales table is shown in the following table.

OrderID	SalesPerson	Amount
101	Tom	199.99
103	Eileen	279.99
108	Enrique	333.42

A sample of the data in the Commission table is shown in the following table.

Person	Commission
Tom	0.04
Eileen	0.05

You need to merge the tables by using Power Query Editor without losing any rows in the Sales table.

How should you complete the query? To answer, select the appropriate options in the answer are a.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Combined = Table.

| Combine |
| InsertRows |
| Join |
| TransformRow |

(Sale, "SalesPerson", Commission, "Name", JoinKind.

| LeftAnti |
| LeftOuter |
| RightAnti |
| RightOuter |

),

Answer Area:

Answer Area

Combined = Table.

| Combine |
| InsertRows |
| Join |
| TransformRow |

(Sale, "SalesPerson", Commission, "Name", JoinKind.

| LeftAnti |
| LeftOuter |
| RightAnti |
| RightOuter |

),

Section:

Explanation:
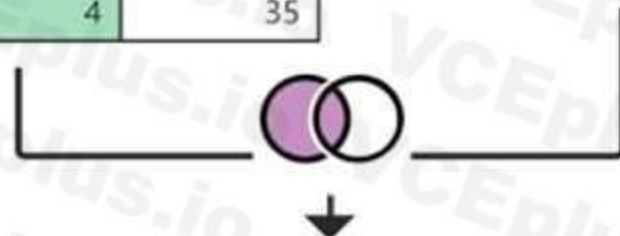
Box 1: Join

Box 2: LeftOuter

Left outer join

One of the join kinds available in the Merge dialog box in Power Query is a left outer join, which keeps all the rows from the left table and brings in any matching rows from the right table.

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	4	35

Right Table

ID	Country
1	USA
2	Canada
3	Panama

Merged Table

Date	CountryID	Units	Country
1/1/2020	1	40	USA
1/2/2020	1	25	USA
1/3/2020	3	30	Panama
1/4/2020	4	35	*null*

Reference: https://docs.microsoft.com/en-us/power-query/merge-queries-left-outer

QUESTION 34
HOTSPOT
You are building a Power Bl dataset that contains a table named Calendar. Calendar contains the following calculated column. pfflag = IF('Calendar'[Date] < TOOAYQ, "Past", "Future") You need to create a measure that will perform a fiscal prior year-to-date calculation that meets the following requirements:
• Returns the fiscal prior year-to-date value for [sales Amount]
• Uses a fiscal year end of June 30
• Produces no result for dates in the future
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.

Hot Area:

Answer Area

```
PYTD =

CALCULATE (

    [Sales Amount],

    ┌─────────────────────┐
    │ CALCULATETABLE      │
    │ CROSSJOIN           │
    │ SUMMARIZECOLUMNS    │
    │ UNION               │
    └─────────────────────┘
    )
                            ┌─────────────────────┐  ( DATESYTD ( 'Calendar'[Date], "6/30/2022" ) ),
                            │ DATEADD             │
        'Ca│ FIRSTDATE     │
                            │ PARALLELPERIOD      │
                            │ SAMEPERIODLASTYEAR  │
                            └─────────────────────┘

        'Calendar'[pfFlag] =   ┌─────────────────────┐
                               │ "Future"            │
            )                  │ NOW()               │
                               │ "Past"              │
            )                  │ TODAY()             │
                               └─────────────────────┘
```

Answer Area:

Answer Area

```
PYTD =
CALCULATE (
    [Sales Amount],
```

| CALCULATETABLE |
| CROSSJOIN |
| SUMMARIZECOLUMNS |
| UNION |

```
)
```

'Ca

| DATEADD |
| FIRSTDATE |
| PARALLELPERIOD |
| SAMEPERIODLASTYEAR |

(DATESYTD ('Calendar'[Date], "6/30/2022")),

```
'Calendar'[pFlag] =
```

| "Future" |
| NOW() |
| "Past" |
| TODAY() |

```
    )
)
```

Section:
Explanation:
Box 1: CALCULATETABLE
CALCULATETABLE evaluates a table expression in a modified filter context.
Syntax: CALCULATETABLE(<expression>[, <filter1> [, <filter2> [, …]]])
Incorrect:
* SUMMARIZECOLUMNS
SUMMARIZECOLUMNS returns a summary table over a set of groups.
Syntax: SUMMARIZECOLUMNS(<groupBy_columnName> [, < groupBy_columnName >]…,
[<filterTable>]…[, <name>, <expression>]…)
* CROSSJOIN returns a table that contains the Cartesian product of all rows from all tables in the arguments. The columns in the new table are all the columns in all the argument tables.
Syntax: CROSSJOIN(<table>, <table>[, <table>]…)
* UNION creates a union (join) table from a pair of tables.
Syntax: UNION(<table_expression1>, <table_expression2> [,<table_expression>]…)
Box 2: SAMEPERIODLASTYEAR
SAMEPERIODLASTYEAR returns a table that contains a column of dates shifted one year back in time from the dates in the specified dates column, in the current context.
Syntax: SAMEPERIODLASTYEAR(<dates>)
The dates returned are the same as the dates returned by this equivalent formula: DATEADD(dates, 1, year)
Example:

The following sample formula creates a measure that calculates the previous year sales of Reseller sales.

= CALCULATE(SUM(ResellerSales_USD[SalesAmount_USD]),

SAMEPERIODLASTYEAR(DateTime[DateKey))

Box 3: TODAY()

TODAY() returns the current date.

The TODAY function is useful when you need to have the current date displayed on a worksheet, regardless of when you open the workbook. It is also useful for calculating intervals.

Example:

The following sample formula creates a measure that calculates the 'Running Total' for Internet sales.

= CALCULATE(SUM(InternetSales_USD[SalesAmount_USD]), DATESYTD(DateTime[DateKey]))

Reference: https://docs.microsoft.com/en-us/dax/calculatetable-function-dax

https://docs.microsoft.com/en-us/dax/sameperiodlastyear-function-dax

https://docs.microsoft.com/en-us/dax/datesytd-function-dax

QUESTION 35

You use the Vertipaq Analyzer to analyze tables in a dataset as shown in the Tables exhibit. (Click the Tables tab.)

The table relationships for the dataset are shown in the Relationships exhibit. (Click the Relationships tab.)

You need to reduce the model size by eliminating invalid relationships.

Which column should you remove?

A. Sales[Sales Amount]

B. Sales[RowID]

C. Sales[Sales ID]

D. Plan[RowID]

Correct Answer: C

Section:

Explanation:

Sales[Row ID] has 858,786 missing keys and 858,789 Max From Cardinality.

Note: The Max From Cardinality column defines the cost of the relationship which is the amount of time DAX needs to transfer the filters from the dimensions table to the fact table.

Reference: https://blog.enterprisedna.co/vertipaq-analyzer-tutorial-relationships-referentialintegrity/

QUESTION 36
You have a sales report as shown in the following exhibit.

The sales report has the following characteristics:
The measures are optimized.
The dataset uses import storage mode.
Data points, hierarchies, and fields cannot be removed or filtered from the report page.
From powerbi.com, users experience slow load times when viewing the report.
You need to reduce how long it takes for the report to load without affecting the data displayed in the report.
Which two actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Change the report theme to monochromatic.

B. Replace the single-value cards with a multi-row card.

C. Replace the product category charts with a bar chart for sales and a hierarchy of Category and Sub Category on the axis.

D. Replace all the filters on the Filters pane with visual slicers on the report page.

Correct Answer: B, C
Section:

QUESTION 37
You have an Azure subscription that contains an Azure Synapse Analytics workspace. You create an Azure Data Lake Storage Gen2 account and upload a CSV file named Filel.csv. You need to use Synapse Studio to query the

data in

Filel.csv by using a serverless SQL pool. Which Transact-SQL operator should you include in the query?

A. STRIMO_SPLIT

B. OPENOUERY

C. OPCNROWSET

D. OPEMDATASOURCE

Correct Answer: C
Section:
Explanation:

QUESTION 38
You are using a Python notebook in an Apache Spark pool in Azure Synapse Analytics.
You need to present the data distribution statistics from a DataFrame in a tabular view.
Which method should you invoke on the DataFrame?

A. rollup

B. cov

C. explain

D. describe

Correct Answer: D
Section:
Explanation:
The aggregating statistic can be calculated for multiple columns at the same time with the describe function.
Example:
titanic[["Age", "Fare"]].describe()
Out[6]:
Age Fare
count 714.000000 891.000000
mean 29.699118 32.204208
std 14.526497 49.693429
min 0.420000 0.000000
25% 20.125000 7.910400
50% 28.000000 14.454200
75% 38.000000 31.000000
max 80.000000 512.329200
Reference:
https://pandas.pydata.org/docs/getting_started/intro_tutorials/06_calculate_statistics.html

QUESTION 39
You are using a Python notebook in an Apache Spark pool in Azure Synapse Analytics.
You need to present the data distribution statistics from a DataFrame in a tabular view.
Which method should you invoke on the DataFrame?

A. sample

B. describe

C. freqItems

D. explain

Correct Answer: B
Section:
Explanation:
pandas.DataFrame.describe
Descriptive statistics include those that summarize the central tendency, dispersion and shape of a dataset's distribution, excluding NaN values.
Analyzes both numeric and object series, as well as DataFrame column sets of mixed data types. The output will vary depending on what is provided.
Reference: https://pandas.pydata.org/pandasdocs/stable/reference/api/pandas.DataFrame.describe.html

QUESTION 40
You have a deployment pipeline for a Power BI workspace. The workspace contains two datasets that use import storage mode.
A database administrator reports a drastic increase in the number of queries sent from the Power Bi service to an Azure SQL database since the creation of the deployment pipeline.
An investigation into the issue identifies the following:
One of the datasets is larger than 1 GB and has a fact table that contains more than 500 million rows.
When publishing dataset changes to development, test, or production pipelines, a refresh is triggered against the entire dataset.
You need to recommend a solution to reduce the size of the queries sent to the database when the dataset changes are published to development, test, or production.
What should you recommend?

A. Request the authors of the deployment pipeline datasets to reduce the number of datasets republished during development.

B. In the dataset, delete the fact table.

C. Configure the dataset to use a composite model that has a DirectQuery connection to the fact table.

D. From Capacity settings in the Power Bi Admin portal, reduce the Max Intermediate Row Set Count setting.

Correct Answer: C
Section:
Explanation:
Previously in Power BI Desktop, when you used a DirectQuery in a report, no other data connections, whether DirectQuery or import, were allowed for that report. With composite models, that restriction is removed. A report can seamlessly include data connections from more than one
DirectQuery or import data connection, in any combination you choose.
The composite models capability in Power BI Desktop consists of three related features:
* Composite models: Allows a report to have two or more data connections from different source groups, such as one or more DirectQuery connections and an import connection, two or more
DirectQuery connections, or any combination thereof.
* Etc.
Reference: https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-composite-models

QUESTION 41
You have a Power BI workspace named workspace1 that contains three reports and two dataflows.
You have an Azure Data Lake Storage account named storage1.
You need to integrate workspace1 and storage1.
What should you do first?

A. In storage1, create a folder named dataflows.

B. Delete the dataflows from workspace1.

C. In storage1, create a folder named reports.

D. Delete the reports from workspace1.

Correct Answer: A
Section:

QUESTION 42
HOTSPOT
You have a Power BI tenant that is configured as shown in the following exhibit.

Export and sharing settings

▸ Allow Azure Active Directory guest users to access Power BI
Enabled for the entire organization

▸ Invite external users to your organization
Disabled for the entire organization

▸ Allow Azure Active Directory guest users to edit and manage content in the organization
Disabled for the entire organization

▸ Show Azure Active Directory guests in lists of suggested people
Enabled for the entire organization

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 answer is worth one point.

Hot Area:

Answer Area

A new external user **[answer choice]**.

| must be added in Azure AD to access Power BI content | ▼ |
| can add other external users to the tenant |
| must be added in Azure AD to access Power BI content |
| will appear in lists of suggested users |

For an existing guest user to edit a report, you must modify a **[answer choice]** setting.

| workspace | ▼ |
| report |
| tenant |
| workspace |

Answer Area:

Answer Area

A new external user **[answer choice]**.

must be added in Azure AD to access Power BI content ▼

| can add other external users to the tenant |
| must be added in Azure AD to access Power BI content |
| will appear in lists of suggested users |

For an existing guest user to edit a report, you must modify a **[answer choice]** setting.

workspace ▼

| report |
| tenant |
| workspace |

Section:
Explanation:

QUESTION 43
DRAG DROP
You have a Power BI data model that contains two tables named Products and Sales.
A one-to-many relationship exists between the tables.
You have a report that contains a report-level filter for Products.
You need to create a DAX measure that will return the percent of total sales for each product. The measure must respect the report-level filter when calculating the total.
How should you complete the measure? To answer, drag the appropriate DAX functions to the correct targets. Each function may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Select and Place:

Functions

| ALL |
| ALLSELECTED |
| CALCULATE |
| FILTER |
| SELECTEDVALUE |

Answer Area

```
Percent of Product Sales =

VAR ProductSales = SUM ('Sales'[Sales])

VAR AllSales =

[                    ] (SUM ('Sales'[Sales]),    [                    ] ('Products'[Product]))

RETURN

DIVIDE (ProductSales, AllSales)
```

Correct Answer:

Functions

ALL

FILTER

SELECTEDVALUE

Answer Area

Percent of Product Sales =

VAR ProductSales = SUM ('Sales'[Sales])

VAR AllSales =

| CALCULATE | (SUM ('Sales'[Sales]), | ALLSELECTED | ('Products'[Product]))

RETURN

DIVIDE (ProductSales, AllSales)

Section:
Explanation:

QUESTION 44
You have an Azure Synapse Analytics notebook.
You run the xxsq1 magic command to render data into an Apache Spark DataFrame named df1, and then you run the following code.
display (df1, summary = true)
Which three attributes will be returned by the command? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. type
B. range
C. missing
D. unique
E. ordinal

Correct Answer: A, B, C
Section:

QUESTION 45
DRAG DROP
You have a shared dataset in Power BI named Dataset1.
You have an on-premises Microsoft SQL Server database named DB1.
You need to ensure that Dataset1 refreshes data from DB1.
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.

Select and Place:

Actions

Actions
From powerbi.com, configure Dataset1 to use a data gateway.
From powerbi.com, add a data source to the gateway clusters.
Create a virtual network data gateway.
Install the on-premises data gateway (standard mode).
Configure a private endpoint.
Install the on-premises data gateway (personal mode).

Answer Area

Correct Answer:

Actions

Actions
Create a virtual network data gateway.
Configure a private endpoint.
Install the on-premises data gateway (personal mode).

Answer Area

Answer Area
Install the on-premises data gateway (standard mode).
From powerbi.com, add a data source to the gateway clusters.
From powerbi.com, configure Dataset1 to use a data gateway.

Section:
Explanation:
Step 1: Install the on-premises data gateway (standard mode)
The personal mode is only for a single user, not to be used for a shared dataset.
Step 2: From powerbi.com, add a data source to the gateway clusters
After you install the on-premises data gateway, you can add data sources that can be used with the gateway.
Add a data source

New data source

Gateway cluster name *

GatewayApril ⌄

Data source name *

New data source

Data source type *

SQL Server ⌄

After you fill in everything, select Create. You can now use this data source for scheduled refresh or
DirectQuery against a SQL Server that's on-premises. You see Created New data source if it succeeded.
Step 3: From powerbi.com, configure Dataset1 to use a data gateway.
Connect a dataset to a SQL Server database
In Power BI Desktop, you connected directly to your on-premises SQL Server database, but the Power
BI service requires a data gateway to act as a bridge between the cloud and your on-premises network. Follow these steps to add your on-premises SQL Server database as a data source to a gateway and then connect your dataset to this data source.
Sign in to Power BI. In the upper-right corner, select the settings gear icon and then select Settings.
On the Datasets tab, select the dataset AdventureWorksProducts, so you can connect to your onpremises
SQL Server database through a data gateway.
Expand Gateway connection and verify that at least one gateway is listed.
Under Actions, expand the toggle button to view the data sources and select the Add to gateway link.
On the Gateways management page, on the Data Source Settings tab, enter and verify the following information, and select Add.
On the Datasets tab, expand the Gateway connection section again. Select the data gateway you configured, which shows a Status of running on the machine where you installed it, and select Apply.
Reference: https://docs.microsoft.com/en-us/power-bi/connect-data/service-gateway-personalmode
https://docs.microsoft.com/en-us/power-bi/connect-data/service-gateway-sql-tutorial
https://docs.microsoft.com/en-us/power-bi/connect-data/service-gateway-enterprise-manage-sql

QUESTION 46
HOTSPOT

You have the Power BI workspaces shown in the following exhibit.

Workspaces

View personal and group workspaces that exist in your organization. To change users' ability to create workspaces, see Tenant settings.

○ Refresh ↓ Export

Name ∨	Description ∨	Type ∨	State ∨	Capacity name ∨	Capacity SKU Tier ∨	Workspace upg... ∨
Web Analytics		Workspace	Active			
Infrastructure Svcs		Workspace	Active		Premium Per User - Re... PP3	
Sales Orders		Workspace	Active	pbie01	A1	
PBI Capacity Metrics		Workspace	Orphaned			
PersonalWorkspace		Personal Group	Active			

Items per page: 10 ▾   1 - 7 of 7 |< < > >|

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

Hot Area:

Answer Area

The [answer choice] workspace will allow users outside the organization to view private reports from a custom application without a Power BI license.

Infrastructure Svcs
PBI Capacity Metrics
Sales Orders
Web Analytics

A user must be assigned the [answer choice] role to change the state of the PBI Capacity Metrics workspace to **Active**.

Admin
Contributor
Member
Viewer

Answer Area:

The [answer choice] workspace will allow users outside the organization to view private
reports from a custom application without a Power BI license.

Infrastructure Svcs
PBI Capacity Metrics
Sales Orders
Web Analytics

A user must be assigned the [answer choice] role to change the state of the PBI Capacity
Metrics workspace to **Active**.

Admin
Contributor
Member
Viewer

Section:

Explanation:

Box 1: Infrastrucrue Svcs

Infrastrucrue Svcs is a Premium workspace.

If users have a free license and the workspace is stored in Premium (dedicated) capacity, they will be able to view and interact with the content in that workspace.

If users have a free license and the workspace is stored in shared capacity (not premium), they will not be able to see the content in shared workspace, only "My workspace".

If users have pro license, they will be able to view and interact with the content in that workspace.

Box 2: Admin

We need to activate the Orpaned workspace.

An orphaned workspace is one that does not have an admin assigned.

If you're a Service Admin, you can now view all of your organization's workspaces through the Admin
Portal in the user interface.

Workspaces

View personal and group workspaces that exist in your organization. To change users' ability to create workspaces, see Tenant settings

◯ Refresh   ⬇ Export

Name ∨ ▽	Description ∨	Type ∨	State ∨	Read only ∨
Wannabe Orphan W...	Orphan testing	Workspace	Orphaned	False

It's easy to Recover an orphan from this screen. Simply select the workspace and click Recover, then add yourself or another user as an admin.

Reference: https://community.powerbi.com/t5/Service/Difference-between-Public-and-Privateworkspace/m-p/1382219

https://docs.microsoft.com/en-us/power-bi/admin/service-admin-portal-workspaces

QUESTION 47

HOTSPOT

You are using Azure Synapse Studio to explore a dataset that contains data about taxi trips.

You need to create a chart that will show the total trip distance according to the number of passengers as shown in the following exhibit.

```
1  SELECT PassengerCount,
2      SUM(TripDistanceMiles) as SumTripDistance,
3      AVG(TripDistanceMiles) as AvgTripDistance
4  FROM   dbo.NYCTaxiTripSmall
5  WHERE TripDistanceMiles > 0 AND PassengerCount > 0
6  GROUP BY PassengerCount
7  ORDER BY PassengerCount;
```

How should you configure the chart? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area:

Category column:

AvgTripDistance
PassengerCount
SumTripDistance
TripDistanceMiles

Legend (series) column:

AvgTripDistance
PassengerCount
SumTripDistance
TripDistanceMiles

Section:

Explanation:

QUESTION 48

HOTSPOT

You have an Azure Synapse workspace named Workspace1.

You need to use PySpark in a notebook to read data from a SQL pool as an Apache Spark DataFrame and display the top five How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
val spark_read =                                   (s"$sql_pool_name.dbo.PublicHoliday")
```

spark.read.load
spark.read.parquet
spark.read.sqlanalytics

```
                                                   (5, truncate = false)
```

spark_read.describe().show
spark_read.show
spark_read.take

Answer Area:

Answer Area

```
val spark_read =  ┌─────────────────────────┐  (s"$sql_pool_name.dbo.PublicHoliday")
                  │ spark.read.load         │
                  │ spark.read.parquet      │
                  │ spark.read.sqlanalytics │
                  └─────────────────────────┘

                  ┌─────────────────────────┐  (5, truncate = false)
                  │ spark_read.describe().show│
                  │ spark_read.show         │
                  │ spark_read.take         │
                  └─────────────────────────┘
```

Section:

Explanation:

Box 1:spark.read.sqlanalytics

Read from a SQL Pool table with Spark

// Read the table we just created in the sql pool as a Spark dataframe

val spark_read = spark.read.

sqlanalytics(s"$sql_pool_name.dbo.PublicHoliday")

spark_read.show(5, truncate = false)

Box 2: spark_read.show

Sample output:

```
spark_read: org.apache.spark.sql.DataFrame = [countryOrRegion: string, holidayName: string ... 3 more fields]
+--------------+-------------------------+-------------------------+------------+-----------------+
|countryOrRegion|holidayName             |normalizeHolidayName     |isPaidTimeOff|countryRegionCode|
+--------------+-------------------------+-------------------------+------------+-----------------+
|Czech         |Den české státnosti      |Den české státnosti      |null        |CZ               |
|Norway        |Søndag                   |Søndag                   |null        |NO               |
|Sweden        |Søndag                   |Søndag                   |null        |SE               |
|India         |Gandhi Jayanti           |Gandhi Jayanti           |true        |IN               |
|Germany       |Tag der Deutschen Einheit|Tag der Deutschen Einheit|null        |DE               |
+--------------+-------------------------+-------------------------+------------+-----------------+
only showing top 5 rows
```

Reference: https://github.com/AzureSamples/Synapse/blob/main/Notebooks/Scala/03%20Read%20and%20write%20from%20SQL%20pool%20table.ipynb

QUESTION 49

DRAG DROP

You plan to create a Power BI report that will use an OData feed as the data source. You will retrieve all the entities from two different collections by using the same service root The OData feed is still in development. The location of the feed will change once development is complete.

The report will be published before the OData feed development is complete.

You need to minimize development effort to change the data source once the location changes.

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.

Select and Place:

Actions

Create a parameter that contains the service root URI.
From Advanced Editor, duplicate the query and change the resource path in the URL.
Get data from an OData feed source and use the parameter to populate the first part of the URL.
From Advanced Editor, get data from an OData feed source and use the parameter to populate the last part of the URL.
From Advanced Editor, reference the query and change the resource path in the URL.

Correct Answer:

Actions

From Advanced Editor, get data from an OData feed source and use the parameter to populate the last part of the URL.
From Advanced Editor, reference the query and change the resource path in the URL.

Answer Area

Create a parameter that contains the service root URI.
Get data from an OData feed source and use the parameter to populate the first part of the URL.
From Advanced Editor, duplicate the query and change the resource path in the URL.

Section:
Explanation:
Step 1: Create a parameter that contains the service root URI
Step 2: Get data from OData feed source and use the parameter to populate the first part of the URL.
The URI is in the first part of the query.
Example: let
Source = OData.Feed ("https://analytics.dev.azure.com/{organization}/{project}/_odata/v3.0preview/WorkItemSnapshot?
"
&"$apply=filter("
&"WorkItemType eq 'Bug' "
&"and StateCategory ne 'Completed' "
&"and startswith(Area/AreaPath,'{areapath}') "
&"and DateValue ge {startdate} "
&") "

```
&"/groupby( "
&"(DateValue,State,WorkItemType,Priority,Severity,Area/AreaPath,Iteration/IterationPath,AreaSK), "
&"aggregate($count as Count) "
&") "
,null, [Implementation="2.0",OmitValues = ODataOmitValues.Nulls,ODataVersion = 4])
in
Source
```

Box 3: From Advanced Editor, duplicate the query and change the resource path in the URL.
Choose Get Data, and then Blank Query.
From the Power BI Query editor, choose Advanced Editor.
The Advanced Editor window opens.
Edit the query.
Etc.
Reference: https://docs.microsoft.com/en-us/azure/devops/report/powerbi/odataquery-connect

QUESTION 50
HOTSPOT
You have an Azure Data Lake Storage Gen 2 container that stores more than 300,000 files representing hourly telemetry dat a. The data is organized in folders by the year, month, and day according to when the telemetry was captured.
You have the following query in Power Query Editor.

```
let
    Source = AzureStorage.Blobs("https://tmppbie01.blob.core.windows.net/logs/"),
    Filtered = Table.SelectRows(Source, each Text.StartsWith([Name], "2019/12/")
        and [Extension] = ".csv"),
    Transformed = Table.AddColumn(Filtered, "Transformed", each TransformFiles([Content])),
    Limited = Table.SelectColumns(Transformed, "Transformed"),
    Expanded = Table.ExpandTableColumn(Limited, "Transformed", {"Date", "Name", "Activity"}),
    Final = Table.TransformColumnTypes(Expanded,
        {{"Date", type date}, {"Name", type text}, {"Activity", type text}})
in
    Final
```

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

Hot Area:

Answer Area		
Statements	**Yes**	**No**
The query uses the hierarchical namespace of the storage account.	○	○
The query uses a custom function to load file data.	○	○
Changing the source to use AzureStorage.DataLake will reduce the load time of the query.	○	○

Answer Area:

Answer Area

Statements	Yes	No
The query uses the hierarchical namespace of the storage account.	☑	○
The query uses a custom function to load file data.	○	☑
Changing the source to use `AzureStorage.DataLake` will reduce the load time of the query.	☑	○

Section:
Explanation:
Box 1: Yes
A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With a hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.
Box 2: No
Table.SelectRows returns a table of rows from the table, that matches the selection condition.
Box 3: Yes
Azure Data Lake Storage has higher throughput and IOPS.
Note: Azure Blob Storage is a general purpose, scalable object store that is designed for a wide variety of storage scenarios. Azure Data Lake Storage is a hyper-scale repository that is optimized for big data analytics workloads.
Azure Data Lake Storage use Cases: Batch, interactive, streaming analytics and machine learning data such as log files, IoT data, click streams, large datasets
Reference: https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace
https://docs.microsoft.com/en-us/powerquery-m/table-selectrows
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-comparison-with-blobstorage

QUESTION 51
HOTSPOT
You manage a dataset that contains the two data sources as shown in the following table.

Data source	Type of data	Privacy level
Azure SQL database	Sensitive company data	Private
Microsoft SharePoint folder	Non-sensitive company data	Private

When you attempt to refresh the dataset in powerbi.com, you receive the following error message:
"[Unable to combine data] Add Columns is accessing data sources that have privacy levels which cannot be used together. Please rebuild this data combination."
You discover that the dataset contains queries that fold data from the SharePoint folder to the Azure
SQL database.
You need to resolve the error. The solution must provide the highest privacy possible.
Which privacy level should you select for each data source? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area:

Azure SQL database:
Organizational
Private
Public

SharePoint folder:
Organizational
Private
Public

Section:
Explanation:
Box 1: Private
This Formula.Firewall error is the result of Power Query's Data Privacy Firewall (aka the Firewall)
Note: Folding is a term that refers to converting expressions in M (such as filters, renames, joins, and so on) into operations against a raw data source (such as SQL, OData, and so on).
Box 2: Organizational
Organizational Limits the visibility of a data source to a trusted group of people. It is isolated from all
Public data sources, but is visible to other Organizational data sources. A common example is a Microsoft Word document on an intranet SharePoint site with permissions enabled for a trusted group.
Reference: https://support.microsoft.com/en-us/office/set-privacy-levels-power-query-cc3ede4d359e-4b28-bc72-9bee7900b540

QUESTION 52
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using an Azure Synapse Analytics serverless SQL pool to query a collection of Apache Parquet files by using automatic schema inference. The files contain more than 40 million rows of UTF-8encoded business names, survey names, and participant counts. The database is configured to use the default collation.
The queries use open row set and infer the schema shown in the following table.
You need to recommend changes to the queries to reduce I/O reads and tempdb usage.
Solution: You recommend defining an external table for the Parquet files and updating the query to use the table
Does this meet the goal?

A. Yes

B. No

Correct Answer: B
Section:

QUESTION 53

You have a deployment pipeline for a Power BI workspace. The workspace contains two datasets that use import storage mode.

A database administrator reports a drastic increase in the number of queries sent from the Power BI service to an Azure SQL database since the creation of the deployment pipeline.

An investigation into the issue identifies the following:

One of the datasets is larger than 1 GB and has a fact table that contains more than 500 million rows.

When publishing dataset changes to development, test, or production pipelines, a refresh is triggered against the entire dataset.

You need to recommend a solution to reduce the size of the queries sent to the database when the dataset changes are published to development, test, or production.

What should you recommend?

A. From Capacity settings in the Power BI Admin portal, reduce the Max Intermediate Row Set Count setting.

B. Configure the dataset to use a composite model that has a DirectQuery connection to the fact table.

C. Enable the large dataset storage format for workspace.

D. From Capacity settings in the Power BI Admin portal, increase the Max Intermediate Row Set Count setting.

Correct Answer: B
Section:
Explanation:
A composite model in Power BI means part of your model can be a DirectQuery connection to a data source (for example, SQL Server database), and another part as Import Data (for example, an Excel file). Previously, when you used
DirectQuery, you couldn't even add another data source into the model.
DirectQuery and Import Data have different advantages.
Now the Composite Model combines the good things of both Import and DirectQuery into one model. Using the Composite Model, you can work with big data tables using DirectQuery, and still import smaller tables using Import Data.
Reference: https://radacad.com/composite-model-directquery-and-import-data-combinedevolution-begins-in-power-bi
https://powerbi.microsoft.com/en-us/blog/five-new-power-bi-premium-capacity-settings-isavailable-on-the-portal-preloaded-with-default-values-admin-can-review-and-override-the-defaultswith-their-preference-to-better-fence-their-capacity/

QUESTION 54
You are using a Python notebook in an Apache Spark pool in Azure Synapse Analytics.
You need to present the data distribution statistics from a DataFrame in a tabular view.
Which method should you invoke on the DataFrame?

A. freqIcems

B. corr

C. summary

D. rollup

Correct Answer: B
Section:
Explanation:
pandas.DataFrame.corr computes pairwise correlation of columns, excluding NA/null values.
Incorrect:
* freqItems
pyspark.sql.DataFrame.freqItems
Finding frequent items for columns, possibly with false positives. Using the frequent element count algorithm described in https://doi.org/10.1145/762471.762473, proposed by Karp, Schenker, and Papadimitriou.'
* summary is used for index.
* There is no panda method for rollup. Rollup would not be correct anyway.
Reference: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html

QUESTION 55

HOTSPOT

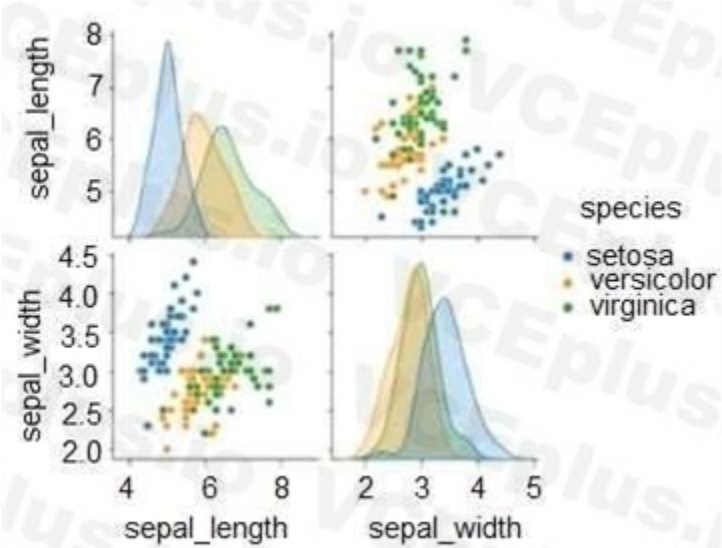You are using an Azure Synapse notebook to create a Python visual.

You run the following code cell to import a dataset named Iris.

```
iris = sns.load_dataset("iris")
iris.head()
```

A sample of the data is shown in the following table.

index	sepal_length	sepal_width	species
0	5.1	3.5	setosa
2	4.9	3	setosa
145	6.7	3	virginica
156	6.3	2.5	virginica

You need to create the visual shown in the exhibit. (Click the Exhibit tab.)

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

NOTE: Each correct selection is worth one point.

Hot Area:

```
sns. [        ▼] (iris, hue= ' [        ▼] ', height=2.5)
      boxplot                  sepal_length
      implot                   sepal_width
      pairplot                 species
      replot

plt.show()
```

Answer Area:

```
sns.  [pairplot ▼] (iris, hue= ' [sepal_width ▼] ', height=2.5)
      boxplot                     sepal_length
      implot                      sepal_width
      pairplot                    species
      replot

plt.show()
```

Section:

Explanation:

Box 1: pairplot A pairs plot allows us to see both distribution of single variables and relationships between two variables. Pair plots are a great method to identify trends for follow-up analysis and, fortunately, are easily implemented in Python!

Example, let's plot data using pairplot:

From the picture below, we can observe the variations in each plot. The plots are in matrix format where the row name represents x axis and column name represents the y axis. The main-diagonal subplots are the univariate histograms (distributions) for each attribute.

Box 2: sepal_width sepal_width is displayed with a height of 2.5 (between 2.0 and 4.5).
Reference: https://medium.com/analytics-vidhya/pairplot-visualization-16325cd725e6

QUESTION 56
HOTSPOT
You use Vertipaq Analyzer to analyze a model.
The Relationships tab contains the results shown in the following exhibit.

Row Labels	Relationship Type	Max From Cardinality	Max to Cardinality	1:M Ratio %	Missing Keys	Invalid Rows	Relationships Size	Bid. Filters MMR
'Date '[Date] ==<–1 'LocalDateTable_39c22ddb-27f3-4e6c-8a44-a3380850fcb4'[Date	M:1	84	2,557	3044.05%	0	0	4,056	
Fact	M:1	90	327	0.69%	22		184	
'Fact'[BU Key]==<–1 'BU'[BU Key]	M:1	26	164	0.34%	0	0	32	
'Fact' [Customer Key] ==<–1 'Customer' [Customer]	M:1	90	327	0.69%	21	1,804	112	
'Fact' [Product Key] ==<–1 'Product' [Product Key]	M:1	7	6	0.01%	1	6,577	8	
'Fact' [Scenario Key] ==<–1 'Scenario' [Scenario Key]	M:1	2	2	0.00%	0	0	8	
'Fact' [YearPeriod] ==<–1 'Date' [YearPeriod]	M:1	16	84	0.18%	0	0	24	
Grand Total	M:1	90	2,557	3044.05%	27		4,320	

IsRowNumber — FALSE
Cardinality (Filter) — All

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

Hot Area:

The [answer choice] table is missing records needed by the Fact table.

- BU Key
- Customer
- Date
- Scenario

There are [answer choice] blank values created by missing dimensional relationships.

- 22
- 1,804
- 6,577
- 8,381

Answer Area:

The **[answer choice]** table is missing records needed by the Fact table.

▼
BU Key
Customer
Date
Scenario

There are **[answer choice]** blank values created by missing dimensional relationships.

▼
22
1,804
6,577
8,381

Section:
Explanation:
Box 1: Customer
There are 1804 invalid rows (records) in the Customer table.
Box 2: 1804 There is a very good video explaining the concept on
https://www.youtube.com/watch?v=XvZyo9Kgyio&list=PLU6II7MW- aiIREc1XdCYviQEj4Ui2pNDS&index=3

QUESTION 57
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Power Query query in a Power BI report. The query maps source database columns to the input columns of a time series machine learning model.
You discover that the model fails to process a source time field.
You need to ensure that the model processes the field.
Solution: You add a Power Query step to convert the time column into an integer column that represents the number of seconds.
Does this meet the requirement?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 58
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question se might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Power Query query in a Power BI report. The query maps source database columns to the input columns of a time series machine learning model.
You discover that the model fails to process a source time field.
You need to ensure that the model processes the field.
Solution: You add a Power Query step to convert the time column into an integer column that represents the number of minutes.
Does this meet the requirement?

A. Yes
B. No

Correct Answer: B
Section: