

# Microsoft

## Exam Questions DP-500

Designing and Implementing Enterprise-Scale Analytics Solutions Using Microsoft Azure and Microsoft Power BI



**NEW QUESTION 1**

- (Exam Topic 3)

You open a Power BI Desktop report that contains an imported data model and a single report page.

You open Performance analyzer, start recording, and refresh the visuals on the page. The recording produces the results shown in the following exhibit

Name	Duration (ms)
Recording started (2/3/2022 10:04:04 PM)	-
Refreshed visual	-
Shape	130
Visual display	48
Other	82
Copy query	-
Actual/Forecast Billable Hours	1649
DAX query	85
Visual display	47
Other	1517
Copy query	-
Actual/Forecast Hours By Type	2083
DAX query	89
Visual display	39
Other	1955
Copy query	-
Projected Utilization %	2311
DAX query	119
Visual display	53
Other	2140
Copy query	-
Actual/Forecast Billable Hrs YTD	2458
DAX query	151

What can you identify from the results?

- A. The Actual/Forecast Hours by Type visual takes a long time to render on the report page when the data is cross-filtered.
- B. The Actual/Forecast Billable Hrs YTD visual displays the most data.
- C. Unoptimized DAX queries cause the page to load slowly.
- D. When all the visuals refresh simultaneously, the visuals spend most of the time waiting on other processes to finish.

**Answer: D**

**Explanation:**

Most time is spent in the category Other - time required by the visual for preparing queries, waiting for other visuals to complete, or performing other background processing.

Note: Each visual's log information includes the time spent (duration) to complete the following categories of tasks:

DAX query - if a DAX query was required, this is the time between the visual sending the query, and for Analysis Services to return the results.

Visual display - time required for the visual to draw on the screen, including time required to retrieve any web images or geocoding.

Other - time required by the visual for preparing queries, waiting for other visuals to complete, or performing other background processing.

Reference: <https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-performance-analyzer>

**NEW QUESTION 2**

- (Exam Topic 3)

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

Vertipaq Analyzer Metrics						
Tables	Columns	Relationships	Partitions	Summary		
Name	Cardinality	Table Size	Col Size	Data	Dictionary	Hier Size
<b>Plan</b>	<b>627,876</b>	<b>22,823,464</b>	<b>21,147,552</b>	<b>6,697,272</b>	<b>10,293,184</b>	<b>4,157,096</b>
Forecast Amount	101,606	22,823,464	7,400,920	1,475,640	5,112,384	812,896
Budget Amount	101,596	22,823,464	7,400,024	1,475,640	5,111,568	812,816
Row ID	627,876	22,823,464	4,185,992	1,674,344	120	2,511,528
ProductKey	628	22,823,464	842,296	818,016	19,208	5,072
<b>Sales</b>	<b>858,789</b>	<b>20,968,092</b>	<b>18,674,660</b>	<b>12,182,384</b>	<b>2,587,004</b>	<b>3,905,272</b>
Row ID	858,789	20,968,092	5,725,408	2,290,112	120	3,435,176
SalesAmount	36,554	20,968,092	2,960,560	1,245,904	1,422,176	292,480
TotalCost	9,711	20,968,092	1,924,272	1,238,488	608,056	77,728
Sales ID	2,000	20,968,092	1,431,192	1,374,064	41,080	16,048
Date	1,095	20,968,092	1,428,968	1,373,856	46,312	8,800

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

Table / Relationship	Size	Max From Cardinality	Max To Cardinality	1:M Ratio %	Missing Keys
<b>Plan</b>	<b>1,675,912</b>	<b>627,876</b>	<b>858,789</b>	<b>136.78%</b>	<b>7</b>
Plan[ProductKey] ∞--1 Product[ProductKey]	848	628	629	0.10%	0
Plan[StoreKey] ∞--1 Store[Store Key]	360	306	299	0.05%	7
Plan[GeographyKey] ∞--1 Geography[GeographyKey]	312	263	263	0.04%	0
Plan[DateKey] ∞--1 Month & Year Distinct[Date]	32	36	36	0.01%	0
<b>Sales</b>	<b>2,293,432</b>	<b>858,789</b>	<b>1,095</b>	<b>0.13%</b>	<b>858,793</b>
Sales[Date] ∞--1 Calendar[Date]	1,760	1,095	1,095	0.13%	0
Sales[GeographyKey] ∞--1 Geography[GeographyKey]	312	263	263	0.03%	0
Sales[PromotionKey] ∞--1 Promotion[Promotion Key]	24	28	28	0.00%	0
Sales[channelKey] ∞--1 Channel[ChannelKey]	8	4	4	0.00%	0
Sales[Row ID] ∞--1 Plan Header Details[Row ID]	0	858,789	3	0.00%	858,786

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]

**Answer: B**

**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-referential-integrity/>

**NEW QUESTION 3**

- (Exam Topic 3)

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.

**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**

⬅  
➡

⬆  
⬇

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Graphical user interface, text, application, chat or text message Description automatically generated

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>

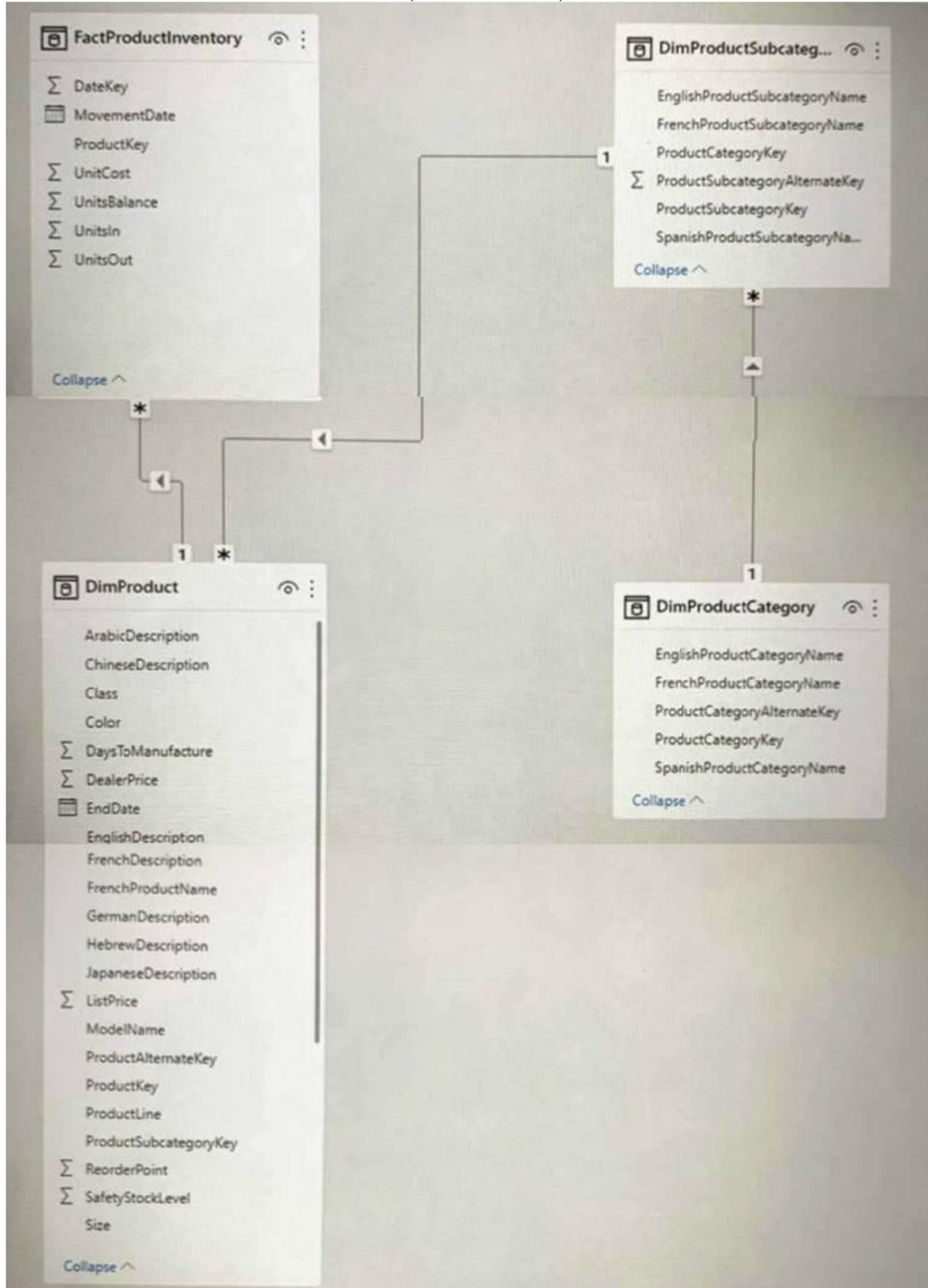
**NEW QUESTION 4**

- (Exam Topic 3)

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 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 normalizing the data model. Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

**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/>

**NEW QUESTION 5**

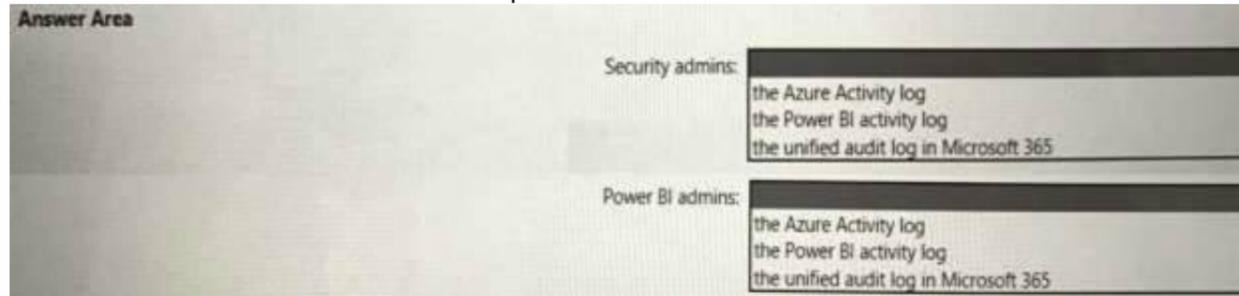
- (Exam Topic 3)

You need to recommend an automated solution to monitor Power BI user activity. The solution must meet the following requirements:

- Security admins must identify when users export reports from Power BI within five days of a new sensitivity label being applied to the artifacts in Power BI.
- Power BI admins must identify updates or changes to the Power BI capacity.
- The principle of least privilege must be used.

Which log should you include in the recommendation for each group? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: the unified audit log in Microsoft 365

Security admins must identify when users export reports from Power BI within five days of a new sensitivity label being applied to the artifacts in Power BI. Use the audit log

If your task is to track user activities across Power BI and Microsoft 365, you work with auditing in Microsoft 365 compliance or use PowerShell. Auditing relies on functionality in Exchange Online, which automatically supports Power BI.

You can filter the audit data by date range, user, dashboard, report, dataset, and activity type. You can also download the activities in a csv (comma-separated value) file to analyze offline.

Box 2: Power BI activity log

Power BI admins must identify updates or changes to the Power BI capacity. Use the activity log

Power BI administrators can analyze usage for all Power BI resources at the tenant level by using custom reports that are based on the Power BI activity log.

Reference: <https://docs.microsoft.com/en-us/power-bi/admin/service-admin-auditing>

**NEW QUESTION 6**

- (Exam Topic 3)

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-8-encoded 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.

name	system_type_name	max_length
businessName	varchar(8000)	8000
surveyName	varchar(8000)	8000
participants	int	4

You need to recommend changes to the queries to reduce I/O reads and tempdb usage.

Solution: You recommend using openrowset with to explicitly define the collation for businessName and surveyName as Latin1\_Generai\_100\_BIN2\_UTF8.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:**

Query Parquet files using serverless SQL pool in Azure Synapse Analytics. Important

Ensure you are using a UTF-8 database collation (for example Latin1\_General\_100\_BIN2\_UTF8) because string values in PARQUET files are encoded using UTF-8 encoding. A mismatch between the text encoding in the PARQUET file and the collation may cause unexpected conversion errors. You can easily change the default collation of the current database using the following T-SQL statement: alter database current collate Latin1\_General\_100\_BIN2\_UTF8'.

Note: If you use the Latin1\_General\_100\_BIN2\_UTF8 collation you will get an additional performance boost compared to the other collations. The Latin1\_General\_100\_BIN2\_UTF8 collation is compatible with parquet string sorting rules. The SQL pool is able to eliminate some parts of the parquet files that will not contain data needed in the queries (file/column-segment pruning). If you use other collations, all data from the parquet files will be loaded into Synapse SQL and the filtering is happening within the SQL process. The Latin1\_General\_100\_BIN2\_UTF8 collation has additional performance optimization that works only for parquet and CosmosDB. The downside is that you lose fine-grained comparison rules like case insensitivity.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-parquet-files>

**NEW QUESTION 7**

- (Exam Topic 3)

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

**Answer:** A

**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.

**NEW QUESTION 8**

- (Exam Topic 3)

You have two Power BI reports named Report1 and Report2.

Report1 connects to a shared dataset named Dataset1.

Report2 connects to a local dataset that has the same structure as Dataset1. Report2 contains several calculated tables and parameters.

You need to prepare Report2 to use Dataset1.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Remove the data source permissions.
- B. Delete all the Power Query Editor objects.
- C. Modify the source of each query.
- D. Update all the parameter values.
- E. Delete all the calculated tables.

**Answer:** CD

**Explanation:**

C: Power BI Desktop also comes with Power Query Editor. Use Power Query Editor to connect to one or many data sources, shape and transform the data to meet your needs, then load that model into Power BI Desktop.

D: Common uses for parameters

Here are some of the most common ways to use parameters. Control paginated report data

\* Filter paginated report data at the data source by writing dataset queries that contain variables.

\* Etc.

Reference: <https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-query-overview> <https://docs.microsoft.com/en-us/learn/modules/dax-power-bi-add-calculated-tables/1-introduction>

**NEW QUESTION 9**

- (Exam Topic 3)

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.

Azure SQL database:

SharePoint folder:

- A. Mastered
- B. Not Mastered

**Answer:** A

**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-cc3ede4d-359e-4b28-bc72-9bee7900>

**NEW QUESTION 10**

- (Exam Topic 3)

You are attempting to configure certification for a Power BI dataset and discover that the certification setting for the dataset is unavailable.

What are two possible causes of the issue? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. The workspace is in shared capacity.
- B. You have insufficient permissions.
- C. Dataset certification is disabled for the Power BI tenant.
- D. The sensitivity level for the dataset is set to Highly Confidential.
- E. Row-level security (RLS) is missing from the dataset.

**Answer:** BC

**Explanation:**

Reference: <https://docs.microsoft.com/en-us/power-bi/admin/service-admin-setup-certification> <https://docs.microsoft.com/en-us/power-bi/collaborate-share/service-endorse-content>

**NEW QUESTION 10**

- (Exam Topic 3)

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

**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-serverless-sql-pool> <https://stackoverflow.com/questions/65320949/parquet-vs-delta-format-in-azure-data-lake-gen-2-store>

**NEW QUESTION 15**

- (Exam Topic 3)

You need to save Power BI dataflows in an Azure Storage account.

Which two prerequisites are required to support the configuration? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. The storage account must be protected by using an Azure Firewall.

- B. The connection must be created by a user that is assigned the Storage Blob Data Owner role.
- C. The storage account must have hierarchical namespace enabled.
- D. Dataflows must exist already for any directly connected Power BI workspaces.
- E. The storage account must be created in a separate Azure region from the Power BI tenant and workspaces.

**Answer:** BC

**Explanation:**

Reference:

<https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-azure-data-lake-storage-integra>

**NEW QUESTION 20**

- (Exam Topic 3)

You are using DAX Studio to analyze a slow-running report query. You need to identify inefficient join operations in the query. What should you review?

- A. the query statistics
- B. the query plan
- C. the query history
- D. the server timings

**Answer:** B

**Explanation:**

Open DAX Studio.

Paste the query there, enable Query Plan display and Server Timings, run your query (with clear cache), and then study the query plan for large row counts. Once the culprit is identified you can decide how to rewrite your DAX to make that part faster.

Reference: <https://community.powerbi.com/t5/Power-Query/DAX-Query-taking-longer-time/td-p/1171961> <https://www.sqlbi.com/wp-content/uploads/DAX-Query-Plans.pdf>

**NEW QUESTION 25**

- (Exam Topic 3)

You plan to modify a Power BI dataset.

You open the Impact analysis panel for the dataset and select Notify contacts. Which contacts will be notified when you use the Notify contacts feature?

- A. any users that accessed a report that uses the dataset within the last 30 days
- B. the workspace admins of any workspace that uses the dataset
- C. the Power BI admins
- D. all the workspace members of any workspace that uses the dataset

**Answer:** D

**Explanation:**

Notify contacts

If you've made a change to a dataset or are thinking about making a change, you might want to contact the relevant users to tell them about it. When you notify contacts, an email is sent to the contact lists of all the impacted workspaces. Your name appears on the email so the contacts can find you and reply back in a new email thread.

Reference: <https://docs.microsoft.com/en-us/power-bi/collaborate-share/service-dataset-impact-analysis>

**NEW QUESTION 27**

- (Exam Topic 3)

You are creating a Power BI single-page report.

Some users will navigate the report by using a keyboard, and some users will navigate the report by using a screen reader.

You need to ensure that the users can consume content on a report page in a logical order. What should you configure on the report page?

- A. the bookmark order
- B. the X position
- C. the layer order
- D. the tab order

**Answer:** D

**Explanation:**

Tab order is the order in which users interact with the items on a page using the keyboard. Generally, we want tab order to be predictable and to closely match the visual order on the page (unless there is a good reason to deviate).

Note: If you are using the keyboard to navigate in a Power BI report, the order in which you arrive at visuals will not follow your vision unless you set the new tab order property. If you have low or no vision, this becomes an even bigger issue because you may not be able to see that you are navigating visuals out of visual order because the screen reader just reads whatever comes next.

Reference: <https://datasavvy.me/2018/12/26/tab-order-enhances-power-bi-report-accessibility/>

**NEW QUESTION 29**

- (Exam Topic 3)

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 BI 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 BI Desktop, you group the measures in a display folder.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

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. Reference: <https://docs.microsoft.com/en-us/power-bi/developer/visuals/capabilities>

**NEW QUESTION 33**

- (Exam Topic 3)

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)

**Answer:** D

**Explanation:**

To schedule refresh of your R visuals or dataset, enable scheduled refresh and install an on-premises data gateway (personal mode) on the computer containing the workbook and R.

Reference: <https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-r-in-query-editor>

**NEW QUESTION 38**

- (Exam Topic 3)

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 BI 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 BI Desktop, you create a hierarchy.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

**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/>

**NEW QUESTION 41**

- (Exam Topic 3)

You have an Azure Synapse Analytics serverless SQL pool and an Azure Data Lake Storage Gen2 account. You need to query all the files in the 'csv/taxi/' folder and all its subfolders. All the files are in CSV format and have a header row.

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

NOTE: Each correct selection is worth one point.

`SELECT*`

`FROM OPENROWSET (`

	▼
BULK 'csv/taxi',	
BULK 'csv/taxi/**',	
BULK 'csv/taxi/*.csv',	
BULK 'csv/taxi/',	

`DATA_SOURCE = 'datalake',`  
`FORMAT = 'CSV', PARSER_VERSION = '2.0',`

	▼
FIRSTROW = 0	
FIRSTROW = 1	
FIRSTROW = -1	
FIRSTROW = 2	

`)`

`WITH (`

`pickup_datetime DATETIME2,`  
`passenger_count INT,`  
`trip_distance FLOAT,`  
`total_amount FLOAT`

`) AS nyc;`

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: BULK 'csv/taxi\*.CSV',

\*.CSV to get all the CSV files. Box 2: FIRSTROW=2

As there is a header we should read from the second line. Note: FIRSTROW = 'first\_row'

Specifies the number of the first row to load. The default is 1 and indicates the first row in the specified data file. The row numbers are determined by counting the row terminators. FIRSTROW is 1-based.

Incorrect:

Not FIRSTROW=1. FIRSTROW=1 is used when there is no header.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-openrowset>

**NEW QUESTION 43**

- (Exam Topic 3)

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.



- A. Mastered
- B. Not Mastered

**Answer:** A

**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-apache-spark-pool-portal>

<https://www.royalcyber.com/blog/data-services/dedicated-sql-pool-vs-serverless-sql/>

**NEW QUESTION 44**

- (Exam Topic 3)

You have a Power BI tenant that contains 10 workspaces.

You need to create dataflows in three of the workspaces. The solution must ensure that data engineers can access the resulting data by using Azure Data Factory.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point

- A. Associate the Power BI tenant to an Azure Data Lake Storage account.
- B. Add the managed identity for Data Factory as a member of the workspaces.
- C. Create and save the dataflows to an Azure Data Lake Storage account.
- D. Create and save the dataflows to the internal storage of Power BI

**Answer:** AC

**Explanation:**

Data used with Power BI is stored in internal storage provided by Power BI by default. With the integration of dataflows and Azure Data Lake Storage Gen 2 (ADLS Gen2), you can store your dataflows in your organization's Azure Data Lake Storage Gen2 account. This essentially allows you to "bring your own storage" to Power BI dataflows, and establish a connection at the tenant or workspace level.

Reference:

<https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-azure-data-lake-storage-integra>

#### NEW QUESTION 49

- (Exam Topic 3)

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

**Answer:** B

#### 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-data-visualization> <https://seaborn.pydata.org/>

#### NEW QUESTION 54

- (Exam Topic 3)

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 BI 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 Tabular Editor, you create a calculation group.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

#### Explanation:

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. Reference: <https://docs.microsoft.com/en-us/power-bi/developer/visuals/capabilities>

#### NEW QUESTION 58

- (Exam Topic 3)

You are using GitHub as a source control solution for an Azure Synapse Studio workspace. You need to modify the source control solution to use an Azure DevOps Git repository. What should you do first?

- A. Disconnect from the GitHub repository.
- B. Create a new pull request.
- C. Change the workspace to live mode.
- D. Change the active branch.

**Answer:** A

#### Explanation:

By default, Synapse Studio authors directly against the Synapse service. If you have a need for collaboration using Git for source control, Synapse Studio allows you to associate your workspace with a Git repository, Azure DevOps, or GitHub.

Prerequisites

Users must have the Azure Contributor (Azure RBAC) or higher role on the Synapse workspace to configure, edit settings and disconnect a Git repository with Synapse.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/cicd/source-control>

#### NEW QUESTION 60

- (Exam Topic 3)

You have a Power BI dataset that has only the necessary fields visible for report development.

You need to ensure that end users see only 25 specific fields that they can use to personalize visuals. What should you do?

- A. From Tabular Editor, create a new role.
- B. Hide all the fields in the dataset.
- C. Configure object-level security (OLS).
- D. From Tabular Editor, create a new perspective.

**Answer:** B

#### NEW QUESTION 62

- (Exam Topic 3)

You are implementing a reporting solution that has the following requirements:

- Reports for external customers must support 500 concurrent requests. The data for these reports is approximately 7 GB and is stored in Azure Synapse Analytics.
  - Reports for the security team use data that must have local security rules applied at the database level to restrict access. The data being reviewed is 2 GB.
- Which storage mode provides the best response time for each group of users?

- A. DirectQuery for the external customers and import for the security team.
- B. DirectQuery for the external customers and DirectQuery for the security team.
- C. Import for the external customers and DirectQuery for the security team.
- D. Import for the external customers and import for the security team.

**Answer:** A

**Explanation:**

With DirectQuery, queries are sent back to your Azure Synapse Analytics in real time as you explore the data. Real-time queries, combined with the scale of Synapse Analytics enables users to create dynamic reports in minutes against terabytes of data.

Need import for the security team for local security rules. Reference:

<https://docs.microsoft.com/en-us/power-bi/connect-data/service-azure-sql-data-warehouse-with-direct-connect>

**NEW QUESTION 67**

- (Exam Topic 3)

You are building a Power BI 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.

**Answer Area**

- A. Mastered
- B. Not Mastered

**Answer:** A

**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>

**NEW QUESTION 72**

- (Exam Topic 3)

You have a group of data scientists who must create machine learning models and run periodic experiments on a large dataset. You need to recommend an Azure Synapse Analytics pool for the data scientists. The solution must minimize costs. Which type of pool should you recommend?

- A. a Data Explorer pool
- B. an Apache Spark pool
- C. a dedicated SQL pool
- D. a serverless SQL pool

**Answer: B**

**Explanation:**

In Azure Synapse, training machine learning models can be performed on the Apache Spark Pools with tools like PySpark/Python, Scala, or .NET. Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/machine-learning/what-is-machine-learning>

**NEW QUESTION 75**

- (Exam Topic 3)

You have a Power BI dataset that contains the following measure.

```

YTD Year-over-Year Var =
DIVIDE (
    (
        [Sales Amount]
        - CALCULATE (
            [Sales],
            SAMEPERIODLASTYEAR ( 'Calendar'[Date] ),
            'Calendar'[Flag] = "YTD"
        )
    ),
    CALCULATE (
        [Sales],
        SAMEPERIODLASTYEAR ( 'Calendar'[Date] ),
        'Calendar'[Flag] = "YTD"
    ),
    BLANK()
)
    
```

You need to improve the performance of the measure without affecting the logic or the results. What should you do?

- A. Replace both calculate functions by using a variable that contains the calculate function.
- B. Remove the alternative result of blank( ) from the divide function.
- C. Create a variable and replace the values for [sales Amount].
- D. Remove "calendar'[Flag] = "YTD" from the code.

**Answer: A**

**NEW QUESTION 76**

- (Exam Topic 3)

You are building a Power BI 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

**Answer: A**

**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>

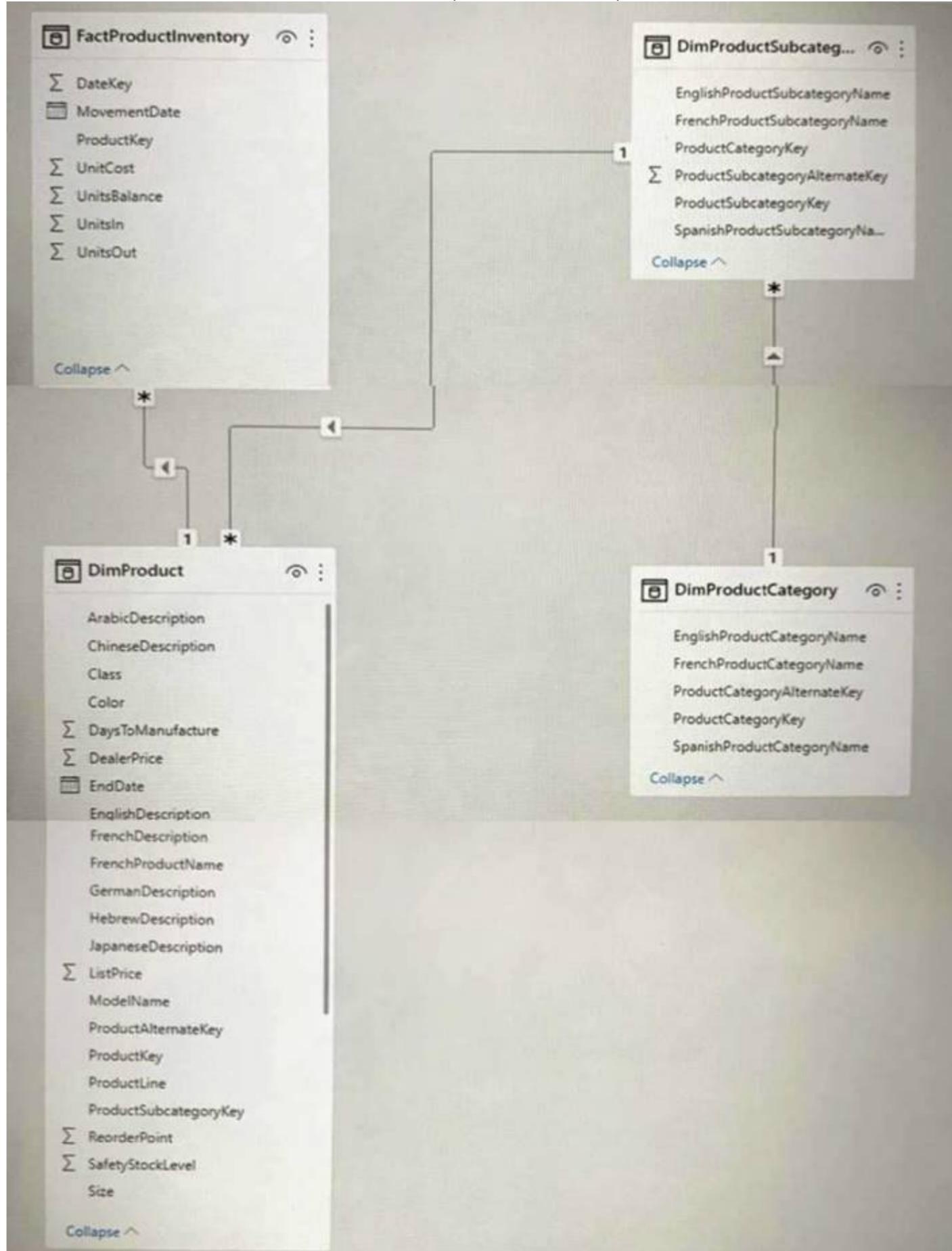
**NEW QUESTION 78**

- (Exam Topic 3)

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 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 moving all the measures to a calculation group. Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**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/>

**NEW QUESTION 80**

- (Exam Topic 3)

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

**Answer:** D

**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](https://pandas.pydata.org/docs/getting_started/intro_tutorials/06_calculate_statistics.html)

**NEW QUESTION 81**

- (Exam Topic 3)

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.

**Answer:** BD

**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-page-refresh>

**NEW QUESTION 85**

- (Exam Topic 3)

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

**Answer:** D

**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](https://matplotlib.org/3.5.0/api/_as_gen/matplotlib.pyplot.fill_between.html)

**NEW QUESTION 87**

- (Exam Topic 3)

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.

**Answer:** CE

**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/dataflows-premium-workload-configurati>

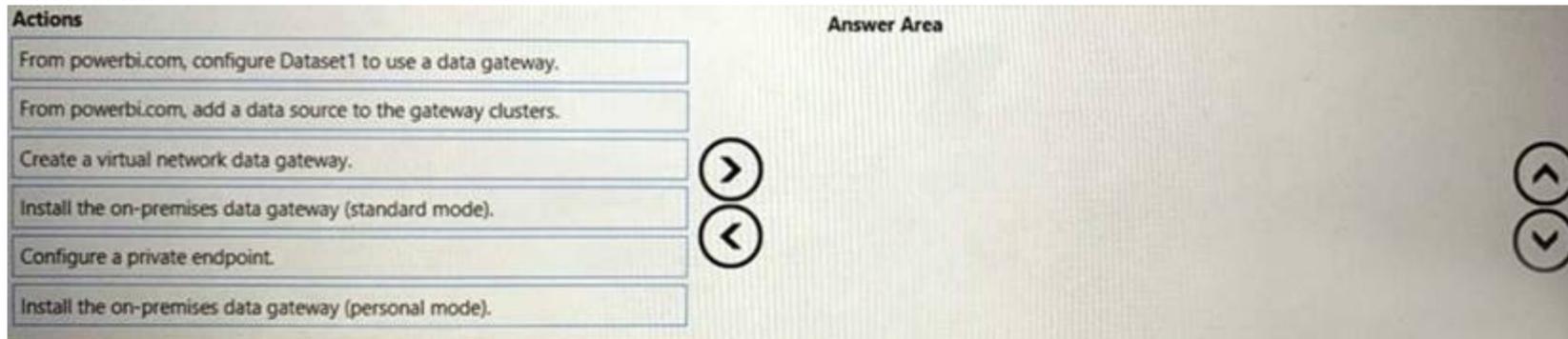
**NEW QUESTION 92**

- (Exam Topic 3)

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.



- A. Mastered
- B. Not Mastered

**Answer:** A

**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

Under Data Source Type, select SQL Server.

**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

on-premises 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-personal-mode> <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>

**NEW QUESTION 97**

- (Exam Topic 3)

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 BI 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 DAX Studio, you write a query that uses grouping sets.

Does this meet the goal?

- A. Yes
- B. No

**Answer: A**

**Explanation:**

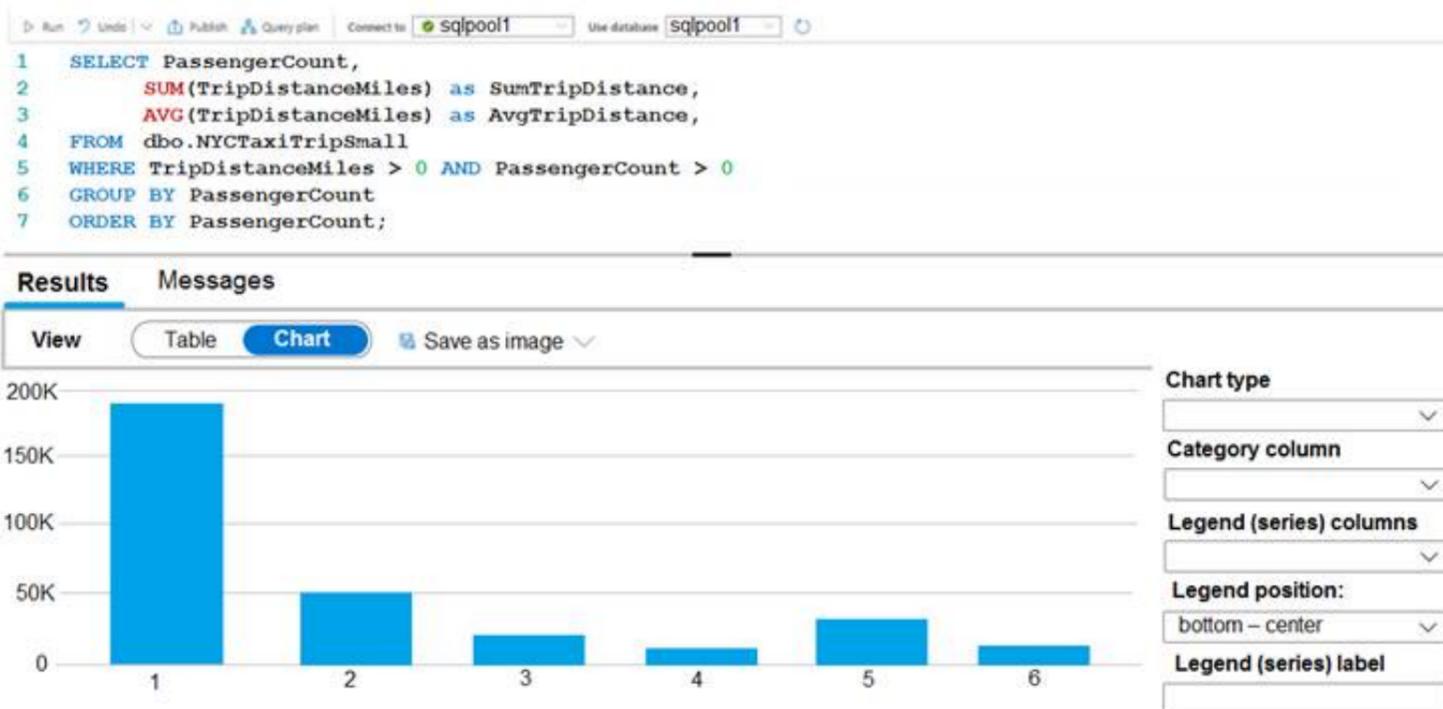
A grouping is a set of discrete values that are used to group measure fields. Reference: <https://docs.microsoft.com/en-us/power-bi/developer/visuals/capabilities>

**NEW QUESTION 102**

- (Exam Topic 3)

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.



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

**Answer Area**

Category column:

▼  
 AvgTripDistance  
 PassengerCount  
 SumTripDistance  
 TripDistanceMiles

Legend (series) column:

▼  
 AvgTripDistance  
 PassengerCount  
 SumTripDistance  
 TripDistanceMiles

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

**Answer Area**

Category column:

	▼
AvgTripDistance	
<b>PassengerCount</b>	
SumTripDistance	
TripDistanceMiles	

Legend (series) column:

	▼
AvgTripDistance	
PassengerCount	
<b>SumTripDistance</b>	
TripDistanceMiles	

**NEW QUESTION 105**

- (Exam Topic 3)

You are optimizing a Power BI data model by using DAX Studio.

You need to capture the query events generated by a Power BI Desktop report. What should you use?

- A. the DMV list
- B. a Query Plan trace
- C. an All Queries trace
- D. a Server Timings trace

**Answer: C**

**Explanation:**

The All Queries trace in Dax Studio supports capturing the query events from all client tools (not just queries sent from DAX Studio like the Query Plan and Server Timings features do). The 'All Queries' trace is really useful when you wish to see the queries that are generated by a client tool like Power BI Desktop.

Reference: <https://daxstudio.org/documentation/features/all-queries-trace/>

**NEW QUESTION 110**

- (Exam Topic 3)

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.

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Graphical user interface, text, application Description automatically generated

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/v3.0-preview/WorkItemSnapshot? "

&"\$apply=filter( "

&"WorkItem Type eq 'Bug' "

&"and StateCategory ne 'Completed' "

&"and startswith(Area/AreaPath, '{areapath}') "

&"and DateValue ge {startdate} "

&") "

```
&"/groupby( "
&"(DateValue,State,WorkItemPriority,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>

**NEW QUESTION 115**

- (Exam Topic 3)

You have the following code in an Azure Synapse notebook.

```
import matplotlib.pyplot as plt
x1 = [2, 3, 4]
y1 = [5, 5, 5]
x2 = [1, 2, 3, 4, 5]
y2 = [2, 3, 2, 3, 4]
y3 = [6, 8, 7, 8, 7]
plt.scatter(x1, y1)
plt.scatter(x2, y2, marker='v', color='r')
plt.scatter(x2, y3, marker='^', color='m')
plt.title('Scatter Plot')
plt.show()
```

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

**Answer Area**

There will be [answer choice] rendered as the output of the code.

There will be [answer choice] used in the output.

- one scatterplot
- two scatterplots
- three scatterplots

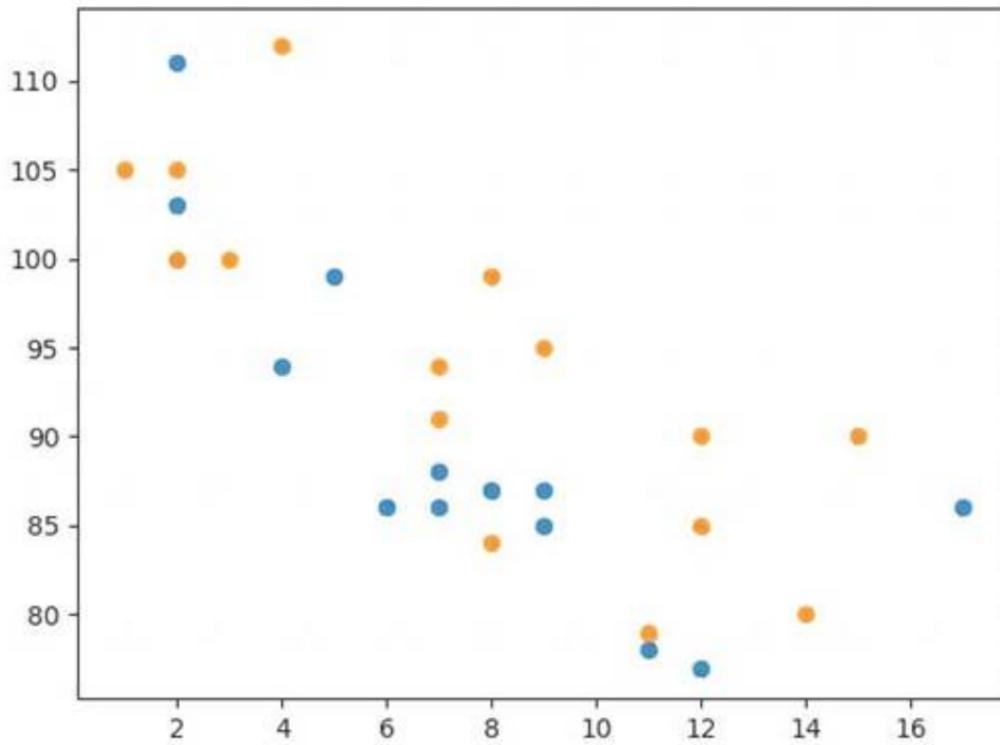
- one marker symbol
- two marker symbols
- three marker symbols

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: three scatterplots Compare Plots  
 Example, Draw two plots on the same figure: import matplotlib.pyplot as plt  
 import numpy as np  
 #day one, the age and speed of 13 cars:  
 x = np.array([5,7,8,7,2,17,2,9,4,11,12,9,6])  
 y = np.array([99,86,87,88,111,86,103,87,94,78,77,85,86])  
 plt.scatter(x, y)  
 #day two, the age and speed of 15 cars:  
 x = np.array([2,2,8,1,15,8,12,9,7,3,11,4,7,14,12])  
 y = np.array([100,105,84,105,90,99,90,95,94,100,79,112,91,80,85])  
 plt.scatter(x, y) plt.show() Result:  
 Chart, scatter chart Description automatically generated



Box 2: three marker symbols

One for each scatterplot. One default, and two defined.

Default is point.

v is triangle down.

^ is triangle up.

Reference: [https://www.w3schools.com/python/matplotlib\\_scatter.asp](https://www.w3schools.com/python/matplotlib_scatter.asp) [https://matplotlib.org/stable/api/markers\\_api.html](https://matplotlib.org/stable/api/markers_api.html)

**NEW QUESTION 117**

- (Exam Topic 3)

You have a Power BI data model.

You need to refresh the data from the source every 15 minutes. What should you do first?

- A. Enable the XMLA endpoint.
- B. Define an incremental refresh policy.
- C. Change the storage mode of the dataset.
- D. Configure a scheduled refresh.

**Answer: D**

**Explanation:**

To get to the Scheduled refresh screen:

\* 1. In the navigation pane, under Datasets, select More options (...) next to a dataset listed.

\* 2. Select Schedule refresh.

Reference: <https://docs.microsoft.com/en-us/power-bi/connect-data/refresh-scheduled-refresh>

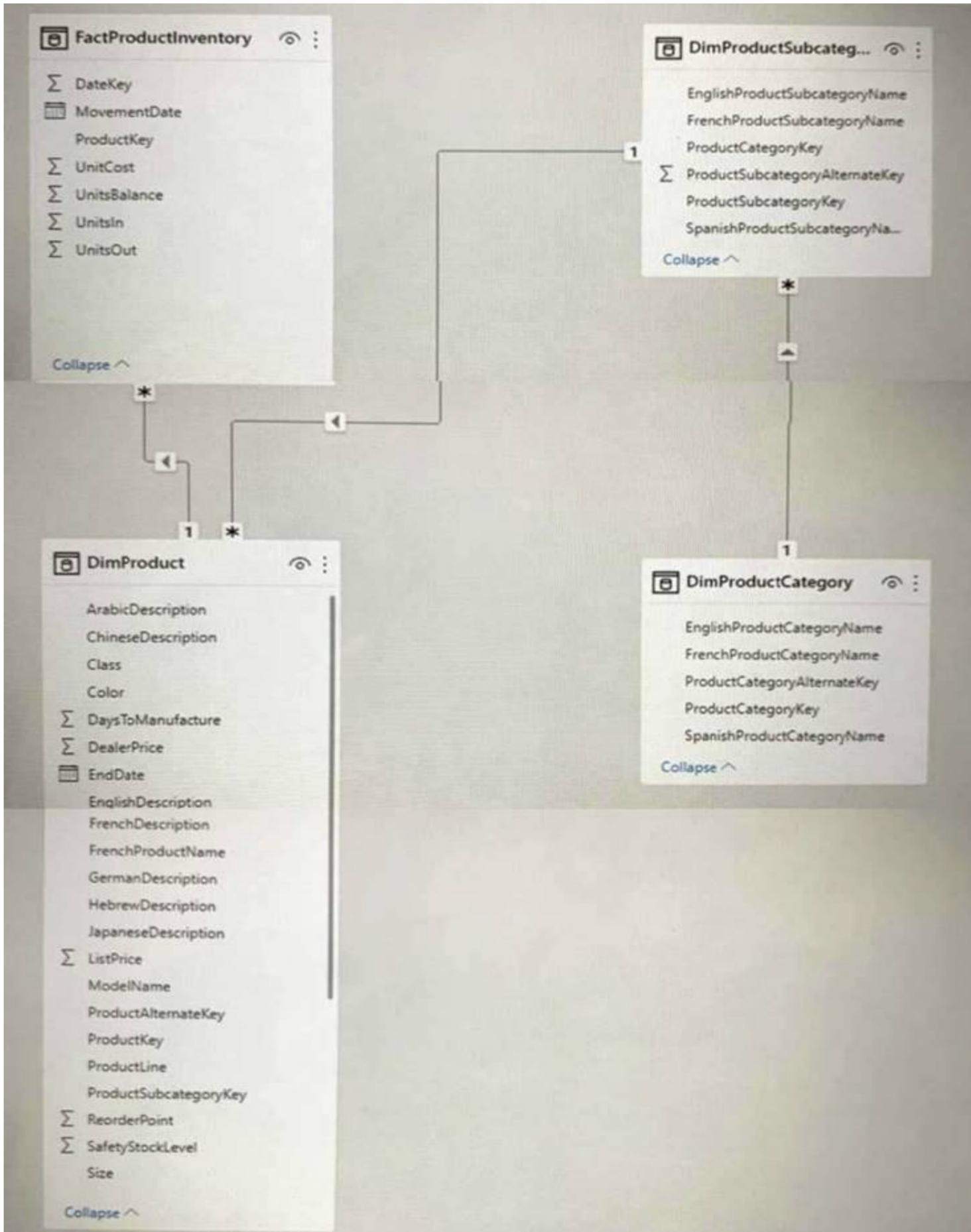
**NEW QUESTION 122**

- (Exam Topic 3)

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 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 denormalizing the data model. Does this meet the goal?

- A. Yes
- B. No

**Answer: A**

**Explanation:**

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/>

**NEW QUESTION 126**

- (Exam Topic 3)

You have a Power BI dataset that uses DirectQuery against an Azure SQL database.

Multiple reports use the dataset.

A database administrator reports that too many queries are being sent from Power BI to the database. You need to reduce the number of queries sent to the database. The solution must meet the following requirements:

- DirectQuery must continue to be used.
- Visual interactions in all the reports must remain as they are configured currently.
- Consumers of the reports must only be allowed to apply filters from the Filter pane. Which two settings should you select? Each correct answer presents part of

the solution. NOTE: Each correct selection is worth one point.

- A. Disabling cross highlighting/filtering by default
- B. Add a single Apply button to the filter pane to apply changes at once
- C. Add an Apply button to each slicer to apply changes when you're ready
- D. Add Apply buttons to all basic filters to apply changes when you're ready
- E. Ignore the Privacy Levels and potentially improve performance

**Answer:** BC

**Explanation:**

Reduce queries

Reduce the number of queries sent by Power BI using the Query reduction settings. For slicers, select the "Add an Apply button to each slicer to apply changes when you're ready" option. For filters, select "Add a single Apply button to the filter pane to apply changes at once (preview)."

Reference: <https://maqsoftware.com/insights/power-bi-best-practices>

**NEW QUESTION 128**

- (Exam Topic 3)

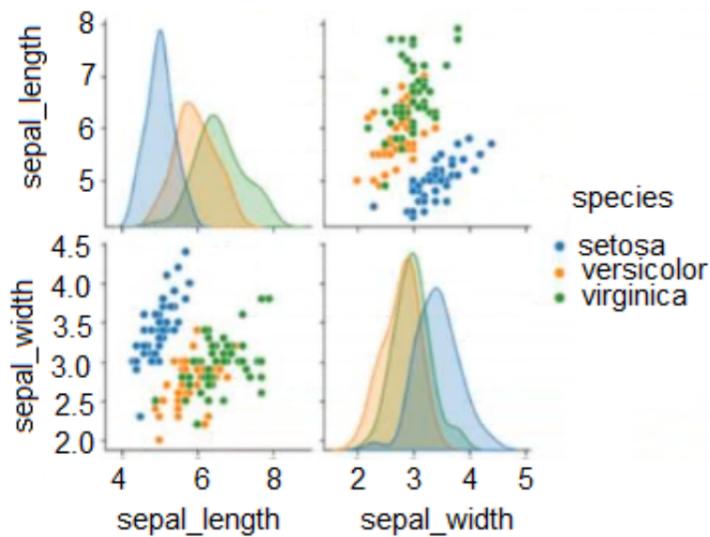
You are using an Azure Synapse notebook to create a Python visual. You run the following code cell to import a dataset named Iris.

```
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.

**Answer Area**

```
sns.  (iris, hue= '', height=2.5)

plt.show()
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

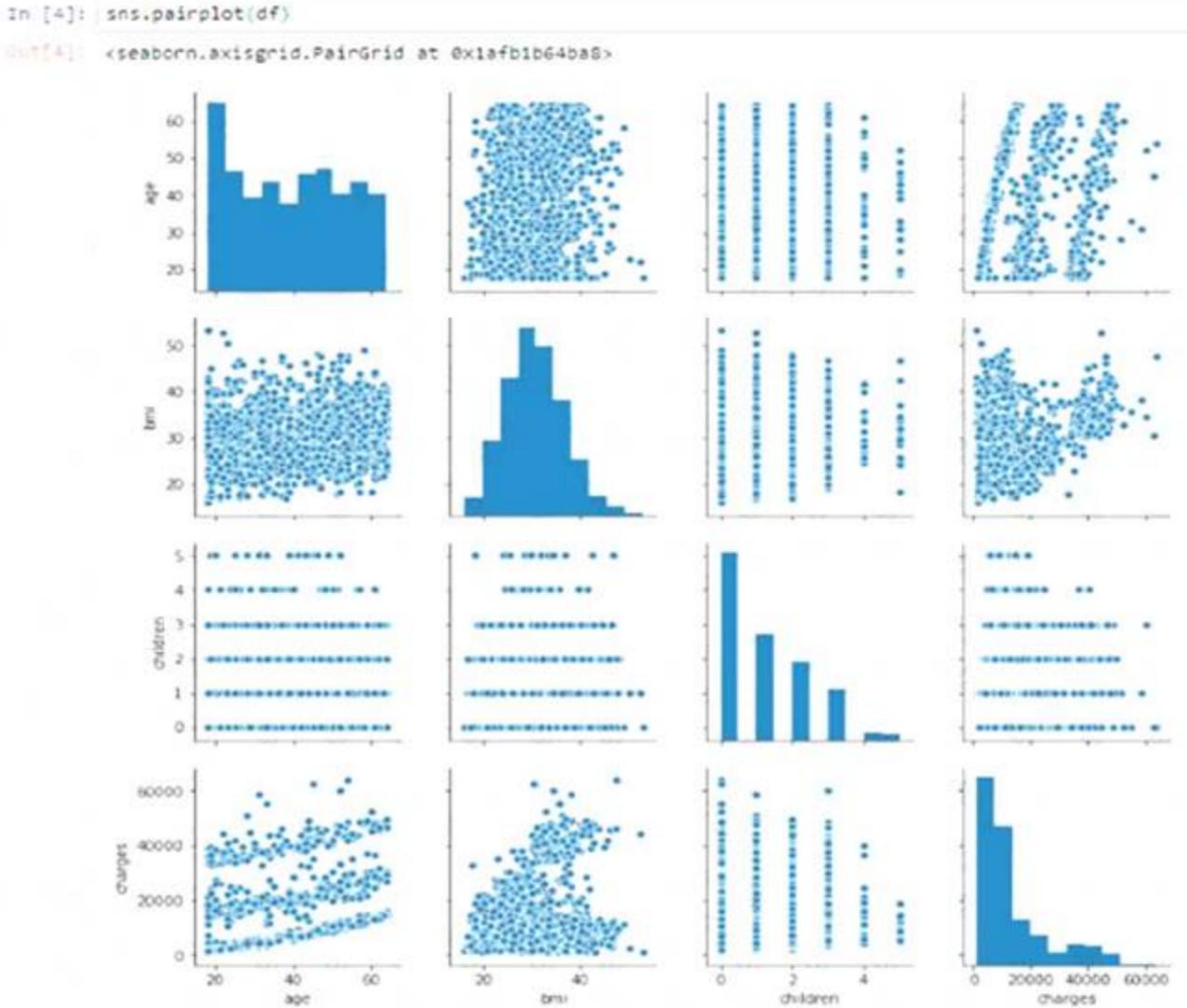
Box 1: pairplot

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.

A picture containing diagram Description automatically generated



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>

**NEW QUESTION 131**

- (Exam Topic 3)

You have an Azure Data Lake Storage Gen 2 container that stores more than 300,000 files representing hourly telemetry data. 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

Statements	Yes	No
The query uses the hierarchical namespace of the storage account.	<input type="radio"/>	<input type="radio"/>
The query uses a custom function to load file data.	<input type="radio"/>	<input type="radio"/>
Changing the source to use AzureStorage.DataLake will reduce the load time of the query.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer: A**

**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-blob-storage>

**NEW QUESTION 132**

- (Exam Topic 3)

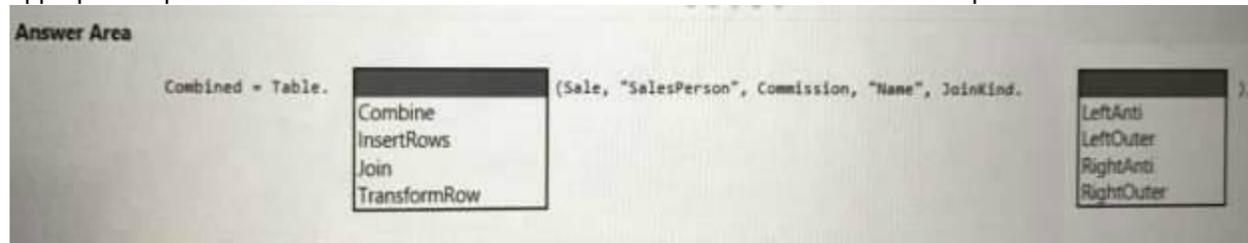
You use Advanced Editor in Power Query Editor to edit a query that references two tables named Sales and Commission. A sample of the data in the Sales table is shown in the following table.

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 area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer: A**

**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.

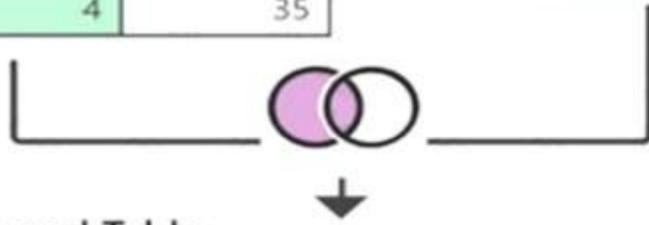
Diagram, table Description automatically generated

**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	<i>null</i>

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

**NEW QUESTION 135**

- (Exam Topic 3)

You have a Power BI 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.

**Answer: D**

**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-synapse-mapping-data-flow/>

**NEW QUESTION 139**

- (Exam Topic 2)

You need to create Power BI reports that will display data based on the customers' subscription level.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

**Actions**

- Create a perspective.
- Enable bidirectional filtering.
- Create a DAX expression.
- Create row-level security (RLS) roles.
- Add members to row-level security (RLS) roles.

**Answer Area**

>  
<

↑  
↓

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Step 1: Create row-level security (RLS) roles Create roles

Note: Provide all the customers with their own Power BI workspace to create their own reports. Each workspace will use the new dataset in the FinData

workspace.

Implement subscription levels for the customers. Each subscription level will provide access to specific rows of financial data.

Deploy prebuilt datasets to Power BI to simplify the query experience of the customers. Step 2: Create a DAX expression

Consider a model with two roles: The first role, named Workers, restricts access to all Payroll table rows by using the following rule expression:

FALSE()

Note: A rule will return no table rows when its expression evaluates to false.

Yet, a second role, named Managers, allows access to all Payroll table rows by using the following rule expression:

TRUE()

Take care: Should a report user map to both roles, they'll see all Payroll table rows. Step 3: Add members to row-level security (RLS) roles

Configure role mappings

Once [the model is] published to Power BI, you must map members to dataset roles. Reference: <https://docs.microsoft.com/en-us/power-bi/guidance/rls-guidance>

**NEW QUESTION 141**

- (Exam Topic 1)

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.

**Answer: D**

**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.

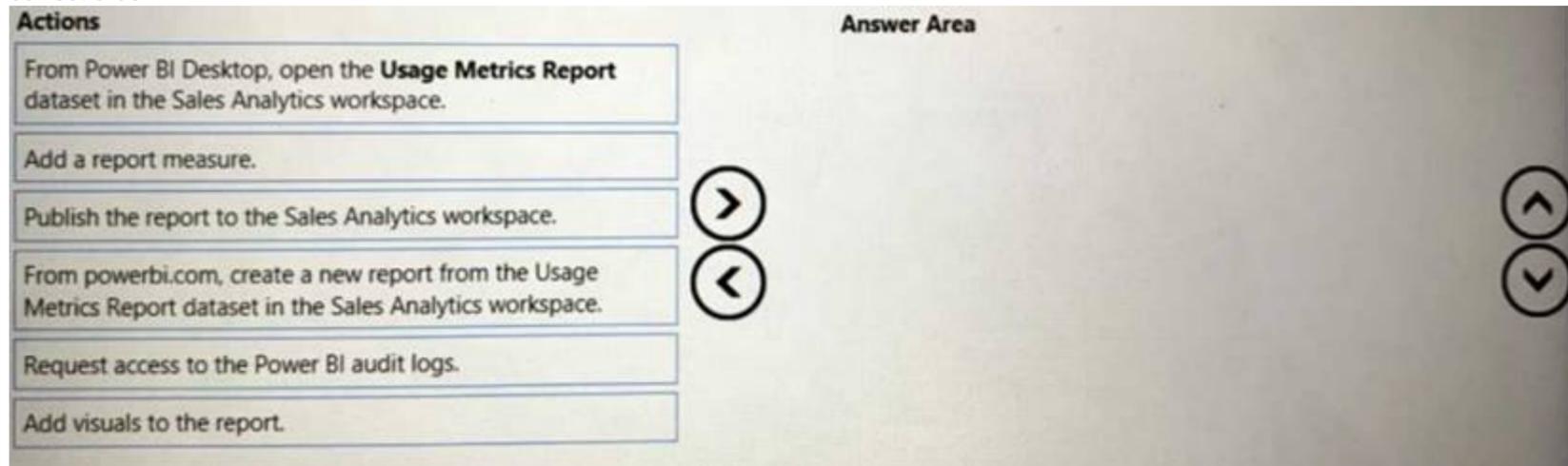
Reference: <https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-templates>

**NEW QUESTION 142**

- (Exam Topic 1)

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.



- A. Mastered
- B. Not Mastered

**Answer: A**

**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>

**NEW QUESTION 143**

- (Exam Topic 1)

You need to recommend a solution to ensure that sensitivity labels are applied. The solution must minimize administrative effort.

Which three actions should you include in the recommendation? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. From the Power BI Admin portal, set Allow users to apply sensitivity labels for Power BI content to Enabled.
- B. From the Power BI Admin portal, set Apply sensitivity labels from data sources to their data in Power BI to Enabled.
- C. In SQLD
- D. apply sensitivity labels to the columns in the Customer and CustomersWithProductScore tables.
- E. In the Power BI datasets, apply sensitivity labels to the columns in the Customer and CustomersWithProductScore tables.
- F. From the Power BI Admin portal, set Make certified content discoverable to Enabled.

**Answer:** ADE

**Explanation:**

A Synapse Analytics dedicated SQL pool is named SQLDW.

Customer contact data in SQLDW and the Power BI dataset must be labeled as Sensitive. Records must be kept of any users that use the sensitive data.

A (not B): Enable sensitivity labels

Sensitivity labels must be enabled on the tenant before they can be used in both the service and in Desktop.

To enable sensitivity labels on the tenant, go to the Power BI Admin portal, open the Tenant settings pane, and find the Information protection section.

In the Information Protection section, perform the following steps:

- > Open Allow users to apply sensitivity labels for Power BI content.
- > Enable the toggle.

D (not C): When data protection is enabled on your tenant, sensitivity labels appear in the sensitivity column in the list view of dashboards, reports, datasets, and dataflows.

E: Power BI Tenant Discovery Setting include Make certified content discoverable.

Reference: <https://docs.microsoft.com/en-us/power-bi/enterprise/service-security-enable-data-sensitivity-labels> <https://docs.microsoft.com/en-us/power-bi/enterprise/service-security-apply-data-sensitivity-labels> <https://support.nhs.net/knowledge-base/power-bi-guidance/>

**NEW QUESTION 145**

- (Exam Topic 1)

You need to populate the CustomersWithProductScore table.

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

**Answer Area**

```

DECLARE @model
SELECT model
FROM MLModel
WHERE model_name = 'PredictPurchase'
);
INSERT INTO CustomersWithProductScore (
    CustomerID
    ,CustomerEmail
    ,ProductID
    ,ProductName
    ,Score
)
SELECT d.CustomerID
    ,d.CustomerEmail
    ,d.ProductID
    ,d.ProductName
    ,p.score
FROM PREDICT(MODEL = @model, DATA =
    WITH (score FLOAT) AS p;
    
```

Box 1: BIT, FLOAT, NVARCHAR(1000), VARBINARY(max)

Box 2: AS d)

dbo.Customer  
 dbo.CustomerPurchases  
 dbo.CustomersWithProductScore  
 dbo.Product

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: FLOAT

Identify which customers should receive promotional emails based on their likelihood of purchasing promoted products.

FLOT is used in the last statement of the code: WITH (score FLOT) as p; From syntax: MODEL

The MODEL parameter is used to specify the model used for scoring or prediction. The model is specified as a variable or a literal or a scalar expression.

Box 2: dbo.CustomerWithProductScore

Identify which customers should receive promotional emails based on their likelihood of purchasing promoted products.

Only table CustomerWithProductScore has the required filed score.

From the syntax: DATA

The DATA parameter is used to specify the data used for scoring or prediction. Data is specified in the form of a table source in the query. Table source can be a table, table alias, CTE alias, view, or table-valued function.

Reference: <https://docs.microsoft.com/en-us/sql/t-sql/queries/predict-transact-sql>

**NEW QUESTION 147**

.....

## **Thank You for Trying Our Product**

### **We offer two products:**

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

2nd - Questions and Answers in PDF Format

### **DP-500 Practice Exam Features:**

- \* DP-500 Questions and Answers Updated Frequently
- \* DP-500 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-500 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-500 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DP-500 Practice Test Here](#)**