

# Databricks

## Exam Questions Databricks-Certified-Data-Analyst-Associate

Databricks Certified Data Analyst Associate Exam



### NEW QUESTION 1

An analyst writes a query that contains a query parameter. They then add an area chart visualization to the query. While adding the area chart visualization to a dashboard, the analyst chooses "Dashboard Parameter" for the query parameter associated with the area chart.

Which of the following statements is true?

- A. The area chart will use whatever is selected in the Dashboard Parameter while all or the other visualizations will remain changed regardless of their parameter use.
- B. The area chart will use whatever is selected in the Dashboard Parameter along with all of the other visualizations in the dashboard that use the same parameter.
- C. The area chart will use whatever value is chosen on the dashboard at the time the area chart is added to the dashboard.
- D. The area chart will use whatever value is input by the analyst when the visualization is added to the dashboard.
- E. The parameter cannot be changed by the user afterwards.
- F. The area chart will convert to a Dashboard Parameter.

**Answer: B**

#### Explanation:

A Dashboard Parameter is a parameter that is configured for one or more visualizations within a dashboard and appears at the top of the dashboard. The parameter values specified for a Dashboard Parameter apply to all visualizations reusing that particular Dashboard Parameter<sup>1</sup>. Therefore, if the analyst chooses "Dashboard Parameter" for the query parameter associated with the area chart, the area chart will use whatever is selected in the Dashboard Parameter along with all of the other visualizations in the dashboard that use the same parameter. This allows the user to filter the data across multiple visualizations using a single parameter widget<sup>2</sup>. References: Databricks SQL dashboards, Query parameters

### NEW QUESTION 2

A data analyst has a managed table `table_name` in database `database_name`. They would now like to remove the table from the database and all of the data files associated with the table. The rest of the tables in the database must continue to exist.

Which of the following commands can the analyst use to complete the task without producing an error?

- A. `DROP DATABASE database_name;`
- B. `DROP TABLE database_name.table_name;`
- C. `DELETE TABLE database_name.table_name;`
- D. `DELETE TABLE table_name FROM database_name;`
- E. `DROP TABLE table_name FROM database_name;`

**Answer: B**

#### Explanation:

The `DROP TABLE` command removes a table from the metastore and deletes the associated data files. The syntax for this command is `DROP TABLE [IF EXISTS] [database_name.]table_name;`. The optional `IF EXISTS` clause prevents an error if the table does not exist. The optional `database_name.` prefix specifies the database where the table resides. If not specified, the current database is used. Therefore, the correct command to remove the table `table_name` from the database `database_name` and all of the data files associated with it is `DROP TABLE database_name.table_name;`. The other commands are either invalid syntax or would produce undesired results. References: Databricks - `DROP TABLE`

### NEW QUESTION 3

Which of the following benefits of using Databricks SQL is provided by Data Explorer?

- A. It can be used to run `UPDATE` queries to update any tables in a database.
- B. It can be used to view metadata and data, as well as view/change permissions.
- C. It can be used to produce dashboards that allow data exploration.
- D. It can be used to make visualizations that can be shared with stakeholders.
- E. It can be used to connect to third party BI tools.

**Answer: B**

#### Explanation:

Data Explorer is a user interface that allows you to discover and manage data, schemas, tables, models, and permissions in Databricks SQL. You can use Data Explorer to view schema details, preview sample data, and see table and model details and properties. Administrators can view and change owners, and admins and data object owners can grant and revoke permissions<sup>1</sup>. References: Discover and manage data using Data Explorer

### NEW QUESTION 4

A data analyst creates a Databricks SQL Query where the result set has the following schema:

`region STRING number_of_customer INT`

When the analyst clicks on the "Add visualization" button on the SQL Editor page, which of the following types of visualizations will be selected by default?

- A. Violin Chart
- B. Line Chart
- C. IBar Chart
- D. Histogram
- E. There is no default
- F. The user must choose a visualization type.

**Answer: C**

#### Explanation:

According to the Databricks SQL documentation, when a data analyst clicks on the "Add visualization" button on the SQL Editor page, the default visualization type is Bar Chart. This is because the result set has two columns: one of type `STRING` and one of type `INT`. The Bar Chart visualization automatically assigns the

STRING column to the X-axis and the INT column to the Y-axis. The Bar Chart visualization is suitable for showing the distribution of a numeric variable across different categories. References: Visualization in Databricks SQL, Visualization types

#### NEW QUESTION 5

Delta Lake stores table data as a series of data files, but it also stores a lot of other information. Which of the following is stored alongside data files when using Delta Lake?

- A. None of these
- B. Table metadata, data summary visualizations, and owner account information
- C. Table metadata
- D. Data summary visualizations
- E. Owner account information

**Answer: C**

#### Explanation:

Delta Lake stores table data as a series of data files in a specified location, but it also stores table metadata in a transaction log. The table metadata includes the schema, partitioning information, table properties, and other configuration details. The table metadata is stored alongside the data files and is updated atomically with every write operation. The table metadata can be accessed using the DESCRIBE DETAIL command or the DeltaTable class in Scala, Python, or Java. The table metadata can also be enriched with custom tags or user-defined commit messages using the TBLPROPERTIES or userMetadata options. References:

- ? Enrich Delta Lake tables with custom metadata
- ? Delta Lake Table metadata - Stack Overflow
- ? Metadata - The Internals of Delta Lake

#### NEW QUESTION 6

A data analyst runs the following command:

```
INSERT INTO stakeholders.suppliers TABLE stakeholders.new_suppliers; What is the result of running this command?
```

- A. The suppliers table now contains both the data it had before the command was run and the data from the new suppliers table, and any duplicate data is deleted.
- B. The command fails because it is written incorrectly.
- C. The suppliers table now contains both the data it had before the command was run and the data from the new suppliers table, including any duplicate data.
- D. The suppliers table now contains the data from the new suppliers table, and the new suppliers table now contains the data from the suppliers table.
- E. The suppliers table now contains only the data from the new suppliers table.

**Answer: B**

#### Explanation:

The command INSERT INTO stakeholders.suppliers TABLE stakeholders.new\_suppliers is not a valid syntax for inserting data into a table in Databricks SQL.

According to the documentation<sup>12</sup>, the correct syntax for inserting data into a table is either:

```
? INSERT { OVERWRITE | INTO } [ TABLE ] table_name [ PARTITION clause ] [ ( column_name [, ...] ) | BY NAME ] query
```

```
? INSERT INTO [ TABLE ] table_name REPLACE WHERE predicate query
```

The command in the question is missing the OVERWRITE or INTO keyword, and the query part that specifies the source of the data to be inserted. The TABLE keyword is optional and can be omitted. The PARTITION clause and the column list are also optional and depend on the table schema and the data source.

Therefore, the command in the question will fail with a syntax error.

References:

- ? INSERT | Databricks on AWS
- ? INSERT - Azure Databricks - Databricks SQL | Microsoft Learn

#### NEW QUESTION 7

A data analyst has been asked to produce a visualization that shows the flow of users through a website.

Which of the following is used for visualizing this type of flow?

- A. Heatmap
- B. Choropleth
- C. Word Cloud
- D. Pivot Table
- E. Sankey

**Answer: E**

#### Explanation:

A Sankey diagram is a type of visualization that shows the flow of data between different nodes or categories. It is often used to represent the movement of users through a website, as it can show the paths they take, the sources they come from, the pages they visit, and the outcomes they achieve. A Sankey diagram consists of links and nodes, where the links represent the volume or weight of the flow, and the nodes represent the stages or steps of the flow. The width of the links is proportional to the amount of flow, and the color of the links can indicate different attributes or segments of the flow. A Sankey diagram can help identify the most common or popular user journeys, the bottlenecks or drop-offs in the flow, and the opportunities for improvement or optimization. References: The answer can be verified from Databricks documentation which provides examples and instructions on how to create Sankey diagrams using Databricks SQL Analytics and Databricks Visualizations. Reference links: Databricks SQL Analytics - Sankey Diagram, Databricks Visualizations - Sankey Diagram

#### NEW QUESTION 8

A data analyst has been asked to count the number of customers in each region and has written the following query:

```
SELECT region, count(*) AS number_of_customers
FROM customers
ORDER BY region;
```

If there is a mistake in the query, which of the following describes the mistake?

- A. The query is using count(). which will count all the customers in the customers table, no matter the region.
- B. The query is missing a GROUP BY region clause.
- C. The query is using ORDER B
- D. which is not allowed in an aggregation.
- E. There are no mistakes in the query.
- F. The query is selecting region but region should only occur in the ORDER BY clause.

**Answer: B**

**Explanation:**

In the provided SQL query, the data analyst is trying to count the number of customers in each region. However, they made a mistake by not including the `GROUP BY` clause to group the results by region. Without this clause, the query will not return counts for each distinct region but rather an error or incorrect result. References: The need for a `GROUP BY` clause in such queries can be understood from Databricks SQL documentation: Databricks SQL.

I also noticed that you uploaded an image with your question. The image shows a snippet of an SQL query written in plain text on a white background. The query is attempting to select regions and count customers from a `customers` table and order the results by region. There's no visible syntax highlighting or any other color - it's monochromatic. The query is the same as the one in your question. I'm not sure why you included the image, but maybe you wanted to show me the exact format of your query. If so, you can also use code blocks to display formatted content such as SQL queries. For example, you can write: `SELECT region, count(*) AS number_of_customers FROM customers ORDER BY region;`

This way, you can avoid uploading images and make your questions more clear and concise. I hope this helps.

**NEW QUESTION 9**

A data analyst created and is the owner of the managed table `my_table`. They now want to change ownership of the table to a single other user using Data Explorer.

Which of the following approaches can the analyst use to complete the task?

- A. Edit the Owner field in the table page by removing their own account
- B. Edit the Owner field in the table page by selecting All Users
- C. Edit the Owner field in the table page by selecting the new owner's account
- D. Edit the Owner field in the table page by selecting the Admins group
- E. Edit the Owner field in the table page by removing all access

**Answer: C**

**Explanation:**

The Owner field in the table page shows the current owner of the table and allows the owner to change it to another user or group. To change the ownership of the table, the owner can click on the Owner field and select the new owner from the drop-down list. This will transfer the ownership of the table to the selected user or group and remove the previous owner from the list of table access control entries<sup>1</sup>. The other options are incorrect because:

- ? A. Removing the owner's account from the Owner field will not change the ownership of the table, but will make the table ownerless<sup>2</sup>.
- ? B. Selecting All Users from the Owner field will not change the ownership of the table, but will grant all users access to the table<sup>3</sup>.
- ? D. Selecting the Admins group from the Owner field will not change the ownership of the table, but will grant the Admins group access to the table<sup>3</sup>.
- ? E. Removing all access from the Owner field will not change the ownership of the table, but will revoke all access to the table<sup>4</sup>. References:

- ? 1: Change table ownership
- ? 2: Ownerless tables
- ? 3: Table access control
- ? 4: Revoke access to a table

**NEW QUESTION 10**

A data analyst has been asked to use the below `tablesales_table` to get the percentage rank of products within region by the sales:

region	product	sales
WEST	A	1880.59
EAST	A	2045.99
EAST	B	4583.23
WEST	B	3391.19

The result of the query should look like this:

region	product	sales
EAST	B	0
EAST	A	1
WEST	B	0
WEST	A	1

Which of the following queries will accomplish this task?

A)

```
SELECT
    region,
    product,
    RANK() OVER (
        PARTITION BY region
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
GROUP BY region, product;
```

B)

```
SELECT
    region,
    product,
    PERCENT_RANK () OVER (
        PARTITION BY region
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
GROUP BY region, product;
```

C)

```
SELECT
    region,
    product,
    PERCENT_RANK () OVER (
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
```

D)

```
SELECT
    region,
    product,
    PERCENT RANK () OVER (
        PARTITION BY product
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
GROUP BY region, product;
```

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

**Answer:** B

**Explanation:**

The correct query to get the percentage rank of products within region by the sales is option B. This query uses the PERCENT\_RANK() window function to calculate the relative rank of each product within each region based on the sales amount. The window function is partitioned by region and ordered by sales in descending order. The result is aliased as rank and displayed along with the region and product columns. The other options are incorrect because:

- ? A. Option A uses the RANK() window function instead of the PERCENT\_RANK() function. The RANK() function returns the rank of each row within the partition, but not the percentage rank. Also, the query does not have a GROUP BY clause, which is required for aggregate functions like SUM().
- ? C. Option C uses the DENSE\_RANK() window function instead of the PERCENT\_RANK() function. The DENSE\_RANK() function returns the rank of each row within the partition, but not the percentage rank. Also, the query does not have a GROUP BY clause, which is required for aggregate functions like SUM().
- ? D. Option D uses the ROW\_NUMBER() window function instead of the PERCENT\_RANK() function. The ROW\_NUMBER() function returns the sequential number of each row within the partition, but not the percentage rank. Also, the query does not have a GROUP BY clause, which is required for aggregate functions like SUM().

References:

- ? 1: PERCENT\_RANK (Transact-SQL)
- ? 2: Window functions in Databricks SQL
- ? 3: Databricks Certified Data Analyst Associate Exam Guide

**NEW QUESTION 10**

Data professionals with varying titles use the Databricks SQL service as the primary touchpoint with the Databricks Lakehouse Platform. However, some users will use other services like Databricks Machine Learning or Databricks Data Science and Engineering. Which of the following roles uses Databricks SQL as a secondary service while primarily using one of the other services?

- A. Business analyst
- B. SQL analyst
- C. Data engineer
- D. Business intelligence analyst
- E. Data analyst

**Answer:** C

**Explanation:**

Data engineers are primarily responsible for building, managing, and optimizing data pipelines and architectures. They use Databricks Data Science and Engineering service to perform tasks such as data ingestion, transformation, quality, and governance. Data engineers may use Databricks SQL as a secondary service to query, analyze, and visualize data from the lakehouse, but this is not their main focus. References: Databricks SQL overview, Databricks Data Science and Engineering overview, Data engineering with Databricks

**NEW QUESTION 11**

A data analyst has created a user-defined function using the following line of code: CREATE FUNCTION price(spend DOUBLE, units DOUBLE) RETURNS DOUBLE RETURN spend / units;

Which of the following code blocks can be used to apply this function to the customer\_spend and customer\_units columns of the table customer\_summary to create column customer\_price?

- A. SELECT PRICE customer\_spend, customer\_units AS customer\_price FROM customer\_summary
- B. SELECT price FROM customer\_summary
- C. SELECT function(price(customer\_spend, customer\_units)) AS customer\_price FROM customer\_summary
- D. SELECT double(price(customer\_spend, customer\_units)) AS customer\_price FROM customer\_summary
- E. SELECT price(customer\_spend, customer\_units) AS customer\_price FROM customer\_summary

**Answer:** E

**Explanation:**

A user-defined function (UDF) is a function defined by a user, allowing custom logic to be reused in the user environment<sup>1</sup>. To apply a UDF to a table, the syntax is `SELECT udf_name(column_name) AS alias FROM table_name2`. Therefore, option E is the correct way to use the UDF price to create a new column customer\_price based on the existing columns customer\_spend and customer\_units from the table customer\_summary. References:

- ? What are user-defined functions (UDFs)?
- ? User-defined scalar functions - SQL V

**NEW QUESTION 12**

A data analyst is processing a complex aggregation on a table with zero null values and their query returns the following result:

group_1	group_2	sum
null	null	100
null	Y	70
null	Z	30
A	null	50
A	Y	30
A	Z	20
B	null	50
B	Y	40
B	Z	10

Which of the following queries did the analyst run to obtain the above result?

A)

```
SELECT
    group_1,
    group_2,
    count(values) AS count
FROM my_table
GROUP BY group_1, group_2 INCLUDING NULL;
```

B)

```
SELECT
    group_1,
    group_2,
    count(values) AS count
FROM my_table
GROUP BY group_1, group_2 WITH ROLLUP;
```

C)

```
SELECT
    group_1,
    group_2,
    count(values) AS count
FROM my_table
GROUP BY group_1, group 2;
```

D)

```
SELECT
    group_1,
    group_2,
    count(values) AS count
FROM my_table
GROUP BY group_1, group_2, (group_1, group_2);
```

E)

```
SELECT
    group_1,
    group_2,
    count(values) AS count
FROM my_table
GROUP BY group_1, group_2 WITH CUBE;
```

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

**Answer:** B

**Explanation:**

The result set provided shows a combination of grouping by two columns ( group\_1 and group\_2) with subtotals for each level of grouping and a grand total. This pattern is typical of a GROUP BY ... WITH ROLLUP operation in SQL, which provides subtotal rows and a grand total row in the result set.

Considering the query options:

- A) Option A: GROUP BY group\_1, group\_2 INCLUDING NULL- This is not a standard SQL clause and would not result in subtotals and a grand total.
- B) Option B: GROUP BY group\_1, group\_2 WITH ROLLUP- This would create subtotals for each unique group\_1, each combination of group\_1 and group\_2, and a grand total, which matches the result set provided.
- C) Option C: GROUP BY group\_1, group\_2- This is a simple GROUP BY and would not include subtotals or a grand total.
- D) Option D: GROUP BY group\_1, group\_2, (group\_1, group\_2)- This syntax is not standard and would likely result in an error or be interpreted as a simple GROUP BY, not providing the subtotals and grand total.
- E) Option E: GROUP BY group\_1, group\_2 WITH CUBE- The WITH CUBE operation produces subtotals for all combinations of the selected columns and a grand total, which is more than what is shown in the result set.

The correct answer is Option B, which uses WITH ROLLUP to generate the subtotals for each level of grouping as well as a grand total. This matches the result set where we have subtotals for each group\_1, each combination of group\_1 and group\_2, and the grand total where both group\_1 and group\_2 are NULL.

**NEW QUESTION 15**

A data analyst runs the following command: SELECT age, country FROM my\_table WHERE age >= 75 AND country = 'canada'; Which of the following tables represents the output of the above command?

A)

age	country
80	canada
NULL	canada
90	NULL

B)

<b>age</b>	<b>country</b>
80	<i>NULL</i>
75	<i>NULL</i>
90	<i>NULL</i>

C)

<b>id</b>	<b>age</b>	<b>country</b>
900	80	canada
901	75	canada
902	90	canada

D)

<b>age</b>	<b>country</b>
80	canada
14	canada
90	canada

E)

<b>age</b>	<b>country</b>
80	canada
75	canada
90	canada

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

**Answer:** E

**Explanation:**

The SQL query provided is designed to filter out records from `my_table` where the age is 75 or above and the country is Canada. Since I can't view the content of the links provided directly, I need to rely on the image attached to this question for context. Based on that, Option E (the image attached) represents a table with columns `age` and `country`, showing records where age is 75 or above and country is Canada. References: The answer can be inferred from understanding SQL queries and their outputs as per Databricks documentation: Databricks SQL

**NEW QUESTION 17**

Which of the following approaches can be used to ingest data directly from cloud-based object storage?

- A. Create an external table while specifying the DBFS storage path to FROM
- B. Create an external table while specifying the DBFS storage path to PATH
- C. It is not possible to directly ingest data from cloud-based object storage
- D. Create an external table while specifying the object storage path to FROM
- E. Create an external table while specifying the object storage path to LOCATION

**Answer:** E

**Explanation:**

External tables are tables that are defined in the Databricks metastore using the information stored in a cloud object storage location. External tables do not manage the data, but provide a schema and a table name to query the data. To create an external table, you can use the CREATE EXTERNAL TABLE statement and specify the object storage path to the LOCATION clause. For example, to create an external table named `ext_table` on a Parquet file stored in S3, you can use the following statement:

```
SQL
CREATE EXTERNAL TABLE ext_table ( col1 INT,
col2 STRING
)
STORED AS PARQUET
LOCATION 's3://bucket/path/file.parquet'
```

AI-generated code. Review and use carefully. More info on FAQ.  
References: External tables

**NEW QUESTION 22**

In which of the following situations will the mean value and median value of variable be meaningfully different?

- A. When the variable contains no outliers
- B. When the variable contains no missing values
- C. When the variable is of the boolean type
- D. When the variable is of the categorical type
- E. When the variable contains a lot of extreme outliers

**Answer:** E

**Explanation:**

The mean value of a variable is the average of all the values in a data set, calculated by dividing the sum of the values by the number of values. The median value of a variable is the middle value of the ordered data set, or the average of the middle two values if the data set has an even number of values. The mean value is sensitive to outliers, which are values that are very different from the rest of the data. Outliers can skew the mean value and make it less representative of the central tendency of the data. The median value is more robust to outliers, as it only depends on the middle values of the data. Therefore, when the variable contains a lot of extreme outliers, the mean value and the median value will be meaningfully different, as the mean value will be pulled towards the outliers, while the median value will remain close to the majority of the data. References: Difference Between Mean and Median in Statistics (With Example) - BYJU'S

**NEW QUESTION 25**

A data analyst is attempting to drop a table `my_table`. The analyst wants to delete all table metadata and data. They run the following command: `DROP TABLE IF EXISTS my_table;` While the object no longer appears when they run `SHOW TABLES`, the data files still exist. Which of the following describes why the data files still exist and the metadata files were deleted?

- A. The table's data was larger than 10 GB
- B. The table did not have a location
- C. The table was external
- D. The table's data was smaller than 10 GB
- E. The table was managed

**Answer:** C

**Explanation:**

An external table is a table that is defined in the metastore, but its data is stored outside of the Databricks environment, such as in S3, ADLS, or GCS. When an external table is dropped, only the metadata is deleted from the metastore, but the data files are not affected. This is different from a managed table, which is a table whose data is stored in the Databricks environment, and whose data files are deleted when the table is dropped. To delete the data files of an external table, the analyst needs to specify the PURGE option in the DROP TABLE command, or manually delete the files from the storage system. References: DROP TABLE, Drop Delta table features, Best practices for dropping a managed Delta Lake table

### NEW QUESTION 28

Which of the following describes how Databricks SQL should be used in relation to other business intelligence (BI) tools like Tableau, Power BI, and Looker?

- A. As an exact substitute with the same level of functionality
- B. As a substitute with less functionality
- C. As a complete replacement with additional functionality
- D. As a complementary tool for professional-grade presentations
- E. As a complementary tool for quick in-platform BI work

**Answer:** E

#### Explanation:

Databricks SQL is not meant to replace or substitute other BI tools, but rather to complement them by providing a fast and easy way to query, explore, and visualize data on the lakehouse using the built-in SQL editor, visualizations, and dashboards. Databricks SQL also integrates seamlessly with popular BI tools like Tableau, Power BI, and Looker, allowing analysts to use their preferred tools to access data through Databricks clusters and SQL warehouses. Databricks SQL offers low-code and no-code experiences, as well as optimized connectors and serverless compute, to enhance the productivity and performance of BI workloads on the lakehouse. References: Databricks SQL, Connecting Applications and BI Tools to Databricks SQL, Databricks integrations overview, Databricks SQL: Delivering a Production SQL Development Experience on the Lakehouse

### NEW QUESTION 30

How can a data analyst determine if query results were pulled from the cache?

- A. Go to the Query History tab and click on the text of the query
- B. The slideout shows if the results came from the cache.
- C. Go to the Alerts tab and check the Cache Status alert.
- D. Go to the Queries tab and click on Cache Status
- E. The status will be green if the results from the last run came from the cache.
- F. Go to the SQL Warehouse (formerly SQL Endpoints) tab and click on Cache
- G. The Cache file will show the contents of the cache.
- H. Go to the Data tab and click Last Query
- I. The details of the query will show if the results came from the cache.

**Answer:** A

#### Explanation:

Databricks SQL uses a query cache to store the results of queries that have been executed previously. This improves the performance and efficiency of repeated queries. To determine if a query result was pulled from the cache, you can go to the Query History tab in the Databricks SQL UI and click on the text of the query. A slideout will appear on the right side of the screen, showing the query details, including the cache status. If the result came from the cache, the cache status will show "Cached". If the result did not come from the cache, the cache status will show "Not cached". You can also see the cache hit ratio, which is the percentage of queries that were served from the cache. References: The answer can be verified from Databricks SQL documentation which provides information on how to use the query cache and how to check the cache status. Reference link: Databricks SQL - Query Cache

### NEW QUESTION 32

Which of the following layers of the medallion architecture is most commonly used by data analysts?

- A. None of these layers are used by data analysts
- B. Gold
- C. All of these layers are used equally by data analysts
- D. Silver
- E. Bronze

**Answer:** B

#### Explanation:

The gold layer of the medallion architecture contains data that is highly refined and aggregated, and powers analytics, machine learning, and production applications. Data analysts typically use the gold layer to access data that has been transformed into knowledge, rather than just information. The gold layer represents the final stage of data quality and optimization in the lakehouse. References: What is the medallion lakehouse architecture?

### NEW QUESTION 33

A data analyst has been asked to provide a list of options on how to share a dashboard with a client. It is a security requirement that the client does not gain access to any other information, resources, or artifacts in the database.

Which of the following approaches cannot be used to share the dashboard and meet the security requirement?

- A. Download the Dashboard as a PDF and share it with the client.
- B. Set a refresh schedule for the dashboard and enter the client's email address in the "Subscribers" box.
- C. Take a screenshot of the dashboard and share it with the client.
- D. Generate a Personal Access Token that is good for 1 day and share it with the client.
- E. Download a PNG file of the visualizations in the dashboard and share them with the client.

**Answer:** D

#### Explanation:

The approach that cannot be used to share the dashboard and meet the security requirement is D. Generating a Personal Access Token that is good for 1 day and sharing it with the client. This approach would give the client access to the Databricks workspace using the token owner's identity and permissions, which could expose other information, resources, or artifacts in the database. The other approaches can be used to share the dashboard and meet the security requirement because:

? A. Downloading the Dashboard as a PDF and sharing it with the client would only provide a static snapshot of the dashboard without any interactive features or access to the underlying data.

? B. Setting a refresh schedule for the dashboard and entering the client's email address in the "Subscribers" box would send the client an email with the

latest dashboard results as an attachment or a link to a secure web page<sup>3</sup>. The client would not be able to access the Databricks workspace or the dashboard itself.

? C. Taking a screenshot of the dashboard and sharing it with the client would also only provide a static snapshot of the dashboard without any interactive features or access to the underlying data<sup>4</sup>.

? E. Downloading a PNG file of the visualizations in the dashboard and sharing them with the client would also only provide a static snapshot of the visualizations without any interactive features or access to the underlying data<sup>5</sup>. References:

? 1: Personal access tokens

? 2: Download as PDF

? 3: Automatically refresh a dashboard

? 4: Take a screenshot

? 5: Download a PNG file

### NEW QUESTION 37

A data analyst is working with gold-layer tables to complete an ad-hoc project. A stakeholder has provided the analyst with an additional dataset that can be used to augment the gold-layer tables already in use.

Which of the following terms is used to describe this data augmentation?

A. Data testing

B. Ad-hoc improvements

C. Last-mile

D. Last-mile ETL

E. Data enhancement

**Answer: E**

### Explanation:

Data enhancement is the process of adding or enriching data with additional information to improve its quality, accuracy, and usefulness. Data enhancement can be used to augment existing data sources with new data sources, such as external datasets, synthetic data, or machine learning models. Data enhancement can help data analysts to gain deeper insights, discover new patterns, and solve complex problems. Data enhancement is one of the applications of generative AI, which can leverage machine learning to generate synthetic data for better models or safer data sharing<sup>1</sup>.

In the context of the question, the data analyst is working with gold-layer tables, which are curated business-level tables that are typically organized in consumption-ready project-specific databases<sup>234</sup>. The gold-layer tables are the final layer of data transformations and data quality rules in the medallion lakehouse architecture, which is a data design pattern used to logically organize data in a lakehouse<sup>2</sup>. The stakeholder has provided the analyst with an additional dataset that can be used to augment the gold-layer tables already in use. This means that the analyst can use the additional dataset to enhance the existing gold-layer tables with more information, such as new features, attributes, or metrics. This data augmentation can help the analyst to complete the ad-hoc project more effectively and efficiently.

References:

? What is the medallion lakehouse architecture? - Databricks

? Data Warehousing Modeling Techniques and Their Implementation on the Databricks Lakehouse Platform | Databricks Blog

? What is the medallion lakehouse architecture? - Azure Databricks

? What is a Medallion Architecture? - Databricks

? Synthetic Data for Better Machine Learning | Databricks Blog

### NEW QUESTION 40

.....

## **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

### **Databricks-Certified-Data-Analyst-Associate Practice Exam Features:**

- \* Databricks-Certified-Data-Analyst-Associate Questions and Answers Updated Frequently
- \* Databricks-Certified-Data-Analyst-Associate Practice Questions Verified by Expert Senior Certified Staff
- \* Databricks-Certified-Data-Analyst-Associate Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* Databricks-Certified-Data-Analyst-Associate Practice Test Questions in Multiple Choice Formats and Updates for 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The Databricks-Certified-Data-Analyst-Associate Practice Test Here](#)**