Databricks-Certified-Data-Analyst-Associate Practice Questions

Shared by Briggs on 18-08-2026

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

A data analyst needs to create an empty managed table table_name in database database_name with a specific schema. The table needs to be recreated and empty, regardless of whether or not the table already exists.

Which command can the analyst use to complete the task?

Options:

A- CREATE TABLE database_name.table_name USING (width INT, length INT, height INT);
B- CREATE OR REPLACE TABLE database_name.table_name USING (width INT, length INT, height INT);
C- CREATE OR REPLACE TABLE database_name.table_name (width INT, length INT, height INT);
D- CREATE OR REPLACE TABLE table_name FROM database_name USING (width INT, length INT, height INT);

Answer:

C

Explanation:

The correct answer is C because the task requires creating or replacing a managed table with an explicitly defined schema. In Databricks SQL, column definitions are placed directly after the table name inside parentheses. The USING clause is for specifying a data source format, not for defining columns. Option C correctly uses CREATE OR REPLACE TABLE database_name.table_name (...), which recreates the table if it already exists and creates an empty table with the specified schema.

Question 2

Question Type: MultipleChoice

A data analyst wants to create a Databricks SQL dashboard with multiple data visualizations and multiple counters. What must be completed before adding the data visualizations and counters to the dashboard?

Options:

A- All data visualizations and counters must be created using Queries.
B- A SQL warehouse (formerly known as SQL endpoint) must be turned on and selected.
C- A markdown-based tile must be added to the top of the dashboard displaying the dashboard's name.
D- The dashboard owner must also be the owner of the queries, data visualizations, and counters.
In Databricks SQL, when creating a dashboard that includes multiple data visualizations and counters, it is imperative that each visualization and counter is based on a query. The process involves the following steps:
Develop Queries:
For each desired visualization or counter, write a SQL query that retrieves the necessary data.
Create Visualizations and Counters:
After executing each query, utilize the results to create corresponding visualizations or counters. Databricks SQL offers a variety of visualization types to represent data effectively.
Assemble the Dashboard:
Add the created visualizations and counters to your dashboard, arranging them as needed to convey the desired insights.
By ensuring that all components of the dashboard are derived from queries, you maintain consistency, accuracy, and the ability to refresh data as needed. This approach also facilitates easier maintenance and updates to the dashboard elements.

Answer:

A

Question 3

Question Type: MultipleChoice

Query History provides Databricks SQL users with a lot of benefits. A data analyst has been asked to share all of these benefits with their team as part of a training exercise. One of the benefit statements the analyst provided to their team is incorrect.

Which statement about Query History is incorrect?

Options:

A- It can be used to view the query plan of queries that have run.
B- It can be used to debug queries.
C- It can be used to automate query execution on multiple warehouses (formerly endpoints).
D- It can be used to troubleshoot slow running queries.

Answer:

C

Explanation:

Query History in Databricks SQL is intended for reviewing executed queries, understanding their execution plans, and identifying performance issues or errors for debugging purposes. It allows users to analyze query duration, resources used, and potential bottlenecks. However, Query History does not provide any capability to automate the execution of queries across multiple warehouses; automation must be handled through jobs or external orchestration tools, not through the Query History feature itself.

Question 4

Question Type: MultipleChoice

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

Options:

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

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.Reference:What is the medallion lakehouse architecture?

Answer:

B

Question 5

Question Type: MultipleChoice

A data analyst has been asked to use the below table sales_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;
```

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

Options:

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

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().Reference:
1: PERCENT_RANK (Transact-SQL)
2: Window functions in Databricks SQL
3: Databricks Certified Data Analyst Associate Exam Guide

Answer:

B

Question 6

Question Type: MultipleChoice

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?

Options:

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
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 entries1. 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 ownerless2.
B .Selecting All Users from the Owner field will not change the ownership of the table, but will grant all users access to the table3.
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 table3.
E .Removing all access from the Owner field will not change the ownership of the table, but will revoke all access to the table4.Reference:

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

Answer:

C

Question 7

Question Type: MultipleChoice

A data analyst has been asked to configure an alert for a query that returns the income in the accounts_receivable table for a date range. The date range is configurable using a Date query parameter.

The Alert does not work.

Which of the following describes why the Alert does not work?

Options:

A- Alerts don't work with queries that access tables.
B- Queries that return results based on dates cannot be used with Alerts.
C- The wrong query parameter is being used. Alerts only work with Date and Time query parameters.
D- Queries that use query parameters cannot be used with Alerts.
E- The wrong query parameter is being used. Alerts only work with drogdown list query parameters, not dates.
The reason the alert is not functioning as expected is because Databricks SQL Alerts do not support query parameters. This limitation applies to all types of parameters, including date parameters.
Here's why:
Alerts require static, deterministic query results so they can compare values consistently during scheduled executions.
When a query includes parameters (e.g., a date range parameter), its results may change based on user input or the default value set in the query editor.
However, Databricks SQL Alerts will always use the default value set for the parameter at the time the alert is created. This means the alert doesn't dynamically adapt to new date ranges and will not reflect changes unless the query is manually updated.
As a result, if the business logic behind the alert depends on changing date ranges or any user input, the alert will not trigger correctly, or may never trigger at all.
Therefore, the correct explanation contradicts Option B, which is incorrect in saying that alerts cannot work with date-based queries at all. In fact, they can---as long as the query is static (i.e., without parameters).

Answer:

D

Explanation:

Databricks SQL Alerts Documentation

Databricks Knowledge: ''You cannot use alerts with queries that contain parameters.''

Question 8

Question Type: MultipleChoice

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?

Options:

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 BY. which is not allowed in an aggregation.
D- There are no mistakes in the query.
E- The query is selecting region but region should only occur in the ORDER BY clause.
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.Reference: 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.

Answer:

B

Question 9

Question Type: MultipleChoice

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?

Options:

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
Delta Lake is a storage layer that enhances data lakes with features like ACID transactions, schema enforcement, and time travel. While it stores table data as Parquet files, Delta Lake also keeps a transaction log (stored in the _delta_log directory) that contains detailed table metadata.
This metadata includes:
Table schema
Partitioning information
Data file paths
Transactional operations like inserts, updates, and deletes
Commit history and version control
This metadata is critical for supporting Delta Lake's advanced capabilities such as time travel and efficient query execution. Delta Lake does not store data summary visualizations or owner account information directly alongside the data files.

Answer:

C

Question 10

Question Type: MultipleChoice

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?

Options:

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. The parameter cannot be changed by the user afterwards.
E- The area chart will convert to a Dashboard Parameter.
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 Parameter1. 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 widget2.Reference:Databricks SQL dashboards,Query parameters

Answer:

B