DUMPSARENA

Databricks Certified Data Analyst Associate Exam

Databricks Databricks-Certified-Data-Analyst-Associate

Version Demo

Total Demo Questions: 10

Total Premium Questions: 65

Buy Premium PDF

https://dumpsarena.co

sales@dumpsarena.co

sales@dumpsarena.co
dumpsarena.co

QUESTION NO: 1

A data analyst has created a Query in Databricks SQL, and now they want to create two data visualizations from that Query and add both of those data visualizations to the same Databricks SQL Dashboard.

Which of the following steps will they need to take when creating and adding both data visualizations to the Databricks SQL Dashboard?

A. They will need to alter the Query to return two separate sets of results.

B. They will need to add two separate visualizations to the dashboard based on the same Query.

C. They will need to create two separate dashboards.

D. They will need to decide on a single data visualization to add to the dashboard.

E. They will need to copy the Query and create one data visualization per query.

ANSWER: B

Explanation:

A data analyst can create multiple visualizations from the same query in Databricks SQL by clicking the + button next to the Results tab and selecting Visualization. Each visualization can have a different type, name, and configuration. To add a visualization to a dashboard, the data analyst can click the vertical ellipsis button beneath the visualization, select + Add to Dashboard, and choose an existing or new dashboard. The data analyst can repeat this process for each visualization they want to add to the same dashboard. Reference: Visualization in Databricks SQL, Visualize queries and create a dashboard in Databricks SQL

QUESTION NO: 2

A data analyst has set up a SQL query to run every four hours on a SQL endpoint, but the SQL endpoint is taking too long to start up with each run.

Which of the following changes can the data analyst make to reduce the start-up time for the endpoint while managing costs?

A. Reduce the SQL endpoint cluster size

B. Increase the SQL endpoint cluster size

C. Turn off the Auto stop feature

D. Increase the minimum scaling value

E. Use a Serverless SQL endpoint

ANSWER: E

Explanation:

A Serverless SQL endpoint is a type of SQL endpoint that does not require a dedicated cluster to run queries. Instead, it uses a shared pool of resources that can scale up and down automatically based on the demand. This means that a Serverless SQL endpoint can start up much faster than a SQL endpoint that uses a cluster, and it can also save costs by only paying for the resources that are used. A Serverless SQL endpoint is suitable for ad-hoc queries and exploratory analysis, but it may not offer the same level of performance and isolation as a SQL endpoint that uses a cluster. Therefore, a data analyst should consider the trade-offs between speed, cost, and quality when choosing between a Serverless SQL endpoint and a SQL endpoint that uses a cluster. Reference: Databricks SQL endpoints, Serverless SQL endpoints, SQL endpoint clusters

---

QUESTION NO: 3

Where in the Databricks SQL workspace can a data analyst configure a refresh schedule for a query when the query is not attached to a dashboard or alert?

A. Data bxplorer

B. The Visualization editor

C. The Query Editor

D. The Dashboard Editor

ANSWER: C

Explanation:

In Databricks SQL, to configure a refresh schedule for a query that is not attached to a dashboard or alert, a data analyst should use the Query Editor. Within the Query Editor, there is an option to set up scheduled executions for queries. This feature enables the query to run at specified intervals, ensuring that the results are updated regularly. By scheduling queries in this manner, analysts can automate data refreshes and maintain up-to-date query results without manual intervention.

Reference: Schedule a query - Databricks Documentation

QUESTION NO: 4

Which location can be used to determine the owner of a managed table?

A. Review the Owner field in the table page using Catalog Explorer B. Review the Owner field in the database page using Data Explorer

B. Review the Owner field in the schema page using Data Explorer D. Review the Owner field in the table page using the SQL Editor

ANSWER: A

Explanation:

In Databricks, to determine the owner of a managed table, you can utilize the Catalog Explorer feature. The steps are as follows: Access Catalog Explorer:

In your Databricks workspace, click on the Catalog icon in the sidebar to open Catalog Explorer.

Navigate to the Table:

Within Catalog Explorer, browse through the catalog and schema to locate the specific managed table whose ownership you wish to verify.

View Table Details:

Click on the table name to open its details page.

Identify the Owner:

On the table's details page, review the Owner field, which displays the principal (user, service principal, or group) that owns the table.

This method provides a straightforward way to ascertain the ownership of managed tables within the Databricks environment. Understanding table ownership is essential for managing permissions and ensuring proper access control.

Reference: Manage Unity Catalog object ownership

QUESTION NO: 5

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 environment1. 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. Reference:

What are user-defined functions (UDFs)?

User-defined scalar functions - SQL

V

QUESTION NO: 6

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)

age	country
80	*NULL*
75	*NULL*
90	*NULL*

C)

id	age	country
900	80	canada
901	75	canada
902	90	canada

D)

age	country
80	canada
14	canada
90	canada

E)

age	country
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 œmy_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 œage and œcountry , showing records where age is 75 or above and country is Canada. Reference: The answer can be inferred from understanding SQL queries and their outputs as per Databricks documentation: Databricks SQL

QUESTION NO: 7

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 documentation12, 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. Reference:

INSERT | Databricks on AWS

INSERT - Azure Databricks - Databricks SQL | Microsoft Learn

QUESTION NO: 8

What does Partner Connect do when connecting Power BI and Tableau?

A. Creates a Personal Access Token. downloads and installs an ODBC driver, and downloads a configuration file for connection by Power BI or Tableau to a SQL Warehouse (formerly known as a SQL Endpoint).

B. Creates a Personal Access Token for authentication into Databricks SQL and emails it to you. C. Downloads a configuration file for connection by Power BI or Tableau to a SQL Warehouse (formerly known as a SQL Endpoint).

C. Downloads and installs an ODBC driver.

ANSWER: A

Explanation:

When connecting Power BI and Tableau through Databricks Partner Connect, the system automates several steps to streamline the integration process:

Personal Access Token Creation: Partner Connect generates a Databricks personal access token, which is essential for authenticating and establishing a secure connection between Databricks and the BI tools.

ODBC Driver Installation: The appropriate ODBC driver is downloaded and installed. This driver facilitates communication between the BI tools and Databricks, ensuring compatibility and optimal performance.

Configuration File Download: A configuration file tailored for the selected BI tool (Power BI or Tableau) is provided. This file contains the necessary connection details, simplifying the setup process within the BI tool.

By automating these steps, Partner Connect ensures a seamless and efficient integration, reducing manual configuration efforts and potential errors. Reference: Connect Tableau and Databricks

---

QUESTION NO: 9

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 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.

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 œGROUP 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 œcustomers 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.

QUESTION NO: 10

What describes the variance of a set of values?

A. Variance is a measure of how far a single observed value is from a set ot va IN

B. Variance is a measure of how far an observed value is from the variable's maximum or minimum value.

C. Variance is a measure of central tendency of a set of values.

D. Variance is a measure of how far a set of values is spread out from the sets central value.

ANSWER: D

Explanation:

Variance is a statistical measure that quantifies the dispersion or spread of a set of values around their mean (central value). It is calculated by taking the average of the squared differences between each value and the mean of the dataset. A higher variance indicates that the data points are more spread out from the mean, while a lower variance suggests that they are closer to the mean. This measure is fundamental in statistics to understand the degree of variability within a dataset. WikipediaWikipedia+1Investopedia+1

Reference: Variance - Wikipedia
