DATABRICKS-CERTIFIED-DATA-ANALYST-ASSOCIATE Q&As

Databricks Certified Data Analyst Associate Exam

Pass Databricks DATABRICKS-CERTIFIED-DATA-ANALYST-ASSOCIATE Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.leads4pass.com/databricks-certified-data-analyst-associate.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Databricks Official Exam Center

QUESTION 1

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

Correct Answer: E

QUESTION 2

The stakeholders.customers table has 15 columns and 3,000 rows of data. The following command is run:

```
CREATE TEMP VIEW stakeholders.eur_customers AS
    SELECT * FROM stakeholders.customers
    WHERE continent = 'eur';
```

After running SELECT * FROM stakeholders.eur_customers, 15 rows are returned. After the command executes completely, the user logs out of Databricks. After logging back in two days later, what is the status of the stakeholders.eur_customers view?

A. The view remains available and SELECT * FROM stakeholders.eur_customers will execute correctly.

B. The view has been dropped.

C. The view is not available in the metastore, but the underlying data can be accessed with SELECT * FROM delta.`stakeholders.eur_customers`.

D. The view remains available but attempting to SELECT from it results in an empty result set because data in views are automatically deleted after logging out.

E. The view has been converted into a table.

Correct Answer: B

The command you sent creates a TEMP VIEW, which is a type of view that is only visible and accessible to the session

that created it.

When the session ends or the user logs out, the TEMP VIEW is automatically dropped and cannot be queried anymore. Therefore, after logging back in two days later, the status of the stakeholders.eur_customers view is that it has been

dropped and SELECT * FROM stakeholders.eur_customers will result in an error. The other options are not correct because:

A) The view does not remain available, as it is a TEMP VIEW that is dropped when the session ends or the user logs out.

C) The view is not available in the metastore, as it is a TEMP VIEW that is not registered in the metastore. The underlying data cannot be accessed with SELECT * FROM delta. stakeholders.eur_customers, as this is not a valid syntax for

querying a Delta Lake table. The correct syntax would be SELECT * FROM delta.dbfs:/stakeholders/eur_customers, where the location path is enclosed in backticks. However, this would also result in an error, as the TEMP VIEW does not

write any data to the file system and the location path does not exist.

D) The view does not remain available, as it is a TEMP VIEW that is dropped when the session ends or the user logs out. Data in views are not automatically deleted after logging out, as views do not store any data. They are only logical

representations of queries on base tables or other views.

E) The view has not been converted into a table, as there is no automatic conversion between views and tables in Databricks. To create a table from a view, you need to use a CREATE TABLE AS statement or a similar command.

QUESTION 3

Which of the following queries can be used to deduplicate the data from table_bronze and write it to a new table table_silver?

A. CREATE TABLE table_silver AS SELECT DISTINCT * FROM table_bronze;

B. CREATE TABLE table_silver AS INSERT * FROM table_bronze;

C. CREATE TABLE table_silver AS MERGE DEDUPLICATE * FROM table_bronze;

D. INSERT INTO TABLE table_silver SELECT * FROM table_bronze;

E. INSERT OVERWRITE TABLE table_silver SELECT * FROM table_bronze;

Correct Answer: A

QUESTION 4

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?

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 dropdown list query parameters, not dates.

Correct Answer: D

QUESTION 5

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.

Correct Answer: B

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.

Correct Answer: B

QUESTION 7

A data engineering team has created a Structured Streaming pipeline that processes data in micro-batches and populates gold-level tables. The microbatches are triggered every minute.

A data analyst has created a dashboard based on this gold-level data. The project stakeholders want to see the results in the dashboard updated within one minute or less of new data becoming available within the gold-level tables.

Which of the following cautions should the data analyst share prior to setting up the dashboard to complete this task?

A. The required compute resources could be costly

B. The gold-level tables are not appropriately clean for business reporting

C. The streaming data is not an appropriate data source for a dashboard

D. The streaming cluster is not fault tolerant

E. The dashboard cannot be refreshed that quickly

Correct Answer: A

QUESTION 8

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

Correct Answer: E

QUESTION 9

Consider the following two statements:

Statement 1:

```
SELECT *
    FROM customers
    LEFT SEMI JOIN orders
    ON customers.customer_id = orders.customer_id;
```

Statement 2:

```
SELECT *
    FROM customers
    LEFT ANTI JOIN orders
    ON customers.customer_id = orders.customer_id;
```

Which of the following describes how the result sets will differ for each statement when they are run in Databricks SQL?

A. The first statement will return all data from the customers table and matching data from the orders table. The second statement will return all data from the orders table and matching data from the customers table. Any missing data will be filled in with NULL.

B. When the first statement is run, only rows from the customers table that have at least one match with the orders table on customer_id will be returned. When the second statement is run, only those rows in the customers table that do not have at least one match with the orders table on customer_id will be returned.

C. There is no difference between the result sets for both statements.

D. Both statements will fail because Databricks SQL does not support those join types.

E. When the first statement is run, all rows from the customers table will be returned and only the customer_id from the orders table will be returned. When the second statement is run, only those rows in the customers table that do not have at least one match with the orders table on customer_id will be returned.

Correct Answer: B

QUESTION 10

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

Correct Answer: B

Reference: https://www.databricks.com/glossary/medallion-architecture