

DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE^{Q&As}

Databricks Certified Data Engineer Associate Exam

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

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

<https://www.pass2lead.com/databricks-certified-data-engineer-associate.html>

100% Passing Guarantee
100% Money Back Assurance

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

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers



QUESTION 1

A single Job runs two notebooks as two separate tasks. A data engineer has noticed that one of the notebooks is running slowly in the Job's current run. The data engineer asks a tech lead for help in identifying why this might be the case.

Which of the following approaches can the tech lead use to identify why the notebook is running slowly as part of the Job?

- A. They can navigate to the Runs tab in the Jobs UI to immediately review the processing notebook.
- B. They can navigate to the Tasks tab in the Jobs UI and click on the active run to review the processing notebook.
- C. They can navigate to the Runs tab in the Jobs UI and click on the active run to review the processing notebook.
- D. There is no way to determine why a Job task is running slowly.
- E. They can navigate to the Tasks tab in the Jobs UI to immediately review the processing notebook.

Correct Answer: C

QUESTION 2

A data analyst has created a Delta table sales that is used by the entire data analysis team. They want help from the data engineering team to implement a series of tests to ensure the data is clean. However, the data engineering team uses Python for its tests rather than SQL.

Which of the following commands could the data engineering team use to access sales in PySpark?

- A. `SELECT * FROM sales`
- B. There is no way to share data between PySpark and SQL.
- C. `spark.sql("sales")`
- D. `spark.delta.table("sales")`
- E. `spark.table("sales")`

Correct Answer: D

QUESTION 3

A dataset has been defined using Delta Live Tables and includes an expectations clause:

```
CONSTRAINT valid_timestamp EXPECT (timestamp > \"2020-01-01\") ON VIOLATION DROP ROW
```

What is the expected behavior when a batch of data containing data that violates these constraints is processed?

- A. Records that violate the expectation are dropped from the target dataset and loaded into a quarantine table.

- B. Records that violate the expectation are added to the target dataset and flagged as invalid in a field added to the target dataset.
- C. Records that violate the expectation are dropped from the target dataset and recorded as invalid in the event log.
- D. Records that violate the expectation are added to the target dataset and recorded as invalid in the event log.
- E. Records that violate the expectation cause the job to fail.

Correct Answer: D

QUESTION 4

Which of the following describes the relationship between Gold tables and Silver tables?

- A. Gold tables are more likely to contain aggregations than Silver tables.
- B. Gold tables are more likely to contain valuable data than Silver tables.
- C. Gold tables are more likely to contain a less refined view of data than Silver tables.
- D. Gold tables are more likely to contain more data than Silver tables.
- E. Gold tables are more likely to contain truthful data than Silver tables.

Correct Answer: C

QUESTION 5

A data engineer has been using a Databricks SQL dashboard to monitor the cleanliness of the input data to an ELT job. The ELT job has its Databricks SQL query that returns the number of input records containing unexpected NULL values. The data engineer wants their entire team to be notified via a messaging webhook whenever this value reaches 100.

Which of the following approaches can the data engineer use to notify their entire team via a messaging webhook whenever the number of NULL values reaches 100?

- A. They can set up an Alert with a custom template.
- B. They can set up an Alert with a new email alert destination.
- C. They can set up an Alert with a new webhook alert destination.
- D. They can set up an Alert with one-time notifications.
- E. They can set up an Alert without notifications.

Correct Answer: C

QUESTION 6

A data engineering team has two tables. The first table march_transactions is a collection of all retail transactions in the

month of March. The second table april_transactions is a collection of all retail transactions in the month of April. There are no duplicate records between the tables.

Which of the following commands should be run to create a new table all_transactions that contains all records from march_transactions and april_transactions without duplicate records?

- A. CREATE TABLE all_transactions AS SELECT * FROM march_transactions INNER JOIN SELECT * FROM april_transactions;
- B. CREATE TABLE all_transactions AS SELECT * FROM march_transactions UNION SELECT * FROM april_transactions;
- C. CREATE TABLE all_transactions AS SELECT * FROM march_transactions OUTER JOIN SELECT * FROM april_transactions;
- D. CREATE TABLE all_transactions AS SELECT * FROM march_transactions INTERSECT SELECT * FROM april_transactions;
- E. CREATE TABLE all_transactions AS SELECT * FROM march_transactions MERGE SELECT * FROM april_transactions;

Correct Answer: B

QUESTION 7

A data engineer is attempting to drop a Spark SQL table my_table. The data engineer 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's data was smaller than 10 GB
- C. The table was external
- D. The table did not have a location
- E. The table was managed

Correct Answer: C

QUESTION 8

Which of the following Git operations must be performed outside of Databricks Repos?

- A. Commit

- B. Pull
- C. Push
- D. Clone
- E. Merge

Correct Answer: D

Reference: <https://docs.databricks.com/repos/repos-setup.html>

QUESTION 9

Which of the following Structured Streaming queries is performing a hop from a Silver table to a Gold table?

A.

```
(spark.readStream.load(rawSalesLocation)
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("append")
  .table("newSales")
)
```

B.

```
(spark.read.load(rawSalesLocation)
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("append")
  .table("newSales")
)
```

C.

```
(spark.table("sales")
  .withColumn("avgPrice", col("sales") / col("units"))
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("append")
  .table("newSales")
)
```

D.

```
(spark.table("sales")
  .filter(col("units") > 0)
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("append")
  .table("newSales")
)
```

E.

```
(spark.table("sales")
  .groupBy("store")
  .agg(sum("sales"))
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("complete")
  .table("newSales")
)
```

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

Correct Answer: E

QUESTION 10

Which of the following is hosted completely in the control plane of the classic Databricks architecture?

- A. Worker node
- B. JDBC data source
- C. Databricks web application
- D. Databricks Filesystem
- E. Driver node

Correct Answer: E

[Latest DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE Dumps](#)

[DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE PDF Dumps](#)

[DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE VCE Dumps](#)