



VCE & PDF

GeekCert.com

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

2024 Latest geekcert DATABRICKS-CERTIFIED-DATA-ENGINEER-

ASSOCIATE PDF and VCE dumps Download

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.geekcert.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



VCE & PDF

GeekCert.com

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

2024 Latest geekcert DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE PDF and VCE dumps Download

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





QUESTION 1

A data analyst has developed a query that runs against Delta table. They want help from the data engineering team to implement a series of tests to ensure the data returned by the query is clean. However, the data engineering team uses Python for its tests rather than SQL.

Which of the following operations could the data engineering team use to run the query and operate with the results in PySpark?

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

Correct Answer: C

Explanation: from pyspark.sql import SparkSession

```
spark = SparkSession.builder.getOrCreate()
```

```
df = spark.sql("SELECT * FROM sales")
```

```
print(df.count())
```

QUESTION 2

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: C

Explanation: With the defined constraint and expectation clause, when a batch of data is processed, any records that violate the expectation (in this case, where the timestamp is not greater than \"2020-01-01\") will be dropped from the target dataset. These dropped records will also be recorded as invalid in the event log, allowing for auditing and tracking



of the data quality issues without causing the entire job to fail. <https://docs.databricks.com/en/delta-live-tables/expectations.html>

QUESTION 3

A data engineer needs to use a Delta table as part of a data pipeline, but they do not know if they have the appropriate permissions. In which of the following locations can the data engineer review their permissions on the table?

- A. Databricks Filesystem
- B. Jobs
- C. Dashboards
- D. Repos
- E. Data Explorer

Correct Answer: E

QUESTION 4

A data analysis team has noticed that their Databricks SQL queries are running too slowly when connected to their always-on SQL endpoint. They claim that this issue is present when many members of the team are running small queries simultaneously. They ask the data engineering team for help. The data engineering team notices that each of the team's queries uses the same SQL endpoint.

Which of the following approaches can the data engineering team use to improve the latency of the team's queries?

- A. They can increase the cluster size of the SQL endpoint.
- B. They can increase the maximum bound of the SQL endpoint's scaling range.
- C. They can turn on the Auto Stop feature for the SQL endpoint.
- D. They can turn on the Serverless feature for the SQL endpoint.
- E. They can turn on the Serverless feature for the SQL endpoint and change the Spot Instance Policy to "Reliability Optimized."

Correct Answer: A

Explanation: When many users are running small queries simultaneously on a SQL endpoint, the database can become overloaded, causing slow query execution times. By increasing the cluster size of the SQL endpoint, the database can handle more simultaneous queries, resulting in faster query execution times.

QUESTION 5

A data engineer is running code in a Databricks Repo that is cloned from a central Git repository. A colleague of the data engineer informs them that changes have been made and synced to the central Git repository. The data engineer now needs to sync their Databricks Repo to get the changes from the central Git repository.



Which of the following Git operations does the data engineer need to run to accomplish this task?

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

Correct Answer: C

Explanation: From the docs:

In Databricks Repos, you can use Git functionality to:

Clone, push to, and pull from a remote Git repository. Create and manage branches for development work, including merging, rebasing, and resolving conflicts.

Create notebooks and including IPYNB notebooks and edit them and other files.

Visually compare differences upon commit and resolve merge conflicts.

Source: <https://docs.databricks.com/en/repos/index.html>

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

[DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE Practice Test](#)

[DATABRICKS-CERTIFIED-DATA-ENGINEER-ASSOCIATE Braindumps](#)