



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 engineer is designing a data pipeline. The source system generates files in a shared directory that is also used by other processes. As a result, the files should be kept as is and will accumulate in the directory. The data engineer needs to identify which files are new since the previous run in the pipeline, and set up the pipeline to only ingest those new files with each run.

Which of the following tools can the data engineer use to solve this problem?

- A. Unity Catalog
- B. Delta Lake
- C. Databricks SQL
- D. Data Explorer
- E. Auto Loader

Correct Answer: E

Explanation: Auto Loader incrementally and efficiently processes new data files as they arrive in cloud storage without any additional setup. <https://docs.databricks.com/en/ingestion/auto-loader/index.html>

QUESTION 2

An engineering manager wants to monitor the performance of a recent project using a Databricks SQL query. For the first week following the project's release, the manager wants the query results to be updated every minute. However, the manager is concerned that the compute resources used for the query will be left running and cost the organization a lot of money beyond the first week of the project's release.

Which of the following approaches can the engineering team use to ensure the query does not cost the organization any money beyond the first week of the project's release?

- A. They can set a limit to the number of DBUs that are consumed by the SQL Endpoint.
- B. They can set the query's refresh schedule to end after a certain number of refreshes.
- C. They cannot ensure the query does not cost the organization money beyond the first week of the project's release.
- D. They can set a limit to the number of individuals that are able to manage the query's refresh schedule.
- E. They can set the query's refresh schedule to end on a certain date in the query scheduler.

Correct Answer: E

If a dashboard is configured for automatic updates, it has a Scheduled button at the top, rather than a Schedule button. To stop automatically updating the dashboard and remove its subscriptions:

Click Scheduled.



In the Refresh every drop-down, select Never.

Click Save. The Scheduled button label changes to Schedule. Source: <https://learn.microsoft.com/en-us/azure/databricks/sql/user/dashboards/>

QUESTION 3

A data engineer wants to schedule their Databricks SQL dashboard to refresh once per day, but they only want the associated SQL endpoint to be running when it is necessary.

Which of the following approaches can the data engineer use to minimize the total running time of the SQL endpoint used in the refresh schedule of their dashboard?

- A. They can ensure the dashboard's SQL endpoint matches each of the queries' SQL endpoints.
- B. They can set up the dashboard's SQL endpoint to be serverless.
- C. They can turn on the Auto Stop feature for the SQL endpoint.
- D. They can reduce the cluster size of the SQL endpoint.
- E. They can ensure the dashboard's SQL endpoint is not one of the included query's SQL endpoint.

Correct Answer: C

QUESTION 4

A data engineer has a single-task Job that runs each morning before they begin working. After identifying an upstream data issue, they need to set up another task to run a new notebook prior to the original task.

Which of the following approaches can the data engineer use to set up the new task?

- A. They can clone the existing task in the existing Job and update it to run the new notebook.
- B. They can create a new task in the existing Job and then add it as a dependency of the original task.
- C. They can create a new task in the existing Job and then add the original task as a dependency of the new task.
- D. They can create a new job from scratch and add both tasks to run concurrently.
- E. They can clone the existing task to a new Job and then edit it to run the new notebook.

Correct Answer: B

Explanation: To set up the new task to run a new notebook prior to the original task in a single-task Job, the data engineer can use the following approach: In the existing Job, create a new task that corresponds to the new notebook that needs to be run. Set up the new task with the appropriate configuration, specifying the notebook to be executed and any necessary parameters or dependencies. Once the new task is created, designate it as a dependency of the original task in the Job configuration. This ensures that the new task is executed before the original task.

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 edit them and other files.

Visually compare differences upon commit and resolve merge conflicts.

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

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

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

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