PRACTICE TEST SOFTWARE

Databricks

Databricks-Certified-Data-Engineer-Associate Exam

Databricks Certified Data Engineer Associate Exam

DEMO Version

Full Version Features:

• 90 Days Free Updates

• 30 Days Money Back Guarantee

• Instant Download Once Purchased

• 24 Hours Live Chat Support

Full version is available at link below with affordable price.

https://www.practicetestsoftware.com/databricks/databricks-certified-data-engineer-associate

Question 1. (Single Select)

A data organization leader is upset about the data analysis team's reports being different from the data engineering team's reports. The leader believes the siloed nature of their organization's data engineering and data analysis architectures is to blame.

Which of the following describes how a data lakehouse could alleviate this issue?

A: Both teams would autoscale their work as data size evolves
B: Both teams would use the same source of truth for their work
C: Both teams would reorganize to report to the same department
D: Both teams would be able to collaborate on projects in real-time
E: Both teams would respond more quickly to ad-hoc requests

Correct Answer: B

Explanation:
A data lakehouse is a data management architecture that combines the flexibility, cost-efficiency, and scale of data lakes with the data management and ACID transactions of data warehouses, enabling business intelligence (BI) and machine learning (ML) on all data12. By using a data lakehouse, both the data analysis and data engineering teams can access the same data sources and formats, ensuring data consistency and quality across their reports. A data lakehouse also supports schema enforcement and evolution, data validation, and time travel to old table versions, which can help resolve data conflicts and errors1. 1: What is a Data Lakehouse? - Databricks 2: What is a data lakehouse? | IBM

Question 2. (Multi Select)

The Delta transaction log for the 'students' tables is shown using the 'DESCRIBE HISTORY students' command. A Data Engineer needs to query the table as it existed before the UPDATE operation listed in the log.

	version	timestamp	operation
1	8	2024–04–22T14:33:31.000	OPTIMIZE
2	7	2024–04–22T14:33:16.000	MERGE
3	6	2024–04–22T14:33:06.000	DELETE
4	5	2024–04–22T14:32:58.000	UPDATE
5	4	2024–04–22T14:32:47.000	WRITE
6	3	2024–04–22T14:32:44.000	WRITE
7	2	2024–04–22T14:32:23.000	WRITE
8	1	2024–04–22T14:32:20.000	WRITE
9	0	2024–04–22T14:31:49.000	CREATE TABLE

Which command should the Data Engineer use to achieve this? (Choose two.)

A: SELECT * FROM students@v4
B: SELECT * FROM students TIMESTAMP AS OF '2024-04-22T 14:32:47.000+00:00'
C: SELECT * FROM students FROM HISTORY VERSION AS OF 3
D: SELECT * FROM students VERSION AS OF 5
E: SELECT * FROM students TIMESTAMP AS OF '2024-04-22T 14:32:58.000+00:00'

Correct Answer: A, B

Question 3. (Single Select)

A data engineer has realized that they made a mistake when making a daily update to a table. They need to use Delta time travel to restore the table to a version that is 3 days old. However, when the data engineer attempts to time travel to the older version, they are unable to restore the data because the data files have been deleted.

Which of the following explains why the data files are no longer present?

A: The VACUUM command was run on the table
B: The TIME TRAVEL command was run on the table
C: The DELETE HISTORY command was run on the table
D: The OPTIMIZE command was nun on the table
E: The HISTORY command was run on the table

Correct Answer: A

Explanation:

The VACUUM command is used to remove files that are no longer referenced by a Delta table and are older than the retention threshold1. The default retention period is 7 days2, but it can be changed by setting the delta.logRetentionDuration and delta.deletedFileRetentionDuration configurations3. If the VACUUM command was run on the table with a retention period shorter than 3 days, then the data files that were needed to restore the table to a 3-day-old version would have been deleted. The other commands do not delete data files from the table. The TIME TRAVEL command is used to query a historical version of the table4. The DELETE HISTORY command is not a valid command in Delta Lake. The OPTIMIZE command is used to improve the performance of the table by compacting small files into larger ones5. The HISTORY command is used to retrieve information about the operations performed on the table. 1: VACUUM | Databricks on AWS 2: Work with Delta Lake table history | Databricks on AWS 3: [Delta Lake configuration | Databricks on AWS] 4: Work with Delta Lake table history - Azure Databricks 5: [OPTIMIZE | Databricks on AWS] : [HISTORY | Databricks on AWS]

Question 4. (Single Select)

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: E

Explanation:
Databricks Repos is a visual Git client and API in Databricks that supports common Git operations such as commit, pull, push, branch management, and visual comparison of diffs when committing1. However, merge is not supported in the Git dialog2. You need to use the Repos UI or your Git provider to merge branches3. Merge is a way to combine the commit history from one branch into another branch1. During a merge, a merge conflict is encountered when Git cannot automatically combine code from one branch into another. Merge conflicts require manual resolution before a merge can be completed1. 4: Run Git operations on Databricks Repos4, 1: CI/CD techniques with Git and Databricks Repos1, 3: Collaborate in Repos3, 2: Databricks Repos - What it is and how we can use it2.
Databricks Repos is a visual Git client and API in Databricks that supports common Git operations such as commit, pull, push, merge, and branch management. However, to clone a remote Git repository to a

Databricks repo, you must use the Databricks UI or API. You cannot clone a Git repo using the CLI through a cluster's web terminal, as the files won't display in the Databricks UI1. 1: Run Git operations on Databricks Repos | Databricks on AWS2

Question 5. (Single Select)

Which of the following data lakehouse features results in improved data quality over a traditional data lake?

A: A data lakehouse provides storage solutions for structured and unstructured data.
B: A data lakehouse supports ACID-compliant transactions.
C: A data lakehouse allows the use of SQL queries to examine data.
D: A data lakehouse stores data in open formats.
E: A data lakehouse enables machine learning and artificial Intelligence workloads.

Correct Answer: B

Explanation:
: A data lakehouse is a data management architecture that combines the flexibility, cost-efficiency, and scale of data lakes with the data management and ACID transactions of data warehouses, enabling business intelligence (BI) and machine learning (ML) on all data12. One of the key features of a data lakehouse is that it supports ACID-compliant transactions, which means that it ensures data integrity, consistency, and isolation across concurrent read and write operations3. This feature results in improved data quality over a traditional data lake, which does not support transactions and may suffer from data corruption, duplication, or inconsistency due to concurrent or streaming data ingestion and processing . 1: What is a Data Lakehouse? - Databricks 2: What is a Data Lakehouse? Definition, features & benefits. - Qlik 3: ACID Transactions - Databricks : [Data Lake vs Data Warehouse: Key Differences] : [Data Lakehouse: The Future of Data Engineering]

Full version is available at link below with affordable price.

https://www.practicetestsoftware.com/databricks/databricks-certified-data-engineer-associate

40% Discount Coupon Code: GET40