

PRACTICE QUESTIONS

Databricks Certified Data Analyst Associate

SHARED BY [u/Smooth-Beacon-7032](#) on aiexamcerts

EDITION Free Set 2

QUESTIONS 10

SHARED October 15, 2026

EXAM Databricks Certified Data Analyst Associate

Shared by [u/Smooth-Beacon-7032](#) on aiexamcerts

10 practice questions for the Databricks Certified Data Analyst Associate exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/databricks/databricks-certified-data-analyst-associate. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. A compliance officer wants analysts to see customer email addresses masked as ***** unless they belong to the privacy group. Which Unity Catalog feature enforces this on the column itself?
 - A A view that hides the table from all users
 - B A Delta table comment that warns readers about personal data
 - C A column mask function attached to the email column
 - D A VACUUM schedule that removes old email values

Answer: C

A column mask is a SQL function attached to a column that returns the real value or a masked value based on the caller's identity or group membership. Because it runs at query time, every access path is covered. A comment only describes the data, and VACUUM deletes old file versions rather than masking current values.

2. Running DROP TABLE on a managed Unity Catalog table removes its underlying data files as well as its metadata.
 - A True
 - B False

Answer: A (True)

A managed table's data is stored in a location Databricks controls, so dropping the table removes both the metadata and the data files. External tables behave differently because their data lives in a path you specify, and dropping them removes only the metadata. Knowing this difference matters before any cleanup.

3. An analyst asks who last changed a Unity Catalog table and which queries read it. Which feature supplies this lineage information?
 - A The Spark UI stage timeline of the most recent cluster run
 - B The notebook revision history for the analyst's own notebook
 - C The table's partition columns listed by SHOW PARTITIONS
 - D Unity Catalog data lineage together with the audit and system tables

Answer: D

Unity Catalog records lineage between tables, notebooks and queries, and the audit and system tables record who accessed or changed objects. The Spark UI shows execution stages for one job rather than governance history. Notebook revisions cover only one notebook, and SHOW PARTITIONS lists storage partitions, not access history.

4. Delta tables with frequently filtered columns can be clustered with Liquid clustering by adding a CLUSTER _____ clause on the table definition.

Answer: BY

Liquid clustering is enabled with CLUSTER BY on the columns that queries commonly filter on, and the layout adapts as data changes. It replaces the older practice of choosing partition columns for most tables. The clause is set when the table is created or altered, not in each query.

5. Which two practices reduce the cost and runtime of a large aggregation query on Delta tables? (Choose two.)

- A Filtering on the clustering or partition columns so that fewer files are read
- B Using SELECT * and dropping columns in the dashboard instead of the query
- C Joining two large tables without any join keys
- D Selecting only the columns the report needs
- E Running VACUUM before every query

Answer: A, D

Filtering on clustering or partition columns lets the engine skip files, and selecting only needed columns reduces the data read from storage. SELECT * followed by dropping columns in the dashboard still reads everything. Joins without keys create cartesian products, and VACUUM before every query deletes history without speeding up reads.

6. An analyst must keep only the most recent order per customer from a table with many orders each. Which pattern is the most direct?

- A SELECT DISTINCT customer_id, with the order_date column removed from the output list
- B A cross join of the orders table with itself and no condition
- C A LIMIT 1 applied to the whole orders table after sorting by customer name
- D ROW_NUMBER() OVER (PARTITION BY customer_id ORDER BY order_date DESC) with QUALIFY = 1

Answer: D

ROW_NUMBER numbers each customer's orders from newest to oldest, and QUALIFY keeps only the first row per customer. DISTINCT on customer_id alone drops the order details the analyst needs. A cross join multiplies rows, and LIMIT 1 returns a single order for the entire table.

7. A chart must show each region's share of total sales as a percentage of the whole. Which visualization fits this best?
- A A scatter plot of region sales against the order date
 - B A table with a row for every single order in the period
 - C A pie or donut chart of region sales, with few regions
 - D A word cloud sized by the number of customers

Answer: C

Pie and donut charts show parts of a whole, which matches a share-of-total question when there are only a few categories. A scatter plot compares two numeric variables and does not show shares. A row-level table hides the proportions, and a word cloud does not encode values precisely.

8. Query history in Databricks SQL can be used to find which queries were slow or failed in the past.
- A True
 - B False

Answer: A (True)

Query history lists past executions with their status, duration and user, which helps analysts find slow or failed queries and tune them. It is a record of past activity, not a live monitor. Filtering that history is a common first step in performance troubleshooting.

9. A dashboard must be shared with executives who should see the same figures, but without the ability to edit the queries. What is the right sharing approach?
- A Give every executive the CAN MANAGE permission on the underlying warehouse
 - B Export the notebook source code and email it to each executive
 - C Grant the executives MODIFY on the source tables
 - D Share the dashboard with view permission

Answer: D

Dashboards can be shared with view permission, which lets people see results without editing the datasets or visualizations. Giving manage rights on the warehouse exposes compute controls the executives do not need. Sending source code and granting MODIFY on tables both go beyond read-only viewing.

10. An alert keeps firing every time a nightly load runs, even though the business does not need that alert. Which change keeps the alert and stops the noise?
- A Delete the Unity Catalog metastore and recreate it from a backup
 - B Change the alert schedule or condition to match when the business cares
 - C Increase the table's VACUUM retention to 720 hours
 - D Disable the query cache on the SQL warehouse

Answer: B

An alert's schedule and its threshold condition decide when it evaluates and whether it notifies, so adjusting them is the direct fix. Deleting a metastore would destroy governance objects and is not a tuning step. VACUUM retention and the query cache do not change when an alert sends its notification.