

PRACTICE QUESTIONS

Databricks Certified Data Analyst Associate

SHARED BY [u/Honest-Pencil-4701](#) on aiexamcerts

EDITION Free Set 1

QUESTIONS 10

SHARED October 15, 2026

EXAM Databricks Certified Data Analyst Associate

Shared by [u/Honest-Pencil-4701](#) 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. An analyst suspects a bad MERGE overwrote values in a Delta table named sales_orders. Which command lists the table's versions with their operations and timestamps so the analyst can pick a version to query?
 - A SHOW TABLE EXTENDED IN retail LIKE 'sales_orders'
 - B ANALYZE TABLE sales_orders COMPUTE STATISTICS
 - C DESCRIBE HISTORY sales_orders
 - D VACUUM sales_orders RETAIN 0 HOURS

Answer: C

DESCRIBE HISTORY returns one row per commit, including the version number, timestamp, user and operation such as MERGE or WRITE, which is what an analyst needs before reading an older version. SHOW TABLE EXTENDED returns metadata about the table rather than its commits, and ANALYZE TABLE only refreshes statistics. VACUUM deletes data files that are no longer referenced, which would remove the very versions the analyst wants to recover.

2. Dashboard widgets that refresh automatically on a set schedule use a refresh _____ that is attached to the dashboard.

Answer: schedule

A scheduled refresh runs the dashboard's underlying queries at an interval you set and can notify subscribers when the refreshed results are ready. Schedules are configured on the dashboard, not on individual visualizations. Viewers still see the data as of the last refresh unless they trigger a new one.

3. An analyst needs a running total of daily sales for each store, with every daily row kept in the output. Which expression should be used?
 - A SUM(sales) OVER (PARTITION BY store_id ORDER BY sale_date)
 - B GROUP BY store_id, with SUM(sales) selected for each store
 - C PIVOT sales over the sale_date values
 - D UNPIVOT the store columns into rows

Answer: A

A window SUM with PARTITION BY store_id and ORDER BY sale_date produces a cumulative total per store while keeping every daily row. A plain GROUP BY collapses each store into one row, so it cannot show the running total by day. PIVOT and UNPIVOT reshape columns and rows, but they do not compute cumulative sums.

4. A Databricks SQL alert can trigger a notification only when a person opens the dashboard that contains it.
- A True
 - B False

Answer: B (False)

Alerts run on their own schedule and send notifications when their condition evaluates to true, without anyone opening a dashboard. That is the purpose of an alert compared with a dashboard that someone must view. The statement describes the opposite of how scheduled alerts behave.

5. An analyst group needs to read the table `retail.sales.orders` and nothing else. Which two grants are needed in addition to the table grant? (Choose two.)
- A USE CATALOG on retail
 - B MODIFY on retail.sales.orders
 - C USE SCHEMA on retail.sales
 - D CREATE TABLE on retail.sales
 - E ALL PRIVILEGES on the metastore

Answer: A, C

Reading a table in Unity Catalog requires SELECT on the table along with USE CATALOG on its catalog and USE SCHEMA on its schema. MODIFY and CREATE TABLE grant write capabilities the analyst group does not need, and ALL PRIVILEGES on the metastore is far broader than the requirement. Least privilege means granting only the three named privileges.

6. A column named `payload` holds JSON text such as `{"sku": "A12", "qty": 3}`. Which function turns it into a struct with typed fields when a schema is supplied?
- A `CAST(payload AS STRING)`
 - B `from_json(payload, schema)`
 - C `EXPLODE(payload)`
 - D `SPLIT(payload, ',')`

Answer: B

`from_json` parses a JSON string with the supplied schema and returns a struct whose fields have the declared types. CAST to STRING leaves the value as text, so no fields are extracted. EXPLODE expects an array or map, and SPLIT breaks text on a delimiter, which does not understand JSON structure.

7. An analyst wants a bar chart of orders per week, where weeks are derived from a timestamp column. Which SQL function groups a timestamp into the start of its week or month?
- A DATEDIFF with the unit set to days
 - B CURRENT_DATE applied to the timestamp column
 - C APPROX_COUNT_DISTINCT over the timestamp column
 - D DATE_TRUNC with the unit set to week or month

Answer: D

DATE_TRUNC returns the timestamp rounded down to the chosen unit, so grouping by it produces one bar per week or month. DATEDIFF returns the number of units between two dates, which is not a grouping key. CURRENT_DATE returns today's date, and APPROX_COUNT_DISTINCT estimates distinct values rather than bucketing time.

8. In Unity Catalog, a table is addressed with a three-level name in the form catalog._____.table.

Answer: schema (also accepted: database)

Unity Catalog uses catalog.schema.table as its three-level namespace, and a schema is also called a database in some contexts. Access is granted at each level, so a user needs USE CATALOG and USE SCHEMA before reading a table. Skipping the middle level leaves the name unresolved.

9. A BI team needs a SQL endpoint that scales to many concurrent dashboard users without the analysts starting clusters. Which option matches this need?
- A A serverless SQL warehouse
 - B A single-node cluster on a laptop
 - C A Delta sharing recipient
 - D A local Spark session in a notebook

Answer: A

Serverless SQL warehouses are managed by Databricks, start quickly and scale for concurrent BI users. A single-node cluster on a laptop is outside the platform and offers no shared endpoint. Delta Sharing is for exchanging data with recipients, and a local Spark session serves one user in one notebook.

10. Which default table format does Databricks use when a CREATE TABLE statement does not name a format?
- A CSV
 - B Delta
 - C Hive text
 - D JSON Lines

Answer: B

Delta is the default table format on Databricks, which is why new tables get transaction logs and time travel without extra syntax. CSV, Hive text and JSON Lines can be read by Databricks, but none is the default for new managed tables. Choosing one of those formats must be stated explicitly.