

PRACTICE QUESTIONS

Databricks Certified Data Engineer Associate

SHARED BY [u/silentbeaver1986](#) on aiexamcerts

EDITION Free Set 2

QUESTIONS 10

SHARED October 15, 2026

EXAM Databricks Certified Data Engineer Associate

Shared by [u/silentbeaver1986](#) on aiexamcerts

10 practice questions for the Databricks Certified Data Engineer 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-engineer-associate. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. A company wants data engineers to keep notebooks, pipeline code and job definitions in Git and deploy them consistently to dev and prod workspaces. Which approach is the most appropriate?
 - A Copy notebooks by hand between workspaces each release
 - B Edit notebooks in the production workspace and sync them back to dev by hand
 - C Deploy Databricks Asset Bundles from Git with the Databricks CLI
 - D Export each job as JSON and recreate it in prod through the UI

Answer: C

Databricks Asset Bundles describe jobs, pipelines and other resources as code in `databricks.yml`, so the same definition deploys to each target workspace from Git. Manual copies and direct production edits make releases inconsistent and hard to audit. Exporting jobs as JSON and recreating them through the UI is manual and error prone.

2. In the medallion architecture, bronze tables should store only aggregated business metrics so that storage costs stay low.
 - A True
 - B False

Answer: B (False)

Bronze holds raw or lightly ingested data in its original form, which allows reprocessing when logic changes. Aggregated metrics belong in the gold layer. The statement is false.

3. A pipeline must combine a streaming table of orders with a customer dimension table and keep only orders from active customers. Which operation expresses this?
 - A A UNION ALL that stacks the order rows on top of the customer rows
 - B An INNER JOIN on customer_id with a filter on the active flag
 - C A CROSS JOIN with no join condition
 - D A window function ordered by order date

Answer: B

An inner join on customer_id returns only matching orders, and the filter on the active flag keeps active customers. UNION stacks rows rather than matching them, and a cross join pairs every order with every customer. A window function ranks rows within a partition and does not filter by customer status.

4. Delta Lake's _____ operation reads through the Delta table history and returns the data as it existed at an earlier version or timestamp.

Answer: time travel (also accepted: time-travel)

Time travel uses the transaction log to read earlier versions, specified with VERSION AS OF or TIMESTAMP AS OF. Retention is governed by table settings and VACUUM. The answer names the Delta feature for reading history.

5. Which TWO statements about Delta Lake time travel and retention are correct? (Choose two.)

- A Old versions remain readable only while their data files are retained
- B Time travel works indefinitely, regardless of VACUUM
- C RESTORE TABLE can roll a table back to an earlier version
- D Time travel requires a separate backup job for each version
- E Time travel is available only for tables stored in CSV format

Answer: A, C

Old versions depend on their data files, which VACUUM removes once they are past the retention threshold, so reads of old versions fail after that point. RESTORE TABLE rolls the table back to an earlier version using the log. Time travel needs no separate backup job and works on Delta tables, not CSV.

6. A gold table must hold daily revenue by region and the team wants fast dashboards. Which pattern is most appropriate?

- A Recompute the full aggregation from raw bronze data on each refresh
- B Join the bronze tables inside every dashboard query at read time
- C Build an aggregated table that the pipeline refreshes on schedule
- D Keep the aggregation only in a temporary view for each session

Answer: C

A gold table with precomputed daily aggregates lets dashboards read small, fast results, and the pipeline refreshes it on schedule. Recomputing from bronze on every refresh, or joining bronze tables in every dashboard query, is slow and costly. A temporary view disappears with the session and stores no results.

7. A team wants an email alert when a pipeline job fails or runs longer than expected. Which Lakeflow Jobs feature handles this?
- A A cluster event log that records each job run
 - B Job notifications for failures and long runtimes sent by email
 - C A Delta table of run history that someone queries once a week
 - D An autoscaling policy on the job cluster

Answer: B

Job notifications send alerts on failure, success or duration thresholds to configured recipients, so problems are noticed quickly. An event log records activity but does not alert anyone, and a run history table queried weekly is slow and unreliable. An autoscaling policy changes cluster size and does not notify anyone.

8. An all-purpose cluster that stays running between jobs is the most cost-efficient choice for scheduled production pipelines.
- A True
 - B False

Answer: B (False)

Job compute starts for the scheduled run and terminates afterward, which avoids paying for idle time. All-purpose clusters are priced for interactive use and cost more when left running. The statement is false.

9. A data engineer wants a table that managed Unity Catalog stores in its own storage location, with the platform handling file layout and lifecycle. Which table type is this?
- A An external table that points to files the team places in storage by hand
 - B A temporary view
 - C A foreign table in a federated catalog
 - D A managed table

Answer: D

A managed table is stored in the metastore-managed location, and Unity Catalog handles its files and drops data when the table is dropped. An external table depends on files the team manages, and a temporary view or a foreign table does not give the platform ownership of the data files. Choose managed when the platform should own the lifecycle.

10. A new hire should be able to query tables in a schema but must not change their structure or delete them. Which grant pattern matches this?
- A GRANT USE SCHEMA and SELECT on the schema's tables
 - B GRANT ALL PRIVILEGES on the catalog to the new hire
 - C Make the new hire the owner of every table in the schema
 - D GRANT USE CATALOG with no other privileges

Answer: A

USE SCHEMA plus SELECT gives read access without changing structure, and it follows least privilege. ALL PRIVILEGES and ownership grant the power to alter or drop tables. USE CATALOG alone lets the new hire see the catalog but not read any table.