

PRACTICE QUESTIONS

Databricks Certified Data Engineer Professional

SHARED BY [u/Crisp-Owl-4281](#) on aiexamcerts

EDITION Free Set 2

QUESTIONS 10

SHARED October 15, 2026

EXAM Databricks Certified Data Engineer Professional

Shared by [u/Crisp-Owl-4281](#) on aiexamcerts

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

1. A table partitioned by event_date is queried mostly by customer_id, a high-cardinality column, and the partitions have become tiny. What is the Databricks-recommended approach for this table design?
 - A Use liquid clustering with CLUSTER BY (customer_id)
 - B Partition the table by customer_id so each customer has its own folder
 - C Partition by event_date and customer_id together
 - D Run OPTIMIZE with ZORDER BY (event_date) after every write

Answer: A

Liquid clustering with CLUSTER BY on the filter column adapts the data layout over time and avoids the small-partition problem for high-cardinality keys. Partitioning by customer_id, alone or together with event_date, would create a very large number of tiny partitions, which is the problem the team already has. Z-ordering on event_date does not help queries that filter on customer_id.

2. A CHECK constraint on a Delta table column only records violating rows in a log and never blocks a write.
 - A True
 - B False

Answer: B (False)

CHECK constraints are enforced on write, so a violating row makes the transaction fail rather than being logged. The idea that they only flag rows comes from warn-level pipeline expectations, which work differently. Constraints are part of the table definition and protect every writer.

3. A pipeline ingests JSON files that arrive in cloud storage, and new fields appear in the source over time. The team wants new columns added automatically without manual schema edits. Which Auto Loader option controls this behavior?

- A `cloudFiles.format`
- B `cloudFiles.schemaEvolutionMode`
- C `cloudFiles.useNotifications`
- D `cloudFiles.includeExistingFiles`

Answer: B

`cloudFiles.schemaEvolutionMode` decides what happens when new columns appear, and `addNewColumns` adds them to the stored schema, after which the stream restarts. `cloudFiles.format` only tells Auto Loader which file type to parse. `cloudFiles.includeExistingFiles` only decides whether files already in the folder are processed on the first run, and it does not affect the schema.

4. To read only the row-level inserts, updates and deletes made to a Delta table since a given version, enable the table property `delta.enableChangeDataFeed` and read the _____ Data Feed.

Answer: Change

Change Data Feed records row-level changes with `_change_type` metadata, which downstream consumers read incrementally. Time travel reads past snapshots but does not describe row-level changes. Liquid clustering organizes data layout and does not track changes.

5. Which two signals help diagnose a slow Structured Streaming query? (Choose two.)

- A Batch duration and input rows per second in the progress metrics
- B The cluster's Databricks Runtime version string
- C The Streaming tab in the Spark UI with per-batch details
- D The catalog's managed storage location
- E The column count of the sink table

Answer: A, C

Progress metrics expose batch duration and input rate, and the Streaming tab charts them over time for each query. The runtime version string, the storage location and the number of sink columns do not describe query performance.

6. Which Databricks CLI command checks a bundle's `databricks.yml` for errors without changing the workspace?
- A `databricks bundle validate`
 - B `databricks bundle deploy`
 - C `databricks bundle run`
 - D `databricks workspace import`

Answer: A

`databricks bundle validate` checks the configuration and reports schema or reference errors without changing the workspace. `deploy` uploads and creates or updates the resources, so it does change the workspace. `databricks workspace import` uploads a single file or notebook and does not read the bundle at all.

7. A notebook in a Git folder needs a Python package for that notebook session only. Which magic command installs it?
- A `%sh apt install <package>`
 - B `%scala import <package>`
 - C `%pip install <package>`
 - D `%fs install <package>`

Answer: C

`%pip` installs a Python package into the notebook's environment, so it applies to that notebook session and not the whole cluster. Using `%sh` with `apt` installs operating system packages and does not make the library available to Python. `%scala` runs Scala code and `%fs` works with the file system, so neither installs a Python package.

8. In Unity Catalog, a user who has `SELECT` on a table also needs `USE CATALOG` on its catalog and `USE SCHEMA` on its schema before the query can run.
- A True
 - B False

Answer: A (True)

Unity Catalog checks the privileges on the containing objects as well, so the principal needs `USE CATALOG` and `USE SCHEMA` in addition to `SELECT` on the table. The tempting assumption is that the table grant alone is enough. Missing parent grants is a common cause of permission-denied errors.

9. Which record shows update starts, flow progress and expectation results for a Lakeflow Declarative Pipelines pipeline?
- A The cluster's syslog file
 - B The pipeline event log
 - C The Spark UI SQL tab
 - D The Unity Catalog lineage graph

Answer: B

Each pipeline has an event log that records update lifecycle, flow progress, data quality expectation outcomes and errors. The cluster syslog holds node-level operating system output, which is too low level for pipeline outcomes. The Spark UI SQL tab shows query plans, and the lineage graph shows table dependencies, so neither records pipeline update events.

10. Which identity should own a production Databricks Workflows job so that it keeps running when the engineer who created it leaves the team?
- A The engineer's personal user account
 - B A shared team account with a common password
 - C The workspace admin who approved the job
 - D A service principal

Answer: D

A service principal is a non-human identity, so the job's permissions do not depend on one person and survive staff changes. A personal account stops working when that user is deactivated. A shared team account and an individual admin are poor owner models for production jobs, since the first weakens auditing and the second ties the job to one person again.