

PRACTICE QUESTIONS

Microsoft

DP-750

Microsoft Certified: Azure Databricks Data Engineer Associate

SHARED BY [u/throwaway_dp750](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	Microsoft DP-750, Microsoft Certified: Azure Databricks Data Engineer Associate

Shared by [u/throwaway_dp750](#) on aiexamcerts

10 practice questions for the Microsoft DP-750 exam, Microsoft Certified: Azure Databricks Data Engineer Associate, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/microsoft/dp-750. Spotted a mistake? Email support@aiexamcerts.com with the exam code DP-750 and the question number.

1. Which Azure service provides the hierarchical data lake storage that Azure Databricks commonly reads and writes?
 - A Azure Cache for Redis, which holds data in memory for fast lookups
 - B Azure Data Lake Storage Gen2 with the hierarchical namespace enabled
 - C Azure Service Bus queues, which carry messages between applications
 - D Azure Front Door, which routes web traffic to origins

Answer: B

Azure Data Lake Storage Gen2 with the hierarchical namespace provides folder-like storage that Spark and Delta Lake use for data files. Redis and Service Bus are messaging or caching services, which makes them poor fits for a data lake. Front Door routes web traffic and does not store data.

2. A Lakeflow Spark Declarative Pipeline requires engineers to write the update order of its tables by hand in a separate job.
 - A True
 - B False

Answer: B (False)

False. Pipelines infer dependencies from the datasets that each definition reads, so the platform works out the update order. Engineers declare the tables and their sources, and the pipeline runs them in the right sequence.

3. A join on a skewed key makes a few tasks run far longer than the rest. Which Spark feature addresses skew in joins?
 - A Increasing the driver memory only
 - B Disabling Adaptive Query Execution
 - C Reading the tables as CSV text
 - D Adaptive skew join handling in Adaptive Query Execution

Answer: D

Adaptive Query Execution can detect skewed partitions at runtime and split them, which spreads the heavy work across more tasks. Driver memory does not affect how the join distributes work, which makes it a tempting but wrong fix. Disabling AQE removes a helpful optimization, and CSV reading does not address skew.

4. The Auto Loader option that controls what happens when a new column appears is named `cloudFiles.schemaEvolutionMode`.

Answer: schemaEvolutionMode

`cloudFiles.schemaEvolutionMode` sets how the stream responds to new columns, such as adding them or failing the stream. It works with the schema location, which stores the inferred schema. Choosing the mode is a deliberate decision because it changes whether pipelines stop or adapt.

5. Which two statements about OPTIMIZE and VACUUM are correct? (Choose two.)

- A OPTIMIZE compacts many small files into fewer larger files
- B VACUUM deletes unreferenced data files older than the retention threshold
- C VACUUM rewrites every table in place to change its file format
- D OPTIMIZE deletes old table versions immediately
- E VACUUM must run before every write to a table

Answer: A, B

OPTIMIZE compacts small files, which speeds up reads, and VACUUM removes files that no retained version references. VACUUM does not change file formats, which makes option C a tempting misreading. OPTIMIZE does not delete old versions directly, and VACUUM is not a prerequisite for writes.

6. Which compute choice fits a development notebook that is used a few hours a day and should stop when idle?

- A An all-purpose cluster with autotermination set
- B A job cluster that never terminates between runs
- C A SQL warehouse that stays always on at its maximum size
- D A fixed cluster with 200 workers that runs all day

Answer: A

An all-purpose cluster suits interactive development, and autotermination stops it after idle time so cost follows use. A job cluster is designed for a scheduled run and is created for it, which makes it a tempting but wrong fit for interactive work. Always-on and fixed large clusters waste money when idle.

7. Which set of privileges lets a user query a table in Unity Catalog?

- A MODIFY on the table only
- B CREATE TABLE on the schema only
- C USE CATALOG, USE SCHEMA and SELECT
- D OWNERSHIP of the metastore

Answer: C

A user needs USE CATALOG on the catalog, USE SCHEMA on the schema and SELECT on the table to read it. MODIFY allows changes rather than reads, which makes it a tempting but incomplete answer. Creating tables and owning the metastore are administrative rights that are not needed to query.

8. Unity Catalog row filters and column masks can restrict which rows and which column values a user sees in a table.
- A True
 - B False

Answer: A (True)

True. A row filter applies a function that returns only permitted rows, and a column mask applies a function that hides or transforms sensitive values for users who are not allowed to see them. Both are enforced when the table is queried.

9. Which Databricks feature runs notebooks, scripts and SQL tasks on a schedule with retries and task dependencies?
- A Databricks Connect, which runs code from a local IDE
 - B Databricks Jobs
 - C Git folders, which sync code with a Git provider
 - D Widgets, which pass parameters into a notebook

Answer: B

Databricks Jobs schedules tasks, retries failed ones and enforces dependencies between them, which is the orchestration layer for production workloads. Databricks Connect is useful for development but does not schedule production runs. Git folders and widgets support code management and parameters, not orchestration.

10. What does the repartition method do to a DataFrame?
- A It deletes duplicate rows from the DataFrame
 - B It returns a view that holds no data
 - C It writes the DataFrame to a Delta table
 - D It changes the number of partitions and can shuffle the data

Answer: D

repartition changes how many partitions hold the data and usually triggers a shuffle to rebalance it. Removing duplicates is the job of dropDuplicates, which makes it a tempting but different method. Writing and creating views are separate operations.