

PRACTICE QUESTIONS

Databricks Certified Machine Learning Professional

SHARED BY [u/politeheron1999](#) on aiexamcerts

EDITION	Free Set 2
---------	------------

QUESTIONS	10
-----------	----

SHARED	October 15, 2026
--------	------------------

EXAM	Databricks Certified Machine Learning Professional
------	--

Shared by [u/politeheron1999](#) on aiexamcerts

10 practice questions for the Databricks Certified Machine Learning 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-machine-learning-professional. Spotted a mistake?

Email support@aiexamcerts.com with the exam name and the question number.

1. A monitored model's accuracy on recent labeled data falls below the agreed threshold. What is the most appropriate next step?
 - A Delete the registered model and redeploy the earliest version without evaluation
 - B Disable inference logging so the drop no longer appears
 - C Retrain on recent labeled data, evaluate against the champion, promote only if it passes
 - D Increase the endpoint's compute size and concurrency, because a larger endpoint retrains the model

Answer: C

A retrained candidate should be checked on held-out recent data against the current champion before any promotion, which keeps the decision evidence-based. Hiding the signal or deleting models removes the evidence, and endpoint size does not change prediction accuracy.

2. A drift metric requires ground-truth labels for every scored row before it can be computed.
 - A True
 - B False

Answer: B (False)

Drift metrics compare input or prediction distributions with a baseline, so they can be computed before labels arrive. Labels are needed for accuracy-style metrics, which is why drift monitoring is an early warning. The statement is false.

3. Which URI loads the model logged under run abc123 in the artifact path model?
 - A `models:/abc123/model`
 - B `experiments:/abc123/model`
 - C `runs:/abc123/model`
 - D `artifacts:/model/abc123`

Answer: C

A `runs:/` URI combines the run ID with the artifact path, which is how MLflow addresses a model logged inside a run. `models:/` URIs refer to the registry, and `experiments:/` and `artifacts:/` are not valid model URI schemes.

4. When Hyperopt runs with SparkTrials, each trial is evaluated on a Spark _____ rather than only on the driver.

Answer: worker (also accepted: worker node)

SparkTrials sends trial evaluations to the cluster so several can run at once across worker nodes. This is what makes distributed tuning faster than running trials one after another on the driver.

5. Which two actions are appropriate when promoting a model for production scoring? (Choose two.)
- A Move the champion alias before any evaluation, so the team can compare later.
 - B Score the candidate on a held-out dataset and compare its metrics with the current champion.
 - C Skip evaluation when the candidate was trained on more rows.
 - D Delete the previous model version so the endpoint cannot roll back.
 - E Keep a model signature and input example with the logged model so requests can be validated.

Answer: B, E

Comparing the candidate with the champion on held-out data is the evidence needed before promotion, and a signature with an input example lets consumers and endpoints validate requests. Moving the alias first, skipping evaluation and deleting versions each remove a safeguard.

6. A categorical column has 10,000 distinct values, which makes one-hot encoding impractical. What is a common alternative?
- A Drop the column before every training run and log the dropped values for audit
 - B Frequency or target encoding fitted on the training split only
 - C Replace each value with a random integer per row
 - D Convert the column to a timestamp in every row

Answer: B

Frequency and target encodings compress high-cardinality categories into a few numeric columns, and fitting them on the training split avoids leakage. Dropping the column throws away signal, and random integers carry no information about the category.

7. Which Unity Catalog privilege lets a user or service principal load and score a registered model?
- A MODIFY
 - B SELECT
 - C CREATE TABLE
 - D EXECUTE

Answer: D

Loading and scoring a registered model is governed by EXECUTE on the model object in Unity Catalog. MODIFY and CREATE TABLE allow changing data or objects, and SELECT applies to tables rather than registered models.

8. `mlflow.autolog()` logs the trained model artifact for supported frameworks without an explicit `log_model` call.
- A True
 - B False

Answer: A (True)

Autologging records parameters, metrics and, by default, the fitted model for supported frameworks. Teams can still turn model logging off in the autolog settings if they want only metrics.

9. A training set joins customer features to labels with a point-in-time lookup. What does point-in-time correctness guarantee?
- A Feature tables are copied to a backup before each run
 - B Only the newest feature value per customer is used for every row, so older values never train the model
 - C Each training row sees only feature values available at its label timestamp
 - D Labels are sorted by customer identifier before joining

Answer: C

A point-in-time lookup matches each label with the feature values recorded up to its timestamp, which prevents future information from entering training. Backups, using only the newest value, and sorting do not address that risk.

10. Which search distribution is most appropriate for a learning rate that varies over several orders of magnitude?
- A `hp.choice` over three fixed integer values
 - B `hp.loguniform` over a log-scaled range
 - C `hp.randint` between zero and one
 - D `hp.quniform` with a step of one

Answer: B

A log-uniform distribution samples evenly across orders of magnitude, so small and large learning rates both get coverage. Integer-based choices and steps do not suit a continuous parameter that spans decades.