

PRACTICE QUESTIONS

Databricks Certified Machine Learning Professional

SHARED BY [u/eagerheron1978](#) on aiexamcerts

EDITION	Free Set 1
---------	------------

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

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

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

Shared by u/eagerheron1978 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. Which call records the input and output schema of a model so that serving clients and the registry can validate requests?

- A `mlflow.set_experiment()` with a workspace path that also writes the schema
- B `mlflow.end_run()` with the FINISHED status
- C `mlflow.models.infer_signature()` passed to `log_model()` as the signature
- D `mlflow.log_param()` called once for each feature name

Answer: C

`infer_signature` builds a schema from sample inputs and outputs, and passing it to `log_model` stores that schema with the model. Setting the experiment, ending a run and logging parameters do not describe the model's inputs or outputs. Logging feature names as parameters records configuration, not a validated schema.

2. Enabling _____ on a Model Serving endpoint writes each request and response to a Delta table for later monitoring.

Answer: inference tables

Inference tables capture request and response payloads from a serving endpoint into Delta tables, which monitoring jobs can then analyze. They are the usual source for checking production drift and prediction quality.

3. Which Spark ML method, called on an Estimator, trains on a DataFrame and returns a fitted Model?

- A `Transformer.transform()`
- B `Evaluator.evaluate()`
- C `ParamGridBuilder.build()`
- D `Estimator.fit()`

Answer: D

Calling `fit` on an Estimator runs training and returns a Model, which is itself a Transformer. `transform` applies an already fitted model, `evaluate` computes a metric, and `ParamGridBuilder` only builds a list of parameter combinations.

4. Lakehouse Monitoring computes profile and drift metrics as Delta tables that you can query alongside the monitored table.
- A True
 - B False

Answer: A (True)

Lakehouse Monitoring writes profile metrics and drift metrics as Delta tables, so teams can query them with SQL and build dashboards or alerts. This is the mechanism behind monitoring a table without exporting data elsewhere.

5. Which two statements about Model Serving endpoints are correct? (Choose two.)
- A An endpoint can be configured to scale to zero when it is idle.
 - B An endpoint replaces the need to register the model in a model registry.
 - C Each endpoint must run on a Spark cluster that stays running at all times.
 - D An endpoint can read its model only from a workspace notebook folder.
 - E An endpoint can split traffic between several served entities by percentage.

Answer: A, E

Serving endpoints support scale to zero for idle capacity and can route traffic percentages across served entities, which enables safe rollouts. A registry is still the source of the model versions being served, and endpoints do not depend on an always-on cluster or a notebook folder.

6. A retraining job must rebuild the exact training data used by a run from last month. Which SQL clause reads a Delta table as it was at an earlier version?
- A `SELECT * FROM sales VERSION AS OF 42`
 - B `SELECT * FROM sales TABLESAMPLE (10 PERCENT)`
 - C `SELECT * FROM sales WHERE version = 42`
 - D `SELECT * FROM sales ROLLBACK TO 42`

Answer: A

Delta time travel uses `VERSION AS OF` or `TIMESTAMP AS OF` to read a historical snapshot of the table. `TABLESAMPLE` returns a random subset of current rows, and a version column filter reads only rows that happen to store a version value. `ROLLBACK TO` is not a read clause for historical data.

7. Which call returns the runs of one experiment whose validation RMSE is below 0.5?
- A `mlflow.get_run(run_id='rmse < 0.5')` returns every run with validation RMSE below 0.5
 - B `mlflow.search_runs(experiment_ids=[...], filter_string='metrics.rmse < 0.5')`
 - C `mlflow.log_metric('rmse', 0.5)`
 - D `mlflow.list_experiments(filter='rmse')`

Answer: B

`mlflow.search_runs` accepts a `filter_string` over metrics, params and tags, and returns a DataFrame of matching runs. `get_run` needs a run ID, `log_metric` writes a value, and `list_experiments` returns experiments rather than runs.

8. A Unity Catalog registered model is addressed with a three-level name of the form `catalog. ....model_name`.

Answer: schema (also accepted: `schema_name`)

Unity Catalog names are catalog, schema and object, so a registered model lives in a schema inside a catalog. The model name is the third level after the schema.

9. A churn feature counts support tickets filed in the 30 days after the label date. Why does this feature cause a problem?
- A It has too few distinct values to be useful
 - B It is categorical and needs one-hot encoding
 - C Spark cannot compute a count over a 30-day window on a tickets table
 - D It leaks future information that would not exist at prediction time

Answer: D

Tickets filed after the label date are future information, so a model trained with this feature would look accurate but could not use it when predicting. Distinct-value counts, encoding and window support are not the issue here.

10. A team wants a Model Serving endpoint to answer an internal application within seconds. Which step must come first?
- A Register the model version in a registry so the endpoint can reference it
 - B Export the model to ONNX format
 - C Store the model inside a long-running Spark cluster
 - D Create the endpoint with the training code pasted into its configuration before training

Answer: A

An endpoint serves a specific registered model version, so the model must be logged and registered first. ONNX export is one possible format but is not a prerequisite, and endpoints do not run on a dedicated Spark cluster.