

PRACTICE QUESTIONS

AWS

MLA-C01

AWS Certified Machine Learning Engineer - Associate

SHARED BY [u/notes_c01](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	AWS MLA-C01, AWS Certified Machine Learning Engineer - Associate

Shared by [u/notes_c01](#) on aiexamcerts

10 practice questions for the AWS MLA-C01 exam, AWS Certified Machine Learning Engineer - Associate, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/aws/mla-c01. Spotted a mistake? Email support@aiexamcerts.com with the exam code MLA-C01 and the question number.

1. Which Amazon SageMaker capability provides a visual interface with built-in transforms for preparing tabular data, and can export the resulting flow to a training pipeline?
- A SageMaker Model Monitor
 - B SageMaker Debugger
 - C SageMaker Data Wrangler
 - D SageMaker Clarify

Answer: C

Data Wrangler lets you build transformation flows visually, check data quality, and export the flow to a notebook, a processing job, or a pipeline. SageMaker Clarify is the tempting alternative because it also analyzes datasets, but its focus is bias and explainability. Model Monitor watches deployed endpoints and Debugger inspects training runs.

2. In Amazon SageMaker Feature Store, the _____ store serves low-latency feature lookups for real-time inference.

Answer: online (also accepted: online store)

The online store returns the latest feature values for an entity with low latency, which suits real-time requests. The offline store keeps feature history in S3 for building training datasets.

3. A team must find and classify personally identifiable information stored in Amazon S3 before that data is used for training. Which service fits this need?
- A Amazon Kinesis Data Streams
 - B Amazon FSx for Lustre
 - C Amazon Athena
 - D Amazon Macie

Answer: D

Macie uses machine learning and pattern matching to discover and classify sensitive data such as PII in S3 buckets. Athena is the tempting choice because it can search the same data, but it queries content without classifying what is sensitive. Kinesis and FSx serve ingestion and high-throughput file storage.

4. Amazon Data Firehose can deliver streaming records to Amazon S3 without you writing and operating consumer application code.
- A True
 - B False

Answer: A (True)

The statement is true. Firehose is a managed delivery service that buffers streaming records and writes them to destinations such as S3, which removes the need for a custom consumer application for that path.

5. Which two statements about SageMaker hyperparameter tuning jobs are correct? (Choose two.)
- A They run multiple training jobs and report the best one according to the objective metric
 - B They always require a GPU instance
 - C They require an objective metric that the tuner optimizes
 - D They replace the need for a validation set
 - E They update the model weights directly during tuning

Answer: A, C

A tuning job launches many training jobs and selects the best by the objective metric it was configured to optimize. The GPU requirement is a tempting assumption, since many tuned models use GPUs, but CPU instances work too. Tuning still depends on a validation signal, and it tunes settings rather than editing weights directly.

6. Which SageMaker Debugger capability helps detect training problems such as vanishing gradients or exploding tensors?
- A Model Registry approvals
 - B Batch Transform jobs
 - C Feature Store offline replication that copies feature history to a second bucket
 - D Built-in rules that evaluate the tensors collected during training

Answer: D

Debugger collects tensors during training and evaluates built-in rules against them, which can raise alerts when a problem appears. Batch Transform is the tempting pick for anything that runs on a model, but it performs offline inference and never sees training tensors. Registry approvals and offline feature replication do not inspect training at all.

7. Which SageMaker built-in algorithm is designed for tabular classification and regression using gradient-boosted decision trees?
- A XGBoost
 - B DeepAR forecasting
 - C Object Detection
 - D Semantic Segmentation

Answer: A

XGBoost is the SageMaker built-in implementation of gradient-boosted trees, which perform well on tabular data. DeepAR is the tempting choice when the data looks like a time series, but it targets forecasting across related series. Object Detection and Semantic Segmentation are computer vision algorithms.

8. SageMaker Automatic Model Tuning searches for the best combination of values for a training job's _____.

Answer: hyperparameters

A tuning job varies hyperparameters across many training jobs and optimizes an objective metric. Learned weights are parameters produced by training, which is a different thing from the settings the tuner changes.

9. A company hosts hundreds of small models that are called rarely. Which SageMaker option reduces the number of endpoints it must run?
- A One dedicated endpoint per model
 - B Copying each model onto its own EC2 instance
 - C A multi-model endpoint that loads model artifacts from S3 on demand
 - D Deleting the rarely used models after exporting their artifacts to an archive bucket

Answer: C

A multi-model endpoint shares instances across many model artifacts stored in S3 and loads them when they are requested. One endpoint per model is the tempting choice for isolation, but it multiplies idle cost. Deleting models removes functionality the company still needs.

10. Which approach best keeps inference dependencies consistent with the environment used for training?
- A Installing packages by hand on each instance, so every host can be tuned individually
 - B Pulling the latest unpinned base image at each deployment
 - C Using a notebook kernel as the serving runtime
 - D Building a versioned container image with pinned dependencies and using it for both

Answer: D

A versioned image with pinned dependencies gives training and serving the same known runtime. Pulling the latest unpinned image is the tempting way to stay current, but it can change behavior between runs. Manual installs and notebook kernels leave no reliable record of the environment.