

PRACTICE QUESTIONS

AWS

# AIP-C01

AWS Certified Generative AI Developer - Professional

SHARED BY [u/grinder\\_c01](#) on aiexamcerts

|           |                                                                   |
|-----------|-------------------------------------------------------------------|
| EDITION   | Free Set 2                                                        |
| QUESTIONS | 10                                                                |
| SHARED    | October 15, 2026                                                  |
| EXAM      | AWS AIP-C01, AWS Certified Generative AI Developer - Professional |

---

Shared by [u/grinder\\_c01](#) on aiexamcerts

10 practice questions for the AWS AIP-C01 exam, AWS Certified Generative AI Developer - Professional, each followed by its answer and a short explanation. More free sets and the full practice exam: [www.aiexamcerts.com/aws/aip-c01](http://www.aiexamcerts.com/aws/aip-c01). Spotted a mistake? Email [support@aiexamcerts.com](mailto:support@aiexamcerts.com) with the exam code AIP-C01 and the question number.

1. A chunk must carry its access level so that queries return only permitted content. What should be stored with each chunk?
  - A The raw document encryption key in plain text
  - B Only the document title, without any attributes
  - C Metadata fields for access attributes, used as retrieval filters
  - D A separate copy of the full document for every query and every user

**Answer: C**

Storing access attributes as metadata on each chunk lets retrieval filter results to what the requester may see. Keys in plain text are a security failure, even though they look convenient for lookups. A title alone cannot enforce permissions, and copying documents per query wastes resources without enforcing access.

2. A Bedrock knowledge base needs its documents converted into a fine-tuning dataset before they can be indexed.
  - A True
  - B False

**Answer: B (False)**

Knowledge bases ingest documents directly, chunk them, and create embeddings, so no fine-tuning dataset is needed. Preparing training pairs is tempting when a team hears about customization, but it solves a different problem. The statement is false.

3. A document pipeline needs retries, parallel branches, and a human approval step that can wait for hours. Which service fits best?
  - A AWS Step Functions state machine
  - B A single Lambda function with a long sleep loop
  - C A cron job on an EC2 instance that writes to a text file
  - D A chain of API Gateway REST calls with no shared state

**Answer: A**

Step Functions provides retry policies, parallel branches, and task tokens that let a workflow wait for an external approval. A single Lambda with a sleep loop is tempting because it is simple, but it runs into runtime limits and has no built-in retry model. A cron job and stateless REST chains have no workflow state.

4. The Step Functions state that runs the same steps over every item in an array is called a \_\_\_\_\_ state.

**Answer: Map**

A Map state iterates over an array and can process items in parallel, which fits batch fan-out. A Parallel state is tempting because it also runs branches concurrently, but it runs a fixed set of branches rather than one branch per item. Choice states decide between paths.

---

5. Which two techniques reduce token usage without hurting a summarization workflow's answers? (Choose two.)
- A Sending the full document history on every turn
  - B Doubling the number of retrieved chunks
  - C Removing duplicate passages before they enter the prompt
  - D Setting the temperature to its maximum
  - E Asking for a summary within a word limit that matches the use case

**Answer: C, E**

Removing duplicate passages and asking for a length-bounded summary both reduce tokens while keeping the needed content. Sending full history on every turn is tempting for context, but it grows cost quickly. Doubling chunks and raising temperature do not reduce usage.

---

6. Which AWS service analyzes logs and signals to detect threats such as suspicious API activity and compromised credentials?
- A Amazon GuardDuty
  - B Amazon Macie
  - C AWS Config
  - D AWS Audit Manager

**Answer: A**

GuardDuty analyzes logs and other signals to detect threats such as suspicious API activity and compromised credentials. Macie is tempting because it is also a security service, but it discovers sensitive data in S3. AWS Config tracks configuration and Audit Manager gathers audit evidence.

---

7. Which control best limits the damage a prompt-injected agent can do when asked to delete customer records?
- A Raising the temperature so the agent refuses more often
  - B A role scoped to read-only permissions, with confirmation for destructive actions
  - C Adding more deletion examples to the prompt so the agent recognizes the request as normal
  - D Removing the guardrail so the agent can reason freely

**Answer: B**

Least-privilege permissions limit what a manipulated agent can reach, and confirmation steps catch destructive actions. Raising temperature is tempting because it changes behavior, but it adds randomness rather than enforcing a refusal. Adding deletion examples and removing guardrails make the risk worse.

8. A passing offline evaluation guarantees that a prompt change will perform well for every production user.
- A True
  - B False

**Answer: B (False)**

An offline test set covers only the cases someone chose, so production traffic can reveal failures the set missed. That is why monitoring after release matters. The statement is false. Treating a passing test run as proof is the tempting error.

9. An assistant gets many simple FAQ questions and a few complex ones. Which practice best reduces its cost?
- A Route simple questions to a smaller model and complex ones to a larger model
  - B Sending every question to the largest model for consistent quality and tone everywhere
  - C Disabling the FAQ so that all questions use search
  - D Caching nothing and retrying every failed call three times

**Answer: A**

Model routing matches each request to the least expensive model that can answer it well, which lowers average cost. Using the largest model everywhere is tempting for consistency, but it inflates cost on simple questions. Disabling the FAQ and retrying every failure both increase cost.

10. Which deployment practice limits the risk of a new prompt version by sending a small, weighted share of traffic to it first?
- A Big-bang cutover of all users at once
  - B Canary release with weighted traffic
  - C Deleting the old version before the new one is tested
  - D Disabling monitoring during the first hour

**Answer: B**

A canary release routes a small share of traffic to the new version and watches metrics before expanding it. A big-bang cutover is tempting for simplicity, but it exposes every user to untested behavior. Deleting the old version and disabling monitoring remove both the fallback and the signal.