

# New Databricks Databricks-Generative-AI-Engineer-Associate Dumps Files | Databricks-Generative-AI-Engineer-Associate Exam Practice



BTW, DOWNLOAD part of Easy4Engine Databricks-Generative-AI-Engineer-Associate dumps from Cloud Storage:  
<https://drive.google.com/open?id=1Wlh3G2Mj--O-fO1u8fHpeeJOcjx4X7o>

Appropriately, we can wrap up this post with the way that the test centers around the material that is essential to handily clear your Databricks Certified Generative AI Engineer Associate certification exam. You can trust the material and set aside an edge to zero in on those before you win eventually over the last Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam dates. To get it, find the source that assists you with getting the right test and spotlight on material agreeable for you for organizing the Databricks Certified Generative AI Engineer Associate exam.

If you are preparing for the exam in order to get the related certification, here comes a piece of good news for you. The Databricks-Generative-AI-Engineer-Associate guide torrent is compiled by our company now has been praised as the secret weapon for candidates who want to pass the Databricks-Generative-AI-Engineer-Associate exam as well as getting the related certification, so you are so lucky to click into this website where you can get your secret weapon. Our reputation for compiling the best Databricks-Generative-AI-Engineer-Associate Training Materials has created a sound base for our future business. We are clearly focused on the international high-end market, thereby committing our resources to the specific product requirements of this key market sector. There are so many advantages of our Databricks-Generative-AI-Engineer-Associate exam torrent, and now, I would like to introduce some details about our Databricks-Generative-AI-Engineer-Associate guide torrent for your reference.

>> New Databricks Databricks-Generative-AI-Engineer-Associate Dumps Files <<

## Databricks Databricks-Generative-AI-Engineer-Associate Exam Practice & Databricks-Generative-AI-Engineer-Associate PDF Questions

With the best quality of Databricks-Generative-AI-Engineer-Associate braindumps pdf from our website, getting certified will be easier and fast. For the preparation of the certification exam, all you have to do is choose the most reliable Databricks-Generative-AI-Engineer-Associate real questions and follow our latest study guide. You can completely rest assured that our Databricks-Generative-AI-Engineer-Associate Dumps Collection will ensure you get high mark in the formal test. You will get lots of knowledge

from our website.

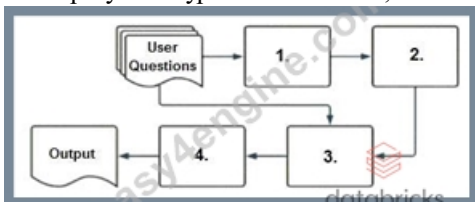
## Databricks Databricks-Generative-AI-Engineer-Associate Exam Syllabus Topics:

Topic	Details
Topic 1	<ul style="list-style-type: none"><li>• Application Development: In this topic, Generative AI Engineers learn about tools needed to extract data, Langchain</li><li>• similar tools, and assessing responses to identify common issues. Moreover, the topic includes questions about adjusting an LLM's response, LLM guardrails, and the best LLM based on the attributes of the application.</li></ul>
Topic 2	<ul style="list-style-type: none"><li>• Governance: Generative AI Engineers who take the exam get knowledge about masking techniques, guardrail techniques, and legal</li><li>• licensing requirements in this topic.</li></ul>
Topic 3	<ul style="list-style-type: none"><li>• Assembling and Deploying Applications: In this topic, Generative AI Engineers get knowledge about coding a chain using a pyfunc mode, coding a simple chain using langchain, and coding a simple chain according to requirements. Additionally, the topic focuses on basic elements needed to create a RAG application. Lastly, the topic addresses sub-topics about registering the model to Unity Catalog using MLflow.</li></ul>
Topic 4	<ul style="list-style-type: none"><li>• Data Preparation: Generative AI Engineers covers a chunking strategy for a given document structure and model constraints. The topic also focuses on filter extraneous content in source documents. Lastly, Generative AI Engineers also learn about extracting document content from provided source data and format.</li></ul>

## Databricks Certified Generative AI Engineer Associate Sample Questions (Q51-Q56):

### NEW QUESTION # 51

A company has a typical RAG-enabled, customer-facing chatbot on its website.



Select the correct sequence of components a user's questions will go through before the final output is returned. Use the diagram above for reference.

- A. 1.embedding model, 2.vector search, 3.context-augmented prompt, 4.response-generating LLM
- B. 1.context-augmented prompt, 2.vector search, 3.embedding model, 4.response-generating LLM
- C. 1.response-generating LLM, 2.vector search, 3.context-augmented prompt, 4.embedding model
- D. 1.response-generating LLM, 2.context-augmented prompt, 3.vector search, 4.embedding model

**Answer: A**

Explanation:

To understand how a typical RAG-enabled customer-facing chatbot processes a user's question, let's go through the correct sequence as depicted in the diagram and explained in option A:

\* Embedding Model (1):The first step involves the user's question being processed through an embedding model. This model converts the text into a vector format that numerically represents the text. This step is essential for allowing the subsequent vector search to operate effectively.

\* Vector Search (2):The vectors generated by the embedding model are then used in a vector search mechanism. This search identifies the most relevant documents or previously answered questions that are stored in a vector format in a database.

\* Context-Augmented Prompt (3):The information retrieved from the vector search is used to create a context-augmented prompt.

This step involves enhancing the basic user query with additional relevant information gathered to ensure the generated response is as accurate and informative as possible.

\* Response-Generating LLM (4): Finally, the context-augmented prompt is fed into a response- generating large language model (LLM). This LLM uses the prompt to generate a coherent and contextually appropriate answer, which is then delivered as the final output to the user.

Why Other Options Are Less Suitable:

\* B, C, D: These options suggest incorrect sequences that do not align with how a RAG system typically processes queries. They misplace the role of embedding models, vector search, and response generation in an order that would not facilitate effective information retrieval and response generation.

Thus, the correct sequence is embedding model, vector search, context-augmented prompt, response- generating LLM, which is option A.

## NEW QUESTION # 52

A Generative AI Engineer is setting up a Databricks Vector Search that will lookup news articles by topic within 10 days of the date specified. An example query might be "Tell me about monster truck news around January 5th 1992". They want to do this with the least amount of effort.

How can they set up their Vector Search index to support this use case?

- A. pass the query directly to the vector search index and return the best articles.
- **B. Include metadata columns for article date and topic to support metadata filtering.**
- C. Create separate indexes by topic and add a classifier model to appropriately pick the best index.
- D. Split articles by 10 day blocks and return the block closest to the query.

**Answer: B**

Explanation:

The task is to set up a Databricks Vector Search index for news articles, supporting queries like "monster truck news around January 5th, 1992," with minimal effort. The index must filter by topic and a 10-day date range. Let's evaluate the options.

\* Option A: Split articles by 10-day blocks and return the block closest to the query

\* Pre-splitting articles into 10-day blocks requires significant preprocessing and index management (e.g., one index per block). It's effort-intensive and inflexible for dynamic date ranges.

\* Databricks Reference: "Static partitioning increases setup complexity; metadata filtering is preferred" ("Databricks Vector Search Documentation").

\* Option B: Include metadata columns for article date and topic to support metadata filtering

\* Adding date and topic as metadata in the Vector Search index allows dynamic filtering (e.g., date

± 5 days, topic = "monster truck") at query time. This leverages Databricks' built-in metadata filtering, minimizing setup effort.

\* Databricks Reference: "Vector Search supports metadata filtering on columns like date or category for precise retrieval with minimal preprocessing" ("Vector Search Guide," 2023).

\* Option C: Pass the query directly to the vector search index and return the best articles

\* Passing the full query (e.g., "Tell me about monster truck news around January 5th, 1992") to Vector Search relies solely on embeddings, ignoring structured filtering for date and topic. This risks inaccurate results without explicit range logic.

\* Databricks Reference: "Pure vector similarity may not handle temporal or categorical constraints effectively" ("Building LLM Applications with Databricks").

\* Option D: Create separate indexes by topic and add a classifier model to appropriately pick the best index

\* Separate indexes per topic plus a classifier model adds significant complexity (index creation, model training, maintenance), far exceeding "least effort." It's overkill for this use case.

\* Databricks Reference: "Multiple indexes increase overhead; single-index with metadata is simpler" ("Databricks Vector Search Documentation").

Conclusion: Option B is the simplest and most effective solution, using metadata filtering in a single Vector Search index to handle date ranges and topics, aligning with Databricks' emphasis on efficient, low-effort setups.

## NEW QUESTION # 53

A Generative AI Engineer interfaces with an LLM with prompt/response behavior that has been trained on customer calls inquiring about product availability. The LLM is designed to output "In Stock" if the product is available or only the term "Out of Stock" if not. Which prompt will work to allow the engineer to respond to call classification labels correctly?

- A. Respond with "Out of Stock" if the customer asks for a product.
- B. You will be given a customer call transcript where the customer inquires about product availability. Respond with "In

Stock" if the product is available or "Out of Stock" if not.

- C. Respond with "In Stock" if the customer asks for a product.
- **D. You will be given a customer call transcript where the customer asks about product availability. The outputs are either "In Stock" or "Out of Stock". Format the output in JSON, for example: {"call\_id": "123", "label": "In Stock"}.**

**Answer: D**

Explanation:

\* Problem Context: The Generative AI Engineer needs a prompt that will enable an LLM trained on customer call transcripts to classify and respond correctly regarding product availability. The desired response should clearly indicate whether a product is "In Stock" or "Out of Stock," and it should be formatted in a way that is structured and easy to parse programmatically, such as JSON.

\* Explanation of Options:

\* Option A: Respond with "In Stock" if the customer asks for a product. This prompt is too generic and does not specify how to handle the case when a product is not available, nor does it provide a structured output format.

\* Option B: This option is correctly formatted and explicit. It instructs the LLM to respond based on the availability mentioned in the customer call transcript and to format the response in JSON.

This structure allows for easy integration into systems that may need to process this information automatically, such as customer service dashboards or databases.

\* Option C: Respond with "Out of Stock" if the customer asks for a product. Like option A, this prompt is also insufficient as it only covers the scenario where a product is unavailable and does not provide a structured output.

\* Option D: While this prompt correctly specifies how to respond based on product availability, it lacks the structured output format, making it less suitable for systems that require formatted data for further processing.

Given the requirements for clear, programmatically usable outputs, Option B is the optimal choice because it provides precise instructions on how to respond and includes a JSON format example for structuring the output, which is ideal for automated systems or further data handling.

#### NEW QUESTION # 54

A Generative AI Engineer is using an LLM to classify species of edible mushrooms based on text descriptions of certain features.

The model is returning accurate responses in testing and the Generative AI Engineer is confident they have the correct list of possible labels, but the output frequently contains additional reasoning in the answer when the Generative AI Engineer only wants to return the label with no additional text.

Which action should they take to elicit the desired behavior from this LLM?

- A. Use few shot prompting to instruct the model on expected output format
- B. Use zero shot prompting to instruct the model on expected output format
- **C. Use a system prompt to instruct the model to be succinct in its answer**
- D. Use zero shot chain-of-thought prompting to prevent a verbose output format

**Answer: C**

Explanation:

The LLM classifies mushroom species accurately but includes unwanted reasoning text, and the engineer wants only the label. Let's assess how to control output format effectively.

\* Option A: Use few shot prompting to instruct the model on expected output format

\* Few-shot prompting provides examples (e.g., input: description, output: label). It can work but requires crafting multiple examples, which is effort-intensive and less direct than a clear instruction.

\* Databricks Reference: "Few-shot prompting guides LLMs via examples, effective for format control but requires careful design" ("Generative AI Cookbook").

\* Option B: Use zero shot prompting to instruct the model on expected output format

\* Zero-shot prompting relies on a single instruction (e.g., "Return only the label") without examples. It's simpler than few-shot but may not consistently enforce succinctness if the LLM's default behavior is verbose.

\* Databricks Reference: "Zero-shot prompting can specify output but may lack precision without examples" ("Building LLM Applications with Databricks").

\* Option C: Use zero shot chain-of-thought prompting to prevent a verbose output format

\* Chain-of-Thought (CoT) encourages step-by-step reasoning, which increases verbosity-opposite to the desired outcome. This contradicts the goal of label-only output.

\* Databricks Reference: "CoT prompting enhances reasoning but often results in detailed responses" ("Databricks Generative AI Engineer Guide").

\* Option D: Use a system prompt to instruct the model to be succinct in its answer

\* A system prompt (e.g., "Respond with only the species label, no additional text") sets a global instruction for the LLM's behavior. It's direct, reusable, and effective for controlling output style across queries.

\* Databricks Reference: "System prompts define LLM behavior consistently, ideal for enforcing concise outputs" ("Generative AI Cookbook," 2023).

Conclusion: Option D is the most effective and straightforward action, using a system prompt to enforce succinct, label-only responses, aligning with Databricks' best practices for output control.

### NEW QUESTION # 55

A Generative AI Engineer is creating an LLM-based application. The documents for its retriever have been chunked to a maximum of 512 tokens each. The Generative AI Engineer knows that cost and latency are more important than quality for this application. They have several context length levels to choose from.

Which will fulfill their need?

- A. context length 32768: smallest model is 14GB and embedding dimension 4096
- B. context length 514; smallest model is 0.44GB and embedding dimension 768
- **C. context length 512: smallest model is 0.13GB and embedding dimension 384**
- D. context length 2048: smallest model is 11GB and embedding dimension 2560

**Answer: C**

Explanation:

When prioritizing cost and latency over quality in a Large Language Model (LLM)-based application, it is crucial to select a configuration that minimizes both computational resources and latency while still providing reasonable performance. Here's why Dis is the best choice:

\* Context length: The context length of 512 tokens aligns with the chunk size used for the documents (maximum of 512 tokens per chunk). This is sufficient for capturing the needed information and generating responses without unnecessary overhead.

\* Smallest model size: The model with a size of 0.13GB is significantly smaller than the other options.

This small footprint ensures faster inference times and lower memory usage, which directly reduces both latency and cost.

\* Embedding dimension: While the embedding dimension of 384 is smaller than the other options, it is still adequate for tasks where cost and speed are more important than precision and depth of understanding.

This setup achieves the desired balance between cost-efficiency and reasonable performance in a latency- sensitive, cost-conscious application.

### NEW QUESTION # 56

.....

If you like to practice Databricks-Generative-AI-Engineer-Associate exam dumps on paper, you should choose us. Our Databricks-Generative-AI-Engineer-Associate PDF version is printable, and you can print them into hard one and take some notes on them. Therefore you can study in anytime and at anyplace. Besides, free demo is available for Databricks-Generative-AI-Engineer-Associate PDF version, and you can have a try before buying. After your payment, you can receive the downloading link and password for Databricks-Generative-AI-Engineer-Associate Exam Dumps within ten minutes, and if you don't receive, you can contact us, we will solve the problem for you as quickly as possible.

**Databricks-Generative-AI-Engineer-Associate Exam Practice:** <https://www.easy4engine.com/Databricks-Generative-AI-Engineer-Associate-test-engine.html>

- Excellent New Databricks-Generative-AI-Engineer-Associate Dumps Files by [www.testkingpass.com](http://www.testkingpass.com) ☐ ► [www.testkingpass.com](http://www.testkingpass.com) ◀ is best website to obtain ► Databricks-Generative-AI-Engineer-Associate ☐ for free download ☐ Databricks-Generative-AI-Engineer-Associate Reliable Dumps Ppt
- Databricks-Generative-AI-Engineer-Associate Latest Dumps ☐ Databricks-Generative-AI-Engineer-Associate Latest Dumps ☐ Databricks-Generative-AI-Engineer-Associate Minimum Pass Score ☐ Download ► Databricks-Generative-AI-Engineer-Associate ◀ for free by simply searching on 「 [www.pdfvce.com](http://www.pdfvce.com) 」 ☐ Databricks-Generative-AI-Engineer-Associate Latest Dumps
- Free PDF Quiz 2026 Latest Databricks-Generative-AI-Engineer-Associate: New Databricks Certified Generative AI Engineer Associate Dumps Files ☐ Easily obtain ► Databricks-Generative-AI-Engineer-Associate ☐ for free download through ⇒ [www.easy4engine.com](http://www.easy4engine.com) ⇐ ☐ Databricks-Generative-AI-Engineer-Associate Test Dumps Demo
- Databricks-Generative-AI-Engineer-Associate Top Dumps ☐ Databricks-Generative-AI-Engineer-Associate Latest Dumps Pdf ☐ Databricks-Generative-AI-Engineer-Associate Latest Dumps Pdf ☐ Open [ [www.pdfvce.com](http://www.pdfvce.com) ] and search for 《 Databricks-Generative-AI-Engineer-Associate 》 to download exam materials for free ☐ New Databricks-



## Test Databricks-Generative-AI-Engineer-Associate

- 91                 /      91                 /      1                 1