

Free PDF Databricks - Newest Databricks-Generative-AI-Engineer-Associate Valid Vce



BONUS!!! Download part of Itcertmaster Databricks-Generative-AI-Engineer-Associate dumps for free:
<https://drive.google.com/open?id=1Q7VXeCcohMBz5MHm4tT4sj0VdFNerSD>

We trounce many peers in this industry by our justifiably excellent Databricks-Generative-AI-Engineer-Associate training guide and considerate services. So our Databricks-Generative-AI-Engineer-Associate exam prep receives a tremendous ovation in market over twenty years. All these years, we have helped tens of thousands of exam candidates achieve success greatly. For all content of our Databricks-Generative-AI-Engineer-Associate Learning Materials are strictly written and tested by our customers as well as the market. Come to try and you will be satisfied!

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

Topic	Details
Topic 1	<ul style="list-style-type: none">• Governance: Generative AI Engineers who take the exam get knowledge about masking techniques, guardrail techniques, and legal• licensing requirements in this topic.
Topic 2	<ul style="list-style-type: none">• Evaluation and Monitoring: This topic is all about selecting an LLM choice and key metrics. Moreover, Generative AI Engineers learn about evaluating model performance. Lastly, the topic includes sub-topics about inference logging and usage of Databricks features.
Topic 3	<ul style="list-style-type: none">• Design Applications: The topic focuses on designing a prompt that elicits a specifically formatted response. It also focuses on selecting model tasks to accomplish a given business requirement. Lastly, the topic covers chain components for a desired model input and output.

Topic 4	<ul style="list-style-type: none"> • 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.
---------	--

>> Databricks-Generative-AI-Engineer-Associate Valid Vce <<

Databricks-Generative-AI-Engineer-Associate Latest Test Camp, Databricks-Generative-AI-Engineer-Associate Latest Exam Notes

Nobody wants to be stranded in the same position in his or her company. And nobody wants to be a normal person forever. Maybe you want to get the Databricks-Generative-AI-Engineer-Associate certification, but daily work and long-time traffic make you busier to improve yourself. However, there is a piece of good news for you. Thanks to our Databricks-Generative-AI-Engineer-Associate Training Materials, you can learn for your Databricks-Generative-AI-Engineer-Associate certification anytime, everywhere. And you will be bound to pass the exam with our Databricks-Generative-AI-Engineer-Associate exam questions.

Databricks Certified Generative AI Engineer Associate Sample Questions (Q15-Q20):

NEW QUESTION # 15

After changing the response generating LLM in a RAG pipeline from GPT-4 to a model with a shorter context length that the company self-hosts, the Generative AI Engineer is getting the following error:

```
{"error_code": "BAD_REQUEST", "message": "Bad request: rpc error: code = InvalidArgument desc = prompt token count (4595) cannot exceed 4096..."}  
+certmaste...COV
```

What TWO solutions should the Generative AI Engineer implement without changing the response generating model? (Choose two.)

- A. Use a smaller embedding model to generate
- B. Retrain the response generating model using ALiBi
- C. Reduce the number of records retrieved from the vector database
- D. Decrease the chunk size of embedded documents
- E. Reduce the maximum output tokens of the new model

Answer: C,D

Explanation:

* Problem Context: After switching to a model with a shorter context length, the error message indicating that the prompt token count has exceeded the limit suggests that the input to the model is too large.

* Explanation of Options:

* Option A: Use a smaller embedding model to generate- This wouldn't necessarily address the issue of prompt size exceeding the model's token limit.

* Option B: Reduce the maximum output tokens of the new model- This option affects the output length, not the size of the input being too large.

* Option C: Decrease the chunk size of embedded documents- This would help reduce the size of each document chunk fed into the model, ensuring that the input remains within the model's context length limitations.

* Option D: Reduce the number of records retrieved from the vector database- By retrieving fewer records, the total input size to the model can be managed more effectively, keeping it within the allowable token limits.

* Option E: Retrain the response generating model using ALiBi- Retraining the model is contrary to the stipulation not to change the response generating model.

Options C and D are the most effective solutions to manage the model's shorter context length without changing the model itself, by adjusting the input size both in terms of individual document size and total documents retrieved.

NEW QUESTION # 16

A team wants to serve a code generation model as an assistant for their software developers. It should support multiple programming languages. Quality is the primary objective.

Which of the Databricks Foundation Model APIs, or models available in the Marketplace, would be the best fit?

- A. MPT-7b
- B. Llama2-70b
- C. BGE-large
- D. **CodeLlama-34B**

Answer: D

Explanation:

For a code generation model that supports multiple programming languages and where quality is the primary objective, CodeLlama-34B is the most suitable choice. Here's the reasoning:

* Specialization in Code Generation: CodeLlama-34B is specifically designed for code generation tasks.

This model has been trained with a focus on understanding and generating code, which makes it particularly adept at handling various programming languages and coding contexts.

* Capacity and Performance: The "34B" indicates a model size of 34 billion parameters, suggesting a high capacity for handling complex tasks and generating high-quality outputs. The large model size typically correlates with better understanding and generation capabilities in diverse scenarios.

* Suitability for Development Teams: Given that the model is optimized for code, it will be able to assist software developers more effectively than general-purpose models. It understands coding syntax, semantics, and the nuances of different programming languages.

* Why Other Options Are Less Suitable:

* A (Llama2-70b): While also a large model, it's more general-purpose and may not be as fine-tuned for code generation as CodeLlama.

* B (BGE-large): This model may not specifically focus on code generation.

* C (MPT-7b): Smaller than CodeLlama-34B and likely less capable in handling complex code generation tasks at high quality.

Therefore, for a high-quality, multi-language code generation application, CodeLlama-34B (option D) is the best fit.

NEW QUESTION # 17

A Generative AI Engineer is using the code below to test setting up a vector store:

```
from databricks.vector_search.client import VectorSearchClient
vsc = VectorSearchClient()
vsc.create_endpoint(
    name= "vector_search_test",
    endpoint_type= "STANDARD"
)
```

Assuming they intend to use Databricks managed embeddings with the default embedding model, what should be the next logical function call?

- A. vsc.create_direct_access_index()
- B. **vsc.create_delta_sync_index()**
- C. vsc.get_index()
- D. vsc.similarity_search()

Answer: B

Explanation:

Context: The Generative AI Engineer is setting up a vector store using Databricks' VectorSearchClient. This is typically done to enable fast and efficient retrieval of vectorized data for tasks like similarity searches.

Explanation of Options:

* Option A: vsc.get_index(): This function would be used to retrieve an existing index, not create one, so it would not be the logical next step immediately after creating an endpoint.

* Option B: vsc.create_delta_sync_index(): After setting up a vector store endpoint, creating an index is necessary to start populating and organizing the data. The create_delta_sync_index() function specifically creates an index that synchronizes with a Delta table, allowing automatic updates as the data changes. This is likely the most appropriate choice if the engineer plans to use dynamic data that is updated over time.

* Option C: vsc.create_direct_access_index(): This function would create an index that directly accesses the data without synchronization. While also a valid approach, it's less likely to be the next logical step if the default setup (typically accommodating

changes) is intended.

* Option D: `vsc.similarity_search()`: This function would be used to perform searches on an existing index; however, an index needs to be created and populated with data before any search can be conducted.

Given the typical workflow in setting up a vector store, the next step after creating an endpoint is to establish an index, particularly one that synchronizes with ongoing data updates, hence Option B.

NEW QUESTION # 18

A Generative AI Engineer has already trained an LLM on Databricks and it is now ready to be deployed.

Which of the following steps correctly outlines the easiest process for deploying a model on Databricks?

- A. Wrap the LLM's prediction function into a Flask application and serve using Gunicorn
- B. Log the model as a pickle object, upload the object to Unity Catalog Volume, register it to Unity Catalog using MLflow, and start a serving endpoint
- C. **Log the model using MLflow during training, directly register the model to Unity Catalog using the MLflow API, and start a serving endpoint**
- D. Save the model along with its dependencies in a local directory, build the Docker image, and run the Docker container

Answer: C

NEW QUESTION # 19

A Generative AI Engineer is developing a RAG application and would like to experiment with different embedding models to improve the application performance.

Which strategy for picking an embedding model should they choose?

- A. Pick the most recent and most performant open LLM released at the time
- **B. Pick an embedding model trained on related domain knowledge**
- C. pick the embedding model ranked highest on the Massive Text Embedding Benchmark (MTEB) leaderboard hosted by HuggingFace
- D. Pick an embedding model with multilingual support to support potential multilingual user questions

Answer: B

Explanation:

The task involves improving a Retrieval-Augmented Generation (RAG) application's performance by experimenting with embedding models. The choice of embedding model impacts retrieval accuracy, which is critical for RAG systems. Let's evaluate the options based on Databricks Generative AI Engineer best practices.

* Option A: Pick an embedding model trained on related domain knowledge

* Embedding models trained on domain-specific data (e.g., industry-specific corpora) produce vectors that better capture the semantics of the application's context, improving retrieval relevance. For RAG, this is a key strategy to enhance performance.

* Databricks Reference: "For optimal retrieval in RAG systems, select embedding models aligned with the domain of your data" ("Building LLM Applications with Databricks," 2023).

* Option B: Pick the most recent and most performant open LLM released at the time

* LLMs are not embedding models; they generate text, not embeddings for retrieval. While recent LLMs may be performant for generation, this doesn't address the embedding step in RAG. This option misunderstands the component being selected.

* Databricks Reference: Embedding models and LLMs are distinct in RAG workflows:

"Embedding models convert text to vectors, while LLMs generate responses" ("Generative AI Cookbook").

* Option C: Pick the embedding model ranked highest on the Massive Text Embedding Benchmark (MTEB) leaderboard hosted by HuggingFace

* The MTEB leaderboard ranks models across general tasks, but high overall performance doesn't guarantee suitability for a specific domain. A top-ranked model might excel in generic contexts but underperform on the engineer's unique data.

* Databricks Reference: General performance is less critical than domain fit. "Benchmark rankings provide a starting point, but domain-specific evaluation is recommended" ("Databricks Generative AI Engineer Guide").

* Option D: Pick an embedding model with multilingual support to support potential multilingual user questions

* Multilingual support is useful only if the application explicitly requires it. Without evidence of multilingual needs, this adds complexity without guaranteed performance gains for the current use case.

* Databricks Reference: "Choose features like multilingual support based on application requirements" ("Building LLM-Powered Applications").

Conclusion: Option A is the best strategy because it prioritizes domain relevance, directly improving retrieval accuracy in a RAG system-aligning with Databricks' emphasis on tailoring models to specific use cases.

NEW QUESTION # 20

In order to help you enjoy the best learning experience, our PDF Databricks-Generative-AI-Engineer-Associate study guide supports you download on your computers and print on papers. In this way, you can make the best use of your spare time. Whatever you are occupied with your work, as long as you really want to learn our Databricks-Generative-AI-Engineer-Associate test engine, you must be inspired by your interests and motivation. Once you print all the contents of our Databricks-Generative-AI-Engineer-Associate Practice Test on the paper, you will find what you need to study is not as difficult as you imagined before. Also, you can make notes on your papers to help you memorize and understand the difficult parts. Maybe you are just scared by yourself. Getting the Databricks-Generative-AI-Engineer-Associate certificate is easy with the help of our test engine. You should seize the opportunities of passing the exam.

Databricks-Generative-AI-Engineer-Associate Latest Test Camp: <https://www.itcertmaster.com/Databricks-Generative-AI-Engineer-Associate.html>

myportal.utt.edu.tt, www.stes.tyc.edu.tw, mpgimer.edu.in, Disposable vapes

P.S. Free & New Databricks-Generative-AI-Engineer-Associate dumps are available on Google Drive shared by Itcertmaster:
<https://drive.google.com/open?id=1Q7VXeCcohMBz5MHm4fT4sj0VdFNerSD>