

Databricks-Generative-AI-Engineer-Associate Lernhilfe & Databricks-Generative-AI-Engineer-Associate Praxisprüfung



2026 Die neuesten Zertpruefung Databricks-Generative-AI-Engineer-Associate PDF-Versionen Prüfungsfragen und Databricks-Generative-AI-Engineer-Associate Fragen und Antworten sind kostenlos verfügbar: <https://drive.google.com/open?id=1JmR99xRsQo-cpeIsM5CIaL5WFtT8OX9>

Wenn Sie Online-Service für die Lerntipps zur Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung kaufen wollen, ist unser Zertpruefung einer der anführenden Websites. Wir bieten die neuesten Schulungsunterlagen von bester Qualität. Alle Lernmaterialien und Schulungsunterlagen zur Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung auf unserer Website entsprechen ihren Kosten. Sie genießen einen einjährigen kostenlosen Update-Service. Wenn alle unseren Produkte Ihnen nicht zum Bestehen der Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung Prüfung verhilft, erstatten wir Ihnen die gesammte Summe zurück.

Databricks Databricks-Generative-AI-Engineer-Associate Prüfungsplan:

Thema	Einzelheiten
Thema 1	<ul style="list-style-type: none"> 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.
Thema 2	<ul style="list-style-type: none"> Application Development: In this topic, Generative AI Engineers learn about tools needed to extract data, Langchain 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.

Thema 3	<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.
Thema 4	<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.
Thema 5	<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 Lernhilfe <<

Databricks Databricks-Generative-AI-Engineer-Associate Quiz - Databricks-Generative-AI-Engineer-Associate Studienanleitung & Databricks-Generative-AI-Engineer-Associate Trainingsmaterialien

Das IT-Expertenteam von Zertprüfung haben eine kurzfristige Schulungsmethode nach ihren Kenntnissen und Erfahrungen bearbeitet. Diese Dumps könne Ihnen effektiv helfen, in kurzer Zeit den erwarteten Effekt zu erzielen, besonders für diejenigen, die arbeiten und zugleich lernen. Zertprüfung kann Ihnen viel Zeit und Energir ersparen. Wählen Sie Zertprüfung und Sie werden Ihre wünschsten Schulungsmaterialien zur Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung bekommen.

Databricks Certified Generative AI Engineer Associate Databricks-Generative-AI-Engineer-Associate Prüfungsfragen mit Lösungen (Q67-Q72):

67. Frage

A Generative AI Engineer has a provisioned throughput model serving endpoint as part of a RAG application and would like to monitor the serving endpoint's incoming requests and outgoing responses. The current approach is to include a micro-service in between the endpoint and the user interface to write logs to a remote server.

Which Databricks feature should they use instead which will perform the same task?

- A. DBSQL
- B. Inference Tables
- C. Vector Search
- D. Lakeview

Antwort: B

Begründung:

Problem Context: The goal is to monitor the serving endpoint for incoming requests and outgoing responses in a provisioned throughput model serving endpoint within a Retrieval-Augmented Generation (RAG) application. The current approach involves using a microservice to log requests and responses to a remote server, but the Generative AI Engineer is looking for a more streamlined solution within Databricks.

Explanation of Options:

* Option A: Vector Search: This feature is used to perform similarity searches within vector databases.

It doesn't provide functionality for logging or monitoring requests and responses in a serving endpoint, so it's not applicable here.

* Option B: Lakeview: Lakeview is not a feature relevant to monitoring or logging request-response cycles for serving endpoints. It might be more related to viewing data in Databricks Lakehouse but doesn't fulfill the specific monitoring requirement.

* Option C: DBSQL: Databricks SQL (DBSQL) is used for running SQL queries on data stored in Databricks, primarily for analytics purposes. It doesn't provide the direct functionality needed to monitor requests and responses in real-time for an inference endpoint.

* Option D: Inference Tables: This is the correct answer. Inference Tables in Databricks are designed to store the results and metadata of inference runs. This allows the system to log incoming requests and outgoing responses directly within Databricks, making it an ideal choice for monitoring the behavior of a provisioned serving endpoint. Inference Tables can be queried and analyzed, enabling easier monitoring and debugging compared to a custom microservice.

Thus, Inference Tables are the optimal feature for monitoring request and response logs within the Databricks infrastructure for a model serving endpoint.

68. Frage

A Generative AI Engineer is creating an LLM-powered application that will need access to up-to-date news articles and stock prices.

The design requires the use of stock prices which are stored in Delta tables and finding the latest relevant news articles by searching the internet.

How should the Generative AI Engineer architect their LLM system?

- A. Query the Delta table for volatile stock prices and use an LLM to generate a search query to investigate potential causes of the stock volatility.
- B. Download and store news articles and stock price information in a vector store. Use a RAG architecture to retrieve and generate at runtime.
- C. Create an agent with tools for SQL querying of Delta tables and web searching, provide retrieved values to an LLM for generation of response.
- D. Use an LLM to summarize the latest news articles and lookup stock tickers from the summaries to find stock prices.

Antwort: C

Begründung:

To build an LLM-powered system that accesses up-to-date news articles and stock prices, the best approach is to create an agent that has access to specific tools (option D).

Agent with SQL and Web Search Capabilities:

By using an agent-based architecture, the LLM can interact with external tools. The agent can query Delta tables (for up-to-date stock prices) via SQL and perform web searches to retrieve the latest news articles. This modular approach ensures the system can access both structured (stock prices) and unstructured (news) data sources dynamically.

Why This Approach Works:

SQL Queries for Stock Prices: Delta tables store stock prices, which the agent can query directly for the latest data.

Web Search for News: For news articles, the agent can generate search queries and retrieve the most relevant and recent articles, then pass them to the LLM for processing.

Why Other Options Are Less Suitable:

A (Summarizing News for Stock Prices): This convoluted approach would not ensure accuracy when retrieving stock prices, which are already structured and stored in Delta tables.

B (Stock Price Volatility Queries): While this could retrieve relevant information, it doesn't address how to obtain the most up-to-date news articles.

C (Vector Store): Storing news articles and stock prices in a vector store might not capture the real-time nature of stock data and news updates, as it relies on pre-existing data rather than dynamic querying.

Thus, using an agent with access to both SQL for querying stock prices and web search for retrieving news articles is the best approach for ensuring up-to-date and accurate responses.

69. Frage

A Generative AI Engineer would like an LLM to generate formatted JSON from emails. This will require parsing and extracting the following information: order ID, date, and sender email. Here's a sample email:

```
Date: April 23, 2024
Time: 4:22 PM
From: anjali.thayer@computex.org
To: cust_service@realtek.com
Subject: Shipment details
```



Hey there,

I have a shipment (order ID is CD34RFT) can you please send me an update?

Thank you,
Anjali

They will need to write a prompt that will extract the relevant information in JSON format with the highest level of output accuracy.

Which prompt will do that?

- A. You will receive customer emails and need to extract date, sender email, and order ID. You should return the date, sender email, and order ID information in JSON format.
- B. You will receive customer emails and need to extract date, sender email, and order ID. Return the extracted information in a human-readable format.
- **C. You will receive customer emails and need to extract date, sender email, and order ID. Return the extracted information in JSON format.**
Here's an example: {"date": "April 16, 2024", "sender_email": "sarah.lee925@gmail.com", "order_id": "RE987D"}
- D. You will receive customer emails and need to extract date, sender email, and order ID. Return the extracted information in JSON format.

Antwort: C

Begründung:

Problem Context: The goal is to parse emails to extract certain pieces of information and output this in a structured JSON format. Clarity and specificity in the prompt design will ensure higher accuracy in the LLM's responses.

Explanation of Options:

- * Option A: Provides a general guideline but lacks an example, which helps an LLM understand the exact format expected.
- * Option B: Includes a clear instruction and a specific example of the output format. Providing an example is crucial as it helps set the pattern and format in which the information should be structured, leading to more accurate results.
- * Option C: Does not specify that the output should be in JSON format, thus not meeting the requirement.
- * Option D: While it correctly asks for JSON format, it lacks an example that would guide the LLM on how to structure the JSON correctly.

Therefore, Option B is optimal as it not only specifies the required format but also illustrates it with an example, enhancing the likelihood of accurate extraction and formatting by the LLM.

70. Frage

A Generative AI Engineer I 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_delta_sync_index()**
- B. vsc.create_direct_access_index()
- C. vsc.get_index()
- D. vsc.similarity_search()

Antwort: A

Begründung:

* 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

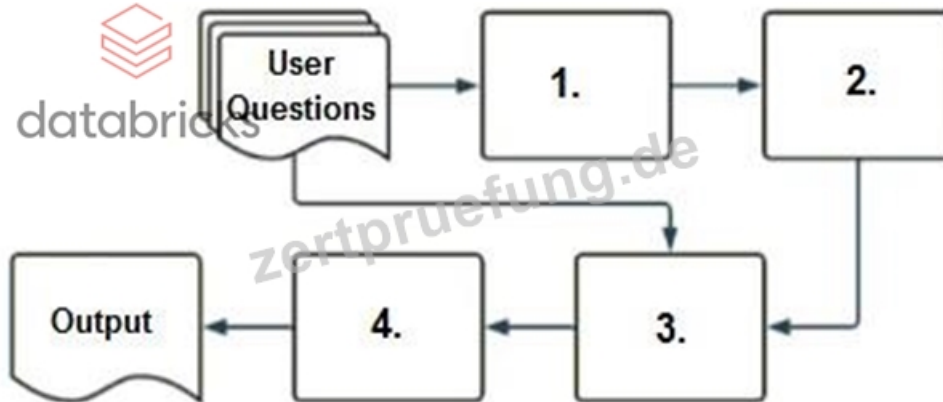
changes) is intended.

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.

71. Frage

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

Antwort: A

Begründung:

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.

72. Frage

.....

Wenn wir am Anfang die Fragenkataloge zur Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung

