Monitor Your Progress with Databricks-Generative-AI-Engineer-Associate Practice Test Software

P.S. Free 2026 Databricks Databricks-Generative-AI-Engineer-Associate dumps are available on Google Drive shared by TorrentValid: https://drive.google.com/open?id=1EpKl1-fdR8KRaNpB22HkDzt8XAlXgyEW

TorrentValid provides proprietary preparation guides for the certification exam offered by the Databricks-Generative-AI-Engineer-Associate exam dumps. In addition to containing numerous questions similar to the Databricks-Generative-AI-Engineer-Associate Exam, the Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam questions are a great way to prepare for the Databricks Databricks-Generative-AI-Engineer-Associate exam dumps.

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

Topic	Details
Topic 1	• 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.

Topic 2	• 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	• 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.

Databricks-Generative-AI-Engineer-Associate Exam Vce - Databricks-Generative-AI-Engineer-Associate Free Braindumps

As you know, the Databricks-Generative-AI-Engineer-Associate certificate is hard to get for most people. But our Databricks-Generative-AI-Engineer-Associate study guide will offer you the most professional guidance. As old saying goes, opportunities are always for those who prepare themselves well. In the end, you will easily pass the Databricks-Generative-AI-Engineer-Associate Exam through our assistance. Then you will find that your work ability is elevated greatly by studying our Databricks-Generative-AI-Engineer-Associate actual exam. In the end, you will become an excellent talent.

Databricks Certified Generative AI Engineer Associate Sample Questions (Q38-Q43):

NEW QUESTION # 38
A Generative AI Engineer is designing a RAG application for answering user questions on technical regulations as they learn a new sport.
What are the steps needed to build this RAG application and deploy it?

- A. Ingest documents from a source -> Index the documents and save to Vector Search -> User submits queries against an LLM -> LLM retrieves relevant documents -> LLM generates a response -> Evaluate model -> Deploy it using Model Serving
- B. Ingest documents from a source -> Index the documents and save to Vector Search -> Evaluate model -> Deploy it using Model Serving
- C. Ingest documents from a source -> Index the documents and saves to Vector Search -> User submits queries against an LLM -> LLM retrieves relevant documents -> Evaluate model -> LLM generates a response -> Deploy it using Model Serving
- D. User submits queries against an LLM -> Ingest documents from a source -> Index the documents and save to Vector Search -> LLM retrieves relevant documents -> LLM generates a response -> Evaluate model -> Deploy it using Model Serving

Answer: A

Explanation:
The Generative AI Engineer needs to follow a methodical pipeline to build and deploy a Retrieval-Augmented Generation (RAG) application. The steps outlined in option B accurately reflect this process:
Ingest documents from a source: This is the first step, where the engineer collects documents (e.g., technical regulations) that will be used for retrieval when the application answers user questions.
Index the documents and save to Vector Search: Once the documents are ingested, they need to be embedded using a technique like embeddings (e.g., with a pre-trained model like BERT) and stored in a vector database (such as Pinecone or FAISS). This enables fast retrieval based on user queries.
User submits queries against an LLM: Users interact with the application by submitting their queries. These queries will be passed to the LLM.
LLM retrieves relevant documents: The LLM works with the vector store to retrieve the most relevant documents based on their vector representations.
LLM generates a response: Using the retrieved documents, the LLM generates a response that is tailored to the user's question.
Evaluate model: After generating responses, the system must be evaluated to ensure the retrieved documents are relevant and the

generated response is accurate. Metrics such as accuracy, relevance, and user satisfaction can be used for evaluation.
Deploy it using Model Serving: Once the RAG pipeline is ready and evaluated, it is deployed using a model-serving platform such as Databricks Model Serving. This enables real-time inference and response generation for users.
By following these steps, the Generative AI Engineer ensures that the RAG application is both efficient and effective for the task of answering technical regulation questions.

NEW QUESTION # 39
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. Download and store news articles and stock price information in a vector store. Use a RAG architecture to retrieve and generate at runtime.
- B. Use an LLM to summarize the latest news articles and lookup stock tickers from the summaries to find stock prices.
- C. 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.
- D. Create an agent with tools for SQL querying of Delta tables and web searching, provide retrieved values to an LLM for generation of response.

Answer: D

Explanation:
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.

NEW QUESTION # 40
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. Lakeview
- C. Vector Search
- D. Inference Tables

Answer: D

Explanation:
Problem Context: The goal is to monitor theserving endpointfor incoming requests and outgoing responses in aprovisioned

throughput model serving endpointwithin aRetrieval-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 Tablesin Databricks are designed to store the results and metadata of inference runs. This allows the system to logincoming requests and outgoing responsesdirectly 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 Tablesare the optimal feature for monitoring request and response logs within the Databricks infrastructure for a model serving endpoint.

NEW QUESTION # 41
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.response-generating LLM, 2.context-augmented prompt, 3.vector search, 4.embedding model
- B. 1.embedding model, 2.vector search, 3.context-augmented prompt, 4.response-generating LLM
- C. 1.response-generating LLM, 2.vector search, 3.context-augmented prompt, 4.embedding model
- D. 1.context-augmented prompt, 2.vector search, 3.embedding model, 4.response-generating LLM

Answer: B

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 isembedding model, vector search, context-augmented prompt, response- generating LLM, which is option A.

NEW QUESTION # 42
A Generative AI Engineer is testing a simple prompt template in LangChain using the code below, but is getting an error.
Assuming the API key was properly defined, what change does the Generative AI Engineer need to make to fix their chain?

- A.
- B.
- C.

- D. □

Answer: C

Explanation:
To fix the error in the LangChain code provided for using a simple prompt template, the correct approach is Option C. Here's a detailed breakdown of why Option C is the right choice and how it addresses the issue:
* Proper Initialization: In Option C, the LLMChain is correctly initialized with the LLM instance specified as OpenAI(), which likely represents a language model (like GPT) from OpenAI. This is crucial as it specifies which model to use for generating responses.
* Correct Use of Classes and Methods:
* The PromptTemplate is defined with the correct format, specifying that adjective is a variable within the template. This allows dynamic insertion of values into the template when generating text.
* The prompt variable is properly linked with the PromptTemplate, and the final template string is passed correctly.
* The LLMChain correctly references the prompt and the initialized OpenAI() instance, ensuring that the template and the model are properly linked for generating output.
Why Other Options Are Incorrect:
* Option A: Misuses the parameter passing in generate method by incorrectly structuring the dictionary.
* Option B: Incorrectly uses prompt.format method which does not exist in the context of LLMChain and PromptTemplate configuration, resulting in potential errors.
* Option D: Incorrect order and setup in the initialization parameters for LLMChain, which would likely lead to a failure in recognizing the correct configuration for prompt and LLM usage.
Thus, Option C is correct because it ensures that the LangChain components are correctly set up and integrated, adhering to proper syntax and logical flow required by LangChain's architecture. This setup avoids common pitfalls such as type errors or method misuses, which are evident in other options.

NEW QUESTION # 43
.....

Valid Exam Bootcamp

- Databricks-Generative-AI-Engineer-Associate Test Papers □ Databricks-Generative-AI-Engineer-Associate Exam Revision Plan □ Regualer Databricks-Generative-AI-Engineer-Associate Update □ Search for （ Databricks-Generative-AI-Engineer-Associate ） and download it for free on [www.troytecdumps.com] website □Databricks-Generative-AI-Engineer-Associate Test Papers
- Databricks Databricks-Generative-AI-Engineer-Associate - Databricks Certified Generative AI Engineer Associate Fantastic Interactive Practice Exam □ Easily obtain " Databricks-Generative-AI-Engineer-Associate " for free download through ➠ www.pdfvce.com □ □Databricks-Generative-AI-Engineer-Associate Exam Revision Plan
- Databricks-Generative-AI-Engineer-Associate Valid Exam Duration □ Databricks-Generative-AI-Engineer-Associate Pass Guarantee □ Exam Databricks-Generative-AI-Engineer-Associate Tests ➡□ Easily obtain □ Databricks-Generative-AI-Engineer-Associate □ for free download through ➠ www.vce4dumps.com □ □Databricks-Generative-AI-Engineer-Associate Exam Revision Plan
- Interactive Databricks-Generative-AI-Engineer-Associate Practice Exam - 100% the Best Accurate Questions Pool ↘ Search for { Databricks-Generative-AI-Engineer-Associate } on ➤ www.pdfvce.com □ immediately to obtain a free download □Exam Databricks-Generative-AI-Engineer-Associate Format
- Interactive Databricks-Generative-AI-Engineer-Associate Practice Exam - 100% the Best Accurate Questions Pool □ Search on ▸ www.pass4test.com ◂ for ➡ Databricks-Generative-AI-Engineer-Associate □□□ to obtain exam materials for free download ◂Databricks-Generative-AI-Engineer-Associate Latest Exam Notes
- gerarditko038975.losblogos.com, henrirdbw345839.csublogs.com, snoopydirectory.com, www.stes.tyc.edu.tw, rsaourt608875.thenerdsblog.com, www.stes.tyc.edu.tw, lewysnleq836686.actoblog.com, sachintamj210726.wikimeglio.com, asiyappts821089.actoblog.com, www.stes.tyc.edu.tw, Disposable vapes

BONUS!!! Download part of TorrentValid Databricks-Generative-AI-Engineer-Associate dumps for free: https://drive.google.com/open?id=1EpKl1-fdR8KRaNpB22HkDzt8XAlXgyEW