

Achieve Success in the Databricks Databricks-Generative-AI-Engineer-Associate Exam with Confidence



P.S. Free 2025 Databricks Databricks-Generative-AI-Engineer-Associate dumps are available on Google Drive shared by FreeDumps: <https://drive.google.com/open?id=1p1QH-jPjXWnUISs9u7MG6cpBcbwTsBmR>

Do you want to pass Databricks-Generative-AI-Engineer-Associate exam and get the related certification within the minimum time and effort? If you would like to give me a positive answer, you really should keep a close eye on our website since you can find the best Databricks-Generative-AI-Engineer-Associate study material in here--our Databricks-Generative-AI-Engineer-Associate Training Materials. We have helped millions of thousands of candidates to prepare for the Databricks-Generative-AI-Engineer-Associate exam and all of them have got a fruitful outcome, we believe you will be the next winner as long as you join in us!

FreeDumps alerts you that the syllabus of the Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) certification exam changes from time to time. Therefore, keep checking the fresh updates released by the Databricks. It will save you from the unnecessary mental hassle of wasting your valuable money and time. FreeDumps announces another remarkable feature to its users by giving them the Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) dumps updates until 1 year after purchasing the Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) certification exam pdf questions.

>> Databricks-Generative-AI-Engineer-Associate Exam Simulator Fee <<

Sample Databricks-Generative-AI-Engineer-Associate Questions | Reliable Databricks-Generative-AI-Engineer-Associate Dumps Free

The Databricks Databricks-Generative-AI-Engineer-Associate exam questions are being offered in three different formats. These formats are Databricks-Generative-AI-Engineer-Associate PDF dumps files, desktop practice test software, and web-based practice test software. All these three Databricks-Generative-AI-Engineer-Associate exam dumps formats contain the Real Databricks-Generative-AI-Engineer-Associate Exam Questions that assist you in your Databricks Certified Generative AI Engineer Associate practice exam preparation and finally, you will be confident to pass the final Databricks Databricks-Generative-AI-Engineer-Associate exam easily.

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"> • 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.
Topic 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.
Topic 4	<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.

Databricks Certified Generative AI Engineer Associate Sample Questions (Q53-Q58):

NEW QUESTION # 53

A Generative AI Engineer is tasked with deploying an application that takes advantage of a custom MLflow Pyfunc model to return some interim results.

How should they configure the endpoint to pass the secrets and credentials?

- A. Add credentials using environment variables
- B. Pass the secrets in plain text
- C. Use spark.conf.set ()
- D. Pass variables using the Databricks Feature Store API

Answer: A

Explanation:

Context: Deploying an application that uses an MLflow Pyfunc model involves managing sensitive information such as secrets and credentials securely.

Explanation of Options:

* Option A: Use spark.conf.set(): While this method can pass configurations within Spark jobs, using it for secrets is not recommended because it may expose them in logs or Spark UI.

* Option B: Pass variables using the Databricks Feature Store API: The Feature Store API is designed for managing features for machine learning, not for handling secrets or credentials.

* Option C: Add credentials using environment variables: This is a common practice for managing credentials in a secure manner, as environment variables can be accessed securely by applications without exposing them in the codebase.

* Option D: Pass the secrets in plain text: This is highly insecure and not recommended, as it exposes sensitive information directly in the code.

Therefore, Option C is the best method for securely passing secrets and credentials to an application, protecting them from exposure.

NEW QUESTION # 54

A Generative AI Engineer has been asked to build an LLM-based question-answering application. The application should take into account new documents that are frequently published. The engineer wants to build this application with the least cost and least development effort and have it operate at the lowest cost possible.

Which combination of chaining components and configuration meets these requirements?

- A. The LLM needs to be frequently with the new documents in order to provide most up-to-date answers.
- B. For the application a prompt, a retriever, and an LLM are required. The retriever output is inserted into the prompt which is given to the LLM to generate answers.
- C. For the question-answering application, prompt engineering and an LLM are required to generate answers.
- D. For the application a prompt, an agent and a fine-tuned LLM are required. The agent is used by the LLM to retrieve relevant content that is inserted into the prompt which is given to the LLM to generate answers.

Answer: B

Explanation:

Problem Context: The task is to build an LLM-based question-answering application that integrates new documents frequently with minimal costs and development efforts.

Explanation of Options:

- * Option A: Utilizes a prompt and a retriever, with the retriever output being fed into the LLM. This setup is efficient because it dynamically updates the data pool via the retriever, allowing the LLM to provide up-to-date answers based on the latest documents without needing to frequently retrain the model. This method offers a balance of cost-effectiveness and functionality.
 - * Option B: Requires frequent retraining of the LLM, which is costly and labor-intensive.
 - * Option C: Only involves prompt engineering and an LLM, which may not adequately handle the requirement for incorporating new documents unless it's part of an ongoing retraining or updating mechanism, which would increase costs.
 - * Option D: Involves an agent and a fine-tuned LLM, which could be overkill and lead to higher development and operational costs.
- Option A is the most suitable as it provides a cost-effective, minimal development approach while ensuring the application remains up-to-date with new information.

NEW QUESTION # 55

A Generative AI Engineer is helping a cinema extend its website's chat bot to be able to respond to questions about specific showtimes for movies currently playing at their local theater. They already have the location of the user provided by location services to their agent, and a Delta table which is continually updated with the latest showtime information by location. They want to implement this new capability in their RAG application.

Which option will do this with the least effort and in the most performant way?

- A. Query the Delta table directly via a SQL query constructed from the user's input using a text-to-SQL LLM in the agent logic / tool
- **B. Create a Feature Serving Endpoint from a FeatureSpec that references an online store synced from the Delta table. Query the Feature Serving Endpoint as part of the agent logic / tool implementation.**
- C. Set up a task in Databricks Workflows to write the information in the Delta table periodically to an external database such as MySQL and query the information from there as part of the agent logic / tool implementation.
- D. implementation. Write the Delta table contents to a text column, then embed those texts using an embedding model and store these in the vector index. Look up the information based on the embedding as part of the agent logic / tool implementation.

Answer: B

Explanation:

The task is to extend a cinema chatbot to provide movie showtime information using a RAG application, leveraging user location and a continuously updated Delta table, with minimal effort and high performance.

Let's evaluate the options.

- * Option A: Create a Feature Serving Endpoint from a FeatureSpec that references an online store synced from the Delta table.

Query the Feature Serving Endpoint as part of the agent logic / tool implementation

* Databricks Feature Serving provides low-latency access to real-time data from Delta tables via an online store. Syncing the Delta table to a Feature Serving Endpoint allows the chatbot to query showtimes efficiently, integrating seamlessly into the RAG agent's tool logic. This leverages Databricks' native infrastructure, minimizing effort and ensuring performance.

* Databricks Reference: "Feature Serving Endpoints provide real-time access to Delta table data with low latency, ideal for production systems" ("Databricks Feature Engineering Guide," 2023).

- * Option B: Query the Delta table directly via a SQL query constructed from the user's input using a text-to-SQL LLM in the agent logic / tool

* Using a text-to-SQL LLM to generate queries adds complexity (e.g., ensuring accurate SQL generation) and latency (LLM inference + SQL execution). While feasible, it's less performant and requires more effort than a pre-built serving solution.

* Databricks Reference: "Direct SQL queries are flexible but may introduce overhead in real-time applications" ("Building LLM Applications with Databricks").

- * Option C: Write the Delta table contents to a text column, then embed those texts using an embedding model and store these in the vector index. Look up the information based on the embedding as part of the agent logic / tool implementation

* Converting structured Delta table data (e.g., showtimes) into text, embedding it, and using vector search is inefficient for structured lookups. It's effort-intensive (preprocessing, embedding) and less precise than direct queries, undermining performance.

* Databricks Reference: "Vector search excels for unstructured data, not structured tabular lookups" ("Databricks Vector Search Documentation").

- * Option D: Set up a task in Databricks Workflows to write the information in the Delta table periodically to an external database such as MySQL and query the information from there as part of the agent logic / tool implementation

* Exporting to an external database (e.g., MySQL) adds setup effort (workflow, external DB management) and latency (periodic updates vs. real-time). It's less performant and more complex than using Databricks' native tools.

* Databricks Reference: "Avoid external systems when Delta tables provide real-time data natively" ("Databricks Workflows Guide").

Conclusion: Option A minimizes effort by using Databricks Feature Serving for real-time, low-latency access to the Delta table, ensuring high performance in a production-ready RAG chatbot.

NEW QUESTION # 56

A Generative AI Engineer is tasked with improving the RAG quality by addressing its inflammatory outputs. Which action would be most effective in mitigating the problem of offensive text outputs?

- A. Restrict access to the data sources to a limited number of users
- B. Inform the user of the expected RAG behavior
- C. Curate upstream data properly that includes manual review before it is fed into the RAG system
- D. Increase the frequency of upstream data updates

Answer: C

Explanation:

Addressing offensive or inflammatory outputs in a Retrieval-Augmented Generation (RAG) system is critical for improving user experience and ensuring ethical AI deployment. Here's why C is the most effective approach:

* Manual data curation: The root cause of offensive outputs often comes from the underlying data used to train the model or populate the retrieval system. By manually curating the upstream data and conducting thorough reviews before the data is fed into the RAG system, the engineer can filter out harmful, offensive, or inappropriate content.

* Improving data quality: Curating data ensures the system retrieves and generates responses from a high-quality, well-vetted dataset. This directly impacts the relevance and appropriateness of the outputs from the RAG system, preventing inflammatory content from being included in responses.

* Effectiveness: This strategy directly tackles the problem at its source (the data) rather than just mitigating the consequences (such as informing users or restricting access). It ensures that the system consistently provides non-offensive, relevant information. Other options, such as increasing the frequency of data updates or informing users about behavior expectations, may not directly mitigate the generation of inflammatory outputs.

NEW QUESTION # 57

A Generative AI Engineer is designing an LLM-powered live sports commentary platform. The platform provides real-time updates and LLM-generated analyses for any users who would like to have live summaries, rather than reading a series of potentially outdated news articles.

Which tool below will give the platform access to real-time data for generating game analyses based on the latest game scores?

- A. Foundation Model APIs
- B. AutoML
- C. DatabricksIQ
- D. Feature Serving

Answer: D

Explanation:

* Problem Context: The engineer is developing an LLM-powered live sports commentary platform that needs to provide real-time updates and analyses based on the latest game scores. The critical requirement here is the capability to access and integrate real-time data efficiently with the platform for immediate analysis and reporting.

* Explanation of Options:

* Option A: DatabricksIQ: While DatabricksIQ offers integration and data processing capabilities, it is more aligned with data analytics rather than real-time feature serving, which is crucial for immediate updates necessary in a live sports commentary context.

* Option B: Foundation Model APIs: These APIs facilitate interactions with pre-trained models and could be part of the solution, but on their own, they do not provide mechanisms to access real-time game scores.

* Option C: Feature Serving: This is the correct answer as feature serving specifically refers to the real-time provision of data (features) to models for prediction. This would be essential for an LLM that generates analyses based on live game data, ensuring that the commentary is current and based on the latest events in the sport.

* Option D: AutoML: This tool automates the process of applying machine learning models to real-world problems, but it does not directly provide real-time data access, which is a critical requirement for the platform.

Thus, Option C (Feature Serving) is the most suitable tool for the platform as it directly supports the real-time data needs of an LLM-

myportal.utt.edu.tt, Disposable vapes

P.S. Free 2025 Databricks Databricks-Generative-AI-Engineer-Associate dumps are available on Google Drive shared by FreeDumps: <https://drive.google.com/open?id=1p1QH-jPjXWnUISs9u7MG6cpBcbwTsBmR>