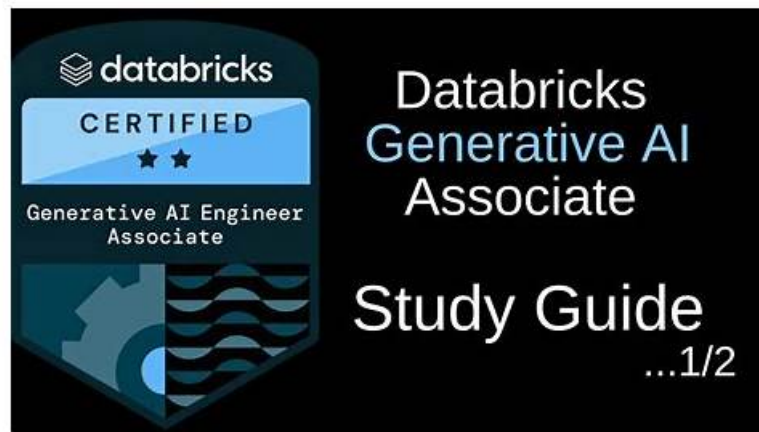


100% Pass-Rate Databricks-Generative-AI-Engineer-Associate Latest Study Notes - Correct Databricks-Generative-AI-Engineer-Associate Exam Tool Guarantee Purchasing Safety



BONUS!!! Download part of Pass4guide Databricks-Generative-AI-Engineer-Associate dumps for free:
https://drive.google.com/open?id=1vrzpICK7kG1m33hby9a_DnCw68BWdOd1

Will you feel nervous in the exam? If you do, just choose us, our Databricks-Generative-AI-Engineer-Associate Soft test engine can stimulate the real exam environment, which will help you know the procedure of the exam, and will strengthen your confidence. Moreover Databricks-Generative-AI-Engineer-Associate exam dumps are high-quality, and we have professional experts to compile them, and they can help you pass the exam just one time. We offer you free demo to have a try for Databricks-Generative-AI-Engineer-Associate Exam Dumps, and free update for one year. If you indeed have questions, just contact with us.

The 21 century is the information century. So there are many changes in the field of the Databricks-Generative-AI-Engineer-Associate exam questions. They are also transforming people's lives and the mode of operation of human society in a profound way. when you are preparing for an Databricks-Generative-AI-Engineer-Associate exam, our company can provide the best electronic Databricks-Generative-AI-Engineer-Associate Exam Torrent for you in this website. I strongly believe that under the guidance of our Databricks-Generative-AI-Engineer-Associate test torrent, you will be able to keep out of troubles way and take everything in your stride.

>> **Databricks-Generative-AI-Engineer-Associate Latest Study Notes** <<

Databricks-Generative-AI-Engineer-Associate Certified Questions & Databricks-Generative-AI-Engineer-Associate Free Learning Cram

Many people dream about occupying a prominent position in the society and being successful in their career and social circle. Thus owning a valuable certificate is of paramount importance to them and passing the test Databricks-Generative-AI-Engineer-Associate certification can help them realize their goals. If you are one of them buying our Databricks-Generative-AI-Engineer-Associate Exam Prep will help you pass the Databricks-Generative-AI-Engineer-Associate exam successfully and easily. Our Databricks-Generative-AI-Engineer-Associate guide torrent provides free download and tryout before the purchase and our purchase procedures are safe.

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

Topic	Details

Topic 1	<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.
Topic 2	<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 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.

Databricks Certified Generative AI Engineer Associate Sample Questions (Q59-Q64):

NEW QUESTION # 59

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:

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. Return the extracted information in JSON format.
Here's an example: {"date": "April 16, 2024", "sender_email": "sarah.lee925@gmail.com", "order_id": "RE987D"}
- 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.
- D. 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.

Answer: A

Explanation:

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.

NEW QUESTION # 60

A Generative AI Engineer has created a RAG application to look up answers to questions about a series of fantasy novels that are being asked on the author's web forum. The fantasy novel texts are chunked and embedded into a vector store with metadata (page number, chapter number, book title), retrieved with the user's query, and provided to an LLM for response generation. The Generative AI Engineer used their intuition to pick the chunking strategy and associated configurations but now wants to more methodically choose the best values.

Which TWO strategies should the Generative AI Engineer take to optimize their chunking strategy and parameters? (Choose two.)

- A. Change embedding models and compare performance.
- B. Create an LLM-as-a-judge metric to evaluate how well previous questions are answered by the most appropriate chunk. Optimize the chunking parameters based upon the values of the metric.
- C. Pass known questions and best answers to an LLM and instruct the LLM to provide the best token count. Use a summary statistic (mean, median, etc.) of the best token counts to choose chunk size.
- D. Add a classifier for user queries that predicts which book will best contain the answer. Use this to filter retrieval.
- E. Choose an appropriate evaluation metric (such as recall or NDCG) and experiment with changes in the chunking strategy, such as splitting chunks by paragraphs or chapters. Choose the strategy that gives the best performance metric.

Answer: B,E

Explanation:

To optimize a chunking strategy for a Retrieval-Augmented Generation (RAG) application, the Generative AI Engineer needs a structured approach to evaluating the chunking strategy, ensuring that the chosen configuration retrieves the most relevant information and leads to accurate and coherent LLM responses.

Here's why C and E are the correct strategies:

Strategy C: Evaluation Metrics (Recall, NDCG)

- * Define an evaluation metric: Common evaluation metrics such as recall, precision, or NDCG (Normalized Discounted Cumulative Gain) measure how well the retrieved chunks match the user's query and the expected response.

- * Recall measures the proportion of relevant information retrieved.

- * NDCG is often used when you want to account for both the relevance of retrieved chunks and the ranking or order in which they are retrieved.

- * Experiment with chunking strategies: Adjusting chunking strategies based on text structure (e.g., splitting by paragraph, chapter, or a fixed number of tokens) allows the engineer to experiment with various ways of slicing the text. Some chunks may better align with the user's query than others.

- * Evaluate performance: By using recall or NDCG, the engineer can methodically test various chunking strategies to identify which one yields the highest performance. This ensures that the chunking method provides the most relevant information when embedding and retrieving data from the vector store.

Strategy E: LLM-as-a-Judge Metric

- * Use the LLM as an evaluator: After retrieving chunks, the LLM can be used to evaluate the quality of answers based on the chunks provided. This could be framed as a "judge" function, where the LLM compares how well a given chunk answers previous user queries.

- * Optimize based on the LLM's judgment: By having the LLM assess previous answers and rate their relevance and accuracy, the engineer can collect feedback on how well different chunking configurations perform in real-world scenarios.

- * This metric could be a qualitative judgment on how closely the retrieved information matches the user's intent.

- * Tune chunking parameters: Based on the LLM's judgment, the engineer can adjust the chunk size or structure to better align with the LLM's responses, optimizing retrieval for future queries.

By combining these two approaches, the engineer ensures that the chunking strategy is systematically evaluated using both quantitative (recall/NDCG) and qualitative (LLM judgment) methods. This balanced optimization process results in improved retrieval relevance and, consequently, better response generation by the LLM.

NEW QUESTION # 61

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.similarity_search()`
- D. `vsc.get_index()`

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 # 62

A Generative AI Engineer wants their (in)etuned LLMs in their prod Databricks workspace available for testing in their dev workspace as well. All of their workspaces are Unity Catalog enabled and they are currently logging their models into the Model Registry in MLflow.

What is the most cost-effective and secure option for the Generative AI Engineer to accomplish their gAi?

- A. Use an external model registry which can be accessed from all workspaces
- B. Setup a duplicate training pipeline in dev, so that an identical model is available in dev.
- C. Setup a script to export the model from prod and import it to dev.
- **D. Use MLflow to log the model directly into Unity Catalog, and enable READ access in the dev workspace to the model.**

Answer: D

Explanation:

The goal is to make fine-tuned LLMs from a production (prod) Databricks workspace available for testing in a development (dev) workspace, leveraging Unity Catalog and MLflow, while ensuring cost-effectiveness and security. Let's analyze the options.

* Option A: Use an external model registry which can be accessed from all workspaces

* An external registry adds cost (e.g., hosting fees) and complexity (e.g., integration, security configurations) outside Databricks' native ecosystem, reducing security compared to Unity Catalog's governance.

* Databricks Reference: "Unity Catalog provides a centralized, secure model registry within Databricks" ("Unity Catalog Documentation," 2023).

* Option B: Setup a script to export the model from prod and import it to dev

- * Export/import scripts require manual effort, storage for model artifacts, and repeated execution, increasing operational cost and risk (e.g., version mismatches, unsecured transfers). It's less efficient than a native solution.
- * Databricks Reference: Manual processes are discouraged when Unity Catalog offers built-in sharing: "Avoid redundant workflows with Unity Catalog's cross-workspace access" ("MLflow with Unity Catalog").
- * Option C: Setup a duplicate training pipeline in dev, so that an identical model is available in dev
- * Duplicating the training pipeline doubles compute and storage costs, as it retrains the model from scratch. It's neither cost-effective nor necessary when the prod model can be reused securely.
- * Databricks Reference: "Re-running training is resource-intensive; leverage existing models where possible" ("Generative AI Engineer Guide").
- * Option D: Use MLflow to log the model directly into Unity Catalog, and enable READ access in the dev workspace to the model
- * Unity Catalog, integrated with MLflow, allows models logged in prod to be centrally managed and accessed across workspaces with fine-grained permissions (e.g., READ for dev). This is cost-effective (no extra infrastructure or retraining) and secure (governed by Databricks' access controls).
- * Databricks Reference: "Log models to Unity Catalog via MLflow, then grant access to other workspaces securely" ("MLflow Model Registry with Unity Catalog," 2023).

Conclusion: Option D leverages Databricks' native tools (MLflow and Unity Catalog) for a seamless, cost-effective, and secure solution, avoiding external systems, manual scripts, or redundant training.

NEW QUESTION # 63

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

Answer: A

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-powered sports commentary system, ensuring that the analyses and updates are based on the latest available information.

NEW QUESTION # 64

.....

As long as you enter the learning interface of our soft test engine of Databricks-Generative-AI-Engineer-Associate quiz guide and start practicing on our Windows software, you will find that there are many small buttons that are designed to better assist you in your learning. When you want to correct the answer after you finish learning, the correct answer for our Databricks-Generative-AI-Engineer-Associate Test Prep is below each question, and you can correct it based on the answer. In addition, we design small buttons, which can also show or hide the Databricks-Generative-AI-Engineer-Associate exam torrent, and you can flexibly and freely choose these two modes according to your habit.

Databricks-Generative-AI-Engineer-Associate Certified Questions: <https://www.pass4guide.com/Databricks-Generative-AI-Engineer-Associate-Certified-Questions/>

- BONUS!!! Download part of Pass4guide Databricks-Generative-AI-Engineer-Associate dumps for free: https://drive.google.com/open?id=1vrzpICK7kG1m33hby9a_DnCw68BWdOd1

BONUS!!! Download part of Pass4guide Databricks-Generative-AI-Engineer-Associate dumps for free: https://drive.google.com/open?id=1vrzpICK7kG1m33hby9a_DnCw68BWdOd1