

Databricks-Generative-AI-Engineer-Associate Reliable Exam Guide | Databricks-Generative-AI-Engineer-Associate Learning Materials



P.S. Free & New Databricks-Generative-AI-Engineer-Associate dumps are available on Google Drive shared by PrepAwayETE: https://drive.google.com/open?id=14HqSCZS_b3t4JFA73ewcGSKxVoy5cHVQ

Each question presents the key information to the learners and each answer provides the detailed explanation and verification by the senior experts. The success of our Databricks-Generative-AI-Engineer-Associate study materials cannot be separated from their painstaking efforts. Our system will do an all-around statistics of the sales volume of our Databricks-Generative-AI-Engineer-Associate Study Materials at home and abroad and our clients' positive feedback rate of our Databricks-Generative-AI-Engineer-Associate study materials. Our system will deal with the clients' online consultation and refund issues promptly and efficiently. So our system is great.

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

Topic	Details
Topic 1	<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 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"> • 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 Reliable Exam Guide <<

Databricks Databricks-Generative-AI-Engineer-Associate Learning Materials, Databricks-Generative-AI-Engineer-Associate Test Practice

Although we have carried out the Databricks-Generative-AI-Engineer-Associate exam questions for customers, it does not mean that we will stop perfecting our study materials. Our experts are still testing new functions for the Databricks-Generative-AI-Engineer-Associate study materials. Even if you have purchased our study materials, you still can enjoy our updated Databricks-Generative-AI-Engineer-Associate Practice Engine. We will soon upload our new version of our Databricks-Generative-AI-Engineer-Associate guide braindumps into our official websites.

Databricks Certified Generative AI Engineer Associate Sample Questions (Q57-Q62):

NEW QUESTION # 57

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. 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.
- B. Change embedding models and compare performance.
- C. Add a classifier for user queries that predicts which book will best contain the answer. Use this to filter retrieval.
- D. 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.
- E. 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.

Answer: A,D

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 # 58

A generative AI engineer is deploying an AI agent authored with MLflow's ChatAgent interface for a retail company's customer support system on Databricks. The agent must handle thousands of inquiries daily, and the engineer needs to track its performance and quality in real-time to ensure it meets service-level agreements. Which metrics are automatically captured by default and made available for monitoring when the agent is deployed using the Mosaic AI Agent Framework?

- A. Operational metrics like request volume, latency, and errors
- B. No metrics are automatically captured
- C. Quality metrics like correctness and guideline adherence
- D. Both operational and quality metrics

Answer: A

Explanation:

When deploying an agent via the Mosaic AI Agent Framework (which leverages Databricks Model Serving), operational metrics are captured automatically by default. These include system-level telemetry such as the number of requests per second (volume), the time taken for the model to respond (latency), and the rate of 4xx/5xx HTTP errors. These are essential for monitoring Service Level Agreements (SLAs). However, Quality metrics (B), such as correctness, groundedness, or adherence to custom guidelines, cannot be determined "automatically" by the serving infrastructure because they require either human feedback or an LLM-as-a-judge evaluation (using Databricks Agent Evaluation). While Databricks makes it easy to generate quality metrics using the `mflow.evaluate` API or the inference table, they are not "default operational metrics" that appear without additional evaluation configuration.

NEW QUESTION # 59

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. 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.
- B. Change embedding models and compare performance.
- C. Add a classifier for user queries that predicts which book will best contain the answer. Use this to filter retrieval.
- D. 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.
- E. 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.

Answer: A,D

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 # 60

A Generative AI Engineer wants to build an LLM-based solution to help a restaurant improve its online customer experience with bookings by automatically handling common customer inquiries. The goal of the solution is to minimize escalations to human intervention and phone calls while maintaining a personalized interaction. To design the solution, the Generative AI Engineer needs to define the input data to the LLM and the task it should perform.

Which input/output pair will support their goal?

- A. Input: Online chat logs; Output: Buttons that represent choices for booking details
- B. Input: Online chat logs; Output: Group the chat logs by users, followed by summarizing each user's interactions
- C. Input: Online chat logs; Output: Cancellation options
- D. Input: Customer reviews; Output: Classify review sentiment

Answer: A

Explanation:

Context: The goal is to improve the online customer experience in a restaurant by handling common inquiries about bookings, minimizing escalations, and maintaining personalized interactions.

Explanation of Options:

* Option A: Grouping and summarizing chat logs by user could provide insights into customer interactions but does not directly address the task of handling booking inquiries or minimizing escalations.

* Option B: Using chat logs to generate interactive buttons for booking details directly supports the goal of facilitating online bookings, minimizing the need for human intervention by providing clear, interactive options for customers to self-serve.

* Option C: Classifying sentiment of customer reviews does not directly help with booking inquiries, although it might provide valuable feedback insights.

* Option D: Providing cancellation options is helpful but narrowly focuses on one aspect of the booking process and doesn't support the broader goal of handling common inquiries about bookings.

Option B best supports the goal of improving online interactions by using chat logs to generate actionable items for customers, helping them complete booking tasks efficiently and reducing the need for human intervention.

NEW QUESTION # 61

A Generative AI Engineer is tasked with developing a RAG application that will help a small internal group of experts at their company answer specific questions, augmented by an internal knowledge base. They want the best possible quality in the answers, and neither latency nor throughput is a huge concern given that the user group is small and they're willing to wait for the best answer. The topics are sensitive in nature and the data is highly confidential and so, due to regulatory requirements, none of the information is allowed to be transmitted to third parties.

Which model meets all the Generative AI Engineer's needs in this situation?

- A. OpenAI GPT-4
- B. BGE-large
- C. Llama2-70B
- D. Dolly 1.5B

Answer: B

Explanation:

Problem Context: The Generative AI Engineer needs a model for a Retrieval-Augmented Generation (RAG) application that provides high-quality answers, where latency and throughput are not major concerns. The key factors are confidentiality and sensitivity of the data, as well as the requirement for all processing to be confined to internal resources without

external data transmission.

Explanation of Options:

* Option A: Dolly 1.5B: This model does not typically support RAG applications as it's more focused on image generation tasks.

* Option B: OpenAI GPT-4: While GPT-4 is powerful for generating responses, its standard deployment involves cloud-based processing, which could violate the confidentiality requirements due to external data transmission.

* Option C: BGE-large: The BGE (Big Green Engine) large model is a suitable choice if it is configured to operate on-premises or within a secure internal environment that meets regulatory requirements.

Assuming this setup, BGE-large can provide high-quality answers while ensuring that data is not transmitted to third parties, thus aligning with the project's sensitivity and confidentiality needs.

* Option D: Llama2-70B: Similar to GPT-4, unless specifically set up for on-premises use, it generally relies on cloud-based services, which might risk confidential data exposure.

Given the sensitivity and confidentiality concerns, BGE-large is assumed to be configurable for secure internal use, making it the optimal choice for this scenario.

NEW QUESTION # 62

.....

PrepAwayETE assists people in better understanding, studying, and passing more difficult certification exams. We take pride in successfully servicing industry experts by always delivering safe and dependable Databricks-Generative-AI-Engineer-Associate exam preparation materials. For your convenience, PrepAwayETE has prepared authentic Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam study material based on a real exam syllabus to help candidates go through their Databricks-Generative-AI-Engineer-Associate exams.

Databricks-Generative-AI-Engineer-Associate Learning Materials: <https://www.prepawayete.com/Databricks/Databricks-Generative-AI-Engineer-Associate-practice-exam-dumps.html>

- Free Databricks-Generative-AI-Engineer-Associate Sample Databricks-Generative-AI-Engineer-Associate Latest Dumps Ebook Free Databricks-Generative-AI-Engineer-Associate Sample Search for ⇒ Databricks-Generative-AI-Engineer-Associate ⇐ and easily obtain a free download on ☀: www.troytecdumps.com ☀: ◀Databricks-Generative-AI-Engineer-Associate Valid Test Pass4sure
- 100% Pass 2026 Databricks-Generative-AI-Engineer-Associate: Databricks Certified Generative AI Engineer Associate Latest Reliable Exam Guide Search for ➡ Databricks-Generative-AI-Engineer-Associate and obtain a free download on (www.pdfvce.com) Reliable Databricks-Generative-AI-Engineer-Associate Exam Pdf
- Free PDF Quiz Databricks - Latest Databricks-Generative-AI-Engineer-Associate - Databricks Certified Generative AI Engineer Associate Reliable Exam Guide Open website ☀: www.troytecdumps.com ☀: and search for (Databricks-Generative-AI-Engineer-Associate) for free download Databricks-Generative-AI-Engineer-Associate Reasonable Exam Price
- 2026 Trustable Databricks-Generative-AI-Engineer-Associate Reliable Exam Guide | 100% Free Databricks-Generative-AI-Engineer-Associate Learning Materials Easily obtain free download of { Databricks-Generative-AI-Engineer-Associate } by searching on www.pdfvce.com Databricks-Generative-AI-Engineer-Associate Dumps Questions
- Correct Databricks-Generative-AI-Engineer-Associate Reliable Exam Guide - Leader in Qualification Exams - Trustable Databricks-Generative-AI-Engineer-Associate: Databricks Certified Generative AI Engineer Associate Easily obtain ▶ Databricks-Generative-AI-Engineer-Associate ◀ for free download through www.pdfdumps.com Databricks-Generative-AI-Engineer-Associate Exam Cram Review
- 100% Pass Quiz 2026 Latest Databricks-Generative-AI-Engineer-Associate: Databricks Certified Generative AI Engineer Associate Reliable Exam Guide Search for “ Databricks-Generative-AI-Engineer-Associate ” and obtain a free download on ✓ www.pdfvce.com ✓ Databricks-Generative-AI-Engineer-Associate Verified Answers
- Highly Authoritative Databricks-Generative-AI-Engineer-Associate Exam Prep Easy for You to Pass Exam Go to website { www.easy4engine.com } open and search for Databricks-Generative-AI-Engineer-Associate to download for free Flexible Databricks-Generative-AI-Engineer-Associate Testing Engine
- Databricks-Generative-AI-Engineer-Associate Latest Braindumps Book Valid Databricks-Generative-AI-Engineer-Associate Test Objectives New Databricks-Generative-AI-Engineer-Associate Exam Topics ✨ The page for free download of Databricks-Generative-AI-Engineer-Associate on ⇒ www.pdfvce.com ⇐ will open immediately Databricks-Generative-AI-Engineer-Associate Verified Answers
- 100% Pass 2026 Databricks-Generative-AI-Engineer-Associate: Databricks Certified Generative AI Engineer Associate Latest Reliable Exam Guide Simply search for { Databricks-Generative-AI-Engineer-Associate } for free download on ▷ www.exam4labs.com ◁ Databricks-Generative-AI-Engineer-Associate Sample Questions Answers
- New Databricks-Generative-AI-Engineer-Associate Exam Topics Databricks-Generative-AI-Engineer-Associate Certified Reliable Databricks-Generative-AI-Engineer-Associate Exam Pdf Download ➡ Databricks-Generative-AI-Engineer-Associate for free by simply entering “ www.pdfvce.com ” website Databricks-Generative-AI-

Engineer-Associate Certified

- Databricks-Generative-AI-Engineer-Associate Valid Study Questions □ Databricks-Generative-AI-Engineer-Associate Dumps Questions □ Databricks-Generative-AI-Engineer-Associate Latest Dumps Ebook □ Open 《 www.examcollectionpass.com 》 and search for ► Databricks-Generative-AI-Engineer-Associate ◀ to download exam materials for free □ Databricks-Generative-AI-Engineer-Associate Sample Questions Answers
- imogenbivq424843.laowaiblog.com, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, haseebsbbn828511.activablog.com, monicaqteg836169.thebindingwiki.com, asiyapsrv202241.aboutyoublog.com, socialtechnet.com, marleypwfx167887.snack-blog.com, tasneemwtmq132432.thebloggers.com, harmonyaawe632080.wikilima.com, Disposable vapes

BTW, DOWNLOAD part of PrepAwayETE Databricks-Generative-AI-Engineer-Associate dumps from Cloud Storage:
https://drive.google.com/open?id=14HqSCZS_b3t4JFA73ewcGSKxVoy5cHVQ