Databricks-Generative-AI-Engineer-Associate Valid Test Guide - Databricks-Generative-AI-Engineer-Associate Latest Test Format

BTW, DOWNLOAD part of PassExamDumps Databricks-Generative-AI-Engineer-Associate dumps from Cloud Storage: https://drive.google.com/open?id=1eAeZrzVCix7vbMOik0CwS_HVnO4qXqyH

Individuals who hold Databricks Databricks-Generative-AI-Engineer-Associate certification exam demonstrate to their employers and clients that they have the knowledge and skills necessary to succeed in the Databricks-Generative-AI-Engineer-Associate exam. PassExamDumps Databricks-Generative-AI-Engineer-Associate Questions have numerous benefits, including the ability to demonstrate to employers and clients that you have the necessary knowledge and skills to succeed in the actual Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam.

The PassExamDumps is a leading platform that is committed to offering make the Databricks Exam Questions preparation simple, smart, and successful. To achieve this objective PassExamDumps has got the services of experienced and qualified Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam trainers. They work together and put all their efforts and ensure the top standard of PassExamDumps Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam dumps all the time.

>> Databricks-Generative-AI-Engineer-Associate Valid Test Guide <<

Reliable Databricks-Generative-AI-Engineer-Associate Valid Test Guide & Leading Provider in Qualification Exams & Verified Databricks-Generative-AI-Engineer-Associate Latest Test Format

PassExamDumps is a reliable study center providing you the valid and correct Databricks-Generative-AI-Engineer-Associate questions & answers for boosting up your success in the actual test. Databricks-Generative-AI-Engineer-Associate PDF file is the common version which many candidates often choose. If you are tired with the screen for study, you can print the Databricks-Generative-AI-Engineer-Associate Pdf Dumps into papers. With the pdf papers, you can write and make notes as you like, which is very convenient for memory. We can ensure you pass with Databricks-Generative-AI-Engineer-Associate study torrent at first time.

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

Topic	Details
Topic 1	• 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.

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	• 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 4	• 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 Certified Generative AI Engineer Associate Sample Questions (Q42-Q47):

NEW QUESTION # 42
A Generative Al Engineer is building an LLM-based application that has an important transcription (speech-to-text) task. Speed is essential for the success of the application Which open Generative Al models should be used?

- A. L!ama-2-70b-chat-hf
- B. DBRX
- C. whisper-large-v3 (1.6B)
- D. MPT-30B-Instruct

Answer: C

Explanation:
The task requires an open generative AI model for a transcription (speech-to-text) task where speed is essential. Let's assess the options based on their suitability for transcription and performance characteristics, referencing Databricks' approach to model selection.
Option A: Llama-2-70b-chat-hf
Llama-2 is a text-based LLM optimized for chat and text generation, not speech-to-text. It lacks transcription capabilities.
Databricks Reference: "Llama models are designed for natural language generation, not audio processing" ("Databricks Model Catalog").
Option B: MPT-30B-Instruct
MPT-30B is another text-based LLM focused on instruction-following and text generation, not transcription. It's irrelevant for speech-to-text tasks.
Databricks Reference: No specific mention, but MPT is categorized under text LLMs in Databricks' ecosystem, not audio models.
Option C: DBRX
DBRX, developed by Databricks, is a powerful text-based LLM for general-purpose generation. It doesn't natively support speech-to-text and isn't optimized for transcription.
Databricks Reference: "DBRX excels at text generation and reasoning tasks" ("Introducing DBRX," 2023)-no mention of audio capabilities.
Option D: whisper-large-v3 (1.6B)
Whisper, developed by OpenAI, is an open-source model specifically designed for speech-to-text transcription. The "large-v3" variant (1.6 billion parameters) balances accuracy and efficiency, with optimizations for speed via quantization or deployment on GPUs-key for the application's requirements.
Databricks Reference: "For audio transcription, models like Whisper are recommended for their speed and accuracy" ("Generative AI Cookbook," 2023). Databricks supports Whisper integration in its MLflow or Lakehouse workflows.
Conclusion: Only D. whisper-large-v3 is a speech-to-text model, making it the sole suitable choice. Its design prioritizes transcription, and its efficiency (e.g., via optimized inference) meets the speed requirement, aligning with Databricks' model deployment best practices.

NEW QUESTION # 43

A Generative Al Engineer is working with a retail company that wants to enhance its customer experience by automatically handling common customer inquiries. They are working on an LLM-powered Al solution that should improve response times while maintaining a personalized interaction. They want to define the appropriate input and LLM task to do this.
Which input/output pair will do this?

- A. Input: Customer service chat logs; Output: Find the answers to similar questions and respond with a summary
- B. Input: Customer service chat logs; Output Group the chat logs by users, followed by summarizing each user's interactions, then respond
- C. Input: Customer reviews; Output Group the reviews by users and aggregate per-user average rating, then respond
- D. Input: Customer reviews: Output Classify review sentiment

Answer: A

Explanation:
The task described in the question involves enhancing customer experience by automatically handling common customer inquiries using an LLM-powered AI solution. This requires the system to process input data (customer inquiries) and generate personalized, relevant responses efficiently. Let's evaluate the options step-by-step in the context of Databricks Generative AI Engineer principles, which emphasize leveraging LLMs for tasks like question answering, summarization, and retrieval-augmented generation (RAG).
Option A: Input: Customer reviews; Output: Group the reviews by users and aggregate per-user average rating, then respond This option focuses on analyzing customer reviews to compute average ratings per user. While this might be useful for sentiment analysis or user profiling, it does not directly address the goal of handling common customer inquiries or improving response times for personalized interactions. Customer reviews are typically feedback data, not real-time inquiries requiring immediate responses.
Databricks Reference: Databricks documentation on LLMs (e.g., "Building LLM Applications with Databricks") emphasizes that LLMs excel at tasks like question answering and conversational responses, not just aggregation or statistical analysis of reviews.
Option B: Input: Customer service chat logs; Output: Group the chat logs by users, followed by summarizing each user's interactions, then respond This option uses chat logs as input, which aligns with customer service scenarios. However, the output-grouping by users and summarizing interactions-focuses on user-specific summaries rather than directly addressing inquiries. While summarization is an LLM capability, this approach lacks the specificity of finding answers to common questions, which is central to the problem.
Databricks Reference: Per Databricks' "Generative AI Cookbook," LLMs can summarize text, but for customer service, the emphasis is on retrieval and response generation (e.g., RAG workflows) rather than user interaction summaries alone.
Option C: Input: Customer service chat logs; Output: Find the answers to similar questions and respond with a summary This option uses chat logs (real customer inquiries) as input and tasks the LLM with identifying answers to similar questions, then providing a summarized response. This directly aligns with the goal of handling common inquiries efficiently while maintaining personalization (by referencing past interactions or similar cases). It leverages LLM capabilities like semantic search, retrieval, and response generation, which are core to Databricks' LLM workflows.
Databricks Reference: From Databricks documentation ("Building LLM-Powered Applications," 2023), an exact extract states: "For customer support use cases, LLMs can be used to retrieve relevant answers from historical data like chat logs and generate concise, contextually appropriate responses." This matches Option C's approach of finding answers and summarizing them.
Option D: Input: Customer reviews; Output: Classify review sentiment
This option focuses on sentiment classification of reviews, which is a valid LLM task but unrelated to handling customer inquiries or improving response times in a conversational context. It's more suited for feedback analysis than real-time customer service.
Databricks Reference: Databricks' "Generative AI Engineer Guide" notes that sentiment analysis is a common LLM task, but it's not highlighted for real-time conversational applications like customer support.
Conclusion: Option C is the best fit because it uses relevant input (chat logs) and defines an LLM task (finding answers and summarizing) that meets the requirements of improving response times and maintaining personalized interaction. This aligns with Databricks' recommended practices for LLM-powered customer service solutions, such as retrieval-augmented generation (RAG) workflows.

NEW QUESTION # 44
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. Vector Search
- B. Lakeview
- C. Inference Tables
- D. DBSQL

Answer: C

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 # 45
After changing the response generating LLM in a RAG pipeline from GPT-4 to a model with a shorter context length that the company self-hosts, the Generative AI Engineer is getting the following error:
What TWO solutions should the Generative AI Engineer implement without changing the response generating model? (Choose two.)

- A. Reduce the maximum output tokens of the new model
- B. Use a smaller embedding model to generate
- C. Reduce the number of records retrieved from the vector database
- D. Retrain the response generating model using ALiBi
- E. Decrease the chunk size of embedded documents

Answer: C,E

Explanation:
* Problem Context: After switching to a model with a shorter context length, the error message indicating that the prompt token count has exceeded the limit suggests that the input to the model is too large.
* Explanation of Options:
* Option A: Use a smaller embedding model to generate- This wouldn't necessarily address the issue of prompt size exceeding the model's token limit.
* Option B: Reduce the maximum output tokens of the new model- This option affects the output length, not the size of the input being too large.
* Option C: Decrease the chunk size of embedded documents- This would help reduce the size of each document chunk fed into the model, ensuring that the input remains within the model's context length limitations.
* Option D: Reduce the number of records retrieved from the vector database- By retrieving fewer records, the total input size to the model can be managed more effectively, keeping it within the allowable token limits.
* Option E: Retrain the response generating model using ALiBi- Retraining the model is contrary to the stipulation not to change the response generating model.
OptionsCandDare the most effective solutions to manage the model's shorter context length without changing the model itself, by adjusting the input size both in terms of individual document size and total documents retrieved.

NEW QUESTION # 46
A Generative Al Engineer has built an LLM-based system that will automatically translate user text between two languages. They now want to benchmark multiple LLM's on this task and pick the best one. They have an evaluation set with known high quality translation examples. They want to evaluate each LLM using the evaluation set with a performant metric.
Which metric should they choose for this evaluation?

- A. RECALL metric
- B. ROUGE metric

- C. BLEU metric
- D. NDCG metric

Answer: C

Explanation:
The task is to benchmark LLMs for text translation using an evaluation set with known high-quality examples, requiring a performant metric. Let's evaluate the options.
Option A: ROUGE metric
ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measures overlap between generated and reference texts, primarily for summarization. It's less suited for translation, where precision and word order matter more.
Databricks Reference: "ROUGE is commonly used for summarization, not translation evaluation" ("Generative AI Cookbook," 2023).
Option B: BLEU metric
BLEU (Bilingual Evaluation Understudy) evaluates translation quality by comparing n-gram overlap with reference translations, accounting for precision and brevity. It's widely used, performant, and appropriate for this task.
Databricks Reference: "BLEU is a standard metric for evaluating machine translation, balancing accuracy and efficiency" ("Building LLM Applications with Databricks").
Option C: NDCG metric
NDCG (Normalized Discounted Cumulative Gain) assesses ranking quality, not text generation. It's irrelevant for translation evaluation.
Databricks Reference: "NDCG is suited for ranking tasks, not generative output scoring" ("Databricks Generative AI Engineer Guide").
Option D: RECALL metric
Recall measures retrieved relevant items but doesn't evaluate translation quality (e.g., fluency, correctness). It's incomplete for this use case.
Databricks Reference: No specific extract, but recall alone lacks the granularity of BLEU for text generation tasks.
Conclusion: Option B (BLEU) is the best metric for translation evaluation, offering a performant and standard approach, as endorsed by Databricks' guidance on generative tasks.

NEW QUESTION # 47
.....

Databricks Databricks-Generative-AI-Engineer-Associate certificate can help you a lot. It can help you improve your job and living standard, and having it can give you a great sum of wealth. Databricks certification Databricks-Generative-AI-Engineer-Associate exam is a test of the level of knowledge of IT professionals. PassExamDumps has developed the best and the most accurate training materials about Databricks Certification Databricks-Generative-AI-Engineer-Associate Exam. Now PassExamDumps can provide you the most comprehensive training materials about Databricks Databricks-Generative-AI-Engineer-Associate exam, including exam practice questions and answers.

□Valid Braindumps Databricks-Generative-AI-Engineer-Associate Ppt

- New Databricks-Generative-AI-Engineer-Associate Test Testking □ Test Databricks-Generative-AI-Engineer-Associate Vce Free □ Valid Braindumps Databricks-Generative-AI-Engineer-Associate Ppt □ Immediately open ▷ www.pdfvce.com ◁ and search for ▶ Databricks-Generative-AI-Engineer-Associate ◀ to obtain a free download □Exam Databricks-Generative-AI-Engineer-Associate Dumps
- HOT Databricks-Generative-AI-Engineer-Associate Valid Test Guide: Databricks Certified Generative AI Engineer Associate - High-quality Databricks Databricks-Generative-AI-Engineer-Associate Latest Test Format □ Copy URL （ www.practicevce.com ） open and search for [Databricks-Generative-AI-Engineer-Associate] to download for free □ □Testing Databricks-Generative-AI-Engineer-Associate Center
- Databricks-Generative-AI-Engineer-Associate Reliable Test Question □ New Databricks-Generative-AI-Engineer-Associate Test Syllabus □ Databricks-Generative-AI-Engineer-Associate Authorized Exam Dumps □ Search for [Databricks-Generative-AI-Engineer-Associate] and easily obtain a free download on ▷ www.pdfvce.com ◁ □Databricks-Generative-AI-Engineer-Associate Reliable Test Question
- Databricks-Generative-AI-Engineer-Associate Exam Papers □ Testing Databricks-Generative-AI-Engineer-Associate Center ↔ Databricks-Generative-AI-Engineer-Associate Latest Test Practice □ Search for 「 Databricks-Generative-AI-Engineer-Associate 」 and download it for free on ➡ www.prepawayete.com □□□ website □Databricks-Generative-AI-Engineer-Associate Valid Study Questions
- Free PDF Quiz 2026 Databricks Trustable Databricks-Generative-AI-Engineer-Associate: Databricks Certified Generative AI Engineer Associate Valid Test Guide □ Search for ➠ Databricks-Generative-AI-Engineer-Associate □ and download exam materials for free through □ www.pdfvce.com □ □Databricks-Generative-AI-Engineer-Associate Exam Papers
- Exam Databricks-Generative-AI-Engineer-Associate Dumps □ Databricks-Generative-AI-Engineer-Associate Discount Code □ Databricks-Generative-AI-Engineer-Associate Exam Voucher □ Download { Databricks-Generative-AI-Engineer-Associate } for free by simply entering ➠ www.practicevce.com □ website □Reliable Databricks-Generative-AI-Engineer-Associate Dumps Ebook
- www.stes.tyc.edu.tw, 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, myportal.utt.edu.tt, ddy.hackp.net, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, bbs.theviko.com, Disposable vapes

BONUS!!! Download part of PassExamDumps Databricks-Generative-AI-Engineer-Associate dumps for free: https://drive.google.com/open?id=1eAeZrzVCix7vbMOik0CwS_HVnO4qXqyH