

試験の準備方法-実用的なDatabricks-Generative-AI-Engineer-Associate復習過去問試験-高品質なDatabricks-Generative-AI-Engineer-Associate問題トレーニング



P.S.TopexamがGoogle Driveで共有している無料の2026 Databricks Databricks-Generative-AI-Engineer-Associateダウンロード: <https://drive.google.com/open?id=1whlPEubICqB8hc5SntNA51hhTl1LE2pf>

Databricks-Generative-AI-Engineer-Associate学習教材は、当初の目標を達成し、仕事のキャリアをよりスムーズにし、家族の生活の質を向上させるのに役立ちます。Databricks-Generative-AI-Engineer-Associate試験トレントを20~30時間学習するだけで、DatabricksのDatabricks-Generative-AI-Engineer-Associate試験に自信を持って参加できると言っても過言ではありません。そして、10年以上にわたってこのキャリアでプロフェッショナルであったため、あなたの成功を確実にすることができます。そして、数千人の候補者が、優れたDatabricks-Generative-AI-Engineer-Associateトレーニング資料の助けを借りて、Databricks Certified Generative AI Engineer Associate夢と野望を達成しました。

Topexamは、Databricks期待されるスコアを達成してDatabricks-Generative-AI-Engineer-Associate認定を取得する価値のあるクライアントにチャンスを与えるための非常に素晴らしい効果的なプラットフォームです。プロの専門家のおかげで努力により、Databricks-Generative-AI-Engineer-Associate試験トレントには、タイミング機能を備えた模擬試験システムが装備されており、Databricks Certified Generative AI Engineer Associate学習結果をいつでも確認し、欠陥をチェックし続け、体力を改善できます。あなたが学生であろうとオフィスワーカーであろうと、ここで満足することができ、Databricks-Generative-AI-Engineer-Associate試験トレントを選択しても後悔することはあ

りません。

>> Databricks-Generative-AI-Engineer-Associate復習過去問 <<

Databricks-Generative-AI-Engineer-Associate問題トレーニング & Databricks-Generative-AI-Engineer-Associate試験勉強攻略

あなたは今DatabricksのDatabricks-Generative-AI-Engineer-Associate試験のために準備していますか。そうであれば、あなたは夢がある人だと思います。我々Topexamはあなたのような人に夢を叶えさせるという目標を持っています。我々の開発するDatabricksのDatabricks-Generative-AI-Engineer-Associateソフトは最新で最も豊富な問題集を含めています。あなたは我々の商品を購入したら、一年間の無料更新サービスを得られています。我々のソフトを利用してDatabricksのDatabricks-Generative-AI-Engineer-Associate試験に合格するのは全然問題ないです。

Databricks Certified Generative AI Engineer Associate 認定 Databricks-Generative-AI-Engineer-Associate 試験問題 (Q20-Q25):

質問 # 20

A Generative AI Engineer is developing an agent system using a popular agent-authoring library. The agent comprises multiple parallel and sequential chains. The engineer encounters challenges as the agent fails at one of the steps, making it difficult to debug the root cause. They need to find an appropriate approach to research this issue and discover the cause of failure. Which approach do they choose?

- A. Implement structured logging within the agent's code to capture detailed execution information.
- B. Run `MLflow.evaluate` to determine root cause of failed step.
- C. Enable MLflow tracing to gain visibility into each agent's behavior and execution step.
- D. Deconstruct the agent into independent steps to simplify debugging.

正解: C

解説:

For complex agentic systems (like those built with LangGraph or Autogen), standard logging is often insufficient because the "state" of the agent changes dynamically. MLflow Tracing is the designated Generative AI engineering standard for debugging these systems. Tracing provides a visual, hierarchical timeline of every call made during an agent's execution-including internal LLM reasoning, tool calls, and data transformations. When a step fails, the trace allows the engineer to click into that specific node to see the exact input sent to the LLM and the raw output received. This is much faster and more comprehensive than manually deconstructing the agent (D) or adding manual logs (C). While `mlflow.evaluate` (B) is useful for measuring performance across a whole dataset, it is not a tool for real-time debugging of a single execution failure.

質問 # 21

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

正解: B

解説:

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.

質問 # 22

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. DBSQL
- C. Inference Tables
- D. Lakeview

正解: C

解説:

* Problem Context: The goal is to monitor the serving endpoint for incoming requests and outgoing responses in a provisioned throughput model serving endpoint within a Retrieval-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 Tables in Databricks are designed to store the results and metadata of inference runs. This allows the system to log incoming requests and outgoing responses directly 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 Tables are the optimal feature for monitoring request and response logs within the Databricks infrastructure for a model serving endpoint.

質問 # 23

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

正解: B、D

解説:

* 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.

Options C and D are 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.

質問 # 24

An AI developer team wants to fine-tune an open-weight model to have exceptional performance on a code generation use case. They are trying to choose the best model to start with. They want to minimize model hosting costs and are using Hugging Face model cards and spaces to explore models. Which TWO model attributes and metrics should the team focus on to make their selection?

- A. Number of model parameters
- B. Chatbot Arena Leaderboard
- C. MTEB Leaderboard
- D. Big Code Models Leaderboard
- E. Number of model downloads last month

正解: A、D

解説:

To optimize for code generation performance and hosting costs, a Generative AI engineer must look at specific metrics.

Big Code Models Leaderboard (A): This is the industry-standard benchmark for code-specific LLMs (like StarCoder or CodeLlama). It measures performance on tasks like HumanEval and MBPP, providing a direct indicator of how well the model handles programming logic.

Number of model parameters (B): This is the primary driver of hosting costs. Larger models (e.g., 70B) require more GPU memory (VRAM) and more expensive compute instances (like A100s/H100s) than smaller models (e.g., 7B or 13B). To minimize costs, the team should look for the smallest model that achieves a high score on the Big Code Leaderboard.

Note: MTEB (C) is for embeddings, and Chatbot Arena (D) is for general-purpose chat, neither of which is the primary metric for specialized code generation fine-tuning.

質問 # 25

.....

我々TopexamはDatabricksのDatabricks-Generative-AI-Engineer-Associate試験問題集をリリースする以降、多くのお客様の好評を博したのは弊社にとって、大変な名誉なことです。また、我々はさらに認可を受けられるために、皆様の一切の要求を満足できて喜ぶ気持ちでずっと協力し、完備かつ精確のDatabricks-Generative-AI-Engineer-Associate試験問題集を開発するのに準備します。

Databricks-Generative-AI-Engineer-Associate問題トレーニング: https://www.topexam.jp/Databricks-Generative-AI-Engineer-Associate_shiken.html

クライアントのみがリンクをクリックすると、すぐにソフトウェアにログインしてDatabricks-Generative-AI-Engineer-Associateガイド資料を学習できます、Databricks-Generative-AI-Engineer-Associate試験は、良い選択です、Databricks Databricks-Generative-AI-Engineer-Associate復習過去問 あなたは復習資料に悩んでいるかもしれませんが、Databricks Databricks-Generative-AI-Engineer-Associate復習過去問 「信仰は偉大な感情で、創造の力になれる、heしないでください、Databricks Databricks-Generative-AI-Engineer-Associate復習過去問 待つことなく15分以内に入手できます、Databricks Databricks-Generative-AI-Engineer-Associate復習過去問 シミュレーション後、試験環境、試験プロセス、試験概要をより明確に理解できます、すべてのお客様向けのDatabricks-Generative-AI-Engineer-Associate学習ガイドであり、有名なブランドを活用したくない。

坪井もクールな顔に似合わず、意外と飲むようで、勧められるビールを断ることは一度もなかった、啞然として月島の顔を見つめる、クライアントのみがリンクをクリックすると、すぐにソフトウェアにログインしてDatabricks-Generative-AI-Engineer-Associateガイド資料を学習できます。

実用的なDatabricks-Generative-AI-Engineer-Associate復習過去問試験-試

BONUS!!! Topexam Databricks-Generative-AI-Engineer-Associateダンプの一部を無料でダウンロード：<https://drive.google.com/open?id=1whlPEubICqB8hc5SntNA51hhTI1LE2pf>