

Neueste Databricks-Generative-AI-Engineer-Associate Pass Guide & neue Prüfung Databricks-Generative-AI-Engineer-Associate braindumps & 100% Erfolgsquote



Übrigens, Sie können die vollständige Version der ZertPruefung Databricks-Generative-AI-Engineer-Associate Prüfungsfragen aus dem Cloud-Speicher herunterladen: https://drive.google.com/open?id=1uFIYKYhKLova_quffHxmzrUzHraBEuc

In den letzten Jahren entwickelt sich die IT-Branche sehr schnell. Viele Leute fangen an, IT-Kenntnisse zu lernen. Sie geben viel Mühe aus, um eine bessere Zukunft zu haben. Die Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung ist eine unentbehrliche Zertifizierungsprüfung in der IT-Branche. Viele Leute machen sich große Sorgen um die Prüfung. Heute empfehle ich Ihnen eine gute Methode, nämlich, die Fragenkataloge zur Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung von ZertPruefung zu kaufen. Sie können Ihnen helfen, die Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung 100% zu bestehen. Sonst geben wir Ihnen eine volle Rückerstattung. Und Sie würden keine Verluste erleiden.

ZertPruefung ist eine Website, die die Erfolgsquote von Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung erhöhen kann. Die erfahrungsreichen IT-Experten entwickeln ständig eine Vielzahl von Programmen, um zu garantieren, dass Sie die Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung 100% erfolgreich bestehen können. Die Trainingsmaterialien von ZertPruefung sind sehr effektiv. Viele IT-Leute, die die Databricks Databricks-Generative-AI-Engineer-Associate Prüfung bestanden haben, haben die Prüfungsfragen und Antworten von ZertPruefung benutzt. Mit Hilfe des ZertPruefung haben viele auch die Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung bestanden. Wenn Sie ZertPruefung wählen, kommt der Erfolg auf Sie zu.

>> Databricks-Generative-AI-Engineer-Associate Lerntipps <<

Kostenlose gültige Prüfung Databricks Databricks-Generative-AI-Engineer-Associate Sammlung - Examcollection

Unser ZertPruefung ist eine Website, die eine lange Geschichte hinter sich hat. So genießt ZertPruefung einen guten Ruf in der IT-Branche. Und wir haben vielen Kandidaten geholfen, die Databricks Databricks-Generative-AI-Engineer-Associate Prüfung zu bestehen. Die Fragen und Antworten zur Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung von ZertPruefung werden von den erfahrungsreichen Expertenteams nach ihren Kenntnissen und Erfahrungen bearbeitet. Wenn Sie an der Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung teilnehmen wollen, ist ZertPruefung zweifellos eine gute Wahl.

Databricks Certified Generative AI Engineer Associate Databricks-Generative-AI-Engineer-Associate Prüfungsfragen mit Lösungen (Q18-Q23):

18. Frage

A Generative AI Engineer is creating an LLM-based application. The documents for its retriever have been chunked to a maximum of 512 tokens each. The Generative AI Engineer knows that cost and latency are more important than quality for this application. They have several context length levels to choose from. Which will fulfill their need?

- A. context length 2048: smallest model is 11GB and embedding dimension 2560
- **B. context length 512: smallest model is 0.13GB and embedding dimension 384**
- C. context length 514; smallest model is 0.44GB and embedding dimension 768
- D. context length 32768: smallest model is 14GB and embedding dimension 4096

Antwort: B

Begründung:

When prioritizing cost and latency over quality in a Large Language Model (LLM)-based application, it is crucial to select a configuration that minimizes both computational resources and latency while still providing reasonable performance. Here's why Dis is the best choice:

* Context length: The context length of 512 tokens aligns with the chunk size used for the documents (maximum of 512 tokens per chunk). This is sufficient for capturing the needed information and generating responses without unnecessary overhead.

* Smallest model size: The model with a size of 0.13GB is significantly smaller than the other options.

This small footprint ensures faster inference times and lower memory usage, which directly reduces both latency and cost.

* Embedding dimension: While the embedding dimension of 384 is smaller than the other options, it is still adequate for tasks where cost and speed are more important than precision and depth of understanding.

This setup achieves the desired balance between cost-efficiency and reasonable performance in a latency- sensitive, cost-conscious application.

19. Frage

A Generative AI Engineer has successfully ingested unstructured documents and chunked them by document sections. They would like to store the chunks in a Vector Search index. The current format of the dataframe has two columns: (i) original document file name (ii) an array of text chunks for each document.

What is the most performant way to store this dataframe?

- A. Split the data into train and test set, create a unique identifier for each document, then save to a Delta table
- B. Store each chunk as an independent JSON file in Unity Catalog Volume. For each JSON file, the key is the document section name and the value is the array of text chunks for that section
- C. First create a unique identifier for each document, then save to a Delta table
- **D. Flatten the dataframe to one chunk per row, create a unique identifier for each row, and save to a Delta table**

Antwort: D

Begründung:

* Problem Context: The engineer needs an efficient way to store chunks of unstructured documents to facilitate easy retrieval and search. The current dataframe consists of document filenames and associated text chunks.

* Explanation of Options:

* Option A: Splitting into train and test sets is more relevant for model training scenarios and not directly applicable to storage for retrieval in a Vector Search index.

* Option B: Flattening the dataframe such that each row contains a single chunk with a unique identifier is the most performant for storage and retrieval. This structure aligns well with how data is indexed and queried in vector search applications, making it easier to retrieve specific chunks efficiently.

* Option C: Creating a unique identifier for each document only does not address the need to access individual chunks efficiently, which is critical in a Vector Search application.

* Option D: Storing each chunk as an independent JSON file creates unnecessary overhead and complexity in managing and querying large volumes of files.

Option B is the most efficient and practical approach, allowing for streamlined indexing and retrieval processes in a Delta table environment, fitting the requirements of a Vector Search index.

20. Frage

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

Antwort: C

Begründung:

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.

21. Frage

A Generative AI Engineer is creating an LLM system that will retrieve news articles from the year 1918 and related to a user's query and summarize them. The engineer has noticed that the summaries are generated well but often also include an explanation of how the summary was generated, which is undesirable.

Which change could the Generative AI Engineer perform to mitigate this issue?

- A. Tune the chunk size of news articles or experiment with different embedding models.
- B. Split the LLM output by newline characters to truncate away the summarization explanation.
- C. Revisit their document ingestion logic, ensuring that the news articles are being ingested properly.
- **D. Provide few shot examples of desired output format to the system and/or user prompt.**

Antwort: D

Begründung:

To mitigate the issue of the LLM including explanations of how summaries are generated in its output, the best approach is to adjust the training or prompt structure. Here's why Option D is effective:

* Few-shot Learning: By providing specific examples of how the desired output should look (i.e., just the summary without explanation), the model learns the preferred format. This few-shot learning approach helps the model understand not only what content to generate but also how to format its responses.

* Prompt Engineering: Adjusting the user prompt to specify the desired output format clearly can guide the LLM to produce summaries without additional explanatory text. Effective prompt design is crucial in controlling the behavior of generative models.

Why Other Options Are Less Suitable:

* A: While technically feasible, splitting the output by newline and truncating could lead to loss of important content or create awkward breaks in the summary.

* B: Tuning chunk sizes or changing embedding models does not directly address the issue of the model's tendency to generate explanations along with summaries.

* C: Revisiting document ingestion logic ensures accurate source data but does not influence how the model formats its output.

By using few-shot examples and refining the prompt, the engineer directly influences the output format, making this approach the most targeted and effective solution.

22. Frage

A Generative AI Engineer is building a production-ready LLM system which replies directly to customers.

The solution makes use of the Foundation Model API via provisioned throughput. They are concerned that the LLM could potentially respond in a toxic or otherwise unsafe way. They also wish to perform this with the least amount of effort.

Which approach will do this?

- A. Ask users to report unsafe responses
- **B. Host Llama Guard on Foundation Model API and use it to detect unsafe responses**
- C. Add a regex expression on inputs and outputs to detect unsafe responses.
- D. Add some LLM calls to their chain to detect unsafe content before returning text

Antwort: B

Begründung:

The task is to prevent toxic or unsafe responses in an LLM system using the Foundation Model API with minimal effort. Let's assess the options.

* Option A: Host Llama Guard on Foundation Model API and use it to detect unsafe responses

* Llama Guard is a safety-focused model designed to detect toxic or unsafe content. Hosting it via the Foundation Model API (a Databricks service) integrates seamlessly with the existing system, requiring minimal setup (just deployment and a check step), and leverages provisioned throughput for performance.

* Databricks Reference: "Foundation Model API supports hosting safety models like Llama Guard to filter outputs efficiently" ("Foundation Model API Documentation," 2023).

* Option B: Add some LLM calls to their chain to detect unsafe content before returning text

* Using additional LLM calls (e.g., prompting an LLM to classify toxicity) increases latency, complexity, and effort (crafting prompts, chaining logic), and lacks the specificity of a dedicated safety model.

* Databricks Reference: "Ad-hoc LLM checks are less efficient than purpose-built safety solutions" ("Building LLM Applications with Databricks").

* Option C: Add a regex expression on inputs and outputs to detect unsafe responses

* Regex can catch simple patterns (e.g., profanity) but fails for nuanced toxicity (e.g., sarcasm, context-dependent harm), requiring significant manual effort to maintain and update rules.

* Databricks Reference: "Regex-based filtering is limited for complex safety needs" ("Generative AI Cookbook").

* Option D: Ask users to report unsafe responses

* User reporting is reactive, not preventive, and places burden on users rather than the system. It doesn't limit unsafe outputs proactively and requires additional effort for feedback handling.

* Databricks Reference: "Proactive guardrails are preferred over user-driven monitoring" ("Databricks Generative AI Engineer Guide").

Conclusion: Option A (Llama Guard on Foundation Model API) is the least-effort, most effective approach, leveraging Databricks' infrastructure for seamless safety integration.

23. Frage

.....

Möchten Sie die Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung mühelos bestehen? Dann sind die Fragenkataloge zur Databricks-Generative-AI-Engineer-Associate Zertifizierung aus ZertPrüfung unerlässlich. Die Fragenpool zur Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung aus ZertPrüfung werden von den erfahrenen Experten durch ständige Praxis entworfen, sie sind eine Kombination aus Fragen und Antworten. Deswegen ist die Webseite ZertPrüfung die Beste. Wählen Sie ZertPrüfung, wartet eine schönere Zukunft auf Sie da.

Databricks-Generative-AI-Engineer-Associate PDF Demo: https://www.zertpruefung.ch/Databricks-Generative-AI-Engineer-Associate_exam.html

Databricks-Generative-AI-Engineer-Associate Lerntipps Wir werden Ihr Produkt in Ihre gültige Mailbox senden, Wir verkaufen nur die neueste Version Databricks-Generative-AI-Engineer-Associate Dumps Guide Materialien, Databricks-Generative-AI-Engineer-Associate Lerntipps Das ist wirklich eine gute Wahl, Databricks-Generative-AI-Engineer-Associate Lerntipps Unsere Erziehungsexperten sind in dieser Reihe viele Jahre erlebt, Databricks-Generative-AI-Engineer-Associate Lerntipps Und Ihre späte Arbeit und Alltagsleben werden sicher interessanter sein.

Zu den Unterbeschäftigungsquoten zählen Arbeitslose, diejenigen, die Teilzeit Databricks-Generative-AI-Engineer-Associate arbeiten, aber einen unfreiwilligen Vollzeitjob in Vollzeit wünschen, und diejenigen, die einen Job wollen, aber das Zuschauen aufgeben.

Databricks-Generative-AI-Engineer-Associate Braindumpsit Dumps PDF & Databricks-Generative-AI-Engineer-Associate Braindumpsit IT-Zertifizierung - Testking Examen Dumps

- Übrigens, Sie können die vollständige Version der ZertPrüfung Databricks-Generative-AI-Engineer-Associate Prüfungsfragen aus dem Cloud-Speicher herunterladen: https://drive.google.com/open?id=1uFYKYhKLova_quffHxmzrUzHraBEuc