

Databricks-Generative-AI-Engineer-Associate zu bestehen mit allseitigen Garantien



2026 Die neuesten ZertPruefung Databricks-Generative-AI-Engineer-Associate PDF-Versionen Prüfungsfragen und Databricks-Generative-AI-Engineer-Associate Fragen und Antworten sind kostenlos verfügbar: <https://drive.google.com/open?id=17TcLCLCSIV6dNv5BIJVmj6fFuFeqnDrh>

Nach der Schulzeit haben wir mehr Verantwortungen und die Zeit fürs Lernen vermindert sich. Wenn Sie sich im IT-Bereich besser entwickeln möchten, dann ist die internationale Zertifizierungsprüfung wie Databricks Databricks-Generative-AI-Engineer-Associate Prüfung zu bestehen sehr notwendig. Wir ZertPruefung bieten Sie mit alle Kräfte vieler IT-Profis die effektivste Hilfe bei der Databricks Databricks-Generative-AI-Engineer-Associate Prüfung. 3 Versionen (PDF, online sowie Software) von Databricks Databricks-Generative-AI-Engineer-Associate Prüfungsunterlagen haben Ihre besondere Überlegenheit. Dadurch, dass Sie die kostenlose Demos probieren, können Sie nach Ihre Gewohnheiten die geeignete Version wählen.

Databricks Databricks-Generative-AI-Engineer-Associate Prüfungsplan:

Thema	Einzelheiten
Thema 1	<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.
Thema 2	<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.

Thema 3	<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.
Thema 4	<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.
Thema 5	<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.

>> Databricks-Generative-AI-Engineer-Associate Quizfragen Und Antworten <<

Databricks-Generative-AI-Engineer-Associate PDF Demo - Databricks-Generative-AI-Engineer-Associate Online Tests

Um die Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung zu bestehen, wählen Sie doch unseren ZertPruefung. Sie werden sicher nicht bereuen, dass Sie mit so wenigem Geld die Prüfung bestehen können. Unser ZertPruefung wird Ihnen helfen, sich auf die Prüfung gut vorzubereiten und die Databricks Databricks-Generative-AI-Engineer-Associate Zertifizierungsprüfung (Databricks Certified Generative AI Engineer Associate) erfolgreich zu bestehen. Außerdem bieten wir Ihnen kostenlos einen einjährigen Update-Service.

Databricks Certified Generative AI Engineer Associate Databricks-Generative-AI-Engineer-Associate Prüfungsfragen mit Lösungen (Q53-Q58):

53. Frage

A Generative AI Engineer at an automotive company would like to build a question-answering chatbot for customers to inquire about their vehicles. They have a database containing various documents of different vehicle makes, their hardware parts, and common maintenance information.

Which of the following components will NOT be useful in building such a chatbot?

- A. Vector database
- B. Response-generating LLM
- C. Embedding model
- D. Invite users to submit long, rather than concise, questions

Antwort: D

Begründung:

The task involves building a question-answering chatbot for an automotive company using a database of vehicle-related documents. The chatbot must efficiently process customer inquiries and provide accurate responses. Let's evaluate each component to determine which is not useful, per Databricks Generative AI Engineer principles.

* Option A: Response-generating LLM

* An LLM is essential for generating natural language responses to customer queries based on retrieved information. This is a core component of any chatbot.

* Databricks Reference: "The response-generating LLM processes retrieved context to produce coherent answers" ("Building LLM Applications with Databricks," 2023).

* Option B: Invite users to submit long, rather than concise, questions

* Encouraging long questions is a user interaction design choice, not a technical component of the chatbot's architecture. Moreover, long, verbose questions can complicate intent detection and retrieval, reducing efficiency and accuracy-counter to best practices for chatbot design. Concise questions are typically preferred for clarity and performance.

* Databricks Reference: While not explicitly stated, Databricks' "Generative AI Cookbook" emphasizes efficient query processing, implying that simpler, focused inputs improve LLM performance. Inviting long questions doesn't align with this.

* Option C: Vector database

* A vector database stores embeddings of the vehicle documents, enabling fast retrieval of relevant information via semantic search. This is critical for a question-answering system with a large document corpus.

* Databricks Reference: "Vector databases enable scalable retrieval of context from large datasets" ("Databricks Generative AI Engineer Guide").

* Option D: Embedding model

* An embedding model converts text (documents and queries) into vector representations for similarity search. It's a foundational component for retrieval-augmented generation (RAG) in chatbots.

* Databricks Reference: "Embedding models transform text into vectors, facilitating efficient matching of queries to documents" ("Building LLM-Powered Applications").

Conclusion: Option B is not a useful component in building the chatbot. It's a user-facing suggestion rather than a technical building block, and it could even degrade performance by introducing unnecessary complexity. Options A, C, and D are all integral to a Databricks-aligned chatbot architecture.

54. Frage

A Generative AI Engineer is responsible for developing a chatbot to enable their company's internal HelpDesk Call Center team to more quickly find related tickets and provide resolution. While creating the GenAI application work breakdown tasks for this project, they realize they need to start planning which data sources (either Unity Catalog volume or Delta table) they could choose for this application. They have collected several candidate data sources for consideration:

call_rep_history: a Delta table with primary keys representative_id, call_id. This table is maintained to calculate representatives' call resolution from fields call_duration and call_start_time.

transcript Volume: a Unity Catalog Volume of all recordings as *.wav files, but also a text transcript as *.txt files.

call_cust_history: a Delta table with primary keys customer_id, call_id. This table is maintained to calculate how much internal customers use the HelpDesk to make sure that the charge back model is consistent with actual service use.

call_detail: a Delta table that includes a snapshot of all call details updated hourly. It includes root_cause and resolution fields, but those fields may be empty for calls that are still active.

maintenance_schedule - a Delta table that includes a listing of both HelpDesk application outages as well as planned upcoming maintenance downtimes.

They need sources that could add context to best identify ticket root cause and resolution.

Which TWO sources do that? (Choose two.)

- A. call_detail
- B. call_rep_history
- C. call_cust_history
- D. maintenance_schedule
- E. transcript Volume

Antwort: A,E

Begründung:

In the context of developing a chatbot for a company's internal HelpDesk Call Center, the key is to select data sources that provide the most contextual and detailed information about the issues being addressed. This includes identifying the root cause and suggesting resolutions. The two most appropriate sources from the list are:

* Call Detail (Option D):

* Contents: This Delta table includes a snapshot of all call details updated hourly, featuring essential fields like root_cause and resolution.

* Relevance: The inclusion of root_cause and resolution fields makes this source particularly valuable, as it directly contains the information necessary to understand and resolve the issues discussed in the calls. Even if some records are incomplete, the data provided is crucial for a chatbot aimed at speeding up resolution identification.

* Transcript Volume (Option E):

* Contents: This Unity Catalog Volume contains recordings in .wav format and text transcripts in .txt files.

* Relevance: The text transcripts of call recordings can provide in-depth context that the chatbot can analyze to understand the nuances of each issue. The chatbot can use natural language processing techniques to extract themes, identify problems, and suggest resolutions based on previous similar interactions documented in the transcripts.

Why Other Options Are Less Suitable:

* A (Call Cust History): While it provides insights into customer interactions with the HelpDesk, it focuses more on the usage metrics rather than the content of the calls or the issues discussed.

* B (Maintenance Schedule): This data is useful for understanding when services may not be available but does not contribute directly to resolving user issues or identifying root causes.

* C (Call Rep History): Though it offers data on call durations and start times, which could help in assessing performance, it lacks direct information on the issues being resolved.

Therefore, Call Detail and Transcript Volume are the most relevant data sources for a chatbot designed to assist with identifying and resolving issues in a HelpDesk Call Center setting, as they provide direct and contextual information related to customer issues.

55. Frage

A Generative AI Engineer is ready to deploy an LLM application written using Foundation Model APIs. They want to follow security best practices for production scenarios. Which authentication method should they choose?

- A. Use a frequently rotated access token belonging to either a workspace user or a service principal
- B. Use an access token belonging to any workspace user
- C. Use OAuth machine-to-machine authentication
- **D. Use an access token belonging to service principals**

Antwort: D

Begründung:

The task is to deploy an LLM application using Foundation Model APIs in a production environment while adhering to security best practices. Authentication is critical for securing access to Databricks resources, such as the Foundation Model API. Let's evaluate the options based on Databricks' security guidelines for production scenarios.

* Option A: Use an access token belonging to service principals

* Service principals are non-human identities designed for automated workflows and applications in Databricks. Using an access token tied to a service principal ensures that the authentication is scoped to the application, follows least-privilege principles (via role-based access control), and avoids reliance on individual user credentials. This is a security best practice for production deployments.

* Databricks Reference: "For production applications, use service principals with access tokens to authenticate securely, avoiding user-specific credentials" ("Databricks Security Best Practices," 2023). Additionally, the "Foundation Model API Documentation" states: "Service principal tokens are recommended for programmatic access to Foundation Model APIs."

* Option B: Use a frequently rotated access token belonging to either a workspace user or a service principal

* Frequent rotation enhances security by limiting token exposure, but tying the token to a workspace user introduces risks (e.g., user account changes, broader permissions). Including both user and service principal options dilutes the focus on application-specific security, making this less ideal than a service-principal-only approach. It also adds operational overhead without clear benefits over Option A.

* Databricks Reference: "While token rotation is a good practice, service principals are preferred over user accounts for application authentication" ("Managing Tokens in Databricks," 2023).

* Option C: Use OAuth machine-to-machine authentication

* OAuth M2M (e.g., client credentials flow) is a secure method for application-to-service communication, often using service principals under the hood. However, Databricks' Foundation Model API primarily supports personal access tokens (PATs) or service principal tokens over full OAuth flows for simplicity in production setups. OAuth M2M adds complexity (e.g., managing refresh tokens) without a clear advantage in this context.

* Databricks Reference: "OAuth is supported in Databricks, but service principal tokens are simpler and sufficient for most API-based workloads" ("Databricks Authentication Guide," 2023).

* Option D: Use an access token belonging to any workspace user

* Using a user's access token ties the application to an individual's identity, violating security best practices. It risks exposure if the user leaves, changes roles, or has overly broad permissions, and it's not scalable or auditable for production.

* Databricks Reference: "Avoid using personal user tokens for production applications due to security and governance concerns" ("Databricks Security Best Practices," 2023).

Conclusion: Option A is the best choice, as it uses a service principal's access token, aligning with Databricks' security best practices for production LLM applications. It ensures secure, application-specific authentication with minimal complexity, as explicitly recommended for Foundation Model API deployments.

56. Frage

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

Antwort: C,D

57. Frage

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

Antwort: A,C

Begründung:

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.

58. Frage

.....

Haben Sie Databricks Databricks-Generative-AI-Engineer-Associate Dumps von ZertPruefung benutzt? Diese Dumps beinhalten die aktualisierten Prüfungsfragen, die auch alle mögliche Prüfungsfragen in der aktuellen Prüfung vorhanden sind. Es kann Ihnen garantieren, nur einmal die Databricks Databricks-Generative-AI-Engineer-Associate Prüfung zu bestehen. Diese Dumps kann Ihnen helfen, unglaubliche Ergebnisse zu bekommen. Wenn Sie in der Databricks Databricks-Generative-AI-Engineer-Associate Prüfung durchgefallen sind, geben wir Ihnen voll Geld zurück. Deshalb müssen Sie sorglos diese Dumps benutzen. Sie können den Erfolg erreichen, wenn Sie die Prüfungsunterlagen von ZertPruefung benutzen.

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

- Databricks-Generative-AI-Engineer-Associate Exam Fragen ♥ Databricks-Generative-AI-Engineer-Associate Exam Fragen Databricks-Generative-AI-Engineer-Associate Übungsmaterialien Suchen Sie jetzt auf ⇒ www.zertfragen.com ⇐ nach [Databricks-Generative-AI-Engineer-Associate] um den kostenlosen Download zu erhalten
- Databricks-Generative-AI-Engineer-Associate Musterprüfungsfragen - Databricks-Generative-AI-Engineer-Associate Zertifizierung - Databricks-Generative-AI-Engineer-Associate Testfragen Suchen Sie jetzt auf ⇒ www.itzert.com nach ➡ Databricks-Generative-AI-Engineer-Associate und laden Sie es kostenlos herunter
- Databricks-Generative-AI-Engineer-Associate examkiller gültige Ausbildung Dumps - Databricks-Generative-AI-Engineer-Associate Prüfung Überprüfung Torrents Erhalten Sie den kostenlosen Download von “ Databricks-Generative-AI-

