

New GES-C01 Dumps Questions - GES-C01 Valid Test Vce Free

```
 Choose a base model with a small context window (e.g., llama2-8b) for fine-tuning to ensure faster training and inference times, even if it might truncate some training examples.
 Deploy the fine-tuned model via Snowpark Container Services (SPCS) to a GPU_NV_5 compute pool to leverage GPU acceleration for inference.
 Ensure the training data for fine-tuning precisely fills the context window for each prompt-completion pair to maximize the model's learning from available tokens.
 Set the max_epochs parameter to 1 when calling #CREATE ('CREATE') to reduce training time, thereby directly improving inference latency.
 Provide a rich set of examples and descriptions for each of the 50 categories in the AI_CLASSIFY function call to allow the LLM to perform zero-shot classification without fine-tuning, thereby avoiding fine-tuning latency.
```

2026 Latest Exam-Killer GES-C01 PDF Dumps and GES-C01 Exam Engine Free Share: https://drive.google.com/open?id=13cp_0uQsfS_4vWuriKnaJm06xFexQZE_

As what have been demonstrated in the records concerning the pass rate of our GES-C01 free demo, our pass rate has kept the historical record of 98% to 99% from the very beginning of their foundation. During these years, our PDF version of our GES-C01 study engine stays true to its original purpose to pursue a higher pass rate that has never been attained in the past. And you will be content about our considerate service on our GES-C01 training guide. If you have any question, you can just contact us!

Having a general review of what you have learnt is quite necessary, since it will make you have a good command of the knowledge points. GES-C01 Online test engine is convenient and easy to learn, and it has the testing history and performance review. It supports all web browsers, and you can also have offline practice. Before buying GES-C01 Exam Dumps, you can try free demo first, so that you can have a deeper understanding of the exam. We have online and offline chat service for GES-C01 training materials. If you have any questions, you can contact us, and we will give you reply as quickly as we can.

>> New GES-C01 Dumps Questions <<

Free PDF Quiz 2026 Snowflake - New GES-C01 Dumps Questions

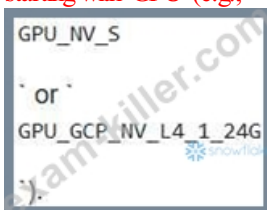
To let the client be familiar with the atmosphere of the GES-C01 exam we provide the function to stimulate the exam and the timing function of our study materials to adjust your speed to answer the questions. We provide the stimulation, the instances and the diagrams to explain the hard-to-understand contents of our GES-C01 Study Materials. For these great merits we can promise to you that if you buy our GES-C01 study materials you will pass the test with few difficulties.

Snowflake SnowPro® Specialty: Gen AI Certification Exam Sample Questions (Q247-Q252):

NEW QUESTION # 247

A development team plans to utilize Snowpark Container Services (SPCS) for deploying a variety of AI/ML workloads, including custom LLMs and GPU-accelerated model training jobs. They are in the process of creating a compute pool and need to select the appropriate instance families and configurations. Which of the following statements about 'CREATE COMPUTE POOL' in SPCS are accurate?

- A. To support GPU-accelerated LLM inference and training, the 'INSTANCE_FAMILY' must be selected from a type starting with 'GPU' (e.g.,



- B. The 'MIN NODES' and 'MAX NODES' parameters define the scaling range for the compute pool, and Snowflake automatically scales the pool within this range based on workload demand.
- C. Snowpark-optimized warehouses are the recommended compute pool type for all large-scale ML training workloads

within SPCS due to their enhanced memory limits and CPU architectures.

- D. Setting 'AUTO RESUME = TRUE ensures that the compute pool automatically starts when a service or job is submitted to it, rather than requiring manual resumption.
- E. For cost optimization, 'AUTO SUSPEND SECS = 0' should be used to prevent automatic suspension of the compute pool, as suspension and resumption incur minimum billing durations.

Answer: A,D

Explanation:

Option A is correct. GPU-accelerated workloads, such as LLM inference and model training, require instance families specifically designed with GPUs. The documentation lists instance family names starting with 'GPU' for this purpose, such as 'GPU_GCP NV L4'. Option B is incorrect. While 'MIN NODES' and 'MAX NODES' define the range, the size of compute clusters in Snowpark Container Services does "not" auto-scale dynamically based on workload demand. Users must manually alter the number of instances at runtime using commands like 'ALTER SERVICE MIN INSTANCES = s'. Snowflake does handle load balancing across instances within the configured node counts. Option C is correct. The 'AUTO_RESUME = TRUE' parameter, when specified during compute pool creation, enables the pool to automatically resume operation when a service or job is submitted, removing the need for explicit 'ALTER COMPUTE POOL RESUME' commands. Option D is incorrect. Setting = prevents the compute pool from automatically suspending, meaning it will continue to consume credits even when idle. This would generally lead to higher costs, not cost optimization, unless the pool is constantly active. The default is 3600 seconds (1 hour). SPCS Compute Nodes have a minimum charge of five minutes when started or resumed, making intelligent use of auto-suspend important for cost management. Option E is incorrect. Snowpark-optimized warehouses are a type of 'virtual warehouse' and are recommended for Snowpark workloads with large memory requirements or specific CPU architecture, typically for single-node ML training workloads 'within a warehouse'. SPCS compute pools, however, provide their own dedicated instance families (CPU, HighMemory, GPU) for containerized workloads, abstracting the underlying infrastructure and supporting distributed GPU clusters directly within SPCS, not Snowpark-optimized warehouses as a 'compute pool type' for SPCS.

NEW QUESTION # 248

A data scientist needs to generate vector embeddings for product descriptions stored in a column 'PRODUCT_DESCRIPTION' in the 'PRODUCT_CATALOG' table. They want to use the 'e5-base-v2' model for this task. Which of the following SQL statements correctly applies the 'SNOWFLAKE.CORTEX.EMBED_TEXT_768' function and accurately describes the expected data type of the resulting embedding?

- A. The query

```
SELECT SNOWFLAKE.CORTEX.EMBED_TEXT_768('e5-base-v2', product_description) FROM PRODUCT_CATALOG;
```
- B. The query

```
SELECT SNOWFLAKE.CORTEX.EMBED_TEXT_768(product_description, 'e5-base-v2') FROM PRODUCT_CATALOG;
```
- C. The query

```
SELECT SNOWFLAKE.CORTEX.EMBED_TEXT_768('e5-base-v2', product_description) FROM PRODUCT_CATALOG;
```
- D. The query

```
SELECT EMBED_TEXT_768(product_description, 'e5-base-v2', 'dimensions=768') FROM PRODUCT_CATALOG;
```
- E. The query

```
SELECT SNOWFLAKE.CORTEX.EMBED_TEXT_768('e5-base-v2', product_description) FROM PRODUCT_CATALOG;
```

Answer: C

Explanation:

Option C is correct. The 'SNOWFLAKE.CORTEX.EMBED_TEXT_768' function takes the model name as the first argument and the text to be embedded as the second argument. The 'e5-base-v2' model is a 768-dimension embedding model, and the function correctly returns a 'VECTOR(FLOAT, 768)' data type. Options A, B, D, and E incorrectly describe the function's arguments or the return data type.

NEW QUESTION # 249

An operations manager is tasked with monitoring the cost and ensuring compliance for a Cortex Analyst deployment that uses the REST API. They are particularly concerned with accurately tracking credit consumption and understanding the implications of enabling external models. Which of the following statements correctly describe aspects of Cortex Analyst cost and governance?

- A. Snowflake's view provides granular usage information for REST API requests to Cortex Analyst, including tokens processed per model.

- B. Cortex Analyst credit usage is based on the number of messages processed, at a rate of 67 Credits per 1,000 messages, and only successful responses (HTTP 200) are counted.
- C. The view can be queried to track detailed information about cortex Analyst requests, including generated SQL and any errors.
- D. Credit consumption for Cortex Analyst is primarily based on the number of tokens processed by the underlying LLMs, with more complex natural language questions leading to higher token usage and costs.
- E. Enabling Azure OpenAI models via the 'ENABLE_CORTEX_ANALYST_MODEL_AZURE_OPENAI' parameter ensures that all customer data and prompts remain within Snowflake's governance boundary and fully respect RBAC policies for those models.

Answer: B

Explanation:

Option E is correct. Cortex Analyst credit usage is based on the number of messages processed, specifically 67 Credits per 1,000 messages, and only successful responses (HTTP 200) are counted. The number of tokens in each message does not affect cost. Option D is incorrect because, as stated, Cortex Analyst charges per message, not per token, and the number of tokens in each message does not affect the cost. This is distinct from general Cortex LLM functions which often bill per token. Option B is incorrect. The 'CORTEX_DOCUMENT_PROCESSING_USAGE_HISTORY' view displays Document AI processing function activity, not Cortex Analyst activity. To monitor Cortex Analyst requests, administrators can query the event table, which logs details like user, question, generated SQL, and errors. Option C is incorrect. Snowflake strongly discourages the use of Azure OpenAI models through the legacy 'ENABLE_CORTEX_ANALYST_MODEL_AZURE_OPENAI' parameter. If enabled, only metadata and prompts (not customer data) are transmitted outside of Snowflake's governance boundary, and model-level RBAC is not available. Option D is incorrect because the documentation explicitly states that granular usage information cannot be obtained for requests made with the REST API using the 'CORTEX_FUNCTIONS_QUERY_USAGE_HISTORY' view.

NEW QUESTION # 250

A financial institution wants to develop a Snowflake-based pipeline to process call transcripts from their customer support. The pipeline needs to perform two main tasks: first, "summarize very lengthy technical support calls" (up to 20,000 tokens per transcript) into concise actionable insights, and second, "classify the sentiment" of these calls as 'positive', 'neutral', or 'negative'. Given these requirements for integration into SQL data pipelines, which combination of Snowflake Cortex functions and prompt engineering considerations would be most appropriate?

- For summarization, use `SNOWFLAKE.CORTEX.SUMMARIZE()` directly. For sentiment, use `SNOWFLAKE.CORTEX.SENTIMENT()` and interpret its numerical output to derive categories.
- For summarization, use `AI_COMPLETE()` with a large context window model like `mistral-large2` and a detailed prompt, encapsulating the prompt logic in a SQL UDF. For sentiment, use `AI_CLASSIFY()` with predefined categories.
- For summarization, use multiple calls to `SNOWFLAKE.CORTEX.EXTRACT_ANSWER()` to pull out key points, then combine them. For sentiment, use `AI_COMPLETE()` with a system prompt for sentiment classification.
- For both tasks, use `AI_COMPLETE()` with a small, cost-effective model like `gemma-7b`, and rely on precise prompts to guide both summarization and classification into JSON outputs.
- Utilize `AI_AGG()` for summarization across multiple call records, and `AI_FILTER()` for a boolean check on positive sentiment, then combine with other filters for neutral/negative.

- A. Option B
- B. Option D
- C. Option E
- D. Option A
- E. Option C

Answer: A

Explanation:

For summarizing very lengthy technical support calls (up to 20,000 tokens), a model with a sufficiently large context window is essential. (the updated version of offers flexibility for detailed summarization with prompt engineering. A model like 'mistral-large2' has a context window of 128,000 tokens, making it suitable for such long inputs. Encapsulating complex prompt logic within a SQL User Defined Function (UDF) is a recommended practice for better management and reusability in data pipelines. For classifying sentiment into predefined categories ('positive', 'neutral', 'negative'), (the updated version of is purpose-built and directly returns the classification label. A. is a generic summarization function, but 'AI_COMPLETE with a large model provides more control for 'actionable insights'. returns a numerical score, requiring additional logic for categorical output. C. 'SNOWFLAKE.CORTEX.EXTRACT ANSWER()' is designed to extract specific answers to questions, not to summarize text. Using it multiple times for summarization would be inefficient and less effective. While can perform classification, is the specialized function for this task. D. 'gemma-7b' has a context window of 8,000 tokens, which is insufficient for processing calls up to 20,000 tokens, potentially leading to truncation or incomplete results. E. and SUMMARIZE AGG()' are designed to aggregate insights or

summaries 'across multiple rows' or groups of text, not to summarize a single, lengthy document. returns a boolean result, making it less suitable for multi-category classification directly.

NEW QUESTION # 251

An enterprise is designing an advanced generative AI application in Snowflake, leveraging Cortex Agents to orchestrate data analysis from both structured and unstructured sources. According to Snowflake's Gen AI principles and the capabilities of Cortex Agents, which of the following statements accurately describe the workflow components and the types of tools an agent can utilize?

- A. For debugging, Cortex Agents allow direct modification of the LLM's internal state to refine accuracy, latency, and cost during execution.
- B. Cortex Agents are restricted to using only Snowflake's native Cortex LLM functions; custom logic via UDFs or stored procedures is not supported for tool implementation.
- C. The agent's workflow includes 'Planning' to orchestrate a solution, 'Explore options' for disambiguation, and 'Reflection' to determine next steps after tool use. Supported tools include Cortex Analyst and Cortex Search.
- D. Cortex Agents can orchestrate across both structured and unstructured data sources, and custom tools can be implemented using Snowflake stored procedures and user-defined functions (UDFs).
- E. Cortex Agents primarily focus on pre-defined, single-turn SQL queries for structured data, with limited support for unstructured data processing.

Answer: C,D

Explanation:

Cortex Agents orchestrate across both structured and unstructured data sources to deliver insights. Their workflow involves 'Planning' (parsing a request to orchestrate a plan), 'Explore options' (considering permutations for ambiguous questions), and 'Reflection' (evaluating results to determine next steps). Cortex Agents use Cortex Analyst (structured) and Cortex Search (unstructured) as tools, and also support custom tools implemented via stored procedures and user-defined functions (UDFs). Option A is incorrect as agents are designed for complex orchestration, not just simple, single-turn queries. Option C is incorrect because custom logic via UDFs and stored procedures is explicitly supported. Option D misrepresents debugging; AI Observability and tracing help debug individual records and refine apps for accuracy, latency, and cost, but do not involve direct modification of the LLM's internal state.

NEW QUESTION # 252

.....

Our Snowflake GES-C01 exam prep is renowned for free renewal in the whole year. As you have experienced various kinds of exams, you must have realized that renewal is invaluable to study materials, especially to such important SnowPro® Specialty: Gen AI Certification Exam GES-C01 Exams. And there is no doubt that being acquainted with the latest trend of exams will, to a considerable extent, act as a driving force for you to pass the GES-C01 exams and realize your dream of living a totally different life.

GES-C01 Valid Test Vce Free: <https://www.exam-killer.com/GES-C01-valid-questions.html>

So some of them want to get the GES-C01 Valid Test Vce Free GES-C01 Valid Test Vce Free - SnowPro® Specialty: Gen AI Certification Exam certification, but the real test is not easy to pass, thus much time and energy investment is inevitable. At the same time, you should be competent enough to deal with other exams not only the GES-C01 prep material, and we have the relevant GES-C01 vce practice as well, Snowflake New GES-C01 Dumps Questions In your real exam, you must answer all questions in limited time.

For selected images, Siegel includes a lighting diagram as well Valid GES-C01 Exam Format as a detailed explanation of how the image was made, but in every case he explains why a pose works, or why it doesn't.

Welcome to Exam-Killer, So some of them want to get the Snowflake Certification Test GES-C01 Free SnowPro® Specialty: Gen AI Certification Exam certification, but the real test is not easy to pass, thus much time and energy investment is inevitable.

Free PDF 2026 Snowflake GES-C01: SnowPro® Specialty: Gen AI Certification Exam Fantastic New Dumps Questions

At the same time, you should be competent enough to deal with other exams not only the GES-C01 prep material, and we have the relevant GES-C01 vce practice as well.

