

SnowPro Advanced: Data Scientist Certification Exam Interactive Testing Engine & DSA-C03 Latest Training Guide & SnowPro Advanced: Data Scientist Certification Exam Self-Paced Training



What's more, part of that Pass4training DSA-C03 dumps now are free: <https://drive.google.com/open?id=1DIPXWmv5MQwXROVzoR2ThEj2YozjzeAL>

In this fast-changing world, the requirements for jobs and talents are higher, and if people want to find a job with high salary they must boost varied skills which not only include the good health but also the working abilities. But if you get the DSA-C03 certification, your working abilities will be proved and you will find an ideal job. We provide you with DSA-C03 Exam Materials of high quality which can help you pass the exam easily. It also saves your much time and energy that you only need little time to learn and prepare for exam

Our latest DSA-C03 vce braindumps are written by our IT experts' wealth of knowledge and experience and can fully meet the demand of DSA-C03 real exam. From related websites or books, you might also see some Snowflake free download study materials, but our DSA-C03 Exam crams are affordable, latest and comprehensive.

>> **Training DSA-C03 Tools** <<

Reliable DSA-C03 Dumps Free | DSA-C03 Exam Dump

To be successful in a professional exam like the Snowflake DSA-C03 exam, you must know the criteria to pass it. You should know the type of SnowPro Advanced: Data Scientist Certification Exam questions, the pattern of the SnowPro Advanced: Data Scientist Certification Exam exam, and the time limit to complete the DSA-C03 Exam. All these factors help you pass the Snowflake DSA-C03 exam. Pass4training is your reliable partner in getting your DSA-C03 certification. The Snowflake DSA-C03 exam dumps help you achieve your professional goals.

Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample Questions (Q55-Q60):

NEW QUESTION # 55

You are tasked with creating a new feature in a machine learning model for predicting customer lifetime value. You have access to a table called 'CUSTOMER ORDERS' which contains order history for each customer. This table contains the following columns: 'CUSTOMER ID', 'ORDER DATE', and 'ORDER AMOUNT'. To improve model performance and reduce the impact of outliers, you plan to bin the 'ORDER AMOUNT' column using quantiles. You decide to create 5 bins, effectively creating quintiles. You also want to create a derived feature indicating if the customer's latest order amount falls in the top quintile. Which of the following approaches, or combination of approaches, is most appropriate and efficient for achieving this in Snowflake? (Choose all that apply)

- A. Use 'WIDTH_BUCKET' function, after finding the boundaries of quantile using 'APPROX_PERCENTILE' or 'PERCENTILE_CONT'. Using MAX(ORDER to determine recent amount is in top quintile.

- B. Create a temporary table storing quintile information, then join this table to original table to find the top quintile order amount.
- C. Use the window function to create quintiles for 'ORDER AMOUNT' and then, in a separate query, check if the latest 'ORDER AMOUNT' for each customer falls within the NTILE that represents the top quintile.
- D. Calculate the 20th, 40th, 60th, and 80th percentiles of the 'ORDER AMOUNT' using 'APPROX PERCENTILE' or 'PERCENTILE CONT' and then use a 'CASE' statement to assign each order to a quintile bin. Calculate and see if on that particular date is in top quintile.
- E. Use a Snowflake UDF (User-Defined Function) written in Python or Java to calculate the quantiles and assign each 'ORDER AMOUNT' to a bin. Later you can use other statement to check the top quintile amount from result set.

Answer: A,C,D

Explanation:

Options A, B, and E are valid and efficient approaches. Option A using 'NTILE' is a direct and efficient way to create quantile bins within Snowflake SQL, and can find the most recent order date for customer with a case statement. Option B calculates the percentiles directly and then uses a CASE statement to assign bins. This is also efficient for explicit boundaries. Option E finds the boundaries of the quantile using 'APPROX_PERCENTILE' or 'PERCENTILE_CONT', after that you can use 'WIDTH_BUCKET' to categorize into quantile bins based on ranges. Option C is possible but generally less efficient due to the overhead of UDF execution and data transfer between Snowflake and the UDF environment. Option D is valid, but creating a temporary table adds complexity and potentially reduces performance compared to window functions or direct quantile calculation within the query.

NEW QUESTION # 56

You are performing exploratory data analysis on a dataset of customer transactions in Snowflake to prepare for a linear regression model that predicts transaction value based on several customer-related features (e.g., age, location, number of previous transactions). You suspect a non-linear relationship between 'customer_age' and 'transaction_value'. Which of the following Snowflake SQL techniques is MOST appropriate for exploring and potentially transforming the 'customer_age' variable to better fit a linear regression model?

- A. Apply a logarithmic transformation to 'customer_age' if a scatter plot of 'customer_age' vs 'transaction_value' shows a curve that flattens out as 'customer_age' increases.
- B. Use the window function to bin 'customer_age' into quartiles and treat each quartile as a categorical variable in the linear regression model.
- C. Calculate the Pearson correlation coefficient between 'customer_age' and 'transaction_value' using the function. If the correlation is low, discard the 'customer_age' variable.
- D. Create polynomial features by adding 'customer_ageA2' and 'customer_ageA3' as new columns to the table, without checking for interaction effects.
- E. Implement a Box-Cox transformation in Snowpark Python, select a suitable transformation parameter based on the data, and apply the transformation on 'customer_age' feature.

Answer: A

Explanation:

Logarithmic transformation is a suitable method when the relationship flattens as the value increases. Creating polynomial features blindly without checking for interaction effects is generally not a good practice. Binning 'customer_age' into quartiles is also a potential solution, it discretizes the continuous data and might lose information, also it's only suitable after confirming it's the best option available. A low correlation does not necessarily mean the variable should be discarded; it could indicate a non-linear relationship that a linear model cannot capture directly. Box-Cox transformation is a good approach but may overcomplicate the task. Since Box-Cox transformations are generally harder than Log transformations.

NEW QUESTION # 57

You are building a product recommendation system using Snowflake Cortex. You have a table 'PRODUCT DESCRIPTIONS' containing product IDs and textual descriptions. You want to generate vector embeddings for these descriptions to perform similarity searches. However, you need to control the cost and latency of the embedding generation process. Which of the following strategies and considerations are MOST important for optimizing performance and cost when generating vector embeddings in Snowflake Cortex using a UDF?

- A. Cache the results of the embedding LJDF. Implement a caching mechanism (e.g., using a Snowflake table) to store the embeddings for frequently accessed product descriptions, avoiding redundant embedding calculations. use a materialized view.

- B. Use a larger Snowflake warehouse size. Increasing the warehouse size always linearly reduces embedding generation time and cost.
- C. Optimize the batch size passed to the embedding UDF. Experiment with different batch sizes to find the optimal trade-off between throughput and latency. Too large batches might cause memory issues, while too small batches increase overhead. Consider using a batch size of 64 or 128 as a starting point, adjusting based on your dataset and resource constraints.
- D. Use the smallest available Cortex embedding model. Smaller models are always faster and cheaper, regardless of the dataset size.
- E. Partition the 'PRODUCT DESCRIPTIONS' table by product category and generate embeddings for each partition separately. This helps to distribute the workload and reduce the size of the data processed by each UDF call. This makes more sense and is faster to re-create the table.

Answer: A,C,E

Explanation:

Optimizing batch size is crucial for throughput and latency (B). Caching embeddings avoids redundant computations (C), and partitioning data helps distribute the workload (D). Using the smallest model may sacrifice accuracy (A), and simply increasing warehouse size isn't always cost-effective (E).

NEW QUESTION # 58

You've trained a sales forecasting model using Snowpark ML and want to deploy it within Snowflake for real-time predictions. You've decided to store the predictions directly in a Snowflake table. The model predicts sales for different product categories based on historical data and promotional activities. Which of the following approaches is the MOST efficient and scalable way to store these predictions, considering a high volume of prediction requests and the need for quick retrieval for downstream dashboards?

- A. Storing predictions in a separate table with a composite key of product category and timestamp, with clustering on the timestamp column and partitioning by product category.
- B. Storing predictions in a single, wide table with all features and predictions as columns. No partitioning or clustering is implemented.
- C. Storing predictions in an external stage (e.g., AWS S3) and querying them using an external table. The external table definition includes the sales prediction as a column.
- D. Storing predictions in a VARIANT column in a single table. All prediction results for a given product category are stored as a JSON document within the VARIANT column.
- E. Storing predictions in a key-value store like Redis and referencing the keys from a Snowflake table. Requires external network access from Snowflake.

Answer: A

Explanation:

Option B is the most efficient and scalable approach. Partitioning by product category allows for faster querying of specific categories. Clustering on the timestamp column ensures that recent predictions are quickly accessible. A composite key of product category and timestamp provides uniqueness. Option A lacks any optimization for querying. Option C can lead to performance issues with large JSON documents and querying specific values within the VARIANT. Option D introduces latency due to external stage access, and external tables are generally slower for frequent queries compared to native Snowflake tables. Option E introduces external dependency and network latency, which is generally not preferred if a native Snowflake solution is possible.

NEW QUESTION # 59

You're building a fraud detection model and want to determine if the average transaction amount for fraudulent transactions is significantly higher than the average transaction amount for legitimate transactions. You have two tables in Snowflake: 'FRAUDULENT TRANSACTIONS' and 'LEGITIMATE TRANSACTIONS', both with a 'TRANSACTION AMOUNT' column. You believe that FRAUDULENT TRANSACTIONS contains fewer than 30 transactions. You don't know the population standard deviations. What are the proper steps to conduct the hypothesis test, and what is the correct hypothesis statement?

- A. Perform a Z-test. Null Hypothesis: The average transaction amount for fraudulent transactions is equal to the average transaction amount for legitimate transactions. Alternative Hypothesis: The average transaction amount for fraudulent transactions is not equal to the average transaction amount for legitimate transactions.
- B. Perform a Z-test. Null Hypothesis: The average transaction amount for fraudulent transactions is less than or equal to the average transaction amount for legitimate transactions. Alternative Hypothesis: The average transaction amount for fraudulent transactions is greater than the average transaction amount for legitimate transactions.

- C. Perform a t-test. Null Hypothesis: The average transaction amount for fraudulent transactions is equal to the average transaction amount for legitimate transactions. Alternative Hypothesis: The average transaction amount for fraudulent transactions is not equal to the average transaction amount for legitimate transactions.
- D. Perform a chi-squared test. Null Hypothesis: There is no relationship between transaction amount and whether a transaction is fraudulent. Alternative Hypothesis: There is a relationship between transaction amount and whether a transaction is fraudulent.
- E. Perform a t-test. Null Hypothesis: The average transaction amount for fraudulent transactions is less than or equal to the average transaction amount for legitimate transactions. Alternative Hypothesis: The average transaction amount for fraudulent transactions is greater than the average transaction amount for legitimate transactions.

Answer: E

Explanation:

The correct answer is C. Since the sample size for fraudulent transactions is less than 30, and the population standard deviations are unknown, a t-test is more appropriate than a Z-test. The null hypothesis should state the assumption that we are trying to disprove (i.e., fraudulent transactions are not, on average, higher). The alternative hypothesis is the claim we are trying to support (i.e., fraudulent transactions ARE, on average, higher). The chi-squared test is used for categorical data, not continuous data like transaction amount. We are interested in knowing if one set of transaction amounts is greater, so its a one tailed t test.

NEW QUESTION # 60

.....

Pass4training provides Snowflake DSA-C03 desktop-based practice software for you to test your knowledge and abilities. The DSA-C03 desktop-based practice software has an easy-to-use interface. You will become accustomed to and familiar with the free demo for Snowflake DSA-C03 Exam Questions. Exam self-evaluation techniques in our DSA-C03 desktop-based software include randomized questions and timed tests. These tools assist you in assessing your ability and identifying areas for improvement to pass the Snowflake certification exam.

Reliable DSA-C03 Dumps Free: <https://www.pass4training.com/DSA-C03-pass-exam-training.html>

According to the feedbacks of previous customers who bought our DSA-C03 updated pdf, the passing rate of our DSA-C03 valid questions reaches up to 98%, even to 100%, so please be assured the purchase, Snowflake Training DSA-C03 Tools Dreams of imaginary make people feel disheartened, Our DSA-C03 test prep embrace latest information, up-to-date knowledge and fresh ideas, encouraging the practice of thinking out of box rather than treading the same old path following a beaten track, Snowflake Training DSA-C03 Tools Now that the network is so developed, we can disclose our information at any time.

This event occurs only when all the objects on the page DSA-C03 have been created and are available for use, Tricky wording, to be sure, According to the feedbacks of previous customers who bought our DSA-C03 updated pdf, the passing rate of our DSA-C03 valid questions reaches up to 98%, even to 100%, so please be assured the purchase.

2026 Snowflake DSA-C03: Professional Training SnowPro Advanced: Data Scientist Certification Exam Tools

Dreams of imaginary make people feel disheartened, Our DSA-C03 test prep embrace latest information, up-to-date knowledge and fresh ideas, encouraging the practice of Reliable DSA-C03 Dumps Free thinking out of box rather than treading the same old path following a beaten track.

Now that the network is so developed, we can disclose our information at any time, Tens of thousands of our loyal customers relayed on our DSA-C03 preparation materials and achieved their dreams.

- DSA-C03 Online Lab Simulation DSA-C03 Exam Question DSA-C03 Online Lab Simulation Search for **>** DSA-C03 and download it for free on { www.examcollectionpass.com } website DSA-C03 Test Simulator Free
- 2026 Accurate Training DSA-C03 Tools | 100% Free Reliable DSA-C03 Dumps Free Simply search for **✓** DSA-C03 for free download on 《 www.pdfvce.com 》 New DSA-C03 Exam Price
- Marvelous Training DSA-C03 Tools for Real Exam Search for DSA-C03 and download it for free on www.pdf.dumps.com website Reliable DSA-C03 Test Duration
- Trustworthy DSA-C03 Dumps Trustworthy DSA-C03 Dumps Examcollection DSA-C03 Dumps Search for **✓** DSA-C03 and easily obtain a free download on **➔** www.pdfvce.com DSA-C03 Exam Question
- Trustworthy DSA-C03 Dumps DSA-C03 Reliable Exam Pass4sure DSA-C03 Valid Test Materials Search for 《 DSA-C03 》 and obtain a free download on **➔** www.prepawayexam.com DSA-C03 Exam Papers

