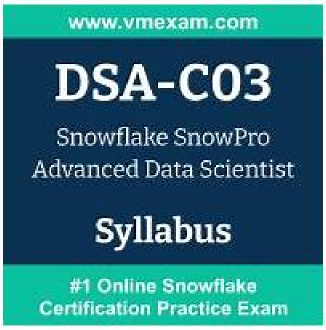
New Snowflake DSA-C03 Test Topics - DSA-C03 Best Preparation Materials



BONUS!!! Download part of Exam4PDF DSA-C03 dumps for free: https://drive.google.com/open?id=1d-rzOHM47qe_OJQl08c_vfTywIMVY-vi

If you want to get a desirable opposition and then achieve your career dream, you are a right place now. Our DSA-C03 study tool can help you pass the exam. So, don't be hesitate, choose the DSA-C03 test torrent and believe in us. Let's strive to our dreams together. Life is short for us, so we all should cherish our life. Our DSA-C03 Guide Torrent can help you to save your valuable time and let you have enough time to do other things you want to do. Just buy our DSA-C03 exam questions, then you will pass the DSA-C03 exam easily.

Our DSA-C03 PDF format is user-friendly and accessible on any smart device, allowing applicants to study from anywhere at any time. We have included actual and updated Snowflake DSA-C03 Questions in this DSA-C03 Dumps PDF file. Our SnowPro Advanced: Data Scientist Certification Exam exam dumps PDF format is designed to help individuals acquire the knowledge necessary to succeed in the test.

>> New Snowflake DSA-C03 Test Topics <<

DSA-C03 Best Preparation Materials - Official DSA-C03 Study Guide

By resorting to our DSA-C03 practice materials, we can absolutely reap more than you have imagined before. We have clear data collected from customers who chose our DSA-C03 actual tests, the passing rate is 98-100 percent. So your chance of getting success will be increased greatly by our DSA-C03 braindump materials. Moreover, there are a series of benefits for you. So the importance of DSA-C03 actual test is needless to say. If you place your order right now, we will send you the free renewals lasting for one year.

Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample Questions (Q52-Q57):

NEW QUESTION # 52
You've deployed a fraud detection model in Snowflake using Snowpark. You are monitoring its performance and notice a significant

decrease in recall, while precision remains high. This means the model is missing many fraudulent transactions. The training data was initially balanced, but you suspect that recent changes in user behavior have skewed the distribution of fraudulent vs. non-fraudulent transactions in production. Which of the following actions are MOST appropriate to address this issue and improve the model's performance, considering best practices for model retraining within the Snowflake ecosystem?

- A. Implement a data drift monitoring system in Snowflake to automatically detect changes in the input features of the model. Trigger an automated retraining pipeline when significant drift is detected. This retraining should include recent production data with updated labels, but only if label data collection can be automated.
- B. Retrain the model using the original training data. Since the precision is high, the model's fundamental logic is still sound. A larger training dataset isn't necessary.
- C. Immediately shut down the model to prevent further inaccurate classifications. Investigate why the recall is low before any retraining is performed.
- D. Retrain the model using a dataset that includes recent production data, being sure to re-balance the dataset to maintain a roughly equal number of fraudulent and non-fraudulent transactions. Prioritize transactions from the last month.
- E. Adjust the model's classification threshold to be more sensitive, even if it means accepting a slightly lower precision. This can be done directly within Snowflake using a SQL UDF that transforms the model's output probabilities.

Answer: A,D,E

Explanation:
Options B, C, and D are the most appropriate. B addresses the data drift by incorporating recent production data with re-balancing to mitigate the skewed distribution. C directly improves recall by adjusting the classification threshold. D establishes a proactive drift detection and retraining system which is a best practice for long-term model maintenance. A is incorrect because the original data doesn't reflect current trends. E is too drastic initially; adjusting the threshold and retraining are preferred first. Retraining with balanced, recent data is critical, especially if the class distribution has shifted. Monitoring for drift provides an automated approach to maintaining model accuracy in a changing environment. Also a low code retraining pipeline is appropriate considering current model performance with SQL udf transformations.

NEW QUESTION # 53
A data scientist uses bootstrapping to estimate the sampling distribution of a statistic calculated from a dataset stored in Snowflake. They observe that the bootstrap distribution is significantly different from the original data distribution. Which of the following statements best describes the possible reasons for this difference, considering both the theoretical underpinnings of bootstrapping and potential limitations?

- A. Bootstrapping is only appropriate for normally distributed data; if the original data is not normal, the bootstrap distribution will inevitably differ significantly.
- B. The original sample may not be representative of the population, and the bootstrap procedure is simply amplifying the biases present in the original sample. Additionally, the statistic itself may be highly sensitive to outliers or specific data points, leading to a distorted bootstrap distribution.
- C. Bootstrapping always provides accurate estimates of sampling distributions, any significant difference indicates an error in the code implementation.
- D. The difference is unexpected; the bootstrap distribution should always closely resemble the original data distribution, regardless of the statistic being estimated.
- E. The statistic being estimated is inherently unstable and has a high variance, causing the bootstrap distribution to be wider and potentially different in shape compared to the original data distribution. This is a normal outcome when dealing with such statistics.

Answer: B,E

Explanation:
Options B and C are correct. Bootstrapping relies on the assumption that the original sample is representative of the population. If it isn't, the bootstrap distribution will reflect the biases of the sample. Also certain statistics, particularly those sensitive to outliers or with high variance, can produce bootstrap distributions that differ significantly from the original data distribution. Option A is incorrect because the bootstrap distribution doesn't necessarily have to be same as sample distribution. Option D is incorrect since Bootstrapping makes no assumptions regarding the distribution of original dataset and can be used for any data distribution. Option E is not correct. Bootstrapping is not always accurate and relies on assumptions to perform correctly.

NEW QUESTION # 54
You are working with a Snowflake table named 'CUSTOMER DATA' containing customer information, including a 'PHONE

NUMBER' column. Due to data entry errors, some phone numbers are stored as NULL, while others are present but in various inconsistent formats (e.g., with or without hyphens, parentheses, or country codes). You want to standardize the 'PHONE NUMBER column and replace missing values using Snowpark for Python. You have already created a Snowpark DataFrame called 'customer df representing the 'CUSTOMER DATA' table. Which of the following approaches, used in combination, would be MOST efficient and reliable for both cleaning the existing data and handling future data ingestion, given the need for scalability?

- A. Create a Snowflake Pipe with a COPY INTO statement and a transformation that uses a SQL function within the COPY INTO statement to format the phone numbers and replace NULL values during data loading. Also, implement a Python UDF for correcting already existing data.
- B. Use a series of and methods on the Snowpark DataFrame to handle NULL values and different phone number formats directly within the DataFrame operations.
- C. Create a Snowflake Stored Procedure in SQL that uses regular expressions and 'CASE statements to format the "PHONE_NUMBER column and replace NULL values. Call this stored procedure from a Snowpark Python script.
- D. Leverage Snowflake's data masking policies to mask any invalid phone number and create a view that replaces NULL values with 'UNKNOWN'. This approach doesn't correct existing data but hides the issue.
- E. Use a UDF (User-Defined Function) written in Python that formats the phone numbers based on a regular expression and applies it to the DataFrame using For NULL values, replace them with a default value of 'UNKNOWN'.

Answer: A,E

Explanation:
Options A and E provide the most robust and scalable solutions. A UDF offers flexibility and reusability for data cleaning within Snowpark (Option A). Option E leverages Snowflake's data loading capabilities to clean data during ingestion and adds a UDF for cleaning existing data providing a comprehensive approach. Using a UDF written in Python and used within Snowpark leverages the power of Python's regular expression capabilities and the distributed processing of Snowpark. Handling data transformations during ingestion with Snowflake's built- in COPY INTO with transformation is highly efficient. Option B is less scalable and maintainable for complex formatting. Option C is viable but executing SQL stored procedures from Snowpark Python loses some of the advantages of Snowpark. Option D addresses data masking not data transformation.

NEW QUESTION # 55
You are using Snowpark for Python to build a feature engineering pipeline for a machine learning model that predicts customer churn. The data is stored in a Snowflake table called 'CUSTOMER DATA' , and you want to create new features based on time-series data within the table. You need to calculate the 'Recency' feature (days since the last transaction) and 'Frequency' feature (number of transactions in the last 3 months). Considering performance and best practices, which Snowpark approach would you choose?

- A. Create a Python UDF using Pandas to calculate 'Recency' and 'Frequency'. Apply this UDF to the 'CUSTOMER DATA' table through Snowpark, processing the data row by row.
- B. Write a stored procedure in SQL that calculates 'Recency' and 'Frequency' using SQL window functions, and then call this stored procedure from your Snowpark Python code.
- C. Fetch the entire 'CUSTOMER DATA table into a Pandas DataFrame using , then use Pandas' time-series functions to calculate 'Recency' and 'Frequency'. After feature engineering, load the Pandas DataFrame back into Snowflake.
- D. Use Snowpark DataFrame API to perform window functions within Snowflake to calculate 'Recency' and 'Frequency' directly, leveraging Snowflake's processing power without transferring data to the client.
- E. Write custom Python code in a Snowpark UDF to retrieve each transaction for a customer and calculate recency and frequency directly in Python without pandas.

Answer: D

Explanation:
Option B is the most efficient approach. Snowpark DataFrame API and window functions allow you to perform the necessary calculations directly within Snowflake's engine, minimizing data transfer and maximizing performance. This leverages the parallel processing capabilities of Snowflake. Option A: Transferring the entire dataset to Pandas can be very slow and resource-intensive, especially for large datasets. Option C: Python UDFs using Pandas might be useful for complex calculations not easily done in SQL, but for window function based feature engineering, using native Snowflake capabilities is better in performance. Additionally, row-by-row processing with UDFs is generally less efficient than set-based operations in SQL. Option D: Calling a stored procedure is a valid approach, but using Snowpark DataFrame API directly often provides better integration with the Python code and can be easier to manage. Option E: Creating the UDF in Python for each transaction is the worst approach and will be very slow.

NEW QUESTION # 56
You are working with a large dataset of sensor readings stored in a Snowflake table. You need to perform several complex feature engineering steps, including calculating rolling statistics (e.g., moving average) over a time window for each sensor. You want to use Snowpark Pandas for this task. However, the dataset is too large to fit into the memory of a single Snowpark Pandas worker. How can you efficiently perform the rolling statistics calculation without exceeding memory limits? Select all options that apply.

- A. Increase the memory allocation for the Snowpark Pandas worker nodes to accommodate the entire dataset.
- B. Break the Snowpark DataFrame into smaller chunks using 'sample' and 'unionAll', process each chunk with Snowpark Pandas, and then combine the results.
- C. Use the 'grouped' method in Snowpark DataFrame to group the data by sensor ID, then download each group as a Pandas DataFrame to the client and perform the rolling statistics calculation locally. Then upload back to Snowflake.
- D. Explore using Snowpark's Pandas user-defined functions (UDFs) with vectorization to apply custom rolling statistics logic directly within Snowflake. UDFs allow you to use Pandas within Snowflake without needing to bring the entire dataset client-side.
- E. Utilize the 'window' function in Snowpark SQL to define a window specification for each sensor and calculate the rolling statistics using SQL aggregate functions within Snowflake. Leverage Snowpark to consume the results of the SQL transformation.

Answer: D,E

Explanation:
Explanation:Options B and D are the most appropriate and efficient solutions for handling large datasets when calculating rolling statistics with Snowpark Pandas. Option B uses the 'window' function in Snowpark SQL. Leverage the 'window' function in Snowpark SQL to define a window specification for each sensor and calculate the rolling statistics using SQL aggregate functions within Snowflake. Option D uses Snowpark's Pandas UDFs. Snowpark's Pandas UDFs with vectorization allow you to bring the processing logic to the data within Snowflake, avoiding the need to move the entire dataset to the client-side and bypassing memory limitations. This approach is generally more scalable and performant for large datasets. Option A is inefficient as it retrieves groups of data from Snowflake to client side before creating the calculations before sending back to snowflake. Option C is correct but complex and not optimal. Option E is possible, but it's not a scalable solution and can be costly.

NEW QUESTION # 57
......

download it for free immediately on www.pdfvce.com DSA-C03 Flexible Learning Mode

- Key DSA-C03 Concepts DSA-C03 Instant Access DSA-C03 Reliable Real Test Copy URL （ www.pass4leader.com ） open and search for [DSA-C03] to download for free DSA-C03 Reliable Study Questions
- Reliable DSA-C03 Study Materials Sample DSA-C03 Questions Pdf New DSA-C03 Test Discount Go to website ▹ www.pdfvce.com ◃ open and search for 【 DSA-C03 】 to download for free Valid DSA-C03 Exam Question
- DSA-C03 Exam Answers Trusted DSA-C03 Exam Resource Key DSA-C03 Concepts Easily obtain DSA-C03 for free download through ➡ www.prep4away.com Practice Test DSA-C03 Fee
- Three formats of Snowflake DSA-C03 practice exams meet the diverse needs Easily obtain free download of ➡ DSA-C03 by searching on { www.pdfvce.com } Test DSA-C03 Question
- Snowflake DSA-C03 exam study materials Search for ➡ DSA-C03 and download it for free on 「 www.getvalidtest.com 」 website DSA-C03 Flexible Learning Mode
- Reliable DSA-C03 Study Materials Test DSA-C03 Question DSA-C03 Certification Materials Search on www.pdfvce.com for 《 DSA-C03 》 to obtain exam materials for free download Valid DSA-C03 Exam Question
- Valid DSA-C03 Exam Question Test DSA-C03 Pass4sure Test DSA-C03 Pass4sure 《 www.dumpsquestion.com 》 is best website to obtain ☀ DSA-C03 ☀ for free download DSA-C03 Reliable Real Test
- Free PDF Quiz Snowflake - Latest DSA-C03 - New SnowPro Advanced: Data Scientist Certification Exam Test Topics   Enter ➠ www.pdfvce.com and search for ➥ DSA-C03 to download for free Valid DSA-C03 Test Cost
- Splendid DSA-C03 Exam Braindumps are from High-quality Learning Quiz - www.examdiscuss.com Search for ➥ DSA-C03 and obtain a free download on ➽ www.examdiscuss.com DSA-C03 Reliable Real Test
- pct.edu.pk, sincerequranicinstitute.com, www.posteezy.com, shortcourses.russellcollege.edu.au, www.stes.tyc.edu.tw, myportal.utt.edu.tt, edu.ais.ind.in, www.stes.tyc.edu.tw, edumente.me, Disposable vapes

DOWNLOAD the newest Exam4PDF DSA-C03 PDF dumps from Cloud Storage for free: https://drive.google.com/open?id=1d-rzOHM47qe_OJQl08c_vfTywIMVY-vi