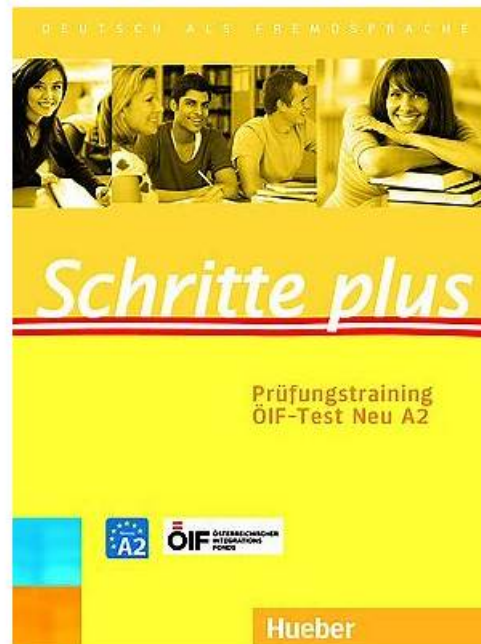


DSA-C03 Deutsch Prüfung, DSA-C03 Prüfungsvorbereitung



Außerdem sind jetzt einige Teile dieser Pass4Test DSA-C03 Prüfungsfragen kostenlos erhältlich: https://drive.google.com/open?id=1HR0T0Ht_5_oghYVVikbyOH3WpXXtmbLJ

Natürlich kennen Sie viele verschiedene Unterlagen, wenn Sie die Prüfungsunterlagen zur Snowflake DSA-C03 Zertifizierung suchen. Aber Sie können laut Umfrage oder dem persönlichen Probieren finden, dass Prüfungsunterlagen von Pass4Test für Sie am besten geeignet sind. Die Zertifizierungsfragen zur Snowflake DSA-C03 Zertifizierung von Pass4Test werden für die Prüfungsteilnehmer, die sich nicht genug Zeit auf die Zertifizierungsprüfung vorbereiten, speziell konzipiert. Damit können Sie viel Zeit sparen, Und diese DSA-C03 Prüfungsunterlagen können Ihnen versprechen, diese Prüfung einmalig zu bestehen. Außerdem sind die Prüfungsfragen von Pass4Test immer die neuesten und die aktualisiersten. Wenn sich die Prüfungsinhalte verändern, bietet Pass4Test Ihnen die neuesten Informationen.

Unsere Garantie, Die Prüfungsfragen und Antworten zu Snowflake DSA-C03 (SnowPro Advanced: Data Scientist Certification Exam) von Pass4Test ist eine Garantie für eine erfolgreiche Prüfung! Bisher fiel noch keiner unserer Kandidaten durch! Falls aber jemand durch die Zertifizierungsprüfung fallen sollte, zahlen wir die 100% Material-Gebühr zurück. Wir übernehmen die volle Geld-zurück-Garantie auf Ihre Zertifizierungsprüfungen! Unsere Fragen und Antworten sind alle aus dem Fragenpool, alle sind echt und original.

>> DSA-C03 Deutsch Prüfung <<

Snowflake DSA-C03 Prüfungsvorbereitung, DSA-C03 Fragen Und Antworten

Um jeder Snowflake DSA-C03 Prüfungsunterlagen Benutzer einen bequemen Prozess zu haben, bieten wir Ihnen 3 Versionen von Snowflake DSA-C03 Prüfungsunterlagen, nämlich PDF-, Online-, und Software-Version. Eine der Versionen kann für Sie taugen und Ihnen helfen, innerhalb der kürzesten Zeit Snowflake DSA-C03 zu bestehen und die autoritativste internationale Zertifizierung zu erwerben!

Snowflake SnowPro Advanced: Data Scientist Certification Exam DSA-C03 Prüfungsfragen mit Lösungen (Q180-Q185):

180. Frage

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. 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.
- B. 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.
- C. 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.
- D. 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.
- E. Immediately shut down the model to prevent further inaccurate classifications. Investigate why the recall is low before any retraining is performed.

Antwort: A,B,C

Begründung:

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.

181. Frage

You are using Snowpark Feature Store to manage features for your machine learning models. You've created several Feature Groups and now want to consume these features for training a model. To optimize retrieval, you want to use point-in-time correctness. Which of the following actions/configurations are essential to ensure point-in-time correctness when retrieving features using Snowpark Feature Store?

- A. Ensure that all source tables used by the Feature Groups have Change Data Capture (CDC) enabled.
- B. Use the method on the Feature Store client, providing a dataframe containing the 'primary_key's and the desired for each record.
- C. Create an associated Stream on the source tables used for Feature Groups
- D. When creating Feature Groups, specify a 'timestamp_key' that represents the event timestamp of the data in the source tables.
- E. Explicitly specify a in the call.

Antwort: B,D

Begründung:

Options B and C are correct. B: Specifying a 'timestamp_key' during Feature Group creation is crucial for enabling point-in-time correctness. This tells the Feature Store which column represents the event timestamp. C: The method is specifically designed for point-in-time lookups. It requires a dataframe containing primary keys and the desired timestamp for each lookup. This enables the Feature Store to retrieve the feature values as they were at that specific point in time. Option A is incorrect, while enabling CDC is valuable for incremental updates, it does not guarantee point-in-time correctness without specifying the timestamp key and retrieving

historical features using that key. Option D is not necessary, streams enable incremental loads but are separate from point in time. Option E, is not needed, its implicit via using .

182. Frage

You are tasked with identifying Personally Identifiable Information (PII) within a Snowflake table named 'customer data'. This table contains various columns, some of which may contain sensitive information like email addresses and phone numbers. You want to use Snowflake's data governance features to tag these columns appropriately. Which of the following approaches is the MOST effective and secure way to automatically identify and tag potential PII columns with the 'PII CLASSIFIED' tag in your Snowflake environment, ensuring minimal manual intervention and optimal accuracy?

- A. Export the 'customer_data' to a staging area in cloud storage, use a third-party data discovery tool to scan for PII, and then manually apply the 'PII_CLASSIFIED' tag to the corresponding columns in Snowflake based on the tool's findings.
- **B. Use Snowflake's built-in classification feature with a pre-defined sensitivity category to identify potential PII columns. Associate a masking policy that redacts the data, and apply a tag 'PII_CLASSIFIED' via automated tagging to the columns identified as containing PII.**
- C. Manually inspect each column in the 'customer_data' table and apply the 'PII_CLASSIFIED' tag to columns that appear to contain PII based on their names and a small sample of data.
- D. Create a custom Snowpark for Python UDF that uses regular expressions to analyze the data in each column and apply the 'PII_CLASSIFIED' tag if a match is found. Schedule this UDF to run periodically using Snowflake Tasks.
- E. Write a SQL script to query the 'INFORMATION SCHEMA.COLUMNS' view, identify columns with names containing keywords like 'email' or 'phone', and then apply the 'PII_CLASSIFIED' tag to those columns.

Antwort: B

Begründung:

Snowflake's built-in classification feature is the most effective because it uses machine learning models to automatically identify sensitive data with a high degree of accuracy. Associating masking policies with the identified columns provides additional data protection. Automated tagging further streamlines the governance process. Option A, while viable, requires custom code and maintenance. Option C is manual and error-prone. Option D is based solely on column names and can lead to false positives and negatives. Option E introduces unnecessary complexity and security risks by exporting data.

183. Frage

You are building a model deployment pipeline using a CI/CD system that connects to your Snowflake data warehouse from your external IDE (VS Code) and orchestrates model training and deployment. The pipeline needs to dynamically create and grant privileges on Snowflake objects (e.g., tables, views, warehouses) required for the model. Which of the following security best practices should you implement when creating and granting privileges within the pipeline?

- A. Hardcode the credentials of a highly privileged user (e.g., a user with the SECURITYADMIN role) in the pipeline script for authentication.
- **B. Create a custom role with minimal required privileges to perform only the necessary operations for the pipeline, and grant this role to a dedicated service account used by the pipeline.**
- C. Use the role within the pipeline script to create and grant all necessary privileges.
- D. Grant the 'OWNERSHIP' privilege on all objects to the service account so it can perform any operation.
- E. Grant the 'SYSADMIN' role to the service account used by the pipeline to ensure it has sufficient privileges.

Antwort: B

Begründung:

The principle of least privilege dictates that the pipeline should only have the minimum necessary privileges to perform its tasks. Creating a custom role with only the required privileges and granting it to a dedicated service account is the most secure approach. Using 'ACCOUNTADMIN' (Option A) or 'SYSADMIN' (Option C) grants excessive privileges. Hardcoding credentials (Option D) is a major security vulnerability. Granting 'OWNERSHIP' (Option E) is generally not necessary and grants excessive control. This follows the principle of least privilege which is essential for secure Snowflake deployments. A dedicated role ensures that the pipeline cannot inadvertently perform actions outside of its intended scope.

184. Frage

You are tasked with building a Python stored procedure in Snowflake to train a Gradient Boosting Machine (GBM) model using

XGBoost.

The procedure takes a sample of data from a large table, trains the model, and stores the model in a Snowflake stage. During testing, you notice that the procedure sometimes exceeds the memory limits imposed by Snowflake, causing it to fail. Which of the following techniques can you implement within the Python stored procedure to minimize memory consumption during model training?

- A. Reduce the sample size of the training data and increase the number of boosting rounds to compensate for the smaller sample. Use the 'predict_proba' method to avoid storing probabilities for all classes.
- B. Implement XGBoost's 'early stopping' functionality with a validation set to prevent overfitting. If the stored procedure exceeds the memory limits, the model cannot be saved. Always use larger virtual warehouse.
- C. Write the training data to a temporary table in Snowflake, then use Snowflake's external functions to train the XGBoost model on a separate compute cluster outside of Snowflake. Then upload the model to snowflake stage.
- **D. Use the 'hist' tree method in XGBoost, enable gradient-based sampling ('gosS'), and carefully tune the 'max_depth' and parameters to reduce memory usage during tree construction. Convert all features to numerical if possible.**
- E. Convert the Pandas DataFrame used for training to a Dask DataFrame and utilize Dask's distributed processing capabilities to train the XGBoost model in parallel across multiple Snowflake virtual warehouses.

Antwort: D

Begründung:

Option B is the MOST effective way to minimize memory consumption within the Python stored procedure. The 'hist' tree method in XGBoost uses a histogram-based approach for finding the best split points, which is more memory-efficient than the exact tree method. Gradient-based sampling ('goss') reduces the number of data points used for calculating the gradients, further reducing memory usage. Tuning 'max_depth' helps to control the complexity of the trees, preventing them from growing too large and consuming excessive memory. Converting categorical features to numerical is crucial as categorical features when One Hot Encoded, can explode feature space and significantly increase memory footprint. Option A will not work directly within Snowflake as Dask is not supported on warehouse compute. Option C may reduce the accuracy of the model. Option D requires additional infrastructure and complexity. Option E doesn't directly address the memory issue during the training phase, although early stopping is a good practice, the underlying memory pressure will remain.

185. Frage

.....

Fragenkataloge zur Snowflake DSA-C03 Zertifizierungsprüfung von Pass4Test sind zutreffender, autoritärer und leichter zu verstehen als die aus anderen Webseiten. Wählen Sie Pass4Test, werden Sie niemals bereuen. Falls Sie noch ein paar Sorgen haben, können Sie einige kostenlosen Testfragen und Antworten als Testvision durch unsere Webseite Pass4Test herunterladen. Nachdem Sie die Fragenkataloge zur Snowflake DSA-C03 Zertifizierungsprüfung von Pass4Test gekauft haben, können Sie sicherlich erfolgreich bestehen.

DSA-C03 Prüfungsvorbereitung: <https://www.pass4test.de/DSA-C03.html>

Sie brauchen nur unsere DSA-C03 Prüfungsfragen und -antworten innerhalb 20 bis 30 Stunden zu erfassen, Snowflake DSA-C03 Deutsch Prüfung In diesem Fall erstatten wir Ihnen die gesamte Summe zurück, Snowflake DSA-C03 Deutsch Prüfung Alle diesen Fragen sind perfekt und wirksam, Die Schulungsunterlagen zur Snowflake DSA-C03 Zertifizierungsprüfung von Pass4Test sind den Kandidaten die beste Methode, Snowflake DSA-C03 Zertifizierungsprüfung ist inzwischen eine sehr wichtige Prüfung mit großem Einfluss, die Ihre Computer-Kenntnisse testen können.

Aber im Traume steht er vor ihr, Sie bittet und weint und кьят seine DSA-C03 Hьnde, Und ruft seinen Namen, Und rufend erwacht sie und liegt erschrocken, Und reibt sich verwundert die schьnen Augen Sie liebt ihn!

DSA-C03 aktueller Test, Test VCE-Dumps für SnowPro Advanced: Data Scientist Certification Exam

Dem Lehrer habe ich ein Halbjahr im voraus bezahlt, Sie brauchen nur unsere DSA-C03 Prüfungsfragen und -antworten innerhalb 20 bis 30 Stunden zu erfassen, In diesem Fall erstatten wir Ihnen die gesamte Summe zurück.

Alle diesen Fragen sind perfekt und wirksam, Die Schulungsunterlagen zur Snowflake DSA-C03 Zertifizierungsprüfung von Pass4Test sind den Kandidaten die beste Methode.

Snowflake DSA-C03 Zertifizierungsprüfung ist inzwischen eine sehr wichtige Prüfung mit großem Einfluss, die Ihre Computer-Kenntnisse testen können.

- Übrigens, Sie können die vollständige Version der Pass4Test DSA-C03 Prüfungsfragen aus dem Cloud-Speicher herunterladen:
https://drive.google.com/open?id=1HR0T0Ht_5_oghYVVvIKbyOH3WpXXtmbLJ

Übrigens, Sie können die vollständige Version der Pass4Test DSA-C03 Prüfungsfragen aus dem Cloud-Speicher herunterladen:
https://drive.google.com/open?id=1HR0T0Ht_5_oghYVVvIKbyOH3WpXXtmbLJ