

Exam DEA-C02 Bible | Perfect SnowPro Advanced: Data Engineer (DEA-C02) 100% Free Exam Material



P.S. Free 2026 Snowflake DEA-C02 dumps are available on Google Drive shared by Pass4Leader: <https://drive.google.com/open?id=1DOfhJO5veNJW3EUTqGAd9AM5DAbdOG09>

Pass4Leader's DEA-C02 exam certification training materials include DEA-C02 exam dumps and answers. The data is worked out by our experienced team and IT professionals through their own exploration and continuous practice, and its authority is unquestioned. You can download DEA-C02 free demo and answers on probation on Pass4Leader website. After you purchase DEA-C02 exam certification training information, we will provide one year free renewal service.

These SnowPro Advanced: Data Engineer (DEA-C02) (DEA-C02) exam questions are a one-time investment to clear the DEA-C02 test in a short time. These DEA-C02 exam questions eliminate the need for candidates to study extra or irrelevant content, allowing them to complete their Snowflake test preparation quickly. By avoiding unnecessary information, you can save time and crack the SnowPro Advanced: Data Engineer (DEA-C02) (DEA-C02) certification exam in one go. Check out the features of the three formats.

>> Exam DEA-C02 Bible <<

Most Probable Real Exam Questions in Snowflake DEA-C02 PDF Dumps Format

The DEA-C02 certification lead you to numerous opportunities in career development and shaping your future. Just imagine that with the DEA-C02 certification, you can get a higher salary and a better position to help you lead a totally different and successful life. And with our DEA-C02 Exam Braindumps, it is easy to pass the exam and get the DEA-C02 certification. According to our data, our pass rate is high as 98% to 100%. You can pass the exam just by your first attempt.

Snowflake SnowPro Advanced: Data Engineer (DEA-C02) Sample Questions (Q258-Q263):

NEW QUESTION # 258

A healthcare provider stores patient data in Snowflake, including 'PATIENT ID', 'NAME', 'MEDICAL HISTORY', and 'INSURANCE ID'. They need to comply with HIPAA regulations. As a data engineer, you need to ensure that PHI (Protected Health Information) is masked appropriately based on user roles. Which of the following steps are NECESSARY to achieve this using Snowflake's data masking features and RBAC? (Select all that apply)

- A. Apply the created masking policies to the corresponding columns in the patient data tables, ensuring that the masking policies are designed to reveal only the necessary information based on the user's role (e.g., doctors see full medical history, nurses see limited medical history, admins see de-identified data).
- B. Grant the 'OWNERSHIP' privilege on the 'PATIENT' table to the 'ACCOUNTADMIN' role, ensuring complete control and management of the data by the administrator.
- C. Identify the columns containing PHI and create appropriate masking policies for each column (e.g., masking 'NAME', 'MEDICAL HISTORY', 'INSURANCE_ID').
- D. Create custom roles representing different user groups within the organization (e.g., 'DOCTOR', 'NURSE', 'ADMIN') and

- grant them the necessary privileges to access the data, including 'SELECT on the tables and views containing patient data.
- E. Enforce multi-factor authentication (MFA) for all users accessing the Snowflake environment to enhance security and prevent unauthorized access to sensitive data.

Answer: A,C,D

Explanation:

Options A, B, and C are all necessary steps for implementing data masking and RBAC for PHI protection. Identifying PHI and creating masking policies is crucial. Defining roles and granting privileges aligns access with job functions. Applying the masking policies enforces role-based data visibility. Option D is not as important, as another admin role may be more suitable (SECURITYADMIN) than the ACCOUNTADMIN, and option E, MFA does enhance security but is not directly related to Data Masking with RBAC in Snowflake.

NEW QUESTION # 259

You are tasked with implementing row-level security (RLS) on a 'SALES' table to restrict access based on the 'REGION' column. Users with the 'NORTH REGION ROLE' should only see data where 'REGION = 'NORTH''. You've created a row access policy named 'north_region_policy'. After applying the policy to the 'SALES' table, users with the 'NORTH REGION ROLE' are still seeing all rows.

Which of the following is the MOST likely reason for this and how can it be corrected?

- A. The ' does not have the USAGE privilege on the database and schema containing the 'SALES' table. Grant the USAGE privilege to the role.
- B. The user has not logged out and back in since the role was granted to them. Force the user to re-authenticate.
- C. The policy needs to be explicitly refreshed. Execute 'REFRESH ROW ACCESS POLICY north_region_policy ON SALES;'
- D. The policy function within is not using the correct context function to determine the user's role. It should use 'CURRENT_ROLE()' instead of 'CURRENT_USER()'
- E. The is not enabled. Execute 'ALTER ROW ACCESS POLICY ON SALES SET ENABLED = TRUE;'

Answer: A

Explanation:

Row access policies require the role to have USAGE privilege on the database and schema. Without this privilege, the policy cannot be enforced. The other options, while potentially relevant in other scenarios, are not the most likely cause for the described issue. Row access policies are automatically enabled when applied and the correct context function would be CURRENT_ROLE(). A refresh command is not required.

NEW QUESTION # 260

You are building a data pipeline that involves ingesting data from AWS S3 into Snowflake using Snowpipe. The data arrives in small files frequently, and you are experiencing performance issues with delayed data availability. You need to optimize the pipeline for near real-time ingestion. Which combination of strategies will MOST effectively address this scenario?

- A. Manually refresh the Snowpipe using the 'ALTER PIPE ... REFRESH' command every few minutes to force ingestion of new data.
- B. Implement a micro-batching process using a third-party tool (like Apache Spark) to aggregate the small files into larger batches before loading them into S3, then configure Snowpipe to ingest the larger files.
- C. Enable auto-ingest on the Snowpipe and increase the frequency of S3 event notifications to Snowflake. Combine this with clustering the target table on a relevant column to optimize query performance after loading.
- D. Increase the warehouse size used for the Snowpipe load process and adjust the 'MAX FILE SIZE' parameter on the pipe definition to match the size of the incoming files.
- E. Configure S3 event notifications to trigger Snowpipe only when a sufficient number of files have arrived in the S3 bucket, using a serverless function (like AWS Lambda) to manage the file count threshold.

Answer: C

Explanation:

Enabling auto-ingest on the Snowpipe and increasing the frequency of S3 event notifications is the most effective way to handle frequent small files. Auto-ingest automatically loads files as soon as they are available, and frequent notifications ensure minimal delay. Clustering the target table improves query performance by organizing data based on a relevant column, leading to faster data

retrieval. The option A related to MAX FILE SIZE is invalid since it does not exist in the context of Snowpipe. Options involving third-party tools add complexity and latency. Manually refreshing the pipe is not a scalable solution. Option C is not ideal because you want low latency, not delaying trigger. Option D provides the lowest latency at scale, while A, B, C and E would likely have a negative impact on latency, add extra steps, or are plain incorrect.

NEW QUESTION # 261

You are responsible for monitoring the performance of a Snowflake data pipeline that loads data from S3 into a Snowflake table named 'SALES DATA. You notice that the COPY INTO command consistently takes longer than expected. You want to implement telemetry to proactively identify the root cause of the performance degradation. Which of the following methods, used together, provide the MOST comprehensive telemetry data for troubleshooting the COPY INTO performance?

- A. Query the 'LOAD_HISTORY' function and monitor the network latency between S3 and Snowflake using an external monitoring tool.
- B. Query the 'COPY_HISTORY' view and the view in 'ACCOUNT_USAG Also, check the S3 bucket for throttling errors.
- C. Query the 'COPY HISTORY' view in the 'INFORMATION SCHEMA' and enable Snowflake's query profiling for the COPY INTO statement.
- D. Use Snowflake's partner connect integrations to monitor the virtual warehouse resource consumption and query the 'VALIDATE' function to ensure data quality before loading.
- E. Query the 'COPY HISTORY' view in the 'INFORMATION SCHEMA' and monitor CPU utilization of the virtual warehouse using the Snowflake web I-JL.

Answer: B,C

Explanation:

To comprehensively troubleshoot COPY INTO performance, you need data on the copy operation itself (COPY HISTORY), overall account and data validation. The COPY_HISTORY view provides details about each COPY INTO execution, including the file size, load time, and any errors encountered. Query profiling offers detailed insight into the internal operations of the COPY INTO command, revealing bottlenecks. Monitoring S3 for throttling ensures that the data source isn't limiting performance. Using helps correlate storage growth with load times. LOAD_HISTORY doesn't exist, 'VALIDATE' function is for data validation not performance. While warehouse CPU utilization is useful, it doesn't provide the specific details needed to diagnose COPY INTO issues. External network monitoring is also less relevant than checking for S3 throttling and analyzing Snowflake's internal telemetry data.

NEW QUESTION # 262

A Snowflake table 'ORDERS' contains billions of records and is frequently queried for reporting purposes. The reporting queries often filter on 'ORDER DATE' and 'CUSTOMER ID'. The data engineering team is considering creating a clustering key to improve query performance. They are evaluating two options: (1) clustering on 'ORDER DATE' alone and (2) clustering on '(ORDER DATE, CUSTOMER ID)'. Which of the following statements best describes the trade-offs between these two options in the context of query performance and data maintenance?

- A. Clustering on '(ORDER DATE, CUSTOMER ID)' is always the best option because it allows for more granular filtering and reduces the need to scan unnecessary micro-partitions, regardless of data distribution.
- B. Clustering on '(ORDER DATE, CUSTOMER ID)' will provide better performance for queries filtering on both columns, but may lead to increased reclustering costs if distribution is skewed within 'ORDER_DATE' partitions.
- C. Neither clustering option will significantly improve performance, and the team should focus on optimizing the queries themselves through techniques like query rewriting and the use of appropriate indexes.
- D. Clustering on ORDER DATE alone is preferable because it eliminates the risk of data skewness associated with 'CUSTOMER ID, leading to more balanced micro-partitions and consistent query performance.
- E. Clustering on 'ORDER_DATE' alone will result in better overall query performance because it's a single dimension, simplifying the clustering process and reducing the need for Snowflake to perform complex data scans.

Answer: B

Explanation:

Clustering on multiple columns (in this case, 'ORDER_DATE' and 'CUSTOMER_ID') improves performance for queries that filter on both columns. However, if the distribution of is uneven within the 'ORDER DATE' partitions (skew), it can lead to some micro-partitions still containing a large range of 'CUSTOMER_ID' values, requiring Snowflake to scan more data. This can also increase reclustering costs as Snowflake tries to optimize the clustering. Clustering on only may be beneficial if most queries primarily filter on date. Option E is incorrect because clustering is a Snowflake feature that improves performance.

