

Here's the Quick Way to Crack Data-Engineer-Associate Certification Exam



What's more, part of that TestsDumps Data-Engineer-Associate dumps now are free: https://drive.google.com/open?id=110uCcT5_7iDdsog1eFBm_yCaETucKLLKv

At TestsDumps, we are committed to providing our clients with the actual and latest Amazon Data-Engineer-Associate exam questions. Our real Data-Engineer-Associate exam questions in three formats are designed to save time and help you clear the Data-Engineer-Associate Certification Exam in a short time. Preparing with TestsDumps's updated Data-Engineer-Associate exam questions is a great way to complete preparation in a short time and pass the Data-Engineer-Associate test in one sitting.

TestsDumps is a very wonderful and effective platform to give chances to our worthy clients who want to achieve their expected scores and gain their Data-Engineer-Associate certifications. With our professional experts' tireless efforts, our Data-Engineer-Associate exam guide is equipped with a simulated examination system with timing function, allowing you to examine your learning results at any time, keep checking for defects, and improve your strength. And you can be satisfied with our Data-Engineer-Associate learning guide.

>> Data-Engineer-Associate Reliable Exam Tips <<

Data-Engineer-Associate Valid Test Cost, Test Data-Engineer-Associate Simulator Free

In a word, you can try our free Data-Engineer-Associate study guide demo before purchasing, AWS Certified Data Engineer - Associate (DEA-C01) Pdf After the researches of many years, we found only the true subject of past-year exam was authoritative and had time-validity, For your benefit, TestsDumps is putting forth you to attempt the free demo and Amazon Data-Engineer-Associate Exam Dumps the best quality highlights of the item, because nobody gives this facility only the TestsDumps Data-

Engineer-Associate Free Learning provide this facility. The example on the right was a simple widget designed Reliable Data-Engineer-Associate Pdf to track points in a rewards program, The pearsonvue website is not affiliated with us, Although computers are great at gathering, manipulating, and calculating raw data, humans prefer their data presented in an orderly fashion.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q36-Q41):

NEW QUESTION # 36

A company uses Amazon Redshift for its data warehouse. A data engineer must query a table named orders.complete_orders_history, which contains 100 columns. The query must return all columns except columns named company_id and unique_system_id.

Which Amazon Redshift SQL statement will meet this requirement?

- A. `SELECT * TRUNCATE company_id, unique_system_id FROM orders.complete_orders_history;`
- B. `SELECT * NOT IN company_id, unique_system_id FROM orders.complete_orders_history;`
- C. `SELECT * EXCLUDE company_id, unique_system_id FROM orders.complete_orders_history;`
- D. `SELECT * EXCEPT company_id, unique_system_id FROM orders.complete_orders_history;`

Answer: C

Explanation:

Option A is correct because Amazon Redshift supports the EXCLUDE clause in the SELECT list for exactly this use case: returning all columns from a wide table except a small set of unwanted columns. AWS documentation states that "EXCLUDE column_list names the columns that are excluded from the query results" and notes that the option is helpful when only a subset of columns must be excluded from a wide table. That matches this question precisely because the table contains 100 columns and only company_id and unique_system_id need to be omitted.

Option B is invalid SQL syntax in Amazon Redshift because NOT IN is used in predicates, not in the select list to remove columns.

Option C is also incorrect because EXCEPT in Redshift is a set operator used between query result sets, not a select-list column exclusion feature. AWS documents EXCEPT alongside UNION and INTERSECT as result-set comparison operators, not as a way to omit columns from SELECT *.

Option D is not valid Redshift SQL syntax for column selection. Therefore, the correct and documented Redshift statement is `SELECT * EXCLUDE company_id, unique_system_id FROM orders.complete_orders_history;`

NEW QUESTION # 37

A company runs an AWS Glue workflow every day to process time series data from an Amazon S3 bucket.

The workflow loads the data into an Amazon Redshift Serverless table. The company observes that some of the jobs in the workflow occasionally fail.

A data engineer must receive a notification when the Redshift table does not contain the most recent data.

Which solution will meet this requirement in the MOST operationally efficient way?

- A. Create an Amazon CloudWatch dashboard that displays a metric named Failed AWS Glue Jobs that counts AWS Glue job failures during the previous day. Set a CloudWatch alarm to send a notification when the metric value exceeds zero.
- B. Load AWS Glue job logs to an Amazon S3 bucket. Configure an Amazon CloudWatch alarm to send a notification when the job logs in the S3 bucket contain Job.State=FAILED.
- C. Schedule an AWS Glue Data Quality job to check the freshness of the data. Create an Amazon EventBridge rule to notify an Amazon Simple Notification Service (Amazon SNS) topic when a data quality rule fails.
- D. Configure an Amazon EventBridge Scheduler to run an Amazon Macie job to scan the Redshift table for data freshness. Configure Macie to notify an Amazon Simple Notification Service (Amazon SNS) topic when an AWS Glue job fails.

Answer: C

Explanation:

Option B is the most operationally efficient because it checks the business requirement directly: whether the target table contains the most recent data, not merely whether a job failed. Monitoring only failures (Options C and D) can produce false positives (a job failure might not impact freshness) and false negatives (a job can succeed but still load stale or incomplete data). The study material emphasizes implementing data quality validation as part of the ETL process so data can be verified before or as it is stored, rather than relying only on pipeline execution status.

Using a data quality rule focused on freshness (for example, validating that a "max event timestamp" or

"latest partition date" meets today's expected value) lets the pipeline detect stale loads even when the workflow runs. Then, an

EventBridge rule can route failures of that data quality check to SNS for immediate notification, keeping operations serverless and centralized. Macie (Option A) is designed for sensitive-data discovery/classification, not operational "freshness" checks on Redshift tables, so it adds unnecessary services and effort compared to a Glue-native data quality validation approach.

NEW QUESTION # 38

A company is uploading log files from on-premises servers to an Amazon S3 bucket. The company needs to validate that the logs from the on-premises servers are the same as the logs that are stored in the S3 bucket.

Which solution will meet this requirement?

- **A. Use the AWS SDK to automatically compute CRC32 checksums during the upload. Store the checksums in S3 object metadata.**
- B. Enable S3 Object Lock in compliance mode on the S3 bucket. Upload the objects to the bucket.
- C. Create an AWS Lambda function to calculate SHA-256 checksums. Store the results in a separate metadata table. Validate the logs after the upload.
- D. After uploading the objects to the S3 bucket, enable S3 Object Lock in governance mode on the S3 objects.

Answer: A

Explanation:

Amazon S3 natively supports checksum validation during uploads when objects are uploaded by using the AWS SDK. The SDK can automatically calculate a checksum, such as CRC32, on the client side and send it with the upload request. Amazon S3 validates the checksum upon receipt and stores it with the object, ensuring that the data uploaded is exactly the same as the source data. Storing the checksum in the S3 object metadata provides a lightweight and cost-effective way to verify data integrity without requiring additional compute, storage, or post-processing steps. This approach validates the logs at upload time, which is the most efficient and reliable method.

Using AWS Lambda to calculate SHA-256 checksums would require additional compute, orchestration, and storage, increasing cost and operational complexity. Amazon S3 Object Lock enforces immutability and retention but does not validate data integrity between source and destination. Governance or compliance mode Object Lock cannot detect whether uploaded data matches the original logs.

Therefore, using the AWS SDK checksum validation during upload is the simplest, most efficient, and AWS- recommended approach for ensuring log integrity.

NEW QUESTION # 39

A financial company wants to use Amazon Athena to run on-demand SQL queries on a petabyte-scale dataset to support a business intelligence (BI) application. An AWS Glue job that runs during non-business hours updates the dataset once every day. The BI application has a standard data refresh frequency of 1 hour to comply with company policies.

A data engineer wants to cost optimize the company's use of Amazon Athena without adding any additional infrastructure costs.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Change the format of the files that are in the dataset to Apache Parquet.
- B. Add an Amazon ElastiCache cluster between the BI application and Athena.
- C. Configure an Amazon S3 Lifecycle policy to move data to the S3 Glacier Deep Archive storage class after 1 day
- **D. Use the query result reuse feature of Amazon Athena for the SQL queries.**

Answer: D

Explanation:

The best solution to cost optimize the company's use of Amazon Athena without adding any additional infrastructure costs is to use the query result reuse feature of Amazon Athena for the SQL queries. This feature allows you to run the same query multiple times without incurring additional charges, as long as the underlying data has not changed and the query results are still in the query result location in Amazon S3. This feature is useful for scenarios where you have a petabyte-scale dataset that is updated infrequently, such as once a day, and you have a BI application that runs the same queries repeatedly, such as every hour. By using the query result reuse feature, you can reduce the amount of data scanned by your queries and save on the cost of running Athena. You can enable or disable this feature at the workgroup level or at the individual query level.

Option A is not the best solution, as configuring an Amazon S3 Lifecycle policy to move data to the S3 Glacier Deep Archive storage class after 1 day would not cost optimize the company's use of Amazon Athena, but rather increase the cost and complexity. Amazon S3 Lifecycle policies are rules that you can define to automatically transition objects between different storage classes based on specified criteria, such as the age of the object. S3 Glacier Deep Archive is the lowest-cost storage class in Amazon S3, designed for long-term data archiving that is accessed once or twice in a year. While moving data to S3 Glacier Deep Archive can

reduce the storage cost, it would also increase the retrieval cost and latency, as it takes up to 12 hours to restore the data from S3 Glacier Deep Archive³. Moreover, Athena does not support querying data that is in S3 Glacier or S3 Glacier Deep Archive storage classes⁴. Therefore, using this option would not meet the requirements of running on-demand SQL queries on the dataset.

Option C is not the best solution, as adding an Amazon ElastiCache cluster between the BI application and Athena would not cost optimize the company's use of Amazon Athena, but rather increase the cost and complexity. Amazon ElastiCache is a service that offers fully managed in-memory data stores, such as Redis and Memcached, that can improve the performance and scalability of web applications by caching frequently accessed data. While using ElastiCache can reduce the latency and load on the BI application, it would not reduce the amount of data scanned by Athena, which is the main factor that determines the cost of running Athena. Moreover, using ElastiCache would introduce additional infrastructure costs and operational overhead, as you would have to provision, manage, and scale the ElastiCache cluster, and integrate it with the BI application and Athena.

Option D is not the best solution, as changing the format of the files that are in the dataset to Apache Parquet would not cost optimize the company's use of Amazon Athena without adding any additional infrastructure costs, but rather increase the complexity. Apache Parquet is a columnar storage format that can improve the performance of analytical queries by reducing the amount of data that needs to be scanned and providing efficient compression and encoding schemes. However, changing the format of the files that are in the dataset to Apache Parquet would require additional processing and transformation steps, such as using AWS Glue or Amazon EMR to convert the files from their original format to Parquet, and storing the converted files in a separate location in Amazon S3. This would increase the complexity and the operational overhead of the data pipeline, and also incur additional costs for using AWS Glue or Amazon EMR. Reference:

Query result reuse

Amazon S3 Lifecycle

S3 Glacier Deep Archive

Storage classes supported by Athena

[What is Amazon ElastiCache?]

[Amazon Athena pricing]

[Columnar Storage Formats]

AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

NEW QUESTION # 40

A data engineer needs to run a data transformation job whenever a user adds a file to an Amazon S3 bucket.

The job will run for less than 1 minute. The job must send the output through an email message to the data engineer. The data engineer expects users to add one file every hour of the day.

Which solution will meet these requirements in the MOST operationally efficient way?

- **A. Create an AWS Lambda function to transform the data. Use Amazon S3 Event Notifications to invoke the Lambda function when a new object is created. Publish the output to an Amazon Simple Notification Service (Amazon SNS) topic. Subscribe the data engineer's email account to the topic.**
- B. Deploy an Amazon EMR cluster. Use EMR File System (EMRFS) to access the files in the S3 bucket. Run transformation code on a schedule to generate the output to a second S3 bucket. Create an Amazon Simple Notification Service (Amazon SNS) topic. Configure Amazon S3 Event Notifications to notify the topic when a new object is created.
- C. Create a small Amazon EC2 instance that polls the S3 bucket for new files. Run transformation code on a schedule to generate the output. Use operating system commands to send email messages.
- D. Run an Amazon Elastic Container Service (Amazon ECS) task to poll the S3 bucket for new files. Run transformation code on a schedule to generate the output. Use operating system commands to send email messages.

Answer: A

NEW QUESTION # 41

.....

In order to pass the exam and fight for a brighter future, these people who want to change themselves need to put their ingenuity and can do spirit to work. More importantly, it is necessary for these people to choose the convenient and helpful Data-Engineer-Associate study materials as their study tool in the next time. Because their time is not enough to prepare for the exam, and a lot of people have difficulty in preparing for the exam, so many people who want to pass the Data-Engineer-Associate Exam and get the related certification in a short time have to pay more attention to the study materials.

Data-Engineer-Associate Valid Test Cost: https://www.testsdumps.com/Data-Engineer-Associate_real-exam-dumps.html

Amazon Data-Engineer-Associate Reliable Exam Tips Every product we have sold to customer will enjoy considerate after-sales service, Amazon Data-Engineer-Associate Reliable Exam Tips Our advanced and convenient learning method, And it is easy and

