

Valid Data-Engineer-Associate Exam Dumps - Data-Engineer-Associate Study Plan



What's more, part of that Exam4Labs Data-Engineer-Associate dumps now are free: <https://drive.google.com/open?id=1FCJURvYsiMmsSMa5cdut-e93pudli-S3>

Up to now we classify our Data-Engineer-Associate exam questions as three different versions. They are pdf, software and the most convenient one APP online. Though the content of these three versions is the same, but their displays are different. Each of them has their respective feature and advantage including new information that you need to know to pass the Data-Engineer-Associate test. So you can choose the version of Data-Engineer-Associate training quiz according to your personal preference.

Rather than pretentious help for customers, our after-sales services are authentic and faithful. Many clients cannot stop praising us in this aspect and become regular customer for good. We have strict criterion to help you with the standard of our Data-Engineer-Associate training materials. Our company has also being Customer First. So we consider the facts of your interest firstly. All the preoccupation based on your needs and all these explain our belief to help you have satisfactory and comfortable purchasing services. We assume all the responsibilities our Data-Engineer-Associate simulating practice may bring you foreseeable outcomes and you will not regret for believing in us assuredly.

>> Valid Data-Engineer-Associate Exam Dumps <<

Data-Engineer-Associate Study Plan - Test Data-Engineer-Associate Dumps Free

Amazon Data-Engineer-Associate dumps PDF version is printable and embedded with valid Amazon Data-Engineer-Associate questions to help you get ready for the Data-Engineer-Associate exam quickly. AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) exam dumps pdf are also usable on several smart devices. You can use it anywhere at any time on your smartphones and tablets.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q222-Q227):

NEW QUESTION # 222

A company is building a data lake for a new analytics team. The company is using Amazon S3 for storage and Amazon Athena for query analysis. All data that is in Amazon S3 is in Apache Parquet format.

The company is running a new Oracle database as a source system in the company's data center. The company has 70 tables in the Oracle database. All the tables have primary keys. Data can occasionally change in the source system. The company wants to ingest the tables every day into the data lake.

Which solution will meet this requirement with the LEAST effort?

- A. Create an AWS Database Migration Service (AWS DMS) task for ongoing replication. Set the Oracle database as the source. Set Amazon S3 as the target. Configure the task to write the data in Parquet format.
- B. Create an AWS Glue connection to the Oracle database. Create an AWS Glue bookmark job to ingest the data incrementally and to write the data to Amazon S3 in Parquet format.

- C. Create an Oracle database in Amazon RDS. Use AWS Database Migration Service (AWS DMS) to migrate the on-premises Oracle database to Amazon RDS. Configure triggers on the tables to invoke AWS Lambda functions to write changed records to Amazon S3 in Parquet format.
- D. Create an Apache Sqoop job in Amazon EMR to read the data from the Oracle database. Configure the Sqoop job to write the data to Amazon S3 in Parquet format.

Answer: A

Explanation:

The company needs to ingest tables from an on-premises Oracle database into a data lake on Amazon S3 in Apache Parquet format. The most efficient solution, requiring the least manual effort, would be to use AWS Database Migration Service (DMS) for continuous data replication.

Option C: Create an AWS Database Migration Service (AWS DMS) task for ongoing replication. Set the Oracle database as the source. Set Amazon S3 as the target. Configure the task to write the data in Parquet format.

AWS DMS can continuously replicate data from the Oracle database into Amazon S3, transforming it into Parquet format as it ingests the data. DMS simplifies the process by providing ongoing replication with minimal setup, and it automatically handles the conversion to Parquet format without requiring manual transformations or separate jobs. This option is the least effort solution since it automates both the ingestion and transformation processes.

Other options:

Option A (Apache Sqoop on EMR) involves more manual configuration and management, including setting up EMR clusters and writing Sqoop jobs.

Option B (AWS Glue bookmark job) involves configuring Glue jobs, which adds complexity. While Glue supports data transformations, DMS offers a more seamless solution for database replication.

Option D (RDS and Lambda triggers) introduces unnecessary complexity by involving RDS and Lambda for a task that DMS can handle more efficiently.

Reference:

AWS Database Migration Service (DMS)

DMS S3 Target Documentation

NEW QUESTION # 223

A retail company stores data from a product lifecycle management (PLM) application in an on-premises MySQL database. The PLM application frequently updates the database when transactions occur.

The company wants to gather insights from the PLM application in near real time. The company wants to integrate the insights with other business datasets and to analyze the combined dataset by using an Amazon Redshift data warehouse.

The company has already established an AWS Direct Connect connection between the on-premises infrastructure and AWS.

Which solution will meet these requirements with the LEAST development effort?

- A. Run a full load plus CDC task in AWS Database Migration Service (AWS DMS) to continuously replicate the MySQL database changes. Set Amazon Redshift as the destination for the task.
- B. Run a scheduled AWS Glue extract, transform, and load (ETL) job to get the MySQL database updates by using a Java Database Connectivity (JDBC) connection. Set Amazon Redshift as the destination for the ETL job.
- C. Use the Amazon AppFlow SDK to build a custom connector for the MySQL database to continuously replicate the database changes. Set Amazon Redshift as the destination for the connector.
- D. Run scheduled AWS DataSync tasks to synchronize data from the MySQL database. Set Amazon Redshift as the destination for the tasks.

Answer: A

Explanation:

* Problem Analysis:

* The company needs near real-time replication of MySQL updates to Amazon Redshift.

* Minimal development effort is required for this solution.

* Key Considerations:

* AWS DMS provides a full load + CDC (Change Data Capture) mode for continuous replication of database changes.

* DMS integrates natively with both MySQL and Redshift, simplifying setup.

* Solution Analysis:

* Option A: AWS Glue Job

* Glue is batch-oriented and does not support near real-time replication.

* Option B: DMS with Full Load + CDC

* Efficiently handles initial database load and continuous updates.

- * Requires minimal setup and operational overhead.
- * Option C: AppFlow SDK
- * AppFlow is not designed for database replication. Custom connectors increase development effort.
- * Option D: DataSync
- * DataSync is for file synchronization and not suitable for database updates.
- * Final Recommendation:
- * Use AWS DMS in full load + CDC mode for continuous replication.

:

AWS Database Migration Service Documentation
Setting Up DMS with Redshift

NEW QUESTION # 224

A company has used an Amazon Redshift table that is named Orders for 6 months. The company performs weekly updates and deletes on the table. The table has an interleaved sort key on a column that contains AWS Regions. The company wants to reclaim disk space so that the company will not run out of storage space. The company also wants to analyze the sort key column.

Which Amazon Redshift command will meet these requirements?

- A. VACUUM SORT ONLY Orders
- **B. VACUUM REINDEX Orders**
- C. VACUUM FULL Orders
- D. VACUUM DELETE ONLY Orders

Answer: B

Explanation:

Amazon Redshift is a fully managed, petabyte-scale data warehouse service that enables fast and cost-effective analysis of large volumes of data. Amazon Redshift uses columnar storage, compression, and zone maps to optimize the storage and performance of data. However, over time, as data is inserted, updated, or deleted, the physical storage of data can become fragmented, resulting in wasted disk space and degraded query performance. To address this issue, Amazon Redshift provides the VACUUM command, which reclaims disk space and resorts rows in either a specified table or all tables in the current schema.

The VACUUM command has four options: FULL, DELETE ONLY, SORT ONLY, and REINDEX. The option that best meets the requirements of the question is VACUUM REINDEX, which re-sorts the rows in a table that has an interleaved sort key and rewrites the table to a new location on disk. An interleaved sort key is a type of sort key that gives equal weight to each column in the sort key, and stores the rows in a way that optimizes the performance of queries that filter by multiple columns in the sort key. However, as data is added or changed, the interleaved sort order can become skewed, resulting in suboptimal query performance. The VACUUM REINDEX option restores the optimal interleaved sort order and reclaims disk space by removing deleted rows. This option also analyzes the sort key column and updates the table statistics, which are used by the query optimizer to generate the most efficient query execution plan²³.

The other options are not optimal for the following reasons:

- A . VACUUM FULL Orders. This option reclaims disk space by removing deleted rows and resorts the entire table. However, this option is not suitable for tables that have an interleaved sort key, as it does not restore the optimal interleaved sort order. Moreover, this option is the most resource-intensive and time-consuming, as it rewrites the entire table to a new location on disk.
- B . VACUUM DELETE ONLY Orders. This option reclaims disk space by removing deleted rows, but does not resort the table. This option is not suitable for tables that have any sort key, as it does not improve the query performance by restoring the sort order. Moreover, this option does not analyze the sort key column and update the table statistics.
- D . VACUUM SORT ONLY Orders. This option resorts the entire table, but does not reclaim disk space by removing deleted rows. This option is not suitable for tables that have an interleaved sort key, as it does not restore the optimal interleaved sort order. Moreover, this option does not analyze the sort key column and update the table statistics.

Reference:

- 1: Amazon Redshift VACUUM
- 2: Amazon Redshift Interleaved Sorting
- 3: Amazon Redshift ANALYZE

NEW QUESTION # 225

A company uses AWS Glue Data Catalog to index data that is uploaded to an Amazon S3 bucket every day.

The company uses a daily batch processes in an extract, transform, and load (ETL) pipeline to upload data from external sources into the S3 bucket.

The company runs a daily report on the S3 data. Some days, the company runs the report before all the daily data has been uploaded to the S3 bucket. A data engineer must be able to send a message that identifies any incomplete data to an existing Amazon Simple Notification Service (Amazon SNS) topic.

Which solution will meet this requirement with the LEAST operational overhead?

- **A. Create data quality checks on the source datasets that the daily reports use. Create data quality actions by using AWS Glue workflows to confirm the completeness and consistency of the datasets. Configure the data quality actions to create an event in Amazon EventBridge if a dataset is incomplete. Configure EventBridge to send the event that informs the data engineer about the incomplete datasets to the Amazon SNS topic.**
- B. Create AWS Lambda functions that run data quality queries on the columns data type and the presence of null values. Orchestrate the ETL pipeline by using an AWS Step Functions workflow that runs the Lambda functions. Configure the Step Functions workflow to send an email notification that informs the data engineer about the incomplete datasets to the SNS topic.
- C. Create data quality checks for the source datasets that the daily reports use. Create a new AWS managed Apache Airflow cluster. Run the data quality checks by using Airflow tasks that run data quality queries on the columns data type and the presence of null values. Configure Airflow Directed Acyclic Graphs (DAGs) to send an email notification that informs the data engineer about the incomplete datasets to the SNS topic.
- D. Create data quality checks on the source datasets that the daily reports use. Create a new Amazon EMR cluster. Use Apache Spark SQL to create Apache Spark jobs in the EMR cluster that run data quality queries on the columns data type and the presence of null values. Orchestrate the ETL pipeline by using an AWS Step Functions workflow. Configure the workflow to send an email notification that informs the data engineer about the incomplete datasets to the SNS topic.

Answer: A

Explanation:

AWS Glue workflows are designed to orchestrate the ETL pipeline, and you can create data quality checks to ensure the uploaded datasets are complete before running reports. If there is an issue with the data, AWS Glue workflows can trigger an Amazon EventBridge event that sends a message to an SNS topic.

* AWS Glue Workflows:

* AWS Glue workflows allow users to automate and monitor complex ETL processes. You can include data quality actions to check for null values, data types, and other consistency checks.

* In the event of incomplete data, an EventBridge event can be generated to notify via SNS.

Reference: AWS Glue Workflows

Alternatives Considered:

A (Airflow cluster): Managed Airflow introduces more operational overhead and complexity compared to Glue workflows.

B (EMR cluster): Setting up an EMR cluster is also more complex compared to the Glue-centric solution.

D (Lambda functions): While Lambda functions can work, using Glue workflows offers a more integrated and lower operational overhead solution.

References:

AWS Glue Workflow Documentation

NEW QUESTION # 226

A company is building an analytics solution. The solution uses Amazon S3 for data lake storage and Amazon Redshift for a data warehouse. The company wants to use Amazon Redshift Spectrum to query the data that is in Amazon S3.

Which actions will provide the FASTEST queries? (Choose two.)

- A. Split the data into files that are less than 10 KB.
- **B. Partition the data based on the most common query predicates.**
- C. Use gzip compression to compress individual files to sizes that are between 1 GB and 5 GB.
- **D. Use a columnar storage file format.**
- E. Use file formats that are not

Answer: B,D

Explanation:

Amazon Redshift Spectrum is a feature that allows you to run SQL queries directly against data in Amazon S3, without loading or transforming the data. Redshift Spectrum can query various data formats, such as CSV, JSON, ORC, Avro, and Parquet.

However, not all data formats are equally efficient for querying. Some data formats, such as CSV and JSON, are row-oriented, meaning that they store data as a sequence of records, each with the same fields. Row-oriented formats are suitable for loading and exporting data, but they are not optimal for analytical queries that often access only a subset of columns. Row-oriented formats also

do not support compression or encoding techniques that can reduce the data size and improve the query performance.

On the other hand, some data formats, such as ORC and Parquet, are column-oriented, meaning that they store data as a collection of columns, each with a specific data type. Column-oriented formats are ideal for analytical queries that often filter, aggregate, or join data by columns. Column-oriented formats also support compression and encoding techniques that can reduce the data size and improve the query performance. For example, Parquet supports dictionary encoding, which replaces repeated values with numeric codes, and run-length encoding, which replaces consecutive identical values with a single value and a count. Parquet also supports various compression algorithms, such as Snappy, GZIP, and ZSTD, that can further reduce the data size and improve the query performance.

Therefore, using a columnar storage file format, such as Parquet, will provide faster queries, as it allows Redshift Spectrum to scan only the relevant columns and skip the rest, reducing the amount of data read from S3. Additionally, partitioning the data based on the most common query predicates, such as date, time, region, etc., will provide faster queries, as it allows Redshift Spectrum to prune the partitions that do not match the query criteria, reducing the amount of data scanned from S3. Partitioning also improves the performance of joins and aggregations, as it reduces data skew and shuffling.

The other options are not as effective as using a columnar storage file format and partitioning the data. Using gzip compression to compress individual files to sizes that are between 1 GB and 5 GB will reduce the data size, but it will not improve the query performance significantly, as gzip is not a splittable compression algorithm and requires decompression before reading. Splitting the data into files that are less than 10 KB will increase the number of files and the metadata overhead, which will degrade the query performance. Using file formats that are not supported by Redshift Spectrum, such as XML, will not work, as Redshift Spectrum will not be able to read or parse the data. Reference:

Amazon Redshift Spectrum

Choosing the Right Data Format

AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide, Chapter 4: Data Lakes and Data Warehouses, Section 4.3: Amazon Redshift Spectrum

NEW QUESTION # 227

.....

The Data-Engineer-Associate exam is on trend but the main problem that every applicant faces while preparing for it is not making the right choice of the Data-Engineer-Associate Questions. They struggle to find the right platform to get actual Data-Engineer-Associate exam questions and achieve their goals. Exam4Labs has made the product after seeing the students struggle to solve their issues and help them pass the Data-Engineer-Associate Certification Exam on the first try. Exam4Labs has designed this Data-Engineer-Associate practice test material after consulting with a lot of professionals and getting their good reviews so our customers can clear Data-Engineer-Associate certification exam quickly and improve themselves.

Data-Engineer-Associate Study Plan: <https://www.exam4labs.com/Data-Engineer-Associate-practice-torrent.html>

Exam4Labs Data-Engineer-Associate Study Plan's products are absolutely exam-oriented, Because our Data-Engineer-Associate exam dumps are frequently updated and reviewed by our experts, Amazon Valid Data-Engineer-Associate Exam Dumps If you choose us, we will let your efforts be paid off, 100% Trustworthy Certification Assistance With own professional certification expert team and customer support, Exam4Labs Data-Engineer-Associate Study Plan provide the most valid exam dumps with the latest real exam questions and accurate answers, We can promise that the three different versions of our Data-Engineer-Associate exam questions are equipment with the high quality.

Adding a user or a service is forbidden, Location where Test Data-Engineer-Associate Dumps Free the widget will ultimately be displayed on the page, Exam4Labs's products are absolutely exam-oriented.

Because our Data-Engineer-Associate Exam Dumps are frequently updated and reviewed by our experts, If you choose us, we will let your efforts be paid off, 100% Trustworthy Certification Assistance With own professional certification expert team and customer Data-Engineer-Associate support, Exam4Labs provide the most valid exam dumps with the latest real exam questions and accurate answers.

100% Pass Quiz Data-Engineer-Associate - Valid AWS Certified Data Engineer - Associate (DEA-C01) Exam Dumps

We can promise that the three different versions of our Data-Engineer-Associate exam questions are equipment with the high quality.

- Free PDF Quiz 2026 Data-Engineer-Associate: Reliable Valid AWS Certified Data Engineer - Associate (DEA-C01) Exam Dumps Immediately open www.validtorrent.com and search for "Data-Engineer-Associate" to obtain a free download Data-Engineer-Associate Accurate Study Material
- Data-Engineer-Associate Latest Exam Dumps Data-Engineer-Associate Latest Exam Dumps Valid Data-Engineer-

Associate Test Simulator □ Download ▷ Data-Engineer-Associate ◁ for free by simply searching on ► www.pdfvce.com
□ □ Trustworthy Data-Engineer-Associate Source

- 100% Pass 2026 Amazon Data-Engineer-Associate –Efficient Valid Exam Dumps □ Immediately open □
www.troytecdumps.com □ and search for ► Data-Engineer-Associate □ to obtain a free download □ Reliable Data-Engineer-Associate Exam Online
- Quiz 2026 Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Accurate Valid Exam Dumps □
□ The page for free download of ⇒ Data-Engineer-Associate ⇐ on ► www.pdfvce.com □ will open immediately □ Latest Data-Engineer-Associate Exam Papers
- Data-Engineer-Associate New Test Camp □ Data-Engineer-Associate Latest Exam Dumps □ Latest Data-Engineer-Associate Test Question □ The page for free download of □ Data-Engineer-Associate □ on □ www.troytecdumps.com □ will open immediately □ Valid Exam Data-Engineer-Associate Practice
- Valid Data-Engineer-Associate Exam Dumps | Pass-Sure Amazon Data-Engineer-Associate Study Plan: AWS Certified Data Engineer - Associate (DEA-C01) □ Search for ► Data-Engineer-Associate □ on [www.pdfvce.com] immediately to obtain a free download □ Reliable Data-Engineer-Associate Exam Online
- Amazon Data-Engineer-Associate Web-Based Practice Test Questions □ Open website ☀ www.pdfdumps.com □ ☀ □ and search for □ Data-Engineer-Associate □ for free download □ Data-Engineer-Associate Sample Questions Pdf
- Trustworthy Data-Engineer-Associate Source □ Data-Engineer-Associate Relevant Questions □ Data-Engineer-Associate Exam Flashcards □ Copy URL { www.pdfvce.com } open and search for □ Data-Engineer-Associate □ to download for free □ Data-Engineer-Associate Certification Exam Cost
- Valid Data-Engineer-Associate Test Simulator □ Top Data-Engineer-Associate Exam Dumps □ Trustworthy Data-Engineer-Associate Source □ Easily obtain ► Data-Engineer-Associate □ for free download through ✓
www.pdfdumps.com □ ✓ □ □ Data-Engineer-Associate Sample Questions Pdf
- Web-Based Practice Tests: The Key to Amazon Data-Engineer-Associate Exam Success □ □ www.pdfvce.com □ is best website to obtain □ Data-Engineer-Associate □ for free download □ Data-Engineer-Associate Latest Exam Testking
- 100% Pass 2026 Trustable Amazon Data-Engineer-Associate: Valid AWS Certified Data Engineer - Associate (DEA-C01) Exam Dumps 📖 Immediately open { www.testkingpass.com } and search for ☀ Data-Engineer-Associate □ ☀ □ to obtain a free download □ Certificate Data-Engineer-Associate Exam
- woodyzbei732239.blogdal.com, sahilycfe236949.shoutmyblog.com, blakefkz897484.blogrenanda.com, alyssabdcg585276.azzablog.com, socialmediainuk.com, exactlybookmarks.com, aprilkzh829954.dgbloggers.com, barryefok592006.tusblogs.com, nanniehiph307887.goabroadblog.com, hainatsws810810.blogoxo.com, Disposable vapes

BTW, DOWNLOAD part of Exam4Labs Data-Engineer-Associate dumps from Cloud Storage: <https://drive.google.com/open?id=1FCJURvYsiMmsSMa5cdut-e93pudIi-S3>