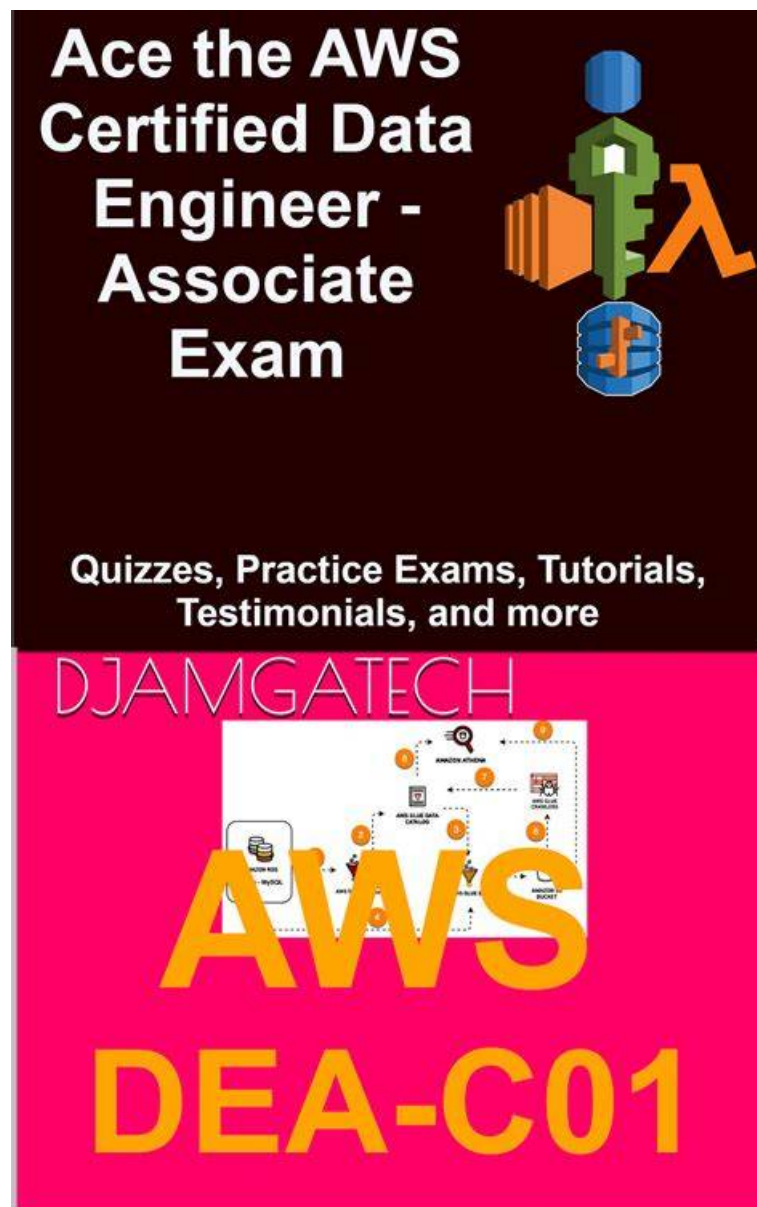


Amazon certification Data-Engineer-Associate exam questions and answers come out



BTW, DOWNLOAD part of VerifiedDumps Data-Engineer-Associate dumps from Cloud Storage: https://drive.google.com/open?id=1QId1j4cU45fDt4A__8RogCJyontn3K4j

One of our outstanding advantages is our high passing rate, which has reached 99%, and much higher than the average pass rate among our peers. Our high passing rate explains why we are the top Data-Engineer-Associate prep guide in our industry. One point does farm work one point harvest, depending on strength speech! The source of our confidence is our wonderful Data-Engineer-Associate exam questions. Passing the exam won't be a problem as long as you keep practice with our Data-Engineer-Associate Study Materials about 20 to 30 hours. Considered many of the candidates are too busy to review, our experts designed the Data-Engineer-Associate question dumps in accord with actual examination questions, which would help you pass the exam with high proficiency.

Amazon Data-Engineer-Associate Exam Syllabus Topics:

Section	Weight	Objectives
Data Store Management	26%	- Optimize storage performance and cost - Select appropriate data storage solutions

Data Operations and Support	22%	- Troubleshoot data workflow issues - Monitor and maintain data pipelines
Data Security and Governance	18%	- Implement data security controls - Apply governance and compliance best practices
Data Ingestion and Transformation	34%	- Build and manage data pipelines - Ingest and transform data using AWS services

>> Accurate Data-Engineer-Associate Prep Material <<

Free PDF 2026 Amazon High-quality Data-Engineer-Associate: Accurate AWS Certified Data Engineer - Associate (DEA-C01) Prep Material

The AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) certification is a valuable credential that every Amazon professional should earn it. The Data-Engineer-Associate certification exam offers a great opportunity for beginners and experienced professionals to demonstrate their expertise. With the AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) certification exam everyone can upgrade their skills and knowledge. There are other several benefits that the Amazon Data-Engineer-Associate exam holders can achieve after the success of the AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) certification exam.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q171-Q176):

NEW QUESTION # 171

A company uploads .csv files to an Amazon S3 bucket. The company's data platform team has set up an AWS Glue crawler to perform data discovery and to create the tables and schemas.

An AWS Glue job writes processed data from the tables to an Amazon Redshift database. The AWS Glue job handles column mapping and creates the Amazon Redshift tables in the Redshift database appropriately.

If the company reruns the AWS Glue job for any reason, duplicate records are introduced into the Amazon Redshift tables. The company needs a solution that will update the Redshift tables without duplicates.

Which solution will meet these requirements?

- A. Modify the AWS Glue job to copy the rows into a staging Redshift table. Add SQL commands to update the existing rows with new values from the staging Redshift table.
- B. Modify the AWS Glue job to load the previously inserted data into a MySQL database. Perform an upsert operation in the MySQL database. Copy the results to the Amazon Redshift tables.
- C. Use Apache Spark's DataFrame dropDuplicates() API to eliminate duplicates. Write the data to the Redshift tables.
- D. Use the AWS Glue ResolveChoice built-in transform to select the value of the column from the most recent record.

Answer: A

Explanation:

To avoid duplicate records in Amazon Redshift, the most effective solution is to perform the ETL in a way that first loads the data into a staging table and then uses SQL commands like MERGE or UPDATE to insert new records and update existing records without introducing duplicates.

* Using Staging Tables in Redshift:

* The AWS Glue job can write data to a staging table in Redshift. Once the data is loaded, SQL commands can be executed to compare the staging data with the target table and update or insert records appropriately. This ensures no duplicates are introduced during re-runs of the Glue job.

NEW QUESTION # 172

A company saves customer data to an Amazon S3 bucket. The company uses server-side encryption with AWS KMS keys (SSE-KMS) to encrypt the bucket. The dataset includes personally identifiable information (PII) such as social security numbers and account details.

Data that is tagged as PII must be masked before the company uses customer data for analysis. Some users must have secure access to the PII data during the preprocessing phase. The company needs a low-maintenance solution to mask and secure the PII data throughout the entire engineering pipeline.

Which combination of solutions will meet these requirements? (Select TWO.)

- A. Configure an Amazon Macie discovery job for the S3 bucket.
- B. Use Amazon GuardDuty to monitor access patterns for the PII data that is used in the engineering pipeline.
- C. Use AWS Identity and Access Management (IAM) to manage permissions and to control access to the PII data.
- D. Write custom scripts in an application to mask the PII data and to control access.
- E. Use AWS Glue DataBrew to perform extract, transform, and load (ETL) tasks that mask the PII data before analysis.

Answer: C,E

Explanation:

To address the requirement of masking PII data and ensuring secure access throughout the data pipeline, the combination of AWS Glue DataBrew and IAM provides a low-maintenance solution.

A . AWS Glue DataBrew for Masking:

AWS Glue DataBrew provides a visual tool to perform data transformations, including masking PII data. It allows for easy configuration of data transformation tasks without requiring manual coding, making it ideal for this use case.

Reference:

D . AWS Identity and Access Management (IAM):

Using IAM policies allows fine-grained control over access to PII data, ensuring that only authorized users can view or process sensitive data during the pipeline stages.

Alternatives Considered:

B (Amazon GuardDuty): GuardDuty is for threat detection and does not handle data masking or access control for PII.

C (Amazon Macie): Macie can help discover sensitive data but does not handle the masking of PII or access control.

E (Custom scripts): Custom scripting increases the operational burden compared to a built-in solution like DataBrew.

AWS Glue DataBrew for Data Masking

IAM Policies for PII Access Control

NEW QUESTION # 173

A company is migrating a legacy application to an Amazon S3 based data lake. A data engineer reviewed data that is associated with the legacy application. The data engineer found that the legacy data contained some duplicate information.

The data engineer must identify and remove duplicate information from the legacy application data.

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

- A. Write an AWS Glue extract, transform, and load (ETL) job. Import the Python dedupe library. Use the dedupe library to perform data deduplication.
- B. Write a custom extract, transform, and load (ETL) job in Python. Import the Python dedupe library. Use the dedupe library to perform data deduplication.
- C. Write an AWS Glue extract, transform, and load (ETL) job. Use the FindMatches machine learning (ML) transform to transform the data to perform data deduplication.
- D. Write a custom extract, transform, and load (ETL) job in Python. Use the DataFrame.drop_duplicates() function by importing the Pandas library to perform data deduplication.

Answer: C

Explanation:

AWS Glue is a fully managed serverless ETL service that can handle data deduplication with minimal operational overhead. AWS Glue provides a built-in ML transform called FindMatches, which can automatically identify and group similar records in a dataset. FindMatches can also generate a primary key for each group of records and remove duplicates. FindMatches does not require any coding or prior ML experience, as it can learn from a sample of labeled data provided by the user. FindMatches can also scale to handle large datasets and optimize the cost and performance of the ETL job. References:

AWS Glue

FindMatches ML Transform

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

NEW QUESTION # 174

A company uses an Amazon S3 bucket to integrate multiple data sources into a central data lake. The company needs to perform multiple transformations and data cleaning processes on the data to make the data accessible to business partners.

The company needs a solution that will give multiple business partners the ability to run SQL queries on the central data lake during normal business hours.

Which solution will meet these requirements MOST cost-effectively?

- A. Use an AWS Lambda function after normal business hours to process the previous day's data, apply all necessary transformations, and load the prepared data into an Amazon Redshift provisioned cluster.
- B. Use an AWS Glue Flex job after normal business hours to process the previous day's data, apply all necessary transformations, and load the prepared data into Amazon Redshift Serverless.
- C. Use an AWS Glue Flex job after normal business hours to process the previous day's data, apply all necessary transformations, and load the prepared data into an Amazon Redshift provisioned cluster.
- D. Use a provisioned Amazon EMR cluster after normal business hours to process the previous day's data, apply all necessary transformations, and load the prepared data into Amazon Redshift Serverless.

Answer: B

Explanation:

Option B is most cost-effective because it combines a serverless ETL service for overnight processing with a serverless warehouse for partner SQL access during business hours. The material describes AWS Glue as a "fully managed, serverless, scalable ETL service" that can extract, transform, and load data with minimal operational overhead, which fits the requirement for multiple transformations and data cleaning.

For the SQL access layer, the study material highlights the cost advantage of Amazon Redshift Serverless: it automatically provisions and scales capacity and you pay only for the compute capacity provisioned, with no compute costs when no workloads are running. That directly supports "multiple business partners" querying during business hours while avoiding the always-on cost of a provisioned cluster.

Options C and D use a provisioned Redshift cluster, which is less cost-effective because capacity must be paid for even when partners are not querying. Option A adds EMR cluster management cost and operational overhead compared to using Glue for ETL.

NEW QUESTION # 175

A company has three subsidiaries. Each subsidiary uses a different data warehousing solution. The first subsidiary hosts its data warehouse in Amazon Redshift. The second subsidiary uses Teradata Vantage on AWS. The third subsidiary uses Google BigQuery.

The company wants to aggregate all the data into a central Amazon S3 data lake. The company wants to use Apache Iceberg as the table format.

A data engineer needs to build a new pipeline to connect to all the data sources, run transformations by using each source engine, join the data, and write the data to Iceberg.

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

- A. Use the Amazon Athena federated query connectors for Amazon Redshift, Teradata, and BigQuery to build the pipeline in Athena. Write a SQL query to read from all the data sources, join the data, and run a Merge operation on the data lake Iceberg table.
- B. Use native Amazon Redshift, Teradata, and BigQuery connectors to build the pipeline in AWS Glue. Use native AWS Glue transforms to join the data. Run a Merge operation on the data lake Iceberg table.
- C. Use the native Amazon Redshift, Teradata, and BigQuery connectors in Amazon Appflow to write data to Amazon S3 and AWS Glue Data Catalog. Use Amazon Athena to join the data. Run a Merge operation on the data lake Iceberg table.
- D. Use the native Amazon Redshift connector, the Java Database Connectivity (JDBC) connector for Teradata, and the open source Apache Spark BigQuery connector to build the pipeline in Amazon EMR. Write code in PySpark to join the data. Run a Merge operation on the data lake Iceberg table.

Answer: A

Explanation:

Amazon Athena provides federated query connectors that allow querying multiple data sources, such as Amazon Redshift, Teradata, and Google BigQuery, without needing to extract the data from the original source. This solution is optimal because it offers the least operational effort by avoiding complex data movement and transformation processes.

* Amazon Athena Federated Queries:

* Athena's federated queries allow direct querying of data stored across multiple sources, including Amazon Redshift, Teradata, and BigQuery. With Athena's support for Apache Iceberg, the company can easily run a Merge operation on the Iceberg table.

* The solution reduces complexity by centralizing the query execution and transformation process in Athena using SQL queries.

Reference: Amazon Athena Federated Query

Alternatives Considered:

A (AWS Glue pipeline): This would work but requires more operational effort to manage and transform the data in AWS Glue.

C (Amazon EMR): Using EMR and writing PySpark code introduces more operational overhead and complexity compared to a SQL-based solution in Athena.

D (Amazon AppFlow): AppFlow is more suitable for transferring data between services but is not as efficient for transformations and joins as Athena federated queries.

NEW QUESTION # 176

Don't miss practicing the Data-Engineer-Associate mock exams and score yourself honestly. You have all the time to try Amazon Data-Engineer-Associate practice exams and then be confident while appearing for the final turn. The desktop software works on Windows and the web-based format works on all operating systems. With PDF questions, you can prepare for the Data-Engineer-Associate Certification Exam while sitting back at our place.

- [illegible]

myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt,
myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt,
myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, Disposable vapes

What's more, part of that VerifiedDumps Data-Engineer-Associate dumps now are free: https://drive.google.com/open?id=1QId1j4cU45fDt4A__8RogCJyontn3K4j