

Guaranteed Data-Engineer-Associate Success & Data-Engineer-Associate Actualtest



P.S. Free 2026 Amazon Data-Engineer-Associate dumps are available on Google Drive shared by PassSureExam:
https://drive.google.com/open?id=1_rJOYxz-x8-SQFcp2cHxjgBkp1kHrCX7

When you decide to pass the Amazon Data-Engineer-Associate exam and get relate certification, you must want to find a reliable exam tool to prepare for exam. That is the reason why I want to recommend our AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate Prep Guide to you, because we believe this is what you have been looking for.

You should not register for the Amazon AWS Certified Data Engineer - Associate (DEA-C01) certification exam without proper preparation. Passing the AWS Certified Data Engineer - Associate (DEA-C01) exam is quite a challenging task. This difficult task becomes easier if you use valid Amazon Data-Engineer-Associate Exam Dumps of PassSureExam. Don't forget that the AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) test registration fee is hefty and your money will go to waste if you don't crack this exam.

>> Guaranteed Data-Engineer-Associate Success <<

Data-Engineer-Associate Actualtest & Authentic Data-Engineer-Associate Exam Questions

The questions of our Data-Engineer-Associate guide questions are related to the latest and basic knowledge. What's more, our Data-Engineer-Associate learning materials are committed to grasp the most knowledgeable points with the fewest problems. So 20-30 hours of study is enough for you to deal with the exam. When you get a Data-Engineer-Associate certificate, you will be more competitive than others, so you can get a promotion and your wages will also rise your future will be controlled by yourselves.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q205-Q210):

NEW QUESTION # 205

A company uses AWS Step Functions to orchestrate a data pipeline. The pipeline consists of Amazon EMR jobs that ingest data from data sources and store the data in an Amazon S3 bucket. The pipeline also includes EMR jobs that load the data to Amazon Redshift.

The company's cloud infrastructure team manually built a Step Functions state machine. The cloud infrastructure team launched an EMR cluster into a VPC to support the EMR jobs. However, the deployed Step Functions state machine is not able to run the EMR jobs.

Which combination of steps should the company take to identify the reason the Step Functions state machine is not able to run the EMR jobs? (Choose two.)

- A. Check for entries in Amazon CloudWatch for the newly created EMR cluster. Change the AWS Step Functions state machine code to use Amazon EMR on EKS. Change the IAM access policies and the security group configuration for the Step Functions state machine code to reflect inclusion of Amazon Elastic Kubernetes Service (Amazon EKS).
- B. Query the flow logs for the VPC. Determine whether the traffic that originates from the EMR cluster can successfully reach the data providers. Determine whether any security group that might be attached to the Amazon EMR cluster allows connections to the data source servers on the informed ports.
- C. Verify that the Step Functions state machine code has all IAM permissions that are necessary to create and run the EMR jobs. Verify that the Step Functions state machine code also includes IAM permissions to access the Amazon S3 buckets that the EMR jobs use. Use Access Analyzer for S3 to check the S3 access properties.
- D. Use AWS CloudFormation to automate the Step Functions state machine deployment. Create a step to pause the state machine during the EMR jobs that fail. Configure the step to wait for a human user to send approval through an email message. Include details of the EMR task in the email message for further analysis.
- E. Check the retry scenarios that the company configured for the EMR jobs. Increase the number of seconds in the interval between each EMR task. Validate that each fallback state has the appropriate catch for each decision state. Configure an Amazon Simple Notification Service (Amazon SNS) topic to store the error messages.

Answer: B,C

Explanation:

To identify the reason why the Step Functions state machine is not able to run the EMR jobs, the company should take the following steps:

Verify that the Step Functions state machine code has all IAM permissions that are necessary to create and run the EMR jobs. The state machine code should have an IAM role that allows it to invoke the EMR APIs, such as `RunJobFlow`, `AddJobFlowSteps`, and `DescribeStep`. The state machine code should also have IAM permissions to access the Amazon S3 buckets that the EMR jobs use as input and output locations. The company can use Access Analyzer for S3 to check the access policies and permissions of the S3 buckets. Therefore, option B is correct.

Query the flow logs for the VPC. The flow logs can provide information about the network traffic to and from the EMR cluster that is launched in the VPC. The company can use the flow logs to determine whether the traffic that originates from the EMR cluster can successfully reach the data providers, such as Amazon RDS, Amazon Redshift, or other external sources. The company can also determine whether any security group that might be attached to the EMR cluster allows connections to the data source servers on the informed ports. The company can use Amazon VPC Flow Logs or Amazon CloudWatch Logs Insights to query the flow logs. Therefore, option D is correct.

Option A is incorrect because it suggests using AWS CloudFormation to automate the Step Functions state machine deployment. While this is a good practice to ensure consistency and repeatability of the deployment, it does not help to identify the reason why the state machine is not able to run the EMR jobs. Moreover, creating a step to pause the state machine during the EMR jobs that fail and wait for a human user to send approval through an email message is not a reliable way to troubleshoot the issue. The company should use the Step Functions console or API to monitor the execution history and status of the state machine, and use Amazon CloudWatch to view the logs and metrics of the EMR jobs.

Option C is incorrect because it suggests changing the AWS Step Functions state machine code to use Amazon EMR on EKS. Amazon EMR on EKS is a service that allows you to run EMR jobs on Amazon Elastic Kubernetes Service (Amazon EKS) clusters. While this service has some benefits, such as lower cost and faster execution time, it does not support all the features and integrations that EMR on EC2 does, such as EMR Notebooks, EMR Studio, and EMRFS. Therefore, changing the state machine code to use EMR on EKS may not be compatible with the existing data pipeline and may introduce new issues.

Option E is incorrect because it suggests checking the retry scenarios that the company configured for the EMR jobs. While this is a good practice to handle transient failures and errors, it does not help to identify the root cause of why the state machine is not able to run the EMR jobs. Moreover, increasing the number of seconds in the interval between each EMR task may not improve the success rate of the jobs, and may increase the execution time and cost of the state machine. Configuring an Amazon SNS topic to store the error messages may help to notify the company of any failures, but it does not provide enough information to troubleshoot

the issue.

Reference:

- 1: Manage an Amazon EMR Job - AWS Step Functions
- 2: Access Analyzer for S3 - Amazon Simple Storage Service
- 3: Working with Amazon EMR and VPC Flow Logs - Amazon EMR
- [4]: Analyzing VPC Flow Logs with Amazon CloudWatch Logs Insights - Amazon Virtual Private Cloud
- [5]: Monitor AWS Step Functions - AWS Step Functions
- [6]: Monitor Amazon EMR clusters - Amazon EMR
- [7]: Amazon EMR on Amazon EKS - Amazon EMR

NEW QUESTION # 206

A company wants to analyze sales records that the company stores in a MySQL database. The company wants to correlate the records with sales opportunities identified by Salesforce.

The company receives 2 GB of sales records every day. The company has 100 GB of identified sales opportunities. A data engineer needs to develop a process that will analyze and correlate sales records and sales opportunities. The process must run once each night.

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

- **A. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use AWS Glue to fetch sales records from the MySQL database. Correlate the sales records with sales opportunities. Use AWS Step Functions to orchestrate the process.**
- B. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use AWS Glue to fetch sales records from the MySQL database. Correlate the sales records with the sales opportunities. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the process.
- C. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to fetch both datasets. Use AWS Lambda functions to correlate the datasets. Use AWS Step Functions to orchestrate the process.
- D. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use Amazon Kinesis Data Streams to fetch sales records from the MySQL database. Use Amazon Managed Service for Apache Flink to correlate the datasets. Use AWS Step Functions to orchestrate the process.

Answer: A

Explanation:

Problem Analysis:

The company processes 2 GB of daily sales records and 100 GB of Salesforce sales opportunities.

The goal is to analyze and correlate the two datasets with low operational overhead.

The process must run once nightly.

Key Considerations:

Amazon AppFlow simplifies data integration with Salesforce.

AWS Glue can extract data from MySQL and perform ETL operations.

Step Functions can orchestrate workflows with minimal manual intervention.

Apache Airflow and Flink add complexity, which conflicts with the requirement for low operational overhead.

Solution Analysis:

Option A: MWAA + Lambda + Step Functions

Requires custom Lambda code for dataset correlation, increasing development and operational complexity.

Option B: AppFlow + Glue + MWAA

MWAA adds orchestration overhead compared to the simpler Step Functions.

Option C: AppFlow + Glue + Step Functions

AppFlow fetches Salesforce data, Glue extracts MySQL data, and Step Functions orchestrate the entire process.

Minimal setup and operational overhead, making it the best choice.

Option D: AppFlow + Kinesis + Flink + Step Functions

Using Kinesis and Flink for batch processing introduces unnecessary complexity.

Final Recommendation:

Use Amazon AppFlow to fetch Salesforce data, AWS Glue to process MySQL data, and Step Functions for orchestration.

Reference:

Amazon AppFlow Overview

AWS Glue ETL Documentation

AWS Step Functions

NEW QUESTION # 207

A company stores server logs in an Amazon S3 bucket. The company needs to keep the logs for 1 year. The logs are not required after 1 year.

A data engineer needs a solution to automatically delete logs that are older than 1 year.

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

- A. Schedule a cron job on an Amazon EC2 instance to delete the logs after 1 year.
- **B. Create an AWS Lambda function to delete the logs after 1 year.**
- C. Configure an AWS Step Functions state machine to delete the logs after 1 year.
- D. Define an S3 Lifecycle configuration to delete the logs after 1 year.

Answer: B

Explanation:

* Problem Analysis:

* The company uses AWS Glue for ETL pipelines and requires automatic data quality checks during pipeline execution.

* The solution must integrate with existing AWS Glue pipelines and evaluate data quality rules based on predefined thresholds.

* Key Considerations:

* Ensure minimal implementation effort by leveraging built-in AWS Glue features.

* Use a standardized approach for defining and evaluating data quality rules.

* Avoid custom libraries or external frameworks unless absolutely necessary.

* Solution Analysis:

* Option A: SQL Transform

* Adding SQL transforms to define and evaluate data quality rules is possible but requires writing complex queries for each rule.

* Increases operational overhead and deviates from Glue's declarative approach.

* Option B: Evaluate Data Quality Transform with DQDL

* AWS Glue provides a built-in Evaluate Data Quality transform.

* Allows defining rules in Data Quality Definition Language (DQDL), a concise and declarative way to define quality checks.

* Fully integrated with Glue Studio, making it the least effort solution.

* Option C: Custom Transform with PyDeequ

* PyDeequ is a powerful library for data quality checks but requires custom code and integration.

* Increases implementation effort compared to Glue's native capabilities.

* Option D: Custom Transform with Great Expectations

* Great Expectations is another powerful library for data quality but adds complexity and external dependencies.

* Final Recommendation:

* Use Evaluate Data Quality transform in AWS Glue.

* Define rules in DQDL for checking thresholds, null values, or other quality criteria.

* This approach minimizes development effort and ensures seamless integration with AWS Glue.

:

AWS Glue Data Quality Overview

DQDL Syntax and Examples

Glue Studio Transformations

NEW QUESTION # 208

A company uses an organization in AWS Organizations to manage multiple AWS accounts. The company uses an enhanced fanout data stream in Amazon Kinesis Data Streams to receive streaming data from multiple producers. The data stream runs in Account A. The company wants to use an AWS Lambda function in Account B to process the data from the stream. The company creates a Lambda execution role in Account B that has permissions to access data from the stream in Account A.

What additional step must the company take to meet this requirement?

- A. Create a service control policy (SCP) to grant the data stream read access to the cross-account Lambda execution role. Attach the SCP to Account A.
- B. Create a service control policy (SCP) to grant the data stream read access to the cross-account Lambda execution role. Attach the SCP to Account B.
- **C. Add a resource-based policy to the data stream to allow read access for the cross-account Lambda execution role.**
- D. Add a resource-based policy to the cross-account Lambda function to grant the data stream read access to the function.

Answer: C

Explanation:

To allow cross-account access to a Kinesis Data Stream, you must add a resource-based policy to the Kinesis stream in Account A, explicitly granting the Lambda execution role in Account B the required permissions.

* SCPs (A & C) set permissions boundaries, but do not grant access.

* Option D incorrectly refers to the Lambda function - but the Kinesis resource must allow access.

"You must add a resource-based policy to the Kinesis Data Stream in Account A to allow a Lambda function in Account B to consume from the stream." Reference: AWS Documentation - Cross-account Lambda access to Kinesis

NEW QUESTION # 209

A data engineer is building an automated extract, transform, and load (ETL) ingestion pipeline by using AWS Glue. The pipeline ingests compressed files that are in an Amazon S3 bucket. The ingestion pipeline must support incremental data processing. Which AWS Glue feature should the data engineer use to meet this requirement?

- A. Triggers
- B. Classifiers
- C. Workflows
- **D. Job bookmarks**

Answer: D

Explanation:

Problem Analysis:

The pipeline processes compressed files in S3 and must support incremental data processing.

AWS Glue features must facilitate tracking progress to avoid reprocessing the same data.

Key Considerations:

Incremental data processing requires tracking which files or partitions have already been processed.

The solution must be automated and efficient for large-scale ETL jobs.

Solution Analysis:

Option A: Workflows

Workflows organize and orchestrate multiple Glue jobs but do not track progress for incremental data processing.

Option B: Triggers

Triggers initiate Glue jobs based on a schedule or events but do not track which data has been processed.

Option C: Job Bookmarks

Job bookmarks track the state of the data that has been processed, enabling incremental processing.

Automatically skip files or partitions that were previously processed in Glue jobs.

Option D: Classifiers

Classifiers determine the schema of incoming data but do not handle incremental processing.

Final Recommendation:

Job bookmarks are specifically designed to enable incremental data processing in AWS Glue ETL pipelines.

AWS Glue Job Bookmarks Documentation

AWS Glue ETL Features

NEW QUESTION # 210

.....

Perhaps you plan to seek a high salary job. But you are not confident enough because of lack of ability. Now, our Data-Engineer-Associate practice guide is able to give you help. You will quickly master all practical knowledge in the shortest time. Also, obtaining the Data-Engineer-Associate certificate fully has no problem. With the high pass rate of our Data-Engineer-Associate exam braindumps as 98% to 100%, we can claim that as long as you study with our Data-Engineer-Associate study materials, you will pass the exam for sure.

Data-Engineer-Associate Actualtest: <https://www.passsureexam.com/Data-Engineer-Associate-pass4sure-exam-dumps.html>

- If you want to Test Quality of the PassSureExam Amazon Data-Engineer-Associate product, When you decide to buy the Data-Engineer-Associate Actualtest - AWS Certified Data Engineer - Associate (DEA-C01) exam dumps, you may still have some doubts and confusion, Amazon Guaranteed Data-Engineer-Associate Success We can totally be trusted, Once you purchase our valid Data-Engineer-Associate dumps torrent, you will not only share high-quality & high pass-rate exam dumps but also rich customer service so that you can clear your exam surely, Amazon Guaranteed Data-Engineer-Associate Success I would like to express my sincere gratitude to you if you can pay attention to my statements for a little while.

Verified Amazon Guaranteed Data-Engineer-Associate Success With Interarctive Test Engine & Efficient Data-Engineer-Associate Actualtest

We can totally be trusted, Once you purchase our valid Data-Engineer-Associate dumps torrent, you will not only share high-quality & high pass-rate exam dumps but also rich customer service so that you can clear your exam surely.

[illegible]

What's more, part of that PassSureExam Data-Engineer-Associate dumps now are free: https://drive.google.com/open?id=1_rJOYxz-x8-SOFcp2cHxigBkp1kHrCX7

