

Valid Data-Engineer-Associate Free Sample & Useful Materials to help you pass Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01)

DATA ENGINEER RESUME

Example by Resume Genius

Memphis, TN • your.email@email.com • (901) 622-7892

Detail-oriented Senior Data Engineer with 6+ years of experience designing, developing, and implementing data solutions for large-scale businesses. Skilled in SQL, Python, Apache Spark, and Hadoop. Proven track record of improving data accuracy and reducing processing time through innovative data warehousing solutions.

Professional Experience

Senior Data Engineer

AutomotiveMastermind, Memphis, TN

March 20XX–Present

- Work with cross-functional teams to devise data security protocols, ensuring sensitive data is protected at all times
- Developed a predictive analytics model using machine learning algorithms, improving customer satisfaction by 12%
- Designed a data warehousing solution that processed 100TB of data daily, reducing processing time by 23% and improving overall data accuracy by 35%
- Develop and maintain a scalable data architecture, supporting the organization's long-term data needs

Data Engineer

FDM Group, Boston, MA

July 20XX–February 20XX

- Automated ETL processes to streamline data transfer between systems, reducing manual effort by 45%
- Implemented a data visualization solution using Tableau, helping business users to better understand and analyze data
- Evaluated and improved data governance policies and procedures
- Spearheaded the implementation of a real-time data streaming solution using Apache Spark, improving data processing speed by 25%

Education

Northeastern University, Boston, MA

May 20XX

Bachelor's Degree in Computer Science

Honors: cum laude (GPA: 3.8/4.0)

Additional Skills

- SQL
- Data visualization
- Cloud computing
- Analytics techniques
- Data modeling

What's more, part of that DumpsTorrent Data-Engineer-Associate dumps now are free: <https://drive.google.com/open?id=1hnuP9KvK6Q-YdKXh2xaEFhuNIIdR7LzJY>

Our Data-Engineer-Associate study prep is classified as three versions up to now. All these versions of our Data-Engineer-Associate exam braindumps are popular and priced cheap with high quality and accuracy rate. They achieved academic maturity so that their quality far beyond other practice materials in the market with high effectiveness and more than 98 percent of former candidates who chose our Data-Engineer-Associate Practice Engine win the exam with their dream certificate.

In the process of using the AWS Certified Data Engineer - Associate (DEA-C01) study training dumps, once users have any questions about our study materials, the user can directly by E-mail us, our products have a dedicated customer service staff to answer for the user, they are 24 hours service for you, we are very welcome to contact us by E-mail and put forward valuable opinion for us. Our Data-Engineer-Associate latest questions already have many different kinds of learning materials, users may be confused about the choice, what is the most suitable Data-Engineer-Associate Test Guide? Believe that users will get the most satisfactory answer after consultation. Our online service staff is professionally trained, and users' needs about Data-Engineer-Associate test guide can be clearly understood by them. The most complete online service of our company will be answered by you, whether it is before the product purchase or the product installation process, or after using the Data-Engineer-Associate latest questions, no matter what problem the user has encountered.

>> Data-Engineer-Associate Free Sample <<

Data-Engineer-Associate Reliable Exam Price & Valuable Data-Engineer-Associate Feedback

To attain this you just need to enroll in the Amazon Data-Engineer-Associate certification exam and put all your efforts to pass this challenging Amazon Data-Engineer-Associate exam with good scores. However, to get success in Data-Engineer-Associate dumps PDF is not an easy task, it is quite difficult to pass it. But with proper planning, firm commitment, and Data-Engineer-Associate Exam Questions, you can pass this milestone easily. The DumpsTorrent is a leading platform that offers real, valid, and updated Data-Engineer-Associate Dumps.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q152-Q157):

NEW QUESTION # 152

A company stores data from an application in an Amazon DynamoDB table that operates in provisioned capacity mode. The workloads of the application have predictable throughput load on a regular schedule. Every Monday, there is an immediate increase in activity early in the morning. The application has very low usage during weekends. The company must ensure that the application performs consistently during peak usage times. Which solution will meet these requirements in the MOST cost-effective way?

- A. Increase the provisioned capacity to the maximum capacity that is currently present during peak load times.
- B. Divide the table into two tables. Provision each table with half of the provisioned capacity of the original table. Spread queries evenly across both tables.
- C. Use AWS Application Auto Scaling to schedule higher provisioned capacity for peak usage times. Schedule lower capacity during off-peak times.
- D. Change the capacity mode from provisioned to on-demand. Configure the table to scale up and scale down based on the load on the table.

Answer: C

Explanation:

Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB offers two capacity modes for throughput capacity: provisioned and on-demand. In provisioned capacity mode, you specify the number of read and write capacity units per second that you expect your application to require. DynamoDB reserves the resources to meet your throughput needs with consistent performance. In on-demand capacity mode, you pay per request and DynamoDB scales the resources up and down automatically based on the actual workload. On-demand capacity mode is suitable for unpredictable workloads that can vary significantly over time¹.

The solution that meets the requirements in the most cost-effective way is to use AWS Application Auto Scaling to schedule higher provisioned capacity for peak usage times and lower capacity during off-peak times. This solution has the following advantages:

- * It allows you to optimize the cost and performance of your DynamoDB table by adjusting the provisioned capacity according to your predictable workload patterns. You can use scheduled scaling to specify the date and time for the scaling actions, and the new minimum and maximum capacity limits. For example, you can schedule higher capacity for every Monday morning and lower capacity for weekends².
- * It enables you to take advantage of the lower cost per unit of provisioned capacity mode compared to on-demand capacity mode. Provisioned capacity mode charges a flat hourly rate for the capacity you reserve, regardless of how much you use. On-demand capacity mode charges for each read and write request you consume, with no minimum capacity required. For predictable workloads, provisioned capacity mode can be more cost-effective than on-demand capacity mode¹.
- * It ensures that your application performs consistently during peak usage times by having enough capacity to handle the increased load. You can also use auto scaling to automatically adjust the provisioned capacity based on the actual utilization of your table, and set a target utilization percentage for your table or global secondary index. This way, you can avoid under-provisioning or over-provisioning your table².

Option A is incorrect because it suggests increasing the provisioned capacity to the maximum capacity that is currently present during peak load times. This solution has the following disadvantages:

- * It wastes money by paying for unused capacity during off-peak times. If you provision the same high capacity for all times, regardless of the actual workload, you are over-provisioning your table and paying for resources that you don't need¹.
- * It does not account for possible changes in the workload patterns over time. If your peak load times increase or decrease in the future, you may need to manually adjust the provisioned capacity to match the new demand. This adds operational overhead and complexity to your application².

Option B is incorrect because it suggests dividing the table into two tables and provisioning each table with half of the provisioned capacity of the original table. This solution has the following disadvantages:

* It complicates the data model and the application logic by splitting the data into two separate tables.

You need to ensure that the queries are evenly distributed across both tables, and that the data is consistent and synchronized between them. This adds extra development and maintenance effort to your application³.

* It does not solve the problem of adjusting the provisioned capacity according to the workload patterns.

You still need to manually or automatically scale the capacity of each table based on the actual utilization and demand. This may result in under-provisioning or over-provisioning your tables².

Option D is incorrect because it suggests changing the capacity mode from provisioned to on-demand. This solution has the following disadvantages:

* It may incur higher costs than provisioned capacity mode for predictable workloads. On-demand capacity mode charges for each read and write request you consume, with no minimum capacity required. For predictable workloads, provisioned capacity mode can be more cost-effective than on-demand capacity mode, as you can reserve the capacity you need at a lower rate¹.

* It may not provide consistent performance during peak usage times, as on-demand capacity mode may take some time to scale up the resources to meet the sudden increase in demand. On-demand capacity mode uses adaptive capacity to handle bursts of traffic, but it may not be able to handle very large spikes or sustained high throughput. In such cases, you may experience throttling or increased latency.

References:

* 1: Choosing the right DynamoDB capacity mode - Amazon DynamoDB

* 2: Managing throughput capacity automatically with DynamoDB auto scaling - Amazon DynamoDB

* 3: Best practices for designing and using partition keys effectively - Amazon DynamoDB

* [4]: On-demand mode guidelines - Amazon DynamoDB

* [5]: How to optimize Amazon DynamoDB costs - AWS Database Blog

* [6]: DynamoDB adaptive capacity: How it works and how it helps - AWS Database Blog

* [7]: Amazon DynamoDB pricing - Amazon Web Services (AWS)

NEW QUESTION # 153

A media company wants to improve a system that recommends media content to customer based on user behavior and preferences. To improve the recommendation system, the company needs to incorporate insights from third-party datasets into the company's existing analytics platform.

The company wants to minimize the effort and time required to incorporate third-party datasets.

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

- A. Use Amazon Kinesis Data Streams to access and integrate third-party datasets from AWS CodeCommit repositories.
- B. Use API calls to access and integrate third-party datasets from AWS
- C. Use Amazon Kinesis Data Streams to access and integrate third-party datasets from Amazon Elastic Container Registry (Amazon ECR).
- **D. Use API calls to access and integrate third-party datasets from AWS Data Exchange.**

Answer: D

Explanation:

AWS Data Exchange is a service that makes it easy to find, subscribe to, and use third-party data in the cloud.

It provides a secure and reliable way to access and integrate data from various sources, such as data providers, public datasets, or AWS services. Using AWS Data Exchange, you can browse and subscribe to data products that suit your needs, and then use API calls or the AWS Management Console to export the data to Amazon S3, where you can use it with your existing analytics platform. This solution minimizes the effort and time required to incorporate third-party datasets, as you do not need to set up and manage data pipelines, storage, or access controls. You also benefit from the data quality and freshness provided by the data providers, who can update their data products as frequently as needed^{1,2}.

The other options are not optimal for the following reasons:

* B. Use API calls to access and integrate third-party datasets from AWS. This option is vague and does not specify which AWS service or feature is used to access and integrate third-party datasets. AWS offers a variety of services and features that can help with data ingestion, processing, and analysis, but not all of them are suitable for the given scenario. For example, AWS Glue is a serverless data integration service that can help you discover, prepare, and combine data from various sources, but it requires you to create and run data extraction, transformation, and loading (ETL) jobs, which can add operational overhead³.

* C. Use Amazon Kinesis Data Streams to access and integrate third-party datasets from AWS CodeCommit repositories. This option is not feasible, as AWS CodeCommit is a source control service that hosts secure Git-based repositories, not a data source that can be accessed by Amazon Kinesis Data Streams. Amazon Kinesis Data Streams is a service that enables you to capture, process, and analyze data streams in real time, such as clickstream data, application logs, or IoT telemetry. It does not support accessing and integrating data from AWS CodeCommit repositories, which are meant for storing and managing code, not data.

* D. Use Amazon Kinesis Data Streams to access and integrate third-party datasets from Amazon Elastic Container Registry (Amazon ECR). This option is also not feasible, as Amazon ECR is a fully managed container registry service that stores, manages,

and deploys container images, not a data source that can be accessed by Amazon Kinesis Data Streams. Amazon Kinesis Data Streams does not support accessing and integrating data from Amazon ECR, which is meant for storing and managing container images, not data.

References:

- * 1: AWS Data Exchange User Guide
- * 2: AWS Data Exchange FAQs
- * 3: AWS Glue Developer Guide
- * : AWS CodeCommit User Guide
- * : Amazon Kinesis Data Streams Developer Guide
- * : Amazon Elastic Container Registry User Guide
- * : Build a Continuous Delivery Pipeline for Your Container Images with Amazon ECR as Source

NEW QUESTION # 154

A data engineer is processing a large amount of log data from web servers. The data is stored in an Amazon S3 bucket. The data engineer uses AWS services to process the data every day. The data engineer needs to extract specific fields from the raw log data and load the data into a data warehouse for analysis.

- **A. Use an AWS Glue crawler to parse the raw log data in the S3 bucket and to generate a schema. Use AWS Glue ETL jobs to extract and transform the data and to load it into Amazon Redshift.**
- B. Use AWS Step Functions to orchestrate a series of AWS Batch jobs to parse the raw log files. Load the specified fields into an Amazon RDS for PostgreSQL database.
- C. Use Amazon EMR to run Apache Hive queries on the raw log files in the S3 bucket to extract the specified fields. Store the output as ORC files in the original S3 bucket.
- D. Use AWS Glue DataBrew to run AWS Glue ETL jobs on a schedule to extract the specified fields from the raw log files in the S3 bucket. Load the data into partitioned tables in Amazon Redshift.

Answer: A

NEW QUESTION # 155

A company uses Amazon S3 to store semi-structured data in a transactional data lake. Some of the data files are small, but other data files are tens of terabytes.

A data engineer must perform a change data capture (CDC) operation to identify changed data from the data source. The data source sends a full snapshot as a JSON file every day and ingests the changed data into the data lake.

Which solution will capture the changed data MOST cost-effectively?

- A. Create an AWS Lambda function to identify the changes between the previous data and the current data. Configure the Lambda function to ingest the changes into the data lake.
- **B. Use an open source data lake format to merge the data source with the S3 data lake to insert the new data and update the existing data.**
- C. Ingest the data into Amazon RDS for MySQL. Use AWS Database Migration Service (AWS DMS) to write the changed data to the data lake.
- D. Ingest the data into an Amazon Aurora MySQL DB instance that runs Aurora Serverless. Use AWS Database Migration Service (AWS DMS) to write the changed data to the data lake.

Answer: B

Explanation:

An open source data lake format, such as Apache Parquet, Apache ORC, or Delta Lake, is a cost-effective way to perform a change data capture (CDC) operation on semi-structured data stored in Amazon S3. An open source data lake format allows you to query data directly from S3 using standard SQL, without the need to move or copy data to another service. An open source data lake format also supports schema evolution, meaning it can handle changes in the data structure over time. An open source data lake format also supports upserts, meaning it can insert new data and update existing data in the same operation, using a merge command. This way, you can efficiently capture the changes from the data source and apply them to the S3 data lake, without duplicating or losing any data.

The other options are not as cost-effective as using an open source data lake format, as they involve additional steps or costs.

Option A requires you to create and maintain an AWS Lambda function, which can be complex and error-prone. AWS Lambda also has some limits on the execution time, memory, and concurrency, which can affect the performance and reliability of the CDC operation. Option B and D require you to ingest the data into a relational database service, such as Amazon RDS or Amazon Aurora, which can be expensive and unnecessary for semi-structured data. AWS Database Migration Service (AWS DMS) can

write the changed data to the data lake, but it also charges you for the data replication and transfer. Additionally, AWS DMS does not support JSON as a source data type, so you would need to convert the data to a supported format before using AWS DMS. References:

What is a data lake?

Choosing a data format for your data lake

Using the MERGE INTO command in Delta Lake

[AWS Lambda quotas]

[AWS Database Migration Service quotas]

NEW QUESTION # 156

A gaming company uses Amazon Kinesis Data Streams to collect clickstream data. The company uses Amazon Kinesis Data Firehose delivery streams to store the data in JSON format in Amazon S3. Data scientists at the company use Amazon Athena to query the most recent data to obtain business insights.

The company wants to reduce Athena costs but does not want to recreate the data pipeline.

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

- A. Change the Firehose output format to Apache Parquet. Provide a custom S3 object YYYYMMDD prefix expression and specify a large buffer size. For the existing data, create an AWS Glue extract, transform, and load (ETL) job. Configure the ETL job to combine small JSON files, convert the JSON files to large Parquet files, and add the YYYYMMDD prefix. Use the ALTER TABLE ADD PARTITION statement to reflect the partition on the existing Athena table.
- B. Integrate an AWS Lambda function with Firehose to convert source records to Apache Parquet and write them to Amazon S3. In parallel, run an AWS Glue extract, transform, and load (ETL) job to combine the JSON files and convert the JSON files to large Parquet files. Create a custom S3 object YYYYMMDD prefix. Use the ALTER TABLE ADD PARTITION statement to reflect the partition on the existing Athena table.
- C. Create an Apache Spark job that combines JSON files and converts the JSON files to Apache Parquet files. Launch an Amazon EMR ephemeral cluster every day to run the Spark job to create new Parquet files in a different S3 location. Use the ALTER TABLE SET LOCATION statement to reflect the new S3 location on the existing Athena table.
- D. Create a Kinesis data stream as a delivery destination for Firehose. Use Amazon Managed Service for Apache Flink (previously known as Amazon Kinesis Data Analytics) to run Apache Flink on the Kinesis data stream. Use Flink to aggregate the data and save the data to Amazon S3 in Apache Parquet format with a custom S3 object YYYYMMDD prefix. Use the ALTER TABLE ADD PARTITION statement to reflect the partition on the existing Athena table.

Answer: A

Explanation:

Step 1: Understanding the Problem

The company collects clickstream data via Amazon Kinesis Data Streams and stores it in JSON format in Amazon S3 using Kinesis Data Firehose. They use Amazon Athena to query the data, but they want to reduce Athena costs while maintaining the same data pipeline.

Since Athena charges based on the amount of data scanned during queries, reducing the data size (by converting JSON to a more efficient format like Apache Parquet) is a key solution to lowering costs.

Step 2: Why Option A is Correct

* Option A provides a straightforward way to reduce costs with minimal management overhead:

* Changing the Firehose output format to Parquet: Parquet is a columnar data format, which is more compact and efficient than JSON for Athena queries. It significantly reduces the amount of data scanned, which in turn reduces Athena query costs.

* Custom S3 Object Prefix (YYYYMMDD): Adding a date-based prefix helps in partitioning the data, which further improves query efficiency in Athena by limiting the data scanned to only relevant partitions.

* AWS Glue ETL Job for Existing Data: To handle existing data stored in JSON format, a one-time AWS Glue ETL job can combine small JSON files, convert them to Parquet, and apply the YYYYMMDD prefix. This ensures consistency in the S3 bucket structure and allows Athena to efficiently query historical data.

* ALTER TABLE ADD PARTITION: This command updates Athena's table metadata to reflect the new partitions, ensuring that future queries target only the required data.

Step 3: Why Other Options Are Not Ideal

* Option B (Apache Spark on EMR) introduces higher management effort by requiring the setup of Apache Spark jobs and an Amazon EMR cluster. While it achieves the goal of converting JSON to Parquet, it involves running and maintaining an EMR cluster, which adds operational complexity.

* Option C (Kinesis and Apache Flink) is a more complex solution involving Apache Flink, which adds a real-time streaming layer to aggregate data. Although Flink is a powerful tool for stream processing, it adds unnecessary overhead in this scenario since the company already uses Kinesis Data Firehose for batch delivery to S3.

* Option D (AWS Lambda with Firehose) suggests using AWS Lambda to convert records in real time.

While Lambda can work in some cases, it's generally not the best tool for handling large-scale data transformations like JSON-to-Parquet conversion due to potential scaling and invocation limitations.

Additionally, running parallel Glue jobs further complicates the setup.

Step 4: How Option A Minimizes Costs

* By using Apache Parquet, Athena queries become more efficient, as Athena will scan significantly less data, directly reducing query costs.

* Firehose natively supports Parquet as an output format, so enabling this conversion in Firehose requires minimal effort. Once set, new data will automatically be stored in Parquet format in S3, without requiring any custom coding or ongoing management.

* The AWS Glue ETL job for historical data ensures that existing JSON files are also converted to Parquet format, ensuring consistency across the data stored in S3.

Conclusion:

Option A meets the requirement to reduce Athena costs without recreating the data pipeline, using Firehose's native support for Apache Parquet and a simple one-time AWS Glue ETL job for existing data. This approach involves minimal management effort compared to the other solutions.

NEW QUESTION # 157

.....

We provide free update and online customer service which works on the line whole day. Our Data-Engineer-Associate study materials provide varied versions of our Data-Engineer-Associate study material for you to choose and the learning costs you little time and energy. You can use our Data-Engineer-Associate exam prep immediately after you purchase them, we will send our Data-Engineer-Associate Exam Questions within 5-10 minutes to you. We treat your time as our own time, as precious as you see, so we never waste a minute or two in some useless process. Please rest assured that use, we believe that you will definitely pass the Data-Engineer-Associate exam.

Data-Engineer-Associate Reliable Exam Price: <https://www.dumpstorrent.com/Data-Engineer-Associate-exam-dumps-torrent.html>

The Amazon Data-Engineer-Associate latest study guide offers a lot of information for your exam guide, including the questions and answers, Amazon Data-Engineer-Associate Free Sample. The possibility to enter in big and famous companies is also raised because they need outstanding talents to serve for them. Customers will receive free updates to the Data-Engineer-Associate exam questions for three months if any changes are made to the AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) exam content after the purchase of the Data-Engineer-Associate practice questions. DumpsTorrent has developed the online Amazon Data-Engineer-Associate practice test to help the candidates get exposure to the actual exam environment.

Follow the path: This approach is based on the Data-Engineer-Associate Reliable Exam Price path that packets take through the network from source to destination, it had equipped me with sufficient understanding to comprehend what Data-Engineer-Associate my new colleagues were talking about and figure out what questions I should be asking.

Unparalleled Amazon Data-Engineer-Associate Free Sample | Try Free Demo before Purchase

The Amazon Data-Engineer-Associate latest study guide offers a lot of information for your exam guide, including the questions and answers, The possibility to enter in big and famous Real Data-Engineer-Associate Dumps Free companies is also raised because they need outstanding talents to serve for them.

Customers will receive free updates to the Data-Engineer-Associate exam questions for three months if any changes are made to the AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) exam content after the purchase of the Data-Engineer-Associate practice questions.

DumpsTorrent has developed the online Amazon Data-Engineer-Associate practice test to help the candidates get exposure to the actual exam environment, Amazon Data-Engineer-Associate Braindumps.

- PDF Data-Engineer-Associate VCE □ Exam Data-Engineer-Associate Pass4sure □ Data-Engineer-Associate Latest Test Answers □ Copy URL 《 www.dumpsmaterials.com 》 open and search for □ Data-Engineer-Associate □ to download for free □ Pass Data-Engineer-Associate Guide
- Amazon Data-Engineer-Associate Free Sample - AWS Certified Data Engineer - Associate (DEA-C01) Realistic Reliable Exam Price 100% Pass Quiz □ Search for “Data-Engineer-Associate ” and download it for free on 【 www.pdfvce.com 】 website □ Data-Engineer-Associate Latest Test Report
- 100% Pass Quiz Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) Fantastic Free Sample

DOWNLOAD the newest DumpsTorrent Data-Engineer-Associate PDF dumps from Cloud Storage for free:
<https://drive.google.com/open?id=1hnuP9KvK6Q-YdKXh2xaEFhuNIdR7LzJY>