

## DAS-C01 Dumps

### AWS Certified Data Analytics - Specialty

<https://www.certleader.com/DAS-C01-dumps.html>



**NEW QUESTION 1**

A company has several Amazon EC2 instances sitting behind an Application Load Balancer (ALB). The company wants its IT Infrastructure team to analyze the IP addresses coming into the company's ALB. The ALB is configured to store access logs in Amazon S3. The access logs create about 1 TB of data each day, and access to the data will be infrequent. The company needs a solution that is scalable, cost-effective and has minimal maintenance requirements. Which solution meets these requirements?

- A. Copy the data into Amazon Redshift and query the data
- B. Use Amazon EMR and Apache Hive to query the S3 data
- C. Use Amazon Athena to query the S3 data
- D. Use Amazon Redshift Spectrum to query the S3 data

**Answer: D**

**NEW QUESTION 2**

A company is reading data from various customer databases that run on Amazon RDS. The databases contain many inconsistent fields. For example, a customer record field that is `place_id` in one database is `location_id` in another database. The company wants to link customer records across different databases, even when many customer record fields do not match exactly. Which solution will meet these requirements with the LEAST operational overhead?

- A. Create an Amazon EMR cluster to process and analyze data in the databases. Connect to the Apache Zeppelin notebook, and use the FindMatches transform to find duplicate records in the data.
- B. Create an AWS Glue crawler to crawl the database.
- C. Use the FindMatches transform to find duplicate records in the data. Evaluate and tune the transform by evaluating performance and results of finding matches.
- D. Create an AWS Glue crawler to crawl the data in the databases. Use Amazon SageMaker to construct Apache Spark ML pipelines to find duplicate records in the data.
- E. Create an Amazon EMR cluster to process and analyze data in the database.
- F. Connect to the Apache Zeppelin notebook, and use Apache Spark ML to find duplicate records in the data.
- G. Evaluate and tune the model by evaluating performance and results of finding duplicates.

**Answer: B**

**NEW QUESTION 3**

An ecommerce company is migrating its business intelligence environment from on premises to the AWS Cloud. The company will use Amazon Redshift in a public subnet and Amazon QuickSight. The tables already are loaded into Amazon Redshift and can be accessed by a SQL tool.

The company starts QuickSight for the first time. During the creation of the data source, a data analytics specialist enters all the information and tries to validate the connection. An error with the following message occurs: "Creating a connection to your data source timed out."

How should the data analytics specialist resolve this error?

- A. Grant the SELECT permission on Amazon Redshift tables.
- B. Add the QuickSight IP address range into the Amazon Redshift security group.
- C. Create an IAM role for QuickSight to access Amazon Redshift.
- D. Use a QuickSight admin user for creating the dataset.

**Answer: A**

**Explanation:**

Connection to the database times out

Your client connection to the database appears to hang or time out when running long queries, such as a COPY command. In this case, you might observe that the Amazon Redshift console displays that the query has completed, but the client tool itself still appears to be running the query. The results of the query might be missing or incomplete depending on when the connection stopped.

**NEW QUESTION 4**

A data analyst is using Amazon QuickSight for data visualization across multiple datasets generated by applications. Each application stores files within a separate Amazon S3 bucket. AWS Glue Data Catalog is used as a central catalog across all application data in Amazon S3. A new application stores its data within a separate S3 bucket. After updating the catalog to include the new application data source, the data analyst created a new Amazon QuickSight data source from an Amazon Athena table, but the import into SPICE failed.

How should the data analyst resolve the issue?

- A. Edit the permissions for the AWS Glue Data Catalog from within the Amazon QuickSight console.
- B. Edit the permissions for the new S3 bucket from within the Amazon QuickSight console.
- C. Edit the permissions for the AWS Glue Data Catalog from within the AWS Glue console.
- D. Edit the permissions for the new S3 bucket from within the S3 console.

**Answer: B**

**NEW QUESTION 5**

A company wants to optimize the cost of its data and analytics platform. The company is ingesting a number of .csv and JSON files in Amazon S3 from various data sources. Incoming data is expected to be 50 GB each day. The company is using Amazon Athena to query the raw data in Amazon S3 directly. Most queries aggregate data from the past 12 months, and data that is older than 5 years is infrequently queried. The typical query scans about 500 MB of data and is expected to return results in less than 1 minute. The raw data must be retained indefinitely for compliance requirements.

Which solution meets the company's requirements?

- A. Use an AWS Glue ETL job to compress, partition, and convert the data into a columnar data format.
- B. Use Athena to query the processed dataset.
- C. Configure a lifecycle policy to move the processed data into the Amazon S3 Standard-Infrequent Access (S3 Standard-IA) storage class 5 years after object creation. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after object creation.
- D. Use an AWS Glue ETL job to partition and convert the data into a row-based data format.

- E. Use Athena to query the processed dataset
- F. Configure a lifecycle policy to move the data into the Amazon S3 Standard- Infrequent Access (S3 Standard-IA) storage class 5 years after object creation
- G. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after object creation.
- H. Use an AWS Glue ETL job to compress, partition, and convert the data into a columnar data format
- I. Use Athena to query the processed dataset
- J. Configure a lifecycle policy to move the processed data into the Amazon S3 Standard-Infrequent Access (S3 Standard-IA) storage class 5 years after the object was last accessed
- K. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after the last date the object was accessed.
- L. Use an AWS Glue ETL job to partition and convert the data into a row-based data format
- M. Use Athena to query the processed dataset
- N. Configure a lifecycle policy to move the data into the Amazon S3 Standard- Infrequent Access (S3 Standard-IA) storage class 5 years after the object was last accessed
- O. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after the last date the object was accessed.

**Answer:** A

#### NEW QUESTION 6

A company recently created a test AWS account to use for a development environment. The company also created a production AWS account in another AWS Region. As part of its security testing, the company wants to send log data from Amazon CloudWatch Logs in its production account to an Amazon Kinesis data stream in its test account.

Which solution will allow the company to accomplish this goal?

- A. Create a subscription filter in the production account's CloudWatch Logs to target the Kinesis data stream in the test account as its destination. In the test account, create an IAM role that grants access to the Kinesis data stream and the CloudWatch Logs resources in the production account.
- B. In the test account, create an IAM role that grants access to the Kinesis data stream and the CloudWatch Logs resources in the production account. Create a destination data stream in Kinesis Data Streams in the test account with an IAM role and a trust policy that allow CloudWatch Logs in the production account to write to the test account.
- C. In the test account, create an IAM role that grants access to the Kinesis data stream and the CloudWatch Logs resources in the production account. Create a destination data stream in Kinesis Data Streams in the test account with an IAM role and a trust policy that allow CloudWatch Logs in the production account to write to the test account.
- D. Create a destination data stream in Kinesis Data Streams in the test account with an IAM role and a trust policy that allow CloudWatch Logs in the production account to write to the test account. Create a subscription filter in the production account's CloudWatch Logs to target the Kinesis data stream in the test account as its destination.

**Answer:** D

#### NEW QUESTION 7

A software company hosts an application on AWS, and new features are released weekly. As part of the application testing process, a solution must be developed that analyzes logs from each Amazon EC2 instance to ensure that the application is working as expected after each deployment. The collection and analysis solution should be highly available with the ability to display new information with minimal delays.

Which method should the company use to collect and analyze the logs?

- A. Enable detailed monitoring on Amazon EC2, use Amazon CloudWatch agent to store logs in Amazon S3, and use Amazon Athena for fast, interactive log analytics.
- B. Use the Amazon Kinesis Producer Library (KPL) agent on Amazon EC2 to collect and send data to Kinesis Data Streams to further push the data to Amazon Elasticsearch Service and visualize using Amazon QuickSight.
- C. Use the Amazon Kinesis Producer Library (KPL) agent on Amazon EC2 to collect and send data to Kinesis Data Firehose to further push the data to Amazon Elasticsearch Service and Kibana.
- D. Use Amazon CloudWatch subscriptions to get access to a real-time feed of logs and have the logs delivered to Amazon Kinesis Data Streams to further push the data to Amazon Elasticsearch Service and Kibana.

**Answer:** D

#### NEW QUESTION 8

A company has a marketing department and a finance department. The departments are storing data in Amazon S3 in their own AWS accounts in AWS Organizations. Both departments use AWS Lake Formation to catalog and secure their data. The departments have some databases and tables that share common names.

The marketing department needs to securely access some tables from the finance department. Which two steps are required for this process? (Choose two.)

- A. The finance department grants Lake Formation permissions for the tables to the external account for the marketing department.
- B. The finance department creates cross-account IAM permissions to the table for the marketing department role.
- C. The marketing department creates an IAM role that has permissions to the Lake Formation tables.

**Answer:** AB

#### Explanation:

Granting Lake Formation Permissions  
Creating an IAM role (AWS CLI)

#### NEW QUESTION 9

A company has 1 million scanned documents stored as image files in Amazon S3. The documents contain typewritten application forms with information including the applicant first name, applicant last name, application date, application type, and application text. The company has developed a machine learning algorithm to extract the metadata values from the scanned documents. The company wants to allow internal data analysts to analyze and find applications using the applicant name, application date, or application text. The original images should also be downloadable. Cost control is secondary to query performance.

Which solution organizes the images and metadata to drive insights while meeting the requirements?

- A. For each image, use object tags to add the metadata.
- B. Use Amazon S3 Select to retrieve the files based on the applicant name and application date.
- C. Index the metadata and the Amazon S3 location of the image file in Amazon Elasticsearch Service. Allow the data analysts to use Kibana to submit queries to the Elasticsearch cluster.

- D. Store the metadata and the Amazon S3 location of the image file in an Amazon Redshift tabl
- E. Allow the data analysts to run ad-hoc queries on the table.
- F. Store the metadata and the Amazon S3 location of the image files in an Apache Parquet file in Amazon S3, and define a table in the AWS Glue Data Catalo
- G. Allow data analysts to use Amazon Athena to submit custom queries.

**Answer:** B

**Explanation:**

<https://aws.amazon.com/blogs/machine-learning/automatically-extract-text-and-structured-data-from-documents>

**NEW QUESTION 10**

A university intends to use Amazon Kinesis Data Firehose to collect JSON-formatted batches of water quality readings in Amazon S3. The readings are from 50 sensors scattered across a local lake. Students will query the stored data using Amazon Athena to observe changes in a captured metric over time, such as water temperature or acidity. Interest has grown in the study, prompting the university to reconsider how data will be stored.

Which data format and partitioning choices will MOST significantly reduce costs? (Choose two.)

- A. Store the data in Apache Avro format using Snappy compression.
- B. Partition the data by year, month, and day.
- C. Store the data in Apache ORC format using no compression.
- D. Store the data in Apache Parquet format using Snappy compression.
- E. Partition the data by sensor, year, month, and day.

**Answer:** CD

**NEW QUESTION 10**

A financial company hosts a data lake in Amazon S3 and a data warehouse on an Amazon Redshift cluster. The company uses Amazon QuickSight to build dashboards and wants to secure access from its on-premises Active Directory to Amazon QuickSight.

How should the data be secured?

- A. Use an Active Directory connector and single sign-on (SSO) in a corporate network environment.
- B. Use a VPC endpoint to connect to Amazon S3 from Amazon QuickSight and an IAM role to authenticate Amazon Redshift.
- C. Establish a secure connection by creating an S3 endpoint to connect Amazon QuickSight and a VPC endpoint to connect to Amazon Redshift.
- D. Place Amazon QuickSight and Amazon Redshift in the security group and use an Amazon S3 endpoint to connect Amazon QuickSight to Amazon S3.

**Answer:** A

**Explanation:**

<https://docs.aws.amazon.com/quicksight/latest/user/directory-integration.html>

**NEW QUESTION 11**

A large financial company is running its ETL process. Part of this process is to move data from Amazon S3 into an Amazon Redshift cluster. The company wants to use the most cost-efficient method to load the dataset into Amazon Redshift.

Which combination of steps would meet these requirements? (Choose two.)

- A. Use the COPY command with the manifest file to load data into Amazon Redshift.
- B. Use S3DistCp to load files into Amazon Redshift.
- C. Use temporary staging tables during the loading process.
- D. Use the UNLOAD command to upload data into Amazon Redshift.
- E. Use Amazon Redshift Spectrum to query files from Amazon S3.

**Answer:** AC

**NEW QUESTION 14**

A medical company has a system with sensor devices that read metrics and send them in real time to an Amazon Kinesis data stream. The Kinesis data stream has multiple shards. The company needs to calculate the average value of a numeric metric every second and set an alarm for whenever the value is above one threshold or below another threshold. The alarm must be sent to Amazon Simple Notification Service (Amazon SNS) in less than 30 seconds.

Which architecture meets these requirements?

- A. Use an Amazon Kinesis Data Firehose delivery stream to read the data from the Kinesis data stream with an AWS Lambda transformation function that calculates the average per second and sends the alarm to Amazon SNS.
- B. Use an AWS Lambda function to read from the Kinesis data stream to calculate the average per second and sent the alarm to Amazon SNS.
- C. Use an Amazon Kinesis Data Firehose deliver stream to read the data from the Kinesis data stream and store it on Amazon S3. Have Amazon S3 trigger an AWS Lambda function that calculates the average per second and sends the alarm to Amazon SNS.
- D. Use an Amazon Kinesis Data Analytics application to read from the Kinesis data stream and calculate the average per secon
- E. Send the results to an AWS Lambda function that sends the alarm to Amazon SNS.

**Answer:** D

**NEW QUESTION 16**

A company has an encrypted Amazon Redshift cluster. The company recently enabled Amazon Redshift audit logs and needs to ensure that the audit logs are also encrypted at rest. The logs are retained for 1 year. The auditor queries the logs once a month.

What is the MOST cost-effective way to meet these requirements?

- A. Encrypt the Amazon S3 bucket where the logs are stored by using AWS Key Management Service (AWS KMS). Copy the data into the Amazon Redshift cluster from Amazon S3 on a daily basi
- B. Query the data as required.
- C. Disable encryption on the Amazon Redshift cluster, configure audit logging, and encrypt the Amazon Redshift cluste
- D. Use Amazon Redshift Spectrum to query the data as required.



- E. Enable default encryption on the Amazon S3 bucket where the logs are stored by using AES-256 encryption
- F. Copy the data into the Amazon Redshift cluster from Amazon S3 on a daily basis
- G. Query the data as required.
- H. Enable default encryption on the Amazon S3 bucket where the logs are stored by using AES-256 encryption
- I. Use Amazon Redshift Spectrum to query the data as required.

**Answer:** A

#### NEW QUESTION 17

A power utility company is deploying thousands of smart meters to obtain real-time updates about power consumption. The company is using Amazon Kinesis Data Streams to collect the data streams from smart meters. The consumer application uses the Kinesis Client Library (KCL) to retrieve the stream data. The company has only one consumer application.

The company observes an average of 1 second of latency from the moment that a record is written to the stream until the record is read by a consumer application. The company must reduce this latency to 500 milliseconds.

Which solution meets these requirements?

- A. Use enhanced fan-out in Kinesis Data Streams.
- B. Increase the number of shards for the Kinesis data stream.
- C. Reduce the propagation delay by overriding the KCL default settings.
- D. Develop consumers by using Amazon Kinesis Data Firehose.

**Answer:** C

#### Explanation:

The KCL defaults are set to follow the best practice of polling every 1 second. This default results in average propagation delays that are typically below 1 second.

#### NEW QUESTION 20

A company wants to enrich application logs in near-real-time and use the enriched dataset for further analysis. The application is running on Amazon EC2 instances across multiple Availability Zones and storing its logs using Amazon CloudWatch Logs. The enrichment source is stored in an Amazon DynamoDB table. Which solution meets the requirements for the event collection and enrichment?

- A. Use a CloudWatch Logs subscription to send the data to Amazon Kinesis Data Firehose
- B. Use AWS Lambda to transform the data in the Kinesis Data Firehose delivery stream and enrich it with the data in the DynamoDB table
- C. Configure Amazon S3 as the Kinesis Data Firehose delivery destination.
- D. Export the raw logs to Amazon S3 on an hourly basis using the AWS CLI
- E. Use AWS Glue crawlers to catalog the log
- F. Set up an AWS Glue connection for the DynamoDB table and set up an AWS Glue ETL job to enrich the data
- G. Store the enriched data in Amazon S3.
- H. Configure the application to write the logs locally and use Amazon Kinesis Agent to send the data to Amazon Kinesis Data Stream
- I. Configure a Kinesis Data Analytics SQL application with the Kinesis data stream as the source
- J. Join the SQL application input stream with DynamoDB records, and then store the enriched output stream in Amazon S3 using Amazon Kinesis Data Firehose.
- K. Export the raw logs to Amazon S3 on an hourly basis using the AWS CLI
- L. Use Apache Spark SQL on Amazon EMR to read the logs from Amazon S3 and enrich the records with the data from DynamoDB
- M. Store the enriched data in Amazon S3.

**Answer:** A

#### Explanation:

<https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/SubscriptionFilters.html#FirehoseExample>

#### NEW QUESTION 24

A company wants to improve user satisfaction for its smart home system by adding more features to its recommendation engine. Each sensor asynchronously pushes its nested JSON data into Amazon Kinesis Data Streams using the Kinesis Producer Library (KPL) in Java. Statistics from a set of failed sensors showed that, when a sensor is malfunctioning, its recorded data is not always sent to the cloud.

The company needs a solution that offers near-real-time analytics on the data from the most updated sensors. Which solution enables the company to meet these requirements?

- A. Set the RecordMaxBufferedTime property of the KPL to "1" to disable the buffering on the sensor side. Use Kinesis Data Analytics to enrich the data based on a company-developed anomaly detection SQL script
- B. Push the enriched data to a fleet of Kinesis data streams and enable the data transformation feature to flatten the JSON file
- C. Instantiate a dense storage Amazon Redshift cluster and use it as the destination for the Kinesis Data Firehose delivery stream.
- D. Update the sensors code to use the PutRecord/PutRecords call from the Kinesis Data Streams API with the AWS SDK for Java
- E. Use Kinesis Data Analytics to enrich the data based on a company-developed anomaly detection SQL script
- F. Direct the output of KDA application to a Kinesis Data Firehose delivery stream, enable the data transformation feature to flatten the JSON file, and set the Kinesis Data Firehose destination to an Amazon Elasticsearch Service cluster.
- G. Set the RecordMaxBufferedTime property of the KPL to "0" to disable the buffering on the sensor side. Connect for each stream a dedicated Kinesis Data Firehose delivery stream and enable the data transformation feature to flatten the JSON file before sending it to an Amazon S3 bucket
- H. Load the S3 data into an Amazon Redshift cluster.
- I. Update the sensors code to use the PutRecord/PutRecords call from the Kinesis Data Streams API with the AWS SDK for Java
- J. Use AWS Glue to fetch and process data from the stream using the Kinesis Client Library (KCL). Instantiate an Amazon Elasticsearch Service cluster and use AWS Lambda to directly push data into it.

**Answer:** B

#### Explanation:

<https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-kpl.html>

The KPL can incur an additional processing delay of up to RecordMaxBufferedTime within the library (user-configurable). Larger values of RecordMaxBufferedTime result in higher packing efficiencies and better performance. Applications that cannot tolerate this additional delay may need to use the AWS SDK directly.

**NEW QUESTION 26**

A company wants to research user turnover by analyzing the past 3 months of user activities. With millions of users, 1.5 TB of uncompressed data is generated each day. A 30-node Amazon Redshift cluster with 2.56 TB of solid state drive (SSD) storage for each node is required to meet the query performance goals. The company wants to run an additional analysis on a year's worth of historical data to examine trends indicating which features are most popular. This analysis will be done once a week.

What is the MOST cost-effective solution?

- A. Increase the size of the Amazon Redshift cluster to 120 nodes so it has enough storage capacity to hold 1 year of data
- B. Then use Amazon Redshift for the additional analysis.
- C. Keep the data from the last 90 days in Amazon Redshift
- D. Move data older than 90 days to Amazon S3 and store it in Apache Parquet format partitioned by date
- E. Then use Amazon Redshift Spectrum for the additional analysis.
- F. Keep the data from the last 90 days in Amazon Redshift
- G. Move data older than 90 days to Amazon S3 and store it in Apache Parquet format partitioned by date
- H. Then provision a persistent Amazon EMR cluster and use Apache Presto for the additional analysis.
- I. Resize the cluster node type to the dense storage node type (DS2) for an additional 16 TB storage capacity on each individual node in the Amazon Redshift cluster
- J. Then use Amazon Redshift for the additional analysis.

**Answer: B**

**NEW QUESTION 28**

An Amazon Redshift database contains sensitive user data. Logging is necessary to meet compliance requirements. The logs must contain database authentication attempts, connections, and disconnections. The logs must also contain each query run against the database and record which database user ran each query.

Which steps will create the required logs?

- A. Enable Amazon Redshift Enhanced VPC Routing
- B. Enable VPC Flow Logs to monitor traffic.
- C. Allow access to the Amazon Redshift database using AWS IAM only
- D. Log access using AWS CloudTrail.
- E. Enable audit logging for Amazon Redshift using the AWS Management Console or the AWS CLI.
- F. Enable and download audit reports from AWS Artifact.

**Answer: C**

**NEW QUESTION 32**

A retail company is building its data warehouse solution using Amazon Redshift. As a part of that effort, the company is loading hundreds of files into the fact table created in its Amazon Redshift cluster. The company wants the solution to achieve the highest throughput and optimally use cluster resources when loading data into the company's fact table.

How should the company meet these requirements?

- A. Use multiple COPY commands to load the data into the Amazon Redshift cluster.
- B. Use S3DistCp to load multiple files into the Hadoop Distributed File System (HDFS) and use an HDFS connector to ingest the data into the Amazon Redshift cluster.
- C. Use LOAD commands equal to the number of Amazon Redshift cluster nodes and load the data in parallel into each node.
- D. Use a single COPY command to load the data into the Amazon Redshift cluster.

**Answer: D**

**Explanation:**

[https://docs.aws.amazon.com/redshift/latest/dg/c\\_best-practices-single-copy-command.html](https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html)

**NEW QUESTION 36**

A company is migrating its existing on-premises ETL jobs to Amazon EMR. The code consists of a series of jobs written in Java. The company needs to reduce overhead for the system administrators without changing the underlying code. Due to the sensitivity of the data, compliance requires that the company use root device volume encryption on all nodes in the cluster. Corporate standards require that environments be provisioned through AWS CloudFormation when possible. Which solution satisfies these requirements?

- A. Install open-source Hadoop on Amazon EC2 instances with encrypted root device volume
- B. Configure the cluster in the CloudFormation template.
- C. Use a CloudFormation template to launch an EMR cluster
- D. In the configuration section of the cluster, define a bootstrap action to enable TLS.
- E. Create a custom AMI with encrypted root device volume
- F. Configure Amazon EMR to use the custom AMI using the CustomAmiId property in the CloudFormation template.
- G. Use a CloudFormation template to launch an EMR cluster
- H. In the configuration section of the cluster, define a bootstrap action to encrypt the root device volume of every node.

**Answer: C**

**NEW QUESTION 37**

A telecommunications company is looking for an anomaly-detection solution to identify fraudulent calls. The company currently uses Amazon Kinesis to stream voice call records in a JSON format from its on-premises database to Amazon S3. The existing dataset contains voice call records with 200 columns. To detect fraudulent calls, the solution would need to look at 5 of these columns only.

The company is interested in a cost-effective solution using AWS that requires minimal effort and experience in anomaly-detection algorithms.

Which solution meets these requirements?

- A. Use an AWS Glue job to transform the data from JSON to Apache Parquet
- B. Use AWS Glue crawlers to discover the schema and build the AWS Glue Data Catalog

- C. Use Amazon Athena to create a table with a subset of column
- D. Use Amazon QuickSight to visualize the data and then use Amazon QuickSight machine learning-powered anomaly detection.
- E. Use Kinesis Data Firehose to detect anomalies on a data stream from Kinesis by running SQL queries, which compute an anomaly score for all calls and store the output in Amazon RD
- F. Use Amazon Athena to build a dataset and Amazon QuickSight to visualize the results.
- G. Use an AWS Glue job to transform the data from JSON to Apache Parquet
- H. Use AWS Glue crawlers to discover the schema and build the AWS Glue Data Catalog
- I. Use Amazon SageMaker to build an anomaly detection model that can detect fraudulent calls by ingesting data from Amazon S3.
- J. Use Kinesis Data Analytics to detect anomalies on a data stream from Kinesis by running SQL queries, which compute an anomaly score for all call
- K. Connect Amazon QuickSight to Kinesis Data Analytics to visualize the anomaly scores.

**Answer:** A

#### NEW QUESTION 40

A company leverages Amazon Athena for ad-hoc queries against data stored in Amazon S3. The company wants to implement additional controls to separate query execution and query history among users, teams, or applications running in the same AWS account to comply with internal security policies. Which solution meets these requirements?

- A. Create an S3 bucket for each given use case, create an S3 bucket policy that grants permissions to appropriate individual IAM user
- B. and apply the S3 bucket policy to the S3 bucket.
- C. Create an Athena workgroup for each given use case, apply tags to the workgroup, and create an IAM policy using the tags to apply appropriate permissions to the workgroup.
- D. Create an IAM role for each given use case, assign appropriate permissions to the role for the given usecase, and add the role to associate the role with Athena.
- E. Create an AWS Glue Data Catalog resource policy for each given use case that grants permissions to appropriate individual IAM users, and apply the resource policy to the specific tables used by Athena.

**Answer:** B

#### Explanation:

<https://docs.aws.amazon.com/athena/latest/ug/user-created-workgroups.html>

Amazon Athena Workgroups - A new resource type that can be used to separate query execution and query history between Users, Teams, or Applications running under the same AWS account [https://aws.amazon.com/about-aws/whats-new/2019/02/athena\\_workgroups/](https://aws.amazon.com/about-aws/whats-new/2019/02/athena_workgroups/)

#### NEW QUESTION 43

A company has a data warehouse in Amazon Redshift that is approximately 500 TB in size. New data is imported every few hours and read-only queries are run throughout the day and evening. There is a particularly heavy load with no writes for several hours each morning on business days. During those hours, some queries are queued and take a long time to execute. The company needs to optimize query execution and avoid any downtime. What is the MOST cost-effective solution?

- A. Enable concurrency scaling in the workload management (WLM) queue.
- B. Add more nodes using the AWS Management Console during peak hour
- C. Set the distribution style to ALL.
- D. Use elastic resize to quickly add nodes during peak time
- E. Remove the nodes when they are not needed.
- F. Use a snapshot, restore, and resize operation
- G. Switch to the new target cluster.

**Answer:** A

#### Explanation:

<https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html>

#### NEW QUESTION 48

A US-based sneaker retail company launched its global website. All the transaction data is stored in Amazon RDS and curated historic transaction data is stored in Amazon Redshift in the us-east-1 Region. The business intelligence (BI) team wants to enhance the user experience by providing a dashboard for sneaker trends. The BI team decides to use Amazon QuickSight to render the website dashboards. During development, a team in Japan provisioned Amazon QuickSight in ap-northeast-1. The team is having difficulty connecting Amazon QuickSight from ap-northeast-1 to Amazon Redshift in us-east-1. Which solution will solve this issue and meet the requirements?

- A. In the Amazon Redshift console, choose to configure cross-Region snapshots and set the destination Region as ap-northeast-1. Restore the Amazon Redshift Cluster from the snapshot and connect to Amazon QuickSight launched in ap-northeast-1.
- B. Create a VPC endpoint from the Amazon QuickSight VPC to the Amazon Redshift VPC so Amazon QuickSight can access data from Amazon Redshift.
- C. Create an Amazon Redshift endpoint connection string with Region information in the string and use this connection string in Amazon QuickSight to connect to Amazon Redshift.
- D. Create a new security group for Amazon Redshift in us-east-1 with an inbound rule authorizing access from the appropriate IP address range for the Amazon QuickSight servers in ap-northeast-1.

**Answer:** B

#### NEW QUESTION 53

A data engineering team within a shared workspace company wants to build a centralized logging system for all weblogs generated by the space reservation system. The company has a fleet of Amazon EC2 instances that process requests for shared space reservations on its website. The data engineering team wants to ingest all weblogs into a service that will provide a near-real-time search engine. The team does not want to manage the maintenance and operation of the logging system.

Which solution allows the data engineering team to efficiently set up the web logging system within AWS?

- A. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatch
- B. Choose Amazon Elasticsearch Service as the end destination of the weblogs.



- C. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Data Firehose delivery stream to CloudWatc
- D. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
- E. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatc
- F. Configure Splunk as the end destination of the weblogs.
- G. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Firehose delivery stream to CloudWatc
- H. Configure Amazon DynamoDB as the end destination of the weblogs.

**Answer:** B

**Explanation:**

[https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL\\_ES\\_Stream.html](https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL_ES_Stream.html)

#### NEW QUESTION 58

A company has an application that uses the Amazon Kinesis Client Library (KCL) to read records from a Kinesis data stream.

After a successful marketing campaign, the application experienced a significant increase in usage. As a result, a data analyst had to split some shards in the data stream. When the shards were split, the application started throwing an `ExpiredIteratorExceptions` error sporadically. What should the data analyst do to resolve this?

- A. Increase the number of threads that process the stream records.
- B. Increase the provisioned read capacity units assigned to the stream's Amazon DynamoDB table.
- C. Increase the provisioned write capacity units assigned to the stream's Amazon DynamoDB table.
- D. Decrease the provisioned write capacity units assigned to the stream's Amazon DynamoDB table.

**Answer:** C

#### NEW QUESTION 60

A media company has been performing analytics on log data generated by its applications. There has been a recent increase in the number of concurrent analytics jobs running, and the overall performance of existing jobs is decreasing as the number of new jobs is increasing. The partitioned data is stored in Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA) and the analytic processing is performed on Amazon EMR clusters using the EMR File System (EMRFS) with consistent view enabled. A data analyst has determined that it is taking longer for the EMR task nodes to list objects in Amazon S3.

Which action would MOST likely increase the performance of accessing log data in Amazon S3?

- A. Use a hash function to create a random string and add that to the beginning of the object prefixes when storing the log data in Amazon S3.
- B. Use a lifecycle policy to change the S3 storage class to S3 Standard for the log data.
- C. Increase the read capacity units (RCUs) for the shared Amazon DynamoDB table.
- D. Redeploy the EMR clusters that are running slowly to a different Availability Zone.

**Answer:** C

**Explanation:**

<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emrfs-metadata.html>

#### NEW QUESTION 65

A retail company stores order invoices in an Amazon OpenSearch Service (Amazon Elasticsearch Service) cluster. Indices on the cluster are created monthly. Once a new month begins, no new writes are made to any of the indices from the previous months. The company has been expanding the storage on the Amazon OpenSearch Service (Amazon Elasticsearch Service) cluster to avoid running out of space, but the company wants to reduce costs. Most searches on the cluster are on the most recent 3 months of data, while the audit team requires infrequent access to older data to generate periodic reports. The most recent 3 months of data must be quickly available for queries, but the audit team can tolerate slower queries if the solution saves on cluster costs.

Which of the following is the MOST operationally efficient solution to meet these requirements?

- A. Archive indices that are older than 3 months by using Index State Management (ISM) to create a policy to store the indices in Amazon S3 Glacier. When the audit team requires the archived data, restore the archived indices back to the Amazon OpenSearch Service (Amazon Elasticsearch Service) cluster.
- B. Archive indices that are older than 3 months by taking manual snapshots and storing the snapshots in Amazon S3. When the audit team requires the archived data, restore the archived indices back to the Amazon OpenSearch Service (Amazon Elasticsearch Service) cluster.
- C. Archive indices that are older than 3 months by using Index State Management (ISM) to create a policy to migrate the indices to Amazon OpenSearch Service (Amazon Elasticsearch Service) UltraWarm storage.
- D. Archive indices that are older than 3 months by using Index State Management (ISM) to create a policy to migrate the indices to Amazon OpenSearch Service (Amazon Elasticsearch Service) UltraWarm storage. When the audit team requires the older data, migrate the indices in UltraWarm storage back to hot storage.

**Answer:** D

#### NEW QUESTION 70

A company needs to collect streaming data from several sources and store the data in the AWS Cloud. The dataset is heavily structured, but analysts need to perform several complex SQL queries and need consistent performance. Some of the data is queried more frequently than the rest. The company wants a solution that meets its performance requirements in a cost-effective manner.

Which solution meets these requirements?

- A. Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon S3. Use Amazon Athena to perform SQL queries over the ingested data.
- B. Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon Redshift. Enable Amazon Redshift workload management (WLM) to prioritize workloads.
- C. Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon Redshift.
- D. Enable Amazon Redshift workload management (WLM) to prioritize workloads.
- E. Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon S3. Load frequently queried data to Amazon Redshift using the COPY command.
- F. Use Amazon Redshift Spectrum for less frequently queried data.

**Answer:** B



**NEW QUESTION 75**

A marketing company collects clickstream data. The company sends the data to Amazon Kinesis Data Firehose and stores the data in Amazon S3. The company wants to build a series of dashboards that will be used by hundreds of users across different departments. The company will use Amazon QuickSight to develop these dashboards. The company has limited resources and wants a solution that could scale and provide daily updates about clickstream activity. Which combination of options will provide the MOST cost-effective solution? (Select TWO.)

- A. Use Amazon Redshift to store and query the clickstream data.
- B. Use QuickSight with a direct SQL query.
- C. Use Amazon Athena to query the clickstream data in Amazon S3.
- D. Use S3 analytics to query the clickstream data.
- E. Use the QuickSight SPICE engine with a daily refresh.

**Answer:** BD

**NEW QUESTION 77**

A marketing company is storing its campaign response data in Amazon S3. A consistent set of sources has generated the data for each campaign. The data is saved into Amazon S3 as .csv files. A business analyst will use Amazon Athena to analyze each campaign's data. The company needs the cost of ongoing data analysis with Athena to be minimized.

Which combination of actions should a data analytics specialist take to meet these requirements? (Choose two.)

- A. Convert the .csv files to Apache Parquet.
- B. Convert the .csv files to Apache Avro.
- C. Partition the data by campaign.
- D. Partition the data by source.
- E. Compress the .csv files.

**Answer:** AC

**Explanation:**

<https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-tips-for-amazon-athena/>

**NEW QUESTION 80**

A streaming application is reading data from Amazon Kinesis Data Streams and immediately writing the data to an Amazon S3 bucket every 10 seconds. The application is reading data from hundreds of shards. The batch interval cannot be changed due to a separate requirement. The data is being accessed by Amazon Athena. Users are seeing degradation in query performance as time progresses.

Which action can help improve query performance?

- A. Merge the files in Amazon S3 to form larger files.
- B. Increase the number of shards in Kinesis Data Streams.
- C. Add more memory and CPU capacity to the streaming application.
- D. Write the files to multiple S3 buckets.

**Answer:** A

**Explanation:**

<https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-tips-for-amazon-athena/>

**NEW QUESTION 84**

A company has developed an Apache Hive script to batch process data stored in Amazon S3. The script needs to run once every day and store the output in Amazon S3. The company tested the script, and it completes within 30 minutes on a small local three-node cluster.

Which solution is the MOST cost-effective for scheduling and executing the script?

- A. Create an AWS Lambda function to spin up an Amazon EMR cluster with a Hive execution step.
- B. Set `KeepJobFlowAliveWhenNoSteps` to false and disable the termination protection flag.
- C. Use Amazon CloudWatch Events to schedule the Lambda function to run daily.
- D. Use the AWS Management Console to spin up an Amazon EMR cluster with Python, Hive, and Apache Oozie.
- E. Hive, and Apache Oozie.
- F. Set the termination protection flag to true and use Spot Instances for the core nodes of the cluster.
- G. Configure an Oozie workflow in the cluster to invoke the Hive script daily.
- H. Create an AWS Glue job with the Hive script to perform the batch operation.
- I. Configure the job to run once a day using a time-based schedule.
- J. Use AWS Lambda layers and load the Hive runtime to AWS Lambda and copy the Hive script. Schedule the Lambda function to run daily by creating a workflow using AWS Step Functions.

**Answer:** C

**NEW QUESTION 88**

An insurance company has raw data in JSON format that is sent without a predefined schedule through an Amazon Kinesis Data Firehose delivery stream to an Amazon S3 bucket. An AWS Glue crawler is scheduled to run every 8 hours to update the schema in the data catalog of the tables stored in the S3 bucket. Data analysts analyze the data using Apache Spark SQL on Amazon EMR set up with AWS Glue Data Catalog as the metastore. Data analysts say that, occasionally, the data they receive is stale. A data engineer needs to provide access to the most up-to-date data.

Which solution meets these requirements?

- A. Create an external schema based on the AWS Glue Data Catalog on the existing Amazon Redshift cluster to query new data in Amazon S3 with Amazon Redshift Spectrum.
- B. Use Amazon CloudWatch Events with the rate (1 hour) expression to execute the AWS Glue crawler every hour.
- C. Using the AWS CLI, modify the execution schedule of the AWS Glue crawler from 8 hours to 1 minute.
- D. Run the AWS Glue crawler from an AWS Lambda function triggered by an `S3:ObjectCreated:*` event notification on the S3 bucket.

**Answer:** D

**Explanation:**

<https://docs.aws.amazon.com/AmazonS3/latest/dev/NotificationHowTo.html> "you can use a wildcard (for example, s3:ObjectCreated:\*) to request notification when an object is created regardless of the API used" "AWS Lambda can run custom code in response to Amazon S3 bucket events. You upload your custom code to AWS Lambda and create what is called a Lambda function. When Amazon S3 detects an event of a specific type (for example, an object created event), it can publish the event to AWS Lambda and invoke your function in Lambda. In response, AWS Lambda runs your function."

**NEW QUESTION 91**

A banking company wants to collect large volumes of transactional data using Amazon Kinesis Data Streams for real-time analytics. The company uses PutRecord to send data to Amazon Kinesis, and has observed network outages during certain times of the day. The company wants to obtain exactly once semantics for the entire processing pipeline.

What should the company do to obtain these characteristics?

- A. Design the application so it can remove duplicates during processing by embedding a unique ID in each record.
- B. Rely on the processing semantics of Amazon Kinesis Data Analytics to avoid duplicate processing of events.
- C. Design the data producer so events are not ingested into Kinesis Data Streams multiple times.
- D. Rely on the exactly once processing semantics of Apache Flink and Apache Spark Streaming included in Amazon EMR.

**Answer:** A

**NEW QUESTION 93**

A real estate company has a mission-critical application using Apache HBase in Amazon EMR. Amazon EMR is configured with a single master node. The company has over 5 TB of data stored on an Hadoop Distributed File System (HDFS). The company wants a cost-effective solution to make its HBase data highly available. Which architectural pattern meets company's requirements?

- A. Use Spot Instances for core and task nodes and a Reserved Instance for the EMR master node. Configure the EMR cluster with multiple master node
- B. Schedule automated snapshots using Amazon EventBridge.
- C. Store the data on an EMR File System (EMRFS) instead of HDFS
- D. Enable EMRFS consistent view. Create an EMR HBase cluster with multiple master node
- E. Point the HBase root directory to an Amazon S3 bucket.
- F. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view. Run two separate EMR clusters in two different Availability Zones
- G. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.
- H. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view. Create a primary EMR HBase cluster with multiple master node
- I. Create a secondary EMR HBase read-replica cluster in a separate Availability Zone
- J. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.

**Answer:** D

**NEW QUESTION 98**

An ecommerce company stores customer purchase data in Amazon RDS. The company wants a solution to store and analyze historical data. The most recent 6 months of data will be queried frequently for analytics workloads. This data is several terabytes large. Once a month, historical data for the last 5 years must be accessible and will be joined with the more recent data. The company wants to optimize performance and cost.

Which storage solution will meet these requirements?

- A. Create a read replica of the RDS database to store the most recent 6 months of data
- B. Copy the historical data into Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3 and Amazon RDS
- C. Run historical queries using Amazon Athena.
- D. Use an ETL tool to incrementally load the most recent 6 months of data into an Amazon Redshift cluster
- E. Run more frequent queries against this cluster
- F. Create a read replica of the RDS database to run queries on the historical data.
- G. Incrementally copy data from Amazon RDS to Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3. Use Amazon Athena to query the data.
- H. Incrementally copy data from Amazon RDS to Amazon S3. Load and store the most recent 6 months of data in Amazon Redshift
- I. Configure an Amazon Redshift Spectrum table to connect to all historical data.

**Answer:** D

**NEW QUESTION 100**

A large university has adopted a strategic goal of increasing diversity among enrolled students. The data analytics team is creating a dashboard with data visualizations to enable stakeholders to view historical trends. All access must be authenticated using Microsoft Active Directory. All data in transit and at rest must be encrypted.

Which solution meets these requirements?

- A. Amazon QuickSight Standard edition configured to perform identity federation using SAML 2.0 and the default encryption settings.
- B. Amazon QuickSight Enterprise edition configured to perform identity federation using SAML 2.0 and the default encryption settings.
- C. Amazon QuickSight Standard edition using AD Connector to authenticate using Active Directory. Configure Amazon QuickSight to use customer-provided keys imported into AWS KMS.
- D. Amazon QuickSight Enterprise edition using AD Connector to authenticate using Active Directory. Configure Amazon QuickSight to use customer-provided keys imported into AWS KMS.

**Answer:** D

**NEW QUESTION 105**

A financial services company needs to aggregate daily stock trade data from the exchanges into a data store. The company requires that data be streamed directly into the data store, but also occasionally allows data to be modified using SQL. The solution should integrate complex, analytic queries running with minimal

latency. The solution must provide a business intelligence dashboard that enables viewing of the top contributors to anomalies in stock prices. Which solution meets the company's requirements?

- A. Use Amazon Kinesis Data Firehose to stream data to Amazon S3. Use Amazon Athena as a data source for Amazon QuickSight to create a business intelligence dashboard.
- B. Use Amazon Kinesis Data Streams to stream data to Amazon Redshift
- C. Use Amazon Redshift as a data source for Amazon QuickSight to create a business intelligence dashboard.
- D. Use Amazon Kinesis Data Firehose to stream data to Amazon Redshift
- E. Use Amazon Redshift as a data source for Amazon QuickSight to create a business intelligence dashboard.
- F. Use Amazon Kinesis Data Streams to stream data to Amazon S3. Use Amazon Athena as a data source for Amazon QuickSight to create a business intelligence dashboard.

**Answer: C**

#### NEW QUESTION 109

An advertising company has a data lake that is built on Amazon S3. The company uses AWS Glue Data Catalog to maintain the metadata. The data lake is several years old and its overall size has increased exponentially as additional data sources and metadata are stored in the data lake. The data lake administrator wants to implement a mechanism to simplify permissions management between Amazon S3 and the Data Catalog to keep them in sync. Which solution will simplify permissions management with minimal development effort?

- A. Set AWS Identity and Access Management (IAM) permissions for AWS Glue
- B. Use AWS Lake Formation permissions
- C. Manage AWS Glue and S3 permissions by using bucket policies
- D. Use Amazon Cognito user pools.

**Answer: B**

#### NEW QUESTION 110

A large energy company is using Amazon QuickSight to build dashboards and report the historical usage data of its customers. This data is hosted in Amazon Redshift. The reports need access to all the fact tables' billions of records to create aggregation in real time grouping by multiple dimensions. A data analyst created the dataset in QuickSight by using a SQL query and not SPICE. Business users have noted that the response time is not fast enough to meet their needs.

Which action would speed up the response time for the reports with the LEAST implementation effort?

- A. Use QuickSight to modify the current dataset to use SPICE
- B. Use AWS Glue to create an Apache Spark job that joins the fact table with the dimension
- C. Load the data into a new table
- D. Use Amazon Redshift to create a materialized view that joins the fact table with the dimensions
- E. Use Amazon Redshift to create a stored procedure that joins the fact table with the dimensions. Load the data into a new table

**Answer: A**

#### NEW QUESTION 115

A financial company uses Apache Hive on Amazon EMR for ad-hoc queries. Users are complaining of sluggish performance.

A data analyst notes the following:

- Approximately 90% of queries are submitted 1 hour after the market opens.
- Hadoop Distributed File System (HDFS) utilization never exceeds 10%.

Which solution would help address the performance issues?

- A. Create instance fleet configurations for core and task node
- B. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metric.
- C. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch CapacityRemainingGB metric.
- D. Create instance fleet configurations for core and task node
- E. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch YARNMemoryAvailablePercentage metric.
- F. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch YARNMemoryAvailablePercentage metric.
- G. Create instance group configurations for core and task node
- H. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metric.
- I. Create an automatic scaling policy to scale in the instance groups based on the CloudWatch CapacityRemainingGB metric.
- J. Create instance group configurations for core and task node
- K. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch YARNMemoryAvailablePercentage metric.
- L. Create an automatic scaling policy to scale in the instance groups based on the CloudWatch YARNMemoryAvailablePercentage metric.

**Answer: D**

#### Explanation:

<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html>

#### NEW QUESTION 118

A retail company leverages Amazon Athena for ad-hoc queries against an AWS Glue Data Catalog. The data analytics team manages the data catalog and data access for the company. The data analytics team wants to separate queries and manage the cost of running those queries by different workloads and teams. Ideally, the data analysts want to group the queries run by different users within a team, store the query results in individual Amazon S3 buckets specific to each team, and enforce cost constraints on the queries run against the Data Catalog.

Which solution meets these requirements?

- A. Create IAM groups and resource tags for each team within the company
- B. Set up IAM policies that control user access and actions on the Data Catalog resources.
- C. Create Athena resource groups for each team within the company and assign users to these groups
- D. Add S3 bucket names and other query configurations to the properties list for the resource groups.

- E. Create Athena workgroups for each team within the company
- F. Set up IAM workgroup policies that control user access and actions on the workgroup resources.
- G. Create Athena query groups for each team within the company and assign users to the groups.

**Answer:** C

**Explanation:**

[https://aws.amazon.com/about-aws/whats-new/2019/02/athena\\_workgroups/](https://aws.amazon.com/about-aws/whats-new/2019/02/athena_workgroups/)

#### NEW QUESTION 120

A human resources company maintains a 10-node Amazon Redshift cluster to run analytics queries on the company's data. The Amazon Redshift cluster contains a product table and a transactions table, and both tables have a product\_sku column. The tables are over 100 GB in size. The majority of queries run on both tables.

Which distribution style should the company use for the two tables to achieve optimal query performance?

- A. An EVEN distribution style for both tables
- B. A KEY distribution style for both tables
- C. An ALL distribution style for the product table and an EVEN distribution style for the transactions table
- D. An EVEN distribution style for the product table and an KEY distribution style for the transactions table

**Answer:** B

#### NEW QUESTION 123

An online gaming company is using an Amazon Kinesis Data Analytics SQL application with a Kinesis data stream as its source. The source sends three non-null fields to the application: player\_id, score, and us\_5\_digit\_zip\_code.

A data analyst has a .csv mapping file that maps a small number of us\_5\_digit\_zip\_code values to a territory code. The data analyst needs to include the territory code, if one exists, as an additional output of the Kinesis Data Analytics application.

How should the data analyst meet this requirement while minimizing costs?

- A. Store the contents of the mapping file in an Amazon DynamoDB table
- B. Preprocess the records as they arrive in the Kinesis Data Analytics application with an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exists
- C. Change the SQL query in the application to include the new field in the SELECT statement.
- D. Store the mapping file in an Amazon S3 bucket and configure the reference data column headers for the .csv file in the Kinesis Data Analytics application
- E. Change the SQL query in the application to include a join to the file's S3 Amazon Resource Name (ARN), and add the territory code field to the SELECT columns.
- F. Store the mapping file in an Amazon S3 bucket and configure it as a reference data source for the Kinesis Data Analytics application
- G. Change the SQL query in the application to include a join to the reference table and add the territory code field to the SELECT columns.
- H. Store the contents of the mapping file in an Amazon DynamoDB table
- I. Change the Kinesis Data Analytics application to send its output to an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exists
- J. Forward the record from the Lambda function to the original application destination.

**Answer:** C

#### NEW QUESTION 127

.....



## Thank You for Trying Our Product

\* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

\* One year free update

You can enjoy free update one year. 24x7 online support.

\* Trusted by Millions

We currently serve more than 30,000,000 customers.

\* Shop Securely

All transactions are protected by VeriSign!

**100% Pass Your DAS-C01 Exam with Our Prep Materials Via below:**

<https://www.certleader.com/DAS-C01-dumps.html>