

DAS-C01 Dumps

AWS Certified Data Analytics - Specialty

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



NEW QUESTION 1

A company that monitors weather conditions from remote construction sites is setting up a solution to collect temperature data from the following two weather stations.

Station A, which has 10 sensors

Station B, which has five sensors

These weather stations were placed by onsite subject-matter experts.

Each sensor has a unique ID. The data collected from each sensor will be collected using Amazon Kinesis Data Streams.

Based on the total incoming and outgoing data throughput, a single Amazon Kinesis data stream with two shards is created. Two partition keys are created based on the station names. During testing, there is a bottleneck on data coming from Station A, but not from Station B. Upon review, it is confirmed that the total stream throughput is still less than the allocated Kinesis Data Streams throughput.

How can this bottleneck be resolved without increasing the overall cost and complexity of the solution, while retaining the data collection quality requirements?

- A. Increase the number of shards in Kinesis Data Streams to increase the level of parallelism.
- B. Create a separate Kinesis data stream for Station A with two shards, and stream Station A sensor data to the new stream.
- C. Modify the partition key to use the sensor ID instead of the station name.
- D. Reduce the number of sensors in Station A from 10 to 5 sensors.

Answer: C

Explanation:

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding.html>

"Splitting increases the number of shards in your stream and therefore increases the data capacity of the stream. Because you are charged on a per-shard basis, splitting increases the cost of your stream"

NEW QUESTION 2

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 3

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 table
- 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 Catalog
- 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 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

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 6

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 7

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 operatio
- 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 8

A company launched a service that produces millions of messages every day and uses Amazon Kinesis Data Streams as the streaming service.

The company uses the Kinesis SDK to write data to Kinesis Data Streams. A few months after launch, a data analyst found that write performance is significantly reduced. The data analyst investigated the metrics and determined that Kinesis is throttling the write requests. The data analyst wants to address this issue without significant changes to the architecture.

Which actions should the data analyst take to resolve this issue? (Choose two.)

- A. Increase the Kinesis Data Streams retention period to reduce throttling.
- B. Replace the Kinesis API-based data ingestion mechanism with Kinesis Agent.
- C. Increase the number of shards in the stream using the UpdateShardCount API.
- D. Choose partition keys in a way that results in a uniform record distribution across shards.
- E. Customize the application code to include retry logic to improve performance.

Answer: CD

Explanation:

<https://aws.amazon.com/blogs/big-data/under-the-hood-scaling-your-kinesis-data-streams/>

NEW QUESTION 9

A hospital uses wearable medical sensor devices to collect data from patients. The hospital is architecting a near-real-time solution that can ingest the data securely at scale. The solution should also be able to remove the patient's protected health information (PHI) from the streaming data and store the data in durable storage.

Which solution meets these requirements with the least operational overhead?

- A. Ingest the data using Amazon Kinesis Data Streams, which invokes an AWS Lambda function using Kinesis Client Library (KCL) to remove all PH
- B. Write the data in Amazon S3.
- C. Ingest the data using Amazon Kinesis Data Firehose to write the data to Amazon S3. Have Amazon S3 trigger an AWS Lambda function that parses the sensor data to remove all PHI in Amazon S3.
- D. Ingest the data using Amazon Kinesis Data Streams to write the data to Amazon S3. Have the data stream launch an AWS Lambda function that parses the sensor data and removes all PHI in Amazon S3.
- E. Ingest the data using Amazon Kinesis Data Firehose to write the data to Amazon S3. Implement a transformation AWS Lambda function that parses the sensor data to remove all PHI.

Answer: D

Explanation:

<https://aws.amazon.com/blogs/big-data/persist-streaming-data-to-amazon-s3-using-amazon-kinesis-firehose-and>

NEW QUESTION 10

A data analytics specialist is building an automated ETL ingestion pipeline using AWS Glue to ingest compressed files that have been uploaded to an Amazon S3 bucket. The ingestion pipeline should support incremental data processing.

Which AWS Glue feature should the data analytics specialist use to meet this requirement?

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

Answer: C

NEW QUESTION 10

A company uses Amazon Redshift for its data warehousing needs. ETL jobs run every night to load data, apply business rules, and create aggregate tables for reporting. The company's data analysis, data science, and business intelligence teams use the data warehouse during regular business hours. The workload management is set to auto, and separate queues exist for each team with the priority set to NORMAL. Recently, a sudden spike of read queries from the data analysis team has occurred at least twice daily, and queries wait in line for cluster resources. The company needs a solution that enables the data analysis team to avoid query queuing without impacting latency and the query times of other teams. Which solution meets these requirements?

- A. Increase the query priority to HIGHEST for the data analysis queue.
- B. Configure the data analysis queue to enable concurrency scaling.
- C. Create a query monitoring rule to add more cluster capacity for the data analysis queue when queries are waiting for resources.
- D. Use workload management query queue hopping to route the query to the next matching queue.

Answer: D

NEW QUESTION 15

Three teams of data analysts use Apache Hive on an Amazon EMR cluster with the EMR File System (EMRFS) to query data stored within each team's Amazon S3 bucket. The EMR cluster has Kerberos enabled and is configured to authenticate users from the corporate Active Directory. The data is highly sensitive, so access must be limited to the members of each team. Which steps will satisfy the security requirements?

- A. For the EMR cluster Amazon EC2 instances, create a service role that grants no access to Amazon S3. Create three additional IAM roles, each granting access to each team's specific bucket.
- B. Add the additional IAM roles to the cluster's EMR role for the EC2 trust policy.
- C. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.
- D. For the EMR cluster Amazon EC2 instances, create a service role that grants no access to Amazon S3. Create three additional IAM roles, each granting access to each team's specific bucket.
- E. Add the service role for the EMR cluster EC2 instances to the trust policies for the additional IAM role.
- F. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.
- G. For the EMR cluster Amazon EC2 instances, create a service role that grants full access to Amazon S3. Create three additional IAM roles, each granting access to each team's specific bucket.
- H. Add the service role for the EMR cluster EC2 instances to the trust policies for the additional IAM role.
- I. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.
- J. For the EMR cluster Amazon EC2 instances, create a service role that grants full access to Amazon S3. Create three additional IAM roles, each granting access to each team's specific bucket.
- K. Add the service role for the EMR cluster EC2 instances to the trust policies for the base IAM role.
- L. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.

Answer: C

NEW QUESTION 19

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 a KEY distribution style for the transactions table

Answer: B

NEW QUESTION 24

A company is migrating from an on-premises Apache Hadoop cluster to an Amazon EMR cluster. The cluster runs only during business hours. Due to a company requirement to avoid intraday cluster failures, the EMR cluster must be highly available. When the cluster is terminated at the end of each business day, the data must persist.

Which configurations would enable the EMR cluster to meet these requirements? (Choose three.)

- A. EMR File System (EMRFS) for storage
- B. Hadoop Distributed File System (HDFS) for storage
- C. AWS Glue Data Catalog as the metastore for Apache Hive
- D. MySQL database on the master node as the metastore for Apache Hive
- E. Multiple master nodes in a single Availability Zone
- F. Multiple master nodes in multiple Availability Zones

Answer: ACE

Explanation:

<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-ha.html> "Note : The cluster can reside only in one Availability Zone or subnet."

NEW QUESTION 29

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 exist
- 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 applicatio
- 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 applicatio
- 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 tabl
- I. Change the Kinesis DataAnalytics 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 exist
- J. Forward the record from the Lambda function to the original application destination.

Answer: C

NEW QUESTION 30

A team of data scientists plans to analyze market trend data for their company's new investment strategy. The trend data comes from five different data sources in large volumes. The team wants to utilize Amazon Kinesis to support their use case. The team uses SQL-like queries to analyze trends and wants to send notifications based on certain significant patterns in the trends. Additionally, the data scientists want to save the data to Amazon S3 for archival and historical re-processing, and use AWS managed services wherever possible. The team wants to implement the lowest-cost solution. Which solution meets these requirements?

- A. Publish data to one Kinesis data strea
- B. Deploy a custom application using the Kinesis Client Library (KCL) for analyzing trends, and send notifications using Amazon SN
- C. Configure Kinesis Data Firehose on the Kinesis data stream to persist data to an S3 bucket.
- D. Publish data to one Kinesis data strea
- E. Deploy Kinesis Data Analytic to the stream for analyzing trends, and configure an AWS Lambda function as an output to send notifications using Amazon SN
- F. Configure Kinesis Data Firehose on the Kinesis data stream to persist data to an S3 bucket.
- G. Publish data to two Kinesis data stream
- H. Deploy Kinesis Data Analytics to the first stream for analyzing trends, and configure an AWS Lambda function as an output to send notifications using Amazon SN
- I. Configure Kinesis Data Firehose on the second Kinesis data stream to persist data to an S3 bucket.
- J. Publish data to two Kinesis data stream
- K. Deploy a custom application using the Kinesis Client Library (KCL) to the first stream for analyzing trends, and send notifications using Amazon SN
- L. Configure Kinesis Data Firehose on the second Kinesis data stream to persist data to an S3 bucket.

Answer: B

NEW QUESTION 35

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 HDFSConnector 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

NEW QUESTION 38

A manufacturing company has been collecting IoT sensor data from devices on its factory floor for a year and is storing the data in Amazon Redshift for daily analysis. A data analyst has determined that, at an expected ingestion rate of about 2 TB per day, the cluster will be undersized in less than 4 months. A long-term solution is needed. The data analyst has indicated that most queries only reference the most recent 13 months of data, yet there are also quarterly reports that need to query all the data generated from the past 7 years. The chief technology officer (CTO) is concerned about the costs, administrative effort, and performance of a long-term solution.

Which solution should the data analyst use to meet these requirements?

- A. Create a daily job in AWS Glue to UNLOAD records older than 13 months to Amazon S3 and delete those records from Amazon Redshif
- B. Create an external table in Amazon Redshift to point to the S3 locatio
- C. Use Amazon Redshift Spectrum to join to data that is older than 13 months.
- D. Take a snapshot of the Amazon Redshift cluste
- E. Restore the cluster to a new cluster using dense storage nodes with additional storage capacity.
- F. Execute a CREATE TABLE AS SELECT (CTAS) statement to move records that are older than 13 months to quarterly partitioned data in Amazon Redshift Spectrum backed by Amazon S3.
- G. Unload all the tables in Amazon Redshift to an Amazon S3 bucket using S3 Intelligent-Tierin
- H. Use AWS Glue to crawl the S3 bucket location to create external tables in an AWS Glue Data Catalo
- I. Create an Amazon EMR cluster using Auto Scaling for any daily analytics needs, and use Amazon Athena for the quarterly reports, with both using the same AWS Glue Data Catalog.

Answer: A

NEW QUESTION 40

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 basis
- B. Query the data as required.
- C. Disable encryption on the Amazon Redshift cluster, configure audit logging, and encrypt the Amazon Redshift cluster
- 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 45

A healthcare company uses AWS data and analytics tools to collect, ingest, and store electronic health record (EHR) data about its patients. The raw EHR data is stored in Amazon S3 in JSON format partitioned by hour, day, and year and is updated every hour. The company wants to maintain the data catalog and metadata in an AWS Glue Data Catalog to be able to access the data using Amazon Athena or Amazon Redshift Spectrum for analytics.

When defining tables in the Data Catalog, the company has the following requirements:

Choose the catalog table name and do not rely on the catalog table naming algorithm. Keep the table updated with new partitions loaded in the respective S3 bucket prefixes.

Which solution meets these requirements with minimal effort?

- A. Run an AWS Glue crawler that connects to one or more data stores, determines the data structures, and writes tables in the Data Catalog.
- B. Use the AWS Glue console to manually create a table in the Data Catalog and schedule an AWS Lambda function to update the table partitions hourly.
- C. Use the AWS Glue API CreateTable operation to create a table in the Data Catalog
- D. Create an AWS Glue crawler and specify the table as the source.
- E. Create an Apache Hive catalog in Amazon EMR with the table schema definition in Amazon S3, and update the table partition with a scheduled job
- F. Migrate the Hive catalog to the Data Catalog.

Answer: C

Explanation:

Updating Manually Created Data Catalog Tables Using Crawlers: To do this, when you define a crawler, instead of specifying one or more data stores as the source of a crawl, you specify one or more existing Data Catalog tables. The crawler then crawls the data stores specified by the catalog tables. In this case, no new tables are created; instead, your manually created tables are updated.

NEW QUESTION 47

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 52

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 57

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 CustomAmild property in the CloudFormation template.
- G. Use a CloudFormation template to launch an EMR cluste
- 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 62

A company wants to improve the data load time of a sales data dashboard. Data has been collected as .csv files and stored within an Amazon S3 bucket that is partitioned by date. The data is then loaded to an Amazon Redshift data warehouse for frequent analysis. The data volume is up to 500 GB per day. Which solution will improve the data loading performance?

- A. Compress .csv files and use an INSERT statement to ingest data into Amazon Redshift.
- B. Split large .csv files, then use a COPY command to load data into Amazon Redshift.
- C. Use Amazon Kinesis Data Firehose to ingest data into Amazon Redshift.
- D. Load the .csv files in an unsorted key order and vacuum the table in Amazon Redshift.

Answer: B

Explanation:

https://docs.aws.amazon.com/redshift/latest/dg/c_loading-data-best-practices.html

NEW QUESTION 67

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 scrip
- B. Push the enriched data to a fleet of Kinesis data streams and enable the data transformation feature to flatten the JSON fil
- 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 Jav
- E. Use Kinesis Data Analytics to enrich the data based on a company-developed anomaly detection SQL scrip
- 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 bucke
- 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 Jav
- 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 results in higher packing efficiencies and better performance. Applications that cannot tolerate this additional delay may need to use the AWS SDK directly.

NEW QUESTION 69

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 70

.....

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>