Amazon-Web-Services

Exam Questions DAS-C01

AWS Certified Data Analytics - Specialty

NEW QUESTION 1
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 2
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 3
A company owns facilities with IoT devices installed across the world. The company is using Amazon Kinesis Data Streams to stream data from the devices to Amazon S3. The company's operations team wants to get insights from the IoT data to monitor data quality at ingestion. The insights need to be derived in near-real time, and the output must be logged to Amazon DynamoDB for further analysis.
Which solution meets these requirements?

A. Connect Amazon Kinesis Data Analytics to analyze the stream dat
B. Save the output to DynamoDB by using the default output from Kinesis Data Analytics.
C. Connect Amazon Kinesis Data Analytics to analyze the stream dat
D. Save the output to DynamoDB by using an AWS Lambda function.
E. Connect Amazon Kinesis Data Firehose to analyze the stream data by using an AWS Lambda function.Save the output to DynamoDB by using the default output from Kinesis Data Firehose.
F. Connect Amazon Kinesis Data Firehose to analyze the stream data by using an AWS Lambda function.Save the data to Amazon S3. Then run an AWS Glue job on schedule to ingest the data into DynamoDB.

Answer: C

NEW QUESTION 4
A company wants to optimize the cost of its data and analytics platform. The company is ingesting a number of .c sv 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 forma
B. Use Athena to query the processed datase
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 forma
E. Use Athena to query the processed datase
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 creatio
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 forma
I. Use Athena to query the processed datase
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 accesse
K. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier forlong-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 forma
M. Use Athena to query the processed datase
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 accesse
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 5
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 accounts CloudWatch Logs to target the Kinesis data stream in the test account as its destination In the test account create an 1AM 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 1AM 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 1AM 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 1AM 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 1AM 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 1AM 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 accounts CloudWatch Logs to target the Kinesis data stream in the test account as its destination

Answer: D

NEW QUESTION 6
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 metadat
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 7
A retail company has 15 stores across 6 cities in the United States. Once a month, the sales team requests a visualization in Amazon QuickSight that provides the ability to easily identify revenue trends across cities and stores. The visualization also helps identify outliers that need to be examined with further analysis.
Which visual type in QuickSight meets the sales team's requirements?

A. Geospatial chart
B. Line chart
C. Heat map
D. Tree map

Answer: A

NEW QUESTION 8
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 9
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 media company is using Amazon QuickSight dashboards to visualize its national sales data. The dashboard is using a dataset with these fields: ID, date, time_zone, city, state, country, longitude, latitude, sales_volume, and number_of_items.
To modify ongoing campaigns, the company wants an interactive and intuitive visualization of which states across the country recorded a significantly lower sales volume compared to the national average.
Which addition to the company's QuickSight dashboard will meet this requirement?

A. A geospatial color-coded chart of sales volume data across the country.
B. A pivot table of sales volume data summed up at the state level.
C. A drill-down layer for state-level sales volume data.
D. A drill through to other dashboards containing state-level sales volume data.

Answer: B

NEW QUESTION 10
A hospital is building a research data lake to ingest data from electronic health records (EHR) systems from multiple hospitals and clinics. The EHR systems are independent of each other and do not have a common patient identifier. The data engineering team is not experienced in machine learning (ML) and has been asked to generate a unique patient identifier for the ingested records.
Which solution will accomplish this task?

A. An AWS Glue ETL job with the FindMatches transform
B. Amazon Kendra
C. Amazon SageMaker Ground Truth
D. An AWS Glue ETL job with the ResolveChoice transform

Answer: A

Explanation:
Matching Records with AWS Lake Formation FindMatches

NEW QUESTION 13
A gaming company is collecting cllckstream data into multiple Amazon Kinesis data streams. The company uses Amazon Kinesis Data Firehose delivery streams to store the data in JSON format in Amazon S3 Data scientists use Amazon Athena to query the most recent data and derive business insights. The company wants to reduce its Athena costs without having to recreate the data pipeline. The company prefers a solution that will require less management effort
Which set of actions can the data scientists take immediately to reduce costs?

A. Change the Kinesis Data Firehose output format to Apache Parquet Provide a custom S3 object YYYYMMDD prefix expression and specify a large buffer size For the existing data, run an AWS Glue ETL job to combine and convert small JSON files to large Parquet files and add the YYYYMMDD prefix Use ALTER TABLE ADD PARTITION to reflect the partition on the existing Athena table.
B. Create an Apache Spark Job that combines and converts JSON files to Apache Parquet files Launch an Amazon EMR ephemeral cluster daily to run the Spark job to create new Parquet files in a different S3 location Use ALTER TABLE SET LOCATION to reflect the new S3 location on the existing Athena table.
C. Create a Kinesis data stream as a delivery target for Kinesis Data Firehose Run Apache Flink on Amazon Kinesis Data Analytics on the stream to read the streaming data, aggregate ikand save it to Amazon S3 in Apache Parquet format with a custom S3 object YYYYMMDD prefix Use ALTER TABLE ADD PARTITION to reflect the partition on the existing Athena table
D. Integrate an AWS Lambda function with Kinesis Data Firehose to convert source records to Apache Parquet and write them to Amazon S3 In parallel, run an AWS Glue ETL job to combine and convert existing JSON files to large Parquet files Create a custom S3 object YYYYMMDD prefix Use ALTER TABLE ADD PARTITION to reflect the partition on the existing Athena table.

Answer: D

NEW QUESTION 17
A media analytics company consumes a stream of social media posts. The posts are sent to an Amazon Kinesis data stream partitioned on user_id. An AWS Lambda function retrieves the records and validates the content before loading the posts into an Amazon Elasticsearch cluster. The validation process needs to receive the posts for a given user in the order they were received. A data analyst has noticed that, during peak hours, the social media platform posts take more than an hour to appear in the Elasticsearch cluster.
What should the data analyst do reduce this latency?

A. Migrate the validation process to Amazon Kinesis Data Firehose.
B. Migrate the Lambda consumers from standard data stream iterators to an HTTP/2 stream consumer.
C. Increase the number of shards in the stream.
D. Configure multiple Lambda functions to process the stream.

Answer: D

NEW QUESTION 20
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 24
A company using Amazon QuickSight Enterprise edition has thousands of dashboards analyses and datasets. The company struggles to manage and assign permissions for granting users access to various items within QuickSight. The company wants to make it easier to implement sharing and permissions management.
Which solution should the company implement to simplify permissions management?

A. Use QuickSight folders to organize dashboards, analyses, and datasets Assign individual users permissions to these folders
B. Use QuickSight folders to organize dashboards analyses, and datasets Assign group permissions by using these folders.
C. Use AWS 1AM resource-based policies to assign group permissions to QuickSight items
D. Use QuickSight user management APIs to provision group permissions based on dashboard naming conventions

Answer: C

NEW QUESTION 25
A company with a video streaming website wants to analyze user behavior to make recommendations to users in real time Clickstream data is being sent to Amazon Kinesis Data Streams and reference data is stored in Amazon S3 The company wants a solution that can use standard SQL quenes The solution must also provide a way to look up pre-calculated reference data while making recommendations
Which solution meets these requirements?

A. Use an AWS Glue Python shell job to process incoming data from Kinesis Data Streams Use the Boto3 library to write data to Amazon Redshift
B. Use AWS Glue streaming and Scale to process incoming data from Kinesis Data Streams Use the AWS Glue connector to write data to Amazon Redshift
C. Use Amazon Kinesis Data Analytics to create an in-application table based upon the reference data Process incoming data from Kinesis Data Streams Use a data stream to write results to Amazon Redshift
D. Use Amazon Kinesis Data Analytics to create an in-application table based upon the reference dataProcess incoming data from Kinesis Data Streams Use an Amazon Kinesis Data Firehose delivery stream to write results to Amazon Redshift

Answer: D

NEW QUESTION 26
A company has a business unit uploading .csv files to an Amazon S3 bucket. The company's data platform team has set up an AWS Glue crawler to do discovery, and create tables and schemas. An AWS Glue job writes processed data from the created tables to an Amazon Redshift database. The AWS Glue job handles column mapping and creating the Amazon Redshift table appropriately. When the AWS Glue job is rerun for any reason in a day, duplicate records are introduced into the Amazon Redshift table.
Which solution will update the Redshift table without duplicates when jobs are rerun?

A. Modify the AWS Glue job to copy the rows into a staging tabl
B. Add SQL commands to replace the existing rows in the main table as postactions in the DynamicFrameWriter class.
C. Load the previously inserted data into a MySQL database in the AWS Glue jo
D. Perform an upsert operation in MySQL, and copy the results to the Amazon Redshift table.
E. Use Apache Spark's DataFrame dropDuplicates() API to eliminate duplicates and then write the data to Amazon Redshift.
F. Use the AWS Glue ResolveChoice built-in transform to select the most recent value of the column.

Answer: A

Explanation:
https://aws.amazon.com/premiumsupport/knowledge-center/sql-commands-redshift-glue-job/ See the section Merge an Amazon Redshift table in AWS Glue (upsert)

NEW QUESTION 31
A company wants to use an automatic machine learning (ML) Random Cut Forest (RCF) algorithm to visualize complex real-world scenarios, such as detecting seasonality and trends, excluding outers, and imputing missing values.
The team working on this project is non-technical and is looking for an out-of-the-box solution that will require the LEAST amount of management overhead.
Which solution will meet these requirements?

A. Use an AWS Glue ML transform to create a forecast and then use Amazon QuickSight to visualize the data.
B. Use Amazon QuickSight to visualize the data and then use ML-powered forecasting to forecast the key business metrics.
C. Use a pre-build ML AMI from the AWS Marketplace to create forecasts and then use Amazon QuickSight to visualize the data.
D. Use calculated fields to create a new forecast and then use Amazon QuickSight to visualize the data.

Answer: A

NEW QUESTION 33
An online retail company with millions of users around the globe wants to improve its ecommerce analytics capabilities. Currently, clickstream data is uploaded directly to Amazon S3 as compressed files. Several times each day, an application running on Amazon EC2 processes the data and makes search options and reports available for visualization by editors and marketers. The company wants to make website clicks and aggregated data available to editors and marketers in minutes to enable them to connect with users more effectively.
Which options will help meet these requirements in the MOST efficient way? (Choose two.)

A. Use Amazon Kinesis Data Firehose to upload compressed and batched clickstream records to Amazon Elasticsearch Service.
B. Upload clickstream records to Amazon S3 as compressed file
C. Then use AWS Lambda to send data to Amazon Elasticsearch Service from Amazon S3.
D. Use Amazon Elasticsearch Service deployed on Amazon EC2 to aggregate, filter, and process the data.Refresh content performance dashboards in near-real time.

E. Use Kibana to aggregate, filter, and visualize the data stored in Amazon Elasticsearch Servic
F. Refresh content performance dashboards in near-real time.
G. Upload clickstream records from Amazon S3 to Amazon Kinesis Data Streams and use a Kinesis Data Streams consumer to send records to Amazon Elasticsearch Service.

Answer: AD

NEW QUESTION 38
A company receives data from its vendor in JSON format with a timestamp in the file name. The vendor uploads the data to an Amazon S3 bucket, and the data is registered into the company's data lake for analysis and reporting. The company has configured an S3 Lifecycle policy to archive all files to S3 Glacier after 5 days.
The company wants to ensure that its AWS Glue crawler catalogs data only from S3 Standard storage and ignores the archived files. A data analytics specialist must implement a solution to achieve this goal without changing the current S3 bucket configuration.
Which solution meets these requirements?

A. Use the exclude patterns feature of AWS Glue to identify the S3 Glacier files for the crawler to exclude.
B. Schedule an automation job that uses AWS Lambda to move files from the original S3 bucket to a new S3 bucket for S3 Glacier storage.
C. Use the excludeStorageClasses property in the AWS Glue Data Catalog table to exclude files on S3 Glacier storage
D. Use the include patterns feature of AWS Glue to identify the S3 Standard files for the crawler to include.

Answer: A

NEW QUESTION 42
An education provider's learning management system (LMS) is hosted in a 100 TB data lake that is built on Amazon S3. The provider's LMS supports hundreds of schools. The provider wants to build an advanced analytics reporting platform using Amazon Redshift to handle complex queries with optimal performance.
System users will query the most recent 4 months of data 95% of the time while 5% of the queries will leverage data from the previous 12 months.
Which solution meets these requirements in the MOST cost-effective way?

A. Store the most recent 4 months of data in the Amazon Redshift cluste
B. Use Amazon Redshift Spectrum to query data in the data lak
C. Use S3 lifecycle management rules to store data from the previous 12 months in Amazon S3 Glacier storage.
D. Leverage DS2 nodes for the Amazon Redshift cluste
E. Migrate all data from Amazon S3 to Amazon Redshif
F. Decommission the data lake.
G. Store the most recent 4 months of data in the Amazon Redshift cluste
H. Use Amazon Redshift Spectrum to query data in the data lak
I. Ensure the S3 Standard storage class is in use with objects in the data lake.
J. Store the most recent 4 months of data in the Amazon Redshift cluste
K. Use Amazon Redshift federated queries to join cluster data with the data lake to reduce cost
L. Ensure the S3 Standard storage class is in use with objects in the data lake.

Answer: C

NEW QUESTION 43
A company uses Amazon Elasticsearch Service (Amazon ES) to store and analyze its website clickstream data. The company ingests 1 TB of data daily using Amazon Kinesis Data Firehose and stores one day's worth of data in an Amazon ES cluster.
The company has very slow query performance on the Amazon ES index and occasionally sees errors from Kinesis Data Firehose when attempting to write to the index. The Amazon ES cluster has 10 nodes running a single index and 3 dedicated master nodes. Each data node has 1.5 TB of Amazon EBS storage attached and the cluster is configured with 1,000 shards. Occasionally, JVMMemoryPressure errors are found in the cluster logs.
Which solution will improve the performance of Amazon ES?

A. Increase the memory of the Amazon ES master nodes.
B. Decrease the number of Amazon ES data nodes.
C. Decrease the number of Amazon ES shards for the index.
D. Increase the number of Amazon ES shards for the index.

Answer: C

Explanation:
https://aws.amazon.com/premiumsupport/knowledge-center/high-jvm-memory-pressure-elasticsearch/

NEW QUESTION 45
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 47
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 49
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 though 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 cluste
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 50
A media content company has a streaming playback application. The company wants to collect and analyze the data to provide near-real-time feedback on playback issues. The company needs to consume this data and return results within 30 seconds according to the service-level agreement (SLA). The company needs the consumer to identify playback issues, such as quality during a specified timeframe. The data will be emitted as JSON and may change schemas over time.
Which solution will allow the company to collect data for processing while meeting these requirements?

A. Send the data to Amazon Kinesis Data Firehose with delivery to Amazon S3. Configure an S3 event trigger an AWS Lambda function to process the dat
B. The Lambda function will consume the data and process it to identify potential playback issue
C. Persist the raw data to Amazon S3.
D. Send the data to Amazon Managed Streaming for Kafka and configure an Amazon Kinesis Analytics for Java application as the consume
E. The application will consume the data and process it to identify potential playback issue
F. Persist the raw data to Amazon DynamoDB.
G. Send the data to Amazon Kinesis Data Firehose with delivery to Amazon S3. Configure Amazon S3 to trigger an event for AWS Lambda to proces
H. The Lambda function will consume the data and process it to identify potential playback issue
I. Persist the raw data to Amazon DynamoDB.
J. Send the data to Amazon Kinesis Data Streams and configure an Amazon Kinesis Analytics for Java application as the consume
K. The application will consume the data and process it to identify potential playback issue
L. Persist the raw data to Amazon S3.

Answer: D

Explanation:
https://aws.amazon.com/blogs/aws/new-amazon-kinesis-data-analytics-for-java/

NEW QUESTION 52
An online retailer is rebuilding its inventory management system and inventory reordering system to automatically reorder products by using Amazon Kinesis Data Streams. The inventory management system uses the Kinesis Producer Library (KPL) to publish data to a stream. The inventory reordering system uses the Kinesis Client Library (KCL) to consume data from the stream. The stream has been configured to scale as needed. Just before production deployment, the retailer discovers that the inventory reordering system is receiving duplicated data.
Which factors could be causing the duplicated data? (Choose two.)

A. The producer has a network-related timeout.
B. The stream's value for the IteratorAgeMilliseconds metric is too high.
C. There was a change in the number of shards, record processors, or both.
D. The AggregationEnabled configuration property was set to true.
E. The max_records configuration property was set to a number that is too high.

Answer: BD

NEW QUESTION 56
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 Parque
B. Use AWS Glue crawlers todiscover the schema and build the AWS Glue Data Catalo
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 Parque
H. Use AWS Glue crawlers to discover the schema and build the AWS Glue Data Catalo
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 61
A data analyst runs a large number of data manipulation language (DML) queries by using Amazon Athena with the JDBC driver. Recently, a query failed after It ran for 30 minutes. The query returned the following message Java.sql.SGLException: Query timeout
The data analyst does not immediately need the query results However, the data analyst needs a long-term solution for this problem
Which solution will meet these requirements?

A. Split the query into smaller queries to search smaller subsets of data.
B. In the settings for Athena, adjust the DML query timeout limit
C. In the Service Quotas console, request an increase for the DML query timeout
D. Save the tables as compressed .csv files

Answer: A

NEW QUESTION 64
A data engineer is using AWS Glue ETL jobs to process data at frequent intervals The processed data is then copied into Amazon S3 The ETL jobs run every 15 minutes. The AWS Glue Data Catalog partitions need to be updated automatically after the completion of each job
Which solution will meet these requirements MOST cost-effectively?

A. Use the AWS Glue Data Catalog to manage the data catalog Define an AWS Glue workflow for the ETL process Define a trigger within the workflow that can start the crawler when an ETL job run is complete
B. Use the AWS Glue Data Catalog to manage the data catalog Use AWS Glue Studio to manage ETL job
C. Use the AWS Glue Studio feature that supports updates to the AWS Glue Data Catalog during job runs.
D. Use an Apache Hive metastore to manage the data catalog Update the AWS Glue ETL code to include the enableUpdateCatalog and partitionKeys arguments.
E. Use the AWS Glue Data Catalog to manage the data catalog Update the AWS Glue ETL code to include the enableUpdateCatalog and partitionKeys arguments.

Answer: A

NEW QUESTION 65
A data analytics specialist is setting up workload management in manual mode for an Amazon Redshift environment. The data analytics specialist is defining query monitoring rules to manage system performance and user experience of an Amazon Redshift cluster.
Which elements must each query monitoring rule include?

A. A unique rule name, a query runtime condition, and an AWS Lambda function to resubmit any failed queries in off hours
B. A queue name, a unique rule name, and a predicate-based stop condition
C. A unique rule name, one to three predicates, and an action
D. A workload name, a unique rule name, and a query runtime-based condition

Answer: C

NEW QUESTION 66
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/

NEW QUESTION 70
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 72
A bank is using Amazon Managed Streaming for Apache Kafka (Amazon MSK) to populate real-time data into a data lake The data lake is built on Amazon S3, and data must be accessible from the data lake within 24 hours Different microservices produce messages to different topics in the cluster The cluster is created with 8 TB of Amazon Elastic Block Store (Amazon EBS) storage and a retention period of 7 days
The customer transaction volume has tripled recently and disk monitoring has provided an alert that the cluster is almost out of storage capacity
What should a data analytics specialist do to prevent the cluster from running out of disk space1?

A. Use the Amazon MSK console to triple the broker storage and restart the cluster
B. Create an Amazon CloudWatch alarm that monitors the KafkaDataLogsDiskUsed metric Automaticallyflush the oldest messages when the value of this metric exceeds 85%
C. Create a custom Amazon MSK configuration Set the log retention hours parameter to 48 Update the cluster with the new configuration file
D. Triple the number of consumers to ensure that data is consumed as soon as it is added to a topic.

Answer: B

NEW QUESTION 73
A marketing company has data in Salesforce, MySQL, and Amazon S3. The company wants to use data from these three locations and create mobile dashboards for its users. The company is unsure how it should create the dashboards and needs a solution with the least possible customization and coding.
Which solution meets these requirements?

A. Use Amazon Athena federated queries to join the data source
B. Use Amazon QuickSight to generate the mobile dashboards.
C. Use AWS Lake Formation to migrate the data sources into Amazon S3. Use Amazon QuickSight to generate the mobile dashboards.
D. Use Amazon Redshift federated queries to join the data source
E. Use Amazon QuickSight to generate the mobile dashboards.
F. Use Amazon QuickSight to connect to the data sources and generate the mobile dashboards.

Answer: C

NEW QUESTION 78
A company uses an Amazon EMR cluster with 50 nodes to process operational data and make the data available for data analysts These jobs run nightly use Apache Hive with the Apache Jez framework as a processing model and write results to Hadoop Distributed File System (HDFS) In the last few weeks, jobs are failing and are producing the following error message
"File could only be replicated to 0 nodes instead of 1"
A data analytics specialist checks the DataNode logs the NameNode logs and network connectivity for potential issues that could have prevented HDFS from replicating data The data analytics specialist rules out these factors as causes for the issue
Which solution will prevent the jobs from failing'?

A. Monitor the HDFSUtilization metri
B. If the value crosses a user-defined threshold add task nodes to the EMR cluster
C. Monitor the HDFSUtilization metri.c If the value crosses a user-defined threshold add core nodes to the EMR cluster
D. Monitor the MemoryAllocatedMB metri
E. If the value crosses a user-defined threshold, add task nodes to the EMR cluster
F. Monitor the MemoryAllocatedMB metri
G. If the value crosses a user-defined threshold, add core nodes to the EMR cluster.

Answer: C

NEW QUESTION 83
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 CloudWatc
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

NEW QUESTION 85
A company is building a data lake and needs to ingest data from a relational database that has time-series data. The company wants to use managed services to accomplish this. The process needs to be scheduled daily and bring incremental data only from the source into Amazon S3.
What is the MOST cost-effective approach to meet these requirements?

A. Use AWS Glue to connect to the data source using JDBC Driver
B. Ingest incremental records only using job bookmarks.
C. Use AWS Glue to connect to the data source using JDBC Driver
D. Store the last updated key in an Amazon DynamoDB table and ingest the data using the updated key as a filter.
E. Use AWS Glue to connect to the data source using JDBC Drivers and ingest the entire datase
F. Use appropriate Apache Spark libraries to compare the dataset, and find the delta.
G. Use AWS Glue to connect to the data source using JDBC Drivers and ingest the full dat
H. Use AWS DataSync to ensure the delta only is written into Amazon S3.

Answer: A

Explanation:
https://docs.aws.amazon.com/glue/latest/dg/monitor-continuations.html

NEW QUESTION 90
A large company receives files from external parties in Amazon EC2 throughout the day. At the end of the day, the files are combined into a single file, compressed into a gzip file, and uploaded to Amazon S3. The total size of all the files is close to 100 GB daily. Once the files are uploaded to Amazon S3, an AWS Batch program executes a COPY command to load the files into an Amazon Redshift cluster.
Which program modification will accelerate the COPY process?

A. Upload the individual files to Amazon S3 and run the COPY command as soon as the files become available.
B. Split the number of files so they are equal to a multiple of the number of slices in the Amazon Redshift cluste
C. Gzip and upload the files to Amazon S3. Run the COPY command on the files.
D. Split the number of files so they are equal to a multiple of the number of compute nodes in the Amazon Redshift cluste
E. Gzip and upload the files to Amazon S3. Run the COPY command on the files.
F. Apply sharding by breaking up the files so the distkey columns with the same values go to the same file.Gzip and upload the sharded files to Amazon S3. Run the COPY command on the files.

Answer: B

NEW QUESTION 95
A company uses Amazon Redshift as its data warehouse. A new table has columns that contain sensitive data. The data in the table will eventually be referenced by several existing queries that run many times a day.
A data analyst needs to load 100 billion rows of data into the new table. Before doing so, the data analyst must ensure that only members of the auditing group can read the columns containing sensitive data.
How can the data analyst meet these requirements with the lowest maintenance overhead?

A. Load all the data into the new table and grant the auditing group permission to read from the tabl
B. Load all the data except for the columns containing sensitive data into a second tabl
C. Grant the appropriate users read-only permissions to the second table.
D. Load all the data into the new table and grant the auditing group permission to read from the tabl
E. Use the GRANT SQL command to allow read-only access to a subset of columns to the appropriate users.
F. Load all the data into the new table and grant all users read-only permissions to non-sensitive columns.Attach an IAM policy to the auditing group with explicit ALLOW access to the sensitive data columns.
G. Load all the data into the new table and grant the auditing group permission to read from the table.Create a view of the new table that contains all the columns, except for those considered sensitive, and grant the appropriate users read-only permissions to the table.

Answer: B

Explanation:
https://aws.amazon.com/blogs/big-data/achieve-finer-grained-data-security-with-column-level-access-control-in

NEW QUESTION 99
A company is streaming its high-volume billing data (100 MBps) to Amazon Kinesis Data Streams. A data analyst partitioned the data on account_id to ensure that all records belonging to an account go to the same Kinesis shard and order is maintained. While building a custom consumer using the Kinesis Java SDK, the data analyst notices that, sometimes, the messages arrive out of order for account_id. Upon further investigation, the data analyst discovers the messages that are out of order seem to be arriving from different shards for the same account_id and are seen when a stream resize runs.
What is an explanation for this behavior and what is the solution?

A. There are multiple shards in a stream and order needs to be maintained in the shar
B. The data analyst needs to make sure there is only a single shard in the stream and no stream resize runs.
C. The hash key generation process for the records is not working correctl
D. The data analyst should generate an explicit hash key on the producer side so the records are directed to the appropriate shard accurately.
E. The records are not being received by Kinesis Data Streams in orde
F. The producer should use the PutRecords API call instead of the PutRecord API call with the SequenceNumberForOrdering parameter.

G. The consumer is not processing the parent shard completely before processing the child shards after a stream resiz
H. The data analyst should process the parent shard completely first before processing the child shards.

Answer: D

Explanation:
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html the parent shards that remain after the reshard could still contain data that you haven't read yet that was added to the stream before the reshard. If you read data from the child shards before having read all data from the parent shards, you could read data for a particular hash key out of the order given by the data records' sequence numbers.
Therefore, assuming that the order of the data is important, you should, after a reshard, always continue to read data from the parent shards until it is exhausted. Only then should you begin reading data from the child shards.

NEW QUESTION 104
A global pharmaceutical company receives test results for new drugs from various testing facilities worldwide. The results are sent in millions of 1 KB-sized JSON objects to an Amazon S3 bucket owned by the company. The data engineering team needs to process those files, convert them into Apache Parquet format, and load them into Amazon Redshift for data analysts to perform dashboard reporting. The engineering team uses AWS Glue to process the objects, AWS Step Functions for process orchestration, and Amazon CloudWatch for job scheduling.
More testing facilities were recently added, and the time to process files is increasing. What will MOST efficiently decrease the data processing time?

A. Use AWS Lambda to group the small files into larger file
B. Write the files back to Amazon S3. Process the files using AWS Glue and load them into Amazon Redshift tables.
C. Use the AWS Glue dynamic frame file grouping option while ingesting the raw input file
D. Process the files and load them into Amazon Redshift tables.
E. Use the Amazon Redshift COPY command to move the files from Amazon S3 into Amazon Redshift tables directl
F. Process the files in Amazon Redshift.
G. Use Amazon EMR instead of AWS Glue to group the small input file
H. Process the files in Amazon EMR and load them into Amazon Redshift tables.

Answer: A

NEW QUESTION 108
A company is planning to do a proof of concept for a machine learning (ML) project using Amazon SageMaker with a subset of existing on-premises data hosted in the company's 3 TB data warehouse. For part of the project, AWS Direct Connect is established and tested. To prepare the data for ML, data analysts are performing data curation. The data analysts want to perform multiple step, including mapping, dropping null fields, resolving choice, and splitting fields. The company needs the fastest solution to curate the data for this project.
Which solution meets these requirements?

A. Ingest data into Amazon S3 using AWS DataSync and use Apache Spark scrips to curate the data in an Amazon EMR cluste
B. Store the curated data in Amazon S3 for ML processing.
C. Create custom ETL jobs on-premises to curate the dat
D. Use AWS DMS to ingest data into Amazon S3 for ML processing.
E. Ingest data into Amazon S3 using AWS DM
F. Use AWS Glue to perform data curation and store the data in Amazon S3 for ML processing.
G. Take a full backup of the data store and ship the backup files using AWS Snowbal
H. Upload Snowball data into Amazon S3 and schedule data curation jobs using AWS Batch to prepare the data for ML.

Answer: C

NEW QUESTION 109
A company has collected more than 100 TB of log files in the last 24 months. The files are stored as raw text in a dedicated Amazon S3 bucket. Each object has a key of the form year-month-day_log_HHmmss.txt where HHmmss represents the time the log file was initially created. A table was created in Amazon Athena that points to the S3 bucket. One-time queries are run against a subset of columns in the table several times an hour.
A data analyst must make changes to reduce the cost of running these queries. Management wants a solution with minimal maintenance overhead.
Which combination of steps should the data analyst take to meet these requirements? (Choose three.)

A. Convert the log files to Apace Avro format.
B. Add a key prefix of the form date=year-month-day/ to the S3 objects to partition the data.
C. Convert the log files to Apache Parquet format.
D. Add a key prefix of the form year-month-day/ to the S3 objects to partition the data.
E. Drop and recreate the table with the PARTITIONED BY claus
F. Run the ALTER TABLE ADD PARTITION statement.
G. Drop and recreate the table with the PARTITIONED BY claus
H. Run the MSCK REPAIR TABLE statement.

Answer: BCF

NEW QUESTION 112
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 113
A company uses the Amazon Kinesis SDK to write data to Kinesis Data Streams. Compliance requirements state that the data must be encrypted at rest using a key that can be rotated. The company wants to meet this encryption requirement with minimal coding effort.
How can these requirements be met?

A. Create a customer master key (CMK) in AWS KM
B. Assign the CMK an alia
C. Use the AWS Encryption SDK, providing it with the key alias to encrypt and decrypt the data.
D. Create a customer master key (CMK) in AWS KM
E. Assign the CMK an alia
F. Enable server-side encryption on the Kinesis data stream using the CMK alias as the KMS master key.
G. Create a customer master key (CMK) in AWS KM
H. Create an AWS Lambda function to encrypt and decrypt the dat
I. Set the KMS key ID in the function's environment variables.
J. Enable server-side encryption on the Kinesis data stream using the default KMS key for Kinesis Data Streams.

Answer: B

NEW QUESTION 116
A company operates toll services for highways across the country and collects data that is used to understand usage patterns. Analysts have requested the ability to run traffic reports in near-real time. The company is interested in building an ingestion pipeline that loads all the data into an Amazon Redshift cluster and alerts operations personnel when toll traffic for a particular toll station does not meet a specified threshold. Station data and the corresponding threshold values are stored in Amazon S3.
Which approach is the MOST efficient way to meet these requirements?

A. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneousl
B. Create a reference data source in Kinesis Data Analytics to temporarily store the threshold values from Amazon S3 and compare the count of vehicles for a particular toll station against its corresponding threshold valu
C. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
D. Use Amazon Kinesis Data Streams to collect all the data from toll station
E. Create a stream in Kinesis Data Streams to temporarily store the threshold values from Amazon S3. Send both streams to Amazon Kinesis Data Analytics to compare the count of vehicles for a particular toll station against its corresponding threshold valu
F. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not me
G. Connect Amazon Kinesis Data Firehose to Kinesis Data Streams to deliver the data to Amazon Redshift.
H. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshif
I. Then, automatically trigger an AWS Lambda function that queries the data in Amazon Redshift, compares the count of vehicles for a particular toll station against its corresponding threshold values read from Amazon S3, and publishes an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
J. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneousl
K. Use Kinesis Data Analytics to compare the count of vehicles against the threshold value for the station stored in a table as an in-application stream based on information stored in Amazon S3. Configure an AWS Lambda function as an output for the application that will publish an Amazon Simple Queue Service (Amazon SQS) notification to alert operations personnel if the threshold is not met.

Answer: D

NEW QUESTION 117
A bank operates in a regulated environment. The compliance requirements for the country in which the bank operates say that customer data for each state should only be accessible by the bank's employees located in the same state. Bank employees in one state should NOT be able to access data for customers who have provided a home address in a different state.
The bank's marketing team has hired a data analyst to gather insights from customer data for a new campaign being launched in certain states. Currently, data linking each customer account to its home state is stored in a tabular .csv file within a single Amazon S3 folder in a private S3 bucket. The total size of the S3 folder is 2 GB uncompressed. Due to the country's compliance requirements, the marketing team is not able to access this folder.
The data analyst is responsible for ensuring that the marketing team gets one-time access to customer data for their campaign analytics project, while being subject to all the compliance requirements and controls.
Which solution should the data analyst implement to meet the desired requirements with the LEAST amount of setup effort?

A. Re-arrange data in Amazon S3 to store customer data about each state in a different S3 folder within the same bucke
B. Set up S3 bucket policies to provide marketing employees with appropriate data access under compliance control
C. Delete the bucket policies after the project.
D. Load tabular data from Amazon S3 to an Amazon EMR cluster using s3DistC
E. Implement a custom Hadoop-based row-level security solution on the Hadoop Distributed File System (HDFS) to provide marketing employees with appropriate data access under compliance control
F. Terminate the EMR cluster after the project.
G. Load tabular data from Amazon S3 to Amazon Redshift with the COPY comman
H. Use the built-in row- level security feature in Amazon Redshift to provide marketing employees with appropriate data access under compliance control
I. Delete the Amazon Redshift tables after the project.
J. Load tabular data from Amazon S3 to Amazon QuickSight Enterprise edition by directly importing it as a data sourc
K. Use the built-in row-level security feature in Amazon QuickSight to provide marketing employees with appropriate data access under compliance control
L. Delete Amazon QuickSight data sources after the project is complete.

Answer: C

NEW QUESTION 118
A manufacturing company uses Amazon Connect to manage its contact center and Salesforce to manage its customer relationship management (CRM) data. The data engineering team must build a pipeline to ingest data from the contact center and CRM system into a data lake that is built on Amazon S3.
What is the MOST efficient way to collect data in the data lake with the LEAST operational overhead?

A. Use Amazon Kinesis Data Streams to ingest Amazon Connect data and Amazon AppFlow to ingest Salesforce data.
B. Use Amazon Kinesis Data Firehose to ingest Amazon Connect data and Amazon Kinesis Data Streams to ingest Salesforce data.
C. Use Amazon Kinesis Data Firehose to ingest Amazon Connect data and Amazon AppFlow to ingest Salesforce data.
D. Use Amazon AppFlow to ingest Amazon Connect data and Amazon Kinesis Data Firehose to ingest Salesforce data.

Answer: B

NEW QUESTION 119
A company uses Amazon Redshift as its data warehouse A new table includes some columns that contain sensitive data and some columns that contain non-sensitive data The data in the table eventually will be referenced by several existing queries that run many times each day
A data analytics specialist must ensure that only members of the company's auditing team can read the columns that contain sensitive data All other users must have read-only access to the columns that contain non-sensitive data
Which solution will meet these requirements with the LEAST operational overhead?

A. Grant the auditing team permission to read from the tabl
B. Load the columns that contain non-sensitive data into a second tabl
C. Grant the appropriate users read-only permissions to the second table.
D. Grant all users read-only permissions to the columns that contain non-sensitive data Use the GRANT SELECT command to allow the auditing team to access the columns that contain sensitive data
E. Grant all users read-only permissions to the columns that contain non-sensitive data Attach an 1AM policy to the auditing team with an explicit Allow action that grants access to the columns that contain sensitive data
F. Grant the auditing team permission to read from the table Create a view of the table that includes the columns that contain non-sensitive data Grant the appropriate users read-only permissions to that view

Answer: B

Explanation:
https://aws.amazon.com/jp/about-aws/whats-new/2020/03/announcing-column-level-access-control-for-amazon

NEW QUESTION 123
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 be 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 one processing semantics of Apache Flink and Apache Spark Streaming included in Amazon EMR.

Answer: A

NEW QUESTION 125
A marketing company is using Amazon EMR clusters for its workloads. The company manually installs third- party libraries on the clusters by logging in to the master nodes. A data analyst needs to create an automated solution to replace the manual process.
Which options can fulfill these requirements? (Choose two.)

A. Place the required installation scripts in Amazon S3 and execute them using custom bootstrap actions.
B. Place the required installation scripts in Amazon S3 and execute them through Apache Spark in Amazon EMR.
C. Install the required third-party libraries in the existing EMR master nod
D. Create an AMI out of that master node and use that custom AMI to re-create the EMR cluster.
E. Use an Amazon DynamoDB table to store the list of required application
F. Trigger an AWS Lambda function with DynamoDB Streams to install the software.
G. Launch an Amazon EC2 instance with Amazon Linux and install the required third-party libraries on the instanc
H. Create an AMI and use that AMI to create the EMR cluster.

Answer: AE

Explanation:
https://aws.amazon.com/about-aws/whats-new/2017/07/amazon-emr-now-supports-launching-clusters-with-cust
https://docs.aws.amazon.com/de_de/emr/latest/ManagementGuide/emr-plan-bootstrap.html

NEW QUESTION 129
A large retailer has successfully migrated to an Amazon S3 data lake architecture. The company's marketing team is using Amazon Redshift and Amazon QuickSight to analyze data, and derive and visualize insights. To ensure the marketing team has the most up-to-date actionable information, a data analyst implements nightly refreshes of Amazon Redshift using terabytes of updates from the previous day.
After the first nightly refresh, users report that half of the most popular dashboards that had been running correctly before the refresh are now running much slower. Amazon CloudWatch does not show any alerts.
What is the MOST likely cause for the performance degradation?

A. The dashboards are suffering from inefficient SQL queries.
B. The cluster is undersized for the queries being run by the dashboards.
C. The nightly data refreshes are causing a lingering transaction that cannot be automatically closed by Amazon Redshift due to ongoing user workloads.
D. The nightly data refreshes left the dashboard tables in need of a vacuum operation that could not be automatically performed by Amazon Redshift due to ongoing user workloads.

Answer: D

Explanation:

https://github.com/awsdocs/amazon-redshift-developer-guide/issues/21

NEW QUESTION 130
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 HDF
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 Zone
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 Zon
J. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.

Answer: D

NEW QUESTION 135
A company is planning to create a data lake in Amazon S3. The company wants to create tiered storage based on access patterns and cost objectives. The solution must include support for JDBC connections from legacy clients, metadata management that allows federation for access control, and batch-based ETL using PySpark and Scala. Operational management should be limited.
Which combination of components can meet these requirements? (Choose three.)

A. AWS Glue Data Catalog for metadata management
B. Amazon EMR with Apache Spark for ETL
C. AWS Glue for Scala-based ETL
D. Amazon EMR with Apache Hive for JDBC clients
E. Amazon Athena for querying data in Amazon S3 using JDBC drivers
F. Amazon EMR with Apache Hive, using an Amazon RDS with MySQL-compatible backed metastore

Answer: BEF

NEW QUESTION 136
An analytics software as a service (SaaS) provider wants to offer its customers business intelligence <BI) reporting capabilities that are self-service The provider is using Amazon QuickSight to build these reports The data for the reports resides in a multi-tenant database, but each customer should only be able to access their own data
The provider wants to give customers two user role options
• Read-only users for individuals who only need to view dashboards
• Power users for individuals who are allowed to create and share new dashboards with other users Which QuickSight feature allows the provider to meet these requirements'?

A. Embedded dashboards
B. Table calculations
C. Isolated namespaces
D. SPICE

Answer: A

NEW QUESTION 137
A reseller that has thousands of AWS accounts receives AWS Cost and Usage Reports in an Amazon S3 bucket The reports are delivered to the S3 bucket in the following format
<examp/e-reporT-prefix>/<examp/e-report-rtame>/yyyymmdd-yyyymmdd/<examp/e-report-name> parquet An AWS Glue crawler crawls the S3 bucket and populates an AWS Glue Data Catalog with a table Business analysts use Amazon Athena to query the table and create monthly summary reports for the AWS accounts
The business analysts are experiencing slow queries because of the accumulation of reports from the last 5 years The business analysts want the operations team to make changes to improve query performance
Which action should the operations team take to meet these requirements?

A. Change the file format to csv.zip.
B. Partition the data by date and account ID
C. Partition the data by month and account ID
D. Partition the data by account ID, year, and month

Answer: B

NEW QUESTION 140
An online retail company is migrating its reporting system to AWS. The company's legacy system runs data processing on online transactions using a complex series of nested Apache Hive queries. Transactional data is exported from the online system to the reporting system several times a day. Schemas in the files are stable between updates.
A data analyst wants to quickly migrate the data processing to AWS, so any code changes should be minimized. To keep storage costs low, the data analyst decides to store the data in Amazon S3. It is vital that the data from the reports and associated analytics is completely up to date based on the data in Amazon S3.
Which solution meets these requirements?

A. Create an AWS Glue Data Catalog to manage the Hive metadat
B. Create an AWS Glue crawler over Amazon S3 that runs when data is refreshed to ensure that data changes are update
C. Create an Amazon EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.
D. Create an AWS Glue Data Catalog to manage the Hive metadat
E. Create an Amazon EMR cluster with consistent view enable
F. Run emrfs sync before each analytics step to ensure data changes are update
G. Create an EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.
H. Create an Amazon Athena table with CREATE TABLE AS SELECT (CTAS) to ensure data is refreshed from underlying queries against the raw datase
I. Create an AWS Glue Data Catalog to manage the Hive metadata over the CTAS tabl
J. Create an Amazon EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.
K. Use an S3 Select query to ensure that the data is properly update
L. Create an AWS Glue Data Catalog to manage the Hive metadata over the S3 Select tabl
M. Create an Amazon EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.

Answer: A

NEW QUESTION 144
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 Redshif
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 Redshif
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 147
An online retail company uses Amazon Redshift to store historical sales transactions. The company is required to encrypt data at rest in the clusters to comply with the Payment Card Industry Data Security Standard (PCI DSS). A corporate governance policy mandates management of encryption keys using an on-premises hardware security module (HSM).
Which solution meets these requirements?

A. Create and manage encryption keys using AWS CloudHSM Classi
B. Launch an Amazon Redshift cluster in a VPC with the option to use CloudHSM Classic for key management.
C. Create a VPC and establish a VPN connection between the VPC and the on-premises networ
D. Create an HSM connection and client certificate for the on-premises HS
E. Launch a cluster in the VPC with theoption to use the on-premises HSM to store keys.
F. Create an HSM connection and client certificate for the on-premises HS
G. Enable HSM encryption on the existing unencrypted cluster by modifying the cluste
H. Connect to the VPC where the Amazon Redshift cluster resides from the on-premises network using a VPN.
I. Create a replica of the on-premises HSM in AWS CloudHS
J. Launch a cluster in a VPC with the option to use CloudHSM to store keys.

Answer: B

NEW QUESTION 151
A regional energy company collects voltage data from sensors attached to buildings. To address any known dangerous conditions, the company wants to be alerted when a sequence of two voltage drops is detected within 10 minutes of a voltage spike at the same building. It is important to ensure that all messages are delivered as quickly as possible. The system must be fully managed and highly available. The company also needs a solution that will automatically scale up as it covers additional cites with this monitoring feature. The alerting system is subscribed to an Amazon SNS topic for remediation.
Which solution meets these requirements?

A. Create an Amazon Managed Streaming for Kafka cluster to ingest the data, and use an Apache Spark Streaming with Apache Kafka consumer API in an automatically scaled Amazon EMR cluster to process the incoming dat
B. Use the Spark Streaming application to detect the known event sequence and send the SNS message.
C. Create a REST-based web service using Amazon API Gateway in front of an AWS Lambda function.Create an Amazon RDS for PostgreSQL database with sufficient Provisioned IOPS (PIOPS). In the Lambda function, store incoming events in the RDS database and query the latest data to detect the known event sequence and send the SNS message.
D. Create an Amazon Kinesis Data Firehose delivery stream to capture the incoming sensor dat
E. Use an AWS Lambda transformation function to detect the known event sequence and send the SNS message.
F. Create an Amazon Kinesis data stream to capture the incoming sensor data and create another stream for alert message
G. Set up AWS Application Auto Scaling on bot
H. Create a Kinesis Data Analytics for Java application to detect the known event sequence, and add a message to the message strea
I. Configure an AWS Lambda function to poll the message stream and publish to the SNS topic.

Answer: D

NEW QUESTION 152
A data analyst is designing a solution to interactively query datasets with SQL using a JDBC connection. Users will join data stored in Amazon S3 in Apache ORC format with data stored in Amazon Elasticsearch Service (Amazon ES) and Amazon Aurora MySQL.
Which solution will provide the MOST up-to-date results?

A. Use AWS Glue jobs to ETL data from Amazon ES and Aurora MySQL to Amazon S3. Query the data with Amazon Athena.
B. Use Amazon DMS to stream data from Amazon ES and Aurora MySQL to Amazon Redshif
C. Query the data with Amazon Redshift.
D. Query all the datasets in place with Apache Spark SQL running on an AWS Glue developer endpoint.
E. Query all the datasets in place with Apache Presto running on Amazon EMR.

Answer: C

NEW QUESTION 157
A transport company wants to track vehicular movements by capturing geolocation records. The records are 10 B in size and up to 10,000 records are captured each second. Data transmission delays of a few minutes are acceptable, considering unreliable network conditions. The transport company decided to use Amazon Kinesis Data Streams to ingest the data. The company is looking for a reliable mechanism to send data to Kinesis Data Streams while maximizing the throughput efficiency of the Kinesis shards.
Which solution will meet the company's requirements?

A. Kinesis Agent
B. Kinesis Producer Library (KPL)
C. Kinesis Data Firehose
D. Kinesis SDK

Answer: B

NEW QUESTION 162
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 ot 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 166
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 metri
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 metri
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 metri
I. Create anautomatic 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 metri
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 170
Once a month, a company receives a 100 MB .csv file compressed with gzip. The file contains 50,000 property listing records and is stored in Amazon S3 Glacier. The company needs its data analyst to query a subset of the data for a specific vendor.
What is the most cost-effective solution?

A. Load the data into Amazon S3 and query it with Amazon S3 Select.
B. Query the data from Amazon S3 Glacier directly with Amazon Glacier Select.
C. Load the data to Amazon S3 and query it with Amazon Athena.
D. Load the data to Amazon S3 and query it with Amazon Redshift Spectrum.

Answer: A

NEW QUESTION 175
A global company has different sub-organizations, and each sub-organization sells its products and services in various countries. The company's senior leadership wants to quickly identify which sub-organization is the strongest performer in each country. All sales data is stored in Amazon S3 in Parquet format.

Which approach can provide the visuals that senior leadership requested with the least amount of effort?

A. Use Amazon QuickSight with Amazon Athena as the data sourc
B. Use heat maps as the visual type.
C. Use Amazon QuickSight with Amazon S3 as the data sourc
D. Use heat maps as the visual type.
E. Use Amazon QuickSight with Amazon Athena as the data sourc
F. Use pivot tables as the visual type.
G. Use Amazon QuickSight with Amazon S3 as the data sourc
H. Use pivot tables as the visual type.

Answer: A

NEW QUESTION 177
A company's data analyst needs to ensure that queries executed in Amazon Athena cannot scan more than a prescribed amount of data for cost control purposes. Queries that exceed the prescribed threshold must be canceled immediately.
What should the data analyst do to achieve this?

A. Configure Athena to invoke an AWS Lambda function that terminates queries when the prescribed threshold is crossed.
B. For each workgroup, set the control limit for each query to the prescribed threshold.
C. Enforce the prescribed threshold on all Amazon S3 bucket policies
D. For each workgroup, set the workgroup-wide data usage control limit to the prescribed threshold.

Answer: B

Explanation:
https://docs.aws.amazon.com/athena/latest/ug/manage-queries-control-costs-with-workgroups.html

NEW QUESTION 179
A company needs to store objects containing log data in JSON format. The objects are generated by eight applications running in AWS. Six of the applications generate a total of 500 KiB of data per second, and two of the applications can generate up to 2 MiB of data per second.
A data engineer wants to implement a scalable solution to capture and store usage data in an Amazon S3 bucket. The usage data objects need to be reformatted, converted to .csv format, and then compressed before they are stored in Amazon S3. The company requires the solution to include the least custom code possible and has authorized the data engineer to request a service quota increase if needed.
Which solution meets these requirements?

A. Configure an Amazon Kinesis Data Firehose delivery stream for each applicatio
B. Write AWS Lambda functions to read log data objects from the stream for each applicatio
C. Have the function perform reformatting and .csv conversio
D. Enable compression on all the delivery streams.
E. Configure an Amazon Kinesis data stream with one shard per applicatio
F. Write an AWS Lambda function to read usage data objects from the shard
G. Have the function perform .csv conversion, reformatting, and compression of the dat
H. Have the function store the output in Amazon S3.
I. Configure an Amazon Kinesis data stream for each applicatio
J. Write an AWS Lambda function to read usage data objects from the stream for each applicatio
K. Have the function perform .csv conversion, reformatting, and compression of the dat
L. Have the function store the output in Amazon S3.
M. Store usage data objects in an Amazon DynamoDB tabl
N. Configure a DynamoDB stream to copy the objects to an S3 bucke
O. Configure an AWS Lambda function to be triggered when objects are written to the S3 bucke
P. Have the function convert the objects into .csv format.

Answer: A

NEW QUESTION 184
A company is sending historical datasets to Amazon S3 for storage. A data engineer at the company wants to make these datasets available for analysis using Amazon Athena. The engineer also wants to encrypt the Athena query results in an S3 results location by using AWS solutions for encryption. The requirements for encrypting the query results are as follows:
Use custom keys for encryption of the primary dataset query results. Use generic encryption for all other query results.
Provide an audit trail for the primary dataset queries that shows when the keys were used and by whom.
Which solution meets these requirements?

A. Use server-side encryption with S3 managed encryption keys (SSE-S3) for the primary datase
B. Use SSE-S3 for the other datasets.
C. Use server-side encryption with customer-provided encryption keys (SSE-C) for the primary dataset.Use server-side encryption with S3 managed encryption keys (SSE-S3) for the other datasets.
D. Use server-side encryption with AWS KMS managed customer master keys (SSE-KMS CMKs) for the primary datase
E. Use server-side encryption with S3 managed encryption keys (SSE-S3) for the other datasets.
F. Use client-side encryption with AWS Key Management Service (AWS KMS) customer managed keys for the primary datase
G. Use S3 client-side encryption with client-side keys for the other datasets.

Answer: A

NEW QUESTION 187
A retail company wants to use Amazon QuickSight to generate dashboards for web and in-store sales. A group of 50 business intelligence professionals will develop and use the dashboards. Once ready, the dashboards will be shared with a group of 1,000 users.
The sales data comes from different stores and is uploaded to Amazon S3 every 24 hours. The data is partitioned by year and month, and is stored in Apache Parquet format. The company is using the AWS Glue Data Catalog as its main data catalog and Amazon Athena for querying. The total size of the uncompressed

data that the dashboards query from at any point is 200 GB.
Which configuration will provide the MOST cost-effective solution that meets these requirements?

A. Load the data into an Amazon Redshift cluster by using the COPY comman
B. Configure 50 author users and 1,000 reader user
C. Use QuickSight Enterprise editio
D. Configure an Amazon Redshift data source with a direct query option.
E. Use QuickSight Standard editio
F. Configure 50 author users and 1,000 reader user
G. Configure an Athena data source with a direct query option.
H. Use QuickSight Enterprise editio
I. Configure 50 author users and 1,000 reader user
J. Configure an Athena data source and import the data into SPIC
K. Automatically refresh every 24 hours.
L. Use QuickSight Enterprise editio
M. Configure 1 administrator and 1,000 reader user
N. Configure an S3 data source and import the data into SPIC
O. Automatically refresh every 24 hours.

Answer: C

NEW QUESTION 192
A utility company wants to visualize data for energy usage on a daily basis in Amazon QuickSight A data analytics specialist at the company has built a data pipeline to collect and ingest the data into Amazon S3 Each day the data is stored in an individual csv file in an S3 bucket This is an example of the naming structure 20210707_datacsv 20210708_datacsv
To allow for data querying in QuickSight through Amazon Athena the specialist used an AWS Glue crawler to create a table with the path "s3 //powertransformer/20210707_data csv" However when the data is queried, it returns zero rows
How can this issue be resolved?

A. Modify the IAM policy for the AWS Glue crawler to access Amazon S3.
B. Ingest the files again.
C. Store the files in Apache Parquet format.
D. Update the table path to "s3://powertransformer/".

Answer: D

NEW QUESTION 195
A retail company's data analytics team recently created multiple product sales analysis dashboards for the average selling price per product using Amazon QuickSight. The dashboards were created from .csv files uploaded to Amazon S3. The team is now planning to share the dashboards with the respective external product owners by creating individual users in Amazon QuickSight. For compliance and governance reasons, restricting access is a key requirement. The product owners should view only their respective product analysis in the dashboard reports.
Which approach should the data analytics team take to allow product owners to view only their products in the dashboard?

A. Separate the data by product and use S3 bucket policies for authorization.
B. Separate the data by product and use IAM policies for authorization.
C. Create a manifest file with row-level security.
D. Create dataset rules with row-level security.

Answer: D

Explanation:
https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html

NEW QUESTION 200
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

DAS-C01 Practice Exam Features:

* DAS-C01 Questions and Answers Updated Frequently

* DAS-C01 Practice Questions Verified by Expert Senior Certified Staff

* DAS-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DAS-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DAS-C01 Practice Test Here](https://www.surepassexam.com/DAS-C01-exam-dumps.html)