

Exam Questions DAS-C01

AWS Certified Data Analytics - Specialty

<https://www.2passeasy.com/dumps/DAS-C01/>



NEW QUESTION 1

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 2

A software company wants to use instrumentation data to detect and resolve errors to improve application recovery time. The company requires API usage anomalies, like error rate and response time spikes, to be detected in near-real time (NRT). The company also requires that data analysts have access to dashboards for log analysis in NRT.

Which solution meets these requirements?

- A. Use Amazon Kinesis Data Firehose as the data transport layer for logging data. Use Amazon Kinesis Data Analytics to uncover the NRT API usage anomalies. Use Kinesis Data Firehose to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring. Use OpenSearch Dashboards (Kibana) in Amazon OpenSearch Service (Amazon Elasticsearch Service) for the dashboards.
- B. Use Amazon Kinesis Data Analytics as the data transport layer for logging data.
- C. Use Amazon Kinesis Data Streams to uncover NRT monitoring metrics.
- D. Use Amazon Kinesis Data Firehose to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring. Use Amazon QuickSight for the dashboards.
- E. Use Amazon Kinesis Data Analytics as the data transport layer for logging data and to uncover NRT monitoring metrics. Use Amazon Kinesis Data Firehose to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring. Use OpenSearch Dashboards (Kibana) in Amazon OpenSearch Service (Amazon Elasticsearch Service) for the dashboards.
- F. Use Amazon Kinesis Data Firehose as the data transport layer for logging data. Use Amazon Kinesis Data Analytics to uncover NRT monitoring metrics. Use Amazon Kinesis Data Streams to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring. Use Amazon QuickSight for the dashboards.

Answer: C

NEW QUESTION 3

A company is building an analytical solution that includes Amazon S3 as data lake storage and Amazon Redshift for data warehousing. The company wants to use Amazon Redshift Spectrum to query the data that is stored in Amazon S3.

Which steps should the company take to improve performance when the company uses Amazon Redshift Spectrum to query the S3 data files? (Select THREE)
Use gzip compression with individual file sizes of 1-5 GB

- A. Use a columnar storage file format.
- B. Partition the data based on the most common query predicates.
- C. Split the data into KB-sized files.
- D. Keep all files about the same size.
- E. Use file formats that are not splittable.

Answer: BCD

NEW QUESTION 4

A large ride-sharing company has thousands of drivers globally serving millions of unique customers every day. The company has decided to migrate an existing data mart to Amazon Redshift. The existing schema includes the following tables.

A trips fact table for information on completed rides. A drivers dimension table for driver profiles. A customers fact table holding customer profile information. The company analyzes trip details by date and destination to examine profitability by region. The drivers data rarely changes. The customers data frequently changes.

What table design provides optimal query performance?

- A. Use DISTSTYLE KEY (destination) for the trips table and sort by date.
- B. Use DISTSTYLE ALL for the drivers and customers tables.
- C. Use DISTSTYLE EVEN for the trips table and sort by date.
- D. Use DISTSTYLE ALL for the drivers table. Use DISTSTYLE EVEN for the customers table.
- E. Use DISTSTYLE KEY (destination) for the trips table and sort by date.
- F. Use DISTSTYLE ALL for the drivers table.
- G. Use DISTSTYLE EVEN for the customers table.
- H. Use DISTSTYLE EVEN for the drivers table and sort by date.
- I. Use DISTSTYLE ALL for both fact tables.

Answer: C

Explanation:

<https://www.matillion.com/resources/blog/aws-redshift-performance-choosing-the-right-distribution-styles/#:~:t>
https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html

NEW QUESTION 5

A company has a marketing department and a finance department. The departments are storing data in Amazon S3 in their own AWS accounts in AWS Organizations. Both departments use AWS Lake Formation to catalog and secure their data. The departments have some databases and tables that share common names.
The marketing department needs to securely access some tables from the finance department. Which two steps are required for this process? (Choose two.)

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

Answer: AB

Explanation:

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

NEW QUESTION 6

An operations team notices that a few AWS Glue jobs for a given ETL application are failing. The AWS Glue jobs read a large number of small JSON files from an Amazon S3 bucket and write the data to a different S3 bucket in Apache Parquet format with no major transformations. Upon initial investigation, a data engineer notices the following error message in the History tab on the AWS Glue console: “Command Failed with Exit Code 1.”
Upon further investigation, the data engineer notices that the driver memory profile of the failed jobs crosses the safe threshold of 50% usage quickly and reaches 90–95% soon after. The average memory usage across all executors continues to be less than 4%.
The data engineer also notices the following error while examining the related Amazon CloudWatch Logs. What should the data engineer do to solve the failure in the MOST cost-effective way?

- A. Change the worker type from Standard to G.2X.
- B. Modify the AWS Glue ETL code to use the ‘groupFiles’: ‘inPartition’ feature.
- C. Increase the fetch size setting by using AWS Glue dynamics frame.
- D. Modify maximum capacity to increase the total maximum data processing units (DPUs) used.

Answer: B

Explanation:

<https://docs.aws.amazon.com/glue/latest/dg/monitor-profile-debug-oom-abnormalities.html#monitor-debug-oom>

NEW QUESTION 7

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

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

Answer: D

NEW QUESTION 8

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 9

A gaming company is collecting clickstream 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 it and 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 10

A company is hosting an enterprise reporting solution with Amazon Redshift. The application provides reporting capabilities to three main groups: an executive group to access financial reports, a data analyst group to run long-running ad-hoc queries, and a data engineering group to run stored procedures and ETL processes. The executive team requires queries to run with optimal performance. The data engineering team expects queries to take minutes. Which Amazon Redshift feature meets the requirements for this task?

- A. Concurrency scaling
- B. Short query acceleration (SQA)
- C. Workload management (WLM)
- D. Materialized views

Answer: D

Explanation:

Materialized views:

NEW QUESTION 10

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 Redshift
- B. Create an external table in Amazon Redshift to point to the S3 location
- C. Use Amazon Redshift Spectrum to join to data that is older than 13 months.
- D. Take a snapshot of the Amazon Redshift cluster
- 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-Tiering
- H. Use AWS Glue to crawl the S3 bucket location to create external tables in an AWS Glue Data Catalog. 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 12

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 IAM 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 13

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 table
- 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 job
- 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 16

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 21

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 22

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

How should the company meet these requirements?

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

Answer: D

Explanation:

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html

NEW QUESTION 23

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 28

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 data
- 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 consumer
- 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 process
- 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 consumer
- 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 31

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 34

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 space?

- 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. Automatically flush 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 38

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

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

- A. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatch
- B. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
- C. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Data Firehose delivery stream to CloudWatch
- 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 CloudWatch
- 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 CloudWatch
- 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 42

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 dataset
- 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 data
- 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 43

An online retailer needs to deploy a product sales reporting solution. The source data is exported from an external online transaction processing (OLTP) system for reporting. Roll-up data is calculated each day for the previous day's activities. The reporting system has the following requirements:

Have the daily roll-up data readily available for 1 year.

After 1 year, archive the daily roll-up data for occasional but immediate access.

The source data exports stored in the reporting system must be retained for 5 years. Query access will be needed only for re-evaluation, which may occur within the first 90 days.

Which combination of actions will meet these requirements while keeping storage costs to a minimum? (Choose two.)

- A. Store the source data initially in the Amazon S3 Standard-Infrequent Access (S3 Standard-IA) storage class
- B. Apply a lifecycle configuration that changes the storage class to Amazon S3 Glacier Deep Archive 90 days after creation, and then deletes the data 5 years after creation.
- C. Store the source data initially in the Amazon S3 Glacier storage class
- D. Apply a lifecycle configuration that changes the storage class from Amazon S3 Glacier to Amazon S3 Glacier Deep Archive 90 days after creation, and then deletes the data 5 years after creation.
- E. Store the daily roll-up data initially in the Amazon S3 Standard storage class
- F. Apply a lifecycle configuration that changes the storage class to Amazon S3 Glacier Deep Archive 1 year after data creation.
- G. Store the daily roll-up data initially in the Amazon S3 Standard storage class
- H. Apply a lifecycle configuration that changes the storage class to Amazon S3 Standard-Infrequent Access (S3 Standard-IA) 1 year after data creation.
- I. Store the daily roll-up data initially in the Amazon S3 Standard-Infrequent Access (S3 Standard-IA) storage class
- J. Apply a lifecycle configuration that changes the storage class to Amazon S3 Glacier 1 year after data creation.

Answer: AD

NEW QUESTION 45

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 steps, 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 scripts to curate the data in an Amazon EMR cluster
- B. Store the curated data in Amazon S3 for ML processing.
- C. Create custom ETL jobs on-premises to curate the data
- D. Use AWS DMS to ingest data into Amazon S3 for ML processing.
- E. Ingest data into Amazon S3 using AWS DMS
- 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 Snowball
- 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 46

A media company wants to perform machine learning and analytics on the data residing in its Amazon S3 data lake. There are two data transformation requirements that will enable the consumers within the company to create reports:

- Daily transformations of 300 GB of data with different file formats landing in Amazon S3 at a scheduled time.
- One-time transformations of terabytes of archived data residing in the S3 data lake.

Which combination of solutions cost-effectively meets the company's requirements for transforming the data? (Choose three.)

- A. For daily incoming data, use AWS Glue crawlers to scan and identify the schema.
- B. For daily incoming data, use Amazon Athena to scan and identify the schema.
- C. For daily incoming data, use Amazon Redshift to perform transformations.
- D. For daily incoming data, use AWS Glue workflows with AWS Glue jobs to perform transformations.
- E. For archived data, use Amazon EMR to perform data transformations.
- F. For archived data, use Amazon SageMaker to perform data transformations.

Answer: ADE

NEW QUESTION 51

A banking company is currently using an Amazon Redshift cluster with dense storage (DS) nodes to store sensitive data. An audit found that the cluster is unencrypted. Compliance requirements state that a database with sensitive data must be encrypted through a hardware security module (HSM) with automated key rotation.

Which combination of steps is required to achieve compliance? (Choose two.)

- A. Set up a trusted connection with HSM using a client and server certificate with automatic key rotation.
- B. Modify the cluster with an HSM encryption option and automatic key rotation.
- C. Create a new HSM-encrypted Amazon Redshift cluster and migrate the data to the new cluster.
- D. Enable HSM with key rotation through the AWS CLI.
- E. Enable Elliptic Curve Diffie-Hellman Ephemeral (ECDHE) encryption in the HSM.

Answer: BD

NEW QUESTION 55

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

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

Answer: BD

NEW QUESTION 57

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

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

Answer: A

Explanation:

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

NEW QUESTION 59

A company analyzes historical data and needs to query data that is stored in Amazon S3. New data is generated daily as .csv files that are stored in Amazon S3. The company's analysts are using Amazon Athena to perform SQL queries against a recent subset of the overall data. The amount of data that is ingested into Amazon S3 has increased substantially over time, and the query latency also has increased. Which solutions could the company implement to improve query performance? (Choose two.)

- A. Use MySQL Workbench on an Amazon EC2 instance, and connect to Athena by using a JDBC or ODBC connector.
- B. Run the query from MySQL Workbench instead of Athena directly.
- C. Use Athena to extract the data and store it in Apache Parquet format on a daily basis.
- D. Query the extracted data.
- E. Run a daily AWS Glue ETL job to convert the data files to Apache Parquet and to partition the converted file.
- F. Create a periodic AWS Glue crawler to automatically crawl the partitioned data on a daily basis.
- G. Run a daily AWS Glue ETL job to compress the data files by using the .gzip format.
- H. Query the compressed data.
- I. Run a daily AWS Glue ETL job to compress the data files by using the .lzo format.
- J. Query the compressed data.

Answer: BC

NEW QUESTION 60

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 bucket.
- 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 command.
- 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 source.
- 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 64

A mortgage company has a microservice for accepting payments. This microservice uses the Amazon DynamoDB encryption client with AWS KMS managed keys to encrypt the sensitive data before writing the data to DynamoDB. The finance team should be able to load this data into Amazon Redshift and aggregate the values within the sensitive fields. The Amazon Redshift cluster is shared with other data analysts from different business units.

Which steps should a data analyst take to accomplish this task efficiently and securely?

- A. Create an AWS Lambda function to process the DynamoDB stream.
- B. Decrypt the sensitive data using the same KMS key.
- C. Save the output to a restricted S3 bucket for the finance team.
- D. Create a finance table in Amazon Redshift that is accessible to the finance team only.
- E. Use the COPY command to load the data from Amazon S3 to the finance table.
- F. Create an AWS Lambda function to process the DynamoDB stream.
- G. Save the output to a restricted S3 bucket for the finance team.
- H. Create a finance table in Amazon Redshift that is accessible to the finance team only.
- I. Use the COPY command with the IAM role that has access to the KMS key to load the data from S3 to the finance table.
- J. Create an Amazon EMR cluster with an EMR_EC2_DefaultRole role that has access to the KMS key. Create Apache Hive tables that reference the data stored in DynamoDB and the finance table in Amazon Redshift.
- K. In Hive, select the data from DynamoDB and then insert the output to the finance table in Amazon Redshift.
- L. Create an Amazon EMR cluster.
- M. Create Apache Hive tables that reference the data stored in DynamoDB.
- N. Insert the output to the restricted Amazon S3 bucket for the finance team.
- O. Use the COPY command with the IAM role that has access to the KMS key to load the data from Amazon S3 to the finance table in Amazon Redshift.

Answer: B

NEW QUESTION 67

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 71

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 table.
- B. Load the columns that contain non-sensitive data into a second table.
- 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 IAM 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 72

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 77

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 78

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 81

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 Classic
- 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 network
- D. Create an HSM connection and client certificate for the on-premises HS
- E. Launch a cluster in the VPC with the option 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 cluster
- 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 CloudHSM
- J. Launch a cluster in a VPC with the option to use CloudHSM to store keys.

Answer: B

NEW QUESTION 83

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 88

A large energy company is using Amazon QuickSight to build dashboards and report the historical usage data of its customers. This data is hosted in Amazon Redshift. The reports need access to all the fact tables' billions of records to create aggregation in real time grouping by multiple dimensions.

A data analyst created the dataset in QuickSight by using a SQL query and not SPICE. Business users have noted that the response time is not fast enough to meet their needs.

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

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

Answer: A

NEW QUESTION 90

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 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 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 94

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 97

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

Answer: A

NEW QUESTION 102

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 104

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 108

A manufacturing company uses Amazon S3 to store its data. The company wants to use AWS Lake Formation to provide granular-level security on those data assets. The data is in Apache Parquet format. The company has set a deadline for a consultant to build a data lake.

How should the consultant create the MOST cost-effective solution that meets these requirements?

- A. Run Lake Formation blueprints to move the data to Lake Formatio
- B. Once Lake Formation has the data, apply permissions on Lake Formation.
- C. To create the data catalog, run an AWS Glue crawler on the existing Parquet dat
- D. Register the Amazon S3 path and then apply permissions through Lake Formation to provide granular-level security.
- E. Install Apache Ranger on an Amazon EC2 instance and integrate with Amazon EM
- F. Using Ranger policies, create role-based access control for the existing data assets in Amazon S3.
- G. Create multiple IAM roles for different users and group
- H. Assign IAM roles to different data assets in Amazon S3 to create table-based and column-based access controls.

Answer: A

Explanation:

<https://aws.amazon.com/blogs/big-data/building-securing-and-managing-data-lakes-with-aws-lake-formation/>

NEW QUESTION 112

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 117

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 122

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 125

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 tabl
- 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 Data Analytics application to send its output to an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exist
- J. Forward the record from the Lambda function to the original application destination.

Answer: C

NEW QUESTION 126

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DAS-C01 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DAS-C01 Product From:

<https://www.2passeasy.com/dumps/DAS-C01/>

Money Back Guarantee

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