

Amazon

Exam Questions AWS-Certified-Data-Analytics-Specialty

AWS Certified Data Analytics - Specialty



NEW QUESTION 1

A company analyzes its data in an Amazon Redshift data warehouse, which currently has a cluster of three dense storage nodes. Due to a recent business acquisition, the company needs to load an additional 4 TB of user data into Amazon Redshift. The engineering team will combine all the user data and apply complex calculations that require I/O intensive resources. The company needs to adjust the cluster's capacity to support the change in analytical and storage requirements.

Which solution meets these requirements?

- A. Resize the cluster using elastic resize with dense compute nodes.
- B. Resize the cluster using classic resize with dense compute nodes.
- C. Resize the cluster using elastic resize with dense storage nodes.
- D. Resize the cluster using classic resize with dense storage nodes.

Answer: C

NEW QUESTION 2

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

Which solution meets these requirements?

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

Answer: C

Explanation:

https://aws.amazon.com/about-aws/whats-new/2019/02/athena_workgroups/

NEW QUESTION 3

A transportation company uses IoT sensors attached to trucks to collect vehicle data for its global delivery fleet. The company currently sends the sensor data in small .csv files to Amazon S3. The files are then loaded into a 10-node Amazon Redshift cluster with two slices per node and queried using both Amazon Athena and Amazon Redshift. The company wants to optimize the files to reduce the cost of querying and also improve the speed of data loading into the Amazon Redshift cluster.

Which solution meets these requirements?

- A. Use AWS Glue to convert all the files from .csv to a single large Apache Parquet file.
- B. COPY the file into Amazon Redshift and query the file with Athena from Amazon S3.
- C. Use Amazon EMR to convert each .csv file to Apache Avro.
- D. COPY the files into Amazon Redshift and query the file with Athena from Amazon S3.
- E. Use AWS Glue to convert the files from .csv to a single large Apache ORC file.
- F. COPY the file into Amazon Redshift and query the file with Athena from Amazon S3.
- G. Use AWS Glue to convert the files from .csv to Apache Parquet to create 20 Parquet files.
- H. COPY the files into Amazon Redshift and query the files with Athena from Amazon S3.

Answer: D

NEW QUESTION 4

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 5

A company is building a service to monitor fleets of vehicles. The company collects IoT data from a device in each vehicle and loads the data into Amazon Redshift in near-real time. Fleet owners upload .csv files containing vehicle reference data into Amazon S3 at different times throughout the day. A nightly process loads the vehicle reference data from Amazon S3 into Amazon Redshift. The company joins the IoT data from the device and the vehicle reference data to power reporting and dashboards. Fleet owners are frustrated by waiting a day for the dashboards to update.

Which solution would provide the SHORTEST delay between uploading reference data to Amazon S3 and the change showing up in the owners' dashboards?

- A. Use S3 event notifications to trigger an AWS Lambda function to copy the vehicle reference data into Amazon Redshift immediately when the reference data is uploaded to Amazon S3.

- B. Create and schedule an AWS Glue Spark job to run every 5 minute
- C. The job inserts reference data into Amazon Redshift.
- D. Send reference data to Amazon Kinesis Data Stream
- E. Configure the Kinesis data stream to directly load the reference data into Amazon Redshift in real time.
- F. Send the reference data to an Amazon Kinesis Data Firehose delivery stream
- G. Configure Kinesis with a buffer interval of 60 seconds and to directly load the data into Amazon Redshift.

Answer: A

NEW QUESTION 6

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

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

Answer: B

NEW QUESTION 7

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 8

A company has developed several AWS Glue jobs to validate and transform its data from Amazon S3 and load it into Amazon RDS for MySQL in batches once every day. The ETL jobs read the S3 data using a DynamicFrame. Currently, the ETL developers are experiencing challenges in processing only the incremental data on every run, as the AWS Glue job processes all the S3 input data on each run.

Which approach would allow the developers to solve the issue with minimal coding effort?

- A. Have the ETL jobs read the data from Amazon S3 using a DataFrame.
- B. Enable job bookmarks on the AWS Glue jobs.
- C. Create custom logic on the ETL jobs to track the processed S3 objects.
- D. Have the ETL jobs delete the processed objects or data from Amazon S3 after each run.

Answer: B

NEW QUESTION 9

A marketing company wants to improve its reporting and business intelligence capabilities. During the planning phase, the company interviewed the relevant stakeholders and discovered that:

- The operations team reports are run hourly for the current month's data.
- The sales team wants to use multiple Amazon QuickSight dashboards to show a rolling view of the last 30 days based on several categories.
- The sales team also wants to view the data as soon as it reaches the reporting backend.
- The finance team's reports are run daily for last month's data and once a month for the last 24 months of data.

Currently, there is 400 TB of data in the system with an expected additional 100 TB added every month. The company is looking for a solution that is as cost-effective as possible.

Which solution meets the company's requirements?

- A. Store the last 24 months of data in Amazon Redshift
- B. Configure Amazon QuickSight with Amazon Redshift as the data source.
- C. Store the last 2 months of data in Amazon Redshift and the rest of the months in Amazon S3. Set up an external schema and table for Amazon Redshift Spectrum
- D. Configure Amazon QuickSight with Amazon Redshift as the data source.
- E. Store the last 24 months of data in Amazon S3 and query it using Amazon Redshift Spectrum. Configure Amazon QuickSight with Amazon Redshift Spectrum as the data source.
- F. Store the last 2 months of data in Amazon Redshift and the rest of the months in Amazon S3. Use a long-running Amazon EMR with Apache Spark cluster to query the data as needed
- G. Configure Amazon QuickSight with Amazon EMR as the data source.

Answer: B

NEW QUESTION 10

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 10

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 11

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 KMS.
- B. Assign the CMK an alias.
- 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 KMS.
- E. Assign the CMK an alias.
- 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 KMS.
- H. Create an AWS Lambda function to encrypt and decrypt the data.
- 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 16

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 17

A data analyst is using AWS Glue to organize, cleanse, validate, and format a 200 GB dataset. The data analyst triggered the job to run with the Standard worker type. After 3 hours, the AWS Glue job status is still RUNNING. Logs from the job run show no error codes. The data analyst wants to improve the job execution time without overprovisioning.

Which actions should the data analyst take?

- A. Enable job bookmarks in AWS Glue to estimate the number of data processing units (DPUs). Based on the profiled metrics, increase the value of the executor-cores job parameter.
- B. Enable job metrics in AWS Glue to estimate the number of data processing units (DPUs). Based on the profiled metrics, increase the value of the maximum capacity job parameter.
- C. Enable job metrics in AWS Glue to estimate the number of data processing units (DPUs). Based on the profiled metrics, increase the value of the spark.yarn.executor.memoryOverhead job parameter.
- D. Enable job bookmarks in AWS Glue to estimate the number of data processing units (DPUs). Based on the profiled metrics, increase the value of the num-executors job parameter.

Answer: B

NEW QUESTION 21

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 Apache 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 clause
- F. Run the ALTER TABLE ADD PARTITION statement.
- G. Drop and recreate the table with the PARTITIONED BY clause
- H. Run the MSCK REPAIR TABLE statement.

Answer: BCF

NEW QUESTION 22

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 24

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 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 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 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 27

A company developed a new elections reporting website that uses Amazon Kinesis Data Firehose to deliver full logs from AWS WAF to an Amazon S3 bucket. The company is now seeking a low-cost option to perform this infrequent data analysis with visualizations of logs in a way that requires minimal development effort. Which solution meets these requirements?

- A. Use an AWS Glue crawler to create and update a table in the Glue data catalog from the log

- B. Use Athena to perform ad-hoc analyses and use Amazon QuickSight to develop data visualizations.
- C. Create a second Kinesis Data Firehose delivery stream to deliver the log files to Amazon Elasticsearch Service (Amazon ES). Use Amazon ES to perform text-based searches of the logs for ad-hoc analyses and use Kibana for data visualizations.
- D. Create an AWS Lambda function to convert the logs into .csv format.
- E. Then add the function to the Kinesis Data Firehose transformation configuration.
- F. Use Amazon Redshift to perform ad-hoc analyses of the logs using SQL queries and use Amazon QuickSight to develop data visualizations.
- G. Create an Amazon EMR cluster and use Amazon S3 as the data source.
- H. Create an Apache Spark job to perform ad-hoc analyses and use Amazon QuickSight to develop data visualizations.

Answer: A

Explanation:

<https://aws.amazon.com/blogs/big-data/analyzing-aws-waf-logs-with-amazon-es-amazon-athena-and-amazon-qu>

NEW QUESTION 29

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 table.
- B. Load all the data except for the columns containing sensitive data into a second table.
- 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 table.
- 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 33

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

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

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

Answer: D

NEW QUESTION 38

A manufacturing company has many IoT devices in different facilities across the world. The company is using Amazon Kinesis Data Streams to collect the data from the devices.

The company's operations team has started to observe many `WroteThroughputExceeded` exceptions. The operations team determines that the reason is the number of records that are being written to certain shards. The data contains device ID, capture date, measurement type, measurement value, and facility ID. The facility ID is used as the partition key.

Which action will resolve this issue?

- A. Change the partition key from facility ID to a randomly generated key.
- B. Increase the number of shards.
- C. Archive the data on the producers' side.
- D. Change the partition key from facility ID to capture date.

Answer: B

NEW QUESTION 43

A data analyst is designing an Amazon QuickSight dashboard using centralized sales data that resides in Amazon Redshift. The dashboard must be restricted so that a salesperson in Sydney, Australia, can see only the Australia view and that a salesperson in New York can see only United States (US) data.

What should the data analyst do to ensure the appropriate data security is in place?

- A. Place the data sources for Australia and the US into separate SPICE capacity pools.
- B. Set up an Amazon Redshift VPC security group for Australia and the US.
- C. Deploy QuickSight Enterprise edition to implement row-level security (RLS) to the sales table.
- D. Deploy QuickSight Enterprise edition and set up different VPC security groups for Australia and the US.

Answer: D

NEW QUESTION 46

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

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

Answer: B

NEW QUESTION 48

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

Answer: C

NEW QUESTION 53

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 54

An IoT company wants to release a new device that will collect data to track sleep overnight on an intelligent mattress. Sensors will send data that will be uploaded to an Amazon S3 bucket. About 2 MB of data is generated each night for each bed. Data must be processed and summarized for each user, and the results need to be available as soon as possible. Part of the process consists of time windowing and other functions. Based on tests with a Python script, every run will require about 1 GB of memory and will complete within a couple of minutes. Which solution will run the script in the MOST cost-effective way?

- A. AWS Lambda with a Python script
- B. AWS Glue with a Scala job
- C. Amazon EMR with an Apache Spark script
- D. AWS Glue with a PySpark job

Answer: A

NEW QUESTION 55

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 56

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

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

Answer: A

Explanation:

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

NEW QUESTION 59

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 64

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 and an AWS Lambda function to process the data
- B. The Lambda function will consume the data and process it to identify potential playback issues
- 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 issues
- 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 issues
- 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 issues
- 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 69

A company stores Apache Parquet-formatted files in Amazon S3. The company uses an AWS Glue Data Catalog to store the table metadata and Amazon Athena to query and analyze the data. The tables have a large number of partitions. The queries are only run on small subsets of data in the table. A data analyst adds new time partitions into the table as new data arrives. The data analyst has been asked to reduce the query runtime. Which solution will provide the MOST reduction in the query runtime?

- A. Convert the Parquet files to the CSV file format. Then attempt to query the data again
- B. Convert the Parquet files to the Apache ORC file format
- C. Then attempt to query the data again
- D. Use partition projection to speed up the processing of the partitioned table
- E. Add more partitions to be used over the table
- F. Then filter over two partitions and put all columns in the WHERE clause

Answer: C

NEW QUESTION 70

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

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

Answer: C

NEW QUESTION 71

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 72

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

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

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

Answer: ACE

Explanation:

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

NEW QUESTION 74

A company wants to provide its data analysts with uninterrupted access to the data in its Amazon Redshift cluster. All data is streamed to an Amazon S3 bucket with Amazon Kinesis Data Firehose. An AWS Glue job that is scheduled to run every 5 minutes issues a COPY command to move the data into Amazon Redshift. The amount of data delivered is uneven throughout the day, and cluster utilization is high during certain periods. The COPY command usually completes within a couple of seconds. However, when load spike occurs, locks can exist and data can be missed. Currently, the AWS Glue job is configured to run without retries, with timeout at 5 minutes and concurrency at 1.

How should a data analytics specialist configure the AWS Glue job to optimize fault tolerance and improve data availability in the Amazon Redshift cluster?

- A. Increase the number of retriee
- B. Decrease the timeout valu
- C. Increase the job concurrency.
- D. Keep the number of retries at 0. Decrease the timeout valu
- E. Increase the job concurrency.
- F. Keep the number of retries at 0. Decrease the timeout valu
- G. Keep the job concurrency at 1.
- H. Keep the number of retries at 0. Increase the timeout valu
- I. Keep the job concurrency at 1.

Answer: B

NEW QUESTION 79

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 80

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 84

A company has an encrypted Amazon Redshift cluster. The company recently enabled Amazon Redshift audit logs and needs to ensure that the audit logs are also encrypted at rest. The logs are retained for 1 year. The auditor queries the logs once a month. What is the MOST cost-effective way to meet these requirements?

- A. Encrypt the Amazon S3 bucket where the logs are stored by using AWS Key Management Service (AWS KMS). Copy the data into the Amazon Redshift cluster from Amazon S3 on a daily basi
- B. Query the data as required.
- C. Disable encryption on the Amazon Redshift cluster, configure audit logging, and encrypt the Amazon Redshift cluste
- D. Use Amazon Redshift Spectrum to query the data as required.
- E. Enable default encryption on the Amazon S3 bucket where the logs are stored by using AES-256 encryptio
- F. Copy the data into the Amazon Redshift cluster from Amazon S3 on a daily basi
- G. Query the data as required.
- H. Enable default encryption on the Amazon S3 bucket where the logs are stored by using AES-256 encryptio
- I. Use Amazon Redshift Spectrum to query the data as required.

Answer: A

NEW QUESTION 88

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 92

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

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

Answer: CD

NEW QUESTION 93

A company currently uses Amazon Athena to query its global datasets. The regional data is stored in Amazon S3 in the us-east-1 and us-west-2 Regions. The data is not encrypted. To simplify the query process and manage it centrally, the company wants to use Athena in us-west-2 to query data from Amazon S3 in both Regions. The solution should be as low-cost as possible. What should the company do to achieve this goal?

- A. Use AWS DMS to migrate the AWS Glue Data Catalog from us-east-1 to us-west-2. Run Athena queries in us-west-2.
- B. Run the AWS Glue crawler in us-west-2 to catalog datasets in all Region
- C. Once the data is crawled, run Athena queries in us-west-2.
- D. Enable cross-Region replication for the S3 buckets in us-east-1 to replicate data in us-west-2. Once the data is replicated in us-west-2, run the AWS Glue crawler there to update the AWS Glue Data Catalog in us-west-2 and run Athena queries.
- E. Update AWS Glue resource policies to provide us-east-1 AWS Glue Data Catalog access to us-west-2. Once the catalog in us-west-2 has access to the catalog in us-east-1, run Athena queries in us-west-2.

Answer: B

NEW QUESTION 98

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 100

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

Answer: D

NEW QUESTION 103

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

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

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

Which solution meets these requirements with minimal effort?

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

Answer: C

Explanation:

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

NEW QUESTION 106

A company's marketing team has asked for help in identifying a high performing long-term storage service for their data based on the following requirements:

- The data size is approximately 32 TB uncompressed.
- There is a low volume of single-row inserts each day.
- There is a high volume of aggregation queries each day.
- Multiple complex joins are performed.
- The queries typically involve a small subset of the columns in a table.

Which storage service will provide the MOST performant solution?

- A. Amazon Aurora MySQL
- B. Amazon Redshift
- C. Amazon Neptune
- D. Amazon Elasticsearch

Answer: B

NEW QUESTION 108

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

Which solution meets these requirements?

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

Answer: B

NEW QUESTION 109

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 110

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 111

A company has an application that ingests streaming data. The company needs to analyze this stream over a 5-minute timeframe to evaluate the stream for anomalies with Random Cut Forest (RCF) and summarize the current count of status codes. The source and summarized data should be persisted for future use. Which approach would enable the desired outcome while keeping data persistence costs low?

- A. Ingest the data stream with Amazon Kinesis Data Stream
- B. Have an AWS Lambda consumer evaluate the stream, collect the number status codes, and evaluate the data against a previously trained RCF model
- C. Persist the source and results as a time series to Amazon DynamoDB.
- D. Ingest the data stream with Amazon Kinesis Data Stream
- E. Have a Kinesis Data Analytics application evaluate the stream over a 5-minute window using the RCF function and summarize the count of status code
- F. Persist the source and results to Amazon S3 through output delivery to Kinesis Data Firehose.
- G. Ingest the data stream with Amazon Kinesis Data Firehose with a delivery frequency of 1 minute or 1 MB in Amazon S3. Ensure Amazon S3 triggers an event to invoke an AWS Lambda consumer that evaluates the batch data, collects the number status codes, and evaluates the data against a previously trained RCF model
- H. Persist the source and results as a time series to Amazon DynamoDB.

I. Ingest the data stream with Amazon Kinesis Data Firehose with a delivery frequency of 5 minutes or 1 MB into Amazon S3. Have a Kinesis Data Analytics application evaluate the stream over a 1-minute window using the RCF function and summarize the count of status code
J. Persist the results to Amazon S3 through a Kinesis Data Analytics output to an AWS Lambda integration.

Answer: B

NEW QUESTION 116

A company has a data lake on AWS that ingests sources of data from multiple business units and uses Amazon Athena for queries. The storage layer is Amazon S3 using the AWS Glue Data Catalog. The company wants to make the data available to its data scientists and business analysts. However, the company first needs to manage data access for Athena based on user roles and responsibilities.

What should the company do to apply these access controls with the LEAST operational overhead?

- A. Define security policy-based rules for the users and applications by role in AWS Lake Formation.
- B. Define security policy-based rules for the users and applications by role in AWS Identity and Access Management (IAM).
- C. Define security policy-based rules for the tables and columns by role in AWS Glue.
- D. Define security policy-based rules for the tables and columns by role in AWS Identity and Access Management (IAM).

Answer: D

NEW QUESTION 121

A market data company aggregates external data sources to create a detailed view of product consumption in different countries. The company wants to sell this data to external parties through a subscription. To achieve this goal, the company needs to make its data securely available to external parties who are also AWS users.

What should the company do to meet these requirements with the LEAST operational overhead?

- A. Store the data in Amazon S3. Share the data by using presigned URLs for security.
- B. Store the data in Amazon S3. Share the data by using S3 bucket ACLs.
- C. Upload the data to AWS Data Exchange for storag
- D. Share the data by using presigned URLs for security.
- E. Upload the data to AWS Data Exchange for storag
- F. Share the data by using the AWS Data Exchange sharing wizard.

Answer: A

NEW QUESTION 126

A company hosts an Apache Flink application on premises. The application processes data from several Apache Kafka clusters. The data originates from a variety of sources, such as web applications mobile apps and operational databases The company has migrated some of these sources to AWS and now wants to migrate the Flink application. The company must ensure that data that resides in databases within the VPC does not traverse the internet The application must be able to process all the data that comes from the company's AWS solution, on-premises resources and the public internet

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

- A. Implement Flink on Amazon EC2 within the company's VPC Create Amazon Managed Streaming for Apache Kafka (Amazon MSK) clusters in the VPC to collect data that comes from applications and databases within the VPC Use Amazon Kinesis Data Streams to collect data that comes from the public internet Configure Flink to have sources from Kinesis Data Streams Amazon MSK and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect
- B. Implement Flink on Amazon EC2 within the company's VPC Use Amazon Kinesis Data Streams to collect data that comes from applications and databases within the VPC and the public internet Configure Flink to have sources from Kinesis Data Streams and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect
- C. Create an Amazon Kinesis Data Analytics application by uploading the compiled Flink jar file Use Amazon Kinesis Data Streams to collect data that comes from applications and databases within the VPC and the public internet Configure the Kinesis Data Analytics application to have sources from Kinesis Data Streams and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect
- D. Create an Amazon Kinesis Data Analytics application by uploading the compiled Flink jar file Create Amazon Managed Streaming for Apache Kafka (Amazon MSK) clusters in the company's VPC to collect data that comes from applications and databases within the VPC Use Amazon Kinesis Data Streams to collect data that comes from the public internet Configure the Kinesis Data Analytics application to have sources from Kinesis Data Stream
- E. Amazon MSK and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect

Answer: D

NEW QUESTION 127

.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questions and Answers in PDF Format

AWS-Certified-Data-Analytics-Specialty Practice Exam Features:

- * AWS-Certified-Data-Analytics-Specialty Questions and Answers Updated Frequently
- * AWS-Certified-Data-Analytics-Specialty Practice Questions Verified by Expert Senior Certified Staff
- * AWS-Certified-Data-Analytics-Specialty Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * AWS-Certified-Data-Analytics-Specialty Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The AWS-Certified-Data-Analytics-Specialty Practice Test Here](#)