

# Amazon-Web-Services

## Exam Questions DAS-C01

AWS Certified Data Analytics - Specialty



**NEW QUESTION 1**

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 2**

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 3**

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

Which solution meets these requirements?

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

**Answer: D**

**Explanation:**

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

**NEW QUESTION 4**

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 5**

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 6**

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

A data analyst notes the following:

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

Which solution would help address the performance issues?

- A. Create instance fleet configurations for core and task node
- B. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB 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 7**

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 8**

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

Which solution meets these requirements?

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

**Answer: B**

**Explanation:**

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

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

**NEW QUESTION 9**

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 10**

A company wants to collect and process events data from different departments in near-real time. Before storing the data in Amazon S3, the company needs to clean the data by standardizing the format of the address and timestamp columns. The data varies in size based on the overall load at each particular point in time. A single data record can be 100 KB-10 MB.

How should a data analytics specialist design the solution for data ingestion?

- A. Use Amazon Kinesis Data Stream
- B. Configure a stream for the raw data
- C. Use a Kinesis Agent to write data to the stream
- D. Create an Amazon Kinesis Data Analytics application that reads data from the raw stream, cleanses it, and stores the output to Amazon S3.
- E. Use Amazon Kinesis Data Firehose
- F. Configure a Firehose delivery stream with a preprocessing AWS Lambda function for data cleansing
- G. Use a Kinesis Agent to write data to the delivery stream
- H. Configure Kinesis Data Firehose to deliver the data to Amazon S3.
- I. Use Amazon Managed Streaming for Apache Kafka
- J. Configure a topic for the raw data
- K. Use a Kafka producer to write data to the topic
- L. Create an application on Amazon EC2 that reads data from the topic by using the Apache Kafka consumer API, cleanses the data, and writes to Amazon S3.
- M. Use Amazon Simple Queue Service (Amazon SQS). Configure an AWS Lambda function to read events from the SQS queue and upload the events to Amazon S3.

**Answer: B**

**NEW QUESTION 10**

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 12**

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

Which solution meets these requirements in the MOST cost-effective way?

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

**Answer: C**

**NEW QUESTION 15**

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 16**

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 20**

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 21**

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 25**

A retail company leverages Amazon Athena for ad-hoc queries against an AWS Glue Data Catalog. The data analytics team manages the data catalog and data access for the company. The data analytics team wants to separate queries and manage the cost of running those queries by different workloads and teams.

Ideally, the data analysts want to group the queries run by different users within a team, store the query results in individual Amazon S3 buckets specific to each team, and enforce cost constraints on the queries run against the Data Catalog.

Which solution meets these requirements?

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

**Answer: C**

**Explanation:**

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

**NEW QUESTION 26**

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

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

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

**Answer:** B

#### NEW QUESTION 31

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

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

**Answer:** B

#### Explanation:

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

#### NEW QUESTION 35

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

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

**Answer:** A

#### Explanation:

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

#### NEW QUESTION 38

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

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

**Answer:** B

#### NEW QUESTION 43

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 45**

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

Which solution meets these requirements?

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

**Answer: B**

**NEW QUESTION 46**

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

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

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

**Answer: D**

**NEW QUESTION 50**

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

Which factors could be causing the duplicated data? (Choose two.)

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

**Answer: BD**

**NEW QUESTION 53**

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 58**

A large company has a central data lake to run analytics across different departments. Each department uses a separate AWS account and stores its data in an Amazon S3 bucket in that account. Each AWS account uses the AWS Glue Data Catalog as its data catalog. There are different data lake access requirements based on roles. Associate analysts should only have read access to their departmental data. Senior data analysts can have access in multiple departments including theirs, but for a subset of columns only.

Which solution achieves these required access patterns to minimize costs and administrative tasks?

- A. Consolidate all AWS accounts into one account.
- B. Create different S3 buckets for each department and move all the data from every account to the central data lake account.
- C. Migrate the individual data catalogs into a central data catalog and apply fine-grained permissions to give to each user the required access to tables and databases in AWS Glue and Amazon S3.
- D. Keep the account structure and the individual AWS Glue catalogs on each account.



- E. Add a central data lake account and use AWS Glue to catalog data from various account
- F. Configure cross-account access for AWS Glue crawlers to scan the data in each departmental S3 bucket to identify the schema and populate the catalog
- G. Add the senior data analysts into the central account and apply highly detailed access controls in the Data Catalog and Amazon S3.
- H. Set up an individual AWS account for the central data lake
- I. Use AWS Lake Formation to catalog the cross-account location
- J. On each individual S3 bucket, modify the bucket policy to grant S3 permissions to the Lake Formation service-linked role
- K. Use Lake Formation permissions to add fine-grained access controls to allow senior analysts to view specific tables and columns.
- L. Set up an individual AWS account for the central data lake and configure a central S3 bucket
- M. Use an AWS Lake Formation blueprint to move the data from the various buckets into the central S3 bucket
- N. On each individual bucket, modify the bucket policy to grant S3 permissions to the Lake Formation service-linked role
- O. Use Lake Formation permissions to add fine-grained access controls for both associate and senior analysts to view specific tables and columns.

**Answer: C**

**Explanation:**

Lake Formation provides secure and granular access to data through a new grant/revoke permissions model that augments AWS Identity and Access Management (IAM) policies. Analysts and data scientists can use the full portfolio of AWS analytics and machine learning services, such as Amazon Athena, to access the data. The configured Lake Formation security policies help ensure that users can access only the data that they are authorized to access. Source : <https://docs.aws.amazon.com/lake-formation/latest/dg/how-it-works.html>

**NEW QUESTION 59**

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 63**

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
- I. Create an Amazon EMR cluster using Auto Scaling for any daily analytics needs, and use Amazon Athena for the quarterly reports, with both using the same AWS Glue Data Catalog.

**Answer: A**

**NEW QUESTION 65**

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 69**

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 Spectru
- 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 neede
- G. Configure Amazon QuickSight with Amazon EMR as the data source.

**Answer:** B

**NEW QUESTION 70**

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 71**

A financial services company needs to aggregate daily stock trade data from the exchanges into a data store.

The company requires that data be streamed directly into the data store, but also occasionally allows data to be modified using SQL. The solution should integrate complex, analytic queries running with minimal latency. The solution must provide a business intelligence dashboard that enables viewing of the top contributors to anomalies in stock prices.

Which solution meets the company's requirements?

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

**Answer:** C

**NEW QUESTION 76**

.....

## 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

### DAS-C01 Practice Exam Features:

- \* DAS-C01 Questions and Answers Updated Frequently
- \* DAS-C01 Practice Questions Verified by Expert Senior Certified Staff
- \* DAS-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DAS-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DAS-C01 Practice Test Here](#)**