

DAS-C01 Dumps

AWS Certified Data Analytics - Specialty

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



NEW QUESTION 1

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

Which solution meets these requirements'?

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

Answer: C

NEW QUESTION 2

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 3

A company uses Amazon Kinesis Data Streams to ingest and process customer behavior information from application users each day. A data analytics specialist notices that its data stream is throttling. The specialist has turned on enhanced monitoring for the Kinesis data stream and has verified that the data stream did not exceed the data limits. The specialist discovers that there are hot shards.

Which solution will resolve this issue?

- A. Use a random partition key to ingest the records.
- B. Increase the number of shards. Split the size of the log records.
- C. Limit the number of records that are sent each second by the producer to match the capacity of the stream.
- D. Decrease the size of the records that are sent from the producer to match the capacity of the stream.

Answer: A

NEW QUESTION 4

A company is reading data from various customer databases that run on Amazon RDS. The databases contain many inconsistent fields. For example, a customer record field that is place_id in one database is location_id in another database. The company wants to link customer records across different databases, even when many customer record fields do not match exactly.

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

- A. Create an Amazon EMR cluster to process and analyze data in the databases. Connect to the Apache Zeppelin notebook, and use the FindMatches transform to find duplicate records in the data.
- B. Create an AWS Glue crawler to crawl the database.
- C. Use the FindMatches transform to find duplicate records in the data. Evaluate and tune the transform by evaluating performance and results of finding matches.
- D. Create an AWS Glue crawler to crawl the data in the databases. Use Amazon SageMaker to construct Apache Spark ML pipelines to find duplicate records in the data.
- E. Create an Amazon EMR cluster to process and analyze data in the database.
- F. Connect to the Apache Zeppelin notebook, and use Apache Spark ML to find duplicate records in the data.
- G. Evaluate and tune the model by evaluating performance and results of finding duplicates.

Answer: B

NEW QUESTION 5

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 6

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

A trips fact table for information on completed rides. A drivers dimension table for driver profiles. A customers fact table holding customer profile information.

The company analyzes trip details by date and destination to examine profitability by region. The drivers data rarely changes. The customers data frequently changes.

What table design provides optimal query performance?

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

Answer: C

Explanation:

<https://www.matillion.com/resources/blog/aws-redshift-performance-choosing-the-right-distribution-styles/#:~:t>

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html

NEW QUESTION 7

A retail company has 15 stores across 6 cities in the United States. Once a month, the sales team requests a visualization in Amazon QuickSight that provides the ability to easily identify revenue trends across cities and stores. The visualization also helps identify outliers that need to be examined with further analysis.

Which visual type in QuickSight meets the sales team's requirements?

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

Answer: A

NEW QUESTION 8

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

Which architecture meets these requirements?

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

Answer: D

NEW QUESTION 9

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

Which Amazon Redshift feature meets the requirements for this task?

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

Answer: D

Explanation:

Materialized views:

NEW QUESTION 10

A 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 queries. 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 data. Process incoming data from Kinesis Data Streams. Use an Amazon Kinesis Data Firehose delivery stream to write results to Amazon Redshift.

Answer: D

NEW QUESTION 10

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 11

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

Which solution meets these requirements with the least operational overhead?

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

Answer: D

Explanation:

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

NEW QUESTION 13

A company is migrating its existing on-premises ETL jobs to Amazon EMR. The code consists of a series of jobs written in Java. The company needs to reduce overhead for the system administrators without changing the underlying code. Due to the sensitivity of the data, compliance requires that the company use root device volume encryption on all nodes in the cluster. Corporate standards require that environments be provisioned through AWS CloudFormation when possible. Which solution satisfies these requirements?

- A. Install open-source Hadoop on Amazon EC2 instances with encrypted root device volume.
- B. Configure the cluster in the CloudFormation template.
- C. Use a CloudFormation template to launch an EMR cluster.
- D. In the configuration section of the cluster, define a bootstrap action to enable TLS.
- E. Create a custom AMI with encrypted root device volume.
- F. Configure Amazon EMR to use the custom AMI using the CustomAmiId property in the CloudFormation template.
- G. Use a CloudFormation template to launch an EMR cluster.
- H. In the configuration section of the cluster, define a bootstrap action to encrypt the root device volume of every node.

Answer: C

NEW QUESTION 16

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 19

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 21

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 25

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

- A. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatch.
- B. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
- C. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Data Firehose delivery stream to CloudWatch.
- D. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
- E. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatch.
- F. Configure Splunk as the end destination of the weblogs.
- G. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Firehose delivery stream to CloudWatch.
- H. Configure Amazon DynamoDB as the end destination of the weblogs.

Answer: B

Explanation:

https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL_ES_Stream.html

NEW QUESTION 26

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

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

Answer: B

NEW QUESTION 29

A company hosts an on-premises PostgreSQL database that contains historical data. An internal legacy application uses the database for read-only activities. The company's business team wants to move the data to a data lake in Amazon S3 as soon as possible and enrich the data for analytics. The company has set up an AWS Direct Connect connection between its VPC and its on-premises network. A data analytics specialist must design a solution that achieves the business team's goals with the least operational overhead. Which solution meets these requirements?

- A. Upload the data from the on-premises PostgreSQL database to Amazon S3 by using a customized batch upload process
- B. Use the AWS Glue crawler to catalog the data in Amazon S3. Use an AWS Glue job to enrich and store the result in a separate S3 bucket in Apache Parquet format
- C. Use Amazon Athena to query the data.
- D. Create an Amazon RDS for PostgreSQL database and use AWS Database Migration Service (AWS DMS) to migrate the data into Amazon RDS
- E. Use AWS Data Pipeline to copy and enrich the data from the Amazon RDS for PostgreSQL table and move the data to Amazon S3. Use Amazon Athena to query the data.
- F. Configure an AWS Glue crawler to use a JDBC connection to catalog the data in the on-premises database
- G. Use an AWS Glue job to enrich the data and save the result to Amazon S3 in Apache Parquet format
- H. Create an Amazon Redshift cluster and use Amazon Redshift Spectrum to query the data.
- I. Configure an AWS Glue crawler to use a JDBC connection to catalog the data in the on-premises database
- J. Use an AWS Glue job to enrich the data and save the result to Amazon S3 in Apache Parquet format
- K. Use Amazon Athena to query the data.

Answer: B

NEW QUESTION 30

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 31

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

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

Answer: BC

NEW QUESTION 33

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

- A. Re-arrange data in Amazon S3 to store customer data about each state in a different S3 folder within the same bucket
- B. Set up S3 bucket policies to provide marketing employees with appropriate data access under compliance control
- C. Delete the bucket policies after the project.
- D. Load tabular data from Amazon S3 to an Amazon EMR cluster using s3DistC

- E. Implement a custom Hadoop-based row-level security solution on the Hadoop Distributed File System (HDFS) to provide marketing employees with appropriate data access under compliance control
- F. Terminate the EMR cluster after the project.
- G. Load tabular data from Amazon S3 to Amazon Redshift with the COPY command
- H. Use the built-in row-level security feature in Amazon Redshift to provide marketing employees with appropriate data access under compliance control
- I. Delete the Amazon Redshift tables after the project.
- J. Load tabular data from Amazon S3 to Amazon QuickSight Enterprise edition by directly importing it as a data source
- K. Use the built-in row-level security feature in Amazon QuickSight to provide marketing employees with appropriate data access under compliance control
- L. Delete Amazon QuickSight data sources after the project is complete.

Answer: C

NEW QUESTION 37

A manufacturing company uses Amazon Connect to manage its contact center and Salesforce to manage its customer relationship management (CRM) data. The data engineering team must build a pipeline to ingest data from the contact center and CRM system into a data lake that is built on Amazon S3.

What is the MOST efficient way to collect data in the data lake with the LEAST operational overhead?

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

Answer: B

NEW QUESTION 41

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 46

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 50

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 51

An ecommerce company stores customer purchase data in Amazon RDS. The company wants a solution to store and analyze historical data. The most recent 6 months of data will be queried frequently for analytics workloads. This data is several terabytes large. Once a month, historical data for the last 5 years must be accessible and will be joined with the more recent data. The company wants to optimize performance and cost.

Which storage solution will meet these requirements?

- A. Create a read replica of the RDS database to store the most recent 6 months of data
- B. Copy the historical data into Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3 and Amazon RDS
- C. Run historical queries using Amazon Athena.
- D. Use an ETL tool to incrementally load the most recent 6 months of data into an Amazon Redshift cluster
- E. Run more frequent queries against this cluster
- F. Create a read replica of the RDS database to run queries on the historical data.
- G. Incrementally copy data from Amazon RDS to Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3. Use Amazon Athena to query the data.
- H. Incrementally copy data from Amazon RDS to Amazon S3. Load and store the most recent 6 months of data in Amazon Redshift
- I. Configure an Amazon Redshift Spectrum table to connect to all historical data.

Answer: D

NEW QUESTION 54

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 DataFrame. 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 59

A company is planning to create a data lake in Amazon S3. The company wants to create tiered storage based on access patterns and cost objectives. The solution must include support for JDBC connections from legacy clients, metadata management that allows federation for access control, and batch-based ETL using PySpark and Scala. Operational management should be limited.

Which combination of components can meet these requirements? (Choose three.)

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

Answer: BEF

NEW QUESTION 61

A company wants to run analytics on its Elastic Load Balancing logs stored in Amazon S3. A data analyst needs to be able to query all data from a desired year, month, or day. The data analyst should also be able to query a subset of the columns. The company requires minimal operational overhead and the most cost-effective solution.

Which approach meets these requirements for optimizing and querying the log data?

- A. Use an AWS Glue job nightly to transform new log files into .csv format and partition by year, month, and day
- B. Use AWS Glue crawlers to detect new partition
- C. Use Amazon Athena to query data.
- D. Launch a long-running Amazon EMR cluster that continuously transforms new log files from Amazon S3 into its Hadoop Distributed File System (HDFS) storage and partitions by year, month, and day
- E. Use Apache Presto to query the optimized format.
- F. Launch a transient Amazon EMR cluster nightly to transform new log files into Apache ORC format and partition by year, month, and day
- G. Use Amazon Redshift Spectrum to query the data.
- H. Use an AWS Glue job nightly to transform new log files into Apache Parquet format and partition by year, month, and day
- I. Use AWS Glue crawlers to detect new partition
- J. Use Amazon Athena to query data.

Answer: C

NEW QUESTION 63

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 68

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 69

An airline has been collecting metrics on flight activities for analytics. A recently completed proof of concept demonstrates how the company provides insights to data analysts to improve on-time departures. The proof of concept used objects in Amazon S3, which contained the metrics in .csv format, and used Amazon Athena for querying the data. As the amount of data increases, the data analyst wants to optimize the storage solution to improve query performance.

Which options should the data analyst use to improve performance as the data lake grows? (Choose three.)

- A. Add a randomized string to the beginning of the keys in S3 to get more throughput across partitions.
- B. Use an S3 bucket in the same account as Athena.
- C. Compress the objects to reduce the data transfer I/O.
- D. Use an S3 bucket in the same Region as Athena.
- E. Preprocess the .csv data to JSON to reduce I/O by fetching only the document keys needed by the query.
- F. Preprocess the .csv data to Apache Parquet to reduce I/O by fetching only the data blocks needed for predicate.

Answer: CDF

Explanation:

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

NEW QUESTION 72

A data analyst is designing a solution to interactively query datasets with SQL using a JDBC connection. Users will join data stored in Amazon S3 in Apache ORC format with data stored in Amazon Elasticsearch Service (Amazon ES) and Amazon Aurora MySQL.

Which solution will provide the MOST up-to-date results?

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

Answer: C

NEW QUESTION 77

Once a month, a company receives a 100 MB .csv file compressed with gzip. The file contains 50,000 property listing records and is stored in Amazon S3 Glacier. The company needs its data analyst to query a subset of the data for a specific vendor.

What is the most cost-effective solution?

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

Answer: A

NEW QUESTION 78

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 79

A company needs to store objects containing log data in JSON format. The objects are generated by eight applications running in AWS. Six of the applications generate a total of 500 KiB of data per second, and two of the applications can generate up to 2 MiB of data per second.

A data engineer wants to implement a scalable solution to capture and store usage data in an Amazon S3 bucket. The usage data objects need to be reformatted, converted to .csv format, and then compressed before they are stored in Amazon S3. The company requires the solution to include the least custom code possible and has authorized the data engineer to request a service quota increase if needed. Which solution meets these requirements?

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

Answer: A

NEW QUESTION 81

A company is sending historical datasets to Amazon S3 for storage. A data engineer at the company wants to make these datasets available for analysis using Amazon Athena. The engineer also wants to encrypt the Athena query results in an S3 results location by using AWS solutions for encryption. The requirements for encrypting the query results are as follows:

Use custom keys for encryption of the primary dataset query results. Use generic encryption for all other query results.

Provide an audit trail for the primary dataset queries that shows when the keys were used and by whom.

Which solution meets these requirements?

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

Answer: A

NEW QUESTION 85

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 90

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DAS-C01 Exam with Our Prep Materials Via below:

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