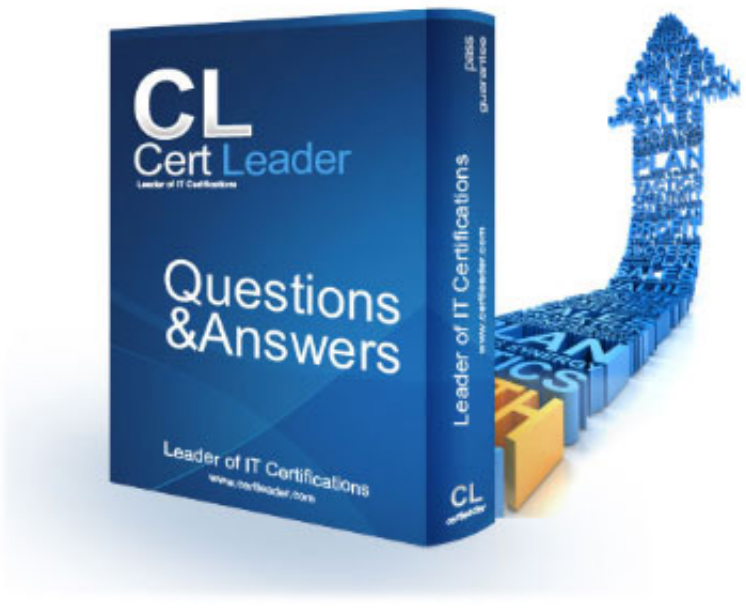
# DAS-C01 Dumps

# AWS Certified Data Analytics - Specialty

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



**NEW QUESTION 1**
A market data company aggregates external data sources to create a detailed view of product consumption in different countries. The company wants to sell this data to external parties through a subscription. To achieve this goal, the company needs to make its data securely available to external parties who are also AWS users.
What should the company do to meet these requirements with the LEAST operational overhead?

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

**Answer:** A

**NEW QUESTION 2**
A company has several Amazon EC2 instances sitting behind an Application Load Balancer (ALB) The company wants its IT Infrastructure team to analyze the IP addresses coming into the company's ALB The ALB is configured to store access logs in Amazon S3 The access logs create about 1 TB of data each day, and access to the data will be infrequent The company needs a solution that is scalable, cost-effective and has minimal maintenance requirements
Which solution meets these requirements?

A. Copy the data into Amazon Redshift and query the data
B. Use Amazon EMR and Apache Hive to query the S3 data
C. Use Amazon Athena to query the S3 data
D. Use Amazon Redshift Spectrum to query the S3 data

**Answer:** D

**NEW QUESTION 3**
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 dat
C. Use Amazon Kinesis Data Streams to uncover NRT monitoring metric
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 logdata to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring Use Amazon QuickSight for the dashboards.

**Answer:** C

**NEW QUESTION 4**
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 Catalo
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 jo
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 5**
A university intends to use Amazon Kinesis Data Firehose to collect JSON-formatted batches of water quality readings in Amazon S3. The readings are from 50 sensors scattered across a local lake. Students will query the stored data using Amazon Athena to observe changes in a captured metric over time, such as water temperature or acidity. Interest has grown in the study, prompting the university to reconsider how data will be stored.

Which data format and partitioning choices will MOST significantly reduce costs? (Choose two.)

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

**Answer:** CD

**NEW QUESTION 6**
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 7**
A data analytics specialist is building an automated ETL ingestion pipeline using AWS Glue to ingest compressed files that have been uploaded to an Amazon S3 bucket. The ingestion pipeline should support incremental data processing.
Which AWS Glue feature should the data analytics specialist use to meet this requirement?

A. Workflows
B. Triggers
C. Job bookmarks
D. Classifiers

**Answer:** C

**NEW QUESTION 8**
An operations team notices that a few AWS Glue jobs for a given ETL application are failing. The AWS Glue jobs read a large number of small JSON files from an Amazon S3 bucket and write the data to a different S3 bucket in Apache Parquet format with no major transformations. Upon initial investigation, a data engineer notices the following error message in the History tab on the AWS Glue console: “Command Failed with Exit Code 1.”
Upon further investigation, the data engineer notices that the driver memory profile of the failed jobs crosses the safe threshold of 50% usage quickly and reaches 90–95% soon after. The average memory usage across all executors continues to be less than 4%.
The data engineer also notices the following error while examining the related Amazon CloudWatch Logs. What should the data engineer do to solve the failure in the MOST cost-effective way?

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

**Answer:** B

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

**NEW QUESTION 9**
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 10**
A company has an encrypted Amazon Redshift cluster. The company recently enabled Amazon Redshift audit logs and needs to ensure that the audit logs are also encrypted at rest. The logs are retained for 1 year. The auditor queries the logs once a month.
What is the MOST cost-effective way to meet these requirements?

A. Encrypt the Amazon S3 bucket where the logs are stored by using AWS Key Management Service (AWS KMS). Copy the data into the Amazon Redshift cluster

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

**Answer:** A

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

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

**Answer:** A

**Explanation:**
Matching Records with AWS Lake Formation FindMatches

**NEW QUESTION 13**
A company that monitors weather conditions from remote construction sites is setting up a solution to collect temperature data from the following two weather stations.

- Station A, which has 10 sensors
- Station B, which has five sensors

These weather stations were placed by onsite subject-matter experts.
Each sensor has a unique ID. The data collected from each sensor will be collected using Amazon Kinesis Data Streams.
Based on the total incoming and outgoing data throughput, a single Amazon Kinesis data stream with two shards is created. Two partition keys are created based on the station names. During testing, there is a bottleneck on data coming from Station A, but not from Station B. Upon review, it is confirmed that the total stream throughput is still less than the allocated Kinesis Data Streams throughput.
How can this bottleneck be resolved without increasing the overall cost and complexity of the solution, while retaining the data collection quality requirements?

A. Increase the number of shards in Kinesis Data Streams to increase the level of parallelism.
B. Create a separate Kinesis data stream for Station A with two shards, and stream Station A sensor data to the new stream.
C. Modify the partition key to use the sensor ID instead of the station name.
D. Reduce the number of sensors in Station A from 10 to 5 sensors.

**Answer:** C

**Explanation:**
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding.html
"Splitting increases the number of shards in your stream and therefore increases the data capacity of the stream. Because you are charged on a per-shard basis, splitting increases the cost of your stream"

**NEW QUESTION 15**
A company uses Amazon Elasticsearch Service (Amazon ES) to store and analyze its website clickstream data. The company ingests 1 TB of data daily using Amazon Kinesis Data Firehose and stores one day's worth of data in an Amazon ES cluster.
The company has very slow query performance on the Amazon ES index and occasionally sees errors from Kinesis Data Firehose when attempting to write to the index. The Amazon ES cluster has 10 nodes running a single index and 3 dedicated master nodes. Each data node has 1.5 TB of Amazon EBS storage attached and the cluster is configured with 1,000 shards. Occasionally, JVMMemoryPressure errors are found in the cluster logs.
Which solution will improve the performance of Amazon ES?

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

**Answer:** C

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

**NEW QUESTION 16**
A company wants to improve the data load time of a sales data dashboard. Data has been collected as .csv files and stored within an Amazon S3 bucket that is partitioned by date. The data is then loaded to an Amazon Redshift data warehouse for frequent analysis. The data volume is up to 500 GB per day.
Which solution will improve the data loading performance?

A. Compress .csv files and use an INSERT statement to ingest data into Amazon Redshift.
B. Split large .csv files, then use a COPY command to load data into Amazon Redshift.
C. Use Amazon Kinesis Data Firehose to ingest data into Amazon Redshift.

D. Load the .csv files in an unsorted key order and vacuum the table in Amazon Redshift.

**Answer:** B

**Explanation:**
https://docs.aws.amazon.com/redshift/latest/dg/c_loading-data-best-practices.html

**NEW QUESTION 21**
A company launched a service that produces millions of messages every day and uses Amazon Kinesis Data Streams as the streaming service.
The company uses the Kinesis SDK to write data to Kinesis Data Streams. A few months after launch, a data analyst found that write performance is significantly reduced. The data analyst investigated the metrics and determined that Kinesis is throttling the write requests. The data analyst wants to address this issue without significant changes to the architecture.
Which actions should the data analyst take to resolve this issue? (Choose two.)

A. Increase the Kinesis Data Streams retention period to reduce throttling.
B. Replace the Kinesis API-based data ingestion mechanism with Kinesis Agent.
C. Increase the number of shards in the stream using the UpdateShardCount API.
D. Choose partition keys in a way that results in a uniform record distribution across shards.
E. Customize the application code to include retry logic to improve performance.

**Answer:** CD

**Explanation:**
https://aws.amazon.com/blogs/big-data/under-the-hood-scaling-your-kinesis-data-streams/

**NEW QUESTION 26**
A telecommunications company is looking for an anomaly-detection solution to identify fraudulent calls. The company currently uses Amazon Kinesis to stream voice call records in a JSON format from its on-premises database to Amazon S3. The existing dataset contains voice call records with 200 columns. To detect fraudulent calls, the solution would need to look at 5 of these columns only.
The company is interested in a cost-effective solution using AWS that requires minimal effort and experience in anomaly-detection algorithms.
Which solution meets these requirements?

A. Use an AWS Glue job to transform the data from JSON to Apache Parque
B. Use AWS Glue crawlers todiscover the schema and build the AWS Glue Data Catalo
C. Use Amazon Athena to create a table with a subset of column
D. Use Amazon QuickSight to visualize the data and then use Amazon QuickSight machine learning-powered anomaly detection.
E. Use Kinesis Data Firehose to detect anomalies on a data stream from Kinesis by running SQL queries, which compute an anomaly score for all calls and store the output in Amazon RD
F. Use Amazon Athena to build a dataset and Amazon QuickSight to visualize the results.
G. Use an AWS Glue job to transform the data from JSON to Apache Parque
H. Use AWS Glue crawlers to discover the schema and build the AWS Glue Data Catalo
I. Use Amazon SageMaker to build an anomaly detection model that can detect fraudulent calls by ingesting data from Amazon S3.
J. Use Kinesis Data Analytics to detect anomalies on a data stream from Kinesis by running SQL queries, which compute an anomaly score for all call
K. Connect Amazon QuickSight to Kinesis Data Analytics to visualize the anomaly scores.

**Answer:** A

**NEW QUESTION 29**
An airline has .csv-formatted data stored in Amazon S3 with an AWS Glue Data Catalog. Data analysts want to join this data with call center data stored in Amazon Redshift as part of a dally batch process. The Amazon Redshift cluster is already under a heavy load. The solution must be managed, serverless, well-functioning, and minimize the load on the existing Amazon Redshift cluster. The solution should also require minimal effort and development activity.
Which solution meets these requirements?

A. Unload the call center data from Amazon Redshift to Amazon S3 using an AWS Lambda function.Perform the join with AWS Glue ETL scripts.
B. Export the call center data from Amazon Redshift using a Python shell in AWS Glu
C. Perform the join with AWS Glue ETL scripts.
D. Create an external table using Amazon Redshift Spectrum for the call center data and perform the join with Amazon Redshift.
E. Export the call center data from Amazon Redshift to Amazon EMR using Apache Sqoo
F. Perform the join with Apache Hive.

**Answer:** C

**Explanation:**
https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-tables.html

**NEW QUESTION 31**
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 33**
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 accoun
B. Create different S3 buckets for each department and move all the data from every account to the central data lake accoun
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 accoun
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 catalo
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 lak
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 rol
K. Use Lake Formation permissions to addfine-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 bucke
M. Use an AWS Lake Formation blueprint to move the data from the various buckets into the central S3 bucke
N. On each individual bucket, modify the bucket policy to grant S3 permissions to the Lake Formation service-linked rol
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 35**
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 37**
An online retailer needs to deploy a product sales reporting solution. The source data is exported from an external online transaction processing (OLTP) system for reporting. Roll-up data is calculated each day for the previous day's activities. The reporting system has the following requirements:
Have the daily roll-up data readily available for 1 year.
After 1 year, archive the daily roll-up data for occasional but immediate access.
The source data exports stored in the reporting system must be retained for 5 years. Query access will be needed only for re-evaluation, which may occur within the first 90 days.
Which combination of actions will meet these requirements while keeping storage costs to a minimum? (Choose two.)

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

**Answer:** AD

**NEW QUESTION 40**
A media company wants to perform machine learning and analytics on the data residing in its Amazon S3 data lake. There are two data transformation requirements that will enable the consumers within the company to create reports:
- Daily transformations of 300 GB of data with different file formats landing in Amazon S3 at a scheduled time.
- One-time transformations of terabytes of archived data residing in the S3 data lake.

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

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

**Answer:** ADE

**NEW QUESTION 44**
A data analyst is using AWS Glue to organize, cleanse, validate, and format a 200 GB dataset. The data analyst triggered the job to run with the Standard worker type. After 3 hours, the AWS Glue job status is still RUNNING. Logs from the job run show no error codes. The data analyst wants to improve the job execution time without overprovisioning.
Which actions should the data analyst take?

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

**Answer:** B

**NEW QUESTION 49**
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 51**
A banking company wants to collect large volumes of transactional data using Amazon Kinesis Data Streams for real-time analytics. The company uses PutRecord to send data to Amazon Kinesis, and has observed network outages during certain times of the day. The company wants to obtain exactly once semantics for the entire processing pipeline.
What should the company do to obtain these characteristics?

A. Design the application so it can remove duplicates during processing be embedding a unique ID in each record.
B. Rely on the processing semantics of Amazon Kinesis Data Analytics to avoid duplicate processing of events.
C. Design the data producer so events are not ingested into Kinesis Data Streams multiple times.
D. Rely on the exactly one processing semantics of Apache Flink and Apache Spark Streaming included in Amazon EMR.

**Answer:** A

**NEW QUESTION 55**
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 Redshif
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 59**
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 61**
A real estate company has a mission-critical application using Apache HBase in Amazon EMR. Amazon EMR is configured with a single master node. The company has over 5 TB of data stored on an Hadoop Distributed File System (HDFS). The company wants a cost-effective solution to make its HBase data highly available. Which architectural pattern meets company's requirements?

A. Use Spot Instances for core and task nodes and a Reserved Instance for the EMR master node.Configure the EMR cluster with multiple master node
B. Schedule automated snapshots using Amazon EventBridge.
C. Store the data on an EMR File System (EMRFS) instead of HDF
D. Enable EMRFS consistent view.Create an EMR HBase cluster with multiple master node
E. Point the HBase root directory to an Amazon S3 bucket.
F. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view.Run two separate EMR clusters in two different Availability Zone
G. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.
H. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view.Create a primary EMR HBase cluster with multiple master node
I. Create a secondary EMR HBase read- replica cluster in a separate Availability Zon
J. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.

**Answer:** D

**NEW QUESTION 64**
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 67**
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 da
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 da
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 da
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 da
I. Use AWS Glue crawlers to detect new partition
J. Use Amazon Athena to querydata.

**Answer:** C

**NEW QUESTION 70**
A company that produces network devices has millions of users. Data is collected from the devices on an hourly basis and stored in an Amazon S3 data lake.

The company runs analyses on the last 24 hours of data flow logs for abnormality detection and to troubleshoot and resolve user issues. The company also analyzes historical logs dating back 2 years to discover patterns and look for improvement opportunities.
The data flow logs contain many metrics, such as date, timestamp, source IP, and target IP. There are about 10 billion events every day.
How should this data be stored for optimal performance?

A. In Apache ORC partitioned by date and sorted by source IP
B. In compressed .csv partitioned by date and sorted by source IP
C. In Apache Parquet partitioned by source IP and sorted by date
D. In compressed nested JSON partitioned by source IP and sorted by date

**Answer:** A

**NEW QUESTION 73**
An online retail company is migrating its reporting system to AWS. The company's legacy system runs data processing on online transactions using a complex series of nested Apache Hive queries. Transactional data is exported from the online system to the reporting system several times a day. Schemas in the files are stable between updates.
A data analyst wants to quickly migrate the data processing to AWS, so any code changes should be minimized. To keep storage costs low, the data analyst decides to store the data in Amazon S3. It is vital that the data from the reports and associated analytics is completely up to date based on the data in Amazon S3.
Which solution meets these requirements?

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

**Answer:** A

**NEW QUESTION 76**
An online retail company uses Amazon Redshift to store historical sales transactions. The company is required to encrypt data at rest in the clusters to comply with the Payment Card Industry Data Security Standard (PCI DSS). A corporate governance policy mandates management of encryption keys using an on-premises hardware security module (HSM).
Which solution meets these requirements?

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

**Answer:** B

**NEW QUESTION 80**
A data analyst is designing a solution to interactively query datasets with SQL using a JDBC connection. Users will join data stored in Amazon S3 in Apache ORC format with data stored in Amazon Elasticsearch Service (Amazon ES) and Amazon Aurora MySQL.
Which solution will provide the MOST up-to-date results?

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

**Answer:** C

**NEW QUESTION 83**
A financial company uses Apache Hive on Amazon EMR for ad-hoc queries. Users are complaining of sluggish performance.
A data analyst notes the following:
- Approximately 90% of queries are submitted 1 hour after the market opens.
- Hadoop Distributed File System (HDFS) utilization never exceeds 10%.

Which solution would help address the performance issues?

A. Create instance fleet configurations for core and task node
B. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metri
C. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch CapacityRemainingGB metric.

D. Create instance fleet configurations for core and task node
E. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch YARNMemoryAvailablePercentage metri
F. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch YARNMemoryAvailablePercentage metric.
G. Create instance group configurations for core and task node
H. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metri
I. Create anautomatic scaling policy to scale in the instance groups based on the CloudWatch CapacityRemainingGB metric.
J. Create instance group configurations for core and task node
K. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch YARNMemoryAvailablePercentage metri
L. Create an automatic scaling policy to scale in the instance groups based on the CloudWatch YARNMemoryAvailablePercentage metric.

**Answer:** D

**Explanation:**
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html

**NEW QUESTION 85**
An online gaming company is using an Amazon Kinesis Data Analytics SQL application with a Kinesis data stream as its source. The source sends three non-null fields to the application: player_id, score, and us_5_digit_zip_code.
A data analyst has a .csv mapping file that maps a small number of us_5_digit_zip_code values to a territory code. The data analyst needs to include the territory code, if one exists, as an additional output of the Kinesis Data Analytics application.
How should the data analyst meet this requirement while minimizing costs?

A. Store the contents of the mapping file in an Amazon DynamoDB tabl
B. Preprocess the records as they arrive in the Kinesis Data Analytics application with an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exist
C. Change the SQL query in the application to include the new field in the SELECT statement.
D. Store the mapping file in an Amazon S3 bucket and configure the reference data column headers for the.csv file in the Kinesis Data Analytics applicatio
E. Change the SQL query in the application to include a join to the file's S3 Amazon Resource Name (ARN), and add the territory code field to the SELECT columns.
F. Store the mapping file in an Amazon S3 bucket and configure it as a reference data source for the Kinesis Data Analytics applicatio
G. Change the SQL query in the application to include a join to the reference table and add the territory code field to the SELECT columns.
H. Store the contents of the mapping file in an Amazon DynamoDB tabl
I. Change the Kinesis Data Analytics application to send its output to an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exist
J. Forward the record from the Lambda function to the original application destination.

**Answer:** C

**NEW QUESTION 89**
.....

# 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