

➤ **Vendor: Amazon**

➤ **Exam Code: DAS-C01**

➤ **Exam Name: AWS Certified Data Analytics - Specialty (DAS-C01) Exam**

➤ **New Updated Questions from [Braindump2go](https://www.braindump2go.com) (Updated in [January/2022](https://www.braindump2go.com))**

[Visit Braindump2go and Download Full Version DAS-C01 Exam Dumps](#)

QUESTION 141

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

QUESTION 142

An analytics software as a service (SaaS) provider wants to offer its customers business intelligence (BI) reporting capabilities that are self-service. The provider is using Amazon QuickSight to build these reports. The data for the reports resides in a multi-tenant database, but each customer should only be able to access their own data.

The provider wants to give customers two user role options:

- Read-only users for individuals who only need to view dashboards
- Power users for individuals who are allowed to create and share new dashboards with other users

Which QuickSight feature allows the provider to meet these requirements?

- A. Embedded dashboards
- B. Table calculations
- C. Isolated namespaces
- D. SPICE

Answer: A

QUESTION 143

A manufacturing company has many IoT devices in different facilities across the world. The company is using Amazon Kinesis Data Streams to collect the data from the devices. The company's operations team has started to observe many WroteThroughputExceeded exceptions. The operations team determines that the reason is the number of records that are being written to certain shards. The data contains device ID, capture date, measurement type, measurement value, and facility ID. The facility ID is used as the partition key.

[DAS-C01 Exam Dumps](#) [DAS-C01 Exam Questions](#) [DAS-C01 PDF Dumps](#) [DAS-C01 VCE Dumps](#)

<https://www.braindump2go.com/das-c01.html>

Which action will resolve this issue?

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

Answer: B

QUESTION 144

A reseller that has thousands of AWS accounts receives AWS Cost and Usage Reports in an Amazon S3 bucket. The reports are delivered to the S3 bucket in the following format:

`<example-report-prefix>/<example-report-name>/yyyymmdd-yyyymmdd/<example-report-name> parquet`

An AWS Glue crawler crawls the S3 bucket and populates an AWS Glue Data Catalog with a table Business analysts use Amazon Athena to query the table and create monthly summary reports for the AWS accounts. The business analysts are experiencing slow queries because of the accumulation of reports from the last 5 years. The business analysts want the operations team to make changes to improve query performance.

Which action should the operations team take to meet these requirements?

- A. Change the file format to csv.zip.
- B. Partition the data by date and account ID
- C. Partition the data by month and account ID
- D. Partition the data by account ID, year, and month

Answer: B

QUESTION 145

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

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

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

Answer: D

QUESTION 146

[DAS-C01 Exam Dumps](#) [DAS-C01 Exam Questions](#) [DAS-C01 PDF Dumps](#) [DAS-C01 VCE Dumps](#)

<https://www.braindump2go.com/das-c01.html>

A hospital uses an electronic health records (EHR) system to collect two types of data.

- Patient information, which includes a patient's name and address
 - Diagnostic tests conducted and the results of these tests
- Patient information is expected to change periodically. Existing diagnostic test data never changes and only new records are added.

The hospital runs an Amazon Redshift cluster with four dc2.large nodes and wants to automate the ingestion of the patient information and diagnostic test data into respective Amazon Redshift tables for analysis. The EHR system exports data as CSV files to an Amazon S3 bucket on a daily basis. Two sets of CSV files are generated. One set of files is for patient information with updates, deletes, and inserts. The other set of files is for new diagnostic test data only. What is the MOST cost-effective solution to meet these requirements?

- A. Use Amazon EMR with Apache Hudi.
Run daily ETL jobs using Apache Spark and the Amazon Redshift JDBC driver
- B. Use an AWS Glue crawler to catalog the data in Amazon S3
Use Amazon Redshift Spectrum to perform scheduled queries of the data in Amazon S3 and ingest the data into the patient information table and the diagnostic tests table.
- C. Use an AWS Lambda function to run a COPY command that appends new diagnostic test data to the diagnostic tests table
Run another COPY command to load the patient information data into the staging tables
Use a stored procedure to handle create, update, and delete operations for the patient information table
- D. Use AWS Database Migration Service (AWS DMS) to collect and process change data capture (CDC) records
Use the COPY command to load patient information data into the staging tables.
Use a stored procedure to handle create, update and delete operations for the patient information table

Answer: B

QUESTION 147

A utility company wants to visualize data for energy usage on a daily basis in Amazon QuickSight.

A data analytics specialist at the company has built a data pipeline to collect and ingest the data into Amazon S3. Each day the data is stored in an individual csv file in an S3 bucket.

This is an example of the naming structure: 20210707_data.csv 20210708_data.csv.

To allow for data querying in QuickSight through Amazon Athena, the specialist used an AWS Glue crawler to create a table with the path "s3://powertransformer/20210707_data.csv".

However, when the data is queried, it returns zero rows.

How can this issue be resolved?

- A. Modify the IAM policy for the AWS Glue crawler to access Amazon S3.
- B. Ingest the files again.
- C. Store the files in Apache Parquet format.
- D. Update the table path to "s3://powertransformer/".

Answer: D

QUESTION 148

A large energy company is using Amazon QuickSight to build dashboards and report the historical usage data of its customers. This data is hosted in Amazon Redshift. The reports need access to all the fact tables' billions of records to create aggregation in real time grouping by multiple dimensions. A data analyst created the dataset in QuickSight by using a SQL query and not SPICE. Business users have noted that the response time is not fast enough to meet their needs.

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

- A. Use QuickSight to modify the current dataset to use SPICE
- B. Use AWS Glue to create an Apache Spark job that joins the fact table with the dimensions.
Load the data into a new table

- C. Use Amazon Redshift to create a materialized view that joins the fact table with the dimensions
- D. Use Amazon Redshift to create a stored procedure that joins the fact table with the dimensions. Load the data into a new table

Answer: A

QUESTION 149

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

Which combination of options will provide the MOST cost-effective solution? (Select TWO)

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

Answer: BD

QUESTION 150

A company uses an Amazon EMR cluster with 50 nodes to process operational data and make the data available for data analysts. These jobs run nightly use Apache Hive with the Apache Jez framework as a processing model and write results to Hadoop Distributed File System (HDFS) In the last few weeks, jobs are failing and are producing the following error message "File could only be replicated to 0 nodes instead of 1". A data analytics specialist checks the DataNode logs the NameNode logs and network connectivity for potential issues that could have prevented HDFS from replicating data. The data analytics specialist rules out these factors as causes for the issue.

Which solution will prevent the jobs from failing'?

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

Answer: C

QUESTION 151

A company recently created a test AWS account to use for a development environment. The company also created a production AWS account in another AWS Region. As part of its security testing the company wants to send log data from Amazon CloudWatch Logs in its production account to an Amazon Kinesis data stream in its test account.

Which solution will allow the company to accomplish this goal?

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

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

Answer: D

QUESTION 152

A bank wants to migrate a Teradata data warehouse to the AWS Cloud. The bank needs a solution for reading large amounts of data and requires the highest possible performance. The solution also must maintain the separation of storage and compute. Which solution meets these requirements?

- A. Use Amazon Athena to query the data in Amazon S3
- B. Use Amazon Redshift with dense compute nodes to query the data in Amazon Redshift managed storage
- C. Use Amazon Redshift with RA3 nodes to query the data in Amazon Redshift managed storage
- D. Use PrestoDB on Amazon EMR to query the data in Amazon S3

Answer: C

QUESTION 153

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

QUESTION 154

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

QUESTION 155

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

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

Answer: C