

➤ **Vendor: Amazon**

➤ **Exam Code: DAS-C01**

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

➤ **New Updated Questions from [Braindump2go](#) (Updated in [October/2021](#))**

Visit Braindump2go and Download Full Version DAS-C01 Exam Dumps

QUESTION 122

A company has a marketing department and a finance department. The departments are storing data in Amazon S3 in their own AWS accounts in AWS Organizations. Both departments use AWS Lake Formation to catalog and secure their data. The departments have some databases and tables that share common names.

The marketing department needs to securely access some tables from the finance department.

Which two steps are required for this process? (Choose two.)

- A. The finance department grants Lake Formation permissions for the tables to the external account for the marketing department.
- B. The finance department creates cross-account IAM permissions to the table for the marketing department role.
- C. The marketing department creates an IAM role that has permissions to the Lake Formation tables.

Answer: AB

Explanation:

Granting Lake Formation Permissions

Creating an IAM role (AWS CLI)

Reference:

<https://docs.aws.amazon.com/lake-formation/latest/dg/lake-formation-permissions.html>

https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_create_for-user.html

QUESTION 123

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

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

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

Answer: B

QUESTION 124

A company receives data from its vendor in JSON format with a timestamp in the file name. The vendor uploads the data to an Amazon S3 bucket, and the data is registered into the company's data lake for analysis and reporting. The company has configured an S3 Lifecycle policy to archive all files to S3 Glacier after 5 days.

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

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

The company wants to ensure that its AWS Glue crawler catalogs data only from S3 Standard storage and ignores the archived files. A data analytics specialist must implement a solution to achieve this goal without changing the current S3 bucket configuration.

Which solution meets these requirements?

- A. Use the exclude patterns feature of AWS Glue to identify the S3 Glacier files for the crawler to exclude.
- B. Schedule an automation job that uses AWS Lambda to move files from the original S3 bucket to a new S3 bucket for S3 Glacier storage.
- C. Use the excludeStorageClasses property in the AWS Glue Data Catalog table to exclude files on S3 Glacier storage.
- D. Use the include patterns feature of AWS Glue to identify the S3 Standard files for the crawler to include.

Answer: A

Explanation:

<https://docs.aws.amazon.com/glue/latest/dg/define-crawler.html#crawler-data-stores-exclude>

QUESTION 125

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. Run the query from MySQL Workbench instead of Athena directly.
- B. Use Athena to extract the data and store it in Apache Parquet format on a daily basis. Query the extracted data.
- C. Run a daily AWS Glue ETL job to convert the data files to Apache Parquet and to partition the converted files. Create a periodic AWS Glue crawler to automatically crawl the partitioned data on a daily basis.
- D. Run a daily AWS Glue ETL job to compress the data files by using the .gzip format. Query the compressed data.
- E. Run a daily AWS Glue ETL job to compress the data files by using the .lzo format. Query the compressed data.

Answer: BC

Explanation:

<https://www.upsolver.com/blog/apache-parquet-why-use>

<https://aws.amazon.com/blogs/big-data/work-with-partitioned-data-in-aws-glue/>

QUESTION 126

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. Use SSE-S3 for the other datasets.
- B. Use server-side encryption with customer-provided encryption keys (SSE-C) for the primary dataset.

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

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

Use server-side encryption with S3 managed encryption keys (SSE-S3) for the other datasets.

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

Answer: A

Explanation:

https://d1.awsstatic.com/product-marketing/S3/Amazon_S3_Security_eBook_2020.pdf

QUESTION 127

A large telecommunications company is planning to set up a data catalog and metadata management for multiple data sources running on AWS. The catalog will be used to maintain the metadata of all the objects stored in the data stores. The data stores are composed of structured sources like Amazon RDS and Amazon Redshift, and semistructured sources like JSON and XML files stored in Amazon S3. The catalog must be updated on a regular basis, be able to detect the changes to object metadata, and require the least possible administration.

Which solution meets these requirements?

- A. Use Amazon Aurora as the data catalog. Create AWS Lambda functions that will connect and gather the metadata information from multiple sources and update the data catalog in Aurora. Schedule the Lambda functions periodically.
- B. Use the AWS Glue Data Catalog as the central metadata repository. Use AWS Glue crawlers to connect to multiple data stores and update the Data Catalog with metadata changes. Schedule the crawlers periodically to update the metadata catalog.
- C. Use Amazon DynamoDB as the data catalog. Create AWS Lambda functions that will connect and gather the metadata information from multiple sources and update the DynamoDB catalog. Schedule the Lambda functions periodically.
- D. Use the AWS Glue Data Catalog as the central metadata repository. Extract the schema for RDS and Amazon Redshift sources and build the Data Catalog. Use AWS crawlers for data stored in Amazon S3 to infer the schema and automatically update the Data Catalog.

Answer: D

Explanation:

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive-metastore-glue.html>

QUESTION 128

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.

Reference: <https://docs.aws.amazon.com/redshift/latest/dg/queries-troubleshooting.html>

QUESTION 129

A power utility company is deploying thousands of smart meters to obtain real-time updates about power consumption. The company is using Amazon Kinesis Data Streams to collect the data streams from smart meters. The consumer application uses the Kinesis Client Library (KCL) to retrieve the stream data. The company has only one consumer application.

The company observes an average of 1 second of latency from the moment that a record is written to the stream until the record is read by a consumer application. The company must reduce this latency to 500 milliseconds.

Which solution meets these requirements?

- A. Use enhanced fan-out in Kinesis Data Streams.
- B. Increase the number of shards for the Kinesis data stream.
- C. Reduce the propagation delay by overriding the KCL default settings.
- D. Develop consumers by using Amazon Kinesis Data Firehose.

Answer: C

Explanation:

The KCL defaults are set to follow the best practice of polling every 1 second. This default results in average propagation delays that are typically below 1 second.

Reference: <https://docs.aws.amazon.com/streams/latest/dev/kinesis-low-latency.html>

QUESTION 130

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

Which solution meets these requirements?

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

Answer: B

Explanation:

<https://aws.amazon.com/about-aws/whats-new/2019/>

QUESTION 131

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

Explanation:

<https://aws.amazon.com/kinesis/data-firehose/?kinesis-blogs.sort-by=item.additionalFields.createdDate&kinesis-blogs.sort-order=desc>

QUESTION 132

A manufacturing company wants to create an operational analytics dashboard to visualize metrics from equipment in near-real time. The company uses Amazon Kinesis Data Streams to stream the data to other applications. The dashboard must automatically refresh every 5 seconds. A data analytics specialist must design a solution that requires the least possible implementation effort.

Which solution meets these requirements?

- A. Use Amazon Kinesis Data Firehose to store the data in Amazon S3. Use Amazon QuickSight to build the dashboard.
- B. Use Apache Spark Streaming on Amazon EMR to read the data in near-real time. Develop a custom application for the dashboard by using D3.js.
- C. Use Amazon Kinesis Data Firehose to push the data into an Amazon Elasticsearch Service (Amazon ES) cluster. Visualize the data by using a Kibana dashboard.
- D. Use AWS Glue streaming ETL to store the data in Amazon S3. Use Amazon QuickSight to build the dashboard.

Answer: B

Explanation:

<https://aws.amazon.com/blogs/big-data/analyze-a-time-series-in-real-time-with-aws-lambda-amazon-kinesis-and-amazon-dynamodb-streams/>