DAS-C01 Q&As

AWS Certified Data Analytics - Specialty (DAS-C01)

Pass Amazon DAS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.com/das-c01.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

A marketing company is storing its campaign response data in Amazon S3. A consistent set of sources has generated the data for each campaign. The data is saved into Amazon S3 as .csv files. A business analyst will use Amazon Athena to analyze each campaign\\'s data. The company needs the cost of ongoing data analysis with Athena to be minimized.

Which combination of actions should a data analytics specialist take to meet these requirements? (Choose two.)

A. Convert the .csv files to Apache Parquet.

B. Convert the .csv files to Apache Avro.

C. Partition the data by campaign.

D. Partition the data by source.

E. Compress the .csv files.

Correct Answer: AC

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

QUESTION 2

An advertising company has a data lake that is built on Amazon S3. The company uses AWS Glue Data Catalog to maintain the metadata. The data lake is several years old and its overall size has increased exponentially as additional data sources and metadata are stored in the data lake. The data lake administrator wants to implement a mechanism to simplify permissions management between Amazon S3 and the Data Catalog to keep them in sync.

Which solution will simplify permissions management with minimal development effort?

A. Set AWS Identity and Access Management (IAM) permissions for AWS Glue

B. Use AWS Lake Formation permissions

C. Manage AWS Glue and S3 permissions by using bucket policies

D. Use Amazon Cognito user pools

Correct Answer: B

Reference: https://docs.aws.amazon.com/lake-formation/latest/dg/how-it-works.html

QUESTION 3

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

Correct Answer: A

QUESTION 4

A mining company is using Amazon S3 as its data lake. The company wants to analyze the data collected by the sensors in its mines. A data pipeline is being built to capture data from the sensors, ingest the data into an S3 bucket, and convert the data to Apache Parquet format. The data pipeline must be processed in near-real time. The data will be used for on-demand queries with Amazon Athena.

Which solution will meet these requirements?

A. Use Amazon Kinesis Data Firehose to invoke an AWS Lambda function that converts the data to Parquet format and stores the data in Amazon S3.

B. Use Amazon Kinesis Data Streams to invoke an AWS Lambda function that converts the data to Parquet format and stores the data in Amazon S3.

C. Use AWS DataSync to invoke an AWS Lambda function that converts the data to Parquet format and stores the data in Amazon S3.

D. Use Amazon Simple Queue Service (Amazon SQS) to stream data directly to an AWS Glue job that converts the data to Parquet format and stores the data in Amazon S3.

Correct Answer: A

QUESTION 5

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.

Correct Answer: AC