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.leads4pass.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

SATISFACTION GUARANTEED
100%
SATISFACTION GUARANTEED

QUESTION 1

A financial services company hosts its data warehouse on a Hadoop cluster located in an on-premises data center. The data is 300 TB in size and grows by 1 TB each day. The data is generated in real time from the company\\'s trading system. The raw data is transformed at the end of the trading day using a custom tool running on the Hadoop cluster.

The company is migrating its data warehouse to AWS using a managed data warehouse product provided by a third party that can ingest data from Amazon S3. The company has already established a 10 Gbps connection with an AWS Region using AWS Direct Connect. The company is required by its security and regulatory compliance policies to not transfer data over the public internet. The company wants to minimize changes to its custom tool for data transformation. The company also plans to eliminate the on-premises Hadoop cluster after the migration.

Which solution MOST cost-effectively meets these requirements?

A. Create a VPC endpoint for Amazon S3. Run a one-time copy job using the DistCp tool to copy existing files from Hadoop to a target S3 bucket over the VPC endpoint Schedule a nightly DistCp job on the Hadoop cluster to copy the incremental files produced by the custom tool to the target S3 bucket

B. Create a VPC endpoint for Amazon S3. Run a one-time copy job using the DistCp tool to copy existing files from Hadoop to a target S3 bucket over the VPC endpoint. Schedule a nightly job on the trading system servers that produces raw data to copy the incremental raw files to the target S3 bucket. Run the data transformation tool on a transient Amazon EMR cluster to output files to Amazon S3.

C. Create a VPC endpoint for Amazon S3. Run a one-time copy job using the DistCp tool to copy existing files from Hadoop to a target S3 bucket over the VPC endpoint. Set up an Amazon Kinesis data stream to ingest raw data from the trading system in real time. Use Amazon Kinesis Data Analytics to transform the raw data and output files to Amazon S3.

D. Complete a one-time transfer of the data using AWS Snowball Edge devices transferring to a target S3 bucket. Schedule a nightly job on the trading system servers that produces raw data to copy the incremental raw files to the target S3 bucket Run the data transformation tool on a transient Amazon EMR cluster to output files to Amazon S3.

Correct Answer: A

QUESTION 2

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. COPY the file into Amazon Redshift and query the file with Athena from Amazon S3.

B. Use Amazon EMR to convert each .csv file to Apache Avro. COPY the files into Amazon Redshift and query the file with Athena from Amazon S3.

C. Use AWS Glue to convert the files from .csv to a single large Apache ORC file. COPY the file into Amazon Redshift and query the file with Athena from Amazon S3.

D. Use AWS Glue to convert the files from .csv to Apache Parquet to create 20 Parquet files. COPY the files into Amazon Redshift and query the files with Athena from Amazon S3.

Correct Answer: D

QUESTION 3

A technology company is creating a dashboard that will visualize and analyze time-sensitive data. The data will come in through Amazon Kinesis Data Firehose with the butter interval set to 60 seconds. The dashboard must support near-realtime data.

Which visualization solution will meet these requirements?

A. Select Amazon OpenSearch Service (Amazon Elasticsearch Service) as the endpoint for Kinesis Data Firehose. Set up an OpenSearch Dashboards (Kibana) using the data in Amazon OpenSearch Service (Amazon ES) with the desired analyses and visualizations.

B. Select Amazon S3 as the endpoint for Kinesis Data Firehose. Read data into an Amazon SageMaker Jupyter notebook and carry out the desired analyses and visualizations.

C. Select Amazon Redshift as the endpoint for Kinesis Data Firehose. Connect Amazon QuickSight with SPICE to Amazon Redshift to create the desired analyses and visualizations.

D. Select Amazon S3 as the endpoint for Kinesis Data Firehose. Use AWS Glue to catalog the data and Amazon Athena to query it. Connect Amazon QuickSight with SPICE to Athena to create the desired analyses and visualizations.

Correct Answer: A

Reference: https://aws.amazon.com/blogs/big-data/ingest-streaming-data-into-amazon-elasticsearch-service-within-the-privacy-of-your-vpc-with-amazon-kinesis-data-firehose/

QUESTION 4

An analytics team uses Amazon OpenSearch Service for an analytics API to be used by data analysts. The OpenSearch Service cluster is configured with three master nodes. The analytics team uses Amazon Managed Streaming for Apache Kafka (Amazon MSK) and a customized data pipeline to ingest and store 2 months of data in an OpenSearch Service cluster. The cluster stopped responding, which is regularly causing timeout requests. The analytics team discovers the cluster is handling too many bulk indexing requests.

Which actions would improve the performance of the OpenSearch Service cluster? (Choose two.)

A. Reduce the number of API bulk requests on the OpenSearch Service cluster and reduce the size of each bulk request.

B. Scale out the OpenSearch Service cluster by increasing the number of nodes.

C. Reduce the number of API bulk requests on the OpenSearch Service cluster, but increase the size of each bulk request

D. Increase the number of master nodes for the OpenSearch Service cluster

E. Scale down the pipeline component that is used to ingest the data into the OpenSearch Service cluster.

Correct Answer: AB

QUESTION 5

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.

Correct Answer: B

The COPY command loads the data in parallel from multiple files, dividing the workload among the nodes in your cluster. When you load all the data from a single large file, Amazon Redshift is forced to perform a serialized load, which is much slower. Split your load data files so that the files are about equal size, between 1 MB and 1 GB after compression. For optimum parallelism, the ideal size is between 1 MB and 125 MB after compression. The number of files should be a multiple of the number of slices in your cluster

[DAS-C01 PDF Dumps](#) [DAS-C01 Practice Test](#) [DAS-C01 Braindumps](#)