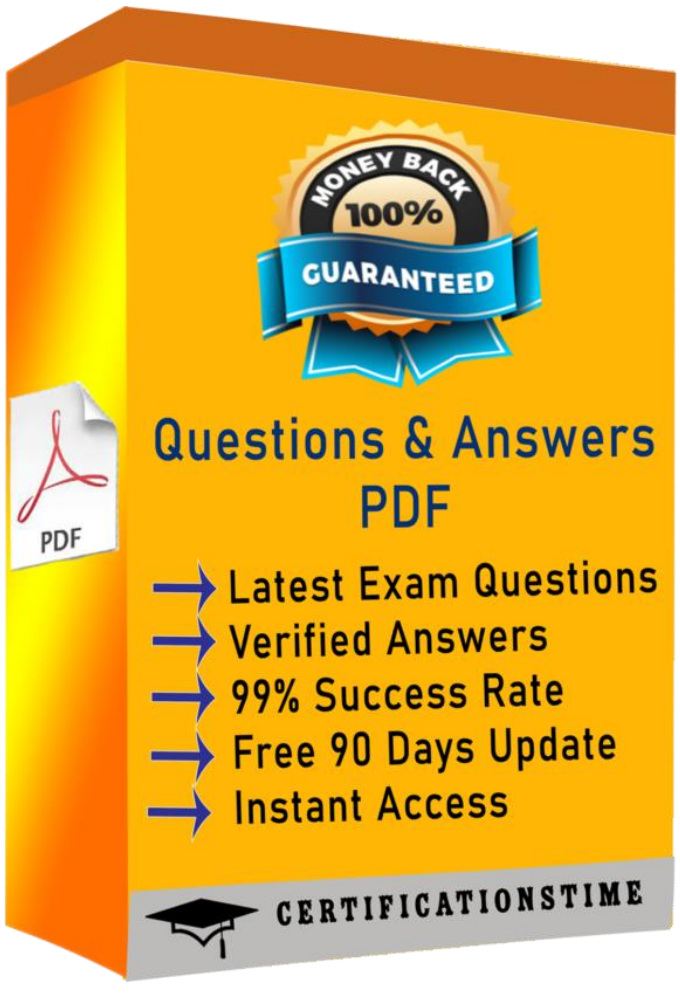
Amazon

Exam Questions DAS-C01

AWS Certified Data Analytics – Specialty

https://certificationstime.com/



QUESTION 1

A financial services company needs to aggregate daily stock trade data from the exchanges into a data store.

The company requires that data be streamed directly into the data store, but also occasionally allows data to be

modified using SQL. The solution should integrate complex, analytic queries running with minimal latency.

The solution must provide a business intelligence dashboard that enables viewing of the top contributors

To anomalies in stock prices.

Which solution meets the company's requirements?

A. Use Amazon Kinesis Data Firehose to stream data to Amazon S3. Use Amazon Athena as a data source for Amazon QuickSight to create a business intelligence dashboard.

B. Use Amazon Kinesis Data Streams to stream data to Amazon Redshift. Use Amazon Redshift as a data source for Amazon QuickSight to create a business intelligence dashboard.

C. Use Amazon Kinesis Data Firehose to stream data to Amazon Redshift. Use Amazon Redshift as a data source for Amazon QuickSight to create a business intelligence dashboard.

D. Use Amazon Kinesis Data Streams to stream data to Amazon S3. Use Amazon Athena as a data source for Amazon QuickSight to create a business intelligence dashboard.

Correct Answer: D

QUESTION 2

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.

Correct Answer: B,C,D

QUESTION 3

A data engineering team within a shared workspace company wants to build a centralized logging system for all weblogs generated by the space reservation system. The company has a fleet of Amazon EC2 instances that process requests for shared space reservations on its website. The data engineering team wants to ingest all weblogs into a service that will provide a near-real-time search engine. The team does not want to manage the maintenance and operation of the logging system.

Which solution allows the data engineering team to efficiently set up the web logging system within AWS?

A. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the

Amazon Kinesis data stream to CloudWatch. Choose Amazon Elasticsearch Service as the end destination of the weblogs.

B. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the

Amazon Kinesis Data Firehose delivery stream to CloudWatch. Choose Amazon Elasticsearch

Service as the end destination of the weblogs.

C. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the

Amazon Kinesis data stream to CloudWatch. Configure Splunk as the end destination of the weblogs.

D. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the

Amazon Kinesis Firehose delivery stream to CloudWatch. Configure Amazon DynamoDB as the end destination of the weblogs.

Correct Answer: A

QUESTION 4

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 nodes. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metric.

Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch

CapacityRemainingGB metric.

B. Create instance fleet configurations for core and task nodes. Create an automatic scaling policy to scale out the instance groups based on the Amazon

CloudWatch YARNMemoryAvailablePercentage metric. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch

YARNMemoryAvailablePercentage metric.

C. Create instance group configurations for core and task nodes. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metric.

Create an automatic scaling policy to scale in the instance groups based on the CloudWatch

CapacityRemainingGB metric.

D. Create instance group configurations for core and task nodes. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch

YARNMemoryAvailablePercentage metric. Create an automatic scaling policy to scale in the instance groups based on the CloudWatch YARNMemoryAvailablePercentage metric.

Correct Answer: C

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.

Correct Answer: C,D

Explanation/Reference:

Reference: https://docs.aws.amazon.com/firehose/latest/dev/record-format-conversion.html