Free Questions for DAS-C01 by certsdeals

Shared by Jackson on 24-05-2024

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

A bank is building an Amazon S3 data lake. The bank wants a single data repository for customer data needs, such as personalized recommendations. The bank needs to use Amazon Kinesis Data Firehose to ingest customers' personal information, bank accounts, and transactions in near real time from a transactional relational database.

All personally identifiable information (PII) that is stored in the S3 bucket must be masked. The bank has enabled versioning for the S3 bucket.

Which solution will meet these requirements?

Options:

A- Invoke an AWS Lambda function from Kinesis Data Firehose to mask the PII before Kinesis Data Firehose delivers the data to the S3 bucket.

B- Use Amazon Macie to scan the S3 bucket. Configure Macie to discover PII. Invoke an AWS Lambda function from S3 events to mask the PII.

C- Configure server-side encryption (SSE) for the S3 bucket. Invoke an AWS Lambda function from S3 events to mask the PII.

D- Create an AWS Lambda function to read the objects, mask the PII, and store the objects back with same key. Invoke the Lambda function from S3 events.

Answer:

A

Question 2

Question Type: MultipleChoice

A gaming company is building a serverless data lake. The company is ingesting streaming data into Amazon Kinesis Data Streams and is writing the data to Amazon S3 through Amazon Kinesis Data Firehose. The company is using 10 MB as the S3 buffer size and is using 90 seconds as the buffer interval. The company runs an AWS Glue ET L job to merge and transform the data to a different format before writing the data back to Amazon S3.

Recently, the company has experienced substantial growth in its data volume. The AWS Glue ETL jobs are frequently showing an OutOfMemoryError error.

Which solutions will resolve this issue without incurring additional costs? (Select TWO.)

Options:

A- Place the small files into one S3 folder. Define one single table for the small S3 files in AWS Glue Data Catalog. Rerun the AWS Glue ET L jobs against this AWS Glue table.

B- Create an AWS Lambda function to merge small S3 files and invoke them periodically. Run the AWS Glue ETL jobs after successful

completion of the Lambda function.

C- Run the S3DistCp utility in Amazon EMR to merge a large number of small S3 files before running the AWS Glue ETL jobs.

D- Use the groupFiles setting in the AWS Glue ET L job to merge small S3 files and rerun AWS Glue E TL jobs.

E- Update the Kinesis Data Firehose S3 buffer size to 128 MB. Update the buffer interval to 900 seconds.

Answer:

A, D

Explanation:

The groupFiles setting is a feature of AWS Glue that enables an ETL job to group files when they are read from an Amazon S3 data store.This can reduce the number of ETL tasks and in-memory partitions, and improve the performance and memory efficiency of the job1. By using the groupFiles setting in the AWS Glue ETL job, the gaming company can merge small S3 files and avoid the OutOfMemoryError error.

The Kinesis Data Firehose S3 buffer size and buffer interval are parameters that determine how much data is buffered before delivering it to Amazon S3.Increasing the buffer size and buffer interval can result in larger files being delivered to Amazon S3, which can reduce the number of small files and improve the performance of downstream processing2. By updating the Kinesis Data Firehose S3 buffer size to 128 MB and buffer interval to 900 seconds, the gaming company can create fewer, larger S3 files and avoid the OutOfMemoryError error.

Question 3

A company hosts its analytics solution on premises. The analytics solution includes a server that collects log files. The analytics solution uses an Apache Hadoop cluster to analyze the log files hourly and to produce output files. All the files are archived to another server for a specified duration.

The company is expanding globally and plans to move the analytics solution to multiple AWS Regions in the AWS Cloud. The company must adhere to the data archival and retention requirements of each country where the data is stored.

Which solution will meet these requirements?

Options:

A- Create an Amazon S3 bucket in one Region to collect the log files. Use S3 event notifications to invoke an AWS Glue job for log analysis. Store the output files in the target S3 bucket. Use S3 Lifecycle rules on the target S3 bucket to set an expiration period that meets the retention requirements of the country that contains the Region.

B- Create a Hadoop Distributed File System (HDFS) file system on an Amazon EMR cluster in one Region to collect the log files. Set up a bootstrap action on the EMR cluster to run an Apache Spark job. Store the output files in a target Amazon S3 bucket. Schedule a job on one of the EMR nodes to delete files that no longer need to be retained.

C- Create an Amazon S3 bucket in each Region to collect log files. Create an Amazon EMR cluster. Submit steps on the EMR cluster for analysis. Store the output files in a target S3 bucket in each Region. Use S3 Lifecycle rules on each target S3 bucket to set an expiration period that meets the retention requirements of the country that contains the Region.

D- Create an Amazon Kinesis Data Firehose delivery stream in each Region to collect log data. Specify an Amazon S3 bucket in each Region as the destination. Use S3 Storage Lens for data analysis. Use S3 Lifecycle rules on each destination S3 bucket to set an expiration period that meets the retention requirements of the country that contains the Region.

Answer:

C

Question 4

Question Type: **MultipleChoice**

A large media company is looking for a cost-effective storage and analysis solution for its daily media recordings formatted with embedded metadat

a. Daily data sizes range between 10-12 TB with stream analysis required on timestamps, video resolutions, file sizes, closed captioning, audio languages, and more. Based on the analysis,

processing the datasets is estimated to take between 30-180 minutes depending on the underlying framework selection. The analysis will be done by using business intelligence (BI) tools that can be connected to data sources with AWS or Java Database Connectivity (JDBC) connectors.

Which solution meets these requirements?

Options:

A- Store the video files in Amazon DynamoDB and use AWS Lambda to extract the metadata from the files and load it to DynamoDB. Use DynamoDB to provide the data to be analyzed by the BI tools.

B- Store the video files in Amazon S3 and use AWS Lambda to extract the metadata from the files and load it to Amazon S3. Use Amazon Athena to provide the data to be analyzed by the BI tools.

C- Store the video files in Amazon DynamoDB and use Amazon EMR to extract the metadata from the files and load it to Apache Hive. Use Apache Hive to provide the data to be analyzed by the BI tools.

D- Store the video files in Amazon S3 and use AWS Glue to extract the metadata from the files and load it to Amazon Redshift. Use Amazon Redshift to provide the data to be analyzed by the BI tools.

Answer:

B

Question 5

Question Type: **MultipleChoice**

A social media company is using business intelligence tools to analyze data for forecasting. The company is using Apache Kafka to ingest dat

a. The company wants to build dynamic dashboards that include machine learning (ML) insights to forecast key business trends.

The dashboards must show recent batched data that is not more than 75 minutes old. Various teams at the company want to view the dashboards by using Amazon QuickSight with ML insights.

Which solution will meet these requirements?

Options:

A- Replace Kafka with Amazon Managed Streaming for Apache Kafka (Amazon MSK). Use AWS Data Exchange to store the data in Amazon S3. Use SPICE in QuickSight Enterprise edition to refresh the data from Amazon S3 each hour. Use QuickSight to create a dynamic dashboard that includes forecasting and ML insights.

B- Replace Kafka with an Amazon Kinesis data stream. Use AWS Data Exchange to store the data in Amazon S3. Use SPICE in QuickSight Standard edition to refresh the data from Amazon S3 each hour. Use QuickSight to create a dynamic dashboard that includes forecasting and ML insights.

C- Configure the Kafka-Kinesis-Connector to publish the data to an Amazon Kinesis Data Firehose delivery stream. Configure the delivery stream to store the data in Amazon S3 with a max buffer size of 60 seconds. Use SPICE in QuickSight Enterprise edition to refresh the data from Amazon S3 each hour. Use QuickSight to create a dynamic dashboard that includes forecasting and ML insights.

D- Configure the Kafka-Kinesis-Connector to publish the data to an Amazon Kinesis Data Firehose delivery stream. Configure the delivery stream to store the data in Amazon S3 with a max buffer size of 60 seconds. Refresh the data in QuickSight Standard edition SPICE from Amazon S3 by using a scheduled AWS Lambda function. Configure the Lambda function to run every 75 minutes and to invoke the QuickSight API to create a dynamic dashboard that includes forecasting and ML insights.

Answer:

C

Question 6

A company has a process that writes two datasets in CSV format to an Amazon S3 bucket every 6 hours. The company needs to join the datasets, convert the data to Apache Parquet, and store the data within another bucket for users to query using Amazon Athen

a. The data also needs to be loaded to Amazon Redshift for advanced analytics. The company needs a solution that is resilient to the failure of any individual job component and can be restarted in case of an error.

Which solution meets these requirements with the LEAST amount of operational overhead?

Options:

A- Use AWS Step Functions to orchestrate an Amazon EMR cluster running Apache Spark. Use PySpark to generate data frames of the datasets in Amazon S3, transform the data, join the data, write the data back to Amazon S3, and load the data to Amazon Redshift.

B- Create an AWS Glue job using Python Shell that generates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift. Use an AWS Glue workflow to orchestrate the AWS Glue job at the desired frequency.

C- Use AWS Step Functions to orchestrate the AWS Glue job. Create an AWS Glue job using Python Shell that creates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon

Redshift.

D- Create an AWS Glue job using PySpark that creates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift. Use an AWS Glue workflow to orchestrate the AWS Glue job.

Answer:

D

Explanation:

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy to prepare and load data for analytics1.It can process datasets from various sources and formats, such as CSV and Parquet, and write them to different destinations, such as Amazon S3 and Amazon Redshift2.

AWS Glue provides two types of jobs: Spark and Python Shell.Spark jobs run on Apache Spark, a distributed processing framework that supports a wide range of data processing tasks3.Python Shell jobs run Python scripts on a managed serverless infrastructure4. Spark jobs are more suitable for complex data transformations and joins than Python Shell jobs.

AWS Glue provides dynamic frames, which are an extension of Apache Spark data frames. Dynamic frames handle schema variations and errors in the data more easily than data frames. They also provide a set of transformations that can be applied to the data, such as join, filter, map, etc.

AWS Glue provides workflows, which are directed acyclic graphs (DAGs) that orchestrate multiple ETL jobs and crawlers. Workflows can handle dependencies, retries, error handling, and concurrency for ETL jobs and crawlers. They can also be triggered by schedules

or events.

By creating an AWS Glue job using PySpark that creates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift, the company can perform the required ETL tasks with a single job. By using an AWS Glue workflow to orchestrate the AWS Glue job, the company can schedule and monitor the job execution with minimal operational overhead.