

Cloud Guru - 2 - Streaming Data Collection Quiz

Let's test your knowledge!

AWS Certified Machine Learning - Specialty 2020 - Streaming Data Collection Quiz

ABOUT THIS QUIZ

NO. OF QUESTIONS
16 Questions

SKILL LEVEL
Intermediate

ATTEMPT HISTORY
No attempts yet!

2nd attempt at Quiz

QUESTION 1

You have been tasked with capturing data from an online gaming platform to run analytics on and process through a machine learning pipeline. The data that you are ingesting is players controller inputs every 1 second (up to 10 players in a game) that is in JSON format. The data needs to be ingested through Kinesis Data Streams and the JSON data blob is 100 KB in size. What is the minimum number of shards you can use to successfully ingest this data?

- ☐ 10 shards
- ☐ Greater than 500 shards, so you'll need to request more shards from AWS
- ☒ 1 shard
- ☐ 100 shards

Good work!

In this scenario, there will be a maximum of 10 records per second with a max payload size of 1000 KB (10 records x 100 KB = 1000KB) written to the shard. A single shard can ingest up to 1 MB of data per second, which is enough to ingest the 1000 KB from the streaming game play. Therefore 1 shard is enough to handle the streaming data.

QUESTION 2

Your organization needs to find a way to capture streaming data from certain events customers are performing. These events are a crucial part of the organization's business development and cannot afford to be lost. You've already set up a Kinesis Data Stream and a consumer EC2 instance to process and deliver the data into S3. You've noticed that the last few days of events are not showing up in S3 and your EC2 instance has been shutdown. What combination of steps can you take to ensure this does not happen again?

☒ Set up CloudWatch monitoring for your EC2 instance as well as AutoScaling if your consumer EC2 instance is shutdown. Next, ensure that the maximum amount of hours are selected (168 hours) for data retention when creating your Kinesis Data Stream. Finally, write logic on the consumer EC2 instance that handles unprocessed data in the Kinesis Data Stream and failed writes to S3.

☐ Set up CloudWatch monitoring for your EC2 instance as well as AutoScaling if your consumer EC2 instance is shutdown. Next, set up a Lambda function to poll the Kinesis Data Stream for failed delivered records and then send those requests back into the consumer EC2 instance.

☐ Set up CloudWatch monitoring for your EC2 instance as well as AutoScaling if your consumer EC2 instance is shutdown. Next, send the data to Kinesis Data Firehose before writing the data into S3. Since Kinesis Data Firehose has retry mechanism built-in, the chances of data being lost is extremely unlikely.

☐ Set up CloudWatch monitoring for your EC2 instance as well as AutoScaling if your consumer EC2 instance is shutdown. Next, set up multiple Kinesis Data Streams to process the data on the EC2 instance.

Good work!

In this setup, the data is being ingested by Kinesis Data Streams and processes and delivered using an EC2 instance. It's best practice to always setup CloudWatch monitoring for your EC2 instance as well as AutoScaling if your consumer EC2 instance is shutdown. Since this data is critical data that we cannot afford to lose, we should set the retention period for the maximum number of hours (168 hours or 7 days). Finally, we need to have reprocessed the failed records that are still in the data stream and that fail to write to S3.

QUESTION 3

Which service in the Kinesis family allows you to easily load streaming data into data stores and analytics tools?

☒ Kinesis Firehose

☐ Kinesis Streams

☐ Kinesis Data Analytics

☐ Kinesis Video Streams

Good work!

Kinesis Firehose is perfect for streaming data into AWS and sending it directly to its final destination - places like S3, Redshift, Elasticsearch, and Splunk Instances.

QUESTION 4

A local university wants to track cars in a parking lot to determine which students are parking in the lot. The university is wanting to ingest videos of the cars parking in near-real time, use machine learning to identify license plates, and store that data in an AWS data store. Which solution meets these requirements with the LEAST amount of development effort?

- ☐ Use Amazon Kinesis Data Streams to ingest the video in near-real time, use the Kinesis Data Streams consumer integrated with Amazon Rekognition Video to process the license plate information, and then store results in DynamoDB.
- ☒ Use Amazon Kinesis Video Streams to ingest the videos in near-real time, use the Kinesis Video Streams integration with Amazon Rekognition Video to identify the license plate information, and then store the results in DynamoDB.
- ☐ Use Amazon Kinesis Data Streams to ingest videos in near-real time, call Amazon Rekognition to identify license plate information, and then store results in DynamoDB.
- ☐ Use Amazon Kinesis Firehose to ingest the video in near-real time and outputs results onto S3. Set up a Lambda function that triggers when a new video is PUT onto S3 to send results to Amazon Rekognition to identify license plate information, and then store results in DynamoDB.

Good work!

Kinesis Video Streams is used to stream videos in near-real time. Amazon Rekognition Video uses Amazon Kinesis Video Streams to receive and process a video stream. After the videos have been processed by Rekognition we can output the results in DynamoDB.

QUESTION 5

Which services in the Kinesis family allows you to analyze streaming data, gain actionable insights, and respond to your business and customer needs in real time?

- ☐ Kinesis Streams
- ☐ Kinesis Firehose
- ☐ Kinesis Video Streams
- ☒ Kinesis Data Analytics

Good work!

Kinesis Data Analytics allows you to run real-time SQL queries on your data to gain insights and respond to events in real time.

QUESTION 6

Which service in the Kinesis family allows you to securely stream video from connected devices to AWS for analytics, machine learning (ML), and other processing?

- ☐ Kinesis Firehose
- ☐ Kinesis Data Analytics
- ☐ Kinesis Streams
- ☒ Kinesis Video Streams

Good work!

Kinesis Video Streams allows you to stream video, images, audio, radar into AWS to further analyze, build custom application around, or store in S3.

QUESTION 7

Which service built by AWS makes it easy to set up a retry mechanism, aggregate records to improve throughput, and automatically submits CloudWatch metrics?

- ☐ Kinesis API (AWS SDK)
- ☒ Kinesis Producer Library (KPL)
- ☐ Kinesis Consumer Library
- ☐ Kinesis Client Library (KCL)

Good work!

Although the Kinesis API built into the AWS SDK can be used for all of this, the Kinesis Producer Library (KPL) makes it easy to integrate all of this into your applications.

QUESTION 8

True or False. If you have mission critical data that must be processed with as minimal delay as possible, you should use the Kinesis API (AWS SDK) over the Kinesis Producer Library.

- ☒ True
- ☐ False

Good work!

The KPL can incur an additional processing delay of up to `RecordMaxBufferedTime` within the library (user-configurable). Larger values of `RecordMaxBufferedTime` results in higher packing efficiencies and better performance. Applications that cannot tolerate this additional delay may need to use the AWS SDK directly.

QUESTION 9

You have been tasked with capturing two different types of streaming events. The first event type includes mission-critical data that needs to immediately be processed before operations can continue. The second event type includes data of less importance, but operations can continue without immediately processing. What is the most appropriate solution to record these different types of events?

- ☐ Capture both events with the PutRecords API call.
- ☐ Capture both event types using the Kinesis Producer Library (KPL).
- ☒ Capture the mission critical events with the PutRecords API call and the second event type with the Kinesis Producer Library (KPL).
- ☐ Capture the mission critical events with the Kinesis Producer Library (KPL) and the second event type with the Putrecords API call.

Good work!

The question is about sending data to Kinesis synchronously vs. asynchronously. PutRecords is a synchronous send function, so it must be used for the first event type (critical events). The Kinesis Producer Library (KPL) implements an asynchronous send function, so it can be used for the second event type. In this scenario, the reason to use the KPL over the PutRecords API call is because: KPL can incur an additional processing delay of up to RecordMaxBufferedTime within the library (user-configurable). Larger values of RecordMaxBufferedTime results in higher packing efficiencies and better performance. Applications that cannot tolerate this additional delay may need to use the AWS SDK directly. For more information about using the AWS SDK with Kinesis Data Streams, see [Developing Producers Using the Amazon Kinesis Data Streams API with the AWS SDK for Java](#). For more information about RecordMaxBufferedTime and other user-configurable properties of the KPL, see [Configuring the Kinesis Producer Library](#).

QUESTION 10

You work for a farming company that has dozens of tractors with build-in IoT devices. These devices stream data into AWS using Kinesis Data Streams. The features associated with the data is tractor Id, latitude, longitude, inside temp, outside temp, and fuel level. As a ML specialist you need to transform the data and store it in a data store. Which combination of services can you use to achieve this?

Choose 3

- ☐ Use Kinesis Data Analytics to run real-time SQL queries to transform the data and immediately write the transformed data into S3.
- ☒ Set up Kinesis Firehose to ingest data from Kinesis Data Streams, then send data to Lambda. Transform the data in Lambda and write the transformed data into S3.
- ☒ Set up Kinesis Data Analytics to ingest the data from Kinesis Data Stream, then run real-time SQL queries on the data to transform it. After the data is transformed, ingest the data with Kinesis Data Firehose and write the data into S3.
- ☒ Immediately send the data to Lambda from Kinesis Data Streams. Transform the data in Lambda and write the transformed data into S3.
- ☐ Use Kinesis Data Streams to immediately write the data into S3. Next, set up a Lambda function that fires any time an object is PUT onto S3. Transform the data from the Lambda function, then write the transformed data into S3.

Good work!

Kinesis Data Streams and Kinesis Data Analytics cannot write data directly to S3. Kinesis Data Firehose is used as the main delivery mechanism for outputting data into S3.

QUESTION 11

Your organization has a standalone Javascript (Node.js) application that streams data into AWS using Kinesis Data Streams. You notice that they are using the Kinesis API (AWS SDK) over the Kinesis Producer Library (KPL). What might be the reasoning behind this?

- ☐ The Kinesis API (AWS SDK) provides greater functionality over the Kinesis Producer Library.
- ☐ The Kinesis API (AWS SDK) runs faster in Javascript applications over the Kinesis Producer Library.
- ☒ The Kinesis Producer Library must be installed as a Java application to use with Kinesis Data Streams.
- ☐ The Kinesis Producer Library cannot be integrated with a Javascript application because of its asynchronous architecture.

Good work!

The KPL must be installed as a Java application before it can be used with your Kinesis Data Streams. There are ways to process KPL serialized data within AWS Lambda, in Java, Node.js, and Python, but not if these answers mentions Lambda.

QUESTION 12

What are your options for storing data into S3?

Choose 4



AWS SDK



The AWS console



AWS CLI



UPLOAD command



PutRecords API call



UNLOAD command

Good work!

You can use the AWS console, the AWS command line interface (cli), or the AWS SDK.

UNLOAD is used by RedShift

QUESTION 13

You are collecting clickstream data from an e-commerce website to make near-real time product suggestions for users actively using the site. Which combination of tools can be used to achieve the quickest recommendations and meets all of the requirements?



Use Kinesis Data Streams to ingest clickstream data, then use Kinesis Data Analytics to run real time SQL queries to gain actionable insights and trigger real-time recommendations with AWS Lambda functions based on conditions.



Use Kinesis Data Firehose to ingest click stream data, then use Kinesis Data Analytics to run real time SQL queries to gain actionable insights and trigger real-time recommendations with AWS Lambda functions based on conditions, then use Lambda to load these results into S3.



Use Kinesis Data Streams to ingest clickstream data, then use Lambda to process that data and write it to S3. Once the data is on S3, use Athena to query based on conditions that data and make real time recommendations to users.



Use the Kinesis Data Analytics to ingest the clickstream data directly and run real time SQL queries to gain actionable insights and trigger real-time recommendations with AWS Lambda functions based on conditions.

Good work!

Kinesis Data Analytics gets its input streaming data from Kinesis Data Streams or Kinesis Data Firehose. You can use Kinesis Data Analytics to run real-time SQL queries on your data. Once certain conditions are met you can trigger Lambda functions to make real time product suggestions to users. It is not important that we store or persist the clickstream data.

QUESTION 14

You are collecting clickstream data from an e-commerce website using Kinesis Data Firehose. You are using the PutRecord API from the AWS SDK to send the data to the stream. What are the required parameters when sending data to Kinesis Data Firehose using the API PutRecord call?

- ☒ DeliveryStreamName and Record (containing the data)
- ☐ DataStreamName, PartitionKey, and Record (containing the data)
- ☐ Data, PartitionKey, StreamName, ShardId
- ☐ Data, PartitionKey, StreamName

Good work!

Kinesis Data Firehose is used as a delivery stream. We do not have to worry about shards, partition keys, etc. All we need is the Firehose DeliveryStreamName and the Record object (which contains the data).

QUESTION 15

Which service in the Kinesis family allows you to build custom applications that process or analyze streaming data for specialized needs?

- ☐ Kinesis Firehose
- ☒ Kinesis Streams
- ☐ Kinesis Video Streams
- ☐ Kinesis Data Analytics

Good work!

Kinesis Streams allows you to stream data into AWS and build custom applications around that streaming data.

QUESTION 16

You are a ML specialist needing to collect data from Twitter tweets. Your goal is to collect tweets that include only the name of your company and the tweet body, and store it off into a data store in AWS. What set of tools can you use to stream, transform, and load the data into AWS with the LEAST amount of effort?



Setup a Kinesis Data Firehose for data ingestion and immediately write that data to S3. Next, setup a Lambda function to trigger when data lands in S3 to transform it and finally write it to DynamoDB.



Setup A Kinesis Data Stream for data ingestion, setup EC2 instances as data consumers to poll and transform the data from the stream. Once the data is transformed, make an API call to write the data to DynamoDB.



Setup Kinesis Data Streams for data ingestion. Next, setup Kinesis Data Firehose to load that data into RedShift. Next, setup a Lambda function to query data using RedShift spectrum and store the results onto DynamoDB.



Create a Kinesis Data Stream to ingest the data. Next, setup a Kinesis Data Firehose and use Lambda to transform the data from the Kinesis Data Stream, then use Lambda to write the data to DynamoDB. Finally, use S3 as the data destination for Kinesis Data Firehose.

Sorry!

Correct Answer

All of these could be used to stream, transform, and load the data into an AWS data store. The setup that requires the LEAST amount of effort and moving parts involves setting up a Kinesis Data Firehose to stream the data into S3, have it transformed by Lambda with an S3 trigger, and then written to DynamoDB.



94%

Congratulations!

You passed AWS Certified Machine Learning - Specialty 2020 - Streaming Data
Collection Quiz!