

Special Offer | Flat 15% OFF on All Courses | Use Coupon - WHIZSITE15

[Home](#) > [My Courses](#) > [AWS Certified Machine Learning Specialty](#) > [Data Engineering](#) > **Report**

Search Courses



Data Engineering

Completed on 28-January-2021

**Attempt**
01**Marks Obtained**
1 / 6**Your score**
16.67%**Time Taken**
00 H 00 M 07 S**Result**
Failed

Domains wise Quiz Performance Report

[Join us on Slack community](#)

No	Domain	Total Question	Correct	Incorrect	Unattempted	Marked as Review
1	Data Engineering	6	1	0	5	0
Total	All Domain	6	1	0	5	0

[Review the Answers](#)

Sorting by

All

Question 1

Unattempted

Domain :Data Engineering

Your machine learning team is responsible for processing video clips posted to your company's Twitter social media account to understand the sentiment of the video clips. Your team takes these video clips and labels them with the appropriate sentiment so that your marketing department can use them in their advertising campaigns. You are now expanding into Spanish and Portuguese speaking regions of the world so you now need to translate video clip audio as a part of your sentiment labeling process.

What AWS services and SageMaker built-in algorithms allow your team to label the foreign language video clips in the most efficient manner?

A. Transcribe -> Comprehend -> Translate

- B. Transcribe -> Comprehend -> SageMaker seq2seq
- C. Transcribe -> Translate -> SageMaker Neural Topic Model (NTM)
- D. Transcribe -> Translate -> SageMaker BlazingText
- E. Translate -> Transcribe -> Comprehend

Explanation:

Answer: A

Option A is CORRECT. This is the most efficient option given. It is accomplished completely using AWS managed services. Comprehend can identify sentiment in Spanish and Portuguese language transcriptions; you don't have to translate to English before using Comprehend.

Option B is incorrect. Using the SageMaker seq2seq built-in algorithm overcomplicates the solution. With this option your team would have to build a machine learning model to translate the Spanish or Portuguese language to English when you could just use the Translate service for this step.

Option C is incorrect. Using the SageMaker Neural Topic Model built-in algorithm overcomplicates the solution. With this option your team would have to build a machine learning model to identify the sentiment in the video clips when you could just use the Comprehend service for this step.

Option D is incorrect. Using the SageMaker BalzingText built-in algorithm overcomplicates the solution. With this option your team would have to build a machine learning model to identify the sentiment in the video clips when you could just use the Comprehend service for this step.

Option E is incorrect. The Amazon Translate service requires the input of a text document. In this option you are attempting to translate directly from the audio of your video clip.

Reference:

Please see the **Amazon Comprehend developer guide** titled **Languages Supported in Amazon Comprehend**.

Please refer to the **Amazon Translate developer guide** titled **What Is Amazon Translate?**.

Please review the **Amazon Translate developer guide** titled **How Amazon Translate Works**.


Ask our Experts

Rate this Question?  

[View Queries](#)[open](#) **Question 2****Unattempted****Domain :Data Engineering**

You work for a machine learning team at a global retail auto parts chain. Your team ingests purchasing data from its 100,000 global auto parts stores to S3 using Kinesis Data Firehose. You are now ready to start training an improved machine learning model that will be used to predict purchasing patterns by global region. The training data requires additional simple transformations. Also, you will need to combine some data attributes. Finally, your team expects to train the model on a daily basis.

Based on the large number of stores plus changing data ingestion, which of the following options will require the least amount of administration and development effort?

- A. Have the stores capture their purchasing data locally on Storage Gateway and then load the data into S3. Transform the data using Glue.
- B. Create an EMR cluster with Apache Spark installed to perform the transformation logic. Run the cluster each day on the incremental records in S3, and write the transformed records to S3.
- C. Create a fleet of EC2 instances that run the transformation logic which transforms the incremental data records on S3, and write the transformed records to S3.
- D. Create a Kinesis Data Analytics stream and use it as the destination of the Kinesis Data Firehose stream. Use Kinesis Data Analytics to transform the raw purchasing data attributes into transformed values using SQL and write the transformed data to S3. 

Explanation:**Answer: D**

Option A is incorrect. Having 100,000 stores use Storage Gateway to move the data to S3 would require a very large administrative effort.

Option B is incorrect. Using EMR for this solution would require administrative cost to build and maintain EMR. Also, your development team would have to write the Apache Spark code to perform the transformations.

Option C is incorrect. Using a fleet of EC2 instances would require the administrative cost of creating and maintaining the EC2 instances. Also, your development team would have to write the transformation logic that runs on the EC2 instances.

Option D is CORRECT. Kinesis Data Analytics can receive your data from Kinesis Data Firehose, transform it, and then write it to S3. The code needed to perform the

transformations in Kinesis Data Analytics would be much simpler than the coding suggested in the other options. Your machine learning model can then use the transformed data in S3 for training.

Reference:

Please see the **Amazon Kinesis Data Analytics developer guide** titled **Example: Writing to an Amazon S3 Bucket**.

Please refer to the **Amazon Kinesis Data Firehose developer guide** titled **Using Amazon Kinesis Data Analytics**.

Ask our Experts

Rate this Question? 😊 😞

View Queries

open ▾

Question 3

Unattempted

Domain :Data Engineering

You are a machine learning specialist working for an oil and gas company. Your company's oil and gas drilling sites around the world are equipped with sensors that stream site equipment status and external conditions like weather. You are responsible for building a machine learning model that predicts equipment failures at the sites. The streaming data from the sites needs to be ingested, transformed and stored in Apache Parquet files for exploration and analysis before you use the data in your model.

Which of the following options would ingest, transform, and store your data in the parquet format with the least amount of effort on your part?

- A. Kinesis Data Streams
- B. Kinesis Data Analytics
- C. Kinesis Data Firehose ✓
- D. Managed Streaming for Apache Kafka (MSK)

Explanation:

Answer: C

Option A is incorrect. While you could use Kinesis Data Streams to ingest your sensor data, you would have to write a Kinesis Client Library application or Lambda function to

transform the sensor data to the parquet format. This involves more work than using Kinesis Data Firehose.

Option B is incorrect. You cannot stream data directly into Kinesis Data Analytics. You would have to stream your sensor data into either Kinesis Data Streams or Kinesis Data Firehose first and then send your data downstream to Kinesis Data Analytics. This involves more work than using Kinesis Data Firehose.

Option C is CORRECT. With Kinesis Data Firehose you can stream your sensor data directly to Kinesis Data Firehose, use its built-in parquet transform, then write the parquet files to S3. This approach requires the least amount of work on your part.

Option D is incorrect. While Kafka could be used to stream your sensor data and transform it, this option requires creating an MSK cluster, creating a client machine, creating a topic, and other effort consuming tasks. This involves more work than using Kinesis Data Firehose.

Reference:

Please see the [Amazon Kinesis Data Streams developer guide](#) titled [What Is Amazon Kinesis Data Streams?](#).

Please refer to the [Amazon Kinesis Data Firehose developer guide](#) titled [What Is Amazon Kinesis Data Firehose?](#).

Please refer to the [Amazon Kinesis Data Analytics for SQL Applications developer guide](#) titled [What Is Amazon Kinesis Data Analytics for SQL Applications?](#)

Ask our Experts

Rate this Question? 😊 😞

View Queries

open ▾


Question 4

Unattempted

Domain :Data Engineering

You are a member of a machine learning team at a large online retailer. Your team is responsible for retail competitor analysis. You have a competitive product data streaming source that you need to ingest into your data lake. You need to use that streaming competitor product data to match to the corresponding product data in your catalog of products. Using this matching, your data scientists can produce competitive analysis dashboards in a BI tool. Which of the following options gives you the best data ingestion and most efficient product comparison solution?

A. Kinesis Data Streams -> S3 -> Lake Formation -> QuickSight

- B. Kinesis Data Streams -> Elasticsearch -> Kibana
- C. Kinesis Data Firehose -> S3 -> Lake Formation -> QuickSight 
- D. Kinesis Data Firehose -> Elasticsearch -> Kibana

Explanation:

Answer: C

Option A is incorrect. Kinesis Data Streams cannot write directly to S3. It needs a Kinesis Consumer Library application to receive the data and then write it to S3.

Option B is incorrect. Kinesis Data Streams cannot write directly to Elasticsearch. Also, Elasticsearch has no built-in data matching capability.

Option C is CORRECT. Kinesis Data Firehose can stream directly to S3. Lake Formation has a FindMatches transform which enables you to identify matching records in your dataset, even when the records do not have a common unique identifier and no fields match exactly. In this scenario you will match products in your product catalog with your competitive product sources even though the product entries are structured differently. QuickSight allows your data scientists to produce their analysis dashboards.

Option D is incorrect. Kinesis Data Firehose can write directly to Elasticsearch, but Elasticsearch has no built-in data matching capability.

Reference:

Please see the [AWS Lake Formation developer guide](#) titled [What Is AWS Lake Formation?](#).

Please see the [AWS Glue developer guide](#) titled [Matching Records with AWS Lake Formation FindMatches](#).


[Ask our Experts](#)[Rate this Question?](#)  

[View Queries](#)[open](#) **Question 5****Unattempted****Domain :Data Engineering**

You are a machine learning specialist working for the digital banking division of a global banking firm. Your bank is in the process of introducing a conversational user interface for their

digital banking service. The service will receive streaming audio from the conversational user interface and will need to converse with the user in real-time. Your machine learning team lead has decided to use the Amazon Transcribe service to convert the streaming audio to streaming text.

In order to handle issues in the network connection when users are on mobile phones, how can you leverage the features of Amazon Transcribe to keep your solution as cost effective as possible?

- A. Use the Transcribe JSON streaming client
- B. Use the Transcribe HTTP/2 streaming client 
- C. Use the Transcribe WebSocket protocol
- D. Use the Transcribe HTTP streaming client

Explanation:

Answer: B

Option A is incorrect. The streaming client available with Transcribe is an HTTP/2 streaming client.

Option B is CORRECT. You can use the Transcribe HTTP/2 streaming client to handle retrying the connection when there are intermittent problems on the network.

Option C is incorrect. The Transcribe WebSocket protocol does not provide retry logic to handle retrying the connection when there are intermittent problems on the network. With this option you would have to code the retry logic yourself.

Option D is incorrect. The streaming client available with Transcribe is an HTTP/2 streaming client.

Reference:

Please see the [Amazon Transcribe developer guide](#) titled [Streaming Transcription](#).

Please see the [Amazon Transcribe developer guide](#) titled [HTTP/2 Streaming Retry Client](#),

Please see the [Amazon Transcribe developer guide](#) titled [Using Amazon Transcribe Streaming with WebSockets](#)

Ask our Experts

Rate this Question?



View Queries

open 

Question 6

Correct

Domain :Data Engineering

You are a machine learning specialist working at a language translation software vendor. Your mobile app team would like to add live transcriptions to the company's mobile language conversation application.

Which AWS services should you use to provide the live transcriptions feature to your mobile app?

A. Stream your audio data to Amazon Transcribe Streaming and use the `StartTranscriptionJob` API call to start a bidirectional HTTP stream that streams your audio to Amazon Transcribe. Amazon Transcribe then streams the transcription results to your application and your app code produces the live transcription.

B. Stream your audio data to Amazon Transcribe Streaming and use the `StartStreamTranscription` API call to start a bidirectional HTTP stream that streams your audio to Amazon Transcribe. Amazon Transcribe then streams the transcription results to your application and your app code produces the live transcription.

✓ C. Stream your audio data to Amazon Transcribe Streaming and use the `StartStreamTranscription` API call to start a bidirectional HTTP/2 stream that streams your audio to Amazon Transcribe. Amazon Transcribe then streams the transcription results to your application and your app code produces the live transcription. ✓

D. Stream your audio data to Amazon Connect Streaming and use the `StartStreamTranscription` API call to start a bidirectional HTTP/2 stream that streams your audio to Amazon Connect. Amazon Connect then streams the transcription results to your application and your app code produces the live transcription.

Explanation:

Answer: C

Option A is incorrect. The `StartTranscriptionJob` API is used for Amazon Transcribe in batch translation mode. You are building a real-time transcription service. Also, Amazon Transcribe operates using HTTP/2, not HTTP.

Option B is incorrect. Amazon Transcribe operates using HTTP/2, not HTTP.

Option C is CORRECT. Use the Amazon Transcribe Streaming service and its `StartStreamTranscription` API. It starts a bidirectional HTTP/2 stream that streams your audio to Amazon Transcribe. Transcribe then streams the transcription results to your application and your app code produces the live transcription to be displayed in your app user interface.

Option D is incorrect. Amazon Connect does not have a transcription feature.

Reference:

Please see the **Amazon Transcribe developer guide** titled **Streaming Transcription**.

Please see the **Amazon Transcribe developer guide** titled **StartTranscriptionJob**,

Please see the **Amazon Transcribe developer guide** titled **StartStreamTranscription**,

Please see the **GitHub repository** titled **Amazon Connect Real-time Transcription Lambda**

[Ask our Experts](#)[Rate this Question?](#)  [View Queries](#)[open](#) [Finish Review](#)**Certification**[Cloud Certification](#)[Java Certification](#)[PM Certification](#)[Big Data Certification](#)**Company**[Become Our Instructor](#)[Support](#)[Discussions](#)[Blog](#)[Business](#)**Support**[Contact Us](#)[Help Topics](#)

Join our open **Slack community** and get your queries answered instantly! Our experts are online to answer your questions!

Follow us