

Quiz: Kinesis

Chandra Lingam, Copyright © 2019 Cloud Wave LLC. All Rights Reserved.

1. Several sensors in a building are monitoring Air Quality. These Sensors are periodically sending data to AWS Kinesis Service for analysis. You would like to analyze the daily air quality trend over past one year across all sensors. This analysis would be considered as:
 - A. Batch Processing and this analysis is not suitable for Kinesis
 - B. Stream Processing and can be easily performed in Kinesis
 - C. Stream Processing and Kinesis retention needs to be increased to support this use case
 - D. Batch Processing and Kinesis retention needs to be increased to support this use case
2. You have several sensors in a building that is monitoring Air Quality. Sensors emit air quality metrics every 30 seconds and send it to your locally hosted data collection server which in-turn forwards the streaming data to AWS Kinesis Service for analysis. You would like to analyze the air quality trend in 5 minute windows and alert of any changes above threshold. This analysis would be considered as:
 - A. Batch Processing
 - B. Stream Processing
3. Kinesis Stream has three shards. One of the shards is very heavily loaded and running into throughput issues. What is the best way to address this?
 - A. Add additional Shards to redistribute the workload
 - B. Split the shard that is heavily loaded and evaluate how partition keys are specified
 - C. Increase Shard capacity soft limit by contacting AWS Support
4. A streaming data record stored in kinesis streams can be processed by:
 - A. Many Consumers
 - B. One Consumer
5. You would like your streaming data to be available for 35 days. What are your options?
 - A. Increase Kinesis Streams retention to 35 days
 - B. Contact AWS Support to increase the retention soft limit
 - C. Store the data in other AWS Services like S3 or Redshift or Elasticsearch
6. Your development team has invested in ETL (Extract Transform Load) to process data available in S3. Another team in your company has started a project for collecting streaming data from client devices. This streaming data needs to be made available in Redshift Data warehouse and they have requested your help. What option would you recommend?
 - A. Use Firehose to send to S3 and use your existing ETL process to load to Redshift
 - B. Use Firehose to automatically publish data to Redshift

- C. Use Kinesis Streams and deploy Lambda function to read from streams and publish to Redshift
 - D. Use Kinesis Streams and route data to S3 and use your existing ETL to send data to Redshift
- 7. You are using Kinesis Firehose for your streaming data collection and usage. You are expecting a 10x increase in data collection.
What are your options to increase scalability and throughput?
 - A. Add additional shards
 - B. Use a different convention for assigning partition keys
 - C. Firehose scalability is automatically managed. Optionally contact AWS Support if you need to increase soft limit
- 8. You are using Firehose for sending streaming data to S3. What will happen if your destination S3 system is unavailable for a few hours?
 - A. Data is lost
 - B. Data is kept in firehose for up to 24 hours and retries loading to S3
 - C. Data is kept in firehose for up to 7 days and retries loading to S3
- 9. How can you read streaming data available in Firehose?
 - A. Using Firehose API
 - B. Using Firehose Command Line Interface
 - C. By routing data to other services like S3
 - D. All the above
- 10. To use Firehose, you need to first collect the streaming data in a separate Kinesis Streams instance and then use firehose to route to appropriate destination
 - A. True
 - B. False

Answers:

- 1. A - Deeper analysis over longer duration should be considered as a batch processing use case. These are not stream processing use cases. Kinesis Streams can store data for a maximum of 7 days
- 2. B - This can be considered as stream processing as new data is continuously monitored and trend analyzed over a short time windows
- 3. B – You can split the shard to distribute incoming data across multiple shards. However, you need to also evaluate how partition keys are specified and ensure sufficient number of partition keys are used.
- 4. A – Kinesis Streams can handle scenarios where many consumers need access to same data. Each Shard preserves the ordering of data and consumers can process data in-order.
- 5. C - Kinesis Streams has a maximum retention of 7 days and Kinesis Firehose has a retention of 1 day
- 6. B - This is the most efficient and fully managed option for publishing to Redshift. Firehose uses S3 as an intermediate store.

7. C – Firehose is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration
8. B – Firehose automatically retries for up to 24 hours
9. C – Firehose data can be read only using Kinesis Analytics or by routing to one of the supported destinations like S3, ElasticSearch or Redshift. You can then read from these destination systems.
10. B - You don't need to create a separate kinesis stream. Firehose can store data records directly under what is known as Firehose Destination Stream.