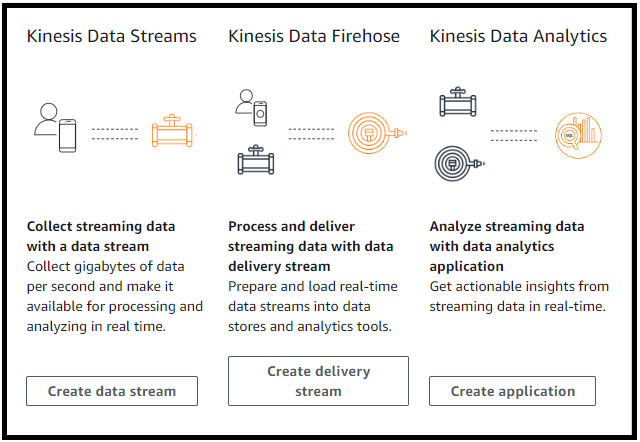
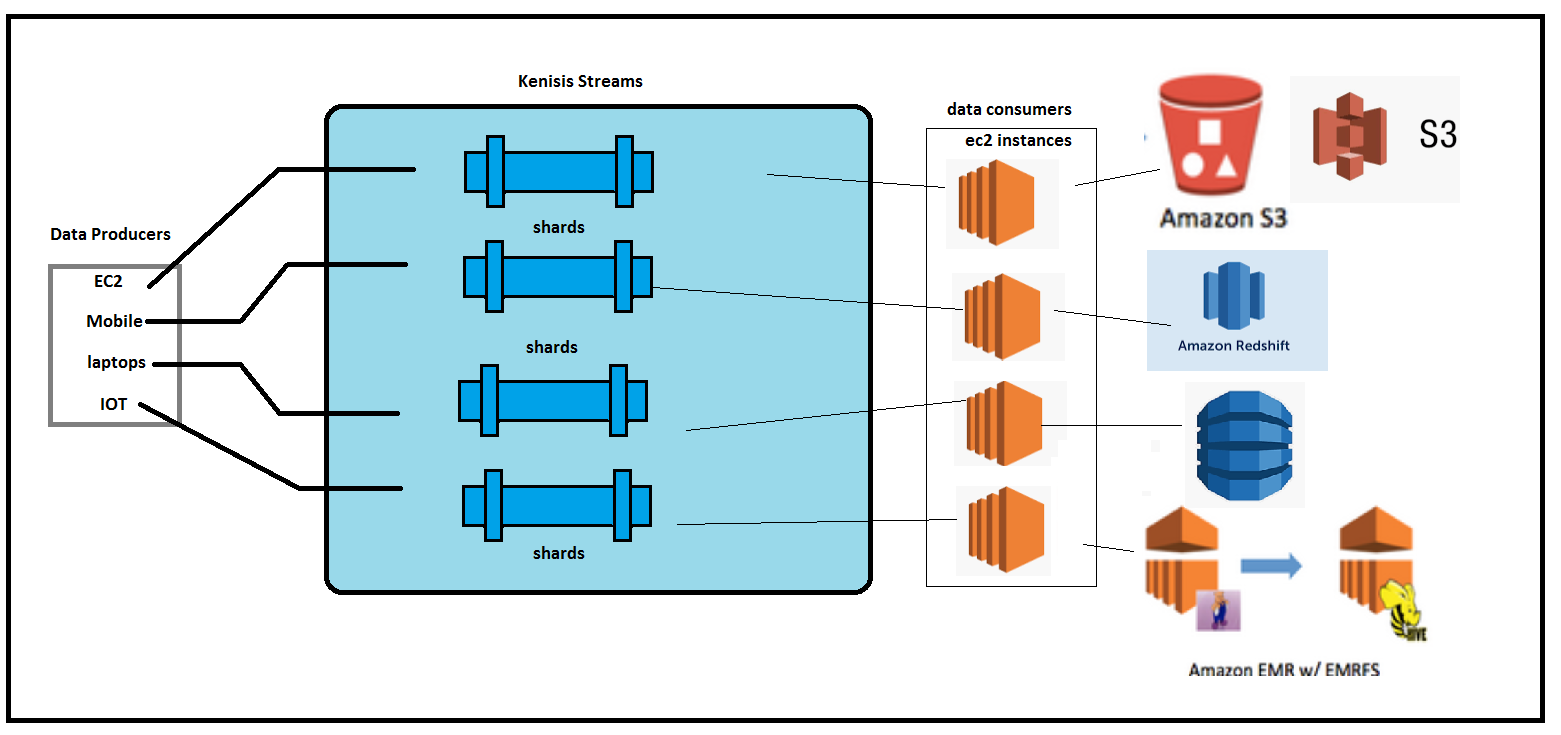
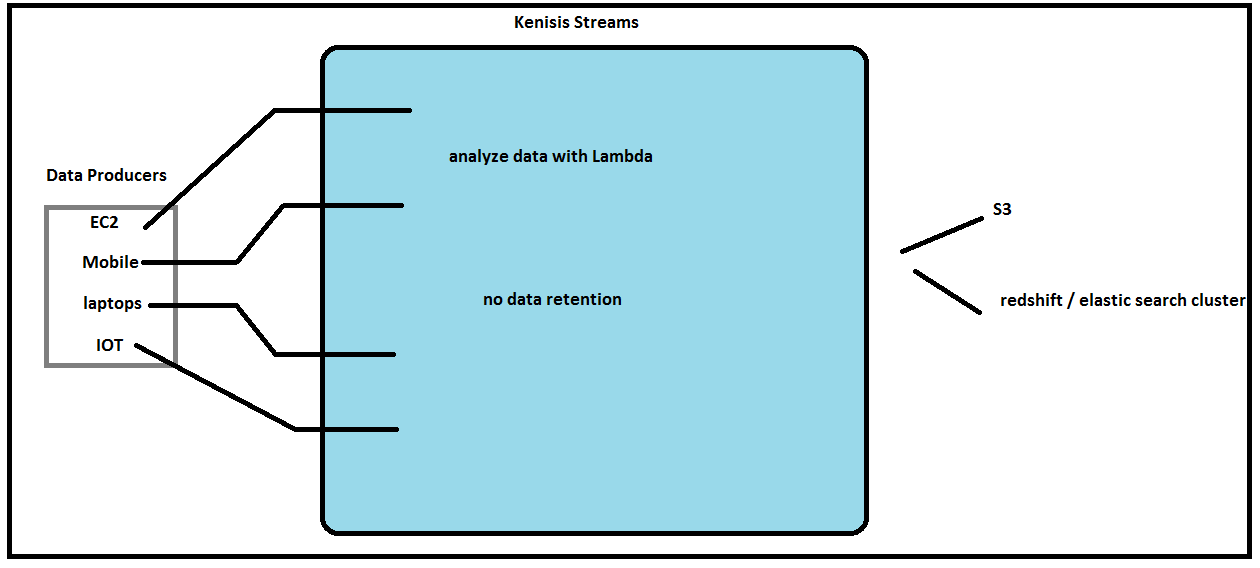
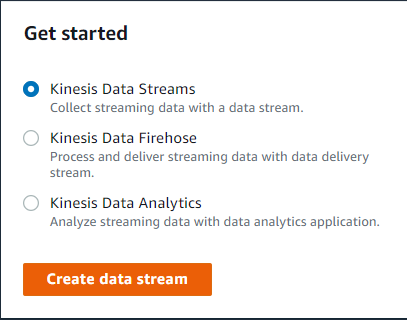
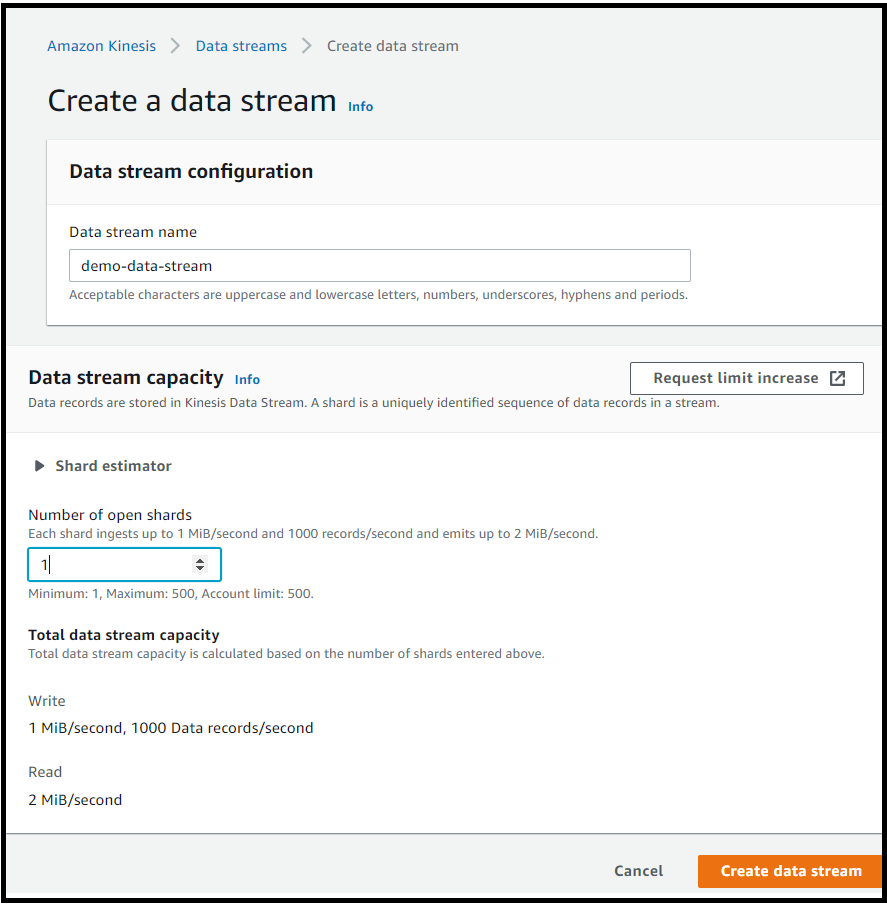
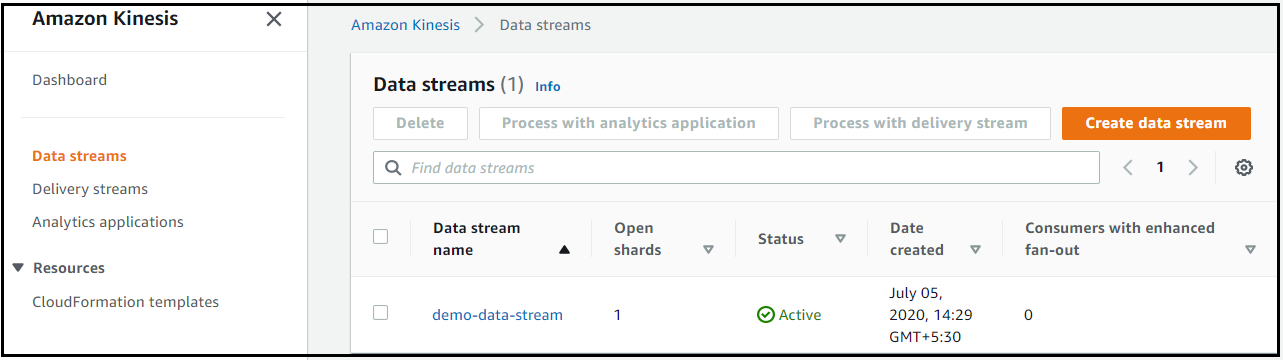
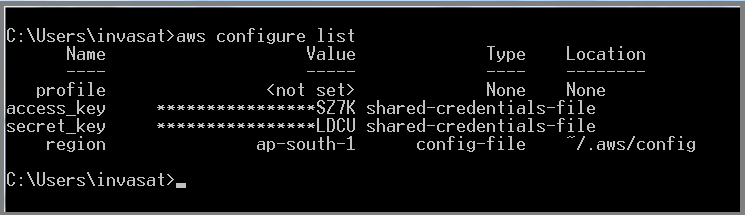
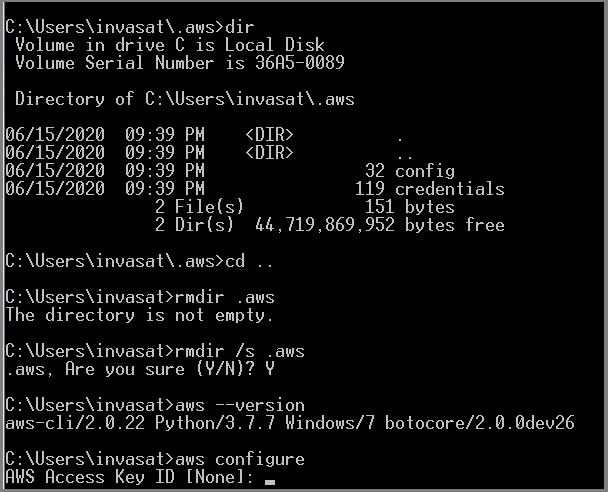
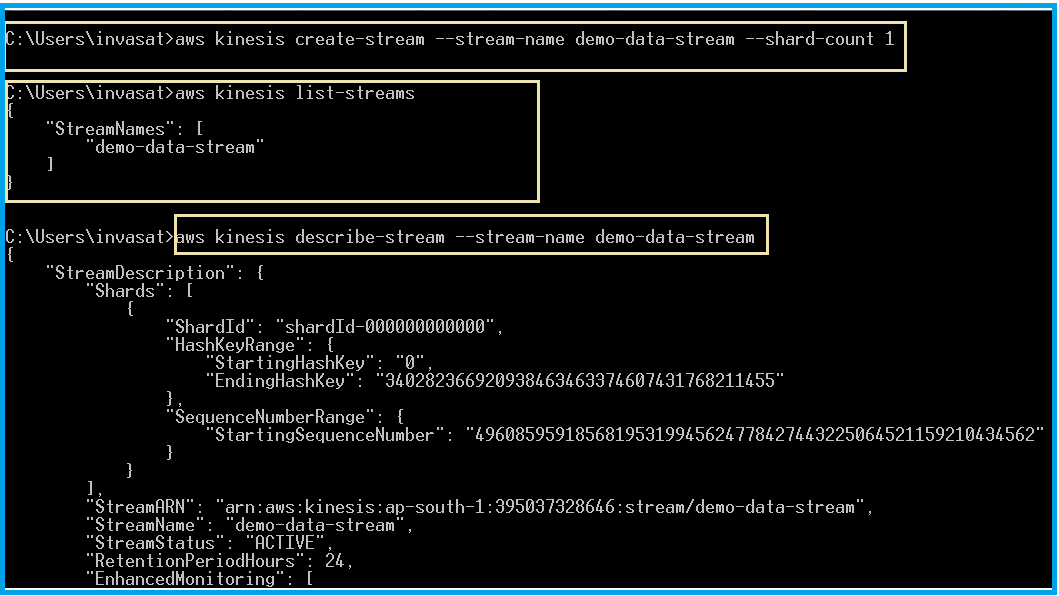
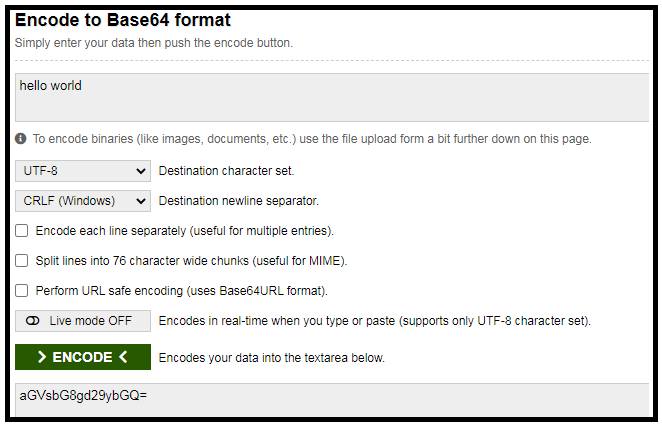
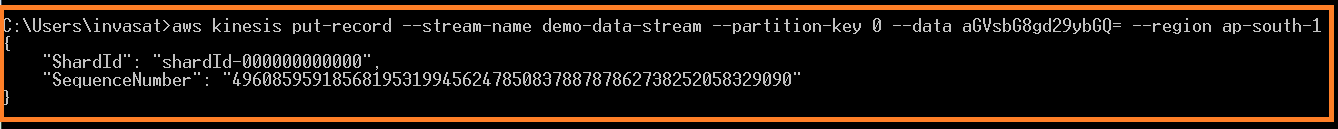
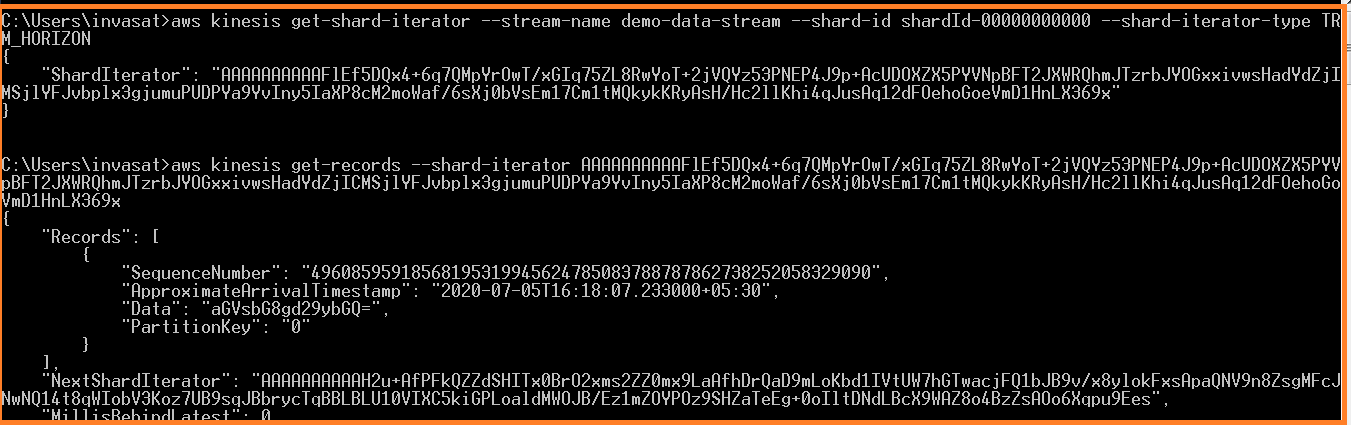
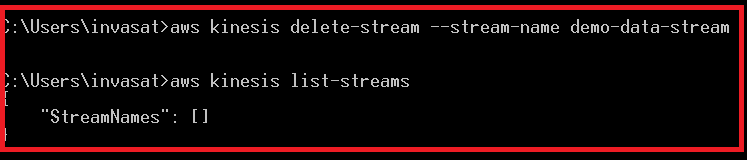
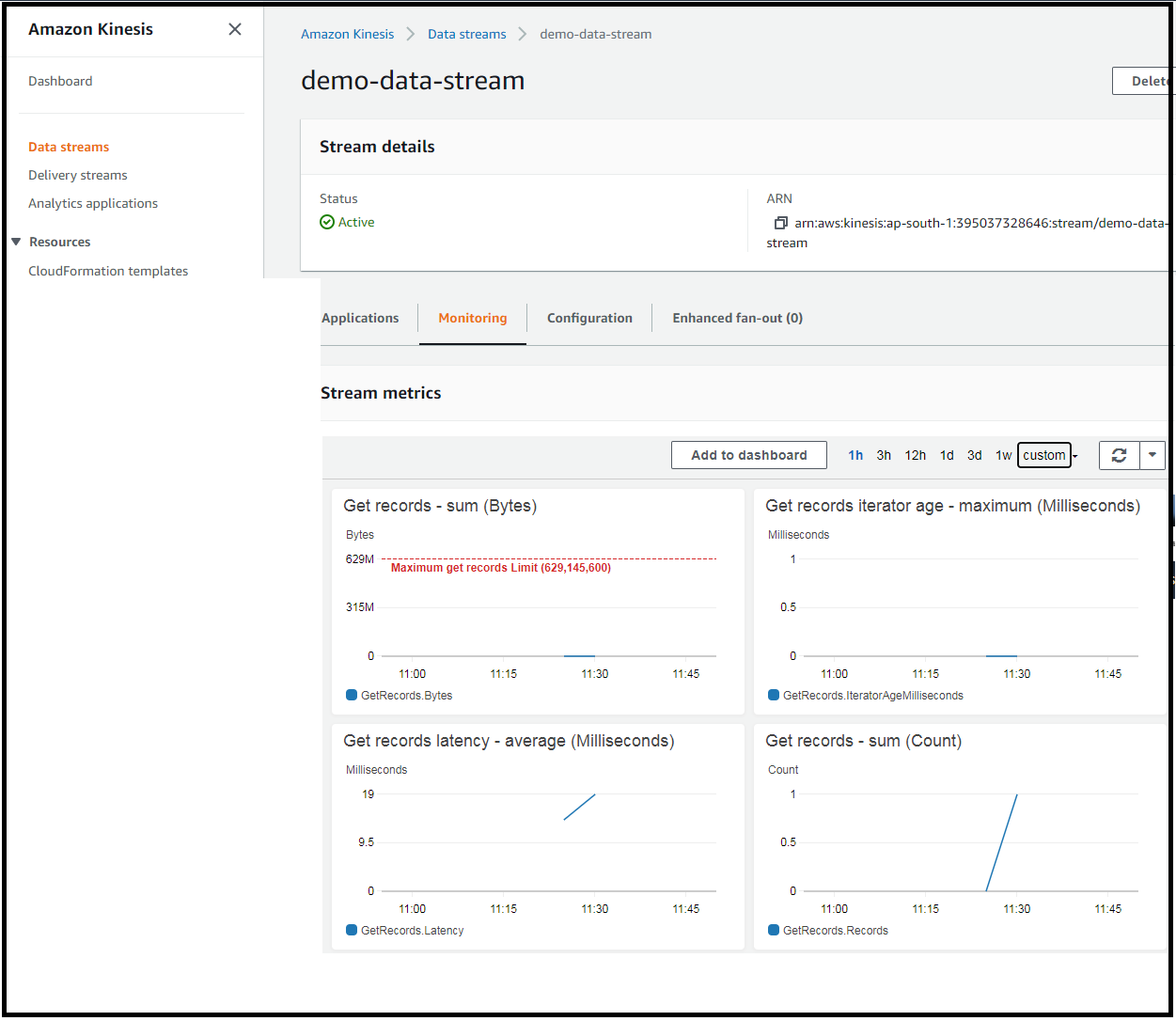
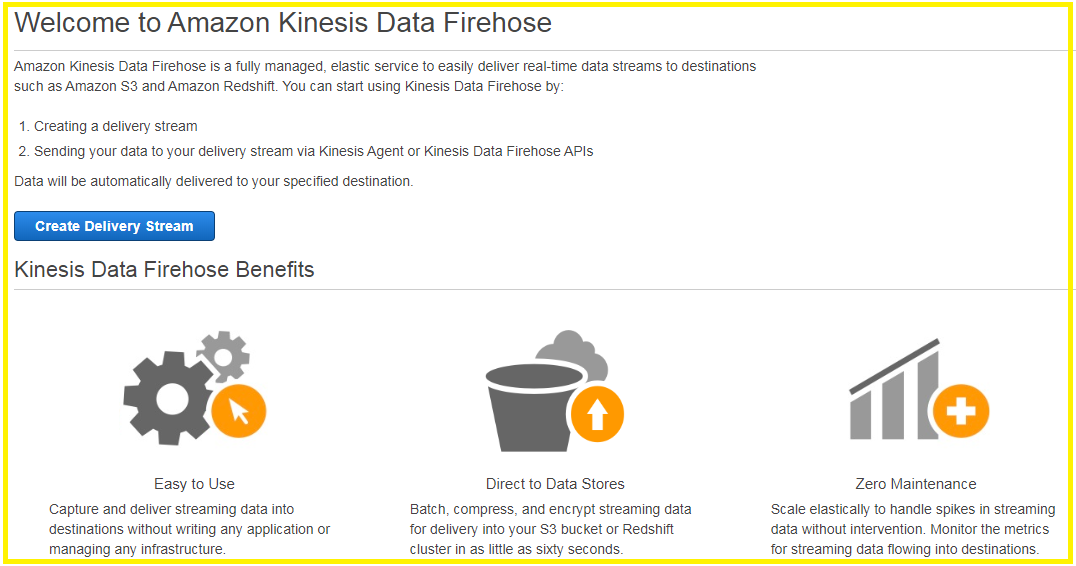
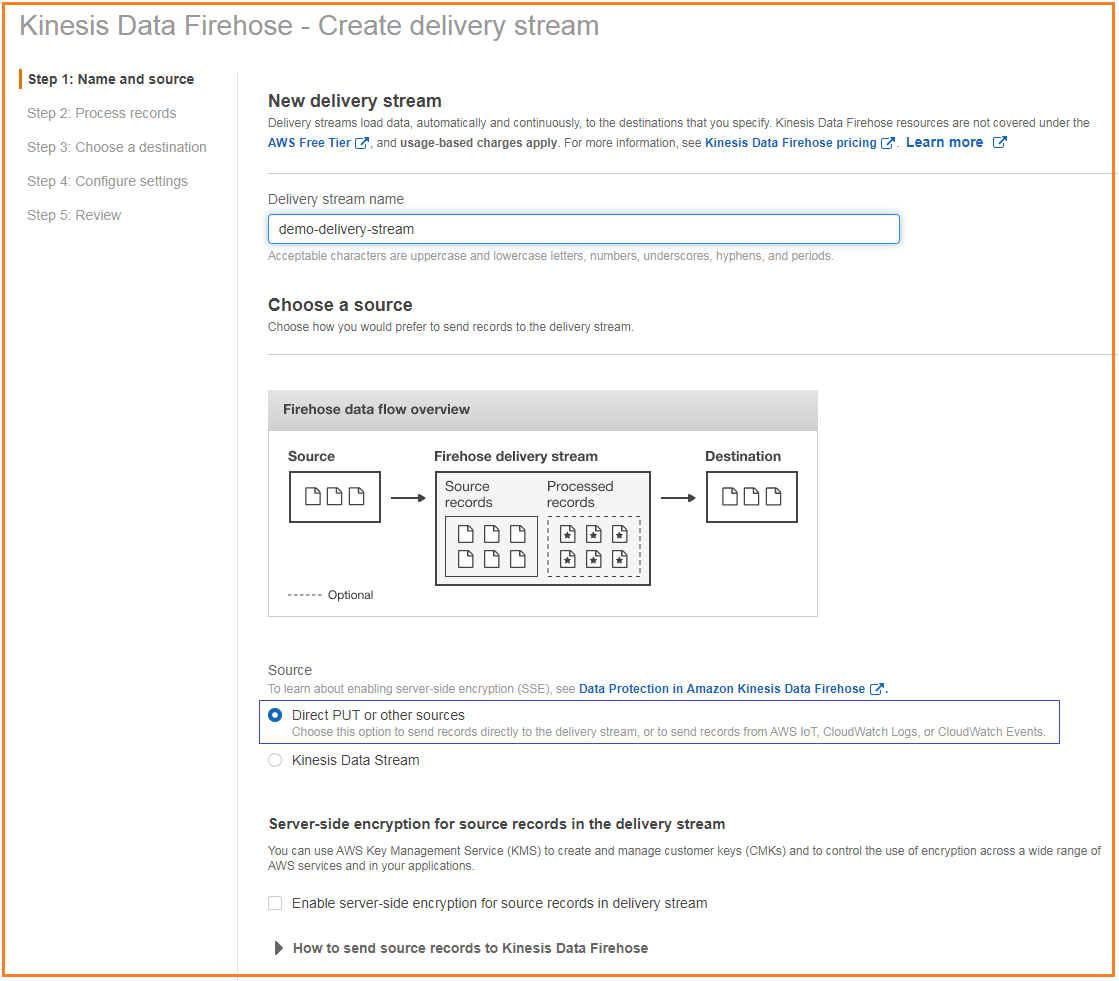
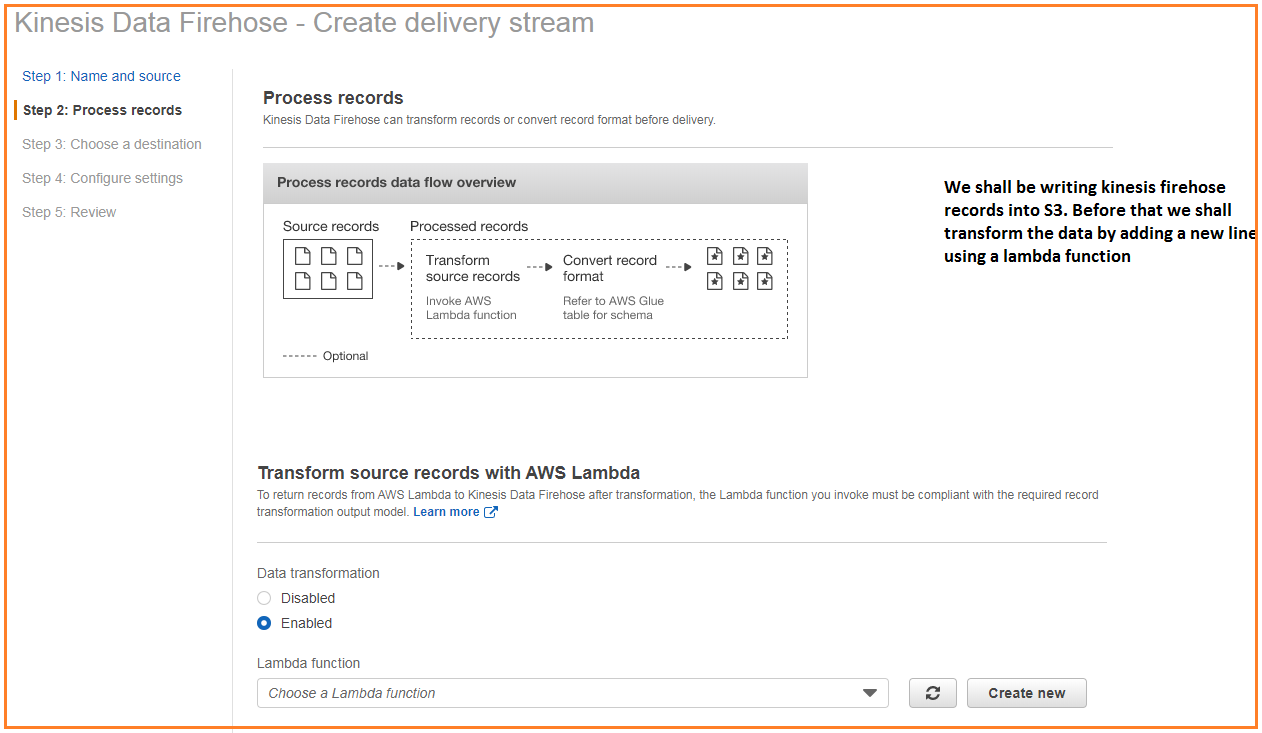
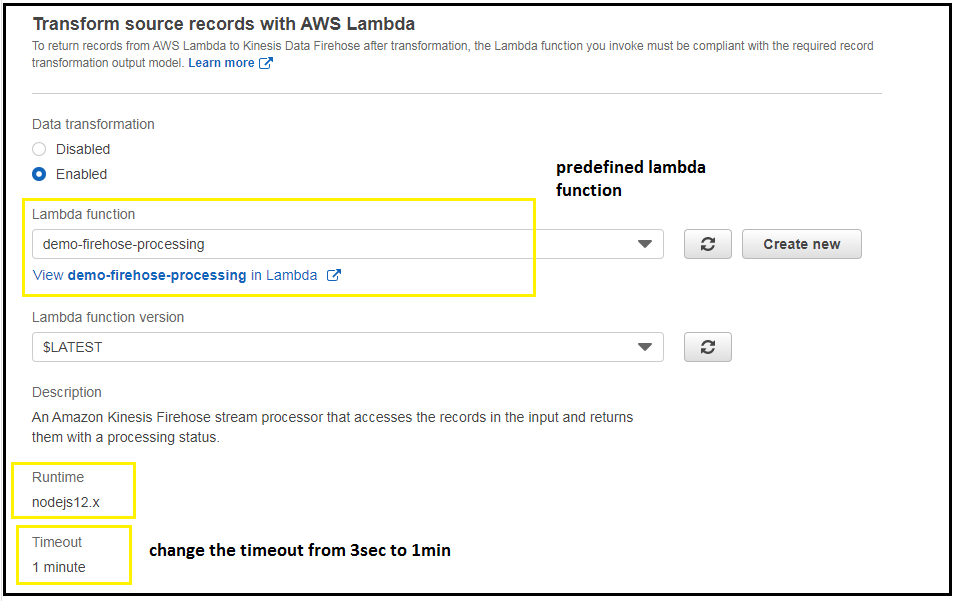
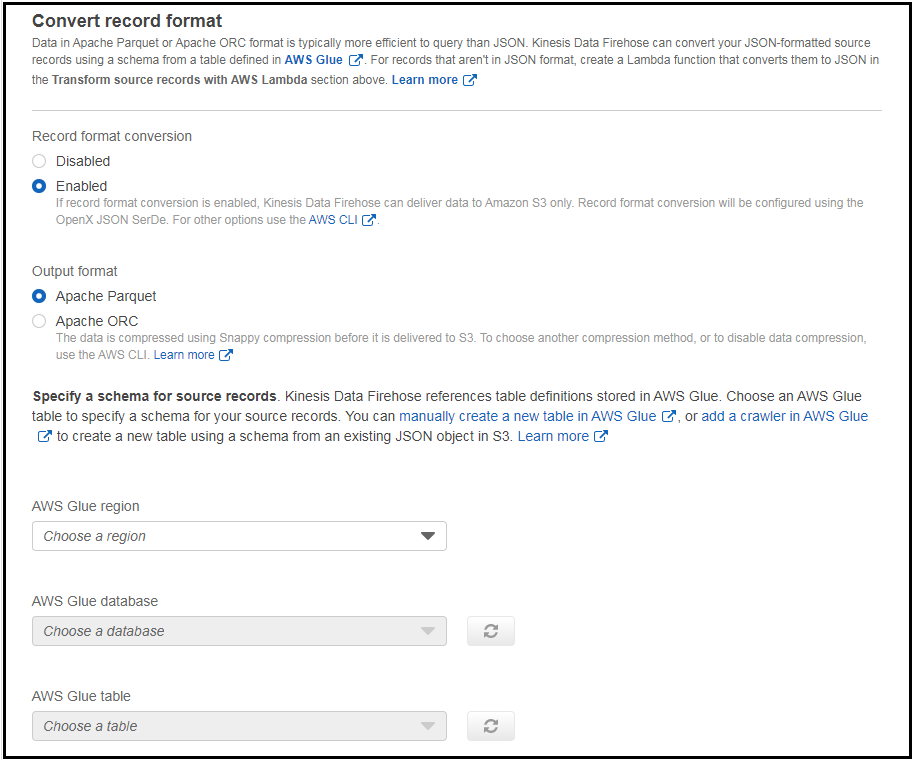
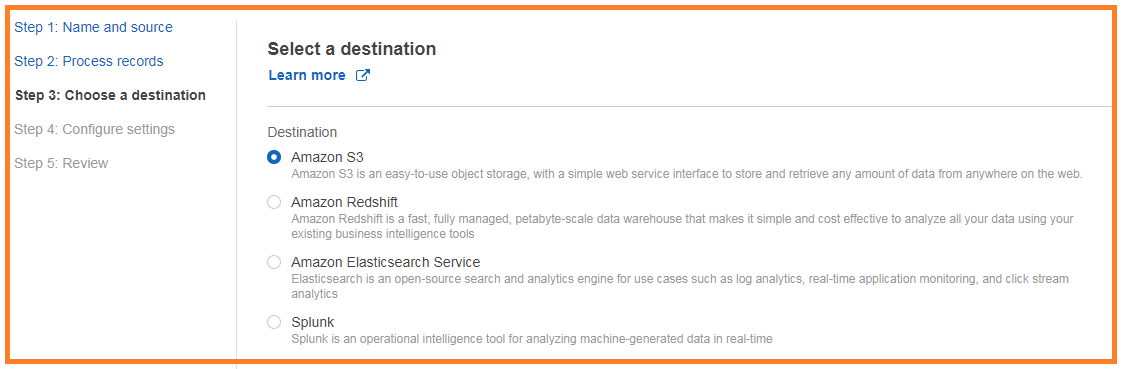
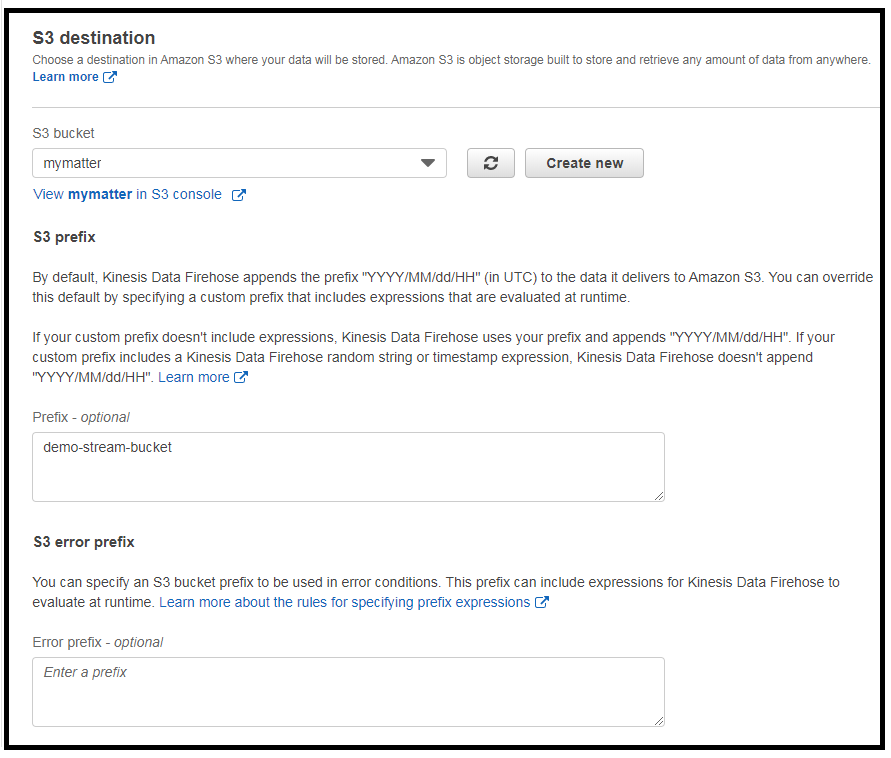
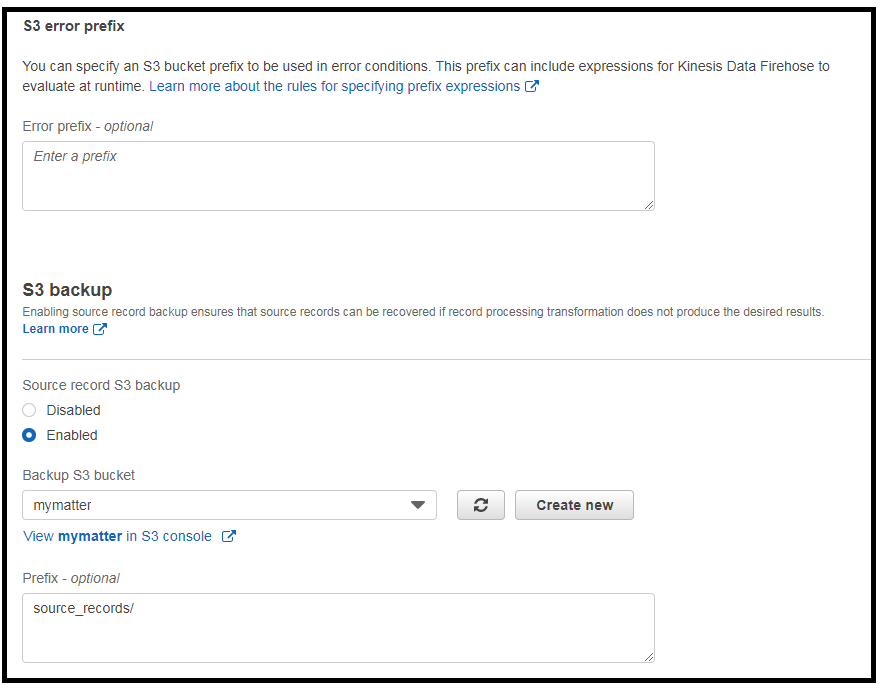
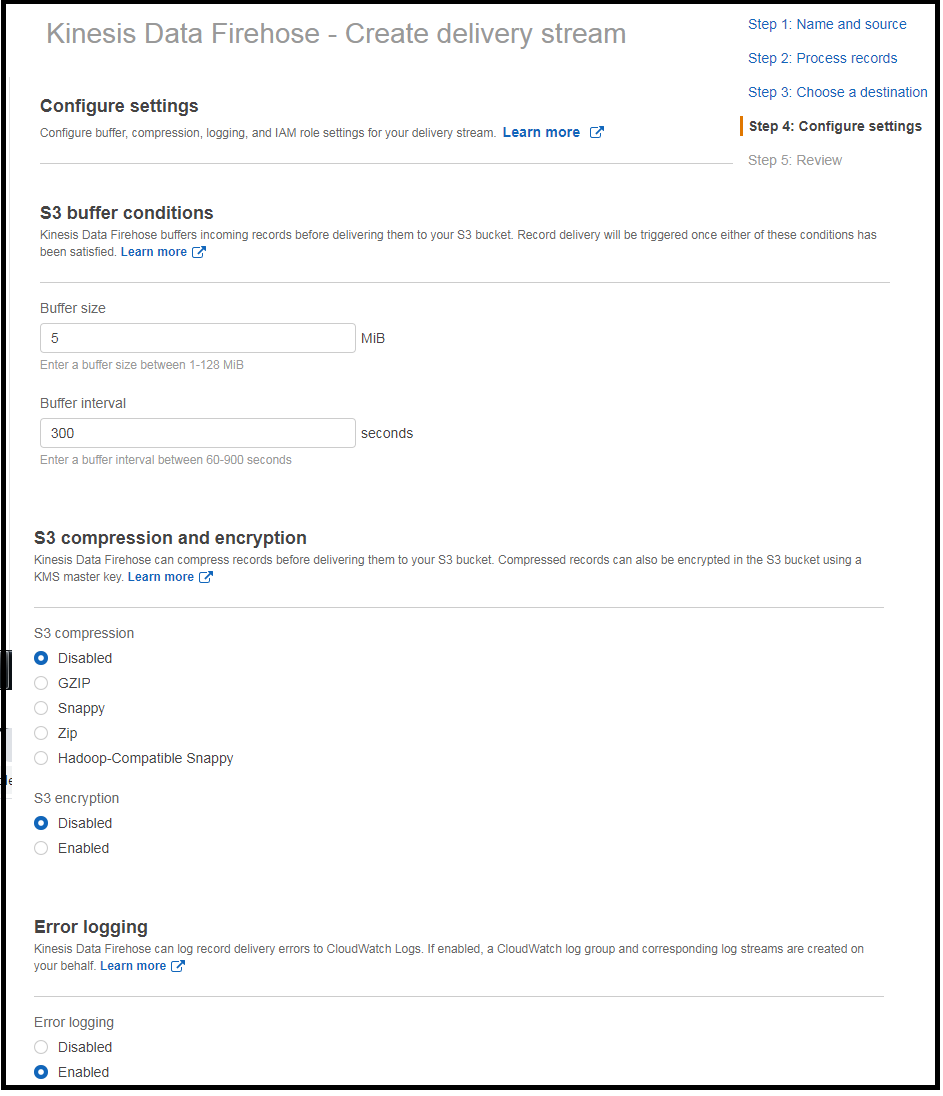
 KINESIS   
  
1. What is streaming data : Streaming data is data that is generated continuously by thousands of data sources, Which typically send in the data records simultaneously and in small sizes of Kb.

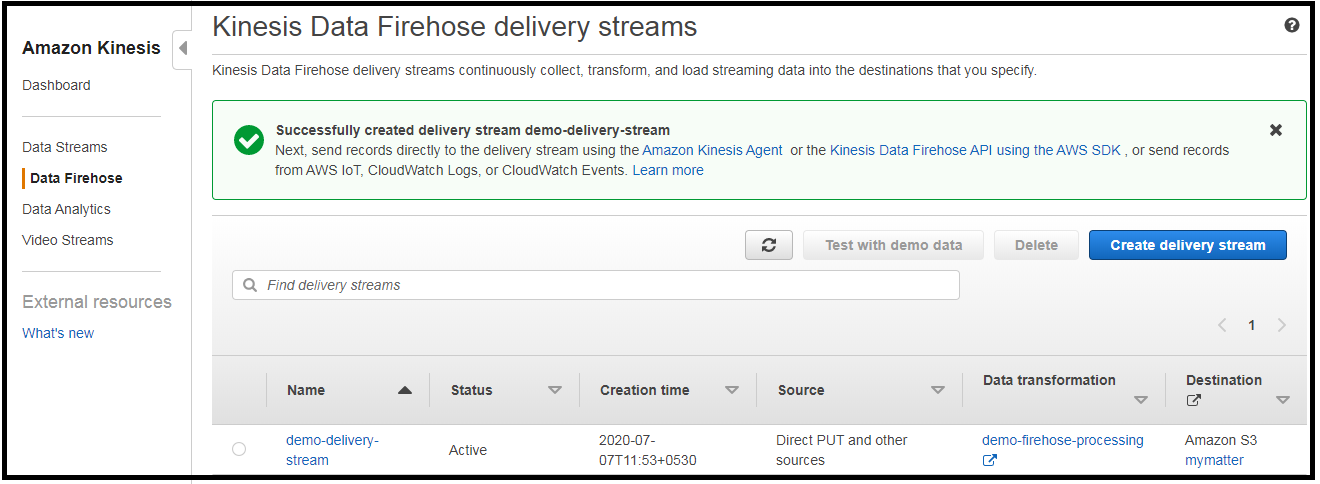
Examples : Purchases from Amazon stores.  
 Stock Prices.  
 Social networking data.  
 Geospatial data (uber.com)

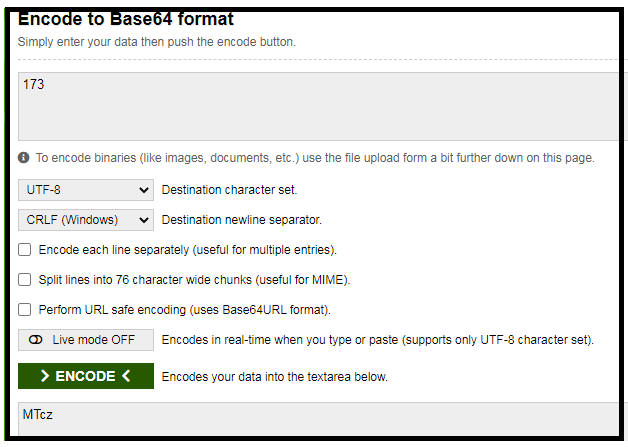
2. What is Kinesis ?  
Amazon’s Kinesis is a platform to which you send your streaming data to.  
Kinesis makes it easy to load and analyze streaming data.

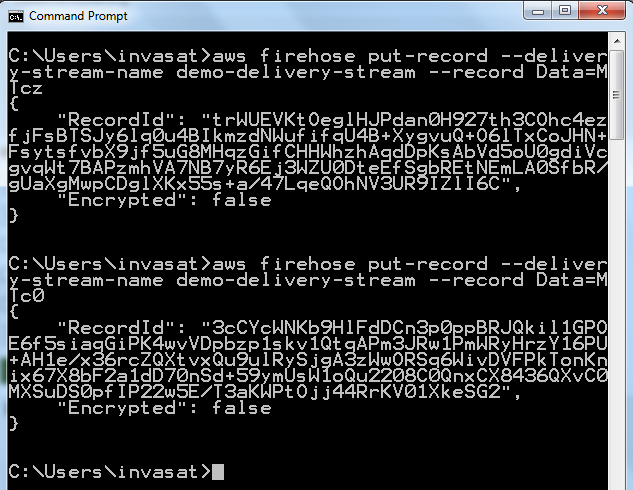
3. Kinesis Services.  
   
  
Kinesis Streams :   
  
- Kinesis Streams consists of shards.  
- A Shard gives you 5 transactions per second for read + not exceeding 2MB/sec.  
- A shard also gives you up to 1000 records per second for writes + not exceeding 1MB per second (including partition keys)  
- You can retain data by default from 24 hours(default) to 7 days.  
- the data consumers take the data from the shard and turn it into something useful.  
- After the data consumers have done processing the data, they can send that data to be stored in a variety of aws services like DynamoDB/S3/Elastic Map Reduce/Redshift.  
  
  
  
Kinesis Firehose :   
- you don’t have to worry about shards and manually adding shards to keep up with the data.  
- You don’t have to worry about data consumers also.  
- There is no automatic data retention window. As soon as the data comes into Kinesis firehose, it is automatically analyzed using AWS Lambda , Or it is sent directly on to S3 Or other locations like RedShift. If you are sending to RedShift via Firehose, we have to write to S3 first and that data is copied to redshift.  
- You can also send this data to elastic search cluster.  
  
  
  
Kinesis Analytics :   
  
- This service allows you to run SQL query of the data, as it exists within ‘firehose’ or ‘streams’.  
- We can store the SQL query result in S3/ Redshift/Elastic Search Cluster.  
- It’s a way of analyzing the data inside Kinesis using SQL.  
  
4. Kinesis LAB.  
Create a Kinesis Stream and interact with using AWS CLI.  
   
  
4.1 Create a data stream.  
   
  
  
  
  
4.2 Make sure you have aws cli installed and a user configured.  
  
  
4.3 Create a stream and list the streams and describe a stream – cli commands  
  
  
4.4 Put a few records in the data stream.  
We need to base64 encode the data.  
  
  
  
  
4.3 Get records.   
- first get the shard iterator and then get the records  
TRIM\_HORIZON : This will fetch the oldest data.  
  
  
  
ApproximateAriivalTimestamp : Time at which the record arrived to Kinesis.  
NextShardIterator : Use this iterator to fetch the next set of records.

4.4 Delete a stream.  


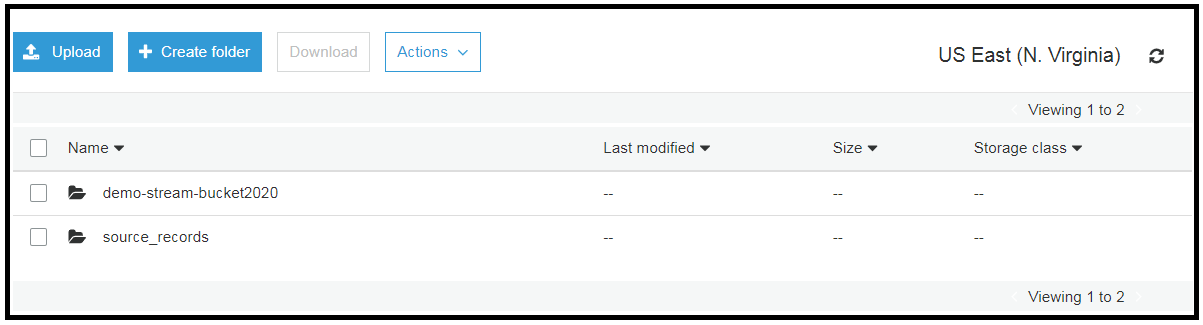
5. Kinesis FIREHOSE lab  
  
  
Step 1 : The input to the delivery stream can be from any of the cli / cloud watch events / cloud watch logs.  
But now we shall be giving the input from CLI.  
  
  
Step 2 :   
  
  
  
Choose a lambda blueprint for kinesis processing. [General Firehose processing]  
We will use the below lambda to add a new line between the records.  
  
  
Step 3 :   
  
  
  
Step 4 : The data that we write will go firehose and to S3.  
  
  
Step 5 : S3 storage configuration.  
  
  
  
Step 6 : Kinesis Firehose buffers incoming data, before delivering it to S3.  
Buffer Size or Buffer Interval – Which ever condition is fulfilled first – triggers data delivery to S3.  


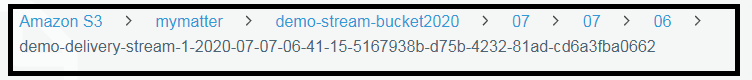
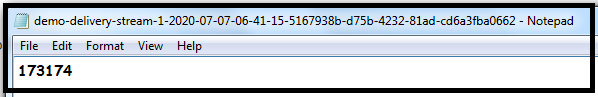
Step 7. Delivery Stream Dashboard after creation.  


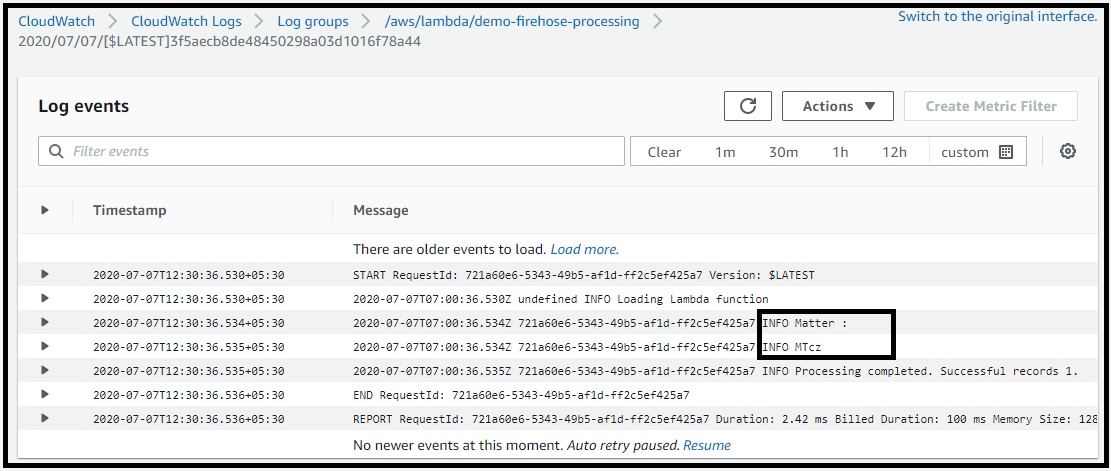
Step 8.  
Encode the data.  


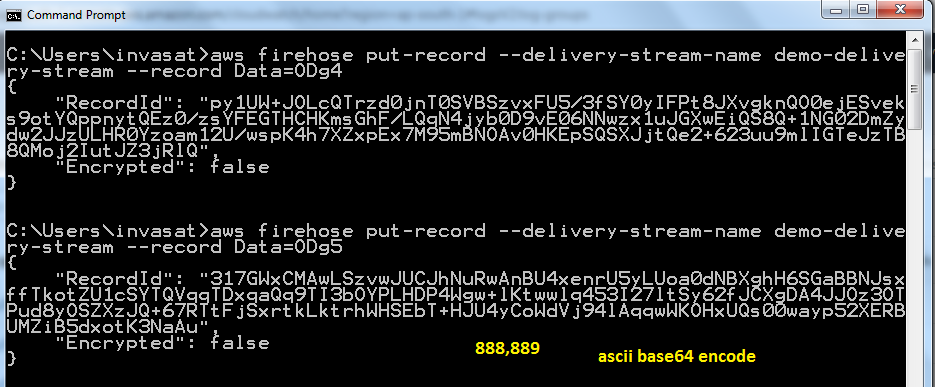
Step 9 : Put records to Firehose.  


Step 10:

Take a look at the data in S3  


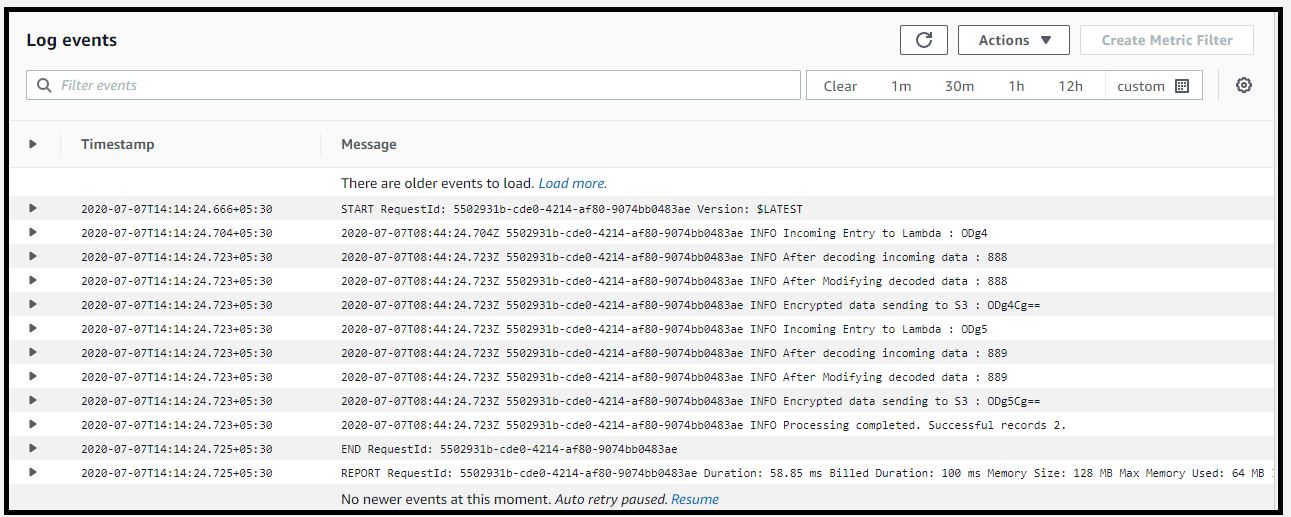
Main Storage Bucket.  
  
  
  
The source records bucket also contains the same file.  
  
  
Step 11. Now let’s find out how to add a new line or process the data.  
  
  
The data is base64 encoded. So in Lambda we need to decode, modify and then append the data.



Step 12 : Lets insert 2 records into firehose  
  
  
Step 13 :



**Step 14 : I am not sure why the lambda is not inserting a new line into the S3 record.**

****  
Step 15 :

Buckets created are   
