**PROJECT**

Name-Shivam Tiwari

Submitted to-Simplilearn

Submission Date-22-07-2021

Project Name-Real-Time Analytics on Streaming Data with Amazon Kinesis and Amazon Elasticsearch Service

Real-Time Analytics on Streaming Data with Amazon Kinesis and Amazon Elasticsearch Service

* **Analysis to be done:**Collect and store streaming data, use Amazon Kinesis Analytics to process and analyze the streaming data, apply machine learning algorithm to detect anomalies in the system.
* **Steps to perform:**

1. Create Kinesis delivery stream
2. Simulate streaming application to detect anomalies
3. Open Elasticsearch service and create a new domain
4. Configure Kinesis Firehose to export the results to Amazon ES
5. Update the buffer size and existing IAM role for the process
6. Open the Amazon Kinesis Analytics console and create a new application
7. Connect to the source for further analysis
8. Launch SQL\_Editor and start the application
9. Load the processed data into Kinesis Firehose delivery stream
10. Visualize the data using Kibana

**Overview**

Nowadays, streaming data is seen and used everywhere—from social networks, to mobile and web applications, IoT devices, instrumentation in data centers, and many other sources. As the speed and volume of this type of data increases, the need to perform data analysis in real time with machine learning algorithms and extract a deeper understanding from the data becomes ever more important. For example, you might want a continuous monitoring system to detect sentiment changes in a social media feed so that you can react to the sentiment in near real time.

Amazon Kinesis Data Streams to collect and store streaming data. We then use Amazon Kinesis Data Analytics to process and analyze the streaming data. Then we use Amazon Kinesis Data Firehose to export the anomalies data to Amazon Elasticsearch Service. We then build a simple dashboard in the open source tool Kibana to visualize the result.

**Amazon Kinesis Data Streams**

You can use Amazon Kinesis Data Streams to build your own streaming application. This application can process and analyze streaming data by continuously capturing and storing terabytes of data per hour from hundreds of thousands of sources.

**Amazon Kinesis Data Analytics**

Kinesis Data Analytics provides an easy and familiar standard SQL language to analyze streaming data in real time. One of its most powerful features is that there are no new languages, processing frameworks, or complex machine learning algorithms that you need to learn.

## Amazon Kinesis Data Firehose

Kinesis Data Firehose is the easiest way to load streaming data into AWS. We will capture, transform, and load streaming data into Amazon Elasticsearch Service in this project.

## Amazon Elasticsearch Service

Amazon Elasticsearch Service is a fully managed service that makes it easy to deploy, operate, and scale Elasticsearch for log analytics, full text search, application monitoring, and more.

**Points to care**

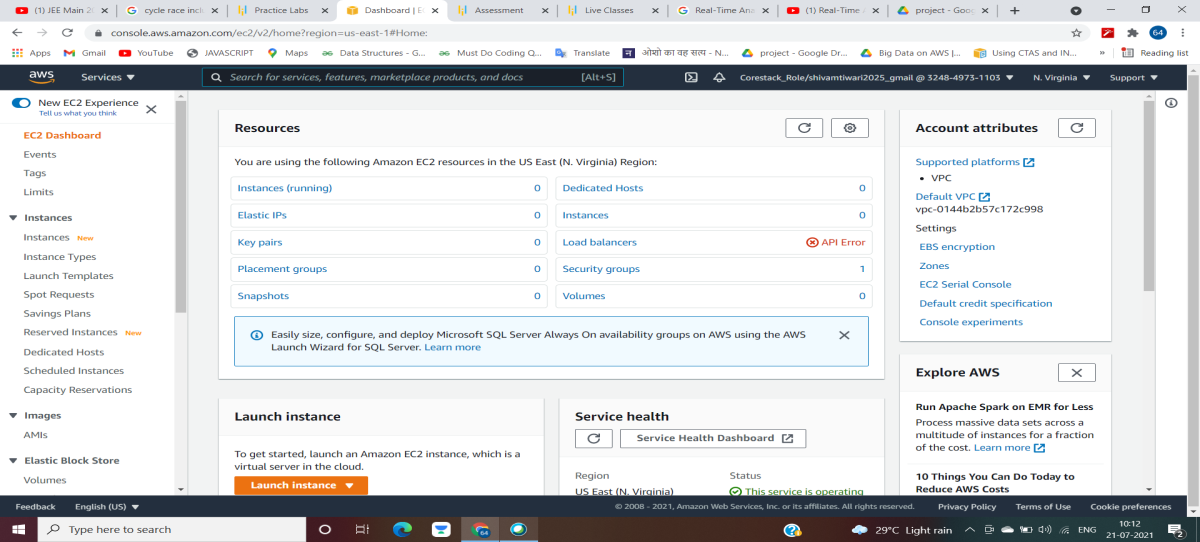
* It is mandatory to solve this problem using proper scripts and setting up the streams
* Components should be coupled properly
* Steps should have to be followed stepwise.

**Solution:**

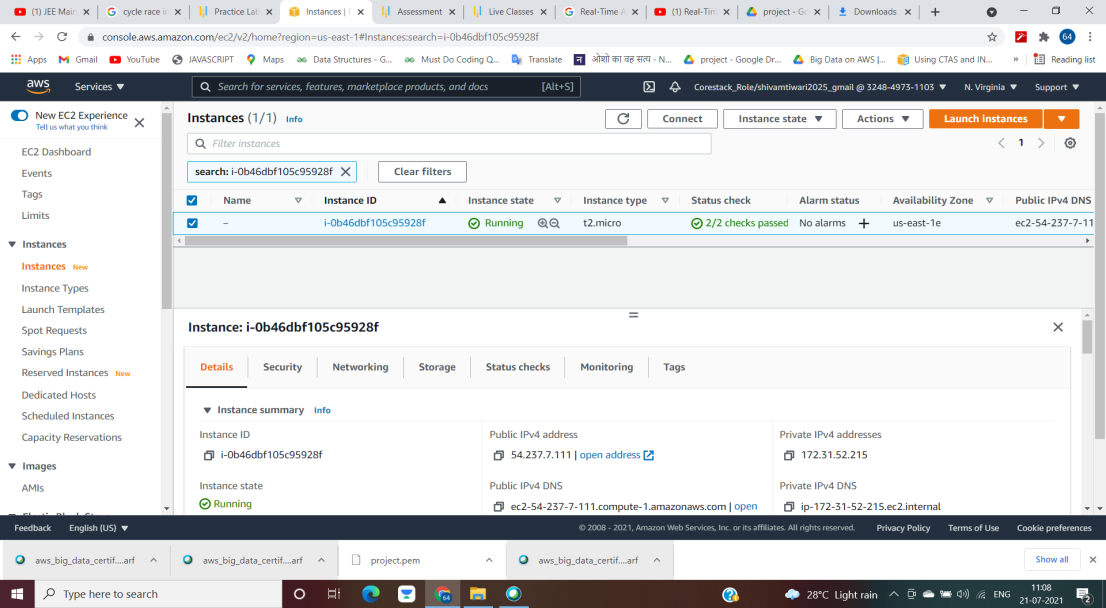
This steps of this project are as follows:

1. Create an EC2 instance.

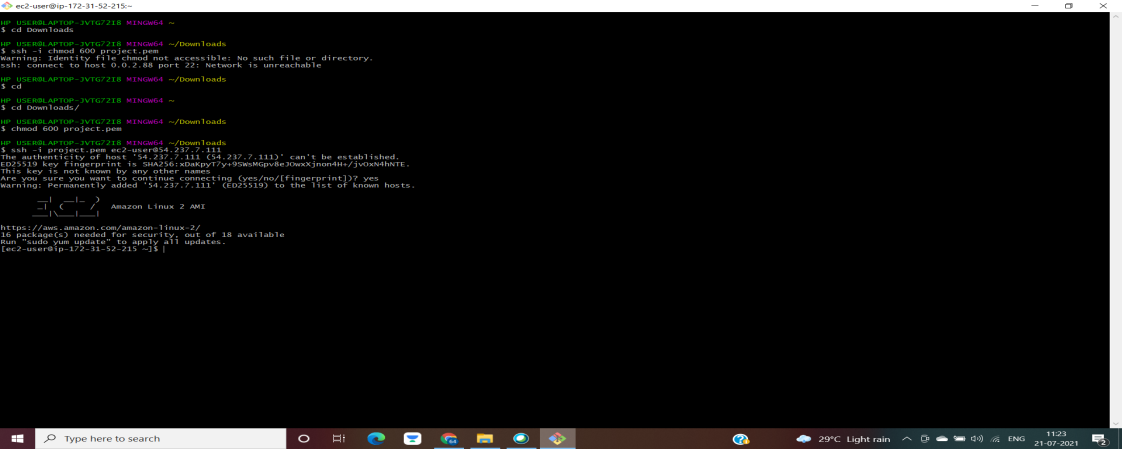
* Open EC2 console and click on launch instance.



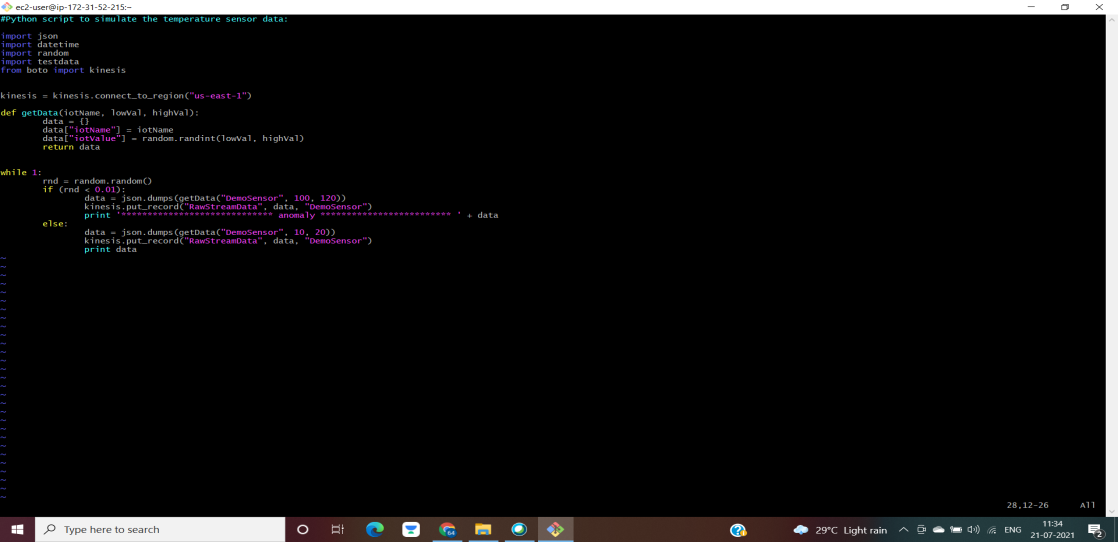
* Now the instance will appear as running.



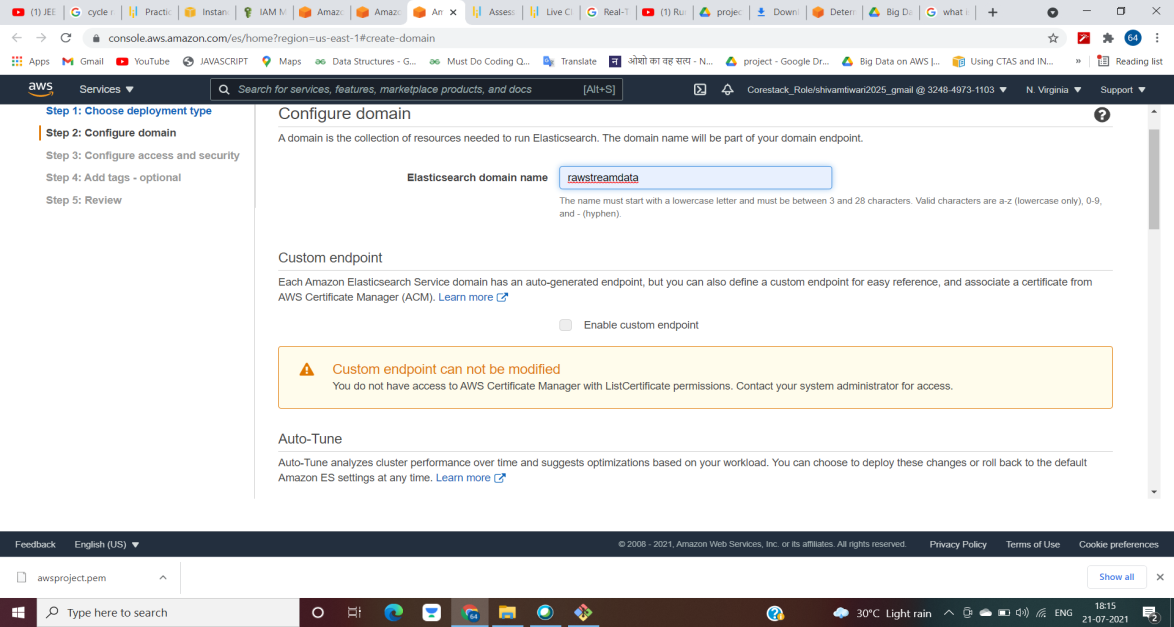
* Connect the Server

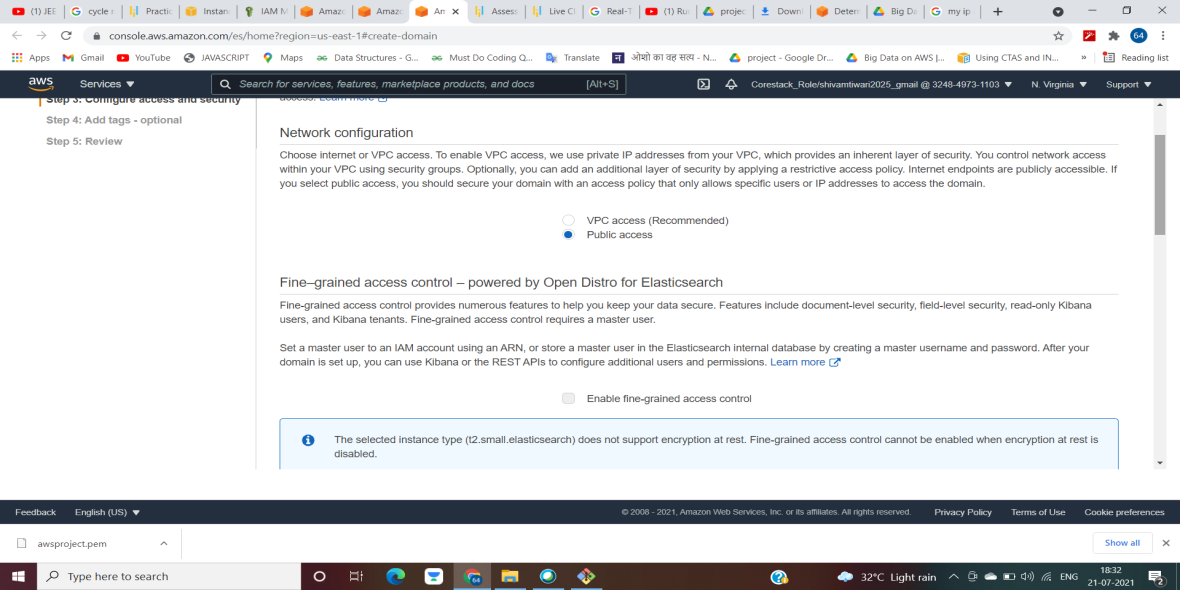


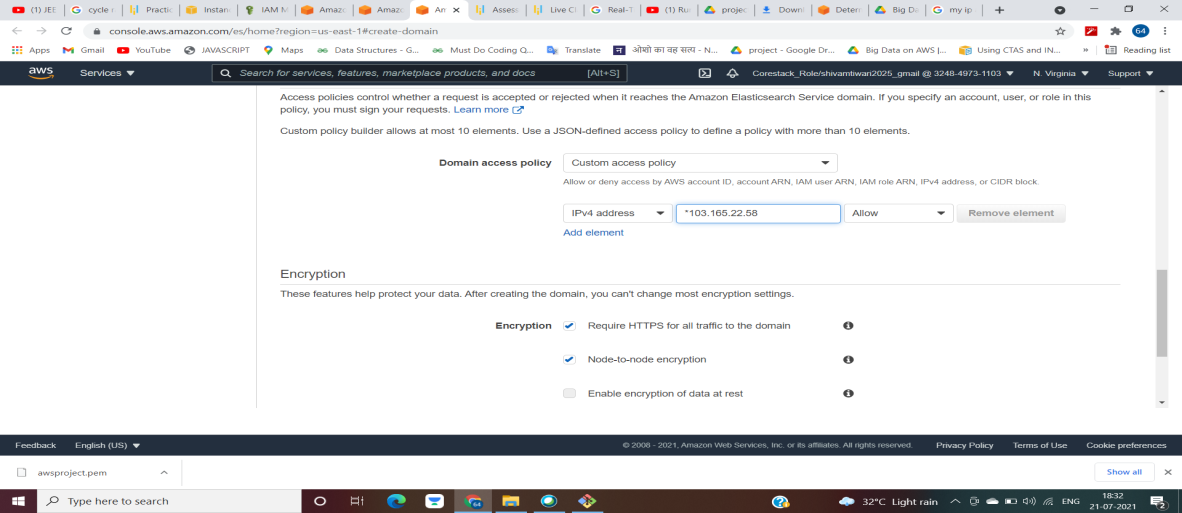
* Install the Python Script on it

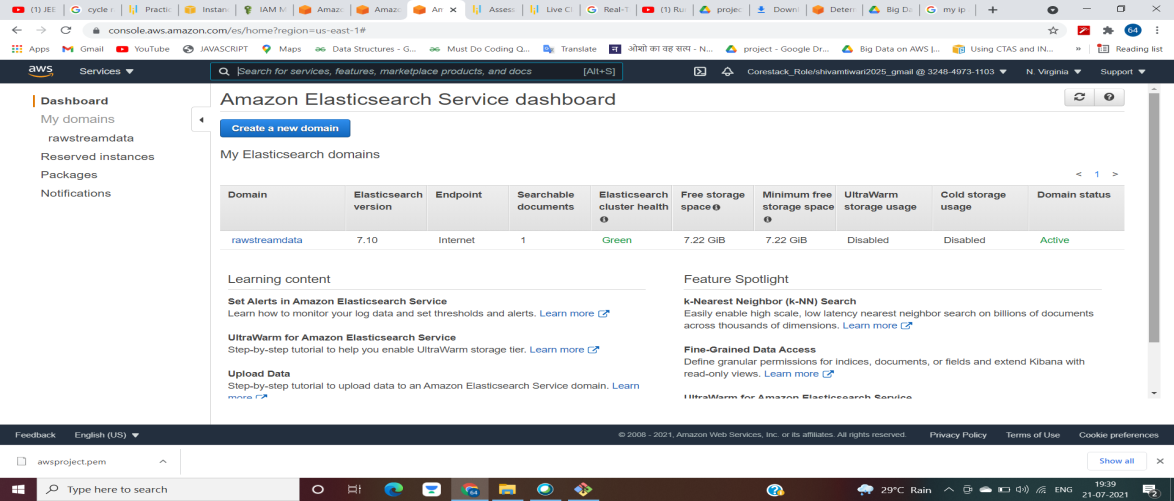


1. Create an Elasticservice domain



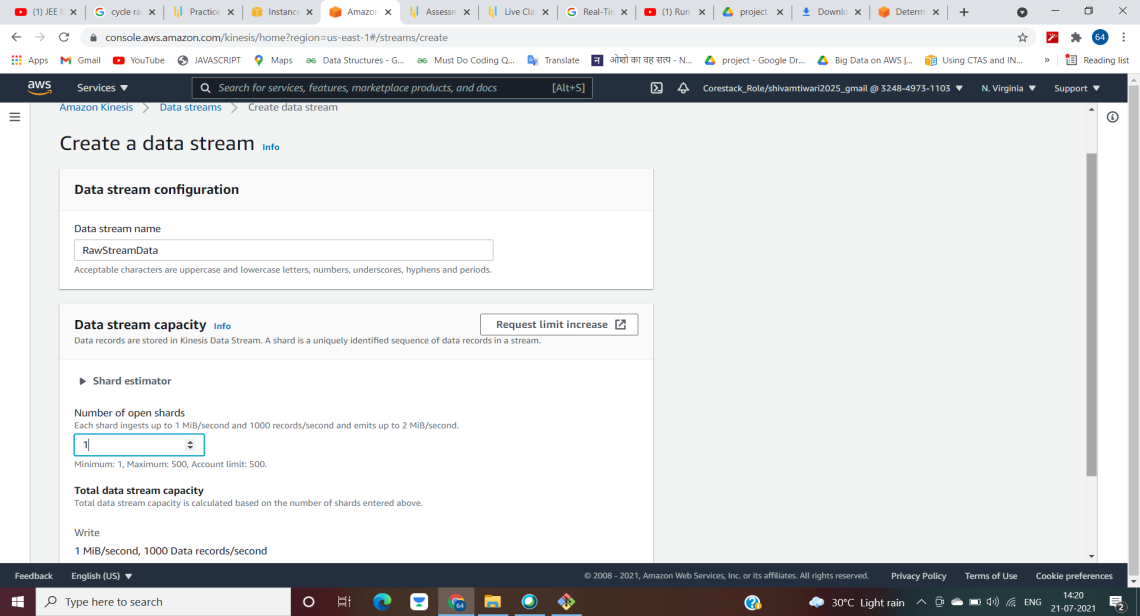






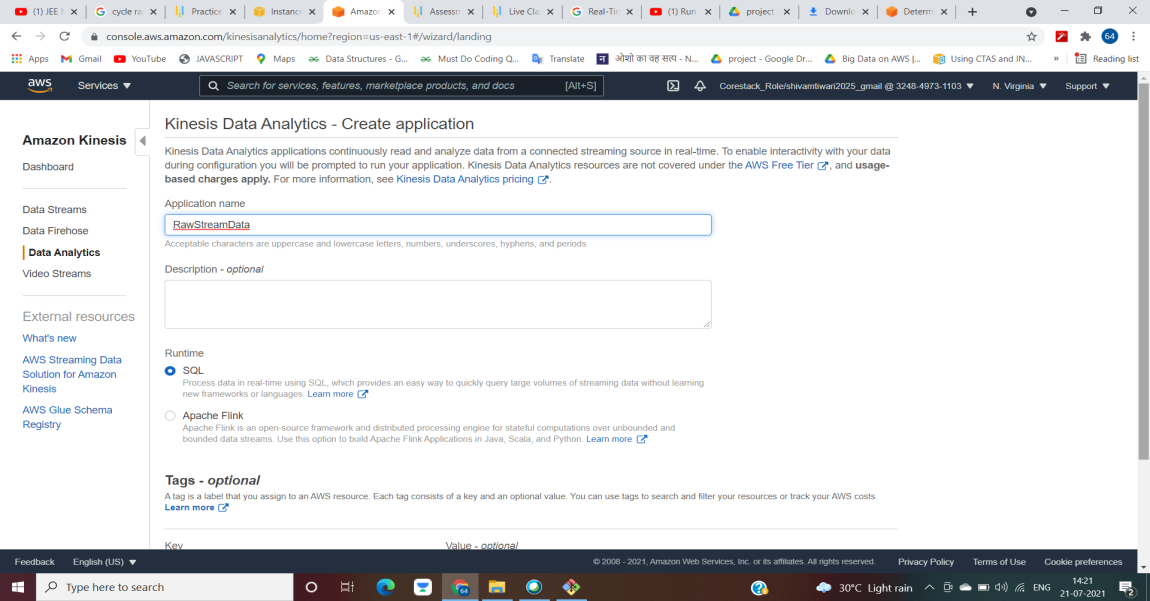
1. Create a Kinesis Data Stream

* We will keep the number of shards as 1
* The name should be same as that of python script

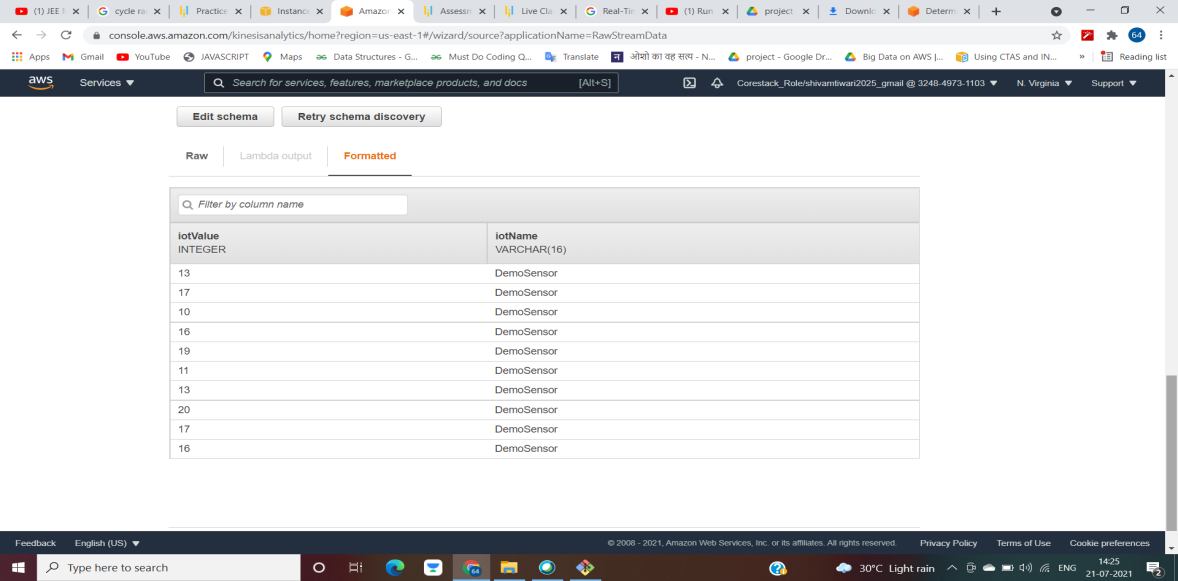


1. Create Kinesis Data Analytics

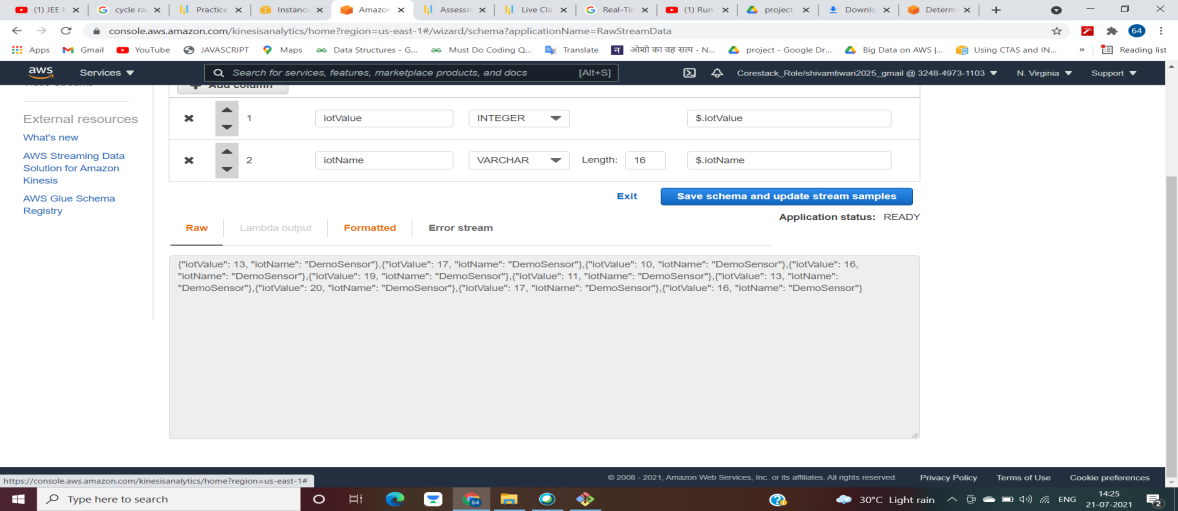
* We will keep the source as Kinetic Data Stream
* The name we will keep the same as that of Kinesis Data Stream and the python script
* Runtime will be SQL



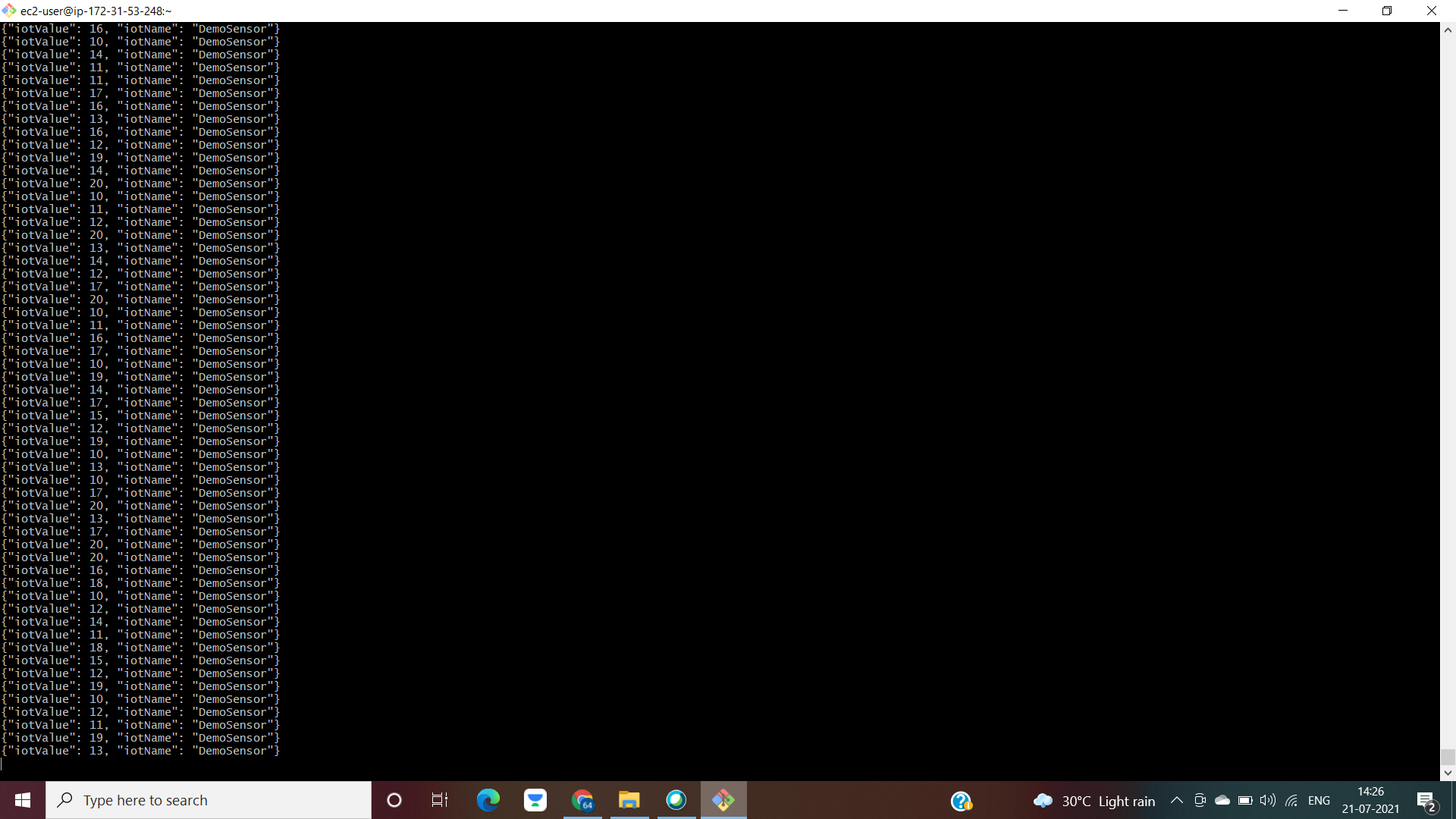
* Then we have to connect for the streaming data and discover schema
* The table will look appear in the same manner as the app running on EC2 instance



* We can also see the raw data as shown below

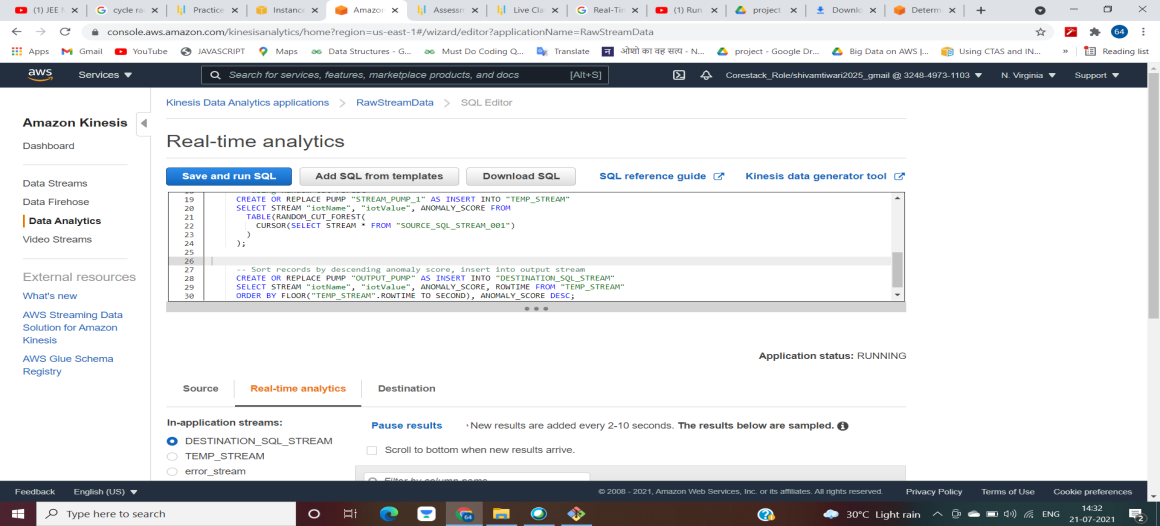


* The data will look like the same in EC2 instance as well

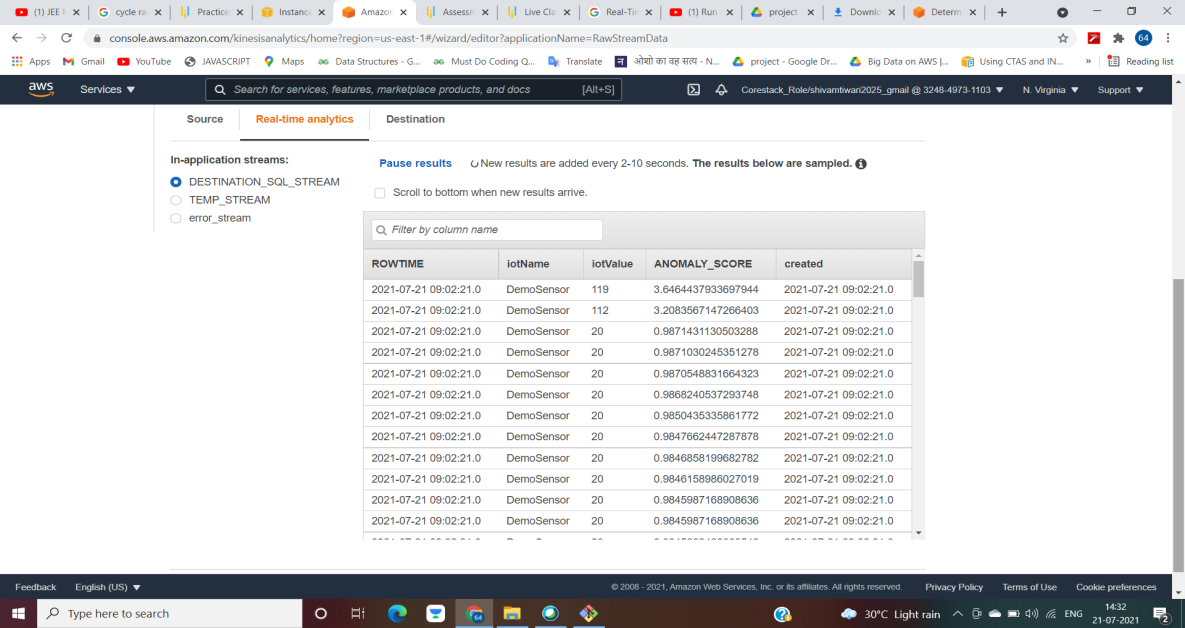


1. Go to Realtime Analytics and enter the SQL command

* The command will look like this

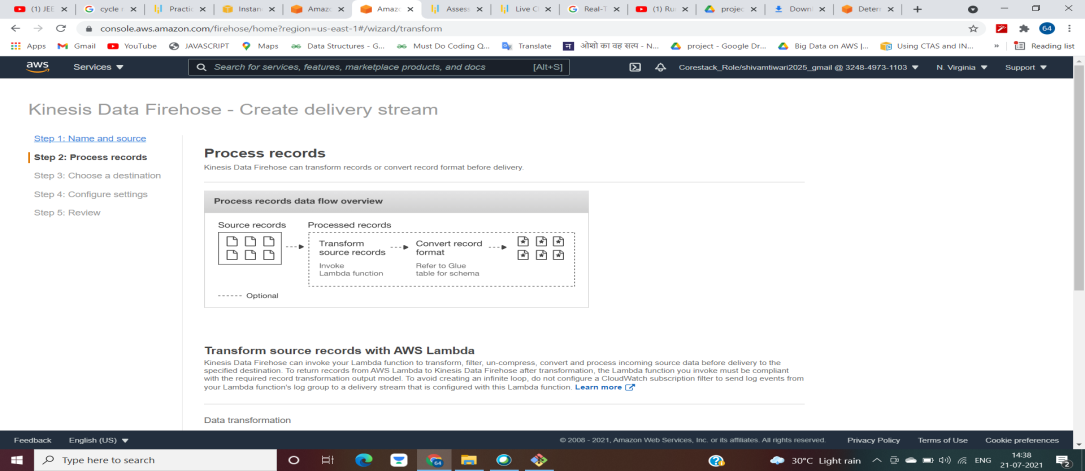


* After running the SQL, the SQL stream will appear as below



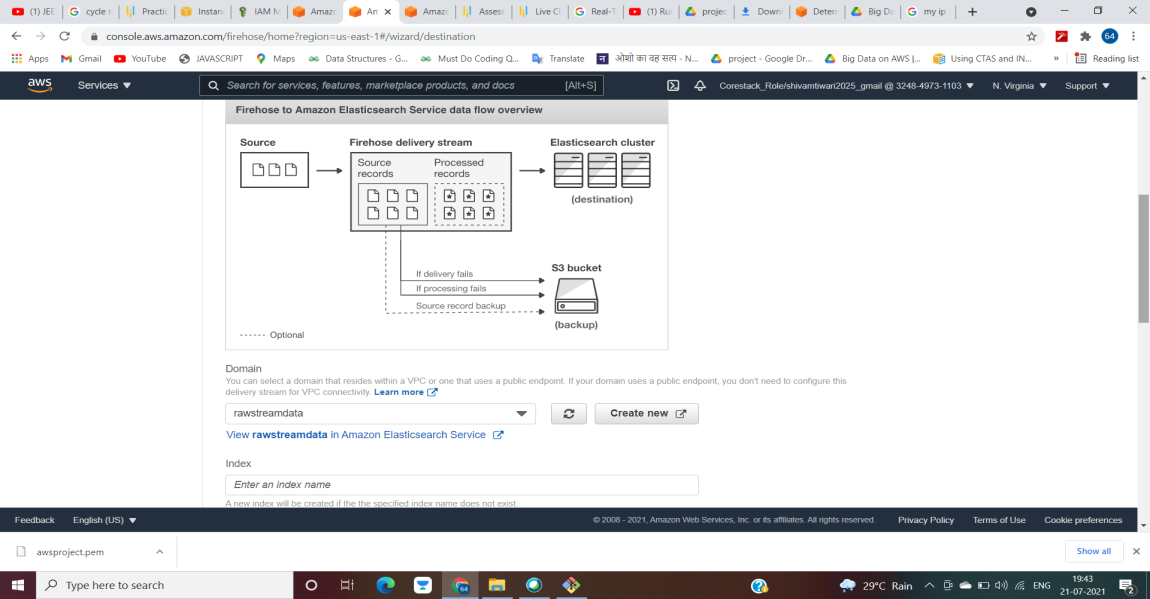
1. Connect the destination to firehose

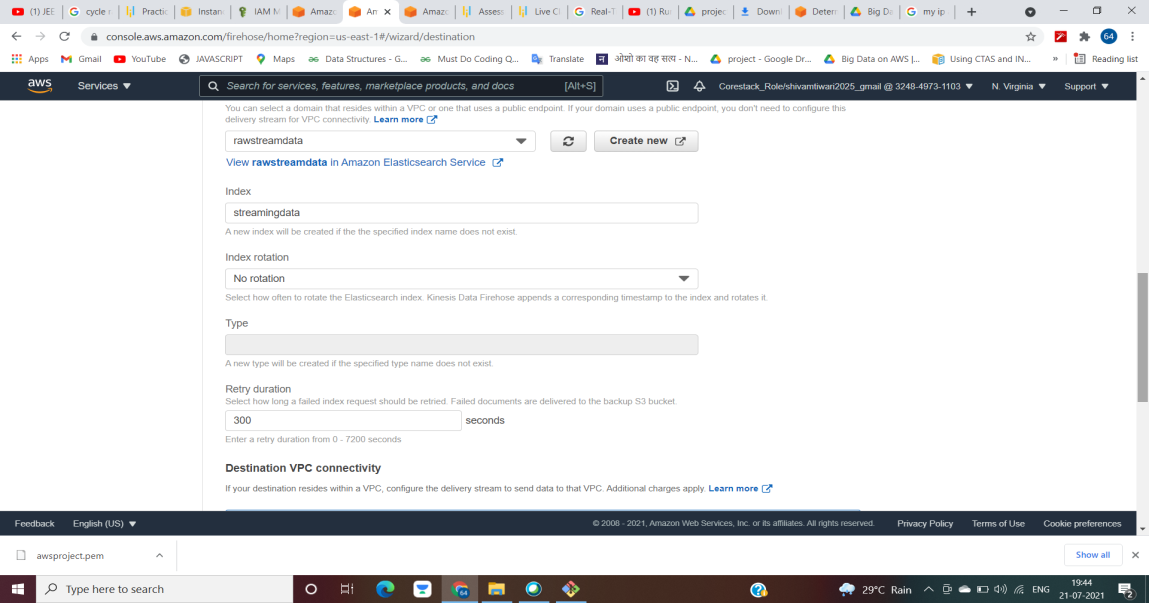
* Fill the details and have to choose the destination as Amzaon Elasticsearch.





* Now provide the Elasticservice domain created

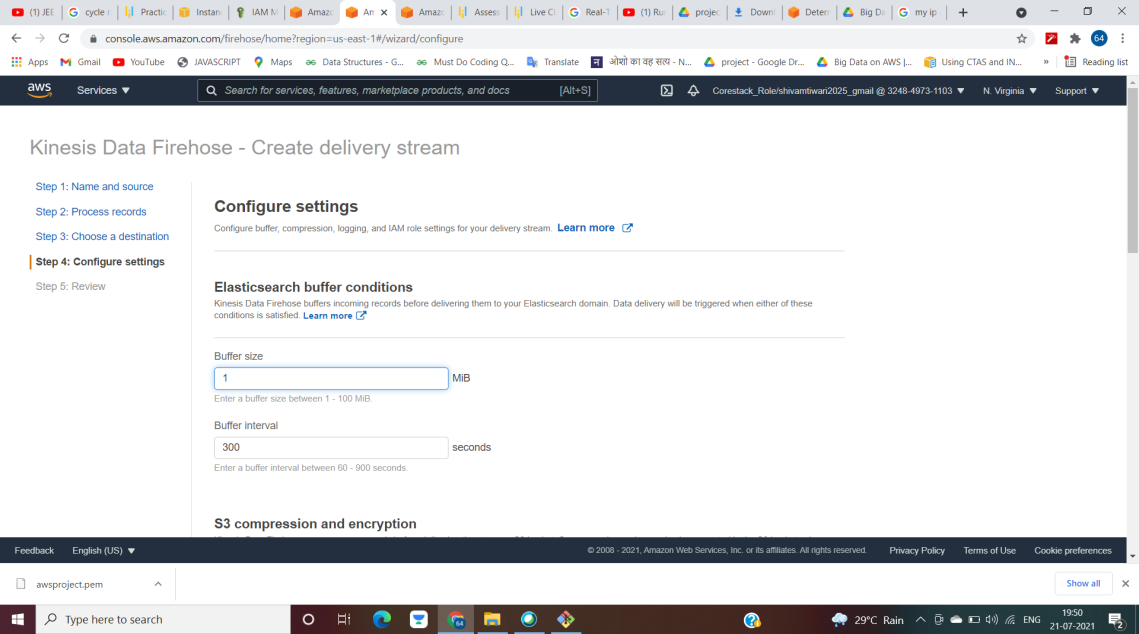




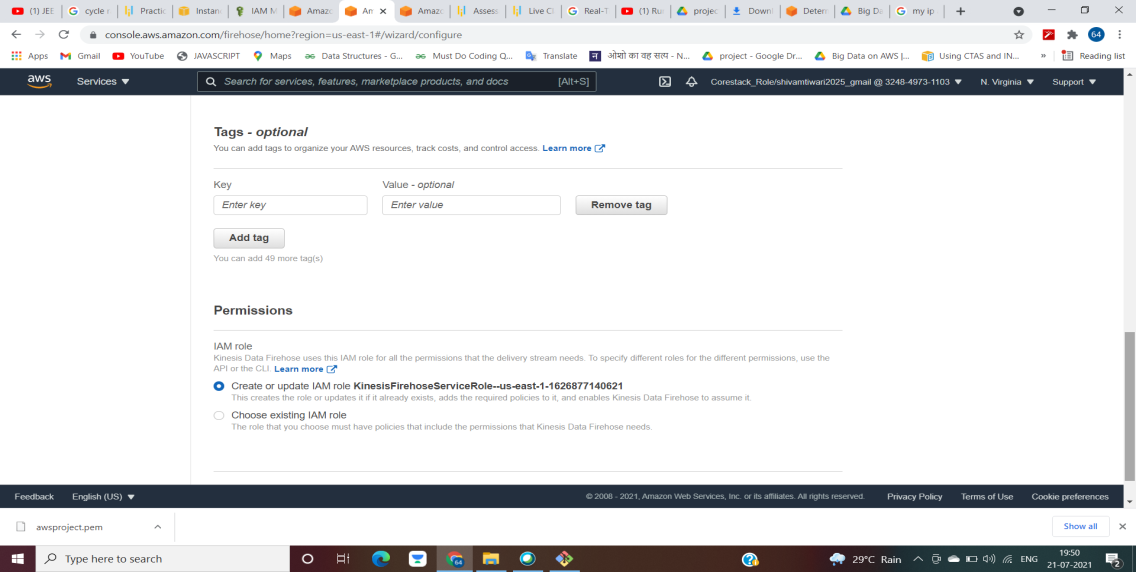
* Choose an S3 bucket

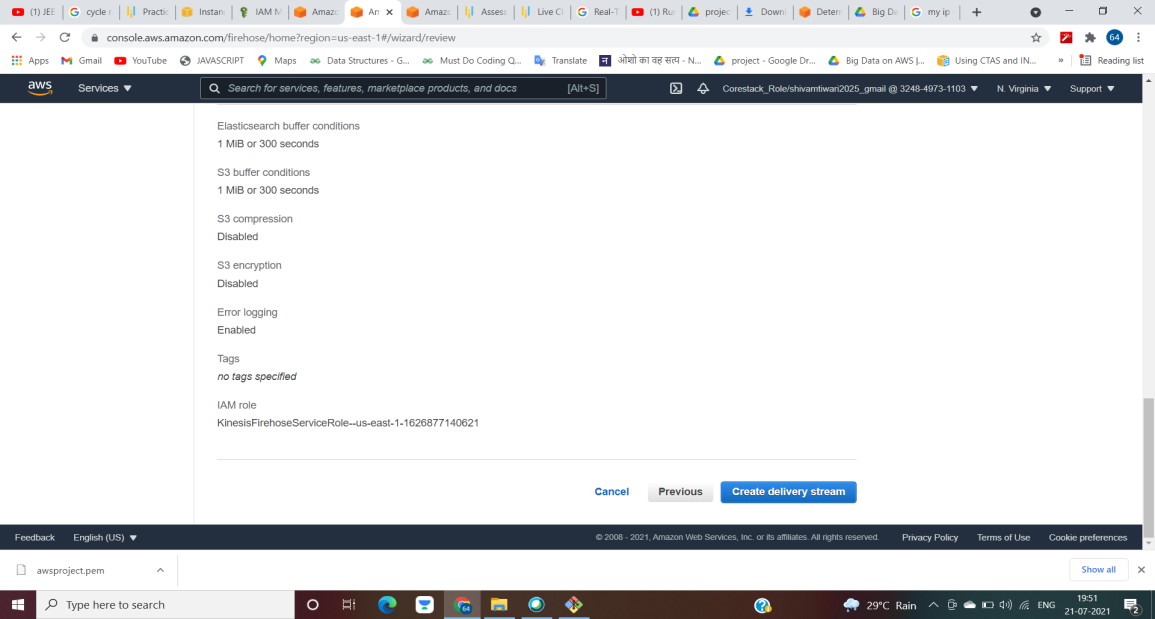


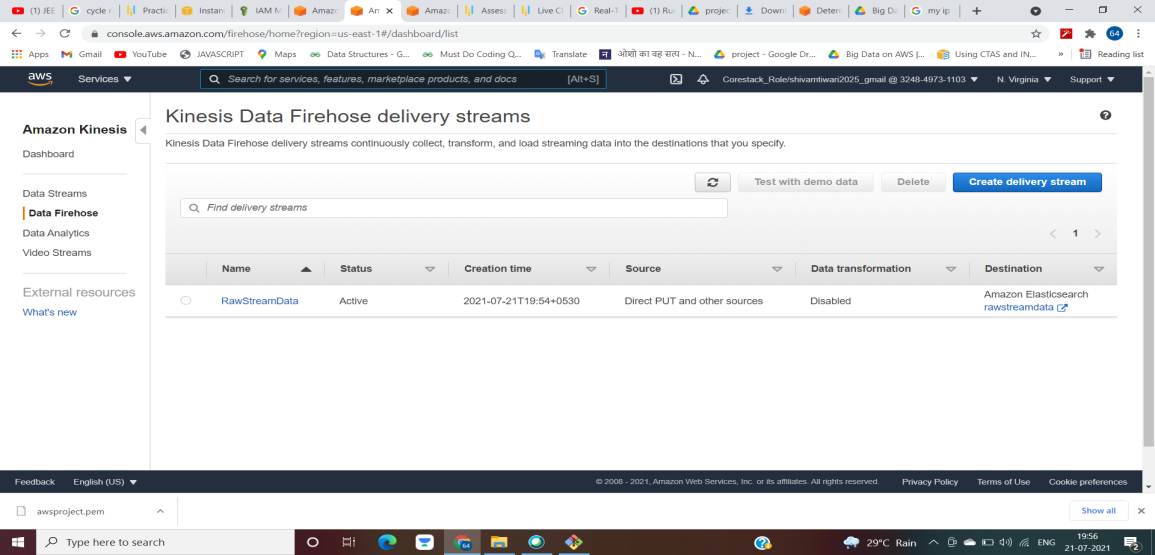
* Then configure settings





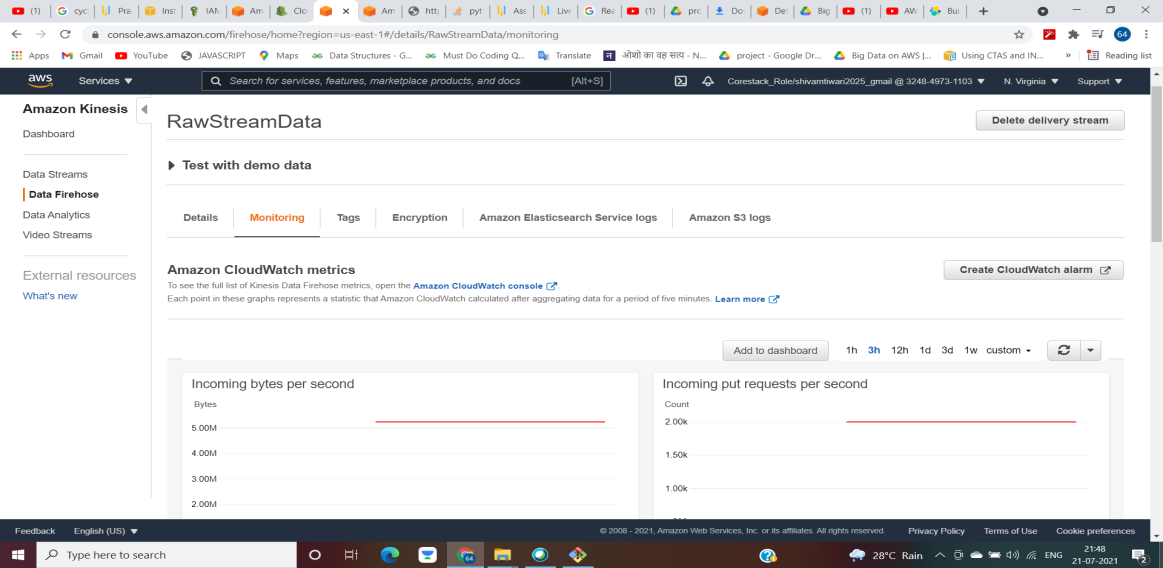




* Kinesis data firehose delivery stream will appear active

1. Go to RawStreamData and click monitoring

* We will see different graphs showing the application to be active and working



1. Open Kibana and we will visualize the data there

