**Elastic Basic – Task 1**

**ELK:**

"ELK" is the acronym for three open source projects: **Elasticsearch, Logstash,** and **Kibana.**

**Elastic Search:**

A highly scalable **open-source full-text search** and **analytics engine**. It allows you to **store, search,** and **analyze big volumes of data quickly** and in near real time. It is generally used as the underlying engine/technology that powers applications that have complex search features and requirements.

**Port:** localhost:9300 runs on TCP

**Command to run:** bin/elasticsearch

**Kibana:**

It is an **open source analytics** and **visualization platform** designed to work with Elasticsearch.

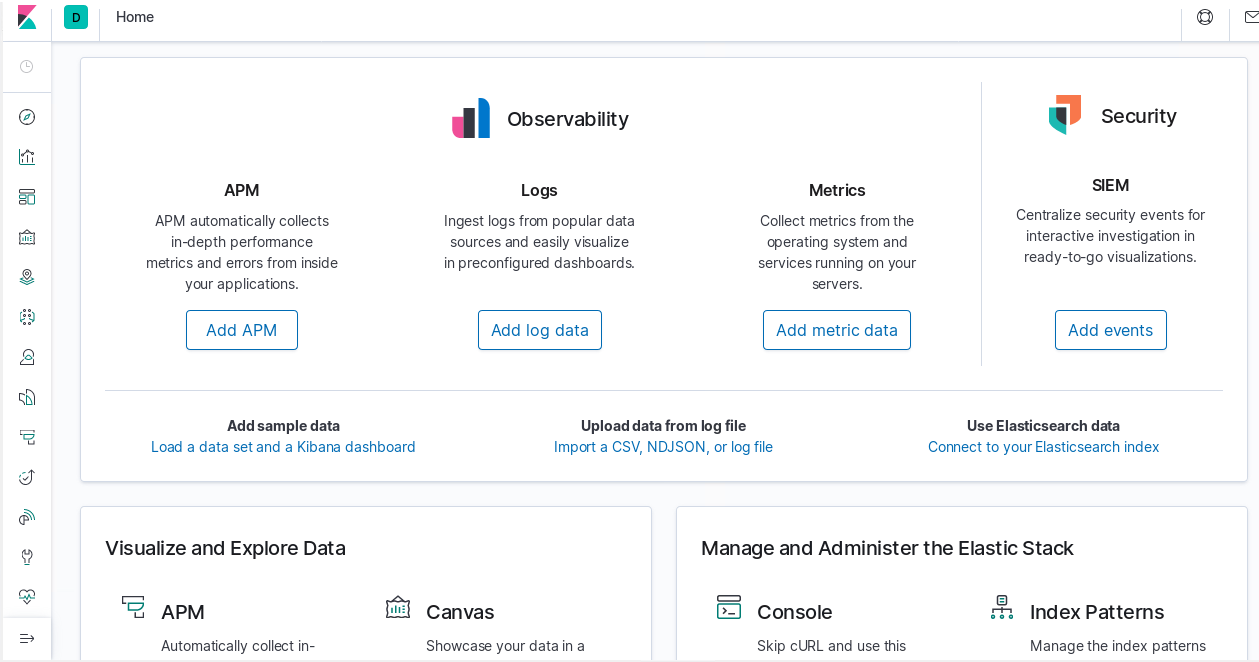
We can use kabana for:

**1. Search**

**2. View**

**3. Interact with data stored in Elastics indices.**

It **performs advanced data analysis** and **visualize your data** in a variety of charts, tables, and maps.

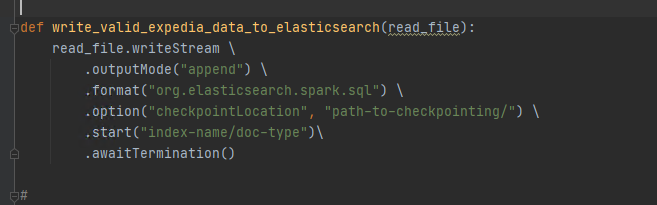


**Port:** http://localhost:5601

**Command to run:** bin/kibana

**Spark Streaming Job:**

To configure spark streaming from elasticsearch, we have added format to org.elasticsearch.spark.sql. It will upload data on default port https://localhost:5601

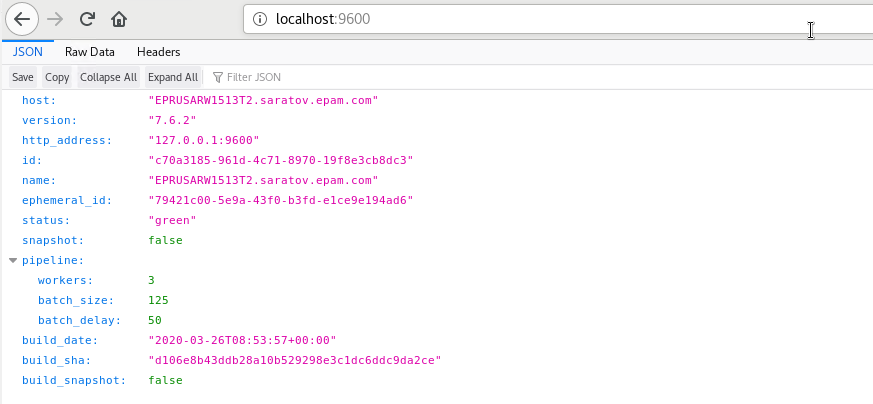
****

**Command to run above piece of code**

spark-submit --packages org.elasticsearch:elasticsearch-hadoop:7.8.0 main.py

**Logstash:**

It is an open source **data collection engine with real time pipeline capabilities** can dynamically unify data from disparate sources normalize the data into destinations of your choice.

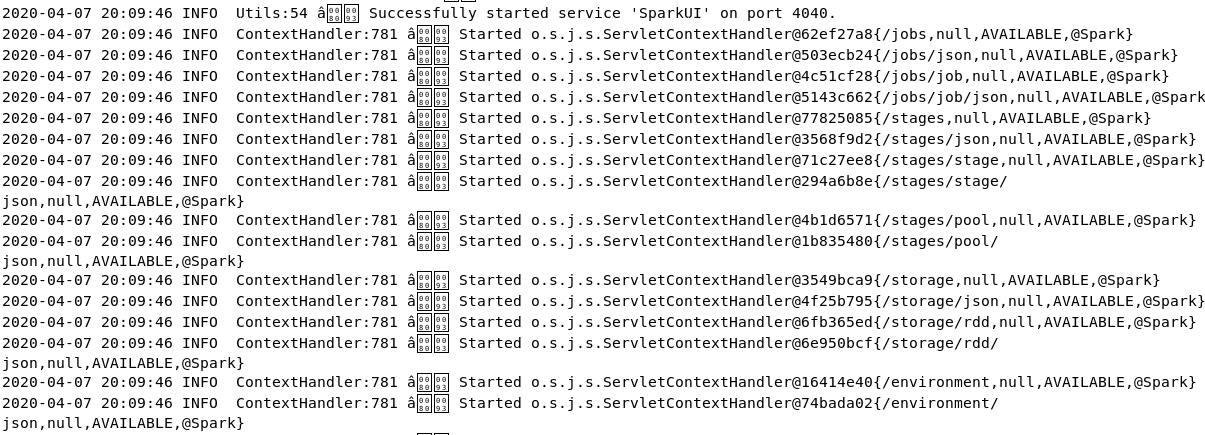


**Port:** http://localhost:9600

**Command to run:** bin/logstash –f {configuration\_file\_name.conf }

**Logs:**

These are logs that are generated by Spark Streaming project.

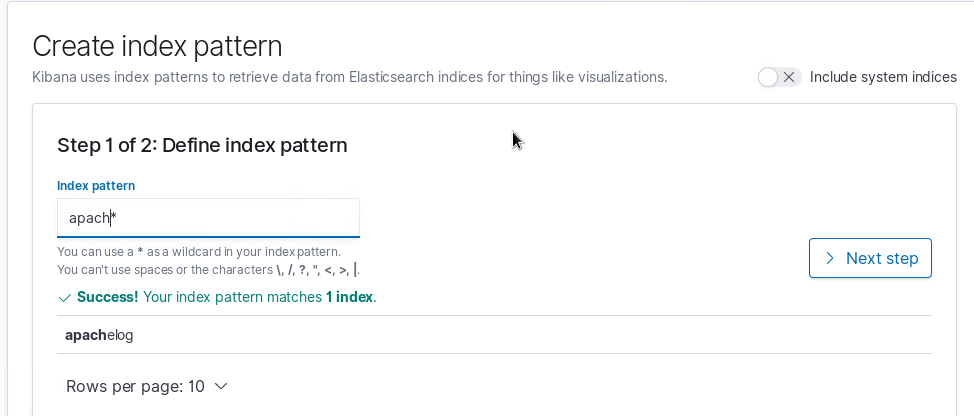


I have written these logs by using log4j.

**Indices:**

An **index** pattern tells **Kibana** which **Elasticsearch indices** contain the data that we have uploaded to kibana.

I have uploaded logs to kibana by creating apachelog indices.

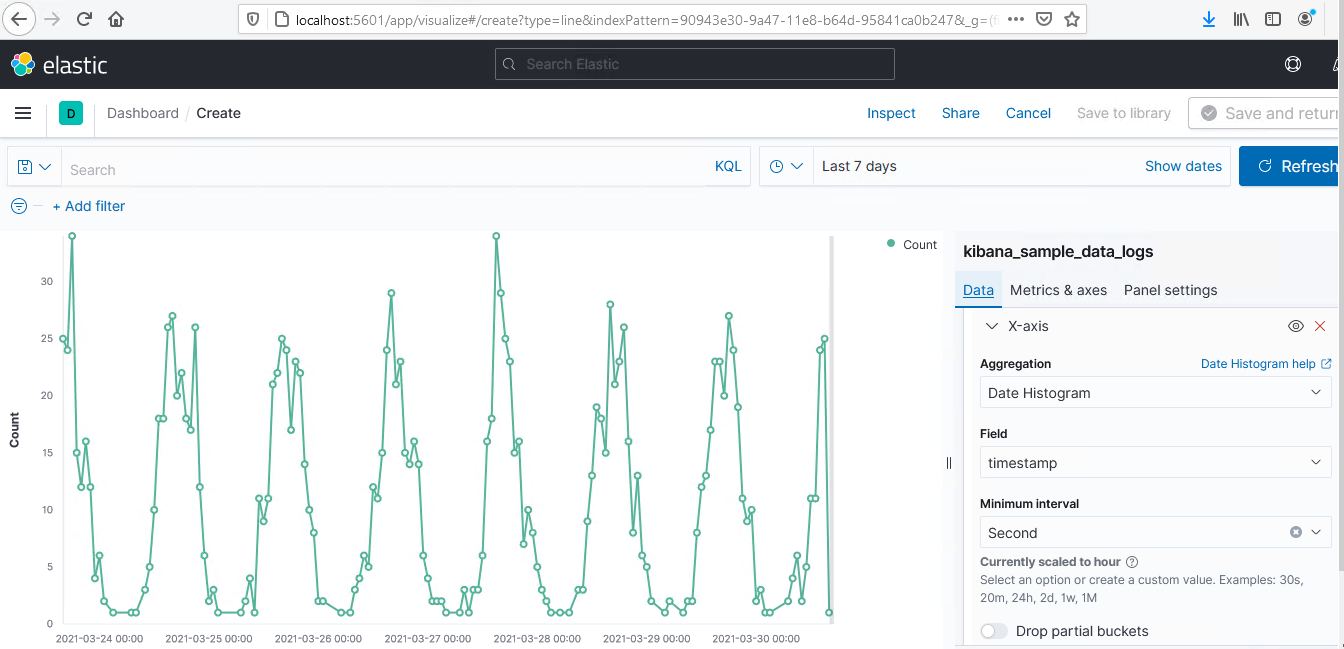


**Dashboard**:

It is a **collection of visualizations, searches,** and **maps, typically in real-time**. Dashboards provide at-a-glance insights into your data and enable you to drill down into details. Some features of kibana are given below:

* Add visualizations, saved searches, and maps for side-by-side analysis
* Arrange dashboard elements to display exactly how you want
* Customize time ranges to display only the data you want
* Inspect and edit dashboard elements to find out exactly what kind of data is displayed

To visualize the logs, I used histogram to show number of events processed per hour.

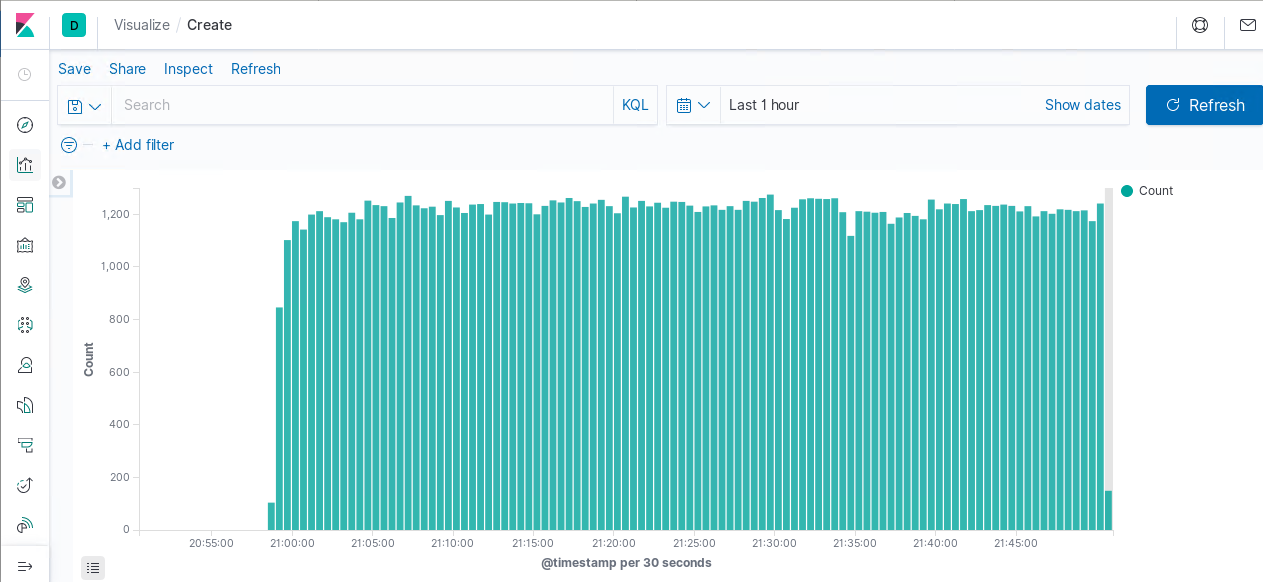


I the above image, we can see that y-axis shows no. of logs & x-axis shows Timestamp per 10 sec.

In the above graph:

On y-axis there is a count.

And on x-axis we have bucket of aggregation. We used data histogram aggregation and showing incoming records in timestamp.



I the above image, we can see that y-axis shows no. of logs & x-axis shows Timestamp per seconds.