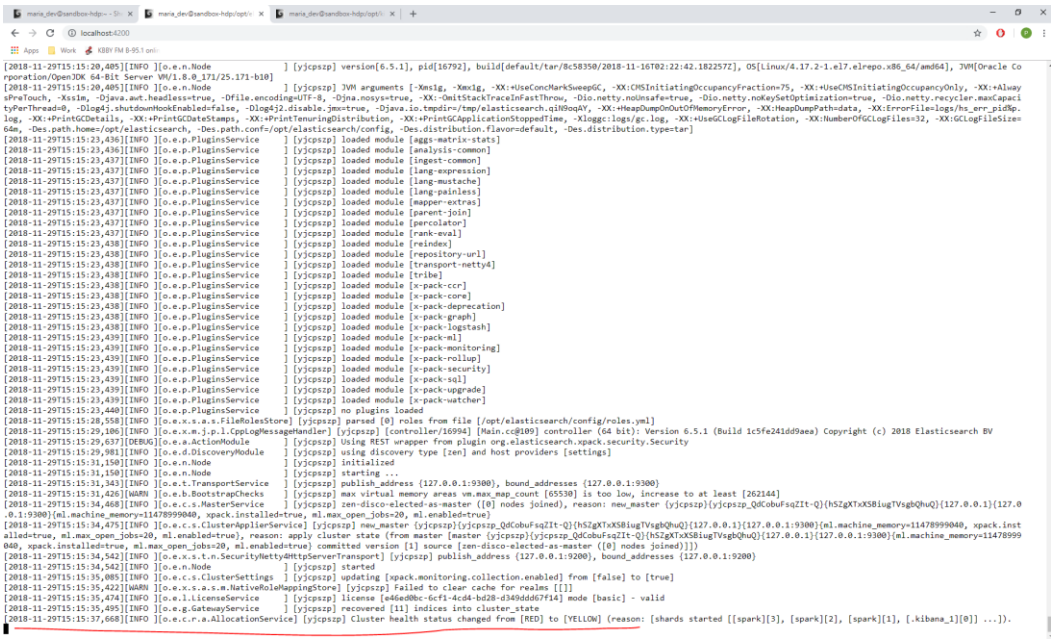
ELK homework.

Configuring kibana and elastic search.

I downloaded both elastic search and kibana, versions 6.5.1 from the official website.

Here's elastic search running:



And here's kibana running:

```
[maria_dev@sandbox-hdp ~]$ cd /opt/kibana/bin
[maria_dev@sandbox-hdp bin]$ ./kibana
  log   [13:46:36.187] [info][status][plugin:kibana@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:36.250] [info][status][plugin:elasticsearch@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.255] [info][status][plugin:xpack_main@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.266] [info][status][plugin:searchprofiler@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.271] [info][status][plugin:ml@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.329] [info][status][plugin:tilemap@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.332] [info][status][plugin:watcher@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.354] [info][status][plugin:license_management@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:36.358] [info][status][plugin:index_management@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.376] [info][status][plugin:rollup@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.638] [info][status][plugin:timelion@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:36.641] [info][status][plugin:graph@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.656] [info][status][plugin:monitoring@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:36.663] [info][status][plugin:spaces@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.678] [warning][security] Generating a random key for xpack.security.encryptionKey. To prevent sessions from being invalidated on restart, please set xpack.security.encryptionKey in kibana.yml
  log   [13:46:36.686] [warning][security] Session cookies will be transmitted over insecure connections. This is not recommended.
  log   [13:46:36.695] [info][status][plugin:security@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.728] [info][status][plugin:grokdebugger@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.736] [info][status][plugin:dashboard_mode@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:36.742] [info][status][plugin:logstash@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.755] [info][status][plugin:beats_management@6.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
  log   [13:46:36.792] [info][status][plugin:apm@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.147] [info][status][plugin:elasticsearch@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.345] [info][status][plugin:canvas@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.361] [info][status][plugin:console@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.366] [info][status][plugin:console_extensions@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.374] [info][status][plugin:notifications@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.439] [info][status][plugin:infra@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.447] [info][status][plugin:metrics@6.5.1] Status changed from uninitialized to green - Ready
  log   [13:46:37.466] [info][license][xpack] Imported license information from Elasticsearch for the [data] cluster: mode: basic | status: active
  log   [13:46:37.471] [info][status][plugin:xpack_main@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.472] [info][status][plugin:searchprofiler@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.473] [info][status][plugin:ml@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.474] [info][status][plugin:tilemap@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.475] [info][status][plugin:watcher@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.476] [info][status][plugin:index_management@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.477] [info][status][plugin:rollup@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.478] [info][status][plugin:graph@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.480] [info][status][plugin:grokdebugger@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.481] [info][status][plugin:logstash@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.481] [info][status][plugin:beats_management@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:37.482] [info][kibana-monitoring][monitoring-ui] Starting monitoring stats collection
  log   [13:46:37.494] [info][status][plugin:security@6.5.1] Status changed from yellow to green - Ready
  log   [13:46:38.065] [warning][reporting] Generating a random key for xpack.reporting.encryptionKey. To prevent pending reports from failing on restart, please set xpack.reporting.encryptionKey in kibana.yml
  log   [13:46:38.074] [info][status][plugin:reporting@6.5.1] Status changed from uninitialized to green - Ready
```

Setting up the spark job from the streaming lesson to publish log events into elastic.

First, we need to create a kafka topic that we will be writing into, mine is called “next779”, here’s how to create it:

```
./kafka-topics.sh --create --zookeeper sandbox-hdp.hortonworks.com:2181 --replication-factor 1
--partitions 1 --topic next779
```

Next, we have to publish events into this topic, I downloaded the “test.csv” dataset and published some of its data into kafka:

```
./spark-submit --driver-memory 550m --num-executors 4 --executor-memory 550m /home/maria_dev/producer.jar
-topic next779 -url sandbox-hdp.hortonworks.com:6667 -filePath /home/maria_dev/test.csv -nThreads 4
```

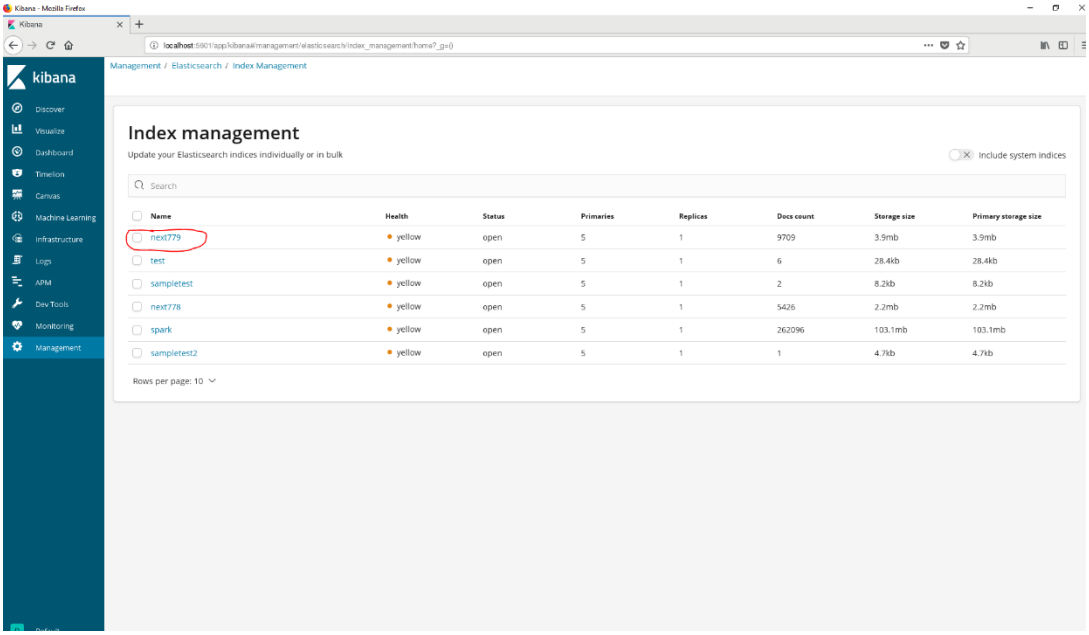
Here’s the execution trace:

```
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9575, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9582, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9588, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9573, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9578, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9580, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9584, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9585, Partition : 0
18/11/29 14:42:59 INFO MessageCallback$: Topic : next779, Offset : 9577, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9589, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9592, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9596, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9600, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9607, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9587, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9590, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9594, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9593, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9595, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9597, Partition : 0
18/11/29 14:43:02 INFO MessageCallback$: Topic : next779, Offset : 9602, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9610, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9608, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9612, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9616, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9609, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9618, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9613, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9619, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9615, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9617, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9601, Partition : 0
18/11/29 14:43:03 INFO MessageCallback$: Topic : next779, Offset : 9591, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9603, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9620, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9623, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9605, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9625, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9627, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9611, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9598, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9629, Partition : 0
18/11/29 14:43:05 INFO MessageCallback$: Topic : next779, Offset : 9614, Partition : 0
```

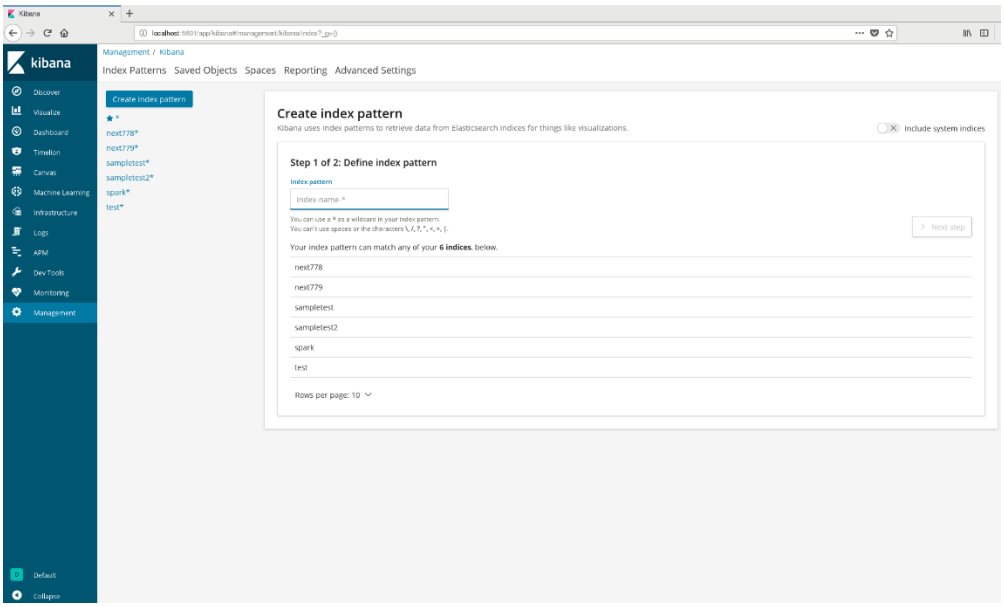
Once we are done publishing events into kafka, we can consume them from kafka into elastic. Like so:

```
./spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.3.2 --master yarn-client
--driver-memory 550m --num-executors 4 --executor-memory 550m /home/maria_dev/consumer.jar -topic next779
-url sandbox-hdp.hortonworks.com:6667 -filePath next779/test_csv -fileFormat org.elasticsearch.spark.sql
-doBatch
false
```

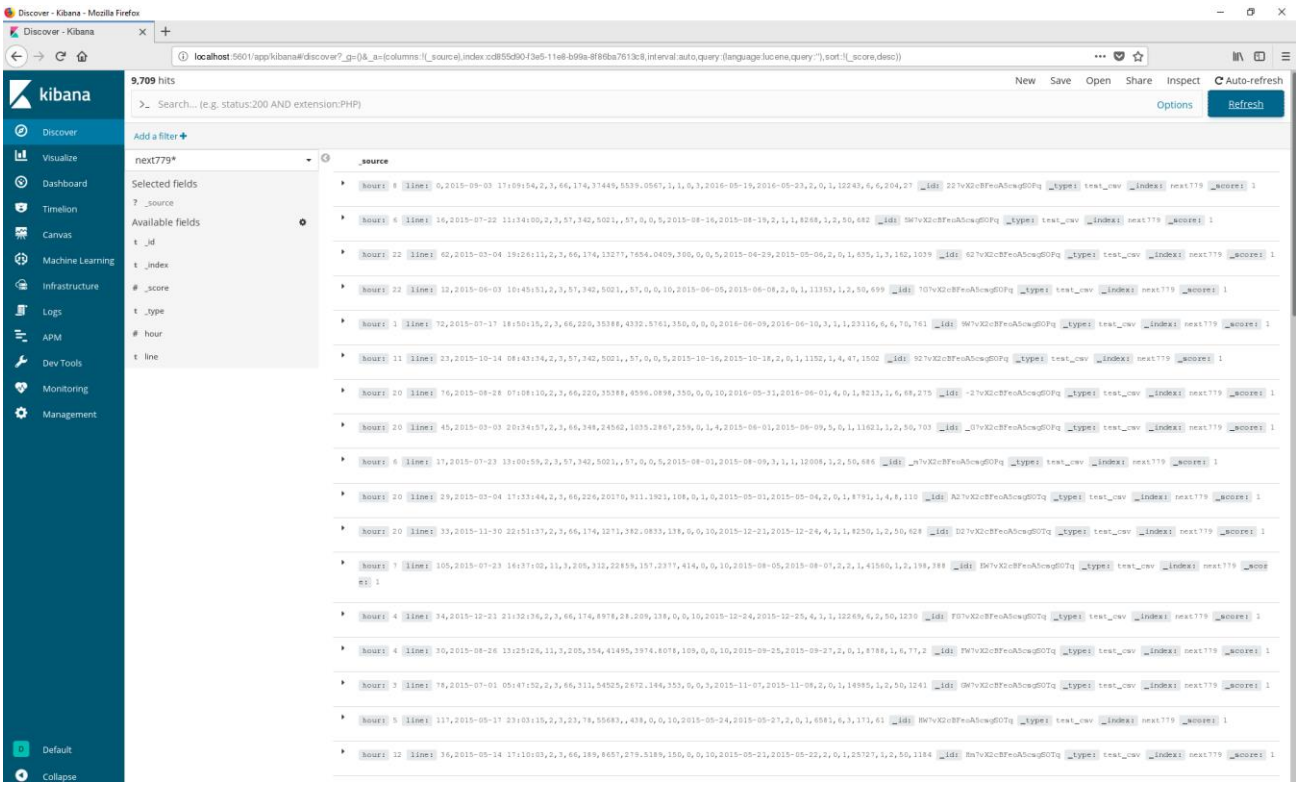
The -filePath argument is in the format “index/type”, so, once this command completes, we will see the “next779” index:



After we uploaded our data into elastic, we need to create an index pattern for kibana, in order to be able to use that data in kibana dashboard:



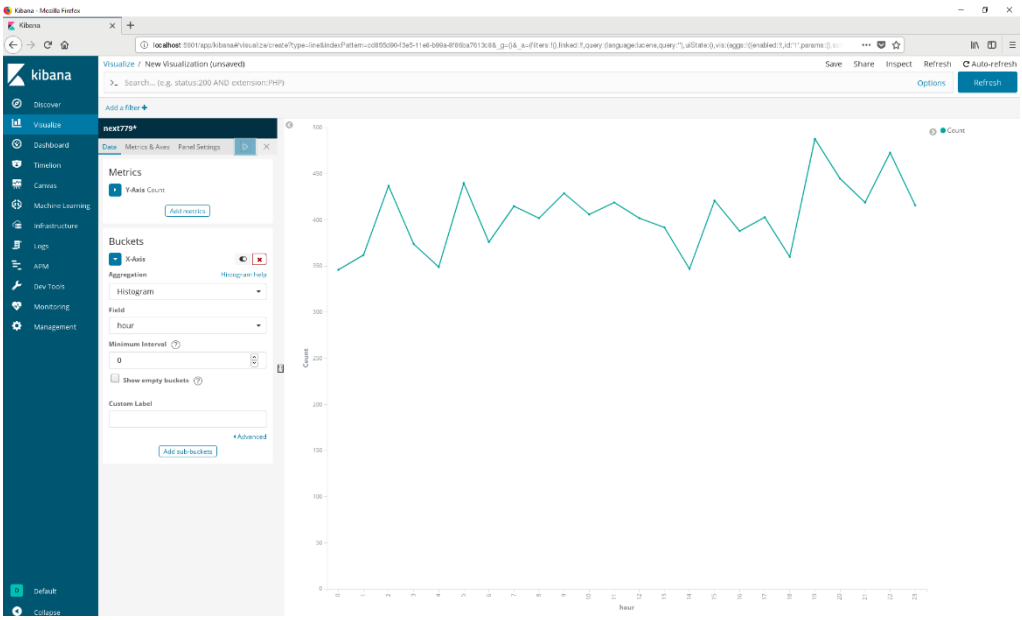
Once created, our data can be viewed in the discover tab:



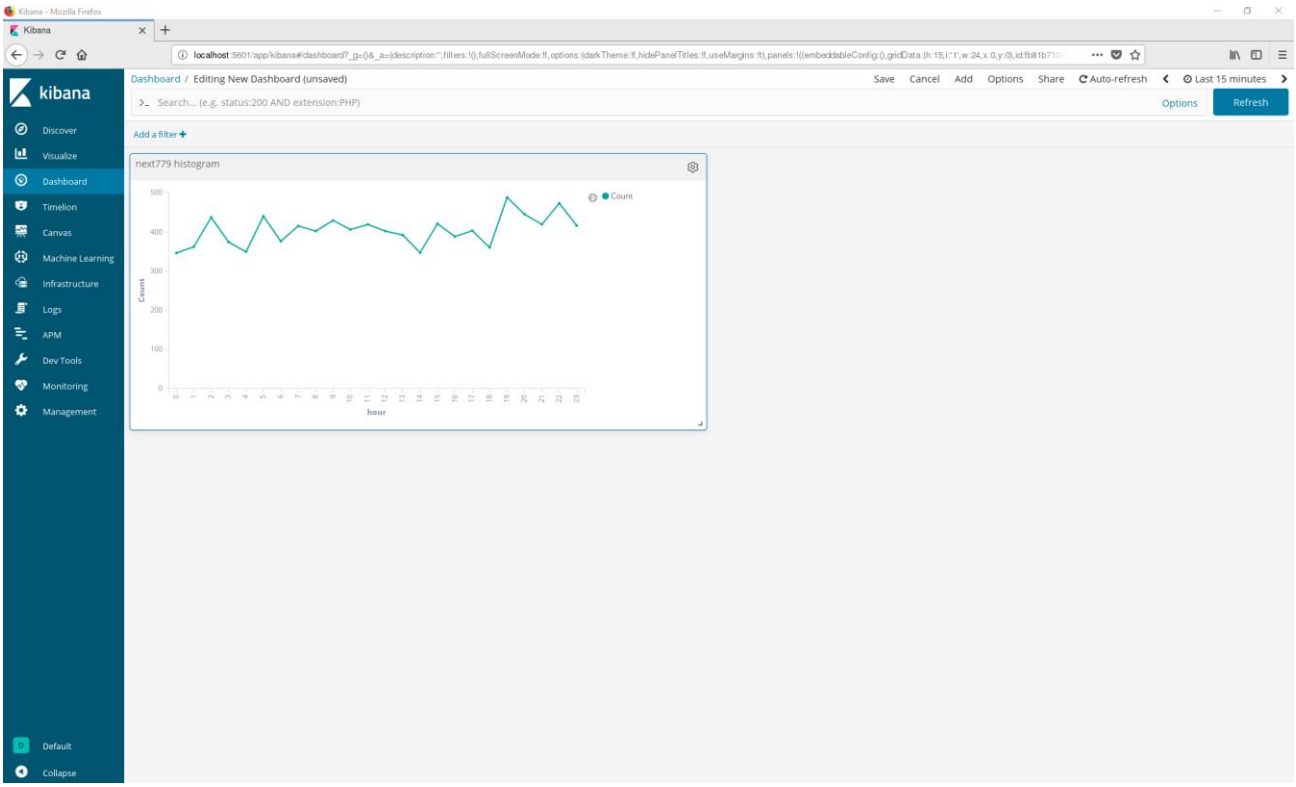
As we can see, we have our data in the proper format. I used the ThreadLocalRandom class in scala to generate random hours that we can use to analyze our “pseudo-data”.

Running the kibana dashboard.

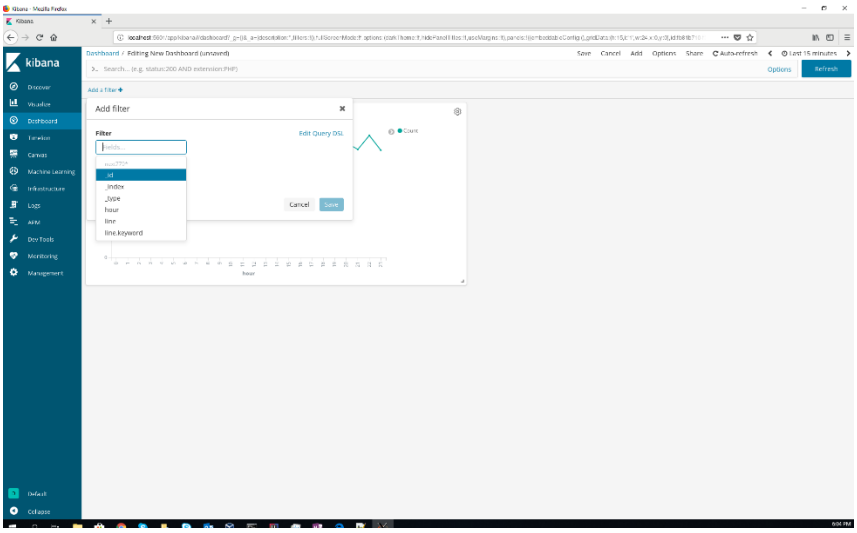
In the visualization tab we can create different kinds of charts to visualize our data, here's one example:



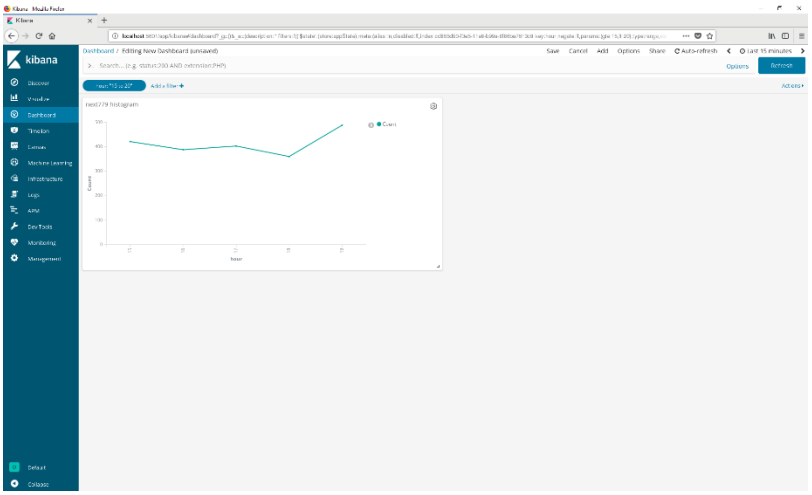
I used the number of logs as the Y-axis, and the X-Axis represents hours. So, each log event belongs to its own hour bucket. The kibana dashboard can be decorated with different kinds of such charts, here's one chart:



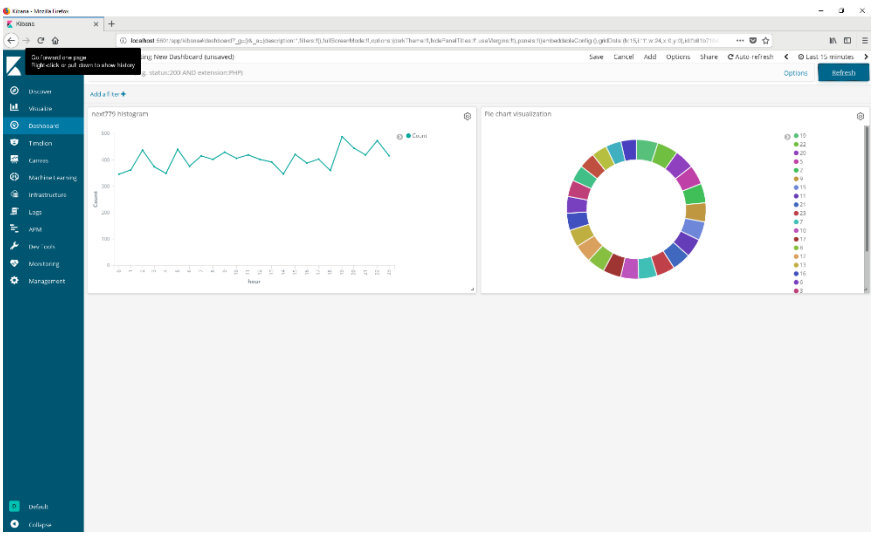
You can add filters in the dashboard to visualize only the subset of your data:



This is what it looks like:



You can also add more visualizations:



And here’s the same visualizations with filters applied: