CSDA 1020 - Big Data Analytics Tools
Project 3: ELK (Elasticsearch, Logstash, Kibana)

Prepared by: Rani Lottey

1.0 Business Problem

In order to provide an analysis on a potential investment in a used car business a dataset called cars.csv, provided by the company, will be used. The following analytics tools, Elasticsearch, Logstash and Kibana, will be installed and configured in a GCP platform in this project. Configuring the Logstash_cars.config file, starting Logstash to ingest the cars.csv data into Elasticsearch and using Kibana to show some initial analysis and visualizations will be the taken away for this project.

2.0 Elasticsearch, Logstash and Kibana (ELK) Tools

Using the GCP platform a single-node cluster and instance named "bigdata" were created to launch and use the ELK tools for analysis for the dataset as shown below in Figure 1.

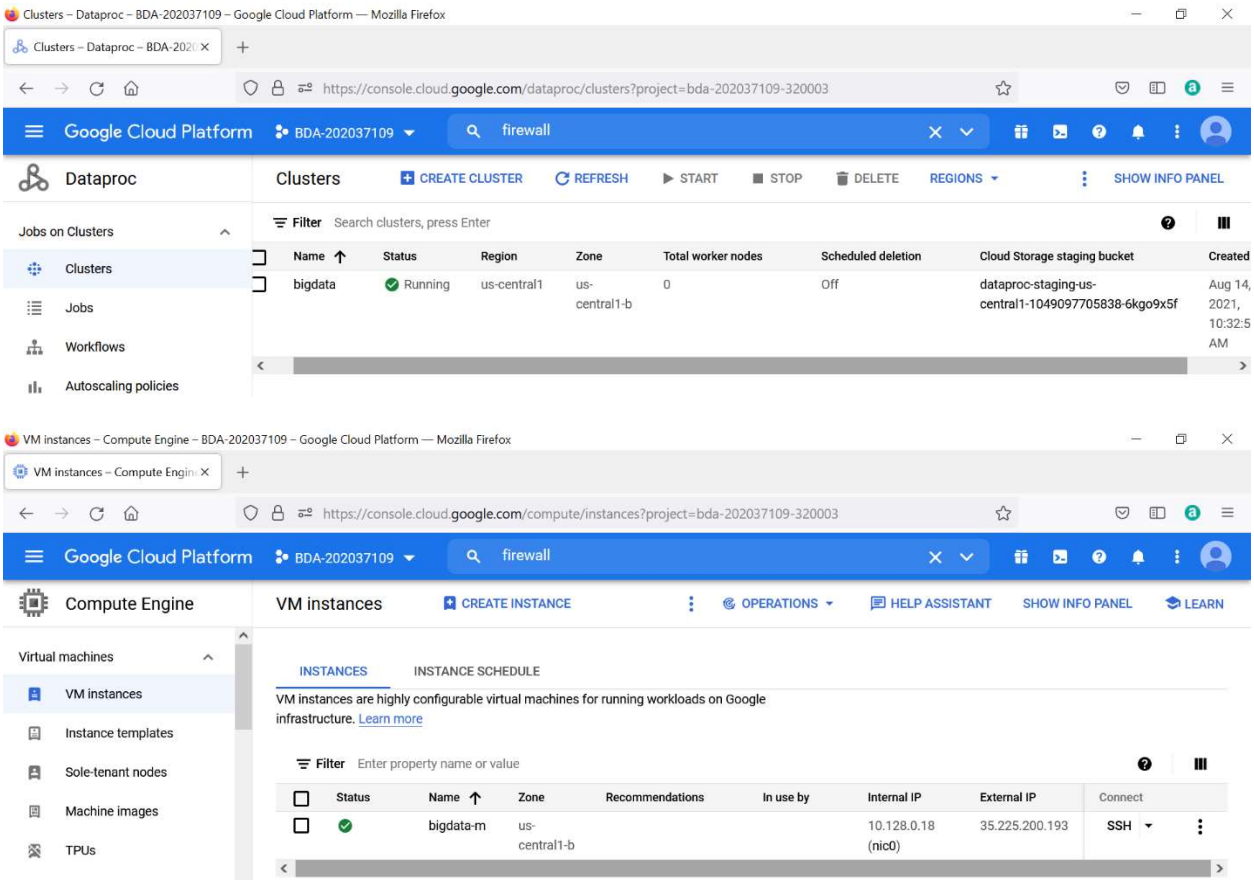


Figure 1: Cluster and Instance Set-up

Following this Elasticsearch, Logstash and Kibana were downloaded into the instance for use as shown in Figure 2.

```
krlottey@bigdata-m: ~/elasticsearch-7.5.1 — Mozilla Firefox
https://ssh.cloud.google.com/projects/bda-202037109-320003/zones/us-central1-b/instances/bigc
Connected, host fingerprint: ssh-rsa 0 5B:F5:D3:8A:D0:3A:2A:B6:B1:EB:21:2E:F8:5D
:D7:E7:D1:DF:91:CA:2B:A1:BC:EA:A3:53:54:BB:5F:43:A2:42
Linux bigdata-m 5.10.0-0.bpo.8-amd64 #1 SMP Debian 5.10.46-2~bpo10+1 (2021-07-22) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
krlottey@bigdata-m:~$ wget https://artifacts.elastic.co/downloads/elasticsearch/e
lasticsearch-7.5.1-linux-x86_64.tar.gz
--2021-08-14 14:54:01--  https://artifacts.elastic.co/downloads/elasticsearch/ela
sticsearch-7.5.1-linux-x86_64.tar.gz
Resolving artifacts.elastic.co (artifacts.elastic.co)... 34.120.127.130, 2600:190
1:0:1d7::
Connecting to artifacts.elastic.co (artifacts.elastic.co)|34.120.127.130|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 290094012 (277M) [application/x-gzip]
Saving to: 'elasticsearch-7.5.1-linux-x86_64.tar.gz'

elasticsearch-7.5.1- 100%[===================>] 276.65M  47.5MB/s    in 6.5s

2021-08-14 14:54:07 (42.7 MB/s) - 'elasticsearch-7.5.1-linux-x86_64.tar.gz' saved
 [290094012/290094012]

krlottey@bigdata-m:~$ wget https://artifacts.elastic.co/downloads/kibana/kibana-7
.5.1-linux-x86_64.tar.gz
--2021-08-14 14:55:09--  https://artifacts.elastic.co/downloads/kibana/kibana-7.5
.1-linux-x86_64.tar.gz
Resolving artifacts.elastic.co (artifacts.elastic.co)... 34.120.127.130, 2600:190
1:0:1d7::
Connecting to artifacts.elastic.co (artifacts.elastic.co)|34.120.127.130|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 238481011 (227M) [application/x-gzip]
Saving to: 'kibana-7.5.1-linux-x86_64.tar.gz'
# The primary way of configuring a node is via this file. This template lists

#
kibana-7.5.1-linux-x 100%[===================>] 227.43M  21.4MB/s    in 11s

2021-08-14 14:55:20 (20.0 MB/s) - 'kibana-7.5.1-linux-x86_64.tar.gz' saved [23848
1011/238481011]

krlottey@bigdata-m:~$ wget https://artifacts.elastic.co/downloads/logstash/logsta
sh-7.5.1.tar.gz
--2021-08-14 14:55:33--  https://artifacts.elastic.co/downloads/logstash/logstash
-7.5.1.tar.gz
Resolving artifacts.elastic.co (artifacts.elastic.co)... 34.120.127.130, 2600:190
1:0:1d7::
Connecting to artifacts.elastic.co (artifacts.elastic.co)|34.120.127.130|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 165760774 (158M) [application/x-gzip]
Saving to: 'logstash-7.5.1.tar.gz'

logstash-7.5.1.tar.g 100%[===================>] 158.08M  18.0MB/s    in 11s

2021-08-14 14:55:44 (14.7 MB/s) - 'logstash-7.5.1.tar.gz' saved [165760774/165760
774]

krlottey@bigdata-m:~$ ls
elasticsearch-7.5.1-linux-x86_64.tar.gz  logstash-7.5.1.tar.gz
kibana-7.5.1-linux-x86_64.tar.gz
krlottey@bigdata-m:~$ sudo sysctl vm.max_map_count=262144
vm.max_map_count = 262144
```

Confirmed loaded

Figure 2: Loading ELK Tools

2.1 Configuration of Elasticsearch

The elasticsearch.yml file in Elasticsearch was configured using the vi editor as shown in Figure 3 and then the service was started as shown in Figure 3.

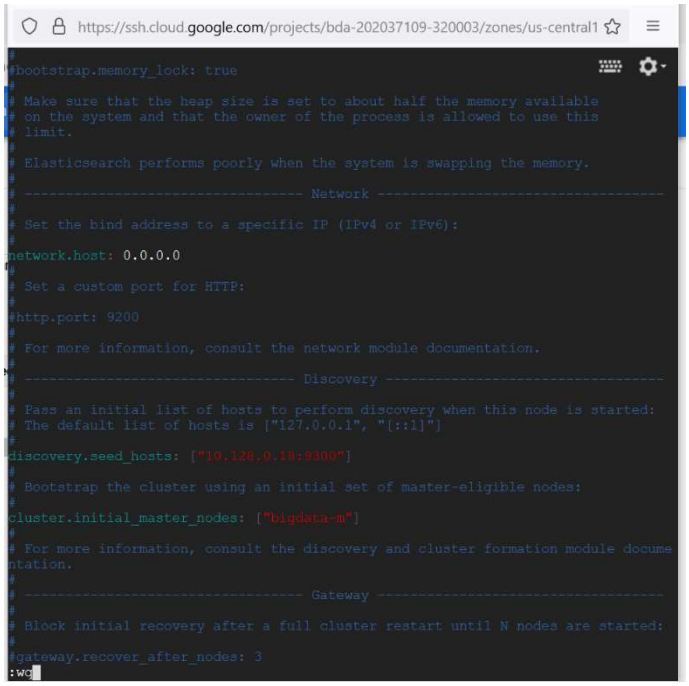
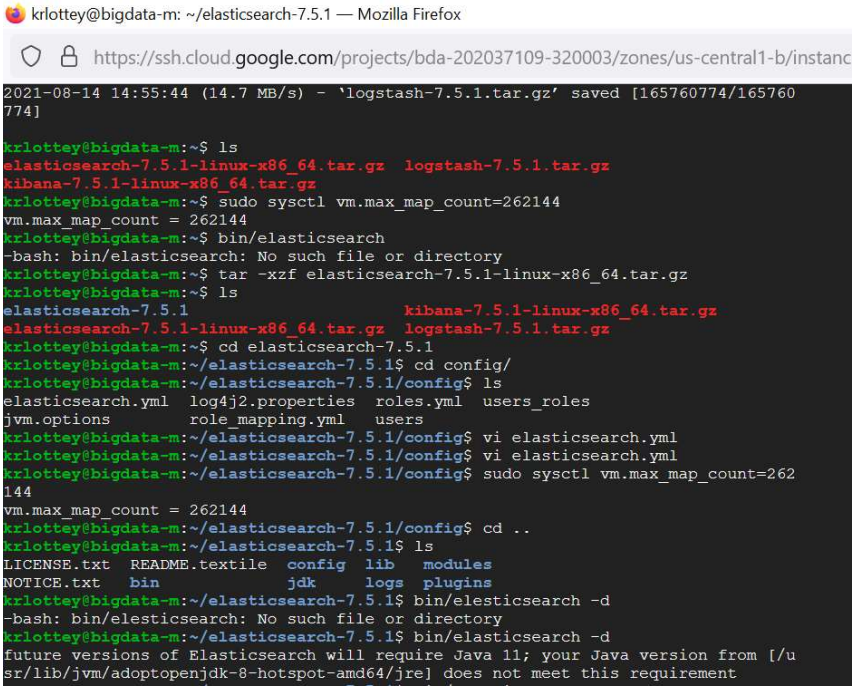


Figure 3: Vi Editor- elasticsearch.yml



Confirmed started

Figure 4: Starting Elasticsearch Services

2.2 Configuration of Kibana

Similar to Elasticsearch, the same steps were taken to start the Kibana services. Editor vi as used to configured and then the Kibana services was started as confimed in Figure 5.

```
krlottey@bigdata-m: ~/kibana-7.5.1-linux-x86_64 — Mozilla Firefox
https://ssh.cloud.google.com/projects/bda-202037109-320003/zones/us-central1-b/inst
logstash-7.5.1.tar.gz
krlottey@bigdata-m:~$ tar -xzf kibana-7.5.1-linux-x86_64.tar.gz
krlottey@bigdata-m:~$ ls
elasticsearch-7.5.1
elasticsearch-7.5.1-linux-x86_64.tar.gz
kibana-7.5.1-linux-x86_64
kibana-7.5.1-linux-x86_64.tar.gz
logstash-7.5.1.tar.gz
krlottey@bigdata-m:~$ cd kibana-7.5.1-linux-x86_64
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64$ ls
LICENSE.txt  built_assets  node_modules  src
NOTICE.txt   config        optimize      webpackShims
README.txt   data          package.json  x-pack
bin          node          plugins
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64$ cd config/
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ bin/kibana
-bash: bin/kibana: No such file or directory
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ cd ..
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64$ ls
LICENSE.txt  bin           data          optimize      src
NOTICE.txt   built_assets  node          package.json  webpackShims
README.txt   config        node_modules  plugins       x-pack
krlottey@bigdata-m:~/kibana-7.5.1-linux-x86_64$ bin/kibana
  log   [15:49:15.921] [info][plugins-system] Setting up [15] plugins: [
licensing,security,code,timelion,features,spaces,translations,uiActions,
inspector,embeddable,advancedUiActions,expressions,newsfeed,eui_utils,da
ta]
  log   [15:49:15.928] [info][licensing][plugins] Setting up plugin
  log   [15:49:15.930] [info][plugins][security] Setting up plugin
  log   [15:49:15.932] [warning][config][plugins][security] Generating a
 random key for xpack.security.encryptionKey. To prevent sessions from b
eing invalidated on restart, please set xpack.security.encryptionKey in
kibana.yml
  log   [15:49:15.932] [warning][config][plugins][security] Session cook
ies will be transmitted over insecure connections. This is not recommend
ed.
  log   [15:49:15.973] [info][code][plugins] Setting up plugin
  log   [15:49:15.974] [info][plugins][timelion] Setting up plugin
  log   [15:49:15.975] [info][features][plugins] Setting up plugin
```

Configuration

Service started

Figure 5: Starting of Kibana Services

2.3 Firewall Configuration

The firewall need to be configured with new rules to allow access to the ports for Elasticsearch and Kibanaso that they could be used. Figure 6 shows these new firewall rules called elasticsearch and kibana.

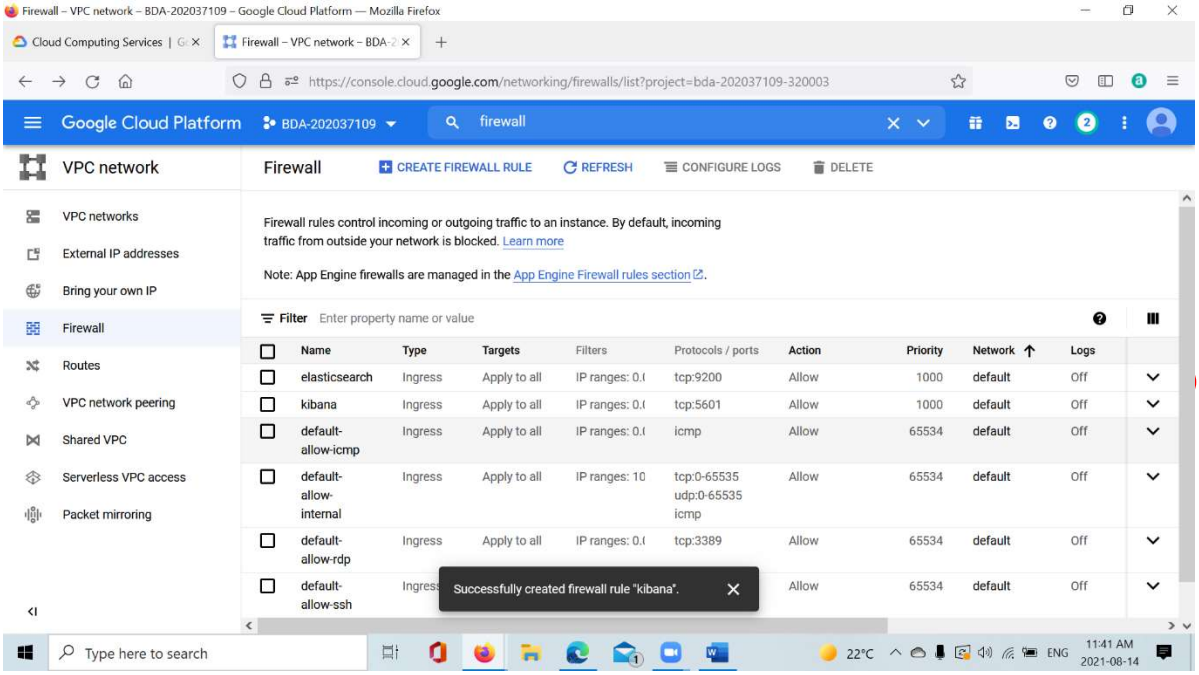


Figure 6: New Firewall Rules

2.4 Configuration of Logstash

The cars.csv dataset was loaded into the database as shown in Figure 7. Along with that the logstash service was started also shown in Figure 8.

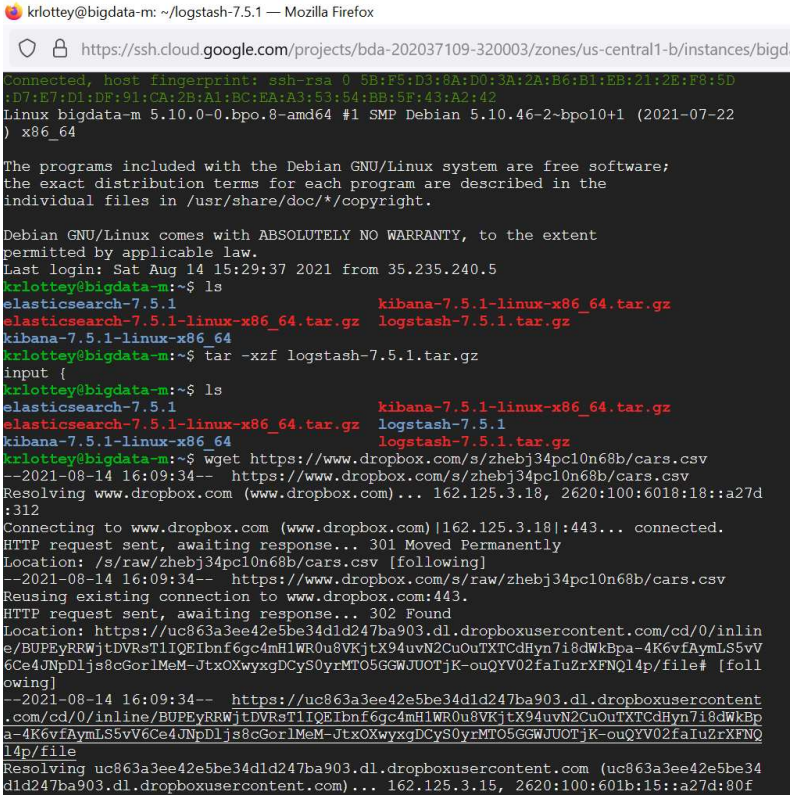


Figure 7: Loading of Dataset cars.csv

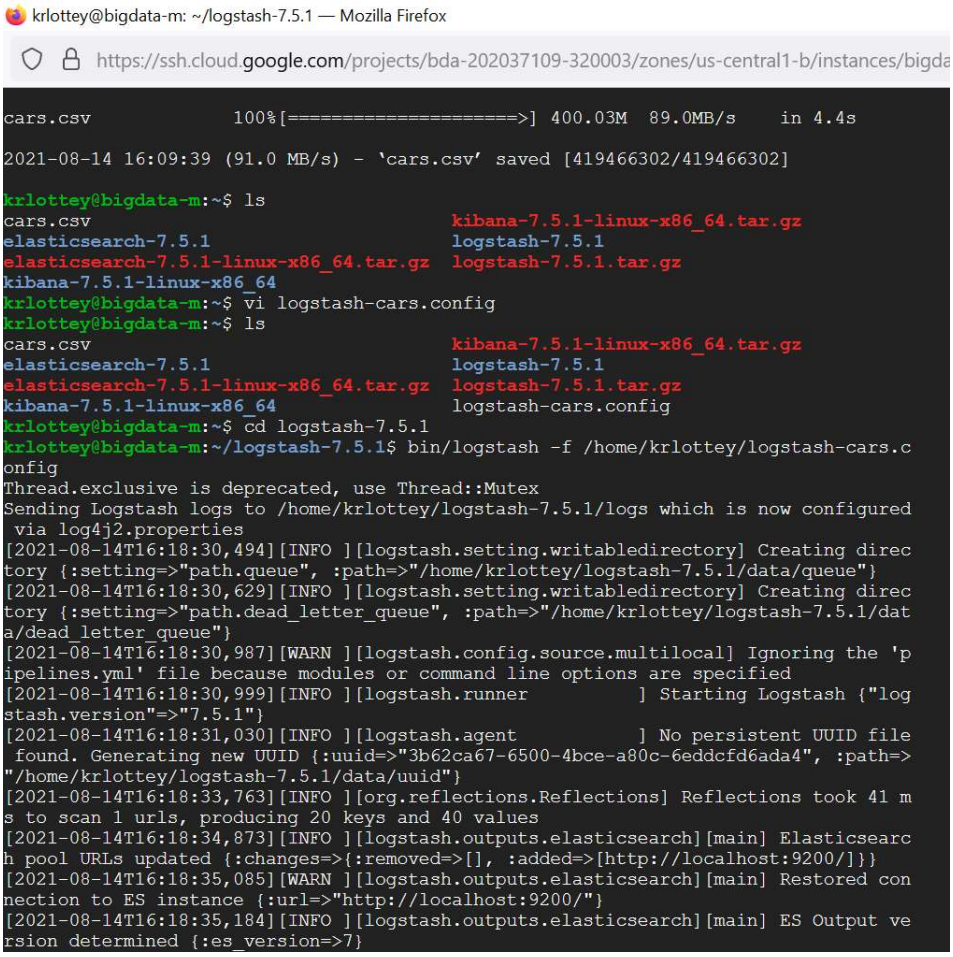


Figure 8: Starting of Logstash Services

3.0 Using Kibana Services

Once all the services were started and the dataset was loaded the Kibana interface was initated by using the extenal IP address from the VM instance and adding :5601 to the IP address. This modified IP address was opened in Google to interact with dataset and conduct some analysis/visualization of the dataset as shown in Figure 9.

Some examples of the visualization that can be conducted are shown in Figures 10 and 11.

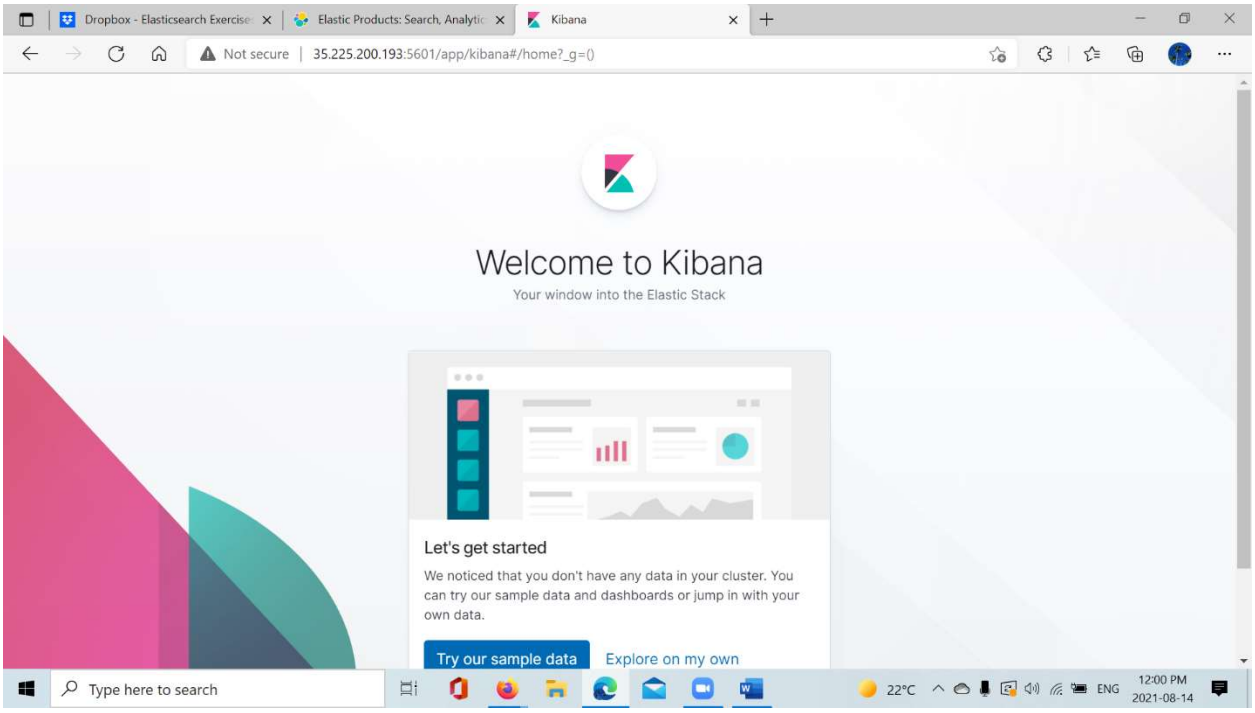


Figure 9: Modified IP Address To Open Kibana in Google

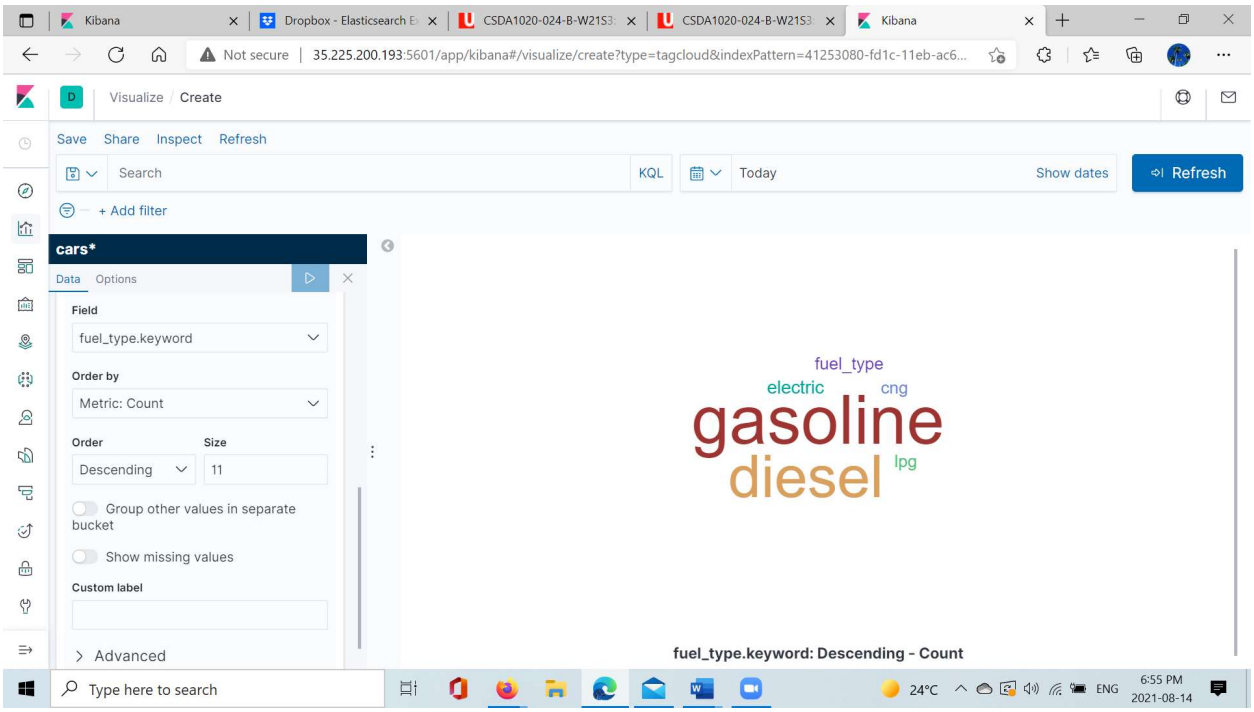


Figure 10: Example of a Tag Cloud Visualization for Fuel_Type from cars.csv Dataset

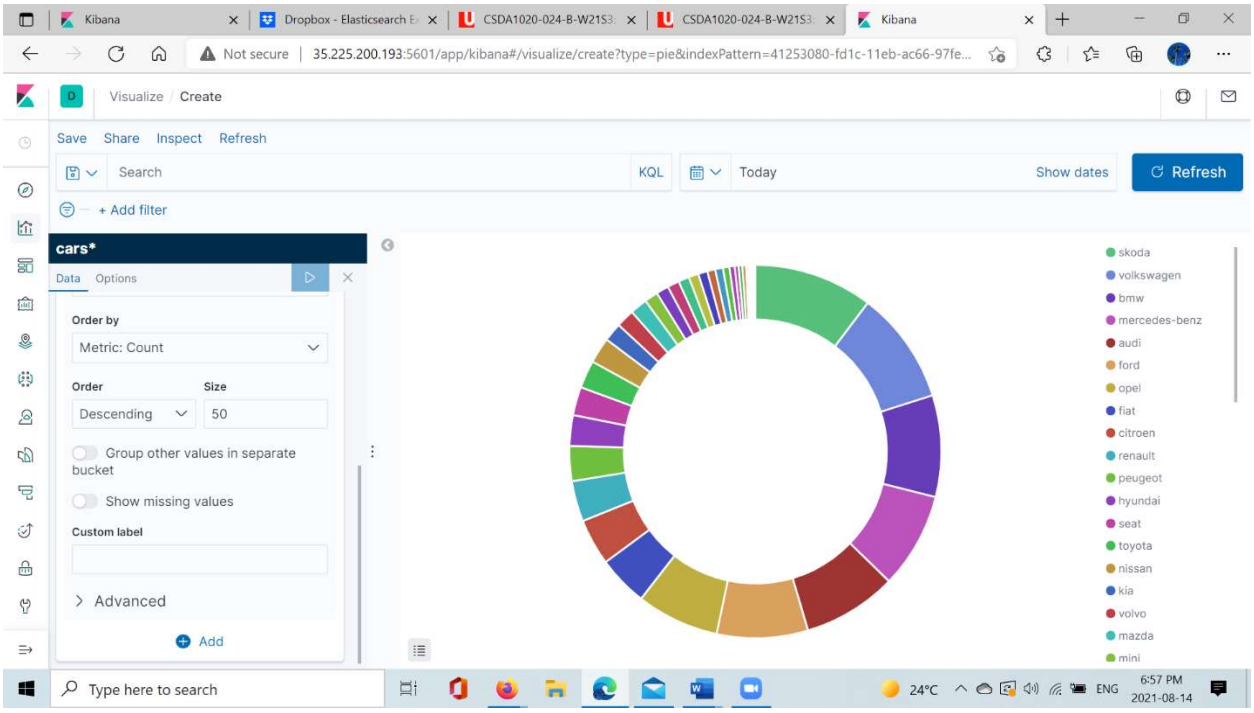


Figure 11: Example of a Pie Chart from the Data cars.csv for Maker Types