

# 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.

The figure consists of two screenshots from the Google Cloud Platform console. The top screenshot shows the 'Clusters' page in the Dataproc section. A single cluster named 'bigdata' is listed with a status of 'Running'. The bottom screenshot shows the 'VM instances' page in the Compute Engine section. A single VM instance named 'bigdata-m' is listed with a status of 'Running'.

Name	Status	Region	Zone	Total worker nodes	Scheduled deletion	Cloud Storage staging bucket	Created
bigdata	Running	us-central1	us-central1-b	0	Off	dataproc-staging-us-central1-1049097705838-6kgo9x5f	Aug 14, 2021, 10:32:5 AM

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
Running	bigdata-m	us-central1-b			10.128.0.18 (nic0)	35.225.200.193	SSH

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/bigdata-m

Connected, host fingerprint: ssh-rsa 0 5B:FS: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:SF:42: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/elasticsearch-7.5.1-linux-x86_64.tar.gz
--2021-08-14 14:54:01-- https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.5.1-linux-x86_64.tar.gz
Resolving artifacts.elastic.co (artifacts.elastic.co)... 34.120.127.130, 2600:1901:0:1d7::
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, 2600:1901:0:1d7::
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 [238481011/238481011]

krlottey@bigdata-m:~$ wget https://artifacts.elastic.co/downloads/logstash/logstash-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, 2600:1901:0:1d7::
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 [165760774/165760774]

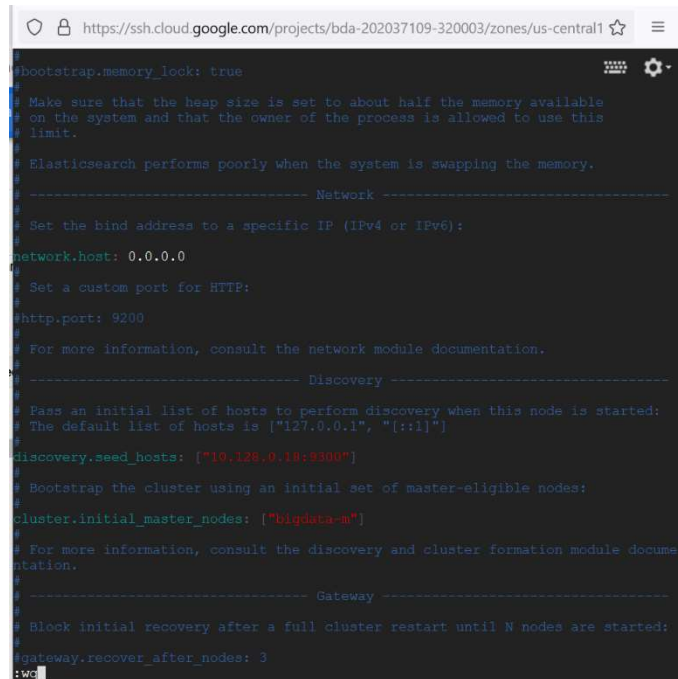
krlottey@bigdata-m:~$ ls
elasticsearch-7.5.1-linux-x86_64.tar.gz  logstash-7.5.1.tar.gz
kibana-7.5.1-linux-x86_64.tar.gz
krlottey@bigdata-m:~$ sudo sysctl vm.max_map_count=262144
vm.max_map_count = 262144
```

← Confirmed loaded

Figure 2: Loading ELK Tools

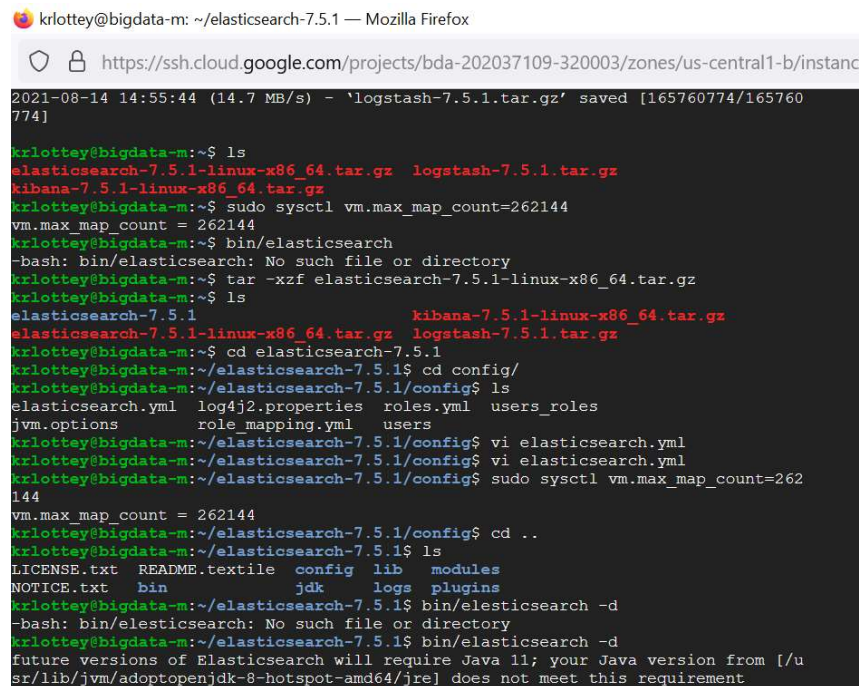
## 2.1 Configuration of Elasticsearch

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.



```
#bootstrap.memory_lock: true
#
# Make sure that the heap size is set to about half the memory available
# on the system and that the owner of the process is allowed to use this
# limit.
#
# Elasticsearch performs poorly when the system is swapping the memory.
#
#----- Network -----
#
# Set the bind address to a specific IP (IPv4 or IPv6):
#
#network.host: 0.0.0.0
#
# Set a custom port for HTTP:
#
#http.port: 9200
#
# For more information, consult the network module documentation.
#
#----- Discovery -----
#
# Pass an initial list of hosts to perform discovery when this node is started:
# The default list of hosts is ["127.0.0.1", "::1"]
#
#discovery.seed_hosts: ["10.128.0.10:9300"]
#
# Bootstrap the cluster using an initial set of master-eligible nodes:
#
#cluster.initial_master_nodes: ["bigdata-m"]
#
# For more information, consult the discovery and cluster formation module documentation.
#
#----- Gateway -----
#
# Block initial recovery after a full cluster restart until N nodes are started:
#
#gateway.recover_after_nodes: 3
```

Figure 3: Vi Editor- elasticsearch.yml



```
krloTTY@bigdata-m: ~/elasticsearch-7.5.1 — Mozilla Firefox
https://ssh.cloud.google.com/projects/bda-202037109-320003/zones/us-central1-b/instance
2021-08-14 14:55:44 (14.7 MB/s) - 'logstash-7.5.1.tar.gz' saved [165760774/165760774]

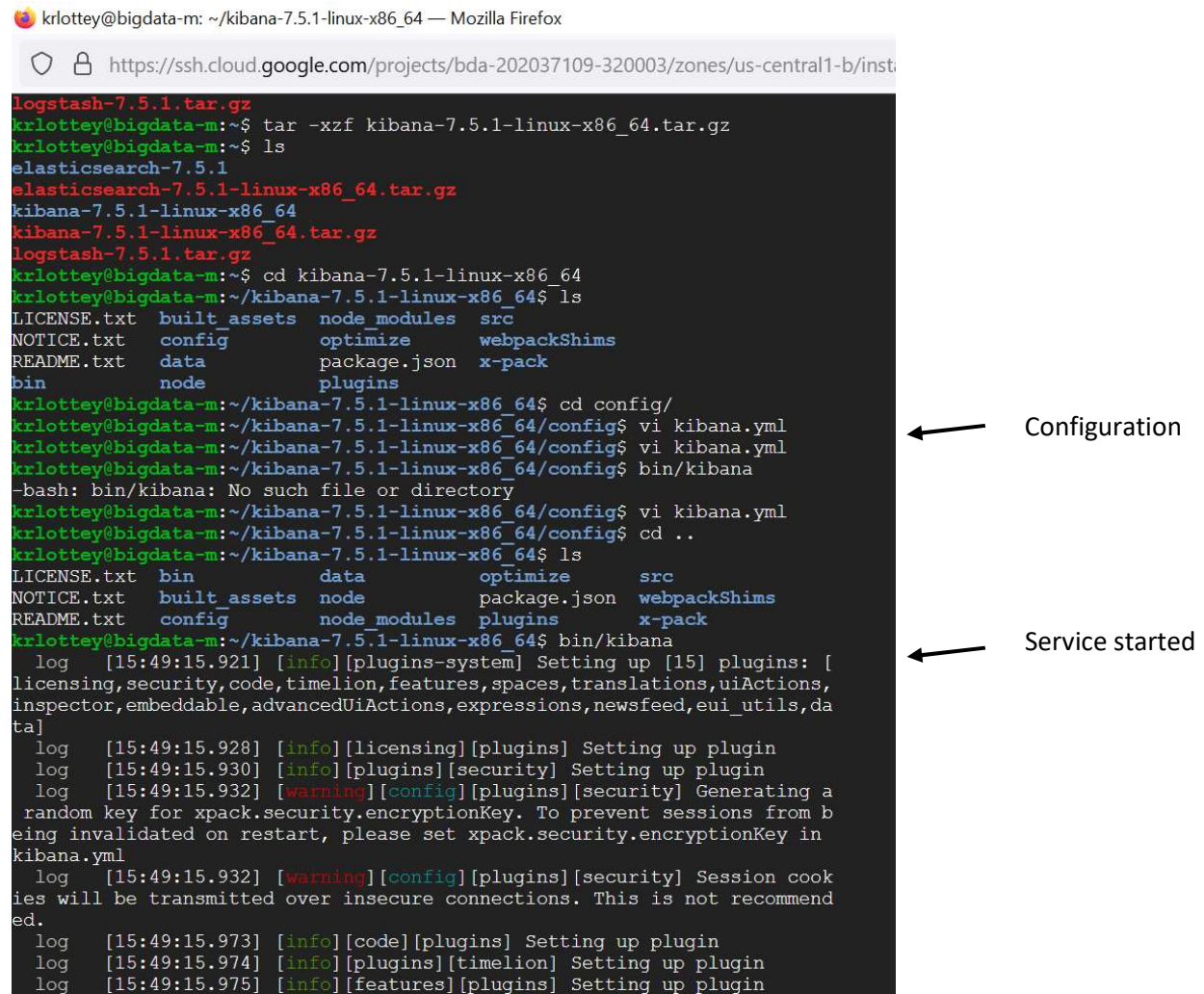
krloTTY@bigdata-m:~$ ls
elasticsearch-7.5.1-linux-x86_64.tar.gz  logstash-7.5.1.tar.gz
kibana-7.5.1-linux-x86_64.tar.gz
krloTTY@bigdata-m:~$ sudo sysctl vm.max_map_count=262144
vm.max_map_count = 262144
krloTTY@bigdata-m:~$ bin/elasticsearch
-bash: bin/elasticsearch: No such file or directory
krloTTY@bigdata-m:~$ tar -xzf elasticsearch-7.5.1-linux-x86_64.tar.gz
krloTTY@bigdata-m:~$ ls
elasticsearch-7.5.1  kibana-7.5.1-linux-x86_64.tar.gz  logstash-7.5.1.tar.gz
krloTTY@bigdata-m:~$ cd elasticsearch-7.5.1
krloTTY@bigdata-m:~/elasticsearch-7.5.1$ cd config/
krloTTY@bigdata-m:~/elasticsearch-7.5.1/config$ ls
elasticsearch.yml  log4j2.properties  roles.yml  users_roles
jvm.options        role_mapping.yml    users
krloTTY@bigdata-m:~/elasticsearch-7.5.1/config$ vi elasticsearch.yml
krloTTY@bigdata-m:~/elasticsearch-7.5.1/config$ vi elasticsearch.yml
krloTTY@bigdata-m:~/elasticsearch-7.5.1/config$ sudo sysctl vm.max_map_count=262144
vm.max_map_count = 262144
krloTTY@bigdata-m:~/elasticsearch-7.5.1/config$ cd ..
krloTTY@bigdata-m:~/elasticsearch-7.5.1$ ls
LICENSE.txt  README.textile  config  lib  modules
NOTICE.txt  bin  jdk  logs  plugins
krloTTY@bigdata-m:~/elasticsearch-7.5.1$ bin/elasticsearch -d
-bash: bin/elasticsearch: No such file or directory
krloTTY@bigdata-m:~/elasticsearch-7.5.1$ bin/elasticsearch -d
future versions of Elasticsearch will require Java 11; your Java version from [/u
s/lib/jvm/adoptopenjdk-8-hotspot-amd64/jre] does not meet this requirement
```

← 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 confined in Figure 5.



```
krлоттеу@bigdata-m: ~/kibana-7.5.1-linux-x86_64 — Mozilla Firefox
https://ssh.cloud.google.com/projects/bda-202037109-320003/zones/us-central1-b/instr

logstash-7.5.1.tar.gz
krлоттеу@bigdata-m:~$ tar -xzf kibana-7.5.1-linux-x86_64.tar.gz
krлоттеу@bigdata-m:~$ ls
elasticsearch-7.5.1
kibana-7.5.1-linux-x86_64
kibana-7.5.1-linux-x86_64.tar.gz
logstash-7.5.1.tar.gz
krлоттеу@bigdata-m:~$ cd kibana-7.5.1-linux-x86_64
krлоттеу@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
krлоттеу@bigdata-m:~/kibana-7.5.1-linux-x86_64$ cd config/
krлоттеу@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
krлоттеу@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
krлоттеу@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ bin/kibana
-bash: bin/kibana: No such file or directory
krлоттеу@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
krлоттеу@bigdata-m:~/kibana-7.5.1-linux-x86_64/config$ cd ..
krлоттеу@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
krлоттеу@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 Kibana so that they could be used. Figure 6 shows these new firewall rules called elasticsearch and kibana.



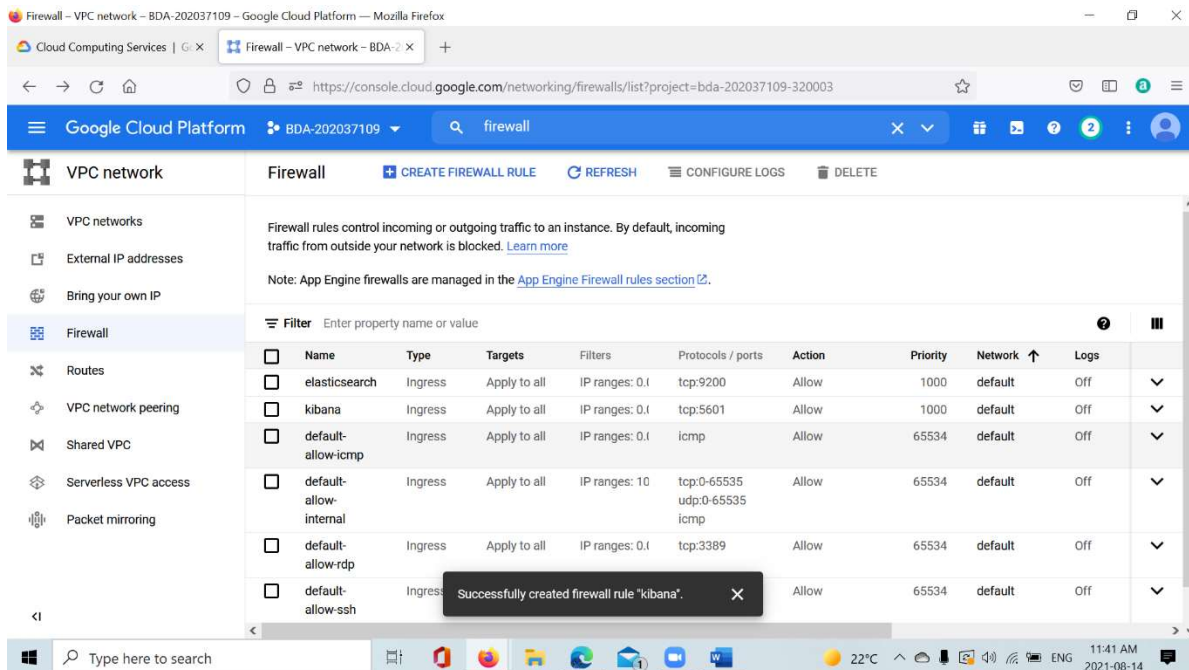


Figure 6: New Firewall Rules

## 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.

```

krlothey@bigdata-m: ~/logstash-7.5.1 — Mozilla Firefox
https://ssh.cloud.google.com/projects/bda-202037109-320003/zones/us-central1-b/instances/bigdata-m

connected, host fingerprint: ssh-rsa 0 5B:F5:D3:9A:D0:3A:2A:B6:B1:EB:21:2E:F8:5D
:D7:E7:D1:DF:91:CA:2B:A1:BC:EA:A3:53:54:BB:5E: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.
Last login: Sat Aug 14 15:29:37 2021 from 35.235.240.5
krlothey@bigdata-m:~$ ls
elasticsearch-7.5.1 kibana-7.5.1-linux-x86_64.tar.gz
elasticsearch-7.5.1-linux-x86_64.tar.gz logstash-7.5.1.tar.gz
kibana-7.5.1-linux-x86_64
krlothey@bigdata-m:~$ tar -xzf logstash-7.5.1.tar.gz
input {
krlothey@bigdata-m:~$ ls
elasticsearch-7.5.1 kibana-7.5.1-linux-x86_64.tar.gz
elasticsearch-7.5.1-linux-x86_64.tar.gz logstash-7.5.1
kibana-7.5.1-linux-x86_64 logstash-7.5.1.tar.gz
krlothey@bigdata-m:~$ wget https://www.dropbox.com/s/zhebj34pc10n68b/cars.csv
--2021-08-14 16:09:34-- https://www.dropbox.com/s/zhebj34pc10n68b/cars.csv
Resolving www.dropbox.com (www.dropbox.com)... 162.125.3.18, 2620:100:6018:18::a27d
:312
Connecting to www.dropbox.com (www.dropbox.com)|162.125.3.18|:443... connected.
HTTP request sent, awaiting response... 301 Moved Permanently
Location: /s/raw/zhebj34pc10n68b/cars.csv [following]
--2021-08-14 16:09:34-- https://www.dropbox.com/s/raw/zhebj34pc10n68b/cars.csv
Reusing existing connection to www.dropbox.com:443.
HTTP request sent, awaiting response... 302 Found
Location: https://uc863a3ee42e5be34d1d247ba903.dl.dropboxusercontent.com/cd/0/inlin
e/BUPEyRRWjtDVRsTlIQEIbnf6gc4mH1WR0u8VKjtX94uvN2CuOuTXTCdHyn7i8dWkBP
a-4K6vfAymLS5vV6Ce4JNpDlj8cGorlMeM-JtxOXwyxgDCyS0yrMT05GGWJUOTjK-ouQYV02faIuZrXFNQ
l4p/file
Resolving uc863a3ee42e5be34d1d247ba903.dl.dropboxusercontent.com (uc863a3ee42e5be34
d1d247ba903.dl.dropboxusercontent.com)... 162.125.3.15, 2620:100:601b:15::a27d:80f

```

Figure 7: Loading of Dataset cars.csv

```
krлоттеу@bigdata-m: ~/logstash-7.5.1 — Mozilla Firefox
https://ssh.cloud.google.com/projects/bda-202037109-320003/zones/us-central1-b/instances/bigda

cars.csv          100%[=====>] 400.03M  89.0MB/s   in 4.4s
2021-08-14 16:09:39 (91.0 MB/s) - 'cars.csv' saved [419466302/419466302]

krлоттеу@bigdata-m:~$ ls
cars.csv          kibana-7.5.1-linux-x86_64.tar.gz
elasticsearch-7.5.1 logstash-7.5.1
elasticsearch-7.5.1-linux-x86_64.tar.gz logstash-7.5.1.tar.gz
kibana-7.5.1-linux-x86_64
krлоттеу@bigdata-m:~$ vi logstash-cars.config
krлоттеу@bigdata-m:~$ ls
cars.csv          kibana-7.5.1-linux-x86_64.tar.gz
elasticsearch-7.5.1 logstash-7.5.1
elasticsearch-7.5.1-linux-x86_64.tar.gz logstash-7.5.1.tar.gz
kibana-7.5.1-linux-x86_64 logstash-cars.config
krлоттеу@bigdata-m:~$ cd logstash-7.5.1
krлоттеу@bigdata-m:~/logstash-7.5.1$ bin/logstash -f /home/krлоттеу/logstash-cars.c
onfig
Thread.exclusive is deprecated, use Thread::Mutex
Sending Logstash logs to /home/krлоттеу/logstash-7.5.1/logs which is now configured
via log4j2.properties
[2021-08-14T16:18:30,494][INFO ][logstash.setting.writabledirectory] Creating direc
tory {:setting=>"path.queue", :path=>"/home/krлоттеу/logstash-7.5.1/data/queue"}
[2021-08-14T16:18:30,629][INFO ][logstash.setting.writabledirectory] Creating direc
tory {:setting=>"path.dead_letter_queue", :path=>"/home/krлоттеу/logstash-7.5.1/dat
a/dead_letter_queue"}
[2021-08-14T16:18:30,987][WARN ][logstash.config.source.multilocal] Ignoring the 'p
ipelines.yml' file because modules or command line options are specified
[2021-08-14T16:18:30,999][INFO ][logstash.runner] Starting Logstash {"log
stash.version"=>"7.5.1"}
[2021-08-14T16:18:31,030][INFO ][logstash.agent] No persistent UUID file
found. Generating new UUID {:uuid=>"3b62ca67-6500-4bce-a80c-6eddcfd6ada4", :path=>
"/home/krлоттеу/logstash-7.5.1/data/uuid"}
[2021-08-14T16:18:33,763][INFO ][org.reflections.Reflections] Reflections took 41 m
s to scan 1 urls, producing 20 keys and 40 values
[2021-08-14T16:18:34,873][INFO ][logstash.outputs.elasticsearch][main] Elasticsearch
h pool URLs updated {:changes=>{:removed=>[], :added=>[http://localhost:9200/]} }
[2021-08-14T16:18:35,085][WARN ][logstash.outputs.elasticsearch][main] Restored con
nection to ES instance {:url=>"http://localhost:9200/" }
[2021-08-14T16:18:35,184][INFO ][logstash.outputs.elasticsearch][main] ES Output ve
rsion determined {:es_version=>7}
```

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 initiated by using the external 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.

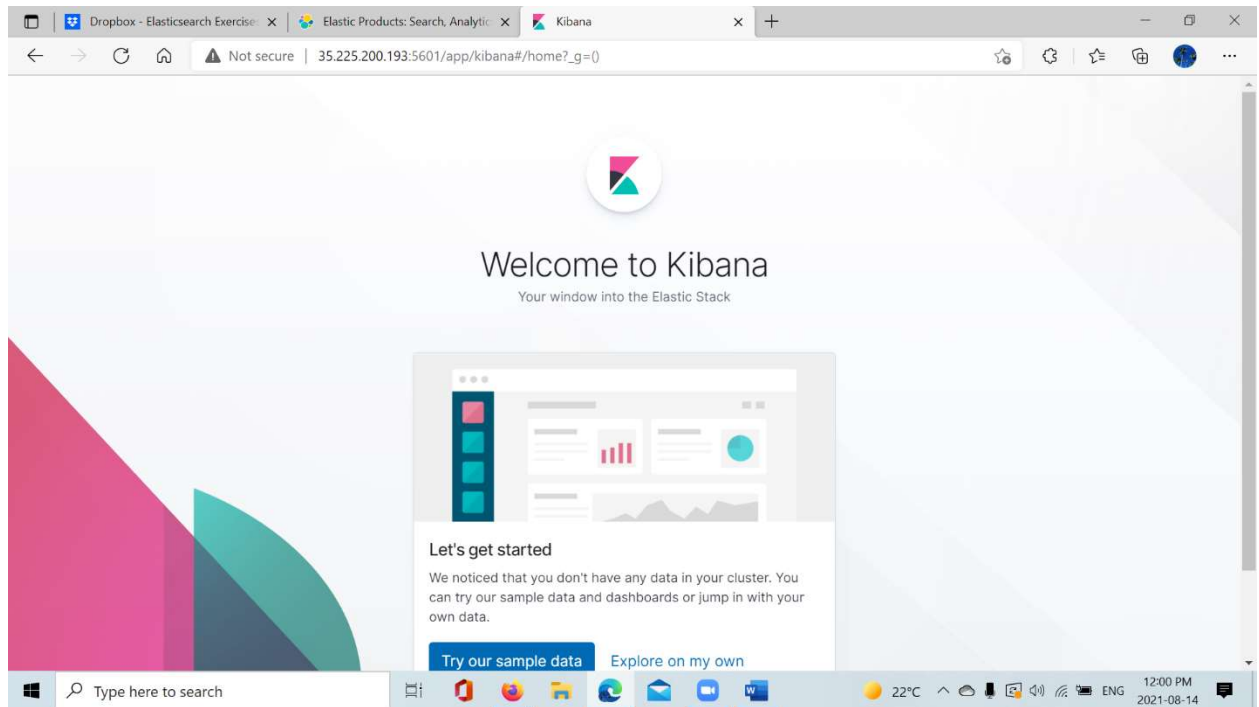


Figure 9: Modified IP Address To Open Kibana in Google

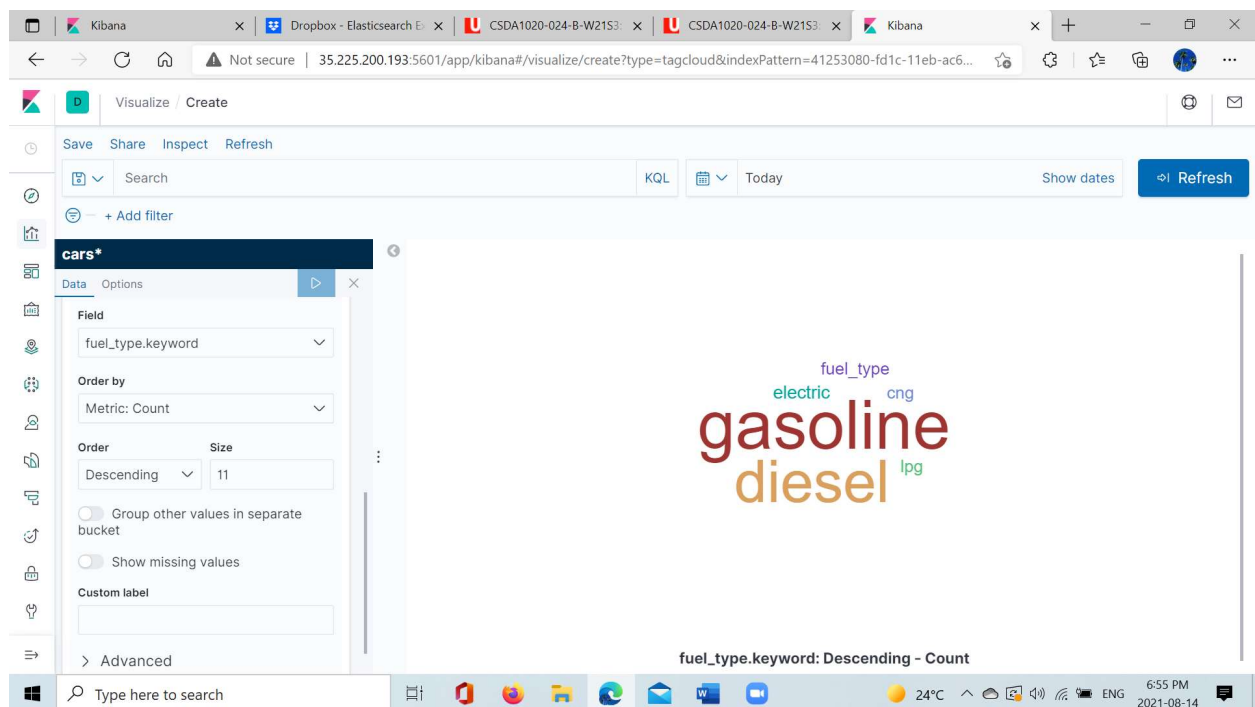


Figure 10: Example of a Tag Cloud Visualization for Fuel\_Type from cars.csv Dataset

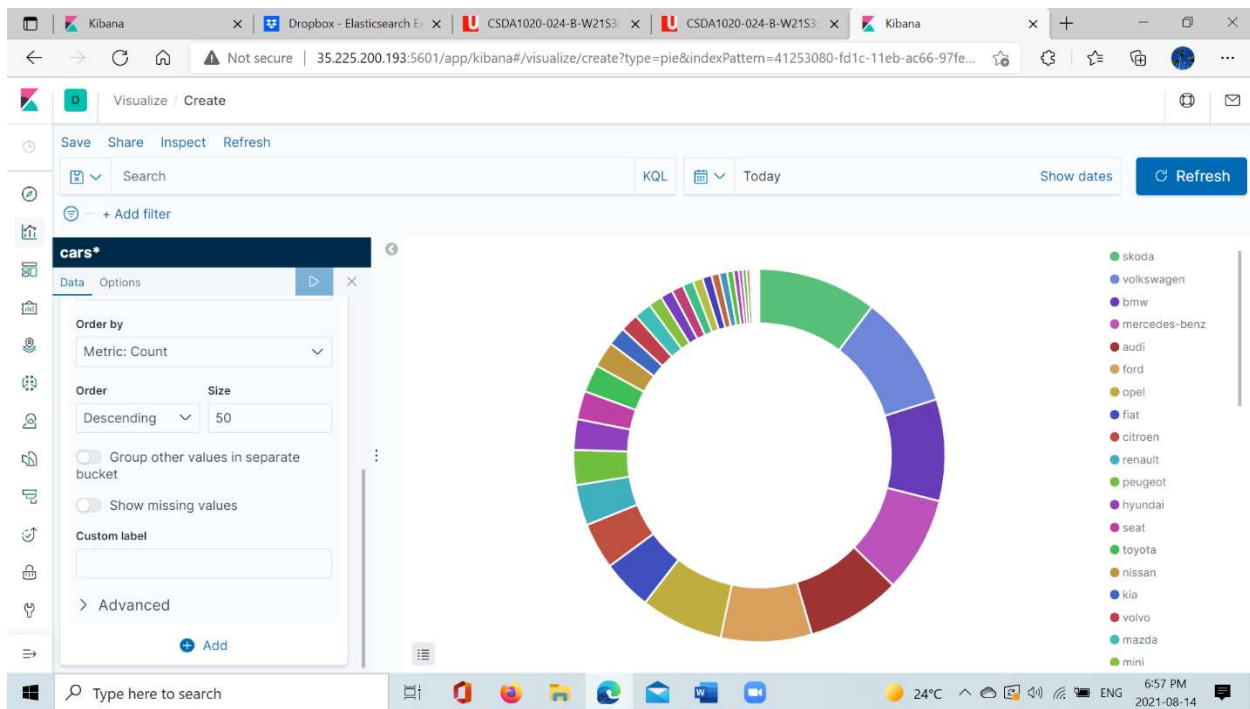


Figure 11: Example of a Pie Chart from the Data cars.csv for Maker Types