

Assignment 2 – *Data Visualisation with Kibana*

Submitted to:

Professor Saber Amini

Submitted by:

Irfan Abdul Rahman (200480839)

Manna Yuna Pagi (200489357)

Divya Pokhriyal (200502928)

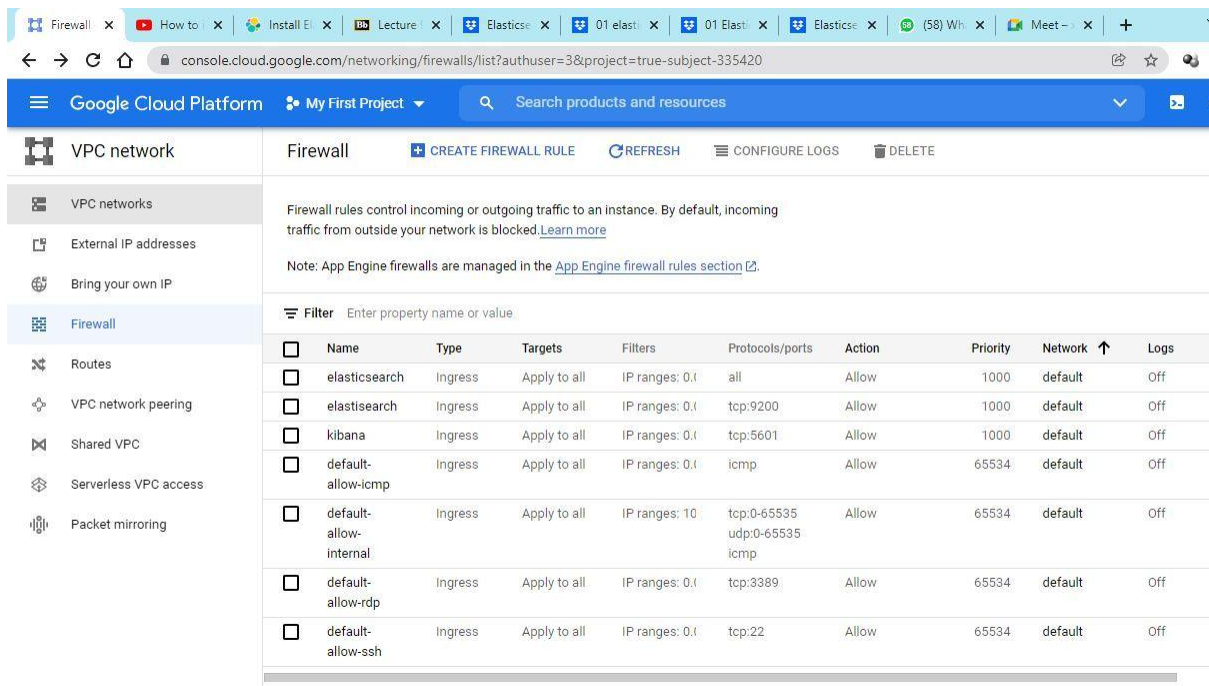
Introduction :

A Service Request is a request for the City to offer assistance, address a problem or perform an inspection. NYC311 can accept Service Requests for a wide range of issues, including over 500 complaint types¹. Service Requests offer a diverse set of services to citizens ranging from - refund for an overpaid parking ticket, complaint against a noisy neighbour, getting a pothole fixed, the heat turned on in an apartment. A citizen can submit a 311 request in various ways including by phone, online, or using mobile.²

In this assignment, we have used the ELK stack as an analytic tool to analyze realistic big data.

Following are the screenshots:

Firewall Rules to open ports for Elasticsearch and Kibana on GCP:

The screenshot shows the Google Cloud Platform console interface. The left sidebar contains navigation links for VPC network, VPC networks, External IP addresses, Bring your own IP, Firewall (selected), Routes, VPC network peering, Shared VPC, Serverless VPC access, and Packet mirroring. The main content area is titled 'Firewall' and includes buttons for 'CREATE FIREWALL RULE', 'REFRESH', 'CONFIGURE LOGS', and 'DELETE'. Below this, there is a filter input field and a table of firewall rules. The table has columns for Name, Type, Targets, Filters, Protocols/ports, Action, Priority, Network, and Logs. The rules listed are: elasticsearch (Ingress, Apply to all, IP ranges: 0.0.0.0/0, all, Allow, 1000, default, Off), elasticsearch (Ingress, Apply to all, IP ranges: 0.0.0.0/0, tcp:9200, Allow, 1000, default, Off), kibana (Ingress, Apply to all, IP ranges: 0.0.0.0/0, tcp:5601, Allow, 1000, default, Off), default-allow-icmp (Ingress, Apply to all, IP ranges: 0.0.0.0/0, icmp, Allow, 65534, default, Off), default-allow-internal (Ingress, Apply to all, IP ranges: 10.0.0.0/8, tcp:0-65535, udp:0-65535, icmp, Allow, 65534, default, Off), default-allow-rdp (Ingress, Apply to all, IP ranges: 0.0.0.0/0, tcp:3389, Allow, 65534, default, Off), and default-allow-ssh (Ingress, Apply to all, IP ranges: 0.0.0.0/0, tcp:22, Allow, 65534, default, Off).

<input type="checkbox"/>	Name	Type	Targets	Filters	Protocols/ports	Action	Priority	Network	↑	Logs
<input type="checkbox"/>	elasticsearch	Ingress	Apply to all	IP ranges: 0.0.0.0/0	all	Allow	1000	default		Off
<input type="checkbox"/>	elasticsearch	Ingress	Apply to all	IP ranges: 0.0.0.0/0	tcp:9200	Allow	1000	default		Off
<input type="checkbox"/>	kibana	Ingress	Apply to all	IP ranges: 0.0.0.0/0	tcp:5601	Allow	1000	default		Off
<input type="checkbox"/>	default-allow-icmp	Ingress	Apply to all	IP ranges: 0.0.0.0/0	icmp	Allow	65534	default		Off
<input type="checkbox"/>	default-allow-internal	Ingress	Apply to all	IP ranges: 10.0.0.0/8	tcp:0-65535 udp:0-65535 icmp	Allow	65534	default		Off
<input type="checkbox"/>	default-allow-rdp	Ingress	Apply to all	IP ranges: 0.0.0.0/0	tcp:3389	Allow	65534	default		Off
<input type="checkbox"/>	default-allow-ssh	Ingress	Apply to all	IP ranges: 0.0.0.0/0	tcp:22	Allow	65534	default		Off

¹ <https://portal.311.nyc.gov/article/?kanumber=KA-03116>

² <https://medium.com/@dmedellin2/a-deep-dive-into-nyc-311-requests-595fb7b31d7a>

Downloading Elasticsearch, Logstash and Kibana into the VM using wget:

```
karthikqwerty247@elk-m: ~/kibana-7.5.1-linux-x86_64 - Google Chrome
ssh.cloud.google.com/projects/true-subject-335420/zones/us-central1-f/instances/elk-m?authuser=3&hl=en_GB&projectNumber=8016...
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
karthikqwerty247@elk-m:~$ wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.5.1-linux-x86_64.tar.gz
--2021-12-20 20:24:20-- 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-linux-x86_64.tar.gz 100%[=====>] 276.65M 37.1MB/s in 11s

#bootstrap.memory_lock: true
2021-12-20 20:24:31 (24.4 MB/s) - 'elasticsearch-7.5.1-linux-x86_64.tar.gz' saved [290094012/290094012]

karthikqwerty247@elk-m:~$ wget https://artifacts.elastic.co/downloads/kibana/kibana-7.5.1-linux-x86_64.tar.gz
--2021-12-20 20:24:40-- 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'

kibana-7.5.1-linux-x86_64.tar.gz 100%[=====>] 227.43M 27.6MB/s in 8.7s

2021-12-20 20:24:49 (26.0 MB/s) - 'kibana-7.5.1-linux-x86_64.tar.gz' saved [238481011/238481011]

karthikqwerty247@elk-m:~$ wget https://artifacts.elastic.co/downloads/logstash/logstash-7.5.1.tar.gz
--2021-12-20 20:24:57-- 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'

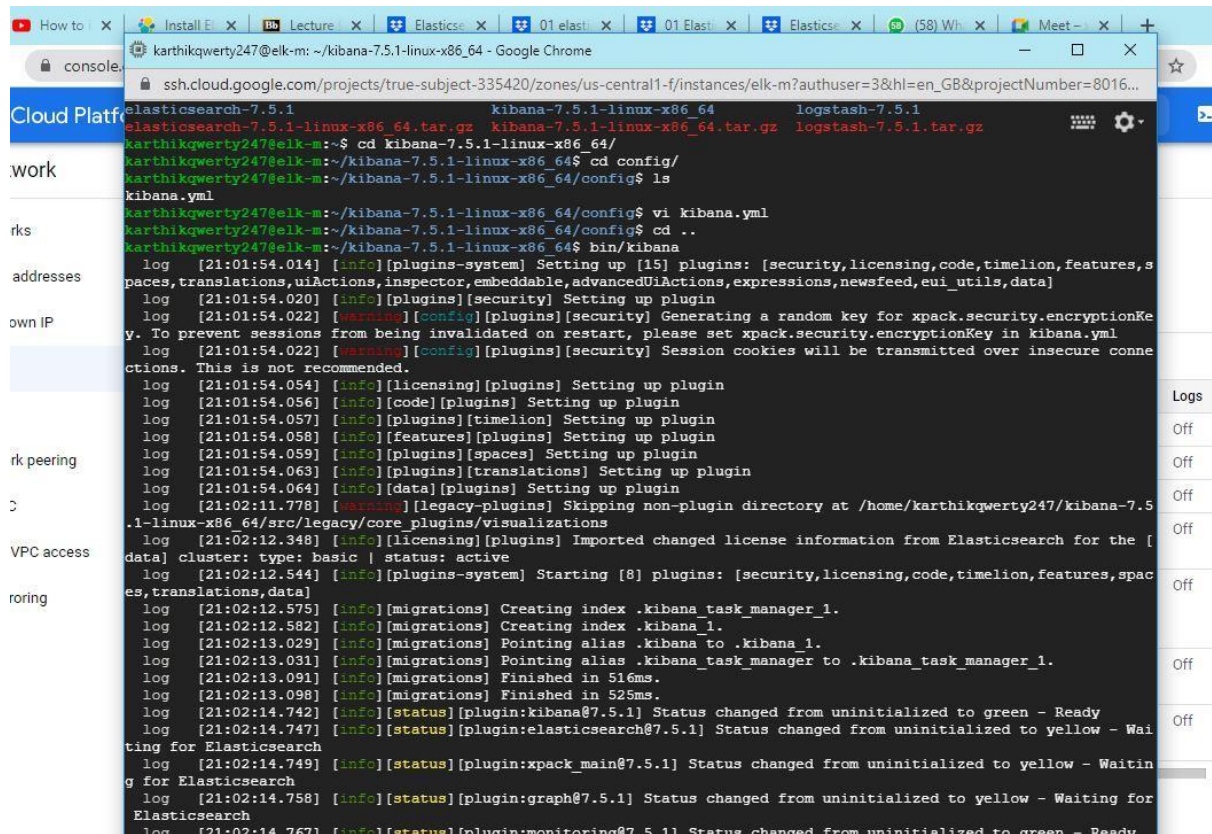
logstash-7.5.1.tar.gz 100%[=====>] 158.08M 53.3MB/s in 3.0s

2021-12-20 20:25:00 (53.3 MB/s) - 'logstash-7.5.1.tar.gz' saved [165760774/165760774]
```

Started Elasticsearch as Daemon:

```
karthikqwerty247@elk-m: ~/kibana-7.5.1-linux-x86_64 - Google Chrome
ssh.cloud.google.com/projects/true-subject-335420/zones/us-central1-f/instances/elk-m?authuser=3&hl=en_GB&projectNumber=8016...
[2021-12-20T20:45:27,981][INFO ][o.e.c.m.MetaDataIndexTemplateService] [elk-m] adding template [.monitoring-kibana-7-*]
[2021-12-20T20:45:28,009][INFO ][o.e.a.s.m.TransportMasterNodeAction] [elk-m] adding index lifecycle policy [watch-history-ilm-policy]
[2021-12-20T20:45:28,044][INFO ][o.e.a.s.m.TransportMasterNodeAction] [elk-m] adding index lifecycle policy [slm-history-ilm-policy]
[2021-12-20T20:45:28,182][INFO ][o.e.l.LicenseService] [elk-m] license [df7a4f3a-9ecc-4f95-8333-03745e1aa5fd] mode [basic] - valid
[2021-12-20T20:45:28,183][INFO ][o.e.x.s.s.SecurityStatusChangeListener] [elk-m] Active license is now [BASIC]; Security is disabled
^C[2021-12-20T20:48:40,045][INFO ][o.e.x.m.p.NativeController] [elk-m] Native controller process has stopped - no new native processes can be started
[2021-12-20T20:48:40,046][INFO ][o.e.n.Node] [elk-m] stopping ...
[2021-12-20T20:48:40,059][INFO ][o.e.x.w.WatcherService] [elk-m] stopping watch service, reason [shutdown initiated]
[2021-12-20T20:48:40,060][INFO ][o.e.x.w.WatcherLifeCycleService] [elk-m] watcher has stopped and shutdown
[2021-12-20T20:48:40,491][INFO ][o.e.n.Node] [elk-m] stopped
[2021-12-20T20:48:40,491][INFO ][o.e.n.Node] [elk-m] closing ...
[2021-12-20T20:48:40,501][INFO ][o.e.n.Node] [elk-m] closed
karthikqwerty247@elk-m:~/elasticsearch-7.5.1$ bin/elasticsearch/d
-bash: bin/elasticsearch/d: Not a directory
karthikqwerty247@elk-m:~/elasticsearch-7.5.1$ bin/elasticsearch -d
future versions of Elasticsearch will require Java 11; your Java version from [/usr/lib/jvm/adoptopenjdk-8-hotspot-amd64/jre] does not meet this requirement
```

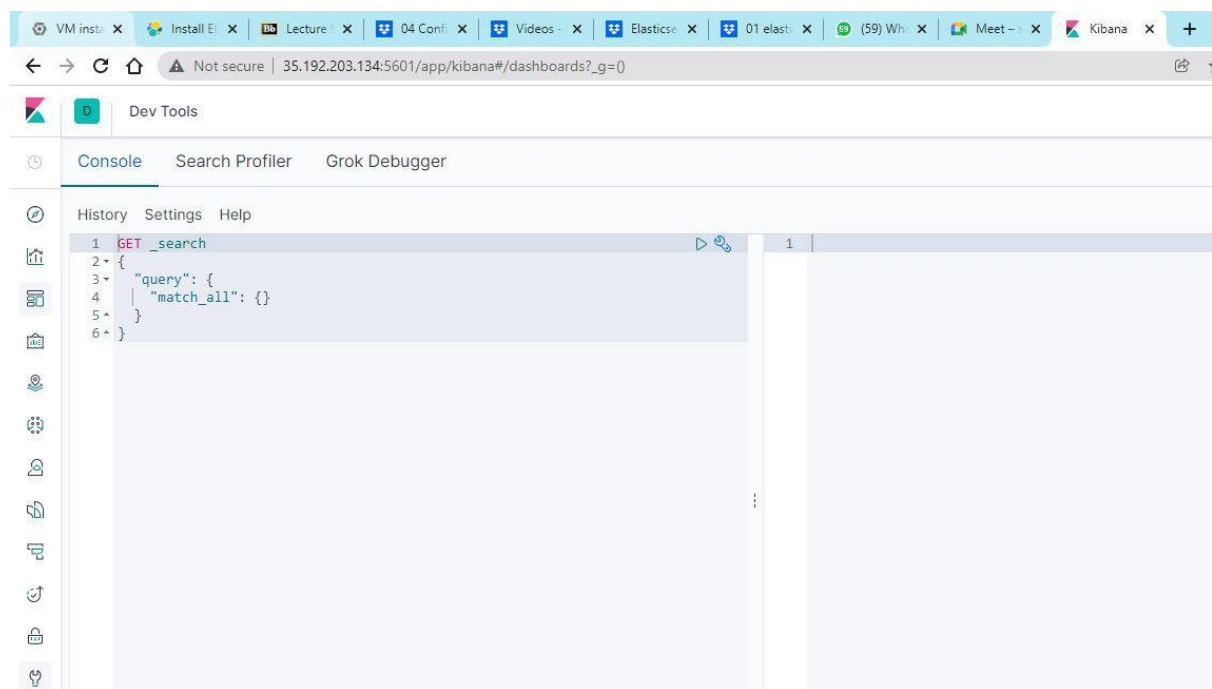
Running Kibana:



A terminal window showing the installation and startup of Kibana 7.5.1 on a Linux system. The user navigates to the Kibana directory, enters the configuration directory, and runs the Kibana binary. The output shows various log messages from the Kibana startup process, including plugin initialization, license information, and status changes for different components like Elasticsearch, Xpack, and Graph.

```
elasticsearch-7.5.1 kibana-7.5.1-linux-x86_64 logstash-7.5.1
elasticsearch-7.5.1-linux-x86_64.tar.gz kibana-7.5.1-linux-x86_64.tar.gz logstash-7.5.1.tar.gz
karthikqwerty247@elk-m:~$ cd kibana-7.5.1-linux-x86_64/
karthikqwerty247@elk-m:~/kibana-7.5.1-linux-x86_64$ cd config/
karthikqwerty247@elk-m:~/kibana-7.5.1-linux-x86_64/config$ ls
kibana.yml
karthikqwerty247@elk-m:~/kibana-7.5.1-linux-x86_64/config$ vi kibana.yml
karthikqwerty247@elk-m:~/kibana-7.5.1-linux-x86_64/config$ cd ..
karthikqwerty247@elk-m:~/kibana-7.5.1-linux-x86_64$ bin/kibana
log [21:01:54.014] [info] [plugins-system] Setting up [15] plugins: [security,licensing,code,timelion,features,spaces,translations,uiActions,inspector,embeddable,advancedUiActions,expressions,newsfeed,eui_utils,data]
log [21:01:54.020] [info] [plugins][security] Setting up plugin
log [21:01:54.022] [warning] [config] [plugins][security] Generating a random key for xpack.security.encryptionKey
y. To prevent sessions from being invalidated on restart, please set xpack.security.encryptionKey in kibana.yml
log [21:01:54.022] [warning] [config] [plugins][security] Session cookies will be transmitted over insecure connections. This is not recommended.
log [21:01:54.054] [info] [licensing][plugins] Setting up plugin
log [21:01:54.056] [info] [code][plugins] Setting up plugin
log [21:01:54.057] [info] [plugins][timelion] Setting up plugin
log [21:01:54.058] [info] [features][plugins] Setting up plugin
log [21:01:54.059] [info] [plugins][spaces] Setting up plugin
log [21:01:54.063] [info] [plugins][translations] Setting up plugin
log [21:01:54.064] [info] [data][plugins] Setting up plugin
log [21:02:11.778] [warning] [legacy-plugins] Skipping non-plugin directory at /home/karthikqwerty247/kibana-7.5.1-linux-x86_64/src/legacy/core_plugins/visualizations
log [21:02:12.348] [info] [licensing][plugins] Imported changed license information from Elasticsearch for the [data] cluster: type: basic | status: active
log [21:02:12.544] [info] [plugins-system] Starting [8] plugins: [security,licensing,code,timelion,features,spaces,translations,data]
log [21:02:12.575] [info] [migrations] Creating index .kibana_task_manager_1.
log [21:02:12.582] [info] [migrations] Creating index .kibana_1.
log [21:02:13.029] [info] [migrations] Pointing alias .kibana to .kibana_1.
log [21:02:13.031] [info] [migrations] Pointing alias .kibana_task_manager to .kibana_task_manager_1.
log [21:02:13.091] [info] [migrations] Finished in 516ms.
log [21:02:13.098] [info] [migrations] Finished in 525ms.
log [21:02:14.742] [info] [status] [plugin:kibana@7.5.1] Status changed from uninitialized to green - Ready
log [21:02:14.747] [info] [status] [plugin:elasticsearch@7.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
log [21:02:14.749] [info] [status] [plugin:xpack_main@7.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
log [21:02:14.758] [info] [status] [plugin:graph@7.5.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
log [21:02:14.767] [info] [status] [plugin:monitoring@7.5.1] Status changed from uninitialized to green - Ready
```

Kibana Console after setup:



Started Logstash and running a quick check:

```
karthikqwerty247@elk-m: ~/logstash-7.5.1 - Google Chrome
ssh.cloud.google.com/projects/true-subject-335420/zones/us-central1-f/instances/elk-m?authuser=3&hl=en_GB&projectNumber=801641889173&useAc

log [21:52:31.579] [info][features][plugins] Stopping plugin
log [21:52:31.579] [info][plugins][timelion] Stopping plugin
log [21:52:31.580] [info][licensing][plugins] Stopping plugin
log [21:52:31.580] [info][plugins][security] Stopping plugin
log [21:52:31.580] [info][code][plugins] Stopping plugin
karthikqwerty247@elk-m:~/kibana-7.5.1-linux-x86_64$ cd
karthikqwerty247@elk-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  log
karthikqwerty247@elk-m:~$ cd logstash-7.5.1/
karthikqwerty247@elk-m:~/logstash-7.5.1$ bin/logstash -e 'input {stdin{}} output {stdout{}}'
Thread.exclusive is deprecated, use Thread::Mutex
Sending Logstash logs to /home/karthikqwerty247/logstash-7.5.1/logs which is now configured via log4j2.properties
[2021-12-20T21:56:29,715][INFO ][logstash.setting.writabledirectory] Creating directory {:setting=>"path.queue", :path=>"/home/
[2021-12-20T21:56:29,847][INFO ][logstash.setting.writabledirectory] Creating directory {:setting=>"path.dead_letter_queue", :
/dead_letter_queue"}
[2021-12-20T21:56:30,178][WARN ][logstash.config.source.multilocal] Ignoring the 'pipelines.yml' file because modules or comma
[2021-12-20T21:56:30,189][INFO ][logstash.runner] Starting Logstash {"logstash.version"=>"7.5.1"}
[2021-12-20T21:56:30,217][INFO ][logstash.agent] No persistent UUID file found. Generating new UUID {:uuid=>"f212dk
rthikqwerty247/logstash-7.5.1/data/uuid"}
[2021-12-20T21:56:31,675][INFO ][org.reflections.Reflections] Reflections took 37 ms to scan 1 urls, producing 20 keys and 40
[2021-12-20T21:56:32,748][WARN ][org.logstash.instrument.metrics.gauge.LazyDelegatingGauge][main] A gauge metric of an unknown
ey: cluster uuids. This may result in invalid serialization. It is recommended to log an issue to the responsible developer/c
[2021-12-20T21:56:32,768][INFO ][logstash.javapipeline] [main] Starting pipeline {:pipeline_id=>"main", "pipeline.workers"=
elay"=>50, "pipeline.max_inflight"=>1000, "pipeline.sources"=>["config string"], :thread=>"#<Thread:0x510290 run>"}
[2021-12-20T21:56:32,864][INFO ][logstash.javapipeline] [main] Pipeline started {"pipeline.id"=>"main"}
The stdin plugin is now waiting for input:
[2021-12-20T21:56:32,924][INFO ][logstash.agent] Pipelines running {:count=>1, :running_pipelines=>[:main], :non_ru
[2021-12-20T21:56:33,248][INFO ][logstash.agent] Successfully started Logstash API endpoint {:port=>9600}
Hi!This Irfan, Divya and Manna
/home/karthikqwerty247/logstash-7.5.1/vendor/bundle/jruby/2.5.0/gems/awesome_print-1.7.0/lib/awesome_print/formatters/base_for
cated
{
  "@timestamp" => 2021-12-20T21:58:18.902Z,
  "message" => "Hi!This Irfan, Divya and Manna",
  "host" => "elk-m",
  "@version" => "1"
}
How is everyone doing?
{
  "@timestamp" => 2021-12-20T21:58:29.945Z,
  "message" => "How is everyone doing?",
  "host" => "elk-m",
  "@version" => "1"
}
```

Downloading dataset using wget command:

```

"@version" => "1"
}
{
"@timestamp" => 2021-12-20T22:19:36.632Z,
"message" => "",
"host" => "elk-m",
"@version" => "1"
}
}
{
"@timestamp" => 2021-12-20T22:19:36.809Z,
"message" => "",
"host" => "elk-m",
"@version" => "1"
}
}
^C[2021-12-20T22:19:43.507] [WARN ] [logstash.runner] SIGINT received. Shutting down.
--2021-12-20T22:19:43.759] [INFO ] [logstash.javapipeline] Pipeline terminated {"pipeline.id"=>"main"}
[2021-12-20T22:19:43.824] [INFO ] [logstash.runner] Logstash shut down.
karthikqwerty247@elk-m:~/logstash-7.5.1$ wget https://www.dropbox.com/s/dzop4gsuwn3esby/311_service.csv
--2021-12-20 22:41:31-- https://www.dropbox.com/s/dzop4gsuwn3esby/311_service.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/dzop4gsuwn3esby/311_service.csv [following]
--2021-12-20 22:41:31-- https://www.dropbox.com/s/raw/dzop4gsuwn3esby/311_service.csv
Reusing existing connection to www.dropbox.com:443.
HTTP request sent, awaiting response... 302 Found
Location: https://ucb63a416c3dae4b2d034b785fe4.dl.dropboxusercontent.com/cd/0/inline/BcM4hdsayUYJGnd1zRdH3F8MQcDgft93AYgHu0iKICJs5pQt
woY3qSAWHQc-gKyZjBxDAQVoZy7TeqN13Xpo5mSVJQu8om7YcNCl_HqFGACydHT/file# [following]
--2021-12-20 22:41:31-- https://ucb63a416c3dae4b2d034b785fe4.dl.dropboxusercontent.com/cd/0/inline/BcM4hdsayUYJGnd1zRdH3F8MQcDgft93A
oR58TuyLwJ_BB53vwoY3qSAWHQc-gKyZjBxDAQVoZy7TeqN13Xpo5mSVJQu8om7YcNCl_HqFGACydHT/file
Resolving ucb63a416c3dae4b2d034b785fe4.dl.dropboxusercontent.com (ucb63a416c3dae4b2d034b785fe4.dl.dropboxusercontent.com)... 162.125.
Connecting to ucb63a416c3dae4b2d034b785fe4.dl.dropboxusercontent.com (ucb63a416c3dae4b2d034b785fe4.dl.dropboxusercontent.com) [162.125
HTTP request sent, awaiting response... 200 OK
Length: 13247207051 (12G) [text/plain]
Saving to: '311_service.csv'

311_service.csv          100%[=====

2021-12-20 22:43:54 (89.5 MB/s) - '311_service.csv' saved [13247207051/13247207051]
```

Loading dataset in kibana using logstash config file

A screenshot of a terminal window. The title bar at the top reads "karthikqwerty247@elk-m: ~/logstash-7.5.1 - Google Chrome". The address bar below it shows "ssh.cloud.google.com/projects/true-subject-335420/zones/us-central1-f/instances/elk-m?authuser=3&hl=en_GB&projectNumber=801641889173&useAdminProxy". The terminal area is dark with a light-colored grid pattern. The prompt "karthikqwerty247@elk-m: ~/logstash-7.5.1" is visible at the top left of the terminal area.

Elasticsearch Index Management page on Kibana:

Management / Index Management

Index Management

Indices Index Templates

Update your Elasticsearch indices individually or in bulk. ☐ Include rollout indices ☐ Include system

Search

Name	Health	Status	Primaries	Replicas	Docs count	Storage size
<input type="checkbox"/> nycinfo	yellow	open	1	1	7300589	12.1gb

Rows per page: 10

Question 1. A table showing the top 10 cities with the highest calls alongside the count of top 10 complaint calls (by Descriptor) in each city.

Visualize / Question - 1

Save Share Inspect Refresh

Search KQL Last 15 minutes Show dates

+ Add filter

nyc*

Data Options

Metrics

Metric Count

Buckets

Split rows City.keyword: De...

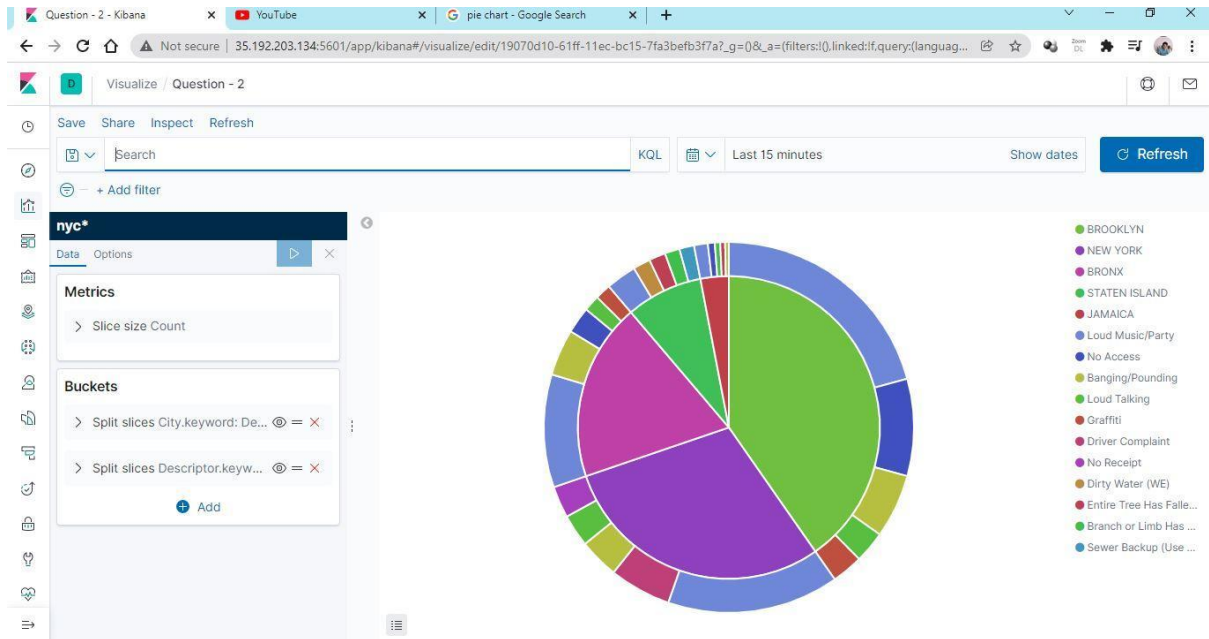
Split rows Descriptor.keywo...

City.keyword: Descending	Descriptor.keyword: Descending	Count
BROOKLYN	Loud Music/Party	36,348
BROOKLYN	No Access	14,898
BROOKLYN	Banging/Pounding	9,925
BROOKLYN	Loud Talking	4,895
BROOKLYN	Graffiti	4,854
BROOKLYN	Sewer Backup (Use Comments) (SA)	4,374
BROOKLYN	Branch or Limb Has Fallen Down	4,305
BROOKLYN	No Receipt	4,230
BROOKLYN	Partial Access	3,941
BROOKLYN	For One Address	3,861

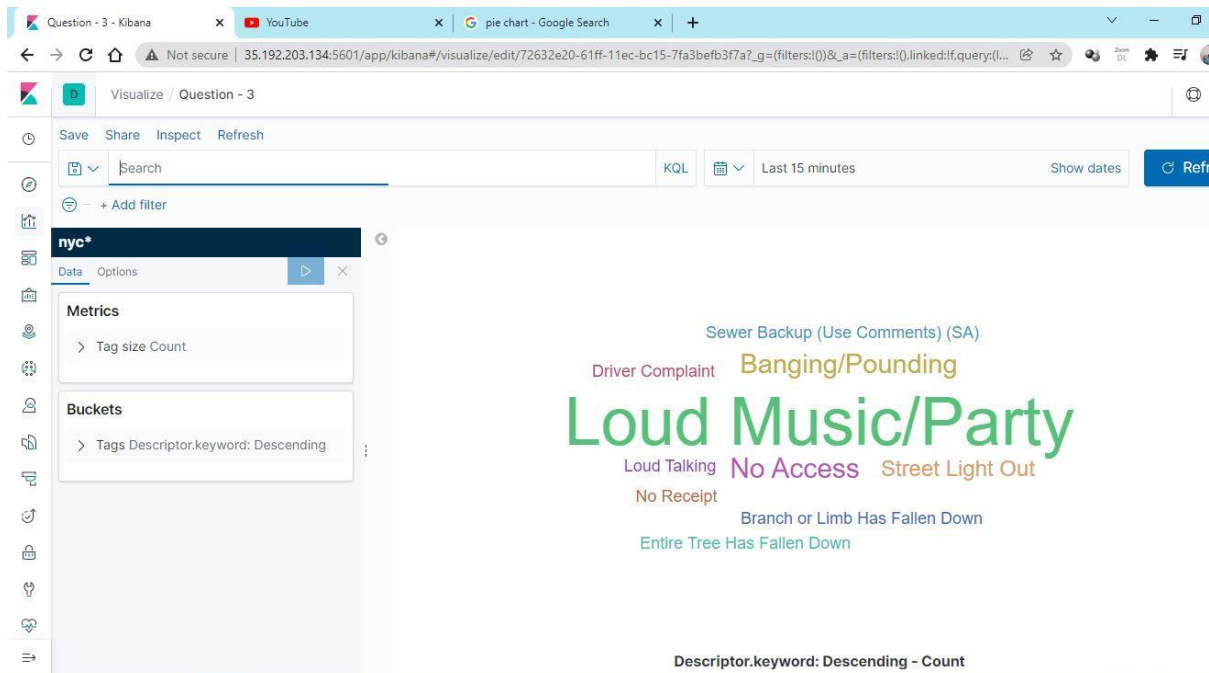
Export: Raw Formatted

1 2 3 4 5 ... 10

Question 2. A pie chart showing the top 5 cities with the highest calls alongside the top five calls (Descriptor) in each city.



Question 3. A tag cloud representing the top 20 call descriptors.



Final Dashboard

