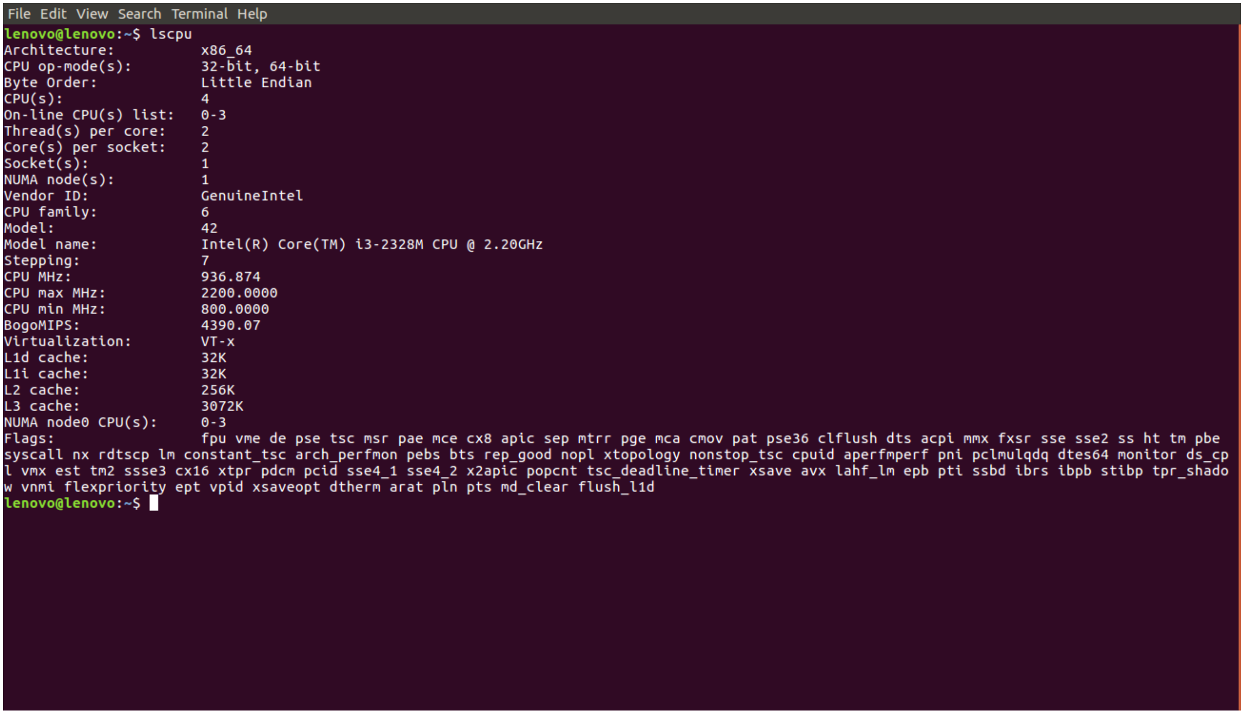
Customer Classification Prediction Model For Online Store

Prepared by
Arhanti Gawde
Arpita Hirlekar
Snehal Surve

CODES AND OUTPUT

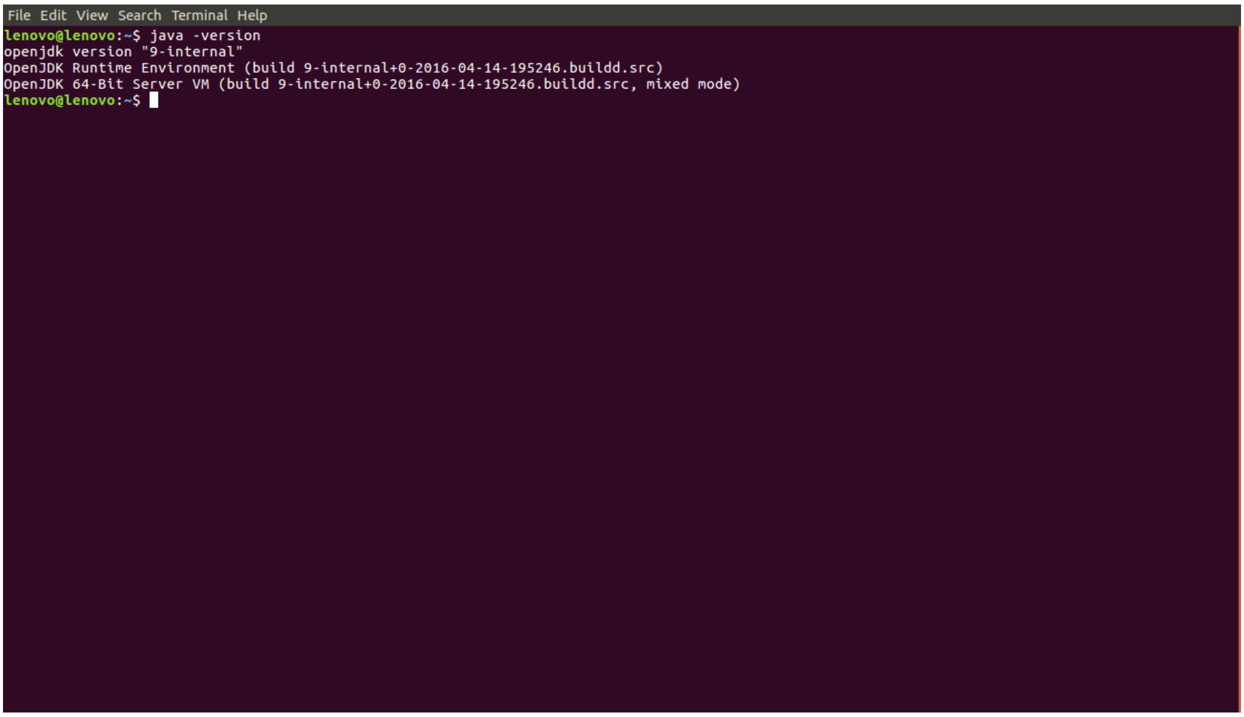
Installing Elasticsearch and Kibana Packages.

- Check the Ubuntu System Architecture using the following command:



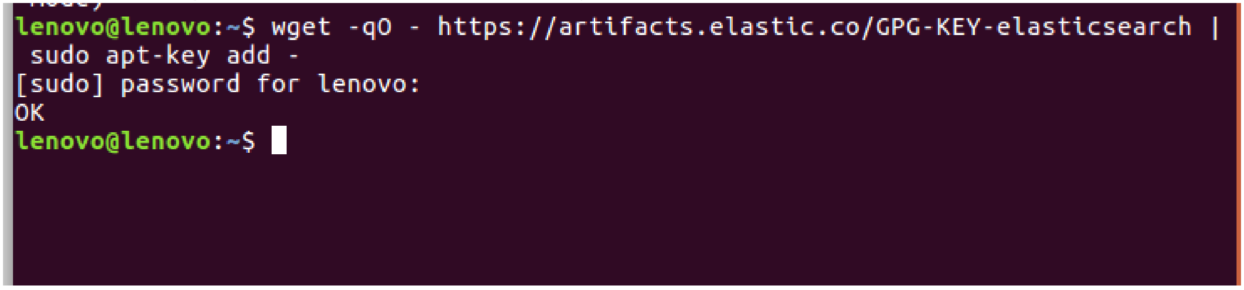
```
lenovo@lenovo:~$ lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                4
On-line CPU(s) list:   0-3
Thread(s) per core:    2
Core(s) per socket:    2
Socket(s):             1
NUMA node(s):          1
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 42
Model name:            Intel(R) Core(TM) i3-2328M CPU @ 2.20GHz
Stepping:              7
CPU MHz:               936.874
CPU max MHz:           2200.0000
CPU min MHz:           800.0000
BogoMIPS:              4390.07
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              256K
L3 cache:              3072K
NUMA node0 CPU(s):     0-3
Flags:                 fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe
syscall nx rdtscp lm constant_tsc arch_perfmon pebs bts rep_good nopl xtopology nonstop_tsc cpuid aperfmperf pni pclmulqdq dtes64 monitor ds_cp
l vmx est tm2 ssse3 cx16 xtpr pdcm pcid sse4_1 sse4_2 x2apic popcnt tsc_deadline_timer xsave avx lahf_lm epb pti ssbd ibrs ibpb stibp tpr_shado
w vnmi flexpriority ept vpid xsaveopt dtherm arat pln pts md_clear flush_l1d
lenovo@lenovo:~$ 
```

- Check the Java Version as shown below:



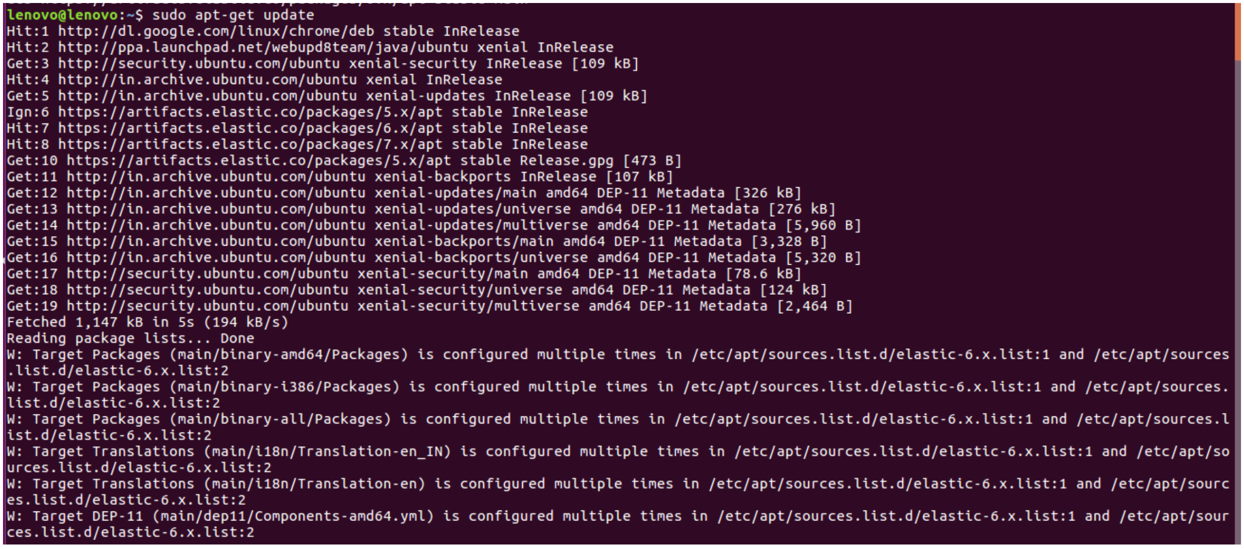
```
lenovo@lenovo:~$ java -version
openjdk version "9-internal"
OpenJDK Runtime Environment (build 9-internal+0-2016-04-14-195246.buildd.src)
OpenJDK 64-Bit Server VM (build 9-internal+0-2016-04-14-195246.buildd.src, mixed mode)
lenovo@lenovo:~$
```

- Importing Elasticsearch public GPG key into APT:



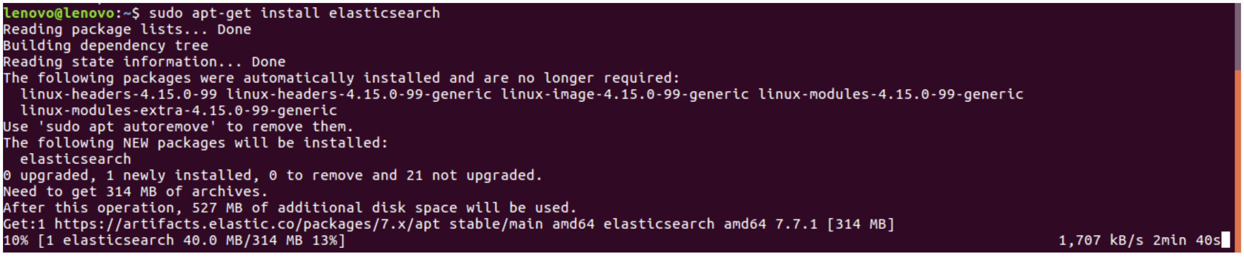
```
lenovo@lenovo:~$ wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch |
 sudo apt-key add -
[sudo] password for lenovo:
OK
lenovo@lenovo:~$
```

- Updating the package lists so that APT can read the new Elastic source:



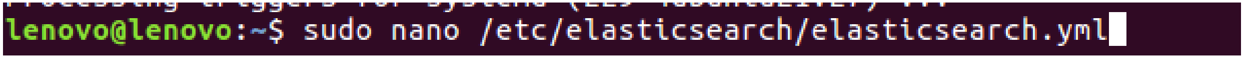
```
lenovo@lenovo:~$ sudo apt-get update
Hit:1 http://dl.google.com/linux/chrome/deb stable InRelease
Hit:2 http://ppa.launchpad.net/webupd8team/java/ubuntu xenial InRelease
Get:3 http://security.ubuntu.com/ubuntu xenial-security InRelease [109 kB]
Hit:4 http://in.archive.ubuntu.com/ubuntu xenial InRelease
Get:5 http://in.archive.ubuntu.com/ubuntu xenial-updates InRelease [109 kB]
Ign:6 https://artifacts.elastic.co/packages/5.x/apt stable InRelease
Hit:7 https://artifacts.elastic.co/packages/6.x/apt stable InRelease
Hit:8 https://artifacts.elastic.co/packages/7.x/apt stable InRelease
Get:10 https://artifacts.elastic.co/packages/5.x/apt stable Release.gpg [473 B]
Get:11 http://in.archive.ubuntu.com/ubuntu xenial-backports InRelease [107 kB]
Get:12 http://in.archive.ubuntu.com/ubuntu xenial-updates/main amd64 DEP-11 Metadata [326 kB]
Get:13 http://in.archive.ubuntu.com/ubuntu xenial-updates/universe amd64 DEP-11 Metadata [276 kB]
Get:14 http://in.archive.ubuntu.com/ubuntu xenial-updates/multiverse amd64 DEP-11 Metadata [5,960 B]
Get:15 http://in.archive.ubuntu.com/ubuntu xenial-backports/main amd64 DEP-11 Metadata [3,328 B]
Get:16 http://in.archive.ubuntu.com/ubuntu xenial-backports/universe amd64 DEP-11 Metadata [5,320 B]
Get:17 http://security.ubuntu.com/ubuntu xenial-security/main amd64 DEP-11 Metadata [78.6 kB]
Get:18 http://security.ubuntu.com/ubuntu xenial-security/universe amd64 DEP-11 Metadata [124 kB]
Get:19 http://security.ubuntu.com/ubuntu xenial-security/multiverse amd64 DEP-11 Metadata [2,464 B]
Fetched 1,147 kB in 5s (194 kB/s)
Reading package lists... Done
W: Target Packages (main/binary-amd64/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-6.x.list:1 and /etc/apt/sources
.list.d/elastic-6.x.list:2
W: Target Packages (main/binary-i386/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-6.x.list:1 and /etc/apt/sources.
list.d/elastic-6.x.list:2
W: Target Packages (main/binary-all/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-6.x.list:1 and /etc/apt/sources.l
ist.d/elastic-6.x.list:2
W: Target Translations (main/i18n/Translation-en_IN) is configured multiple times in /etc/apt/sources.list.d/elastic-6.x.list:1 and /etc/apt/so
urces.list.d/elastic-6.x.list:2
W: Target Translations (main/i18n/Translation-en) is configured multiple times in /etc/apt/sources.list.d/elastic-6.x.list:1 and /etc/apt/sourc
es.list.d/elastic-6.x.list:2
W: Target DEP-11 (main/dep11/Components-amd64.yml) is configured multiple times in /etc/apt/sources.list.d/elastic-6.x.list:1 and /etc/apt/sour
ces.list.d/elastic-6.x.list:2
```

- Installing Elasticsearch:



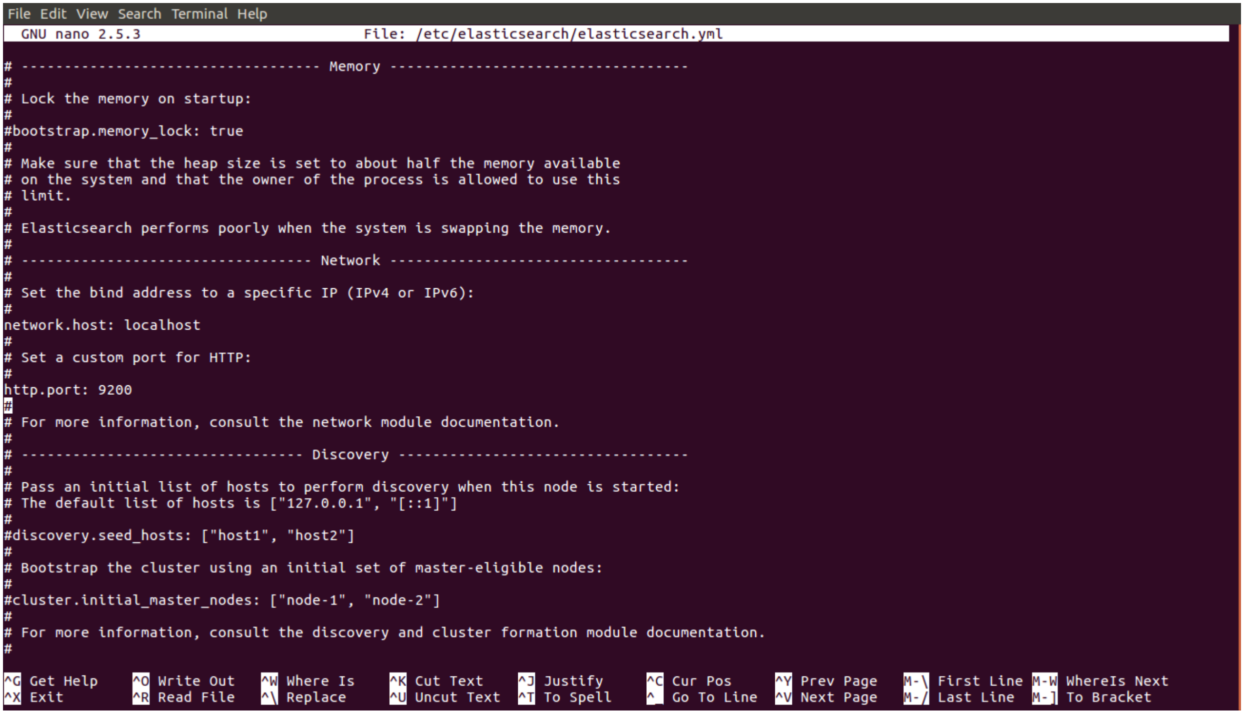
```
lenovo@lenovo:~$ sudo apt-get install elasticsearch
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-4.15.0-99 linux-headers-4.15.0-99-generic linux-image-4.15.0-99-generic linux-modules-4.15.0-99-generic
  linux-modules-extra-4.15.0-99-generic
Use 'sudo apt autoremove' to remove them.
The following NEW packages will be installed:
  elasticsearch
0 upgraded, 1 newly installed, 0 to remove and 21 not upgraded.
Need to get 314 MB of archives.
After this operation, 527 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 elasticsearch amd64 7.7.1 [314 MB]
10% [1 elasticsearch 40.0 MB/314 MB 13%]                                                          1,707 kB/s 2min 40s
```

- To edit Elasticsearch’s main configuration file:



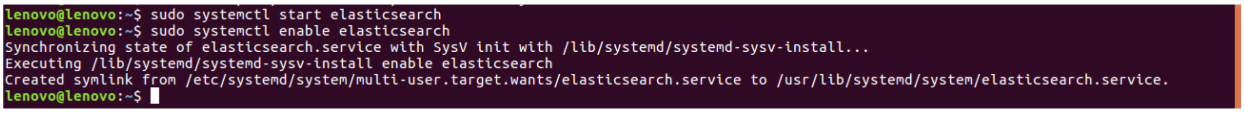
```
lenovo@lenovo:~$ sudo nano /etc/elasticsearch/elasticsearch.yml
```

- Uncomment network host and http port:



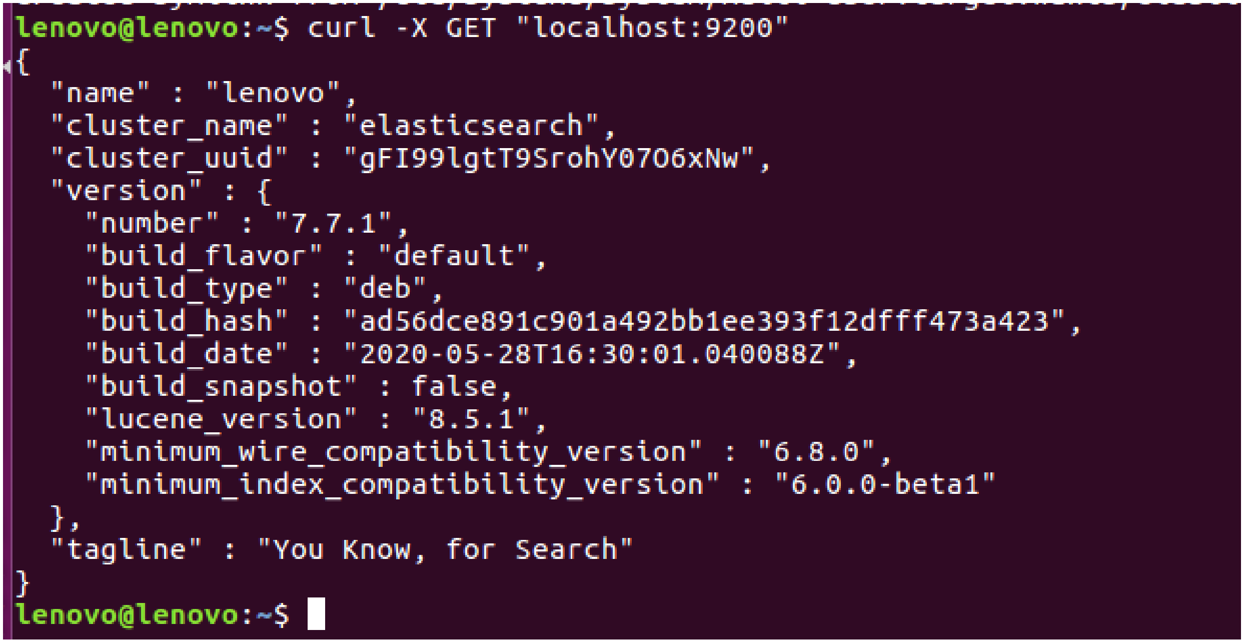
```
  GNU nano 2.5.3                          File: /etc/elasticsearch/elasticsearch.yml

# ----------------------------------- Memory -----------------------------------
#
# Lock the memory on startup:
#
#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: localhost
#
# 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: ["host1", "host2"]
#
# Bootstrap the cluster using an initial set of master-eligible nodes:
#
#cluster.initial_master_nodes: ["node-1", "node-2"]
#
# For more information, consult the discovery and cluster formation module documentation.
#

^G Get Help    ^O Write Out   ^W Where Is    ^K Cut Text    ^J Justify     ^C Cur Pos     ^Y Prev Page   M-\ First Line M-W WhereIs Next
^X Exit        ^R Read File   ^\ Replace     ^U Uncut Text  ^T To Spell    ^_ Go To Line  ^V Next Page   M-/ Last Line  M-] To Bracket
```

- Start Elasticsearch:



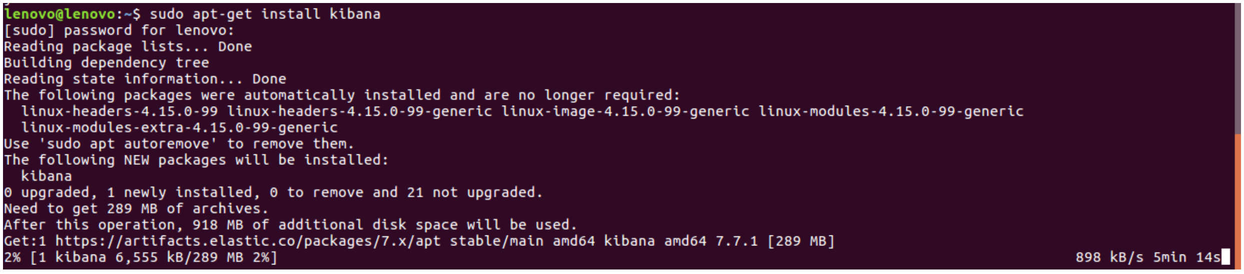
```
lenovo@lenovo:~$ sudo systemctl start elasticsearch
lenovo@lenovo:~$ sudo systemctl enable elasticsearch
Synchronizing state of elasticsearch.service with SysV init with /lib/systemd/systemd-sysv-install...
Executing /lib/systemd/systemd-sysv-install enable elasticsearch
Created symlink from /etc/systemd/system/multi-user.target.wants/elasticsearch.service to /usr/lib/systemd/system/elasticsearch.service.
lenovo@lenovo:~$
```

- Testing the Elasticsearch:



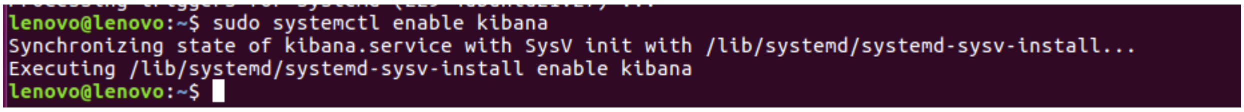
```
lenovo@lenovo:~$ curl -X GET "localhost:9200"
{
  "name" : "lenovo",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "gFI99lgtT9SrohY0706xNw",
  "version" : {
    "number" : "7.7.1",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "ad56dce891c901a492bb1ee393f12dfff473a423",
    "build_date" : "2020-05-28T16:30:01.040088Z",
    "build_snapshot" : false,
    "lucene_version" : "8.5.1",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
lenovo@lenovo:~$
```

- Installing Kibana

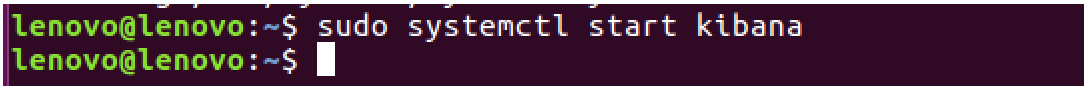


```
lenovo@lenovo:~$ sudo apt-get install kibana
[sudo] password for lenovo:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-4.15.0-99 linux-headers-4.15.0-99-generic linux-image-4.15.0-99-generic linux-modules-4.15.0-99-generic
  linux-modules-extra-4.15.0-99-generic
Use 'sudo apt autoremove' to remove them.
The following NEW packages will be installed:
  kibana
0 upgraded, 1 newly installed, 0 to remove and 21 not upgraded.
Need to get 289 MB of archives.
After this operation, 918 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 kibana amd64 7.7.1 [289 MB]
2% [1 kibana 6,555 kB/289 MB 2%]                                                                          898 kB/s 5min 14s
```

- Enable and Start Kibana Services:

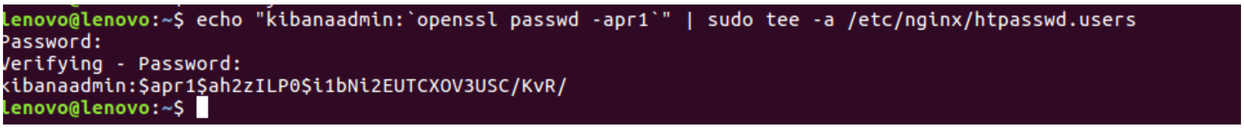


```
lenovo@lenovo:~$ sudo systemctl enable kibana
Synchronizing state of kibana.service with SysV init with /lib/systemd/systemd-sysv-install...
Executing /lib/systemd/systemd-sysv-install enable kibana
lenovo@lenovo:~$
```



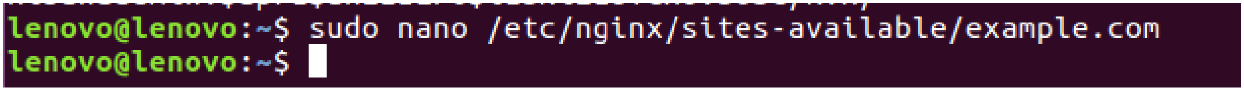
```
lenovo@lenovo:~$ sudo systemctl start kibana
lenovo@lenovo:~$
```

- Create an administrative Kibana user:

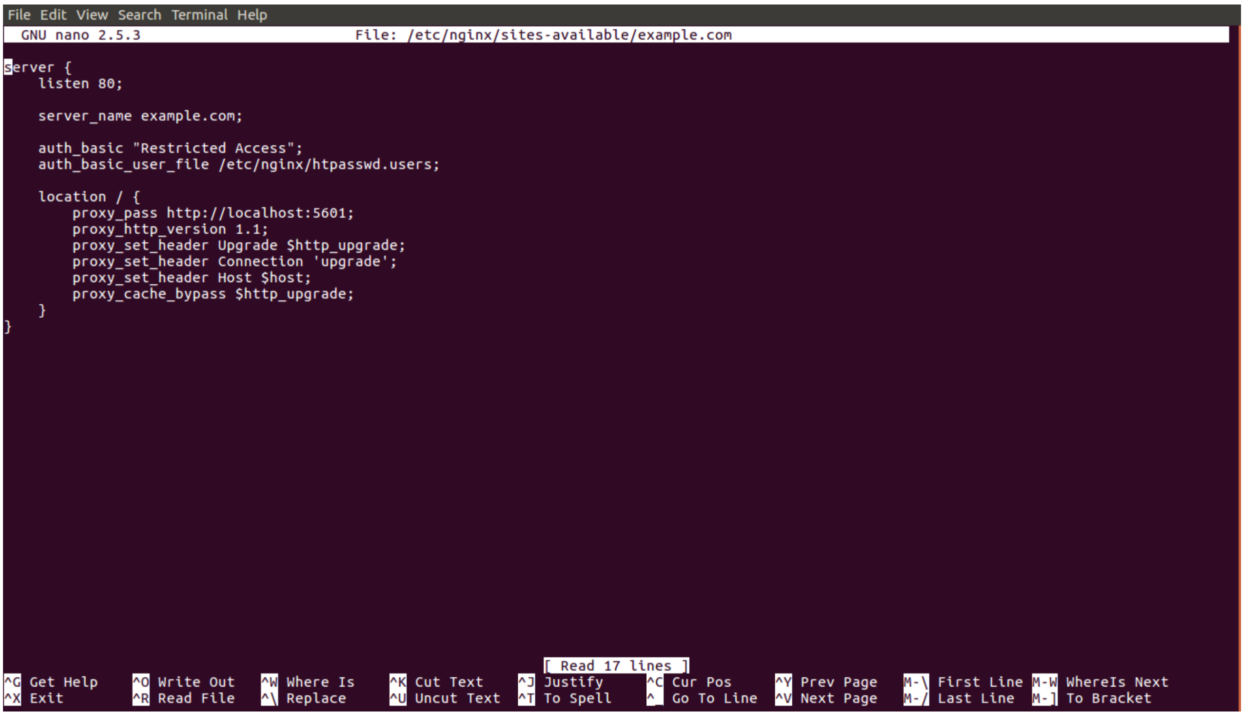


```
lenovo@lenovo:~$ echo "kibanaadmin:`openssl passwd -apr1`" | sudo tee -a /etc/nginx/htpasswd.users
Password:
Verifying - Password:
kibanaadmin:$apr1$ah2zILP0$i1bNi2EUTCXOV3USC/KvR/
lenovo@lenovo:~$
```

- Check the Nginx:

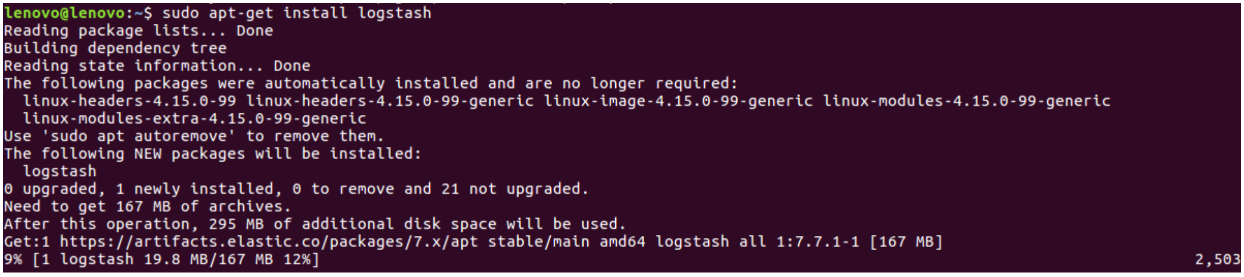


```
lenovo@lenovo:~$ sudo nano /etc/nginx/sites-available/example.com
lenovo@lenovo:~$ 
```



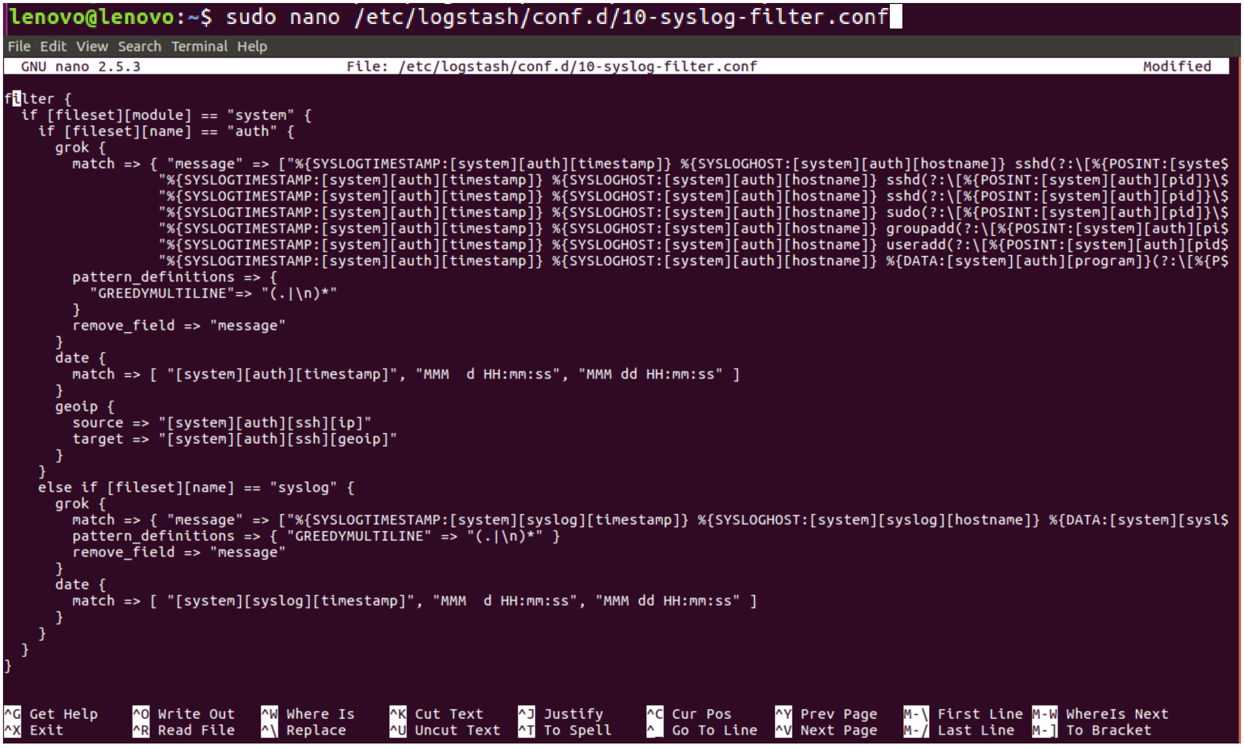
```
server {
    listen 80;

    server_name example.com;

    auth_basic "Restricted Access";
    auth_basic_user_file /etc/nginx/htpasswd.users;

    location / {
        proxy_pass http://localhost:5601;
        proxy_http_version 1.1;
        proxy_set_header Upgrade $http_upgrade;
        proxy_set_header Connection 'upgrade';
        proxy_set_header Host $host;
        proxy_cache_bypass $http_upgrade;
    }
}
```

- Installing Logstash:



```
lenovo@lenovo:~$ sudo apt-get install logstash
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-4.15.0-99 linux-headers-4.15.0-99-generic linux-image-4.15.0-99-generic linux-modules-4.15.0-99-generic
  linux-modules-extra-4.15.0-99-generic
Use 'sudo apt autoremove' to remove them.
The following NEW packages will be installed:
  logstash
0 upgraded, 1 newly installed, 0 to remove and 21 not upgraded.
Need to get 167 MB of archives.
After this operation, 295 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 logstash all 1:7.7.1-1 [167 MB]
9% [1 logstash 19.8 MB/167 MB 12%]                                                                                     2,503
```

- Add a Filter to the system logs:

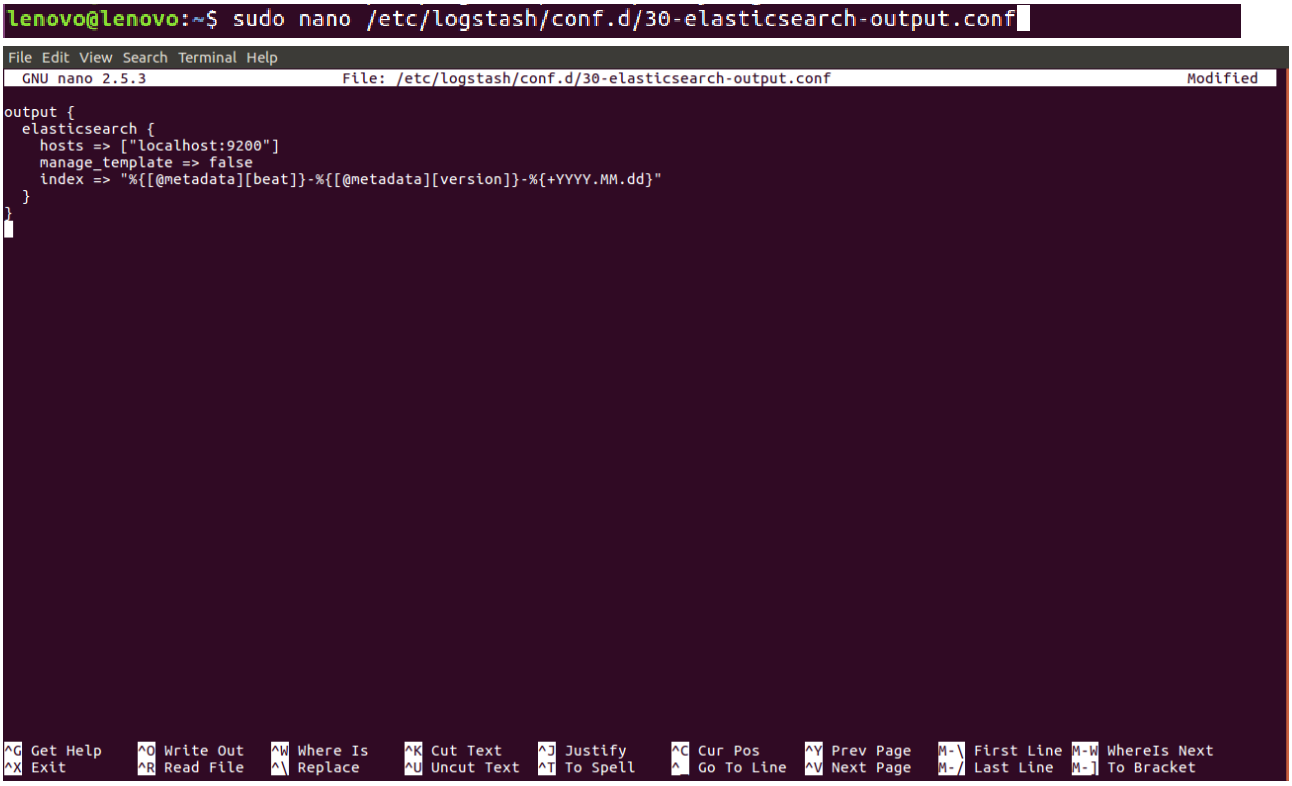


```
lenovo@lenovo:~$ sudo nano /etc/logstash/conf.d/10-syslog-filter.conf
```

File Edit View Search Terminal Help
GNU nano 2.5.3
File: /etc/logstash/conf.d/10-syslog-filter.conf
Modified

```
filter {
  if [fileset][module] == "system" {
    if [fileset][name] == "auth" {
      grok {
        match => { "message" => ["%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} sshd(?:\[%{POSINT:[syste$
                  "%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} sshd(?:\[%{POSINT:[system][auth][pid]}\$
                  "%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} sshd(?:\[%{POSINT:[system][auth][pid]}\$
                  "%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} sudo(?:\[%{POSINT:[system][auth][pid]}\$
                  "%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} groupadd(?:\[%{POSINT:[system][auth][pi$
                  "%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} useradd(?:\[%{POSINT:[system][auth][pid$
                  "%{SYSLOGTIMESTAMP:[system][auth][timestamp]} %{SYSLOGHOST:[system][auth][hostname]} %{DATA:[system][auth][program]}(?:\[%{P$
        pattern_definitions => {
          "GREEDYMULTILINE"=> "(.|\n)*"
        }
        remove_field => "message"
      }
      date {
        match => [ "[system][auth][timestamp]", "MMM  d HH:mm:ss", "MMM dd HH:mm:ss" ]
      }
      geoip {
        source => "[system][auth][ssh][ip]"
        target => "[system][auth][ssh][geoip]"
      }
    }
    else if [fileset][name] == "syslog" {
      grok {
        match => { "message" => ["%{SYSLOGTIMESTAMP:[system][syslog][timestamp]} %{SYSLOGHOST:[system][syslog][hostname]} %{DATA:[system][sysl$
        pattern_definitions => { "GREEDYMULTILINE" => "(.|\n)*" }
        remove_field => "message"
      }
      date {
        match => [ "[system][syslog][timestamp]", "MMM  d HH:mm:ss", "MMM dd HH:mm:ss" ]
      }
    }
  }
}
```

- Create a config file:

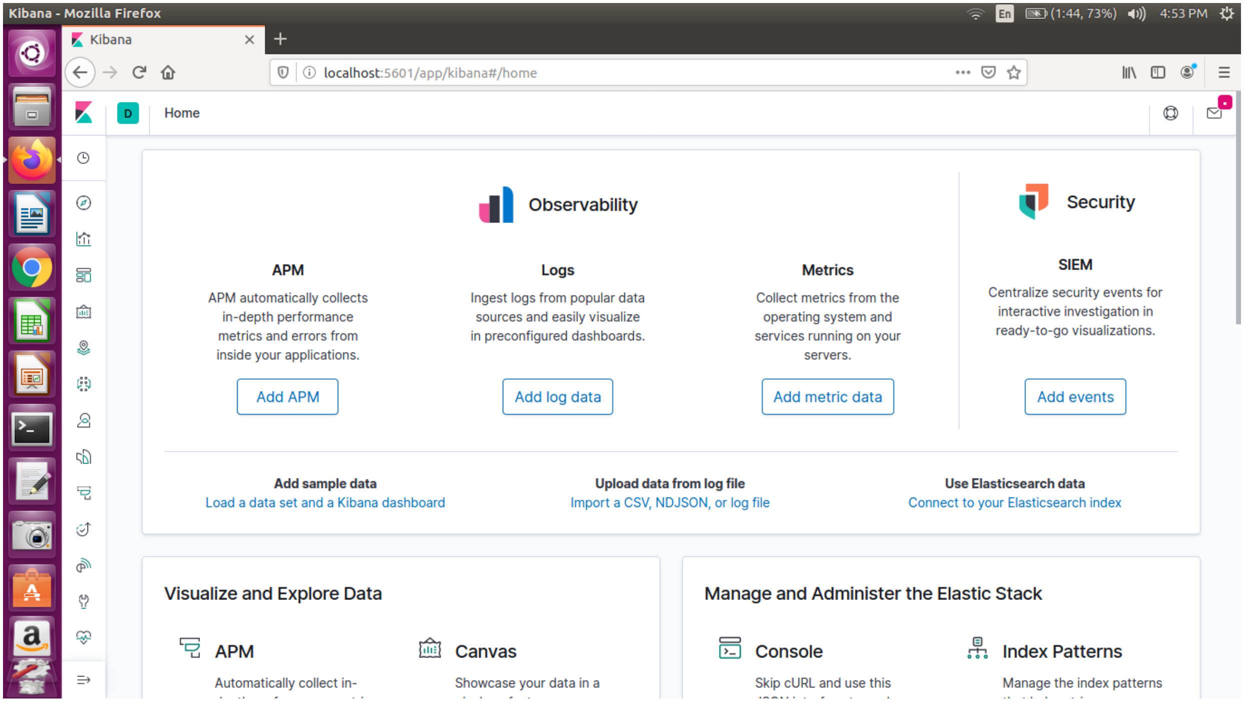


```
lenovo@lenovo:~$ sudo nano /etc/logstash/conf.d/30-elasticsearch-output.conf
```

File Edit View Search Terminal Help
GNU nano 2.5.3
File: /etc/logstash/conf.d/30-elasticsearch-output.conf
Modified

```
output {
  elasticsearch {
    hosts => ["localhost:9200"]
    manage_template => false
    index => "%{[@metadata][beat]}-%{[@metadata][version]}-%{+YYYY.MM.dd}"
  }
}
```

- Kibana Home Page:



APM

APM automatically collects in-depth performance metrics and errors from inside your applications.

Logs

Ingest logs from popular data sources and easily visualize in preconfigured dashboards.

Metrics

Collect metrics from the operating system and services running on your servers.

SIEM

Centralize security events for interactive investigation in ready-to-go visualizations.

Add sample data
Load a data set and a Kibana dashboard

Upload data from log file
Import a CSV, NDJSON, or log file

Use Elasticsearch data
Connect to your Elasticsearch index

Visualize and Explore Data

APM
Automatically collect in-

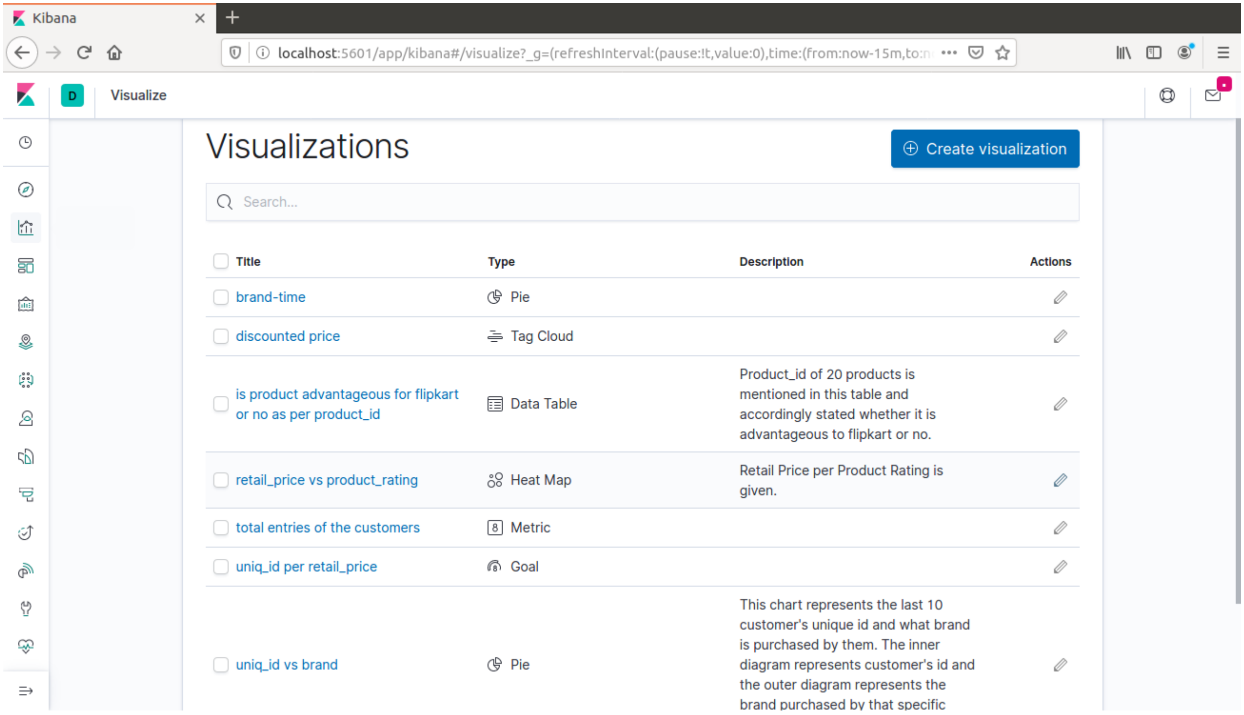
Canvas
Showcase your data in a

Manage and Administer the Elastic Stack

Console
Skip cURL and use this

Index Patterns
Manage the index patterns

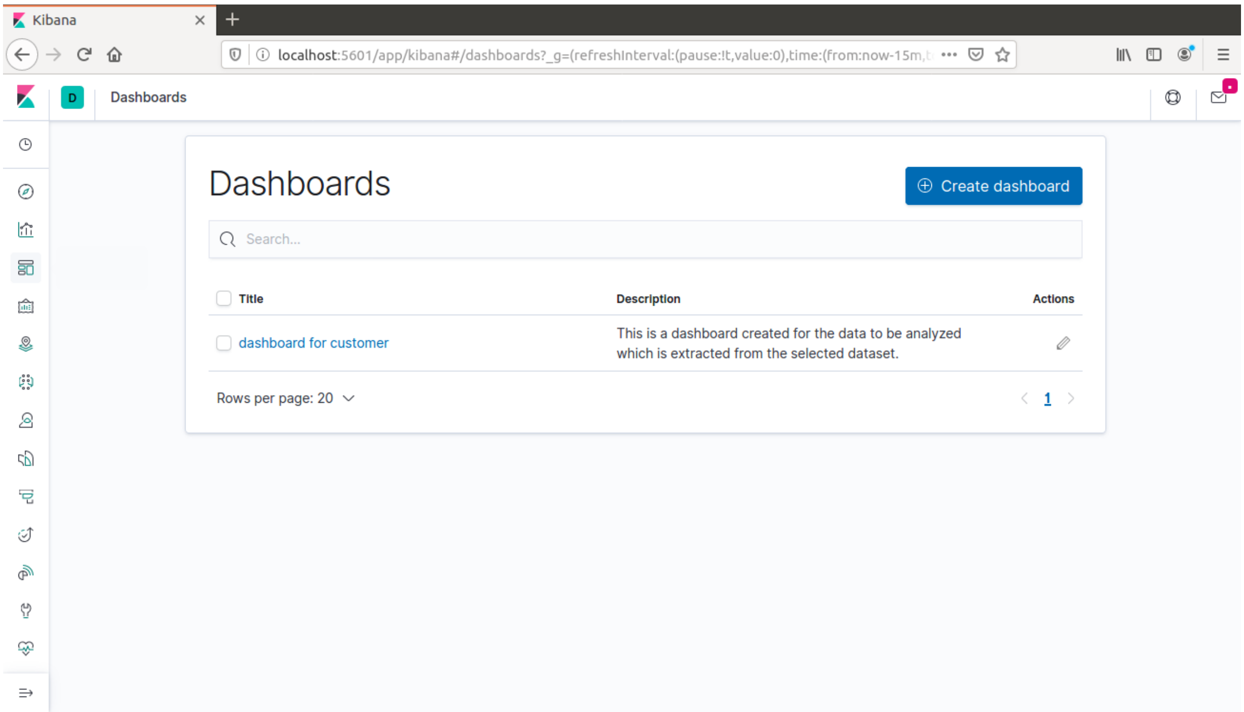
- Visualizations Page:


Kibana
localhost:5601/app/kibana#/visualize?_g=(refreshInterval:(pause:!t,value:0),time:(from:now-15m,to:n
D
Visualize
Create visualization
Search...

Visualizations

☐ Title	Type	Description	Actions
☐ brand-time	Pie		✎
☐ discounted price	Tag Cloud		✎
☐ is product advantageous for flipkart or no as per product_id	Data Table	Product_id of 20 products is mentioned in this table and accordingly stated whether it is advantageous to flipkart or no.	✎
☐ retail_price vs product_rating	Heat Map	Retail Price per Product Rating is given.	✎
☐ total entries of the customers	Metric		✎
☐ uniq_id per retail_price	Goal		✎
☐ uniq_id vs brand	Pie	This chart represents the last 10 customer's unique id and what brand is purchased by them. The inner diagram represents customer's id and the outer diagram represents the brand purchased by that specific	✎

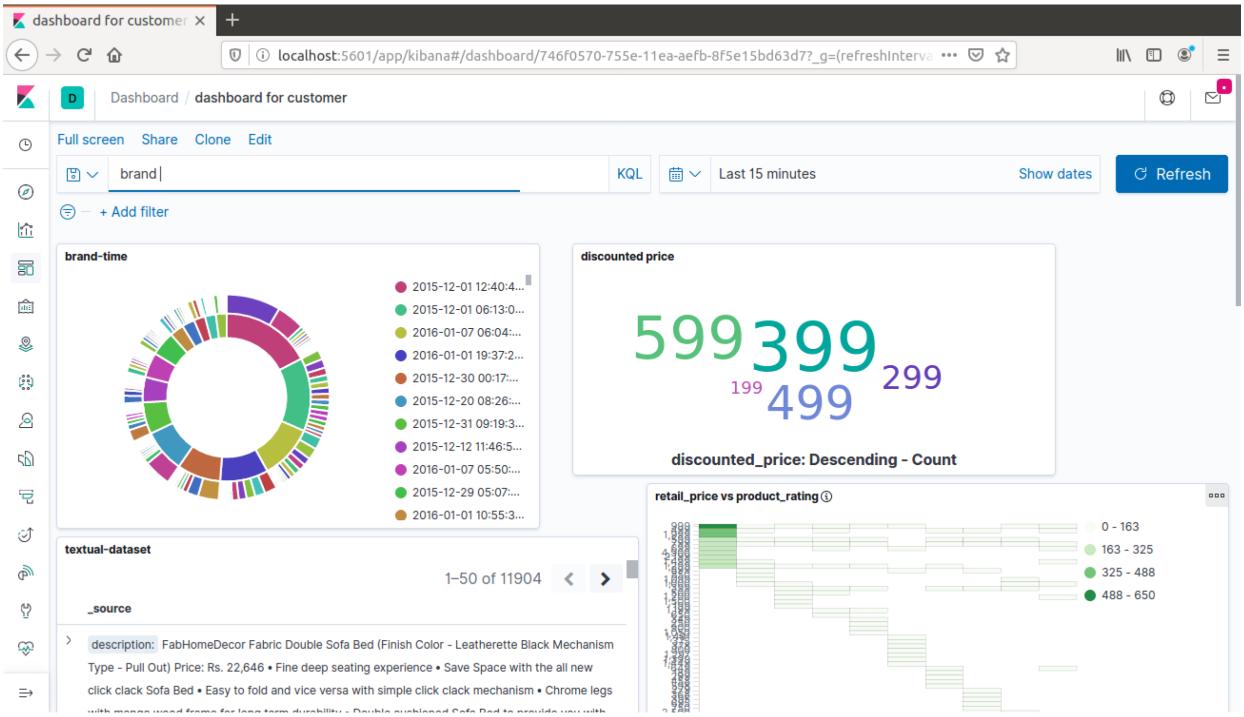
- Dashboard:


Kibana
localhost:5601/app/kibana#/dashboards?_g=(refreshInterval:(pause:!t,value:0),time:(from:now-15m,t
D
Dashboards
Create dashboard
Search...
Rows per page: 20
1

Dashboards

Title	Description	Actions
dashboard for customer	This is a dashboard created for the data to be analyzed which is extracted from the selected dataset.	

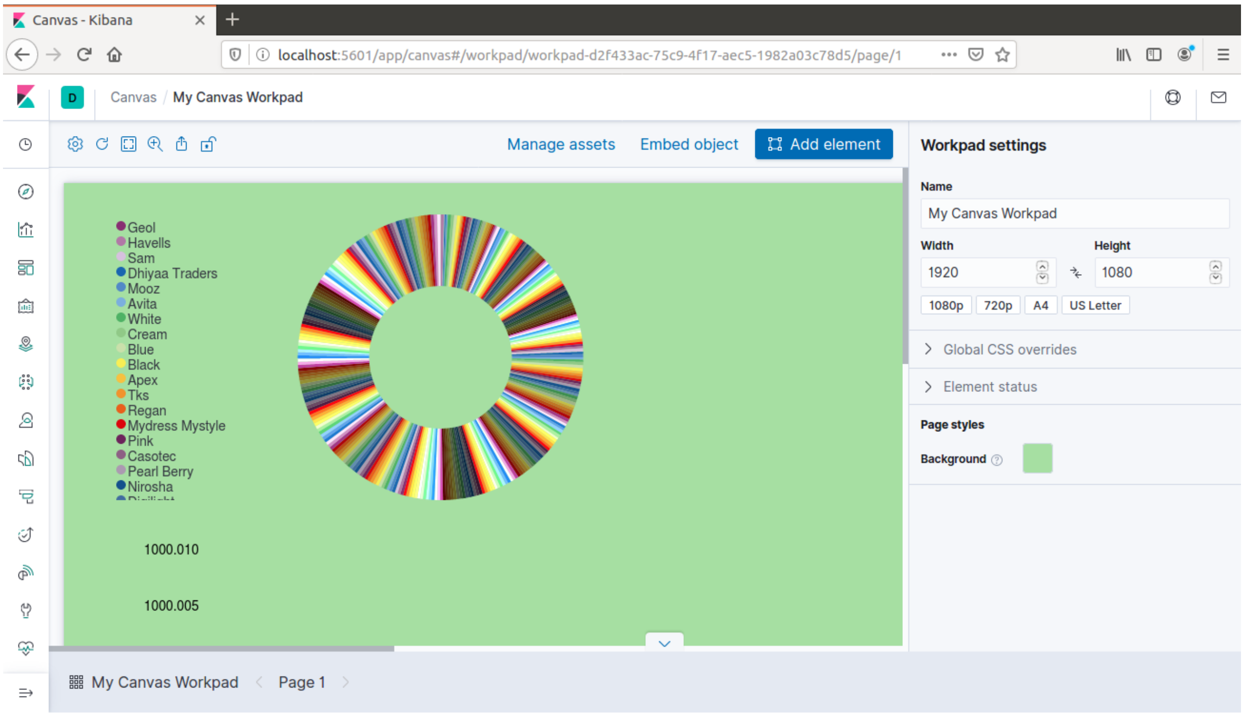
- Dashboard:



brand

description: FabHomeDecor Fabric Double Sofa Bed (Finish Color - Leatherette Black Mechanism Type - Pull Out) Price: Rs. 22,646 • Fine deep seating experience • Save Space with the all new click clack Sofa Bed • Easy to fold and vice versa with simple click clack mechanism • Chrome legs

- Canvas:



Workpad settings

Name: My Canvas Workpad

Width: 1920

Height: 1080

1080p | 720p | A4 | US Letter

Global CSS overrides

Element status

Page styles

Background