CE706 - Information Retrieval

2005569

Introduction:

COVID-19 is an infectious disease impacting all over the world. Clinicians and researchers continuously are searching for reliable data on disease and its impact . The COVID-19 data would be retrieved using Elasticsearch . Elastic search is an open-source, free and also a search engine for all types of data such as structured and unstructured and geospatial. Elastic search was introduced in 2010 and was built on Apache Lucene. Mainly , Elastic search is a database that is used to manage ,retrieve semi-structured , documented-oriented data and store the data . Using dataset from https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge) for performing all the operations.

To start,

Version and tool required -

Name	Version or Tool used
OS	Ubuntu Linux&Windows (Postman, Chrome)
ElasticSearch	7.10.2
User Access	Root user or should have sudio access
Logstash	7.10
Browser	Chrome
RESTFUL	Postman
JDK	Latest Version

Data Flow :

```
DataSource(Metadata.csv) → Logstash → ElasticSearch → Postman
                                                     → Browser
```
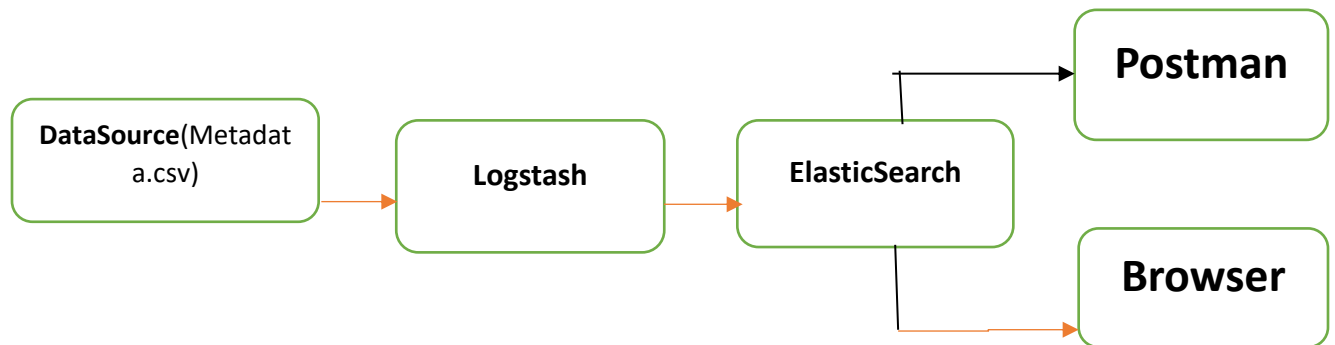
Installation & Configuration:

Download and install the Debian package manually

```
root@md2cd4qc:/etc# wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.10.2-amd64.deb
--2021-02-07 23:24:22--  https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.10.2-amd64.deb
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: 318852724 (304M) [application/octet-stream]
Saving to: 'elasticsearch-7.10.2-amd64.deb'

elasticsearch-7.10.2-amd64.deb          100%[===================================================================================>] 304.08M  2.15MB/s    in 2m 24s

2021-02-07 23:26:47 (2.11 MB/s) - 'elasticsearch-7.10.2-amd64.deb' saved [318852724/318852724]

root@md2cd4qc:/etc# wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.10.2-amd64.deb.sha512
--2021-02-07 23:28:06--  https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.10.2-amd64.deb.sha512
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: 160 [application/octet-stream]
Saving to: 'elasticsearch-7.10.2-amd64.deb.sha512'

elasticsearch-7.10.2-amd64.deb.sha512   100%[===================================================================================>]     160  --.-KB/s    in 0s

2021-02-07 23:28:07 (4.36 MB/s) - 'elasticsearch-7.10.2-amd64.deb.sha512' saved [160/160]

root@md2cd4qc:/etc# shasum -a 512 -c elasticsearch-7.10.2-amd64.deb.sha512
elasticsearch-7.10.2-amd64.deb: OK
root@md2cd4qc:/etc# sudo dpkg -i elasticsearch-7.10.2-amd64.deb
Selecting previously unselected package elasticsearch.
(Reading database ... 62349 files and directories currently installed.)
Preparing to unpack elasticsearch-7.10.2-amd64.deb ...
Creating elasticsearch group... OK
Creating elasticsearch user... OK
Unpacking elasticsearch (7.10.2) ...
Setting up elasticsearch (7.10.2) ...
System has not been booted with systemd as init system (PID 1). Can't operate.
Created elasticsearch keystore in /etc/elasticsearch/elasticsearch.keystore
Processing triggers for ureadahead (0.100.0-21) ...
Processing triggers for systemd (237-3ubuntu10.39) ...
```
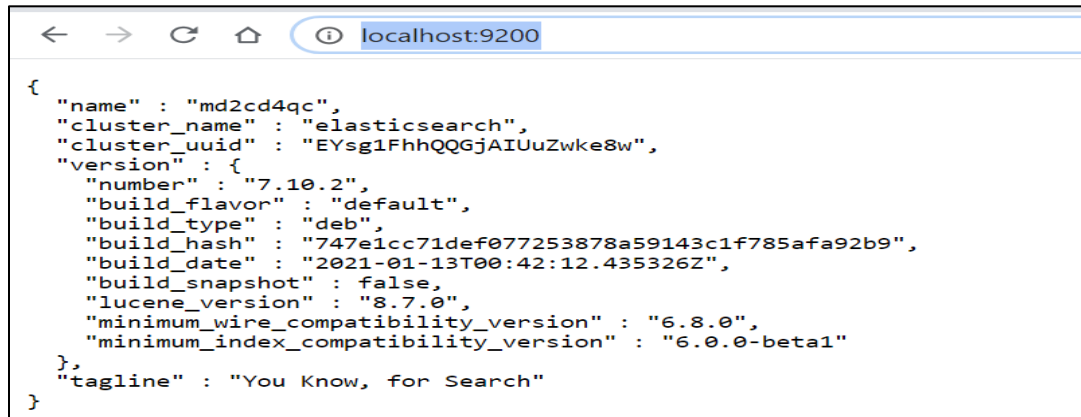
Starting the Elasticsearch

Command: Service elasticsearch start

```
root@md2cd4qc:/etc# service elasticsearch status
 * elasticsearch is not running
root@md2cd4qc:/etc# service elasticsearch start
 * Starting Elasticsearch Server                                                                    [ OK ]
```

Validating the elasticsearch

```
←  →  C  ⌂    ⓘ  localhost:9200

{
  "name" : "md2cd4qc",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "EYsg1FhhQQGjAIUuZwke8w",
  "version" : {
    "number" : "7.10.2",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "747e1cc71def077253878a59143c1f785afa92b9",
    "build_date" : "2021-01-13T00:42:12.435326Z",
    "build_snapshot" : false,
    "lucene_version" : "8.7.0",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
```

Indexing the CSV data into elasticsearch using Logstash

Installation:

```
root@md2cd4qc:/etc# echo "deb https://artifacts.elastic.co/packages/7.x/apt stable main" | sudo tee -a /etc/apt/sources.list.d/elastic-7.x.list
deb https://artifacts.elastic.co/packages/7.x/apt stable main
root@md2cd4qc:/etc# sudo apt-get update
Hit:1 http://nginx.org/packages/mainline/ubuntu bionic InRelease
Hit:2 http://archive.ubuntu.com/ubuntu bionic InRelease
Hit:3 http://security.ubuntu.com/ubuntu bionic-security InRelease
Get:4 http://ppa.launchpad.net/ansible/ansible/ubuntu bionic InRelease [15.9 kB]
Hit:5 http://archive.ubuntu.com/ubuntu bionic-updates InRelease
Hit:6 http://archive.ubuntu.com/ubuntu bionic-backports InRelease
Ign:7 https://download.arangodb.com/9c169fe900ff79790395784287bfa82f0dc0059375a34a2881b9b745c8efd42e/arangodb35/DEBIAN  InRelease
Hit:8 https://download.arangodb.com/9c169fe900ff79790395784287bfa82f0dc0059375a34a2881b9b745c8efd42e/arangodb35/DEBIAN  Release
Get:9 https://artifacts.elastic.co/packages/7.x/apt stable InRelease [10.4 kB]
Get:10 http://ppa.launchpad.net/ansible/ansible/ubuntu bionic/main amd64 Packages [820 B]
Get:12 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 Packages [53.6 kB]
Fetched 80.6 kB in 11s (7361 B/s)
Reading package lists... Done
W: Target Packages (main/binary-amd64/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target Packages (main/binary-all/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target Translations (main/i18n/Translation-en) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target CNF (main/cnf/Commands-amd64) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target CNF (main/cnf/Commands-all) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target Packages (main/binary-amd64/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target Packages (main/binary-all/Packages) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target Translations (main/i18n/Translation-en) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target CNF (main/cnf/Commands-amd64) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
W: Target CNF (main/cnf/Commands-all) is configured multiple times in /etc/apt/sources.list.d/elastic-7.x.list:1 and /etc/apt/sources.list.d/elastic-7.x.list:2
```

```
root@md2cd4qc:/etc# sudo apt-get install logstash
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  logstash
0 upgraded, 1 newly installed, 0 to remove and 179 not upgraded.
Need to get 353 MB of archives.
After this operation, 613 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 logstash amd64 1:7.10.2-1 [353 MB]
Fetched 353 MB in 3min 10s (1856 kB/s)
Selecting previously unselected package logstash.
(Reading database ... 63401 files and directories currently installed.)
Preparing to unpack .../logstash_1%3a7.10.2-1_amd64.deb ...
Unpacking logstash (1:7.10.2-1) ...


Setting up logstash (1:7.10.2-1) ...
Using bundled JDK: /usr/share/logstash/jdk
Using provided startup.options file: /etc/logstash/startup.options
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC was deprecated in version 9.0 and will likely be removed in a future release.
```

Navigated to path: cd /etc , cd logstash/

Selecting   first 1500 entries from metadata.csv

	A	B	C	D	E	F	G	H	
1	cord_uid	sha	source_x	title	doi	pmcid	pubmed_i	license	abstract
2	ug7v899j	d1aafb70c	PMC	Clinical fe	10.1186/1471-2334-1-6	PMC35282	11472636	no-cc	OBJECTIVE: This retrospective chart review describes the epidemiology and clinical fe
3	02tnwd4m	6b056772!	PMC	Nitric oxid	10.1186/rr14	PMC59543	11667967	no-cc	Inflammatory diseases of the respiratory tract are commonly associated with elevate
4	ejv2xln0	06ced00a!	PMC	Surfactant	10.1186/rr19	PMC59549	11667972	no-cc	Surfactant protein-D (SP-D) participates in the innate response to inhaled microorgar
5	2b73a28n	34805564!	PMC	Role of en	10.1186/rr44	PMC59574	11686871	no-cc	Endothelin-1 (ET-1) is a 21 amino acid peptide with diverse biological activity that has
6	9785vg6d	5f48792a5	PMC	Gene expr	10.1186/rr61	PMC59580	11686888	no-cc	Respiratory syncytial virus (RSV) and pneumonia virus of mice (PVM) are viruses of th
7	zjufx4fo	b2897e12'	PMC	Sequence	10.1093/emboj/20.24.7220	PMC125340	11742998	green-oa	Nidovirus subgenomic mRNAs contain a leader sequence derived from the 5â€² end (
8	5yhe786e	3bb07ea1!	PMC	Debate: T	10.1186/cc987	PMC137267	11299062	no-cc	Recent evidence suggests that critically ill patients are able to tolerate lower levels of
9	8zchiykl	5806726a;	PMC	The 21st I	10.1186/cc1013	PMC137274	11353930	no-cc	The 21st International Symposium on Intensive Care and Emergency Medicine was d
10	8qnrcgnk	faaf1022c	PMC	Heme oxy	10.1186/1465-9921-4-7	PMC193681	12964953	no-cc	Heme oxygenase-1 (HO-1), an inducible stress protein, confers cytoprotection agains
11	jg13scgo	5b44feca5	PMC	Technical	10.1197/jamia.m1345	PMC212776	12807803	bronze-oa	This report describes the design and implementation of the Real-time Outbreak and I
12	5tkvsudh	9d4e3e8el	PMC	Conservat	10.1093/emboj/19.8.1907	PMC302018	10775274	no-cc	Regulation of ornithine decarboxylase in vertebrates involves a negative feedback me
13	6lvn10f4	14e0cac6e	PMC	Heteroger	10.1093/emboj/19.17.4701	PMC302072	10970862	green-oa	Heterogeneous nuclear ribonucleoprotein (hnRNP A1) is involved in pre-mRNA splicir
14	tvxpckxo	d09b7902(	PMC	A Method	10.1251/bpo66	PMC302190	14702098	no-cc	The UBA domain is a conserved sequence motif among polyubiquitin binding protein
15	mcuixluu	44102e3e(	PMC	Vaccinia vi	10.1093/emboj/19.15.3932	PMC306617	10921875	green-oa	We examined the role of the microtubule cytoskeleton during vaccinia virus infection
16	6iu1dtvl	6e8517ch;	PMC	The site of	10.1186/1479-5876-2-3	PMC340389	14733617	no-cc	

Copying the metadata.csv file to logstash path /etc/logstash

Furthermore , we need to update the logstah_1.conf file with the columns & file name

```
1   # Sample Logstash configuration for creating a simple
2   # Beats -> Logstash -> Elasticsearch pipeline.
3
4   input {
5     file {
6       path => "/etc/logstash/metadata.csv"
7       start_position => "beginning"
8       sincedb_path => "/dev/null"
9     }
10  }
11
12  filter {
13    csv {
14        separator => ","
15        skip_header => "true"
16        columns =>
          ["covid_uid","Sha","source_x","title","doi","pmcid","pubmed_id","license","publish_time","authors","journal","mag_id","who_covidence_id","arxiv_id"
          ,"pdf_json_files","pmc_json_files","url","s2_id"]
17    }
18  }
19  output {
20    elasticsearch {
21      hosts => ["http://localhost:9200"]
22      index => "covid"
23      #user => "elastic"
24      #password => "changeme"
25    }
26  }
```

Modify the file with input, filter & output

```
root@md2cd4qc:/usr/share/logstash/bin# cd /usr/share/logstash/bin/
```

```
root@md2cd4qc:/etc/logstash# cd /usr/share/logstash/bin/
root@md2cd4qc:/usr/share/logstash/bin# ./logstash -f /etc/logstash/logstash_1.conf
Using bundled JDK: /usr/share/logstash/jdk
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC was deprecated in version 9.0 and will likely be removed in a future release.
```

Output:

```
        "_source" : {
            "covid_uid" : "02tnwd4m",
            "source_x" : "PMC",
            "publish_time" : "Inflammatory diseases of the respiratory tract are commonly associated with elevated production of
-dependent oxidative stress. Although NO• is known to have anti-microbial, anti-inflammatory and anti-oxidant properties, vario
of NO• to lung injury in several disease models. On the basis of biochemical evidence, it is often presumed that such NO• -depe
the oxidant peroxynitrite, although alternative mechanisms involving the phagocyte-derived heme proteins myeloperoxidase and eo
conditions of inflammation. Because of the overwhelming literature on NO• generation and activities in the respiratory tract, i
to review this area comprehensively. Instead, it focuses on recent evidence and concepts of the presumed contribution of NO• to
            "message" : "02tnwd4m,6b0567729c2143a66d737eb0a2f63f2dce2e5a7d,PMC,Nitric oxide: a pro-inflammatory mediator in lung
cc,\"Inflammatory diseases of the respiratory tract are commonly associated with elevated production of nitric oxide (NO•) and
stress. Although NO• is known to have anti-microbial, anti-inflammatory and anti-oxidant properties, various lines of evidence
in several disease models. On the basis of biochemical evidence, it is often presumed that such NO• -dependent oxidations are c
peroxynitrite, although alternative mechanisms involving the phagocyte-derived heme proteins myeloperoxidase and eosinophil per
inflammation. Because of the overwhelming literature on NO• generation and activities in the respiratory tract, it would be bey
area comprehensively. Instead, it focuses on recent evidence and concepts of the presumed contribution of NO• to inflammatory c
Albert van der; Eiserich, Jason P; Cross, Carroll E\",Respir
Res,,,,document_parses/pdf_json/6b0567729c2143a66d737eb0a2f63f2dce2e5a7d.json,document_parses/pmc_json/PMC59543.xml.json,https:
r",
            "mag_id" : "Respir Res",
            "title" : "Nitric oxide: a pro-inflammatory mediator in lung disease?",
            "path" : "/etc/logstash/metadata.csv",
            "@version" : "1",
            "pdf_json_files" : null,
            "journal" : "Vliet, Albert van der; Eiserich, Jason P; Cross, Carroll E",
            "Sha" : "6b0567729c2143a66d737eb0a2f63f2dce2e5a7d",
            "authors" : "2000-08-15",
            "license" : "no-cc",
            "arxiv_id" : null,
            "doi" : "10.1186/rr14",
            "pmcid" : "PMC59543",
            "pmc_json_files" : "document_parses/pdf_json/6b0567729c2143a66d737eb0a2f63f2dce2e5a7d.json",
            "column19" : null,
            "@timestamp" : "2021-02-12T02:54:49.883Z".
```

Again, starting the logstash before that we have to delete data and index created

Using postman to delete the index

```
DELETE    ▼    http://localhost:9200/covid/

Params    Authorization    Headers (6)    Body    Pre-request Script    Tests    Settings

Body    Cookies    Headers (3)    Test Results                          200 OK    3

Pretty    Raw    Preview    Visualize    JSON ▼    ⇉

1  {
2      "acknowledged": true
3  }
```

```
root@md2cd4qc:/usr/share/logstash/bin# cd /usr/share/logstash/bin/
```

```
root@md2cd4qc:/etc/logstash# cd /usr/share/logstash/bin/
root@md2cd4qc:/usr/share/logstash/bin# ./logstash -f /etc/logstash/logstash_1.conf
Using bundled JDK: /usr/share/logstash/jdk
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC was deprecated in version 9.0 and will likely be removed in a future release.
```
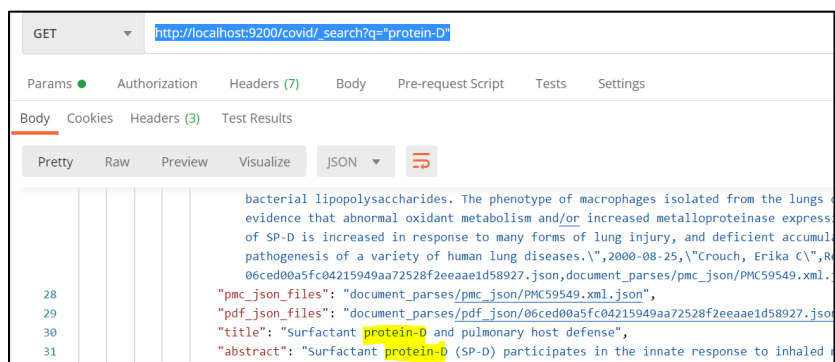
Using Postman updated the classic tokenizer, stopwords& stemming section

Copying the content from covid_mapping and pasting it to the Body tab then run it.

With the updated mapping file, deleting the existing index and creating the new index
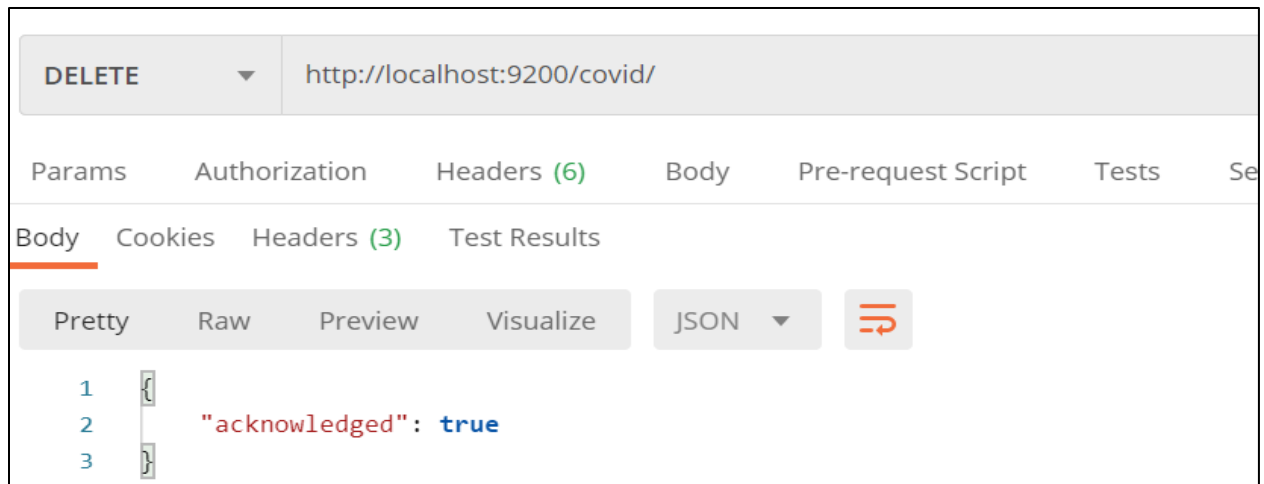
Running logstash to get data with all the customization used :

Tokenizer: Classic Tokenizer→ It is a good tokenizer for the English language which splits the words based on hyphens.

Filters: Stop words→ It removes the undesirable and useless words that are not relevant to the search index for example:

a, are, for, if, in, into, is, it, not, on, or, such, that, they, this, to, was, will etc.

Filters: Stemming → It indexes the data with a different variant of the data.

Tokenizer to Analyzer :
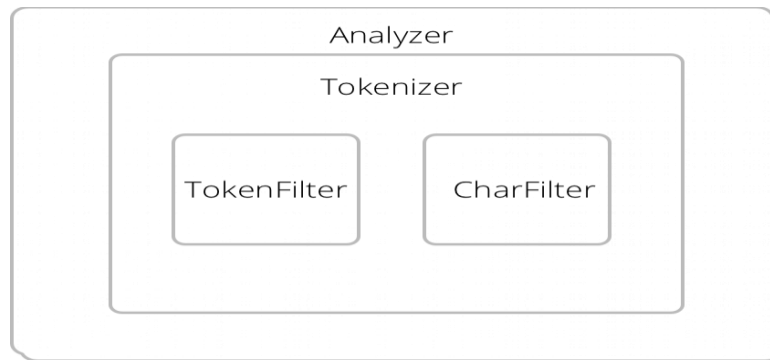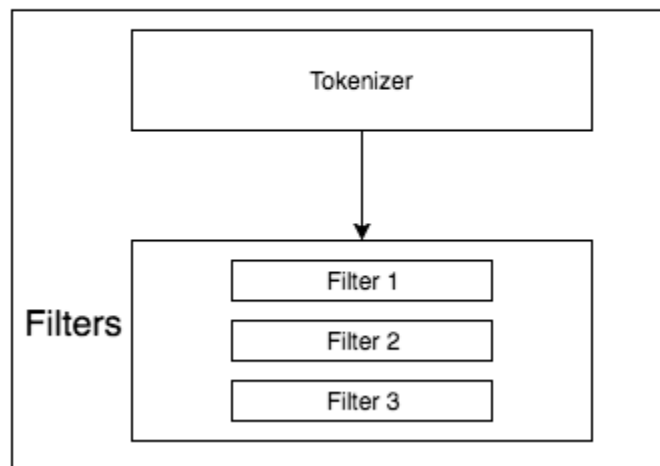
Flow diagram Analyze the sentence using analyzer:
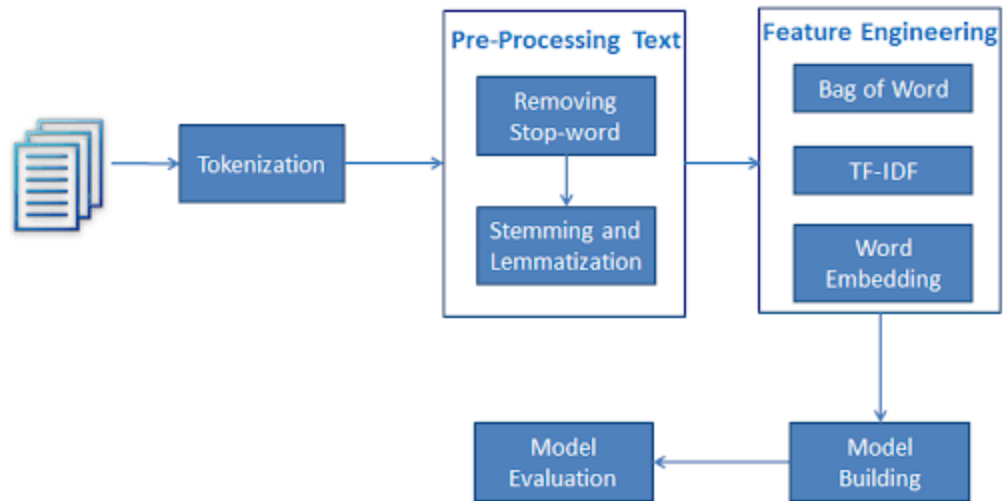
The screenshot below shows the section for different section for Tokenizer & filters

```
 4        "settings": {
 5          "analysis": {
 6            "analyzer": {
 7              "my_analyzer": {
 8                "tokenizer": "my_tokenizer"
 9                "filter": [ "stop","stemmer" ]
10              }
11            },
12            "tokenizer": {
13              "my_tokenizer": {
14                "type": "classic",
15                "max_token_length": 5
16              }
17            }
```
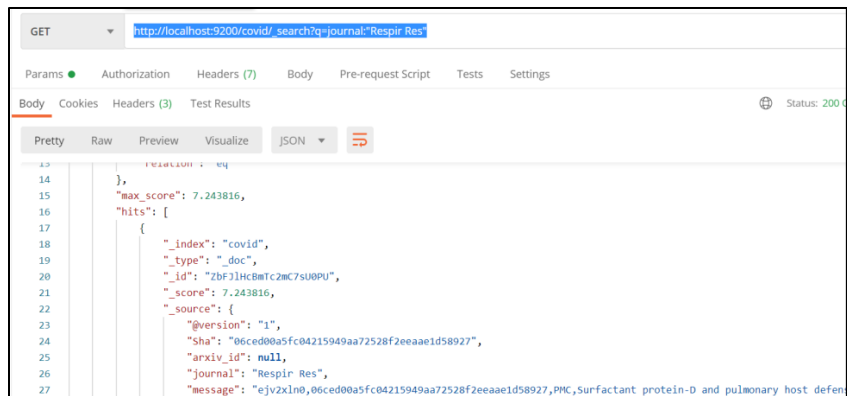
Searching through Postman:

1. Searching the entire content indexed using the logstash

2. Searching the field **journal** with value **Respir Res**

3. Searching fields with value **protein-D**

4. Searching documents using a wildcard with value**alveo**

Download the files from https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

In that csv files in 1_population:

- Management of patients who are underhoused otherwise lower social-economic status.csv

- Methods to reach marginalized and disadvantaged population.csv

- Measures to control the spread in communities.csv

- Means of communication with target high-risk population.csv

- What are recommendations for combating resource failures_.csv

- Different ways to create hospital infrastructure to prevent nosocomial outbreaks_.csv

Copying all of the above files into Covid folder

```
root@md2cd4qc:/etc/logstash# ls
Covid    input        log4j2.properties    logstash-wiki-movies.conf  pipelines.yml    wiki_movie_plots_deduped.csv
conf.d   jvm.options  logstash-sample.conf  logstash.yml               startup.options
```

Now creating five logstash configs for each .csv file

```
root@md2cd4qc:/etc/logstash# mv logstash-wiki-movies.conf logstash_1.conf
root@md2cd4qc:/etc/logstash# cp logstash_1.conf logstash_2.conf
root@md2cd4qc:/etc/logstash# cp logstash_1.conf logstash_3.conf
root@md2cd4qc:/etc/logstash# cp logstash_1.conf logstash_4.conf
root@md2cd4qc:/etc/logstash# cp logstash_1.conf logstash_5.conf
root@md2cd4qc:/etc/logstash# ls
Covid    input        log4j2.properties    logstash.yml     logstash_2.conf  logstash_4.conf  pipelines.yml    wiki_movie_plots_deduped.csv
conf.d   jvm.options  logstash-sample.conf  logstash_1.conf  logstash_3.conf  logstash_5.conf  startup.options
```

Updated the each logstash .conf file with respective csv file and fields

Updated the each csv file with column names =>

["Date","Study","Study_Link","Journal","Study_Type","Addressed_Population","Challenge","Solution","Evidence","Added_on"]

	A	B	C	D	E	F	G	H	I	J	K
1		Date	Study	Study_Link	Journal	Study_Typ	Addressed	Challenge	Solution	Evidence	Added_on
2		0	########	COVID-19	https://do	medRxiv	Retrospec	Homeless	Evolving C	Universal	patients: 4	########

D1 fx Study_Link

```
 1  # Sample Logstash configuration for creating a simple
 2  # Beats -> Logstash -> Elasticsearch pipeline.
 3
 4  input {
 5    file {
 6      path => "/etc/logstash/Covid/Management of patients who are underhoused or otherwise lower social economic status.csv"
 7      start_position => "beginning"
 8      sincedb_path => "/dev/null"
 9    }
10  }
11
12  filter {
13    csv {
14      separator => ","
15      skip_header => "true"
16      columns => ["Date","Study","Study_Link","Journal","Study_Type","Addressed_Population","Challenge","Solution","Evidence","Added_on"]
17    }
18  }
19  output {
20    elasticsearch {
21      hosts => ["http://localhost:9200"]
22      index => "Covid"
23      #user => "elastic"
24      #password => "changeme"
25    }
26  }
27
```

Starting the Elastic using the below command.

```
root@md2cd4qc:/usr/share/logstash/bin# ^C
root@md2cd4qc:/usr/share/logstash/bin# service elasticsearch status
 * elasticsearch is not running
root@md2cd4qc:/usr/share/logstash/bin# service elasticsearch start
 * Starting Elasticsearch Server
root@md2cd4qc:/usr/share/logstash/bin# ./logstash -f /etc/logstash/logstash_1.conf
Using bundled JDK: /usr/share/logstash/jdk
```

Index name should to be small case

```
output {
   elasticsearch {
      hosts => ["http://localhost:9200"]
      index => "covid"
      #user => "elastic"
      #password => "changeme"
   }
}
```

All the documents are indexed properly

localhost:9200/covid/_search/?pretty=true

```
      "failed" : 0
    },
    "hits" : {
      "total" : {
        "value" : 18,
        "relation" : "eq"
      },
      "max_score" : 1.0,
      "hits" : [
        {
          "_index" : "covid",
          "_type" : "_doc",
          "_id" : "_rGlkXcBmTc2mC7spzAV",
          "_score" : 1.0,
          "_source" : {
            "Study" : "2020-04-02",
            "@timestamp" : "2021-02-11T15:11:27.928Z",
            "host" : "md2cd4qc",
            "column11" : "4/22/2020",
            "Date" : "7",
            "message" : "7,2020-04-02,Flattening the Curve for Incarcerated Populations — Covid-19 in Jails and Pris
England Journal of Medicine,Expert Review,Incarcerated,\"People entering jails are among the most vulnerable in ou
exacerbated by restricted movement, confined spaces, and limited medical care.\",\"We believe in decarcerating, fo
crimes, but also on the elderly and infirm. Also urging police and courts to immediately suspend arresting and sen
and misdemeanors.\",-,4/22/2020",
            "Journal" : "https://www.nejm.org/doi/full/10.1056/NEJMp2005687",
            "Challenge" : "Incarcerated",
            "Solution" : "People entering jails are among the most vulnerable in our society, and during incarcerati
movement, confined spaces, and limited medical care.",
            "Addressed_Population" : "Expert Review",
            "@version" : "1",
            "Study_Link" : "Flattening the Curve for Incarcerated Populations — Covid-19 in Jails and Prisons",
            "Evidence" : "We believe in decarcerating, focusing on those who are least likely to commit additional c
police and courts to immediately suspend arresting and sentencing people, as much as possible, for low-level crime
            "Added_on" : "-",
            "path" : "/etc/logstash/Covid/Management of patients who are underhoused or otherwise lower social econo
            "Study_Type" : "New England Journal of Medicine"
          }
```
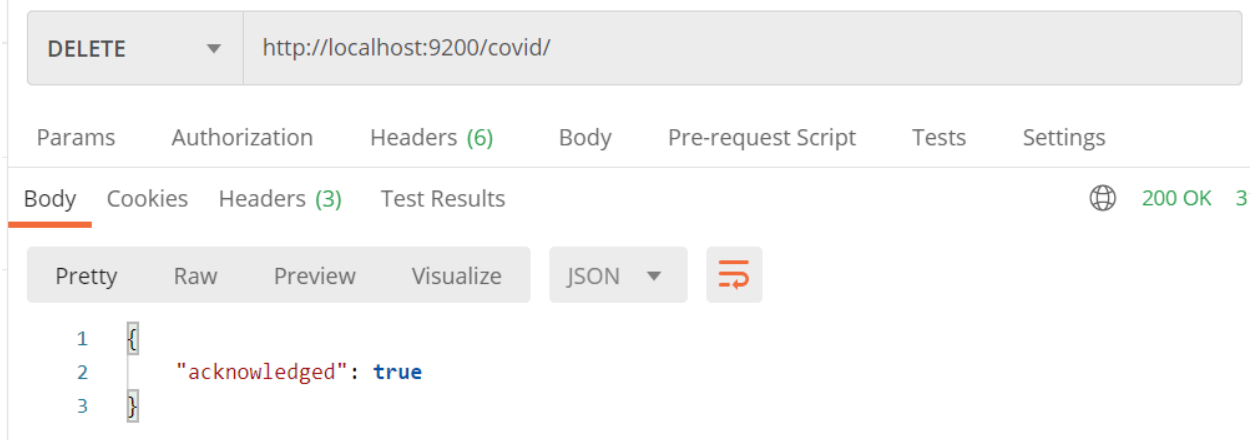
Using RESTFul Postman for checking the results

Copying the content from covid_mapping then paste it in Body tab then running the URL.
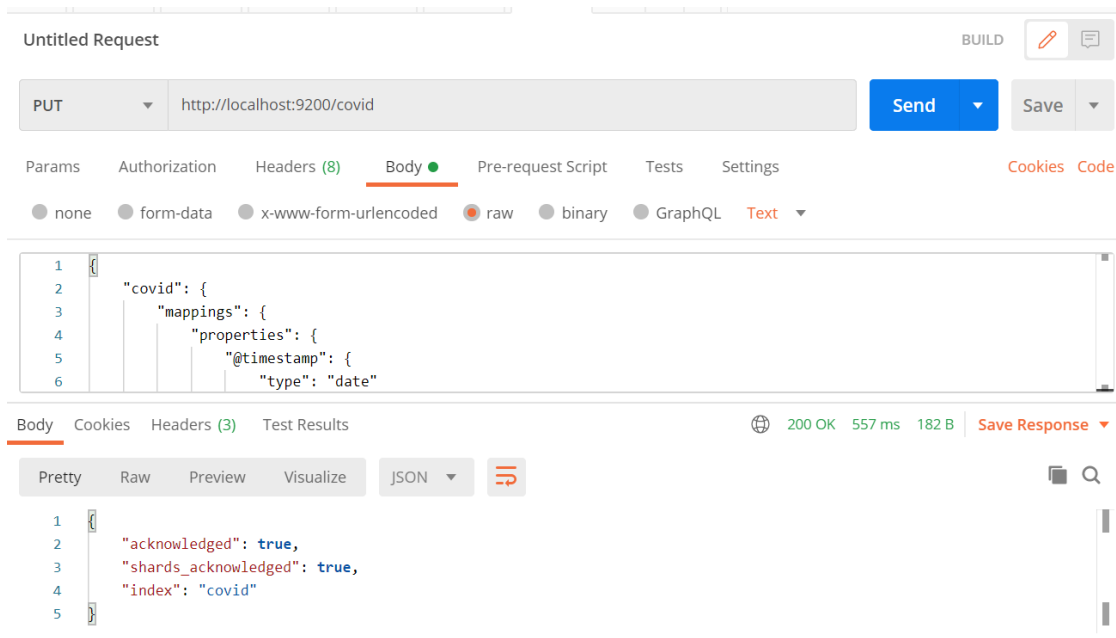
Now the index is created with proper mapping for date & added on

Once completed ,Run the command again to get the fresh data into Elasticsearch

```
root@md2cd4qc:/usr/share/logstash/bin# ./logstash -f /etc/logstash/logstash_1.conf
Using bundled JDK: /usr/share/logstash/jdk
```

Looks like csv file empty first column without column name for all the CSV file so I have removed it in all the files

After removing it looks like:

	A	B	C	D	E	F	G	H	I	J
1	Date	Study	Study_Lin	Journal	Study_Typ	Addressed	Challenge	Solution	Evidence	Added_on
2	########	COVID-19	https://dc	medRxiv	Retrospec	Homeless	Evolving C	Universal	patients: 4	########
3	########	Multidisci	https://dc	The Lance	Expert Rev	Psychiatric	Impact on	Multidisci	-	########
4	########	U.S. count	https://dc	MedRxiv	Simulatior	Poverty ar	Not all Am	Both the f	countries:	########
5	########	Epidemiol	http://hdl	Annals of	Case Serie	Homeless	People exp	Municipal	patients: 1	########
6	4/7/2020	The Challe	https://jar	Journal of	Editorial	Incarcerat	Social dist	Releasing	-	########
7	4/6/2020	Mitigating	https://dc	Addiction	Expert Rev	Substance	The novel	There is ar	-	########
8	4/3/2020	Addressin	https://jar	Journal of	Expert Rev	Psychiatric	Disasters (	Expand ar	-	########
9	4/2/2020	Flattening	https://wv	New Engla	Expert Rev	Incarcerat	People en	We believe	-	########
10	4/2/2020	Fighting cc	https://dc	Boston Me	Expert Rev	Incarcerat	Poor priso	Health edu	-	########
11	########	Addressin	http://uds	Center for	Expert Rev	Homeless	The econc	Income as	-	########
12	########	Estimated	https://es	UCLA: Can	Simulatior	Homeless	The rapid	The currer	cities: New	########
13	########	Respondir	https://dc	The Hastir	Expert Rev	Vulnerable	When the	Decisions	-	########
14	########	Efforts esc	https://dc	The Lance	Editorial	Homeless	Homeless	UK govern	-	########
15	########	Covid-19:	https://dc	Boston Me	Expert Rev	Vulnerable	Failure to	Fair and e	-	########
16	########	Securing tl	https://jar	Journal of	Expert Rev	Vulnerable	COVID-19	Medicaid	-	########
17	########	Prisons an	https://dc	The Lance	Expert Rev	Incarcerat	Prisons ar	In order fc	-	########
18	########	The Coron	https://ne	Next City	Expert Rev	Homeless	But amid a	Prudent ac	-	5/9/2020
19	########	COVID-19:	https://dc	The Lance	Expert Rev	Homeless	It is proba	Testing kit	-	########

Now again delete the index and create index mapping file

DELETE	▼	http://localhost:9200/covid/

Params   Authorization   Headers (6)   Body   Pre-request Script   Tests   Se

Body   Cookies   Headers (3)   Test Results

Pretty   Raw   Preview   Visualize   JSON ▼

```
1  {
2      "acknowledged": true
3  }
```

PUT ▼	http://localhost:9200/covid

Params   Authorization   Headers (8)   **Body** ●   Pre-request Script   Tests   S

○ none   ○ form-data   ○ x-www-form-urlencoded   ● raw   ○ binary   ○ GraphQL

```
1  {
2      "covid": {
3          "mappings": {
4              "properties": {
5                  "@timestamp": {
6                      "type": "date"
```

Body   Cookies   Headers (3)   Test Results

Pretty   Raw   Preview   Visualize   JSON ▼

```
1  {
2      "acknowledged": true,
3      "shards_acknowledged": true,
4      "index": "covid"
```

Run the command again

```
root@md2cd4qc:/usr/share/logstash/bin# ./logstash -f /etc/logstash/logstash_1.conf
Using bundled JDK: /usr/share/logstash/jdk
```

localhost:9200/covid/_search/?pretty=true

```
    skipped : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 18,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "covid",
        "_type" : "_doc",
        "_id" : "TbHYkXcBmTc2mC7sBzFV",
        "_score" : 1.0,
        "_source" : {
          "@version" : "1",
          "Study_Type" : "Retrospective Observational Study",
          "Evidence" : "patients: 408",
          "Challenge" : "Evolving Cluster of COVID Cases in a Homeless Shelter in Boston",
          "Addressed_Population" : "Homeless",
          "message" : "4/15/2020,COVID-19 outbreak at a large homeless shelter in Boston: Implications for universal
testing,https://doi.org/10.1101/2020.04.12.20059618,medRxiv,Retrospective Observational Study,Homeless,Evolving Cluster of COVID Cases in a Homeless S
Boston,\"Universal PCR testing, rather than a symptom triggered approach, may be a better strategy for identifying and mitigating COVID-19 among peopl
homelessness\",patients: 408,4/22/2020\r",
          "Study" : "COVID-19 outbreak at a large homeless shelter in Boston: Implications for universal testing",
          "host" : "md2cd4qc",
          "Solution" : "Universal PCR testing, rather than a symptom triggered approach, may be a better strategy for identifying and mitigating COVID
experiencing homelessness",
          "Journal" : "medRxiv",
          "Added_on" : "4/22/2020",
          "path" : "/etc/logstash/Elastic/Management of patients who are underhoused or otherwise lower social economic status.csv",
          "Date" : "4/15/2020",
          "Study_Link" : "https://doi.org/10.1101/2020.04.12.20059618",
          "@timestamp" : "2021-02-11T16:06:29.588Z"
        }
      },
```

References

Gormley, C. and Tong, Z., 2015. *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine*. " O'Reilly Media, Inc.".

Hiscott, J., Alexandridi, M., Muscolini, M., Tassone, E., Palermo, E., Soultsioti, M. and Zevini, A., 2020. The global impact of the coronavirus pandemic. *Cytokine & growth factor reviews*, *53*, pp.1-9. , Kaggle.com, https://www.amazon.com/Elasticsearch-Definitive-Distributed-Real-Time-Analytics/dp/1449358543 ,
https://pubmed.ncbi.nlm.nih.gov/32487439/,https://www.ibm.com/cloud/learn/elasticsearch.