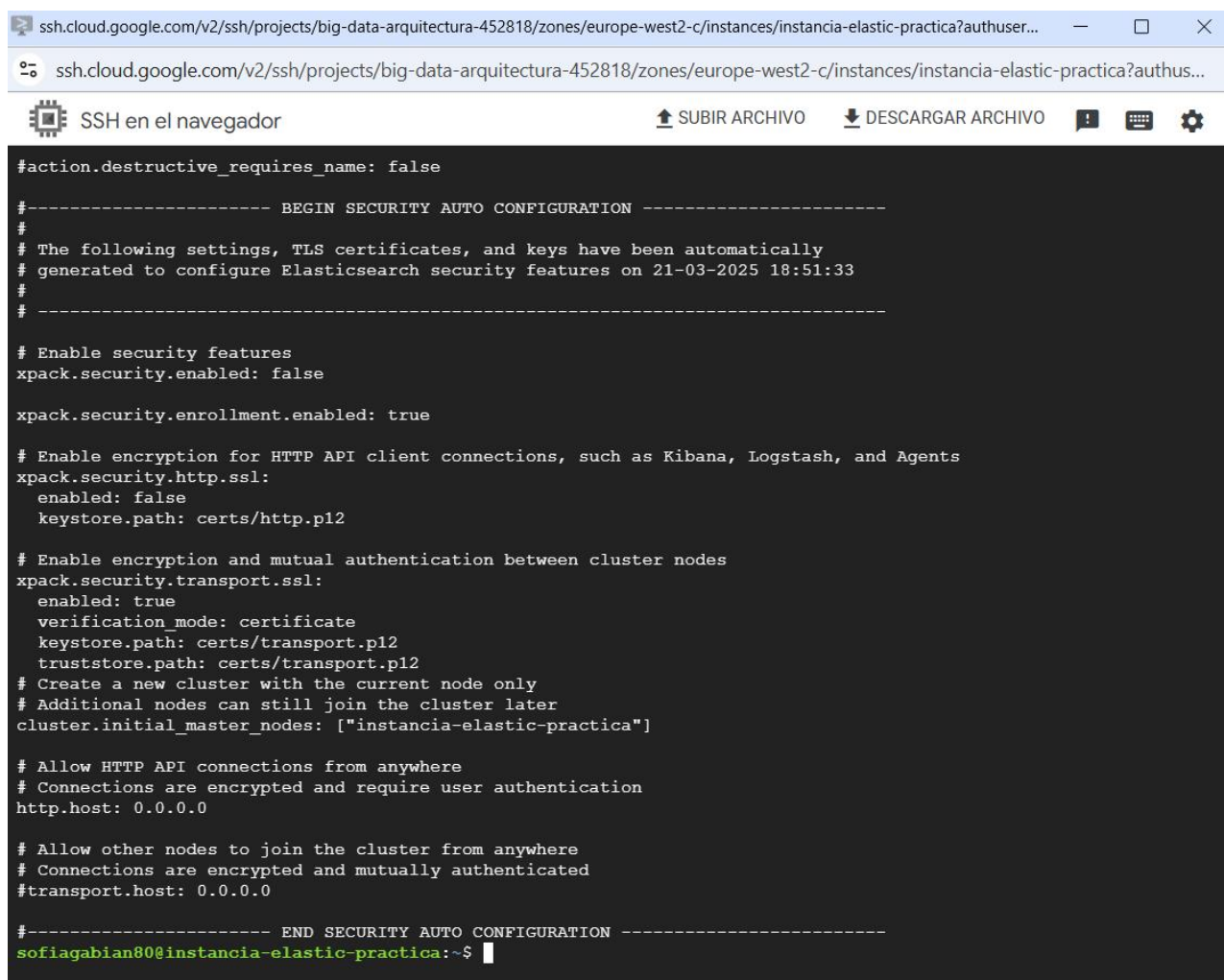


Práctica en el que conectaremos Hadoop con ElasticSearch. En concreto, haremos una conexión entre Hive y un índice (o index) de ElasticSearch.

**ENTREGABLE PARTE 1:** Captura de pantalla de la consola SSH del cluster Hadoop una vez finalizada la configuración y carga.

```
sofiagabian80@cluster-hadoop-practica-m:~$ gsutil cp gs://bucket-elastic-practica/elasticsearch-hadoop-8.14.1.jar .
Copying gs://bucket-elastic-practica/elasticsearch-hadoop-8.14.1.jar...
/ [1 files][ 2.1 MiB/ 2.1 MiB]
Operation completed over 1 objects/2.1 MiB.
sofiagabian80@cluster-hadoop-practica-m:~$ gsutil cp gs://bucket-elastic-practica/commons-httpclient-3.1.jar .
Copying gs://bucket-elastic-practica/commons-httpclient-3.1.jar...
/ [1 files][297.8 KiB/297.8 KiB]
Operation completed over 1 objects/297.8 KiB.
sofiagabian80@cluster-hadoop-practica-m:~$
```

**ENTREGABLE PARTE 2:** Captura de pantalla de la consola del server Elastic donde se vea la configuración de Elastic, desde 'Enable security features' hasta el final (el fichero elasticsearch.yml) .



```
#action.destructive_requires_name: false

#----- BEGIN SECURITY AUTO CONFIGURATION -----
#
# The following settings, TLS certificates, and keys have been automatically
# generated to configure Elasticsearch security features on 21-03-2025 18:51:33
#
# -----

# Enable security features
xpack.security.enabled: false

xpack.security.enrollment.enabled: true

# Enable encryption for HTTP API client connections, such as Kibana, Logstash, and Agents
xpack.security.http.ssl:
  enabled: false
  keystore.path: certs/http.p12

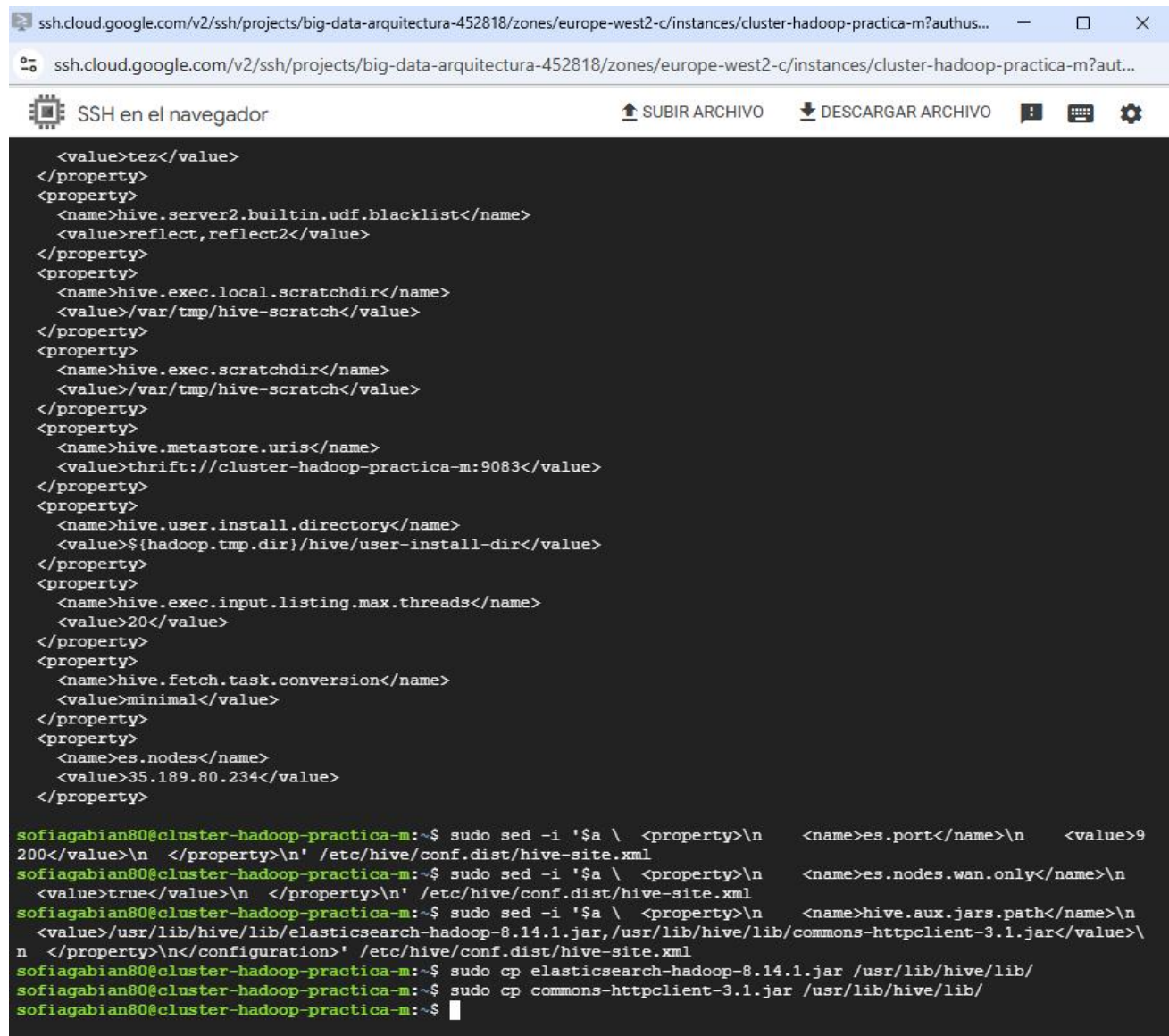
# Enable encryption and mutual authentication between cluster nodes
xpack.security.transport.ssl:
  enabled: true
  verification_mode: certificate
  keystore.path: certs/transport.p12
  truststore.path: certs/transport.p12
# Create a new cluster with the current node only
# Additional nodes can still join the cluster later
cluster.initial_master_nodes: ["instancia-elastic-practica"]

# Allow HTTP API connections from anywhere
# Connections are encrypted and require user authentication
http.host: 0.0.0.0

# Allow other nodes to join the cluster from anywhere
# Connections are encrypted and mutually authenticated
#transport.host: 0.0.0.0

#----- END SECURITY AUTO CONFIGURATION -----
sofiagabian80@instancia-elastic-practica:~$
```

### ENTREGABLE PARTE 3: Captura de pantalla del proceso de configuración en Cluster Hadoop de Conexión con ES completo.



```
ssh.cloud.google.com/v2/ssh/projects/big-data-arquitectura-452818/zones/europe-west2-c/instances/cluster-hadoop-practica-m?authus...  
ssh.cloud.google.com/v2/ssh/projects/big-data-arquitectura-452818/zones/europe-west2-c/instances/cluster-hadoop-practica-m?aut...  
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO  
<value>tez</value>  
</property>  
<property>  
  <name>hive.server2.builtin.udf.blacklist</name>  
  <value>reflect,reflect2</value>  
</property>  
<property>  
  <name>hive.exec.local.scratchdir</name>  
  <value>/var/tmp/hive-scratch</value>  
</property>  
<property>  
  <name>hive.exec.scratchdir</name>  
  <value>/var/tmp/hive-scratch</value>  
</property>  
<property>  
  <name>hive.metastore.uris</name>  
  <value>thrift://cluster-hadoop-practica-m:9083</value>  
</property>  
<property>  
  <name>hive.user.install.directory</name>  
  <value>${hadoop.tmp.dir}/hive/user-install-dir</value>  
</property>  
<property>  
  <name>hive.exec.input.listing.max.threads</name>  
  <value>20</value>  
</property>  
<property>  
  <name>hive.fetch.task.conversion</name>  
  <value>minimal</value>  
</property>  
<property>  
  <name>es.nodes</name>  
  <value>35.189.80.234</value>  
</property>  
  
sofiagabian80@cluster-hadoop-practica-m:~$ sudo sed -i '$a \ <property>\n <name>es.port</name>\n <value>9  
200</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml  
sofiagabian80@cluster-hadoop-practica-m:~$ sudo sed -i '$a \ <property>\n <name>es.nodes.wan.only</name>\n  
<value>true</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml  
sofiagabian80@cluster-hadoop-practica-m:~$ sudo sed -i '$a \ <property>\n <name>hive.aux.jars.path</name>\n  
<value>/usr/lib/hive/lib/elasticsearch-hadoop-8.14.1.jar,/usr/lib/hive/lib/commons-httpclient-3.1.jar</value>\n  
</property>\n</configuration>' /etc/hive/conf.dist/hive-site.xml  
sofiagabian80@cluster-hadoop-practica-m:~$ sudo cp elasticsearch-hadoop-8.14.1.jar /usr/lib/hive/lib/  
sofiagabian80@cluster-hadoop-practica-m:~$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/  
sofiagabian80@cluster-hadoop-practica-m:~$
```

The screenshot shows a terminal window with the following commands and output:

```
sofiagabian80@cluster-hadoop-practica-m:~$ sudo systemctl stop hive-server2
sofiagabian80@cluster-hadoop-practica-m:~$ sudo systemctl stop hive-metastore
sofiagabian80@cluster-hadoop-practica-m:~$ sudo systemctl start hive-metastore
sofiagabian80@cluster-hadoop-practica-m:~$ sudo systemctl start hive-server2
sofiagabian80@cluster-hadoop-practica-m:~$ sudo systemctl status hive-server2
```

● hive-server2.service - LSB: Hive Server2

Loaded: loaded (/etc/init.d/hive-server2; generated)

Drop-In: /etc/systemd/system/hive-server2.service.d

└─hive-hbase.conf, restart.conf

Active: active (running) since Fri 2025-03-21 20:27:23 UTC; 4s ago

Docs: man:systemd-sysv-generator(8)

Process: 50198 ExecStart=/etc/init.d/hive-server2 start (code=exited, status=0/SUCCESS)

Main PID: 50209 (java)

Tasks: 0 (limit: 19180)

Memory: 4.0K

CGroup: /system.slice/hive-server2.service

└─ 50209 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc\_jar -Dhive.log.dir=/var/log/hive -Dhive.l

```
Mar 21 20:27:20 cluster-hadoop-practica-m systemd[1]: Starting LSB: Hive Server2...
Mar 21 20:27:20 cluster-hadoop-practica-m su[50207]: (to hive) root on none
Mar 21 20:27:20 cluster-hadoop-practica-m su[50207]: pam_unix(su:session): session opened for user hive(uid=119)
Mar 21 20:27:20 cluster-hadoop-practica-m su[50207]: pam_unix(su:session): session closed for user hive
Mar 21 20:27:23 cluster-hadoop-practica-m hive-server2[50198]: * Started Hive Server2 (hive-server2):
Mar 21 20:27:23 cluster-hadoop-practica-m systemd[1]: Started LSB: Hive Server2.
sofiagabian80@cluster-hadoop-practica-m:~$ sudo systemctl status hive-metastore
```

● hive-metastore.service - LSB: Hive Metastore

Loaded: loaded (/etc/init.d/hive-metastore; generated)

Drop-In: /etc/systemd/system/hive-metastore.service.d

└─restart.conf

Active: active (running) since Fri 2025-03-21 20:27:15 UTC; 22s ago

Docs: man:systemd-sysv-generator(8)

Process: 50015 ExecStart=/etc/init.d/hive-metastore start (code=exited, status=0/SUCCESS)

Main PID: 50038 (java)

Tasks: 0 (limit: 19180)

Memory: 4.0K

CGroup: /system.slice/hive-metastore.service

└─ 50038 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc\_jar -Dhive.log.dir=/var/log/hive -Dhive.l

```
Mar 21 20:27:12 cluster-hadoop-practica-m systemd[1]: Starting LSB: Hive Metastore...
Mar 21 20:27:12 cluster-hadoop-practica-m hive-metastore[50015]: * Starting Hive Metastore (hive-metastore):
Mar 21 20:27:12 cluster-hadoop-practica-m su[50024]: (to hive) root on none
Mar 21 20:27:12 cluster-hadoop-practica-m su[50024]: pam_unix(su:session): session opened for user hive(uid=119)
Mar 21 20:27:12 cluster-hadoop-practica-m su[50024]: pam_unix(su:session): session closed for user hive
Mar 21 20:27:15 cluster-hadoop-practica-m systemd[1]: Started LSB: Hive Metastore.
sofiagabian80@cluster-hadoop-practica-m:~$
```

Para reiniciar el HIVE es mejor:

1.- Detenerlo con los comandos:

```
sudo systemctl stop hive-server2
sudo systemctl stop hive-metastore
```

2.- Y luego iniciarlo con:

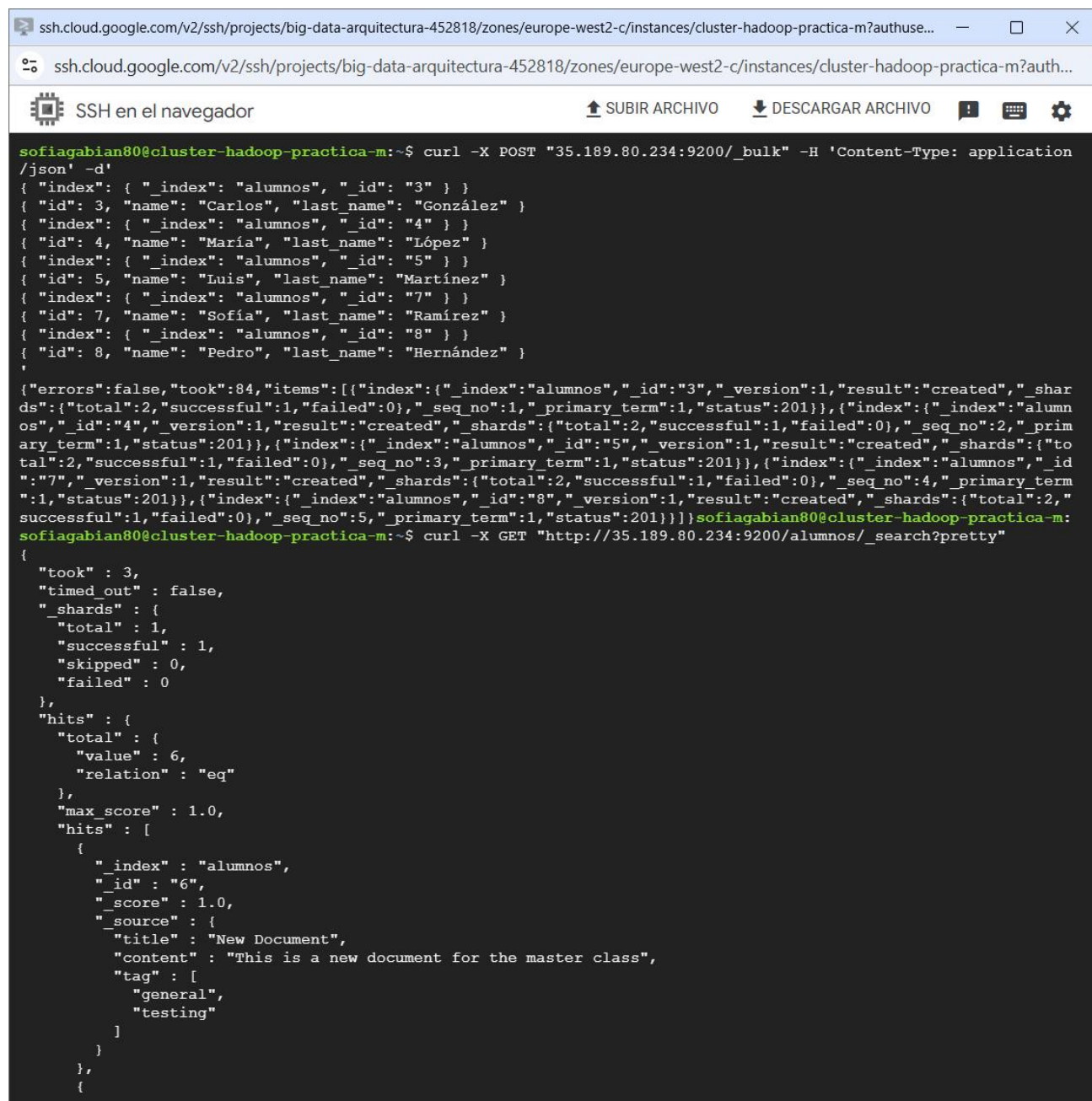
```
sudo systemctl start hive-metastore
sudo systemctl start hive-server2
```

3.- Podremos verificar que se está ejecutando correctamente con:

```
sudo systemctl status hive-server2
sudo systemctl status hive-metastore
```



## ENTREGABLE PARTE 4: Captura de pantalla de la consola del cluster Hadoop con el resultado la consulta.



```
ssh.cloud.google.com/v2/ssh/projects/big-data-arquitectura-452818/zones/europe-west2-c/instances/cluster-hadoop-practica-m?authuse...
ssh.cloud.google.com/v2/ssh/projects/big-data-arquitectura-452818/zones/europe-west2-c/instances/cluster-hadoop-practica-m?auth...
SSH en el navegador
SUBIR ARCHIVO
DESCARGAR ARCHIVO

sofiagabian80@cluster-hadoop-practica-m:~$ curl -X POST "35.189.80.234:9200/_bulk" -H 'Content-Type: application
/json' -d'
{ "index": { "_index": "alumnos", "_id": "3" } }
{ "id": 3, "name": "Carlos", "last_name": "González" }
{ "index": { "_index": "alumnos", "_id": "4" } }
{ "id": 4, "name": "María", "last_name": "López" }
{ "index": { "_index": "alumnos", "_id": "5" } }
{ "id": 5, "name": "Luis", "last_name": "Martínez" }
{ "index": { "_index": "alumnos", "_id": "7" } }
{ "id": 7, "name": "Sofía", "last_name": "Ramírez" }
{ "index": { "_index": "alumnos", "_id": "8" } }
{ "id": 8, "name": "Pedro", "last_name": "Hernández" }
'
{"errors":false,"took":84,"items":[{"index":{"_index":"alumnos","_id":"3","_version":1,"result":"created","_shar
ds":{"total":2,"successful":1,"failed":0},"_seq_no":1,"_primary_term":1,"status":201}},{"index":{"_index":"alumn
os","_id":"4","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":2,"_prim
ary_term":1,"status":201}},{"index":{"_index":"alumnos","_id":"5","_version":1,"result":"created","_shards":{"to
tal":2,"successful":1,"failed":0},"_seq_no":3,"_primary_term":1,"status":201}},{"index":{"_index":"alumnos","_id
":"7","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":4,"_primary term
":1,"status":201}},{"index":{"_index":"alumnos","_id":"8","_version":1,"result":"created","_shards":{"total":2,"
successful":1,"failed":0},"_seq_no":5,"_primary_term":1,"status":201}}]}sofiagabian80@cluster-hadoop-practica-m:
sofiagabian80@cluster-hadoop-practica-m:~$ curl -X GET "http://35.189.80.234:9200/alumnos/_search?pretty"
{
  "took" : 3,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 6,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "alumnos",
        "_id" : "6",
        "_score" : 1.0,
        "_source" : {
          "title" : "New Document",
          "content" : "This is a new document for the master class",
          "tag" : [
            "general",
            "testing"
          ]
        }
      }
    ]
  },
}
```

ssh.cloud.google.com/v2/ssh/projects/big-data-arquitectura-452818/zones/europe-west2-c/instances/cluster-hadoop-practica-m?authuse... — □ ×

ssh.cloud.google.com/v2/ssh/projects/big-data-arquitectura-452818/zones/europe-west2-c/instances/cluster-hadoop-practica-m?auth...



SSH en el navegador

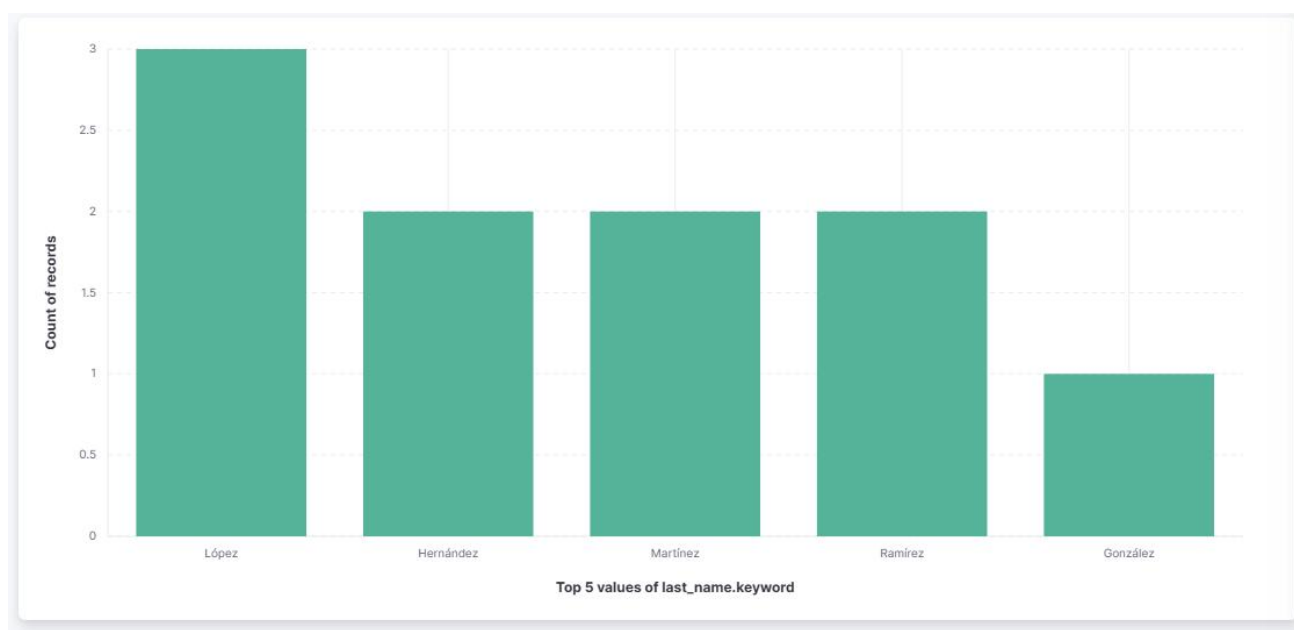
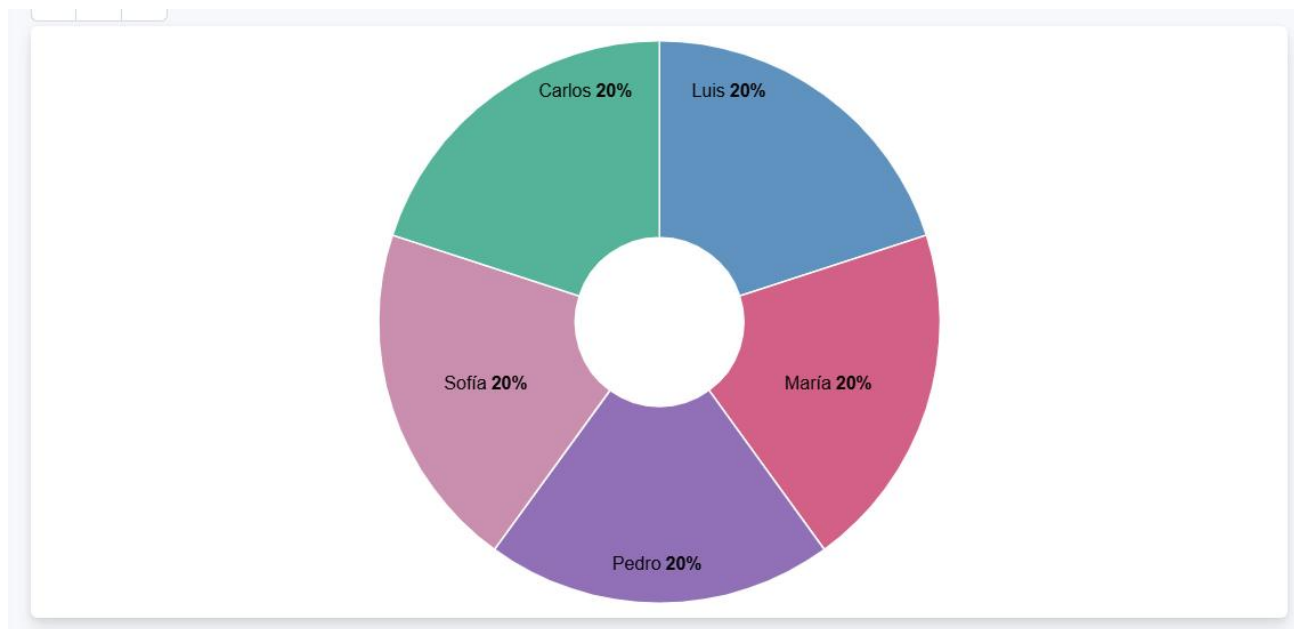
↑ SUBIR ARCHIVO

↓ DESCARGAR ARCHIVO



```
"_id" : "3",
"_score" : 1.0,
"_source" : {
  "id" : 3,
  "name" : "Carlos",
  "last_name" : "González"
}
},
{
  "_index" : "alumnos",
  "_id" : "4",
  "_score" : 1.0,
  "_source" : {
    "id" : 4,
    "name" : "María",
    "last_name" : "López"
  }
},
{
  "_index" : "alumnos",
  "_id" : "5",
  "_score" : 1.0,
  "_source" : {
    "id" : 5,
    "name" : "Luis",
    "last_name" : "Martínez"
  }
},
{
  "_index" : "alumnos",
  "_id" : "7",
  "_score" : 1.0,
  "_source" : {
    "id" : 7,
    "name" : "Sofía",
    "last_name" : "Ramírez"
  }
},
{
  "_index" : "alumnos",
  "_id" : "8",
  "_score" : 1.0,
  "_source" : {
    "id" : 8,
    "name" : "Pedro",
    "last_name" : "Hernández"
  }
}
]
}
}
sofiagabian80@cluster-hadoop-practica-m:~$
```

**ENTREGABLE 5:** Opcional. Captura de pantalla de la consola de Kibana con alguna visualización sencilla.





Los 5 valores principales de name.keyword: recuento de registros

