

This file corresponds to a brief summary of Practica BDA Walkthrough (**Click me!**)

Felipe Gallegos, Big Data, ML & AI Fullstack Bootcamp student

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

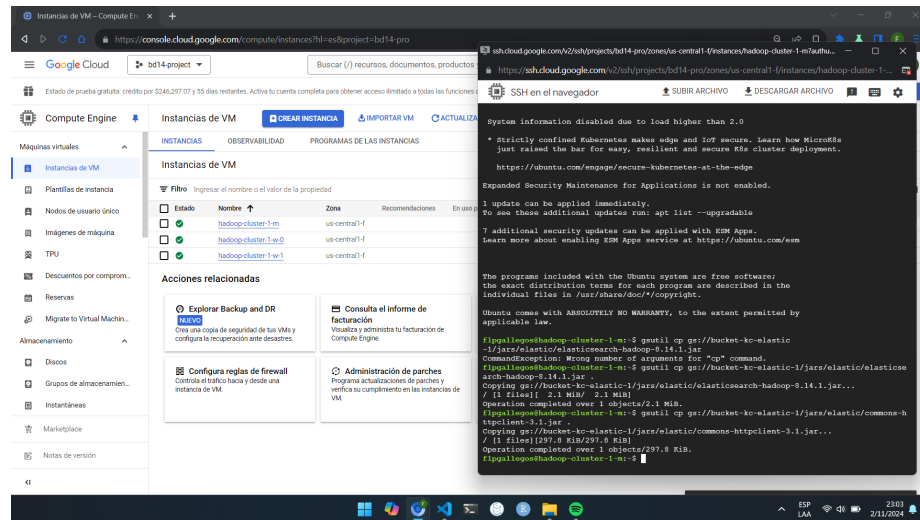
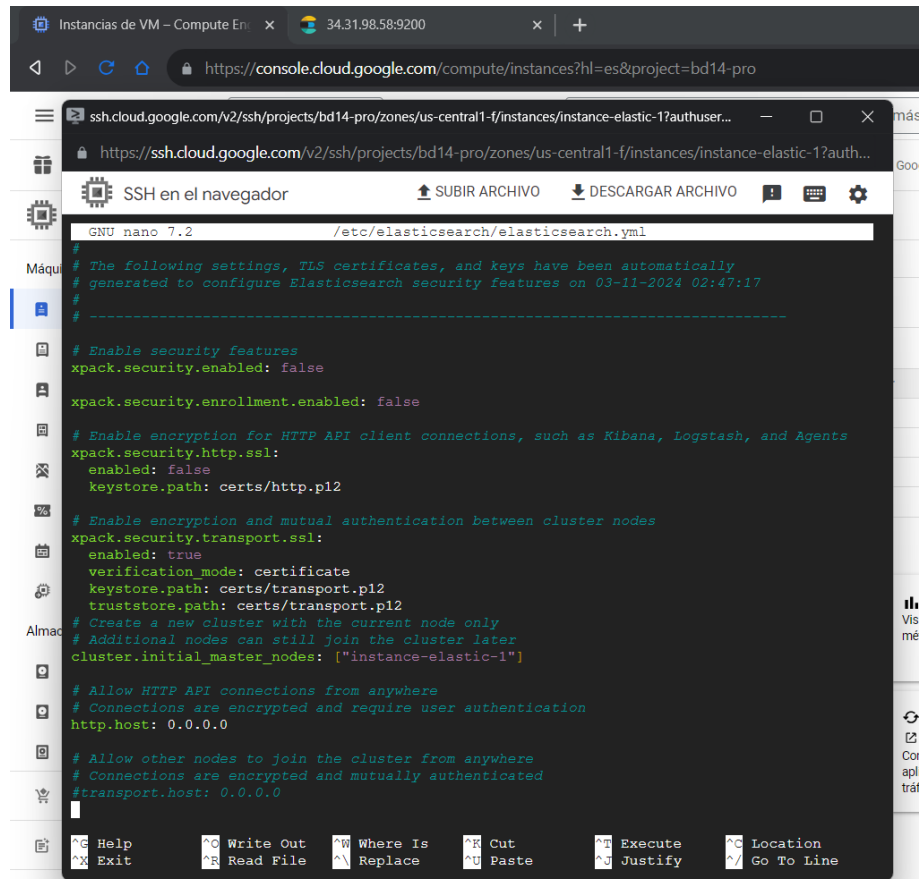


Figure 1: Hadoop configured

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



```
GNU nano 7.2 /etc/elasticsearch/elasticsearch.yml
#
# The following settings, TLS certificates, and keys have been automatically
# generated to configure Elasticsearch security features on 03-11-2024 02:47:17
#
# -----
#
# Enable security features
xpack.security.enabled: false
xpack.security.enrollment.enabled: false
# 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: ["instance-elastic-1"]
# 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
```

Figure 2: Elastic Configured

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

```

# Saved Objects: Migrations
# Saved object migrations run at startup. If you run into migration-related issues, you might need to adjust these settings.
# The number of documents migrated at a time.
# If Kibana can't start up or upgrade due to an Elasticsearch "circumventing exception", use a smaller batchSize value to reduce the memory pressure. Defaults to 1000 objects per batch.
# Migrations.batchSize: 1000
# The maximum payload size for indexing batches of upgraded saved objects.
# To avoid migrations failing due to a 413 Request Entity Too Large response from Elasticsearch, this value should be lower than or equal to your Elasticsearch cluster's 'http.max_content_length' configuration option. Default: 100mb
# Migrations.maxBatchSizeBytes: 100mb
# The number of times to retry temporary migration failures. Increase the setting if migrations fail frequently with a message such as 'Unable to complete the [...] step of ...'
# 15 attempts, terminating. Defaults to 15
# Migrations.retryAttempts: 15

# ===== Search Autocomplete =====
# Time in milliseconds to wait for autocomplete suggestions from Elasticsearch.
# This value must be a whole number greater than zero. Defaults to 100ms
# UnifiedSearch.autocomplete.valuesuggestions.timeout: 1000
# Maximum number of documents loaded by each shard to generate autocomplete suggestions.
# This value must be a whole number greater than zero. Defaults to 100,000
# UnifiedSearch.autocomplete.valuesuggestions.timeoutAfter: 100000
server.host: 0.0.0.0

[ippallogeshadoop-cluster-1-m]:$ sudo nano /etc/elasticsearch/elasticsearch.yml
[ippallogeshadoop-cluster-1-m]:$ sudo service elasticsearch restart
[ippallogeshadoop-cluster-1-m]:$ sudo service kibana restart
[ippallogeshadoop-cluster-1-m]:$ sudo nano /etc/elasticsearch/elasticsearch.yml
[ippallogeshadoop-cluster-1-m]:$

[ippallogeshadoop-cluster-1-m]:$ gsuutil cp gs://bucket-kc-elastic-1/jars/elasticsearch-hadoop-8.14.1.jar
CommandException: Wrong number of arguments for "cp" command.
[ippallogeshadoop-cluster-1-m]:$ gsuutil cp gs://bucket-kc-elastic-1/jars/elasticsearch-hadoop-8.14.1.jar .
Copying gs://bucket-kc-elastic-1/jars/elasticsearch-hadoop-8.14.1.jar...
/ (1 files) [ 2.1 MiB / 2.1 MiB ]
Operation completed over 1 objects/2.1 Min.
[ippallogeshadoop-cluster-1-m]:$ gsuutil cp gs://bucket-kc-elastic-1/jars/elasticsearch-commons-httpclient-3.1.jar .
Copying gs://bucket-kc-elastic-1/jars/elasticsearch-commons-httpclient-3.1.jar...
/ (1 files) [297.8 KiB/297.8 KiB]
Operation completed over 1 objects/297.8 KiB.
[ippallogeshadoop-cluster-1-m]:$ curl -I http://34.31.98.58:9200
curl: (7) Failed to connect to 34.31.98.58 port 9200 after 6 ms: Couldn't connect to server
[ippallogeshadoop-cluster-1-m]:$ sudo systemctl status elasticsearch
Unit elasticsearch.service could not be found.
[ippallogeshadoop-cluster-1-m]:$ curl -I http://34.31.98.58:9200
HTTP/1.1 200 OK
X-elastic-product: Elasticsearch
content-type: application/json
content-length: 543

[ippallogeshadoop-cluster-1-m]:$ sudo sed -i 's#/etc/hive/conf/dist/hive-site.xml#n
[ippallogeshadoop-cluster-1-m]:$ sudo sed -i 's#\<property>\n\<name>\n\<value>34.31.98.58/</value>\n\</property>\n\</etc/hive/conf/dist/hive-site.xml#n
[ippallogeshadoop-cluster-1-m]:$ sudo sed -i 's#\<property>\n\<name>\n\<value>9200/</value>\n\</property>\n\</etc/hive/conf/dist/hive-site.xml#n
[ippallogeshadoop-cluster-1-m]:$ sudo sed -i 's#\<property>\n\<name>\n\<value>run/</value>\n\</property>\n\</etc/hive/conf/dist/hive-site.xml#n
[ippallogeshadoop-cluster-1-m]:$ sudo sed -i 's#\<property>\n\<name>\n\<value>hadoop-8.14.1.jar/</property>\n\</etc/hive/conf/dist/hive-site.xml#n
[ippallogeshadoop-cluster-1-m]:$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
[ippallogeshadoop-cluster-1-m]:$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
[ippallogeshadoop-cluster-1-m]:$ sudo service hive-server2 restart
[ippallogeshadoop-cluster-1-m]:$

```

Figure 3: ES-Hadoop connection

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

The screenshot displays the Google Cloud Platform console for a project named 'bd14-pro'. The left sidebar shows the 'Instances' section, with the 'elastic-17a...' instance selected. The main panel shows the 'SSH en el navegador' (SSH in the browser) window, which displays the terminal output of the 'ipgallagos@hadoop-cluster-1-m1:~\$' command. The output shows the results of a query, including the 'source' field with details like 'code', 'name', 'last_name', 'age', 'email', and 'enrollment_date'. The results are displayed in a table format, with columns for 'source', 'id', 'score', 'name', 'last_name', 'age', 'email', and 'enrollment_date'. The results show three entries: 'source' (code: '1008', name: 'Ana', last_name: 'Garcia', age: 15, email: 'ana.garcia@example.com', enrollment_date: '2022-06-10'), 'source' (code: '1009', name: 'Javier', last_name: 'Morales', age: 22, email: 'javier.morales@example.com', enrollment_date: '2022-07-18'), and 'source' (code: '1010', name: 'Lucia', last_name: 'Castillo', age: 21, email: 'lucia.castillo@example.com', enrollment_date: '2022-08-25').

```
ipgallagos@hadoop-cluster-1-m1:~$ curl -X POST 'localhost:9200/_search' -d '{
  "index": "alumnos",
  "id": "1",
  "source": {
    "code": "1008",
    "name": "Ana",
    "last_name": "Garcia",
    "age": 15,
    "email": "ana.garcia@example.com",
    "enrollment_date": "2022-06-10"
  }
}'

{"_source": {
  "code": "1008",
  "name": "Ana",
  "last_name": "Garcia",
  "age": 15,
  "email": "ana.garcia@example.com",
  "enrollment_date": "2022-06-10"
},
  "index": "alumnos",
  "id": "1",
  "score": 1.0,
  "source": {
    "code": "1009",
    "name": "Javier",
    "last_name": "Morales",
    "age": 22,
    "email": "javier.morales@example.com",
    "enrollment_date": "2022-07-18"
  }
},
  "index": "alumnos",
  "id": "10",
  "score": 1.0,
  "source": {
    "code": "1010",
    "name": "Lucia",
    "last_name": "Castillo",
    "age": 21,
    "email": "lucia.castillo@example.com",
    "enrollment_date": "2022-08-25"
  }
}
}
```

Figure 4: Hadoop Query

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

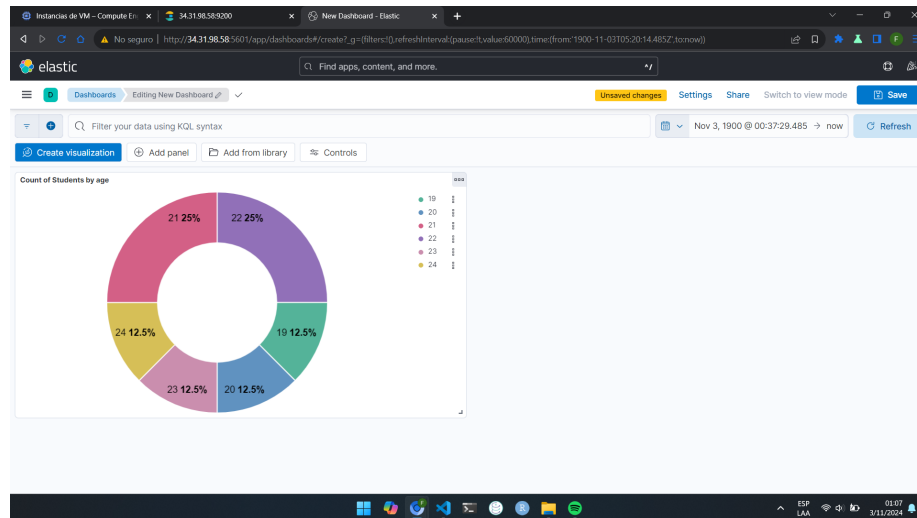


Figure 5: Kibana Dashboard