

PRÁCTICA ELASTICSEARCH HADOOP

Sandy Rodríguez Aponte

1. Configuración ES-Hadoop

The screenshot displays three main windows from the Google Cloud Platform:

- Google Cloud Home:** Shows the DataProc cluster "cluster-hadoop-2" with three workers (m1) and one master (m2). The master node has an SSH terminal open, showing the command: `sandy_73@cluster-hadoop-2-mr:~$ hadoop fs -cp gs://hadoop_elastic_practica_backup/elasticsearch-hadoop-8.14.1.jar . / { files} | 2.1 MB/ 2.1 MB [jars]`.
- DataProc Cluster Details:** Provides cluster-level information such as name, UUID, type (DataProc), and state (En ejecución).
- Cloud Storage Bucket Details:** Shows the bucket "hadoop_elastic_practica_backup" with two objects: "commons-httpclient-3.1.jar" (297.9 KB) and "elasticsearch-hadoop-8.14.1.jar" (2.1 MB).

2. Configuración Server ElasticSearch

Nueva regla firewall

Estado de prueba gratuita: crédito por €207.37 y 74 días restantes. Activa tu cuenta completa para obtener acceso ilimitado a todas las funciones de Google Cloud. Usa los créditos restantes y paga solo por lo que usas.

Seguridad de red

Detalles de la regla de firewall

[EDITAR](#) [BORRAR](#)

Proxy web seguro

Cloud Armor

- Panel de DDoS
- Políticas de Cloud Armor
- Protección adaptable
- Nivel de servicio de Cloud Ar...

IDS de Cloud

- Panel del IDS
- Extremos de IDS
- Amenazas en IDS

Cloud NGFW

- Panel
- Políticas de firewall**
- Amenazas
- Extremos de firewall
- Estadísticas

Componentes comunes

- Grupos de direcciones
- Perfiles de seguridad
- Políticas de inspección de T...
- Políticas de SSL
- Autenticación de clientes

elastic-kibana-hadoop-practica

Registros [?](#)
Desactivada
[Ver en el Explorador de registros](#)

Red
default

Prioridad
1000

Dirección
Entrada

Acción en caso de coincidencia
Permitir

Filtros de fuente

Rangos de IP
34.55.74.126

Protocolos y puertos
tcp:5601, 8088, 9200

Aplicación
Habilitada

Estadísticas
Ninguno

Supervisión de aciertos [?](#)
—

Aplicable a las instancias

cluster-hadoop-2-m En la siguiente tabla, no se muestran instancias en entorno flexible de App Engine

Filtro [Filtrar según nombre de instancia, proyecto o subred](#)

Nombre	Subred	Rangos de IP internas	Rangos de IP externas	Etiquetas	Cuentas de servicio	Proyecto	Etiquetas	Detalles de la red
cluster-hadoop-2-m	default	10.128.0.10	34.55.74.126	Ninguno	53645706888-compute@developer.gserviceaccount.com	bd14-project-438920	goog-datap... enabled	VER DETALLES
cluster-hadoop-2-w-0	default	10.128.0.9	34.45.149.114	Ninguno	53645706888-compute@developer.gserviceaccount.com	bd14-project-438920	goog-datap... enabled	VER DETALLES
cluster-hadoop-2-w-1	default	10.128.0.8	34.121.181.96	Ninguno	53645706888-compute@developer.gserviceaccount.com	bd14-project-438920	goog-datap... enabled	VER DETALLES
elastic-1	default	10.128.0.11	Ninguno	http-server, https-server	53645706888-compute@developer.gserviceaccount.com	bd14-project-438920		VER DETALLES

Configuro el fichero /etc/elasticsearch/elasticsearch.yml

The screenshot displays two browser windows side-by-side. The left window is the Google Cloud Compute Engine interface, showing a list of VM instances including 'cluster-hadoop-2-m', 'cluster-hadoop-2-w-0', 'cluster-hadoop-2-w-1', and 'elasti-1'. The right window is an SSH session on port 438920, showing the Elasticsearch configuration file (`/etc/elasticsearch/elasticsearch.yml`). The configuration includes network settings like `network.host: 192.168.0.1`, port 9200, and discovery settings. A specific section for security features is highlighted, showing `xpack.security.enabled: false`. The bottom of the configuration file shows the end of the security auto-configuration block.

```
# Elasticsearch performs poorly when the system is swapping the memory.
#----- Network -----
# By default Elasticsearch is only accessible on localhost. Set a different
# address here to expose this node on the network:
#network.host: 192.168.0.1

# By default Elasticsearch listens for HTTP traffic on the first free port it
# finds starting at 9200. Set a specific HTTP port here:
#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: ["node1", "node2"]

# For more information, consult the discovery and cluster formation module documentation.
#----- Various -----
# Allow wildcard deletion of indices:
#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 01-11-2024 17:45:15
#----- Security Features -----
# Enable security features
xpack.security.enabled: false

xpack.security.enrollment.enabled: true

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

# Create a new cluster with the current node only
# Additional nodes can still join the cluster later
cluster.initial_master_nodes: ["elasti-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

#----- END SECURITY AUTO CONFIGURATION -----
sandy_73c@elasti-1:~$
```

Reinicio el servicio de elasticsearch y kibana

```

# By default Elasticsearch is only accessible on localhost. Set a different
# address here to expose this node on the network:
#network.host: 192.168.0.1

# By default Elasticsearch listens for HTTP traffic on the first free port it
# finds starting at 9200. Set a specific HTTP port here:
#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", "([::])"]
#discovery.seed_hosts: ["host1", "host2"]

# Bootstrap the cluster using an initial set of master-eligible nodes:
#cluster.initial_master_nodes: ["node1", "node2"]

# For more information, consult the discovery and cluster formation module documentation.

# ----- Various -----
# Allow wildcard deletion of indices:
#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 01-11-2024 17:43:15

# Enable security features
xpack.security.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
  verificationMode: 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: ["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

#----- END SECURITY AUTO CONFIGURATION -----
#andy_73@elasti-1:~$ sudo service elasticsearch restart
#andy_73@elasti-1:~$ sudo service elasticsearch restart
#andy_73@elasti-1:~$ 

```

Hago una comprobación desde el cluster hadoop con hive al server de elastic.

```

# By default Elasticsearch is only accessible on localhost. Set a different
# address here to expose this node on the network:
#network.host: 192.168.0.1

# By default Elasticsearch listens for HTTP traffic on the first free port it
# finds starting at 9200. Set a specific HTTP port here:
#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", "([::])"]
#discovery.seed_hosts: ["host1", "host2"]

# Bootstrap the cluster using an initial set of master-eligible nodes:
#cluster.initial_master_nodes: ["node1", "node2"]

# For more information, consult the discovery and cluster formation module documentation.

# ----- Various -----
# Allow wildcard deletion of indices:
#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 01-11-2024 17:43:15

# Enable security features
xpack.security.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
  verificationMode: 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: ["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

#----- END SECURITY AUTO CONFIGURATION -----
#andy_73@elasti-1:~$ curl -I http://35.232.206.132:9200
#HTTP/2.0 200 OK
#date: Mon, 11 Nov 2024 17:43:15 GMT
#content-type: application/json
#content-length: 334
#andy_73@elasti-1:~$ 

```

3. Configuración en Cluster Hadoop de conexión con ES

Cargo los ficheros jars en la configuración de HIVE, después, reinicio el servicio de hive para que se apliquen los cambios.

The screenshot shows two windows. On the left, the Google Cloud Compute Engine interface lists instances: 'cluster-hadoop-2-m' (status: healthy), 'cluster-hadoop-2-w-0' (status: healthy), 'cluster-hadoop-2-w-1' (status: healthy), and 'elastic-1' (status: healthy). On the right, an SSH terminal window titled 'SSH en el navegador' shows the configuration file /etc/hive/conf.dist/hive-site.xml. A red box highlights the section where auxiliary JAR paths are added:

```
<configuration>
<property>
<name>hive.aux.jars.path</name>
<value>/usr/lib/hive/lib/elasticsearch-hadoop-8.14.1.jar, /usr/lib/hive/lib/commons-httpclient-3.1.jar</value>
</property>
</configuration>
```

Below this, the terminal shows the command to restart the hive service:

```
sandy_73@cluster-hadoop-2-m:~$ sudo cp elasticsearch-hadoop-8.14.1.jar /usr/lib/hive/lib/
sandy_73@cluster-hadoop-2-m:~$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
sandy_73@cluster-hadoop-2-m:~$ sudo service hive-server2 restart
sandy_73@cluster-hadoop-2-m:~$
```

4. A conectar datos

Desde Server elasticsearch creo un índice. Despues me conectare desde el cluster de hadoop.

The screenshot shows the Google Cloud Compute Engine interface and an SSH session. On the left, the 'Instances de VM' section lists several instances, including 'cluster-hadoop-2-m' and 'elasticsearch-1'. On the right, an SSH terminal window titled 'SSH en el navegador' is open, showing the command:

```
sandy_7@elastic-1:~$ curl -X POST "localhost:9200/alumnos/_doc/6" -H "Content-Type: application/json" -d
{
  "title": "New Document",
  "content": "This is a new document for the master class",
  "tag": ["general", "testing"]
}
```

The terminal output shows the document was created successfully.

Desde cluster Hadoop, agrego documentos al index de alumnos que he creado en server elasticsearch.

The screenshot shows the Google Cloud Compute Engine interface and an SSH session. The instance 'cluster-hadoop-2-m' is selected. On the right, an SSH terminal window titled 'SSH en el navegador' is open, showing the command:

```
sandy_7@cluster-hadoop-2-m:~$ curl -X POST "http://35.232.206.132:9200/alumnos/_search?pretty"
```

The terminal output shows the search results for the 'alumnos' index, including the newly inserted document.

Hagamos una consulta para ver los datos insertados:

The screenshot shows the Google Cloud Platform interface. On the left, the 'Compute Engine' section is open, displaying a list of VM instances. The instances listed are:

- cluster-hadoop-2-m (IP: 10.128.0.10)
- cluster-hadoop-2-w-0 (IP: 10.128.0.9)
- cluster-hadoop-2-w-1 (IP: 10.128.0.8)
- elasto-1 (IP: 10.128.0.11)

On the right, a terminal window titled 'SSH en el navegador' is running a curl command to search an Elasticsearch index named 'alumnos'. The command is:

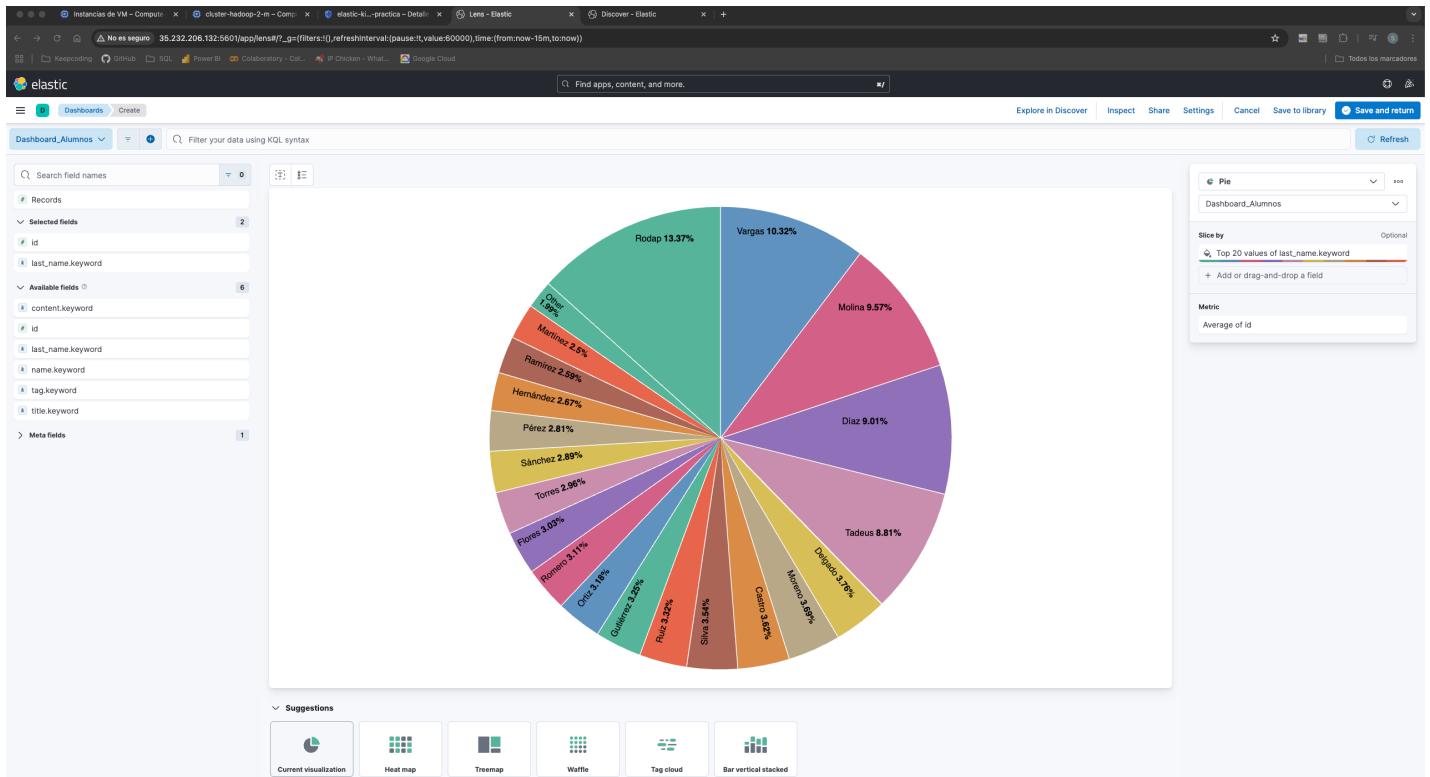
```
sandy_73@cluster-hadoop-2-m:~$ curl -X GET "http://35.232.206.132:9200/alumnos/_search?pretty"
```

The terminal output shows the results of the Elasticsearch search, which includes documents like 'new document' and various student records.

5. Opcional, KIBANA!

Para esta práctica he añadido 100 registros más de alumnos.

Hago el porcentaje de los apellidos de los alumnos



Otra forma de presentación

