

Práctica de **Arquitectura Big Data** en el que conectaremos un Clúster Hadoop con ElasticSearch desde Google Cloud.

1. Configuración del clúster Hadoop

- Creación del clúster

Nombre	practica-bd-kc
UUID del clúster	c1e4348a-8ea9-44cf-8a88-04a362dadead
Tipo	Clúster de Dataproc
Estado	✓ En ejecución



< MONITORING TRABAJOS INSTANCIAS DE VM CONFIGURACIÓN			
☰ Filtro Filtrar instancias			
●	Nombre	Rol	
✓	practica-bd-kc-m	Instancia principal	SSH ▾
✓	practica-bd-kc-w-0	Trabajador	
✓	practica-bd-kc-w-1	Trabajador	

- creacion del Bucket

📁 practica-bd-hadoop

Ubicación	Clase de almacenamiento	Acceso público	Protección
europa-west1 (Bélgica)	Standard	No público	Borrar de forma no definitiva

OBJETOS	CONFIGURACIÓN	PERMISOS	PROTECCIÓN	CICLO DE VIDA	OBSERVABILIDAD	NUEVO
---------	---------------	----------	------------	---------------	----------------	-------

I> Depósitos > practica-bd-hadoop 📁					
CREAR CARPETA SUBIR ▾ TRANSFERIR LOS DATOS ▾ OTROS SERVICIOS ▾					
Filtrar solo por prefijo de nombre ▾ ☰ Filtro Filtrar objetos y carpetas					
<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación ?	Clase de aln
<input type="checkbox"/>	 commons-httpclient-3.1.jar	305 KB	application/octet-stream	19 mar 2025 19:01:37	Standard
<input type="checkbox"/>	 elasticsearch-hadoop-8.14.1.jar	2.2 MB	application/octet-stream	19 mar 2025 19:03:33	Standard

- Carga de documento a la consola del clúster

```

oridiciaccio@practica-bd-kc-m:~$ ls
snap
oridiciaccio@practica-bd-kc-m:~$ gsutil cp gs://practica-bd-hadoop/elasticsearch-hadoop-8.14.1.jar .

Copying gs://practica-bd-hadoop/elasticsearch-hadoop-8.14.1.jar...
/ [0 files][ 0.0 B/ 2.1 MiB]
/ [1 files][ 2.1 MiB/ 2.1 MiB]

Operation completed over 1 objects/2.1 MiB.

oridiciaccio@practica-bd-kc-m:~$
oridiciaccio@practica-bd-kc-m:~$ gsutil ls gs://practica-bd-hadoop/

gs://practica-bd-hadoop/commons-httpclient-3.1.jar
gs://practica-bd-hadoop/elasticsearch-hadoop-8.14.1.jar
oridiciaccio@practica-bd-kc-m:~$
oridiciaccio@practica-bd-kc-m:~$ gsutil cp gs://practica-bd-hadoop/commons-httpclient-3.1.jar .
Copying gs://practica-bd-hadoop/commons-httpclient-3.1.jar...
/ [0 files][ 0.0 B/297.8 KiB]
/ [1 files][297.8 KiB/297.8 KiB]

Operation completed over 1 objects/297.8 KiB.

oridiciaccio@practica-bd-kc-m:~$ ls
commons-httpclient-3.1.jar elasticsearch-hadoop-8.14.1.jar snap
oridiciaccio@practica-bd-kc-m:~$ █

```

- Firewall

Nombre	Tipo	Destinos	Filtros	Protocolos/puertos	Acción	Prioridad	Red ↑	Registros
abrir- puertos- 8088-9870	Entrada	Aplicar a	Intervalos de	tcp:8088, 9870	Permitir	1000	default	Desactivado

2. Configuración Server Elasticsearch

```
# 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: false
  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: ["elastic-prueba"]

# 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 -----
```

- Firewall

elastic-prueba	Entrada	Aplicar a	Intervalos de	tcp:5601, 9200	Permitir	1000	default	Desactivado
--------------------------------	---------	-----------	---------------	----------------	----------	------	-------------------------	-------------

- Confirmar desde nodo master que VM ElasticS es accesible

```
oridiciaccio@practica-bd-kc-m:~$ ping 34.175.158.177 -c 4
PING 34.175.158.177 (34.175.158.177) 56(84) bytes of data.
64 bytes from 34.175.158.177: icmp_seq=1 ttl=61 time=3.39 ms
64 bytes from 34.175.158.177: icmp_seq=2 ttl=61 time=0.312 ms
64 bytes from 34.175.158.177: icmp_seq=3 ttl=61 time=0.325 ms
64 bytes from 34.175.158.177: icmp_seq=4 ttl=61 time=0.448 ms

--- 34.175.158.177 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3043ms
rtt min/avg/max/mdev = 0.312/1.119/3.391/1.312 ms
oridiciaccio@practica-bd-kc-m:~$
```

```
oridiciaccio@practica-bd-kc-m:~$ curl -X GET "http://34.175.158.177:9200"
{
  "name" : "elastic-prueba",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "AnnYeAtHTBKzAEgKBiSGgQ",
  "version" : {
    "number" : "8.14.1",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "93a57a1a76f556d8aee6a90d1a95b06187501310",
    "build_date" : "2024-06-10T23:35:17.114581191Z",
    "build_snapshot" : false,
    "lucene_version" : "9.10.0",
    "minimum_wire_compatibility_version" : "7.17.0",
    "minimum_index_compatibility_version" : "7.0.0"
  },
  "tagline" : "You Know, for Search"
}
```

- Confirmar Conexion de Elastic con HTTP

```
{
  "name" : "elastic-prueba",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "AnnYeAtHTBKzAEgKBiSGgQ",
  "version" : {
    "number" : "8.14.1",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "93a57a1a76f556d8aee6a90d1a95b06187501310",
    "build_date" : "2024-06-10T23:35:17.114581191Z",
    "build_snapshot" : false,
    "lucene_version" : "9.10.0",
    "minimum_wire_compatibility_version" : "7.17.0",
    "minimum_index_compatibility_version" : "7.0.0"
  },
  "tagline" : "You Know, for Search"
}
```

3. Configuración Hive en Cluster Hadoop con conexión a ElasticSearch

- Configuración y reinicio de Hive

```
oridiciaccio@practica-bd-kc-m:~$ sudo sed -i '$d' /etc/hive/conf.dist/hive-site.xml
oridiciaccio@practica-bd-kc-m:~$ sudo sed -i '$a \ <property>\n <name>es.nodes</name>\n <value>AQUÍ LA IP DE ELASTIC</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
oridiciaccio@practica-bd-kc-m:~$ sudo sed -i '$a \ <property>\n <name>es.port</name>\n <value>9200</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
oridiciaccio@practica-bd-kc-m:~$ sudo sed -i '$a \ <property>\n <name>es.nodes.wan.only</name>\n <value>>true</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
oridiciaccio@practica-bd-kc-m:~$ sudo sed -i '$a \ <property>\n <name>hive.aux.jars.path</name>\n <value>/usr/lib/hive/lib/elasticsearch-hadoop-8.14.1.jar,/usr/lib/hive/lib/commons-httpclient-3.1.jar</value>\n </property>\n</configuration>' /etc/hive/conf.dist/hive-site.xml
oridiciaccio@practica-bd-kc-m:~$ sudo cp elasticsearch-hadoop-8.14.1.jar /usr/lib/hive/lib/
oridiciaccio@practica-bd-kc-m:~$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
oridiciaccio@practica-bd-kc-m:~$ sudo systemctl restart hive-metastore
oridiciaccio@practica-bd-kc-m:~$ sudo systemctl restart hive-server2
```

4. Creacion de index y consulta de datos

- Agregar los documentos al index

```
"took" : 9,
"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"
        ]
      }
    },
    {
      "_index" : "alumnos",
      "_id" : "3",
      "_score" : 1.0,
      "_source" : {
        "id" : 3,
        "name" : "Carlos",
        "last_name" : "González"
      }
    },
    {
      "_index" : "alumnos",
      "_id" : "4",
      "_score" : 1.0,
      "_source" : {
```

```
    {
      "_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"
      }
    }
  ]
}
```

5. Dashboard con kibana, únicamente representativo de la herramienta, sin valor analítico

María Pedro
Sofía Carlos
Luis

