

PRÁCTICA ARQUITECTURA BIG DATA: ElasticSearch Hadoop.

ALUMNO: Jose Ignacio González Tejedor

PARTE 1 - Configuración ES-Hadoop

Free trial status: €255.97 credit and 80 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISS

ACTIVATE

Google Cloud

bd14-project

🔍 ⚡ 📄 1 ? ⋮ J

Cluster details

SUBMIT JOB

REFRESH

START

STOP

DELETE

⋮

1

The firewall rules for specified network or subnetwork would allow ingress traffic from 0.0.0.0/0, which could be a security risk.

MORE

Name

practica-hadoop

Cluster UUID

800ca33c-c2d8-40a6-8f86-95ae4223193c

Type

Dataproc Cluster

Status

Running

MONITORING

JOB

VM INSTANCES

CONFIGURATION

WEB INTERFACES

Filter

Filter instances

?

⋮

●

Name

Role

Machine type

✓

practica-hadoop-m

Master

SSH

e2-standard-4

✓

practica-hadoop-w-0

Worker

e2-standard-2

✓

practica-hadoop-w-1

Worker

e2-standard-2

practica-hadoop-1

Location

Storage class

Public access

Protection

us-central1 (Iowa)

Standard

Not public

Soft Delete

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

Buckets > practica-hadoop-1

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

⋮

Name

Size

Type

Created

commons-httpclient-3.1.jar

297.9 KB

application/octet-stream

Oct 28, 2024, 11:49:21 PM

elasticsearch-hadoop-8.14.1.jar

2.1 MB

application/octet-stream

Oct 28, 2024, 11:49:35 PM

Mozilla Firefox

https://ssh.cloud.google.com/v2/ssh/projects/bd14-project-439221/zones/us-central1-a/instances/practica-hadoop-m?authuser=

SSH-in-browser

UPLOAD FILE

DOWNLOAD FILE

⚠

⌵

nacho\_sql\_training@practica-hadoop-m:~\$ gsutil cp gs://practica-hadoop-1/elasticsearch-hadoop-8.14.1.jar .

Copying gs://practica-hadoop-1/elasticsearch-hadoop-8.14.1.jar...

/ [1 files] [ 2.1 MiB / 2.1 MiB]

Operation completed over 1 objects/2.1 MiB.

nacho\_sql\_training@practica-hadoop-m:~\$

nacho\_sql\_training@practica-hadoop-m:~\$ gsutil cp gs://practica-hadoop-1/commons-httpclient-3.1.jar .

Copying gs://practica-hadoop-1/commons-httpclient-3.1.jar...

/ [1 files] [297.8 KiB / 297.8 KiB]

Operation completed over 1 objects/297.8 KiB.

nacho\_sql\_training@practica-hadoop-m:~\$

## PARTE 2 - Configuración server Elasticsearch

Google Cloud bd14-project 1

VM instances CREATE INSTANCE IMPORT VM REFRESH LEARN

INSTANCES OBSERVABILITY INSTANCE SCHEDULES

VM instances

Filter Enter property name or value

Status	Name ↑	Zone	Recommendations	In use by	Internal IP	Connect
<input checked="" type="checkbox"/>	<a href="#">practica-elastic-1</a>	us-central1-f			10.128.0.16 (nic0)	SSH
<input checked="" type="checkbox"/>	<a href="#">practica-hadoop-m</a>	us-central1-a			10.128.0.13 (nic0)	SSH
<input checked="" type="checkbox"/>	<a href="#">practica-hadoop-w-0</a>	us-central1-a			10.128.0.14 (nic0)	SSH
<input checked="" type="checkbox"/>	<a href="#">practica-hadoop-w-1</a>	us-central1-a			10.128.0.15 (nic0)	SSH

SSH-in-browser UPLOAD FILE DOWNLOAD FILE

```
nacho_sql_training@practica-hadoop-m:~$ gsutil cp gs://practica-hadoop-1/elasticsearch-hadoop-8.14.1.jar .
Copying gs://practica-hadoop-1/elasticsearch-hadoop-8.14.1.jar...
/ [1 files][ 2.1 MiB/ 2.1 MiB]
Operation completed over 1 objects/2.1 MiB.
nacho_sql_training@practica-hadoop-m:~$
nacho_sql_training@practica-hadoop-m:~$ gsutil cp gs://practica-hadoop-1/commons-httpclient-3.1.jar .
Copying gs://practica-hadoop-1/commons-httpclient-3.1.jar...
/ [1 files][297.8 KiB/297.8 KiB]
Operation completed over 1 objects/297.8 KiB.
nacho_sql_training@practica-hadoop-m:~$ curl -I http://34.133.109.252:9200
HTTP/1.1 200 OK
X-elastic-product: Elasticsearch
content-type: application/json
content-length: 543
nacho_sql_training@practica-hadoop-m:~$
```

```
#----- BEGIN SECURITY AUTO CONFIGURATION -----
#
# The following settings, TLS certificates, and keys have been automatically
# generated to configure Elasticsearch security features on 28-10-2024 23:23:55
#
# -----

# 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: ["practica-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 -----
nacho_sql_training@practica-elastic-1:~$
```

## PARTE 3 - Configuración en Cluster Hadoop de Conexión con ES

```
SSH-in-browser  UPLOAD FILE  DOWNLOAD FILE  [ ] [ ] [ ]

nacho_sql_training@practica-hadoop-m:~$ gsutil cp gs://practica-hadoop-1/elasticsearch-hadoop-8.14.1.jar .
Copying gs://practica-hadoop-1/elasticsearch-hadoop-8.14.1.jar...
/ [1 files][ 2.1 MiB/ 2.1 MiB]
Operation completed over 1 objects/2.1 MiB.
nacho_sql_training@practica-hadoop-m:~$
nacho_sql_training@practica-hadoop-m:~$ gsutil cp gs://practica-hadoop-1/commons-httpclient-3.1.jar .
Copying gs://practica-hadoop-1/commons-httpclient-3.1.jar...
/ [1 files][297.8 KiB/297.8 KiB]
Operation completed over 1 objects/297.8 KiB.
nacho_sql_training@practica-hadoop-m:~$ curl -I http://34.133.109.252:9200
HTTP/1.1 200 OK
X-elastic-product: Elasticsearch
content-type: application/json
content-length: 543

nacho_sql_training@practica-hadoop-m:~$ sudo sed -i 's/d' /etc/hive/conf.dist/hive-site.xml
nacho_sql_training@practica-hadoop-m:~$ sudo sed -i 's/a \ <property>\n <name>es.nodes</name>\n <value>34.133
.109.252</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
nacho_sql_training@practica-hadoop-m:~$ sudo sed -i 's/a \ <property>\n <name>es.port</name>\n <value>9200</v
alue>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
nacho_sql_training@practica-hadoop-m:~$ sudo sed -i 's/a \ <property>\n <name>es.nodes.wan.only</name>\n <val
ue>true</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
nacho_sql_training@practica-hadoop-m:~$ sudo sed -i 's/a \ <property>\n <name>hive.aux.jars.path</name>\n <val
ue>/usr/lib/hive/lib/elasticsearch-hadoop-8.14.1.jar,/usr/lib/hive/lib/commons-httpclient-3.1.jar</value>\n </prop
erty>\n</configuration>' /etc/hive/conf.dist/hive-site.xml
nacho_sql_training@practica-hadoop-m:~$ sudo cp elasticsearch-hadoop-8.14.1.jar /usr/lib/hive/lib/
nacho_sql_training@practica-hadoop-m:~$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
nacho_sql_training@practica-hadoop-m:~$ sudo service hive-server2 restart
nacho_sql_training@practica-hadoop-m:~$
```

## PARTE 4 - A conectar datos!

```
nacho_sql_training@practica-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"]
> }
>
{"_index":"alumnos","_id":"6","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq
_no":0,"_primary_term":1}nacho_sql_training@practica-elastic-1:~$
```

```
nacho_sql_training@practica-hadoop-m:~$ curl -X POST "34.133.109.252:9200/_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": "Maria", "last_name": "López" }
{"_index": { " _index": "alumnos", " _id": "5" } }
{"_id": 5, "name": "Luis", "last_name": "Martínez" }
{"_index": { " _index": "alumnos", " _id": "7" } }
{"_id": 7, "name": "Sofia", "last_name": "Ramírez" }
{"_index": { " _index": "alumnos", " _id": "8" } }
{"_id": 8, "name": "Pedro", "last_name": "Hernández" }
}

{"errors":false,"took":83,"items":[{"_index":{"_index":"alumnos","_id":"3","_version":1,"result":"created","_shards":
{"total":2,"successful":1,"failed":0},"_seq_no":1,"_primary_term":1,"status":201}},{"_index":{"_index":"alumnos","_
id":"4","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":2,"_primary_term"
:1,"status":201}},{"_index":{"_index":"alumnos","_id":"5","_version":1,"result":"created","_shards":{"total":2,"succ
essful":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}}]}nacho_sql_training@practica-hadoop-m:~$
```

```
nacho_sql_training@practica-hadoop-m:~$ curl -X GET "http://34.133.109.252:9200/alumnos/_search?pretty"
{
  "took" : 8,
  "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",
```

```
{
  "_index" : "alumnos",
  "_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"
  }
}
]
}
}
nacho_sql_training@practica-hadoop-m:~$
```

PARTE 5 - Opcional. KIBANA!

