Big Data Arquitecture XIV

PROFESOR : ALEJANDRO SAEZ
ALUMNO: PABLO AGUINAGA CABREDO

PARTE 1 - Configuración ES-Hadoop:

```
Last login: Tue Oct 29 19:35:53 2024 from 35.235.245.129
p_aguinaga@cluster-hadoop-m:~$ gsutil cp gs://bucket-elastic1/jars/elastic/elasticsearch-hadoop-8.14.1.jar .
CommandException: No URLs matched: gs://bucket-elastic1/jars/elastic/elasticsearch-hadoop-8.14.1.jar
p_aguinaga@cluster-hadoop-m:~$ gsutil cp gs://bucket-elastic1/elasticsearch-hadoop-8.14.1.jar .
Copying gs://bucket-elastic1/elasticsearch-hadoop-8.14.1.jar...
- [1 files][  2.1 MiB/  2.1 MiB]
Operation completed over 1 objects/2.1 MiB.
p_aguinaga@cluster-hadoop-m:~$ gsutil cp gs://bucket-elastic1/commons-httpclient-3.1.jar .
Copying gs://bucket-elastic1/commons-httpclient-3.1.jar...
- [1 files][297.8 KiB/297.8 KiB]
Operation completed over 1 objects/297.8 KiB.
p_aguinaga@cluster-hadoop-m:~$
```

PARTE 2- Configuración 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: 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: ["elasticsearch1"]

# 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 ------------------------
p_aguinaga@elasticsearch1:~$
```

PARTE 2- comprobación acceso desde el cluster Hadoop.

```
Last login: Tue Oct 29 23:19:42 2024 from 35.235.245.128
p_aguinaga@cluster-hadoop-m:~$ curl -I http://34.175.245.88:9200
HTTP/1.1 200 OK
X-elastic-product: Elasticsearch
content-type: application/json
content-length: 539

p_aguinaga@cluster-hadoop-m:~$
```

PARTE 3 - Configuración en Cluster Hadoop de Conexión con ES

```
p_aguinaga@cluster-hadoop-m:~$ sudo sed -i '$d' /etc/hive/conf.dist/hive-site.xml
p_aguinaga@cluster-hadoop-m:~$ sudo sed -i '$a \  <property>\n    <name>es.nodes</name>\n    <value>34.175.245.88</valu
e>\n  </property>\n' /etc/hive/conf.dist/hive-site.xml
p_aguinaga@cluster-hadoop-m:~$ sudo sed -i '$a \  <property>\n    <name>es.port</name>\n    <value>9200</value>\n  </pr
operty>\n' /etc/hive/conf.dist/hive-site.xml
p_aguinaga@cluster-hadoop-m:~$ sudo sed -i '$a \  <property>\n    <name>es.nodes.wan.only</name>\n    <value>true</valu
e>\n  </property>\n' /etc/hive/conf.dist/hive-site.xml
p_aguinaga@cluster-hadoop-m:~$ sudo sed -i '$a \  <property>\n    <name>hive.aux.jars.path</name>\n   <value>/usr/lib/h
ive/lib/elasticsearch-hadoop-8.14.1.jar,/usr/lib/hive/lib/commons-httpclient-3.1.jar</value>\n  </property>\n</configur
ation>' /etc/hive/conf.dist/hive-site.xml
p_aguinaga@cluster-hadoop-m:~$ sudo cp elasticsearch-hadoop-8.14.1.jar /usr/lib/hive/lib/
p_aguinaga@cluster-hadoop-m:~$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
p_aguinaga@cluster-hadoop-m:~$ sudo service hive-server2 restart
```

PARTE 4 Conexión de datos 1/2

```
p_aguinaga@cluster-hadoop-m:/usr/lib/hive/lib$ sudo service hive-server2 restart
p_aguinaga@cluster-hadoop-m:/usr/lib/hive/lib$ curl -X GET "http://34.175.18.121:9200/alumnos/_search?pretty"
{
  "took" : 2,
  "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" : {
          "id" : 4,
          "name" : "María",
          "last_name" : "López"
        }
      },
      {
        "_index" : "alumnos",
        "_id" : "5",
        "_score" : 1.0,
        "_source" : {
          "id" : 5,
          "name" : "Luis",
          "last_name" : "Martínez"
        }
      },
```

```
          "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"
        }
      }
    ]
  }
}
```

```
_aguinaga@cluster-hadoop-m:/usr/lib/hive/lib$ ||
```