

Elasticsearch for Apache Hadoop

Parte 1 – Configuración ES-Hadoop

Se crea clúster single-node con nombre hive-cluster, en la foto se realiza una petición al servicio de Elasticsearch.

```
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO
admba@hive-cluster-m:~$ curl -I http://34.73.127.59:9200
HTTP/1.1 200 OK
X-elastic-product: Elasticsearch
content-type: application/json
content-length: 547
admba@hive-cluster-m:~$
```

Parte 2 – Configuración server Elasticsearch

Se crea instancia instancia-elasticsearch se agrega a la red las ip de mi local y la de la instancia hive-cluster donde se encuentra el clúster, y se crean dos reglas firewall, allow-elasticsearch y allow-kibana. Por otra parte, se configura el archivo elasticsearch.yml

```
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO
# 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: ["node-1", "node-2"]
#
# 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 16-10-2024 23:52:21
#
# -----
#
# 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: true
  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-elasticsearch"]
#
# Allow HTTP API connections from anywhere
# Connections are encrypted and require user authentication
http.host: 0.0.0.0
http.port: 9200
# 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 -----
admba@instance-elasticsearch:~$
```

Parte 3 – Configuración del Clúster Hadoop de conexión

```
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO

admahhive-cluster-m1:~$ sudo sed -i /etc/hive/conf.dist/hive-site.xml
sed: -e expression #1, char 7: extra characters after command
admahhive-cluster-m1:~$ sudo sed -i 's# \<property>\n <name>es.nodes</name>\n <value>34.73.127.59</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
admahhive-cluster-m1:~$ sudo sed -i 's# \</property>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
admahhive-cluster-m1:~$ sudo sed -i 's# \<property>\n <name>es.port</name>\n <value>9200</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
admahhive-cluster-m1:~$ sudo sed -i 's# \<property>\n <name>es.nodes.wan.only</name>\n <value>true</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
admahhive-cluster-m1:~$ sudo sed -i 's# \<property>\n <name>hive.aux.jars.path</name>\n <value>/usr/lib/hive/lib/elasticsearch-hadoop-3.14.1-jar,/usr/lib/hive/lib/commons-httpclient-3.1.jar</value>\n </property>\n' /etc/hive/conf.dist/hive-site.xml
admahhive-cluster-m1:~$ cd /usr/local/share/google/dataproc/lib
admahhive-cluster-m1:~$ cp elasticsearch-hadoop-3.14.1-jar /usr/lib/hive/lib/
admahhive-cluster-m1:~$ cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
admahhive-cluster-m1:~$ sudo systemctl restart hive-server2
admahhive-cluster-m1:~$
```

Parte 4 – A conectar datos

Se crea a través del POST la creación del índice de alumnos y se realiza una petición GET al `_search` para obtener todos los documentos que pertenecen al índice.

```
admahhive-cluster-m1:~$ curl -X GET "http://34.73.127.59:9200/alumnos/_search?pretty"
{
  "took": 94,
  "timed_out": false,
  "_source": [
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  ],
  "hits": [
    {
      "total": 1,
      "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": "Maria",
            "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": "Sofia",
            "last_name": "Ramírez"
          }
        },
        {
          "index": "alumnos",
          "id": "8",
          "score": 1.0,
          "source": {
            "id": 8,
            "name": "Pedro",
            "last_name": "Hernández"
          }
        }
      ]
    }
  ]
}
```

Parte 5 – Kibana

