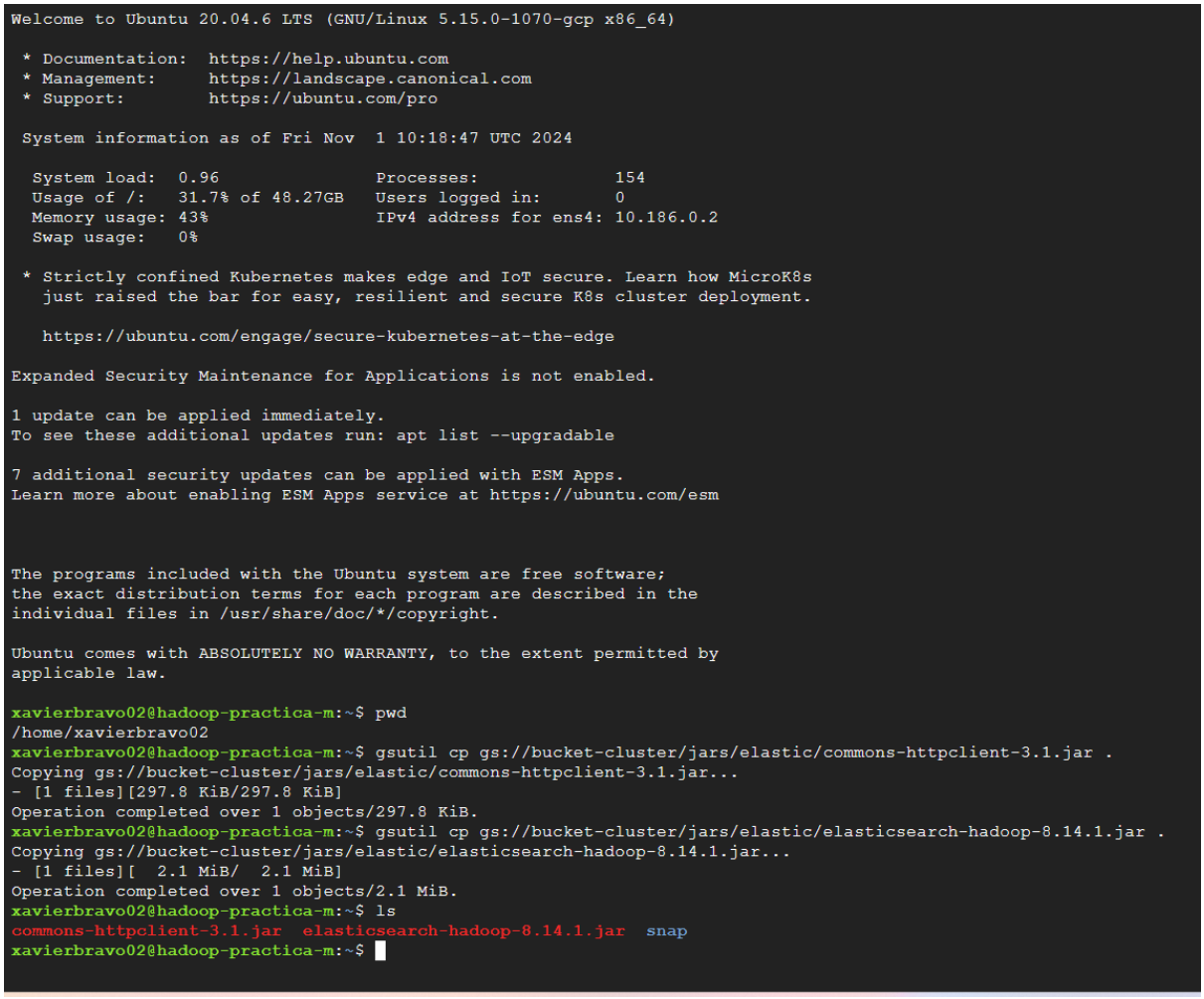
Xavier Bravo Cayancela

Práctica ElasticSearch_Hadoop

Primera entrega

En esta primera imagen, se muestra la copia de los archivos .jar “httpclient” y “elasticsearch-hadoop”, que fueron subidos al bucket de Google Cloud Storage y descargados al sistema local en una máquina con Ubuntu. Estos archivos son muy importantes, ya que permiten integrar el clúster de Hadoop con Elasticsearch, facilitando la comunicación mediante seguridad o conexión HTTP.



Segunda entrega

En la segunda imagen se puede ver una parte del archivo de configuración elasticsearch.yml en un servidor de Elasticsearch. Aquí se han ajustado algunos parámetros de seguridad, como activar SSL para las conexiones HTTP y la comunicación entre nodos. También se ha indicado un nodo maestro inicial y se ha permitido que el servidor acepte conexiones HTTP desde cualquier dirección IP, aunque la seguridad general de Elasticsearch (xpack.security) sigue desactivada.

```
path.data: /var/lib/elasticsearch
path.logs: /var/log/elasticsearch
network.host: 0.0.0.0
http.port: 9200
# Enable security features
xpack.security.enabled: false
xpack.security.enrollment.enabled: false
# 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: ["elk-stack-ubuntu20-04"]

# Allow HTTP API connections from anywhere
# Connections are encrypted and require user authentication
http.host: 0.0.0.0
```

Tercera entrega

En esta imagen se muestra el resultado del proceso de configuración, se integró el clúster Hadoop con Elasticsearch utilizando Hive. Para ello, se modificó el archivo `hive-site.xml` de Hive, agregando propiedades que definen la conexión al nodo de Elasticsearch, incluyendo la IP, el puerto, y el modo WAN. Además, se especificó la ruta de los JARs auxiliares necesarios (`elasticsearch-hadoop` y `commons-httpclient`) y se copiaron estos archivos a la carpeta de librerías de Hive (`/usr/lib/hive/lib/`). Finalmente, se reinició el servicio `hive-server2` para aplicar los cambios y asegurar la conexión de Hive con Elasticsearch, habilitando así la interoperabilidad entre ambos sistemas.

```
xavierbravo02@hadoop-practica-m:~$ sudo sed -i '$d' /etc/hive/conf.dist/hive-site.xml
xavierbravo02@hadoop-practica-m:~$ sudo sed -i '$a \  <property>\n    <name>es.nodes</name>\n    <value>10.186.0.3:9200</value>\n  </propert
y>\n' /etc/hive/conf.dist/hive-site.xml
xavierbravo02@hadoop-practica-m:~$ sudo sed -i '$a \  <property>\n    <name>es.port</name>\n    <value>9200</value>\n  </property>\n' /etc/h
ive/conf.dist/hive-site.xml
xavierbravo02@hadoop-practica-m:~$ sudo sed -i '$a \  <property>\n    <name>es.nodes.wan.only</name>\n    <value>true</value>\n  </property>
\n' /etc/hive/conf.dist/hive-site.xml
xavierbravo02@hadoop-practica-m:~$ sudo sed -i '$a \  <property>\n    <name>hive.aux.jars.path</name>\n    <value>/usr/lib/hive/lib/elasticse
arch-hadoop-8.14.1.jar,/usr/lib/hive/lib/commons-httpclient-3.1.jar</value>\n  </property>\n</configuration>' /etc/hive/conf.dist/hive-site.
xml
xavierbravo02@hadoop-practica-m:~$ sudo cp elasticsearch-hadoop-8.14.1.jar /usr/lib/hive/lib/
xavierbravo02@hadoop-practica-m:~$ sudo cp commons-httpclient-3.1.jar /usr/lib/hive/lib/
xavierbravo02@hadoop-practica-m:~$ sudo service hive-server2 restart
Warning: The unit file, source configuration file or drop-ins of hive-server2.service changed on disk. Run 'systemctl daemon-reload' to relo
ad units.
xavierbravo02@hadoop-practica-m:~$ sudo systemctl daemon-reload
xavierbravo02@hadoop-practica-m:~$ sudo service hive-server2 restart
xavierbravo02@hadoop-practica-m:~$
```

Entrega 4

La imagen adjuntada muestra la consulta desde el clúster Hadoop a un índice llamado `alumnos` en Elasticsearch, comprobando así que el índice fue creado y contiene datos accesibles. El comando 'curl' utilizado verifica que la conexión entre Hadoop y Elasticsearch está funcionando correctamente, ya que devuelve documentos con información de prueba, como nombres y apellidos. Confirmandoa que ambos requisitos, tanto la creación del índice en Elasticsearch como la consulta exitosa desde Hadoop.

```
xavierbravo02@hadoop-practica-m:~$ curl -X GET "http://10.186.0.3:9200/alum
nos/_search?pretty"
{
  "took" : 13,
  "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"
        }
      },
      {
```

Entrega 5

La imagen aquí muestra una visualización en la consola de Kibana en forma de gráfico de dona, cumpliendo con el requisito de "Captura de pantalla de la consola de Kibana con alguna visualización sencilla". El gráfico representa la distribución de nombres en el índice alumnos, con porcentajes asociados a cada nombre: Pedro (29.63%), Sofía (25.93%), Luis (18.52%), María (14.81%) y Carlos (11.11%). Esta visualización permite ver de manera rápida y clara la proporción de cada nombre dentro de los registros almacenados en Elasticsearch.