Instalación de un cluster Hadoop en una máquina virtual Ubuntu.

A la hora de realizar una instalación de un cluster Hadoop, se nos plantea la posibilidad de realizar dicha instalación en distintos soportes, por una parte, tenemos Windows o Ubuntu, otra opción puede ser un soporte Cloud qué nos permitirá usar el cluster sin necesidad de la instalación, sin embargo, en este caso, elegiremos Ubuntu por las ventajas que nos ofrece a la hora de rendimiento frente a Windows.

Materiales necesarios:

- Máquina virtual de Ubuntu 64 bits.
- Conexión a internet.
- Usuario administrador.

Pasos a realizar

1. Creación de usuario:

   Añadiremos un usuario nuevo al sistema que nos simplificará el trabajo y usará Hadoop, para ello utilizaremos comandos básicos de Linux y crearemos un usuario.

```
root@Ubuntu:/usr/local# sudo adduser hdoop
Adding user `hdoop' ...
Adding new group `hdoop' (1002) ...
Adding new user `hdoop' (1002) with group `hdoop' ...
Creating home directory `/home/hdoop' ...
Copying files from `/etc/skel' ...
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
passwd: password updated successfully
Changing the user information for hdoop
Enter the new value, or press ENTER for the default
        Full Name []: hdoop
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
```

2. Generamos un par de claves ssh para comunicarnos con el cluster.

   ```
   ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
   ```

   Guardaremos la clave pública con el siguiente comando:

   ```
   cat ~/.ssh/clave.pub >> ~/.ssh/authorized_key
   ```

Sin olvidar los permisos:

```
chmod 0600 ~/.ssh/authorized_keys
```

Testear con ssh localhost.

```
hdoop@Ubuntu:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ED25519 key fingerprint is SHA256:9nkpmIzGv1hy/VRk8JbDWMawR4WcZZsSKuhH35JVVkE.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.15.0-50-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

170 updates can be applied immediately.
86 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable


The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
```

3. Instalamos Java en el usuario principal.

   Tenemos dos opciones:

   JDK Builds from Oracle (java.net)

   La carpeta debe ir en la ruta /usr/local/.
   Añadir variables de entorno:

```
export JAVA_HOME=/usr/local/*OPENJDK*
export PATH=$PATH:$JAVA_HOME/bin
```

   Guardamos las variables de entorno con el siguiente comando:

```
source ~/.bashrc
```

   Verificamos si es correcto con `java -version`

```
root@Ubuntu:/home/vboxuser# java -version
openjdk version "19.0.1" 2022-10-18
OpenJDK Runtime Environment (build 19.0.1+10-21)
OpenJDK 64-Bit Server VM (build 19.0.1+10-21, mixed mode, sharing)
```

Otra opción es con el siguiente comando:
sudo apt install openjdk-8-jdk -y

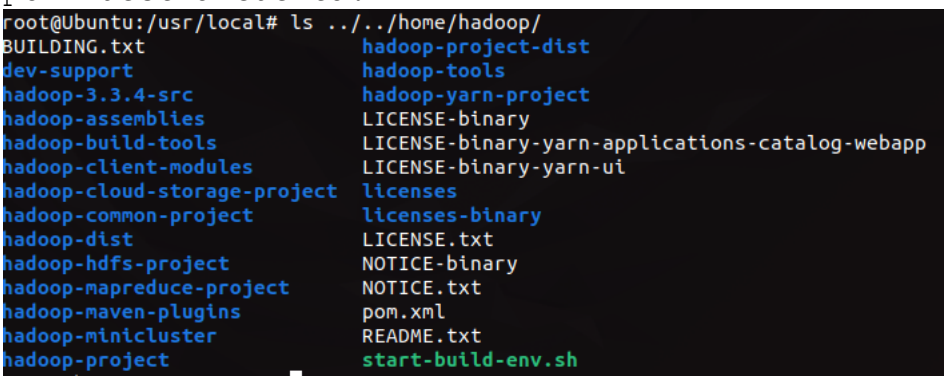
4. Descargamos Hadoop en su usuario.

   Para la descarga efectuaremos los siguientes comandos:

   Wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz

   tar xzf *Carpeta comprimida*

   También podemos descargar el código fuente y compilarlo por nuestra cuenta:



5. Definir modos, en este caso, trabajaremos en local. Para funcionar en modo local deberemos hacer lo siguiente:

   Abrimos con un editor de texto el archivo .bashrc
   Y añadimos las siguientes variables de entorno.

```
export HADOOP_HOME=/home/hdoop/hadoop...
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export
HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export
HADOOP_OPTS"-Djava.library.path=$HADOOP_HOME/lib/native"
```

   Para guardarlas utilizaremos el siguiente comando:

```
source ~/.bashrc
```

Si todo ha ido bien ejecuta el comando hadoop version, tiene que dar el siguiente resultado:

```
hdoop@Ubuntu:~$ hadoop version
Hadoop 3.3.4
Source code repository https://github.com/apache/hadoop.git -r a585a73c3e02ac62350c136643a5e7f6095a3dbb
Compiled by stevel on 2022-07-29T12:32Z
Compiled with protoc 3.7.1
From source with checksum fb9dd8918a7b8a5b430d61af858f6ec
This command was run using /home/hdoop/hadoop/share/hadoop/common/hadoop-common-3.3.4.jar
```