



Royaume du Maroc
Ministère de l'Enseignement Supérieur,
de la Recherche Scientifique et de l'Innovation
Université Sultan Moulay Slimane
L'Ecole Supérieure de Technologie -- Fkih Ben Salah



Filière : licence BIG DATA
Année Universitaire : 2023-2024

Sujet

MISE EN ŒUVRE DU HBASE DANS UN CLUSTER HADOOP

Réalisé par :
Iliasse KABLANI

Dans ce guide, nous allons configurer HBase sur un Hadoop Multi-Node Cluster de 3 machines, dans lesquels l'un sera notre nœud master et les autres sont des workers.

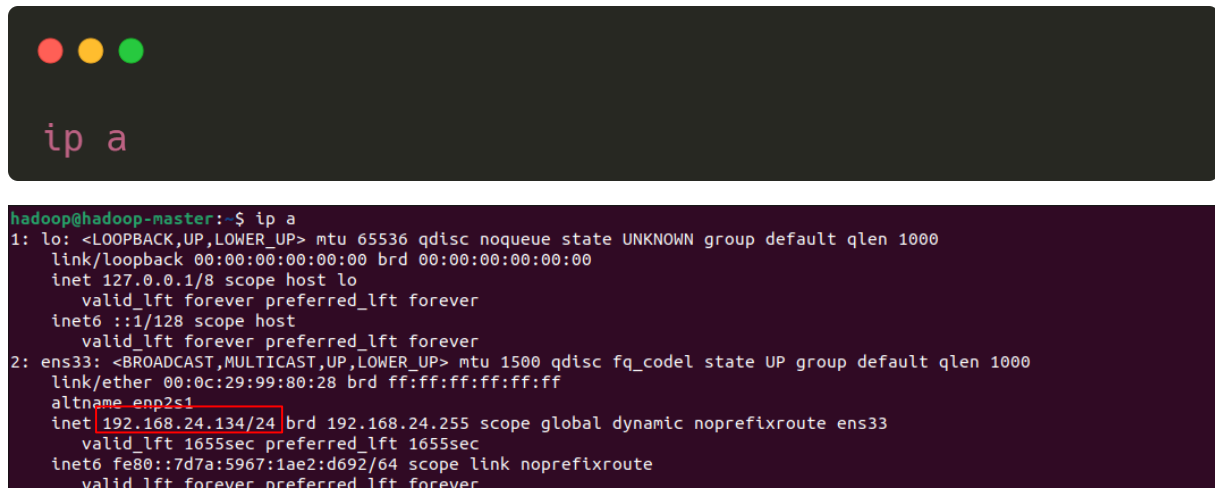
Nous avons besoin de 3 machines avec Ubuntu en tant que système d'exploitation.

Les 3 nœuds sont sous les noms suivants :

Hadoop-master, hadoop-worker, hadoop-worker2.

1-Mise en œuvre d'un cluster hadoop en 3 nœuds :

Etape 1 : Récupérer les adresses IP des 3 machines : Sur les 3 machines.



```
hadoop@hadoop-master:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens33: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 00:0c:29:99:80:28 brd ff:ff:ff:ff:ff:ff
    altname enp2s1
    inet 192.168.24.134/24 brd 192.168.24.255 scope global dynamic noprefixroute ens33
        valid_lft 1655sec preferred_lft 1655sec
    inet6 fe80::7d7a:5967:1ae2:d692/64 scope link noprefixroute
        valid_lft forever preferred_lft forever
```

Pour moi j'ai récupéré les adresses suivants :

192.168.24.134

192.168.24.136

192.168.24.137

Etape 2 : Attribuer des noms à ces adresses IP : Sur les 3 machines

On accède au fichier hosts et on ajoute les adresses IP avec ses noms pour qu'ils puissent communiquer entre eux à travers le nom au lieu d'adresse IP.

Dans notre cas :

192.168.24.134 hadoop-master

192.168.24.136 hadoop-worker2

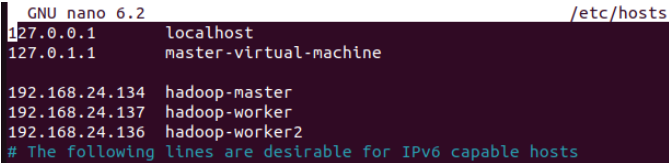
192.168.24.137 hadoop-worker

On accède au fichier hosts :




```
sudo nano /etc/hosts
```

Et on ajout ces lignes :



```
GNU nano 6.2 /etc/hosts
127.0.0.1 localhost
127.0.1.1 master-virtual-machine
192.168.24.134 hadoop-master
192.168.24.137 hadoop-worker
192.168.24.136 hadoop-worker2
# The following lines are desirable for IPv6 capable hosts
```

Etape 3 : On crée utilisateur sous le nom hadoop : Sur les 3 machines



```
sudo adduser hadoop
sudo adduser hadoop sudo
```

Etape 4 : On change le nom des machines : Sur les 3 machines



```
sudo nano /etc/hostname
```



```
GNU nano 6.2 /etc/hostname *
hadoop-master
```

Et on fait la même chose pour hadoop-worker et hadoop-worker2

Etape 5 : On redémarre les machines : Sur les 3 machines

Etape 6 : Configuration de SSH : Sur hadoop-master :



```
ssh-keygen -t rsa
```

```
hadoop@hadoop-master:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa):
Created directory '/home/hadoop/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:a6j8BUZZETiEgsJa67afly44SytgmTeuHeyZ5GCdyITU hadoop@hadoop-master
The key's randomart image is:
+---[RSA 3072]-----+
|+ .....=|
|oE = |
|.+. = |
|+ . o |
|.. .. S|
|oo=... ..|
|=B.O+ . o|
|=OX=*o .|
|=*@*.|
+---[SHA256]-----+
```

```
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop@hadoop-master
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop@hadoop-slave
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop@hadoop-slave2
```

- Pour chaque ligne ils vont demander le mot de passe de l'utilisateur hadoop de la machine mentionner dans la ligne.
- Cette configuration permet les machines de communiquer entre eux sans besoin de taper le mot de passe à chaque fois.

Etape 7 : Installation de JAVA sur les machines : Sur les 3 machines

```
sudo apt install openjdk-8-jdk
```

Etape 8 : Téléchargement et extraction du hadoop : Sur les 3 nœuds

```
wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.6.tar.gz
tar -xzf hadoop-3.3.6.tar.gz
mv hadoop-3.3.6 hadoop
```

Etape 9 :Ajouter Path Java dans l'environnement hadoop: Sur 3 machines



```
nano hadoop/etc/hadoop/hadoop-env.sh
```

Et on ajoute ce chemin dans le fichier :

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line
```

Etape 10 : Ajouter les variables d'environnements de hadoop dans .bashrc :
Sur les 3 machines



```
nano ~/.bashrc
```

```
#Hadoop Related Options
export HADOOP_HOME="/home/hadoop/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"
```



```
source ~/.bashrc
```

Etape 11 : Créer les dossiers du namenode et datanode : sur les 3 machines



```
mkdir data
mkdir data/namenode data/datanode
```

Etape 12 : Configuration du hadoop : sur les 3 machines



```
nano hadoop/etc/hadoop/core-site.xml
```

- On ajout cette configuration :

```
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://hadoop-master:9000</value>
</property>
</configuration>
```

- On accede maintenant au hdfs-site.xml :



```
nano hadoop/etc/hadoop/hdfs-site.xml
```

- On ajout la configuration suivante :

```
<configuration>
<property>
<name>dfs.name.dir</name>
<value>/home/hadoop/data/namenode</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/home/hadoop/data/datanode</value>
</property>
<property>
<name>dfs.replication</name>
<value>2</value>
</property>
</configuration>
```

- On accede maintenant au fichier mapred-site.xml :



```
nano hadoop/etc/hadoop/mapred-site.xml
```

- On ajout cette configuration :

```
<configuration>
<property>
<name>mapred.job.tracker</name>
```

```
<value>hadoop-master:9001</value>
</property>
</configuration>
```

- On accede au fichier workers :



```
nano hadoop/etc/hadoop/workers
```

- On ajout les nom des nœuds workers :

Pour notre cas :

hadoop-worker


hadoop-worker2



```
GNU nano 6.2 hadoop/etc/hadoop/workers
hadoop-worker
hadoop-worker2
```

Etape 13 : Configuration de yarn sur les workers : Seulement les workers

- On accede au fichier yarn-site.xml :

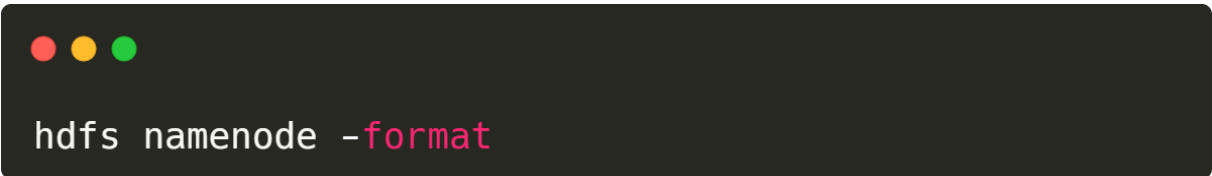


```
nano hadoop/etc/hadoop/yarn-site.xml
```

- On ajout cette configuration :

```
<configuration>
<property>
<name>yarn.resourcemanager.hostname</name>
<value>hadoop-master</value>
</property>
</configuration>
```

Etape 14 : Formatage du namenode : Seulement dans master



```
hdfs namenode -format
```

Etape 15 : Démarrer Hadoop : Seulement dans master



```
start-all.sh
```

2-Configuration de zookeeper sur ces 3 nœuds :

Etape 1 :Créer un dossier pour les données de zookeeper : sur les 3 nœuds

On accede au repertoide courant

```
mkdir data/zookeeper
```

Etape 2 : Téléchargement de zookeeper : sur les 3 nœuds :

```
wget https://www.apache.org/dyn/closer.lua/zookeeper/zookeeper-3.8.3/apache-zookeeper-3.8.3-bin.tar.gz
tar -xvf apache-zookeeper-3.8.3-bin.tar.gz
mv apache-zookeeper-3.8.3-bin hadoop/zookeeper
```

Etape 3 : Configuration de zookeeper : sur les 3 nœuds

On créer un fichier de configuration :

```
nano hadoop/zookeeper/conf/zoo.cfg
```

On ajoute cette configuration :

```
tickTime=2000
dataDir=/home/hadoop/data/zookeeper
clientPort=2181
maxClientCnxns=60
initLimit=10
syncLimit=5
server.1=hadoop-master:2888:3888
server.2=hadoop-worker2:2888:3888
server.3=hadoop-worker:2888:3888
```

On crée un fichier myid dans le repertoire des données de zookeeper :

- Pour hadoop-master

```
echo "1" > /home/hadoop/data/zookeeper/myid
```


- Pour hadoop-worker2 :

```
echo "2" > /home/hadoop/data/zookeeper/myid
```

- Pour hadoop-worker :

```
echo "3" > /home/hadoop/data/zookeeper/myid
```

Etape 4 : Les variables d'environnement dans .bashrc : Sur les 3 nœuds :

Ouvre le fichier .bashrc :

```
nano ~/.bashrc
```

Ajouter ces lignes dans ce fichier :

```
# Set ZOOKEEPER_HOME
export ZOOKEEPER_HOME=/home/hadoop/hadoop/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin
```

```
source ~/.bashrc
```

Etape 5 : Demarrer zookeeper : Sur les 3 nœuds :

```
zkServer.sh start
```

3-Installation et configuration HBase sur 3 nœuds :

Etape 1 :Téléchargement de hbase dans les 3 nœuds :

```
wget https://www.apache.org/dyn/closer.lua/hbase/2.4.17/hbase-2.4.17-bin.tar.gz
tar -xzf hbase-2.4.17-bin.tar.gz
mv hbase-2.4.17 hadoop/hbase
```

Etape 2 : Variables d'environnement au fichier .bashrc : Sur les 3 nœuds :

```
nano ~/.bashrc
```

Ajouter ces variables dans .bashrc :

```
#HBASE CONFIGS
export HBASE_HOME="/home/hadoop/hadoop/hbase"
export PATH="$HBASE_HOME/bin:$PATH"
```

```
source ~/.bashrc
```

Etape 3 : Ajout de chemin Java dans hbase-env.sh : Sur les 3 nœuds

```
nano ~/hadoop/hbase/conf/hbase-env.sh
```

Ajouter chemin dans ce fichier :

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
export HBASE_MANAGES_ZK=false
```

Etape 4 : Configuration de hadoop-master : Seulement pour hadoop-master

Accéder au fichier hbase-site.xml :

```
nano ~/hadoop/hbase/conf/hbase-site.xml
```

Ajouter cette configuration :

```
<configuration>
<property>
  <name>hbase.rootdir</name>
  <value>hdfs://hadoop-master:9000/hbase</value>
</property><property>
  <name>hbase.cluster.distributed</name>
  <value>true</value>
</property><property>
  <name>hbase.zookeeper.property.dataDir</name>
  <value>/home/hadoop/data/zookeeper</value>
</property><property>
  <name>hbase.zookeeper.quorum</name>
  <value>hadoop-master, hadoop-worker, hadoop-worker2</value>
</property><property>
  <name>hbase.zookeeper.property.clientPort</name>
  <value>2181</value>
</property>
</configuration>
```

Puis Ajouter les deux workers au fichier regionservers :

```
echo 'hadoop-worker' > ~/hadoop/hbase/conf/regionservers
echo 'hadoop-worker2' >> ~/hadoop/hbase/conf/regionservers
```

Etape 5 : Configuration des workers :

Accéder au fichier hbase-site.xml :

```
nano ~/hadoop/hbase/conf/hbase-site.xml
```

Ajouter cette configuration :

```
<configuration>
<property>
<name>hbase.rootdir</name>
<value>hdfs://hadoop-master:9000/hbase</value>
</property>
<property>
<name>hbase.cluster.distributed</name>
<value>true</value>
</property>
</configuration>
```

Puis Ajouter les deux workers au fichier regionservers :

```
echo 'hadoop-worker' > ~/hadoop/hbase/conf/regionservers  
echo 'hadoop-worker2' >> ~/hadoop/hbase/conf/regionservers
```

Etape 6 : Donner certaines permissions a l'utilisateur hadoop :

```
sudo chmod -R 777 /home/hadoop/hadoop/hbase  
sudo chmod -R 777 /var
```

Etape 7 : Démarrer hadoop , zookeeper et HBase :

- Démarrage de hadoop : Seulement dans hadoop-master :

```
start-all.sh
```

- Démarrage de zookeeper : Sur les 3 nœuds :

```
zkServer.sh start
```

- Démarrage de HBase :

```
start-hbase.sh
```

- L'interface de commande de HBase :

```
hbase shell
```