INPT
المعهد الوطني للبريد والمواصلات
ⵜⴰⵙⴷⴰⵡⵜ ⵏ ⵜⵓⵔⵉⵜ ⴷ ⵜⵙⵡⵉⵍⵜ
Institut National des Postes et Télécommunications

anrt
agence nationale de réglementation des télécommunications
الوكالة الوطنية لتقنين المواصلات

19 NOVEMBRE 2022

Rapport : Installation et configuration d'Apache Hadoop

et exécution d'un programme MapReduce dans un cluster hadoop à nœud unique et à nœuds multiples.

Réalisé par

Ayoub TALBI
Mounir Bencherif

Encadré par

Mme. Dounia ZAIDOUNI

Pré-requis :

Après avoir créé une machine virtuelle avec le système d'exploitation open source **ubuntu-22.04.1 LTS** en utilisant VirtualBox, on installe le package dkms avec la commande :
$ sudo apt-get install dkms
et on lance les commandes suivantes avant de redémarrer la machine :
$ sudo groupadd vboxusers
$ sudo usermod -G vboxusers -a ayoub

1) Installation et configuration d'un nœud unique d'Apache Hadoop 3.3.4 :

1ère étape : Ajout d'un utilisateur sudo **'hdtalbibencherif'**
2ème étape : Mise en place de la clé SSH :

Installation de paquet openssh-server :

```
hdtalbibencherif@ayoub-VirtualBox:~$ sudo apt-get install openssh-serv
er
[sudo] password for hdtalbibencherif:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  ncurses-term openssh-sftp-server ssh-import-id
Suggested packages:
  molly-guard monkeysphere ssh-askpass
The following NEW packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
0 upgraded, 4 newly installed, 0 to remove and 0 not upgraded.
Need to get 751 kB of archives.
After this operation, 6,046 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://ma.archive.ubuntu.com/ubuntu jammy/main amd64 openssh-sft
p-server amd64 1:8.9p1-3 [38.8 kB]
Get:2 http://ma.archive.ubuntu.com/ubuntu jammy/main amd64 openssh-ser
ver amd64 1:8.9p1-3 [434 kB]
Get:3 http://ma.archive.ubuntu.com/ubuntu jammy/main amd64 ncurses-ter
```

Mise en place de la clé ssh :

```
hdtalbibencherif@ayoub-VirtualBox:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hdtalbibencherif/.ssh/id_rs
a):
Created directory '/home/hdtalbibencherif/.ssh'.
Your identification has been saved in /home/hdtalbibencherif/.ssh/id_r
sa
Your public key has been saved in /home/hdtalbibencherif/.ssh/id_rsa.p
ub
The key fingerprint is:
SHA256:xqglkuhmlyHmqk7zkCNFzjyCABsoYbNTio6WQDcGHWI hdtalbibencherif@ay
oub-VirtualBox
The key's randomart image is:
+---[RSA 3072]----+
|=Eo*.            |
|OoO..            |
|*+.              |
|**o.   o         |
|=*O.. o S        |
|=oooo+ .         |
|.@ o.            |
|* *              |
|=. .             |
+----[SHA256]-----+
```

1On copie la clé public sur le server localhost

```
hdtalbibencherif@ayoub-VirtualBox:~$ ssh-copy-id -i /home/hdtalbibench
erif/.ssh/id_rsa.pub hdtalbibencherif@ayoub-VirtualBox
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/h
dtalbibencherif/.ssh/id_rsa.pub"
The authenticity of host 'ayoub-virtualbox (127.0.1.1)' can't be estab
lished.
ED25519 key fingerprint is SHA256:oQld88scwQ6JgUdr37D73+eYRqx+UVWByd9o
sfySn7Y.
This host key is known by the following other names/addresses:
    ~/.ssh/known_hosts:1: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])? y
es
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s),
to filter out any that are already installed

/usr/bin/ssh-copy-id: WARNING: All keys were skipped because they alre
ady exist on the remote system.
                (if you think this is a mistake, you may want to use -
f option)
```

On teste la connexion a localhost par la commande « ssh localhost »

```
hdtalbibencherif@ayoub-VirtualBox:~$ ssh localhost
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.15.0-43-generic x86_64)

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

0 updates can be applied immediately.

Last login: Fri Nov 18 17:28:22 2022 from 127.0.0.1
hdtalbibencherif@ayoub-VirtualBox:~$ exit
logout
Connection to localhost closed.
hdtalbibencherif@ayoub-VirtualBox:~$
```

3ème étape : Installation de JAVA 8
On crée le répertoire /opt/java et on extrait l'archive

```
hdtalbibencherif@ayoub-VirtualBox:~$ sudo -i
[sudo] password for hdtalbibencherif:
root@ayoub-VirtualBox:~# mkdir /opt/java
root@ayoub-VirtualBox:~# cd /home/hdtalbibencherif/Documents
root@ayoub-VirtualBox:/home/hdtalbibencherif/Documents# tar -zxvf jdk-
8u71-linux-x64.tar.gz
jdk1.8.0_71/
jdk1.8.0_71/db/
jdk1.8.0_71/db/lib/
jdk1.8.0_71/db/lib/derbyLocale_pl.jar
jdk1.8.0_71/db/lib/derbyLocale_zh_TW.jar
jdk1.8.0_71/db/lib/derbyoptionaltools.jar
jdk1.8.0_71/db/lib/derbyLocale_cs.jar
jdk1.8.0_71/db/lib/derbyLocale_de_DE.jar
jdk1.8.0_71/db/lib/derbytools.jar
jdk1.8.0_71/db/lib/derbyrun.jar
jdk1.8.0_71/db/lib/derbyLocale_hu.jar
jdk1.8.0_71/db/lib/derbynet.jar
jdk1.8.0_71/db/lib/derby.jar
jdk1.8.0_71/db/lib/derbyLocale_ko_KR.jar
jdk1.8.0_71/db/lib/derbyLocale_it.jar
```

On utilise la commande update-alternatives pour dire au système où java et ses exécutables sont installés.

```
root@ayoub-VirtualBox:/opt/java/jdk1.8.0_71# update-alternatives --ins
tall /usr/bin/java java /opt/java/jdk1.8.0_71/bin/java 100
update-alternatives: using /opt/java/jdk1.8.0_71/bin/java to provide /
usr/bin/java (java) in auto mode
root@ayoub-VirtualBox:/opt/java/jdk1.8.0_71# update-alternatives --con
fig java
There is only one alternative in link group java (providing /usr/bin/j
ava): /opt/java/jdk1.8.0_71/bin/java
Nothing to configure.
root@ayoub-VirtualBox:/opt/java/jdk1.8.0_71#
```

On modifie le fichier **/etc/profile** comme ci-dessous

Après avoir enregistré le fichier profile, on exécute la commande suivante pour recharger le fichier (en tant que root et avec l'utilisateur hadoop):

$ source /etc/profile

```
root@ayoub-VirtualBox:/opt/java/jdk1.8.0_71# update-alternatives --ins
tall /usr/bin/java java /opt/java/jdk1.8.0_71/bin/java 100
update-alternatives: using /opt/java/jdk1.8.0_71/bin/java to provide /
usr/bin/java (java) in auto mode
root@ayoub-VirtualBox:/opt/java/jdk1.8.0_71# update-alternatives --con
fig java
There is only one alternative in link group java (providing /usr/bin/j
ava): /opt/java/jdk1.8.0_71/bin/java
Nothing to configure.
root@ayoub-VirtualBox:/opt/java/jdk1.8.0_71#
```

On modifie le fichier **/etc/profile** comme ci-dessous

Après avoir enregistré le fichier profile, on exécute la commande suivante pour recharger le fichier (en tant que root et avec l'utilisateur hadoop):

$ source /etc/profile

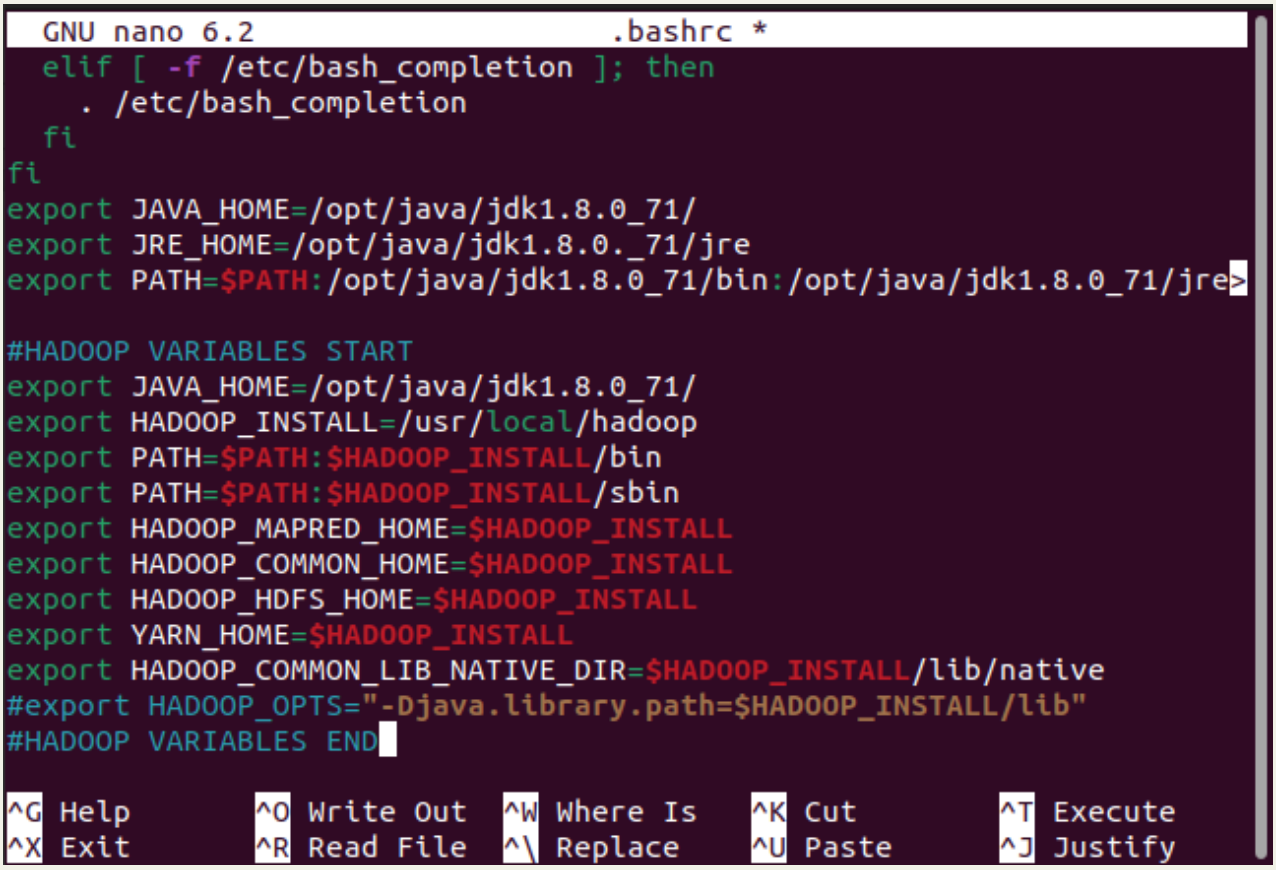
et on ajoute les mêmes lignes dans le fichier **~/.bashrc** pour mettre en place mettre en place de manière permanente les variables d'environnement JAVA pour un utilisateur unique.

4ème étape : Installation d'Apache Hadoop 3.3.4
On se déplace vers le répertoire /Documents Puis on exécute la commande
tar -zxvf hadoop-3.3.1.tar.gz

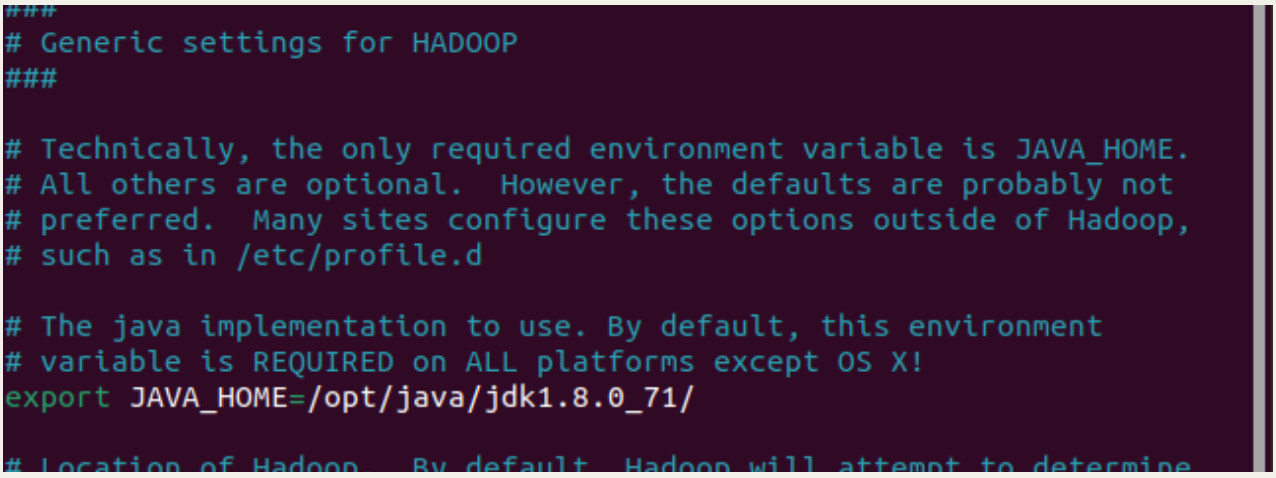
```
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logo_maven.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_warning_sml.
gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logos/
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logos/build-by-ma
ven-black.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logos/maven-feath
er.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logos/build-by-ma
ven-white.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/banner.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/h5.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_error_sml.gi
f
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_success_sml.
gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/expanded.gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/external.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/icon_info_sml.gif
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/logo_apache.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/bg.jpg
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/newwindow.png
hadoop-3.3.4/share/doc/hadoop/hadoop-hdfs-nfs/images/h3.jpg
hdtalbibencherif@ayoub-VirtualBox:~/Documents$
```

```
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ mv hadoop-3.3.4 hadoop
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo mv hadoop /usr/loc
al/hadoop/
[sudo] password for hdtalbibencherif:
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo chown -R hduser /u
sr/local/hadoop
chown: invalid user: ‘hduser’
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo chown -R hdtalbibe
ncherif /usr/local/hadoop
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo mkdir -p /usr/loca
l/hadoop_store/hdfs/namenode
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo mkdir -p /usr/loca
l/hadoop_store/hdfs/datanode
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo chown -R hdtalbibe
ncherif /usr/local/hadoop_store
hdtalbibencherif@ayoub-VirtualBox:~/Documents$
```

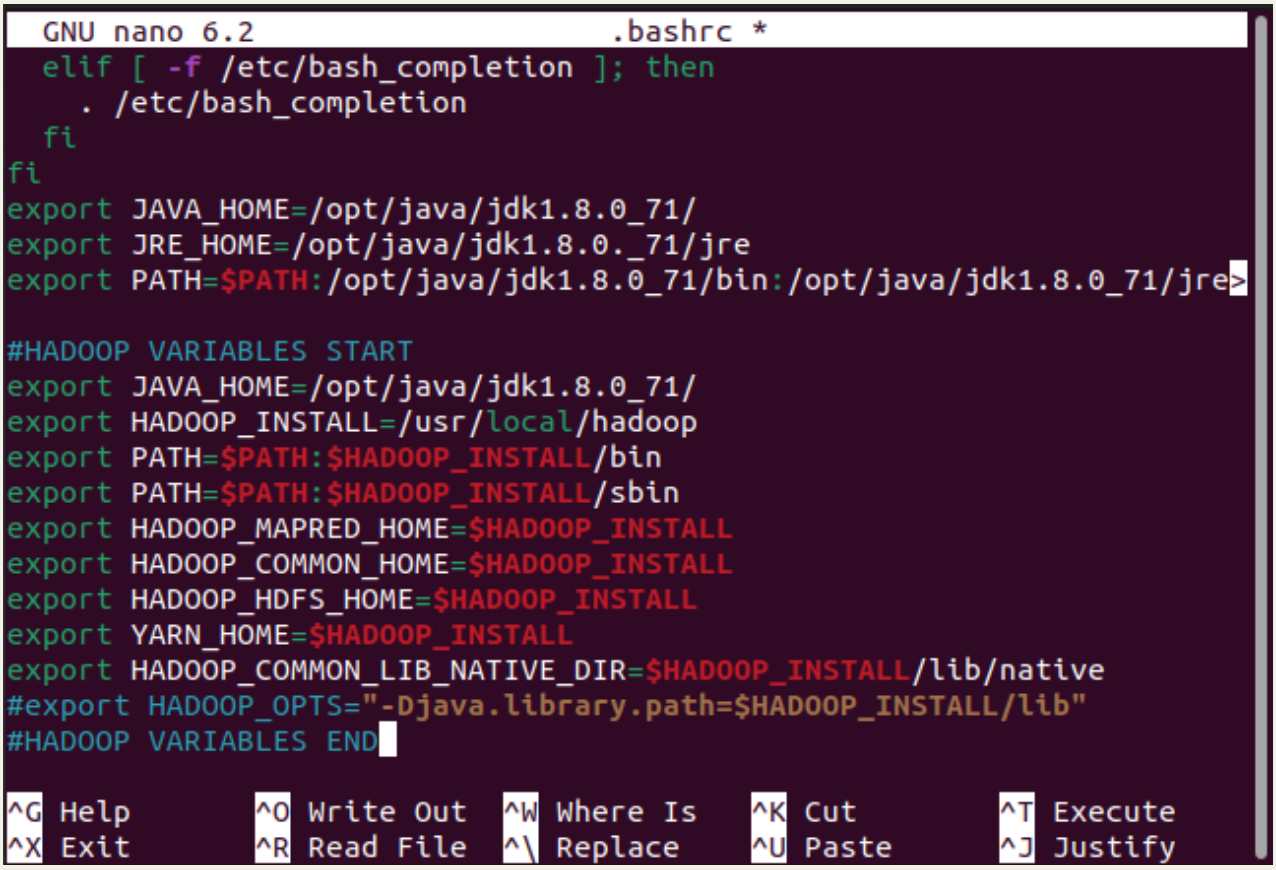
5ème étape : Configuration d'Apache Hadoop 3.3.4
- Mise en place des variables d'environnements :
On modifie le fichier **.bashrc** en ajoutant les lignes suivantes à la fin du fichier :



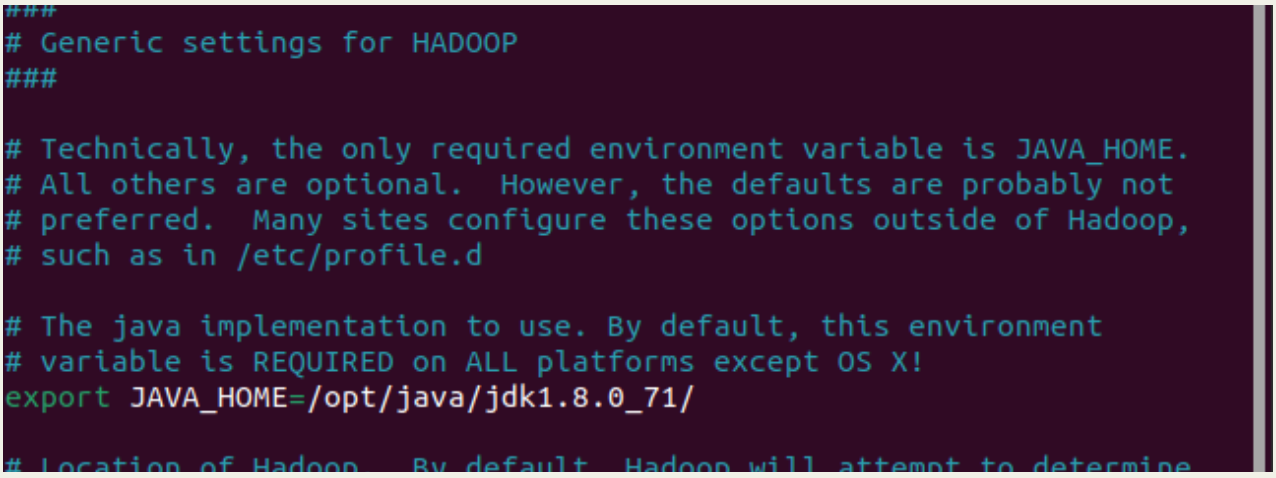
On recharge le fichier **~/.bashrc** en utilisant la fonction **source**
Maintenant, on ouvre le fichier **/usr/local/hadoop/etc/hadoop/hadoop-env.sh** avec **nano** et modifiez la variable d'environnement **JAVA_HOME** :



5ème étape : Configuration d'Apache Hadoop 3.3.4
- Mise en place des variables d'environnements :
On modifie le fichier **.bashrc** en ajoutant les lignes suivantes à la fin du fichier :



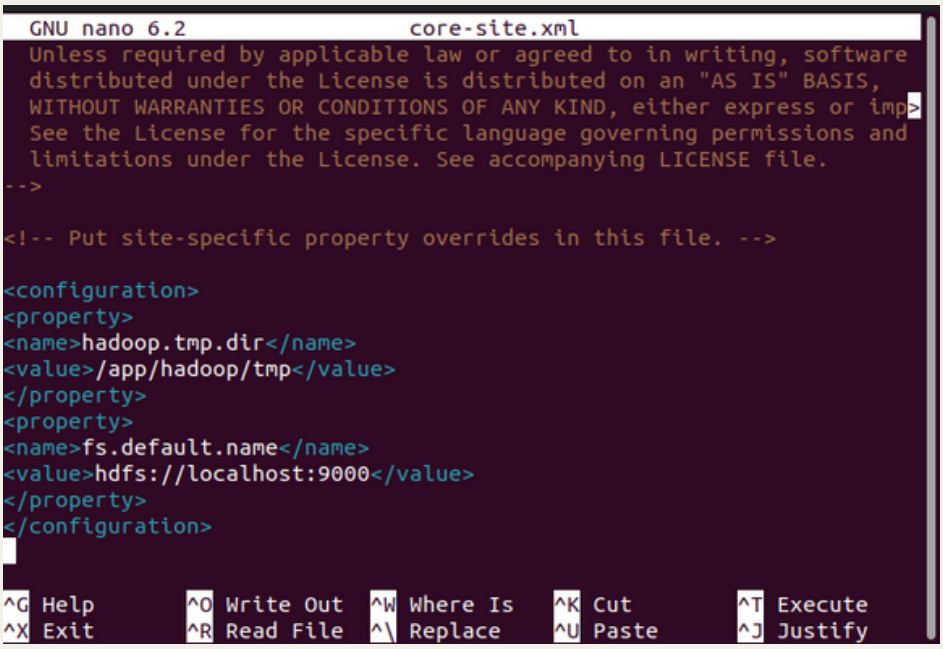
On recharge le fichier **~/.bashrc** en utilisant la fonction **source**
Maintenant, on ouvre le fichier **/usr/local/hadoop/etc/hadoop/hadoop-env.sh** avec **nano** et modifiez la variable d'environnement **JAVA_HOME** :

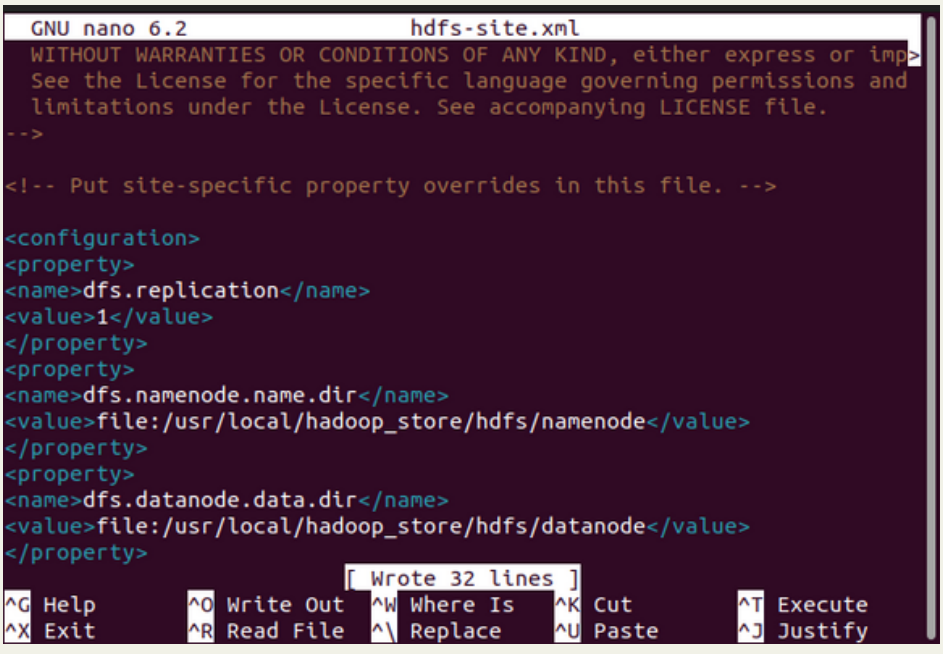


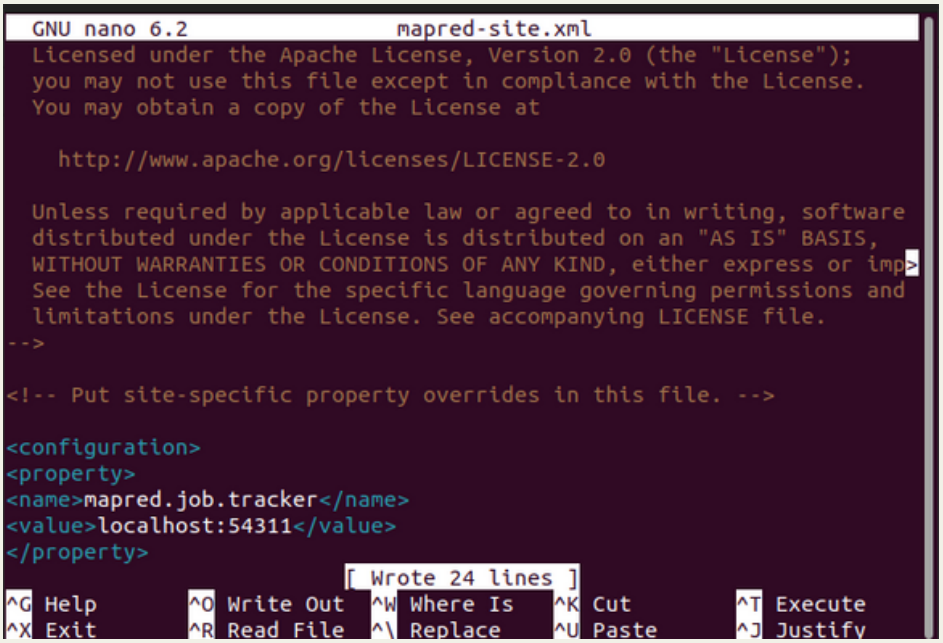
On crée le répertoire des fichiers temporaires de hadoop :

```
hdtalbibencherif@ayoub-VirtualBox:~$ sudo mkdir -p /app/hadoop/tmp
hdtalbibencherif@ayoub-VirtualBox:~$ sudo chown hduser /app/hadoop/tmp
chown: invalid user: ‘hduser’
hdtalbibencherif@ayoub-VirtualBox:~$ sudo chown hdtalbibencherif /app/
hadoop/tmp
hdtalbibencherif@ayoub-VirtualBox:~$
```

On se déplace dans **/usr/local/hadoop/etc/hadoop/** et on modifie les fichiers **core-site.xml** , **hdfs-site.xml** , **mapred-site.xml** et **yarn-site.xml** respectivement comme ci-dessous :






GNU nano 6.2 mapred-site.xml
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at
http://www.apache.org/licenses/LICENSE-2.0
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or imp>
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:54311</value>
</property>
[Wrote 24 lines]
^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify


GNU nano 6.2 yarn-site.xml *
You may obtain a copy of the License at
http://www.apache.org/licenses/LICENSE-2.0
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or imp>
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify

Formatage du Namenode :

```
ntries
2022-11-18 17:33:53,994 INFO namenode.FSImage: Allocated new BlockPool
Id: BP-1887966517-127.0.1.1-1668792833862
2022-11-18 17:33:54,101 INFO common.Storage: Storage directory /usr/lo
cal/hadoop_store/hdfs/namenode has been successfully formatted.
2022-11-18 17:33:54,317 INFO namenode.FSImageFormatProtobuf: Saving im
age file /usr/local/hadoop_store/hdfs/namenode/current/fsimage.ckpt_00
0000000000000000 using no compression
2022-11-18 17:33:54,882 INFO namenode.FSImageFormatProtobuf: Image fil
e /usr/local/hadoop_store/hdfs/namenode/current/fsimage.ckpt_000000000
0000000000 of size 411 bytes saved in 0 seconds .
2022-11-18 17:33:54,932 INFO namenode.NNStorageRetentionManager: Going
 to retain 1 images with txid >= 0
2022-11-18 17:33:55,070 INFO namenode.FSNamesystem: Stopping services
started for active state
2022-11-18 17:33:55,073 INFO namenode.FSNamesystem: Stopping services
started for standby state
2022-11-18 17:33:55,095 INFO namenode.FSImage: FSImageSaver clean chec
kpoint: txid=0 when meet shutdown.
2022-11-18 17:33:55,096 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ayoub-VirtualBox/127.0.1.1
************************************************************/
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop/etc/hadoop$
```

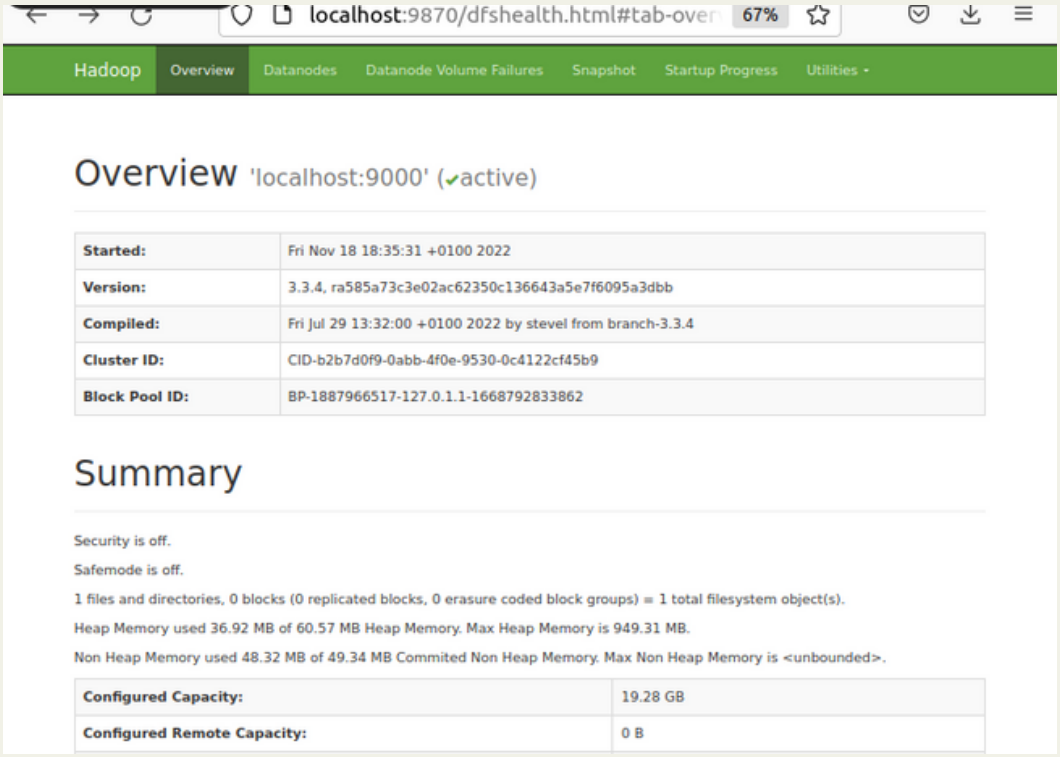
Maintenant, il est temps de démarrer le cluster à nœud unique nouvellement installé. Pour cela, on lance les scripts start-dfs.sh et start-yarn.sh :

```
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop/etc/hadoop$ start-
dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ayoub-VirtualBox]
2022-11-18 17:35:53,165 WARN util.NativeCodeLoader: Unable to load nat
ive-hadoop library for your platform... using builtin-java classes whe
re applicable
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop/etc/hadoop$
```

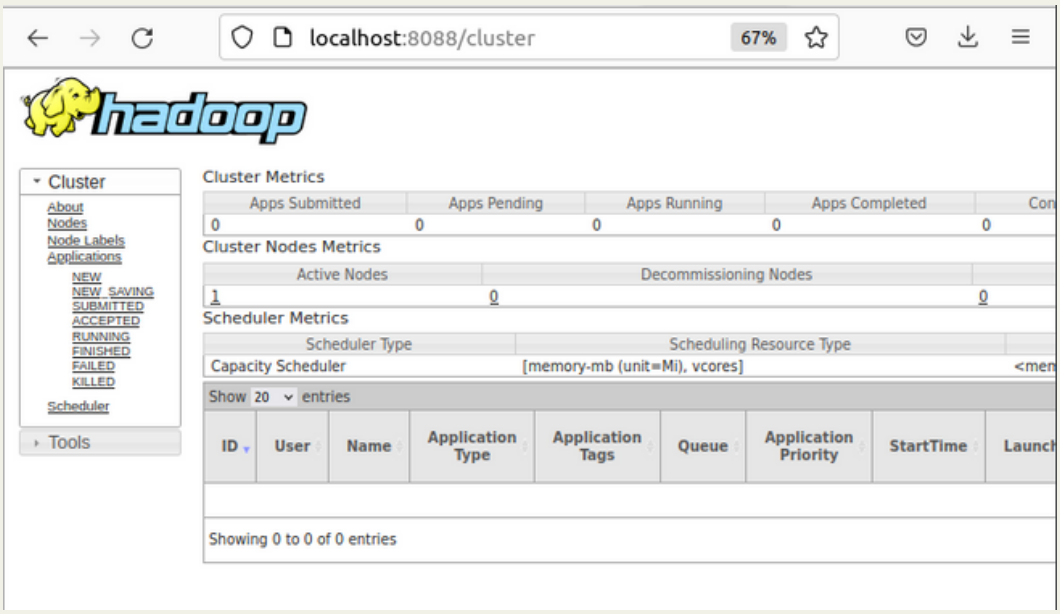
on peut vérfier si tous les composants du cluster fonctionnent avec la commance jps

```
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop/etc/hadoop$ jps
7680 ResourceManager
7425 SecondaryNameNode
7175 NameNode
8119 Jps
7784 NodeManager
7273 DataNode
```

On accède aux interfaces graphiques de Hadoop via le navigateur : Hadoop NameNode démarre sur le port 9870 par défaut. Accédez à http://localhost: 9870 / dans votre navigateur web préféré :



On accède à l'interface du ResourceManager à : http://localhost:8088/ pour obtenir les informations sur le cluster et toutes les applications :



2) Exécution d'un programme Map/Reduce dans un cluster à nœud unique :

hdfs dfsadmin -report

```
-------------------------------------------------
Live datanodes (1):

Name: 127.0.0.1:9866 (localhost)
Hostname: ayoub-VirtualBox
Decommission Status : Normal
Configured Capacity: 20703952896 (19.28 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 12358660096 (11.51 GB)
DFS Remaining: 7267594240 (6.77 GB)
DFS Used%: 0.00%
DFS Remaining%: 35.10%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Fri Nov 18 17:43:25 WET 2022
Last Block Report: Fri Nov 18 17:35:55 WET 2022
Num of Blocks: 0


hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop/etc/hadoop$
```

On exécute les commandes suivantes :

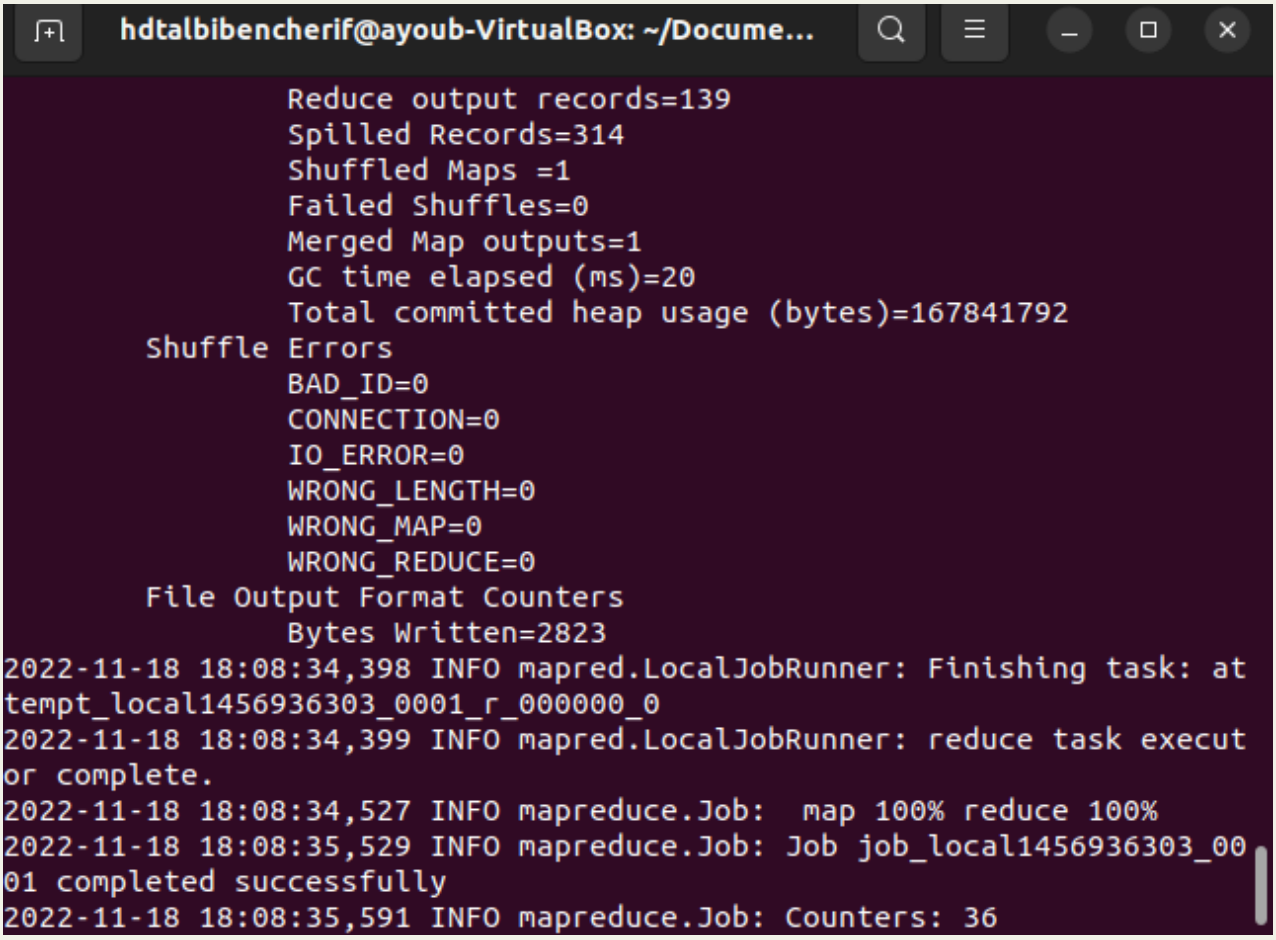
```
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ ls
code  code.zip  hadoop-3.3.4.tar.gz  jdk-8u71-linux-x64.tar.gz
hdtalbibencherif@ayoub-VirtualBox:~/Documents$ sudo unzip code.zip -d
code
Archive:  code.zip
   creating: code/code/
  inflating: code/code/wcount.jar
  inflating: code/code/classpath
  inflating: code/code/WCount.java
  inflating: code/code/WCountMap.java
  inflating: code/code/WCountReduce.java
   creating: code/code/org/
   creating: code/code/org/hadoop/
   creating: code/code/org/hadoop/wordcount/
  inflating: code/code/org/hadoop/wordcount/WCount.class
  inflating: code/code/org/hadoop/wordcount/WCountMap.class
  inflating: code/code/org/hadoop/wordcount/WCountReduce.class
hdtalbibencherif@ayoub-VirtualBox:~/Documents$
```

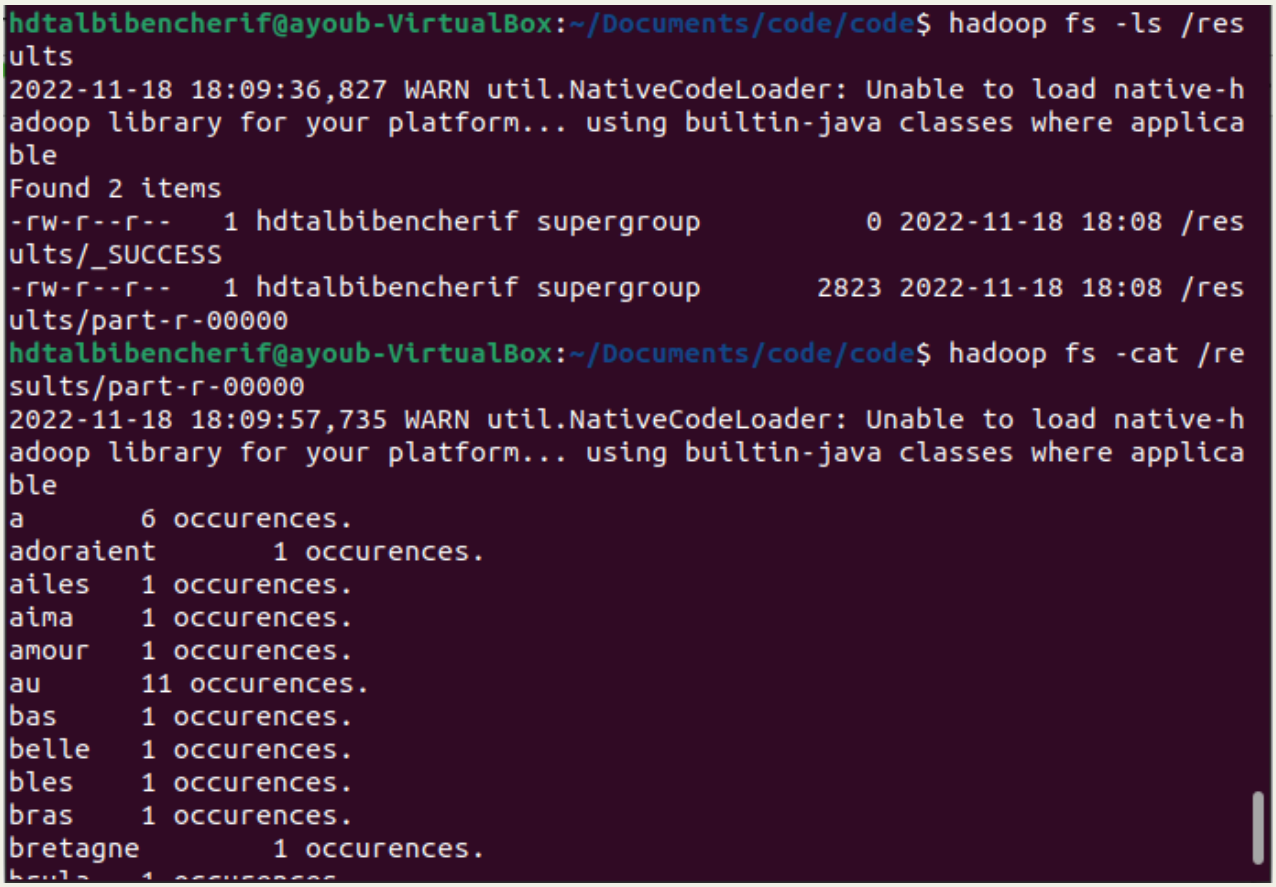
```
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ . classpath
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ mkdir -p org/
hadoop/wordcount/
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ javac WCount*
java
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ mv *.class or
g/hadoop/wordcount/
```

```
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ jar -cvf wcou
nt.jar . /home/hdtalbibencherif/Documents/org
/home/hdtalbibencherif/Documents/org : no such file or directory
added manifest
adding: WCount.class(in = 1646) (out= 859)(deflated 47%)
adding: classpath(in = 305) (out= 114)(deflated 62%)
adding: WCountReduce.class(in = 1834) (out= 775)(deflated 57%)
adding: org/(in = 0) (out= 0)(stored 0%)
adding: org/hadoop/(in = 0) (out= 0)(stored 0%)
adding: org/hadoop/wordcount/(in = 0) (out= 0)(stored 0%)
adding: org/hadoop/wordcount/WCount.class(in = 1646) (out= 859)(deflat
ed 47%)
adding: org/hadoop/wordcount/WCountReduce.class(in = 1834) (out= 775)(
deflated 57%)
adding: org/hadoop/wordcount/WCountMap.class(in = 1674) (out= 722)(def
lated 56%)
adding: WCount.java(in = 2005) (out= 890)(deflated 55%)
adding: WCountMap.java(in = 1057) (out= 565)(deflated 46%)
adding: WCountMap.class(in = 1674) (out= 722)(deflated 56%)
adding: WCountReduce.java(in = 1142) (out= 589)(deflated 48%)
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$
```

```
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop$ bin/hdfs dfs -put
 /home/hdtalbibencherif/Documents/poeme.txt /
2022-11-18 18:05:42,609 WARN util.NativeCodeLoader: Unable to load nat
ive-hadoop library for your platform... using builtin-java classes whe
re applicable
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop$ bin/hdfs dfs -ls
/
2022-11-18 18:06:38,720 WARN util.NativeCodeLoader: Unable to load nat
ive-hadoop library for your platform... using builtin-java classes whe
re applicable
Found 1 items
-rw-r--r--   1 hdtalbibencherif supergroup       1668 2022-11-18 18:05
 /poeme.txt
hdtalbibencherif@ayoub-VirtualBox:/usr/local/hadoop$
```

$ hadoop jar wcount.jar org.hadoop.wordcount.WCount /poeme.txt /results





On utilise les scripts stop-dfs.sh et stop-yarn.sh pour arrêter tous les daemons en cours d'exécution sur votre machine virtuelle

```
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ stop-dfs.sh
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [ayoub-VirtualBox]
2022-11-18 18:11:49,275 WARN util.NativeCodeLoader: Unable to load native-h
adoop library for your platform... using builtin-java classes where applica
ble
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$ stop-yarn.sh
Stopping nodemanagers
Stopping resourcemanager
hdtalbibencherif@ayoub-VirtualBox:~/Documents/code/code$
```

3) Configuration d'un cluster multi-noeuds d'Apache Hadoop :

Nous allons travaillé avec la machine virtuelle précédemment configuré en node unique dans la section précédente. Dans cette VM : - On modifie le fichier **/etc/hostname** de la machine (sudo nano /etc/hostname) , et on supprime son contenu et ajoutez le nouveau hostname de la machine master :

```
  GNU nano 6.2                    /etc/hostname
hadoopmaster
```

On modifie le fichier **/etc/hosts** comme suit :

```
  GNU nano 6.2                        /etc/hosts
127.0.0.1       localhost
127.0.1.1       ayoub-VirtualBox
192.168.0.1     hadoopmaster
192.168.0.2     slave1
192.168.0.3     slave2
# The following lines are desirable for IPv6 capable hosts
::1     ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
```

Afin de configurer de manière permanente l'adresse IP fixe de la machine hadoopmaster, il faut utiliser Netplan qui est un utilitaire de ligne de commande utilisé pour configurer la mise en réseau sur les distributions Ubuntu modernes.

On se déplace dans le fichier **/etc/netplan/**
On modifie le fichier YAML avec la commande :
sudo nano 01-network-manager-all.yaml

```
  GNU nano 6.2              01-network-manager-all.yaml
# Let NetworkManager manage all devices on this system
network:
  version: 2
  ethernets:
   enp0s3:
    dhcp4: false
    addresses: [192.168.0.1/24]
    routes:
      - to: default
        via: 192.168.0.254
```

Enfin, pour prendre en considération ces modifications, On redémarre la machine pour prendre en compte les configurations
Suppression des fichiers du répertoire de stockage de données créer par l'installation single node de Hadoop :

```
hdtalbibencherif@hadoopmaster:~$ cd /usr/local/hadoop_store/
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ rm -rf *
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ mkdir -p /usr
/local/hadoop_store/hdfs/namenode
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ chown -R hdta
lbibencherif /usr/local/hadoop_store/hdfs/namenode
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ cd /usr/local
/hadoop/etc/hadoop/
hdtalbibencherif@hadoopmaster:/usr/local/hadoop/etc/hadoop$
```

Modification des fichiers de configuration de hadoop :

```
  GNU nano 6.2                  core-site.xml *
  Unless required by applicable law or agreed to in writing, softwa>
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or i>
  See the License for the specific language governing permissions a>
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuratio>
<property>
<name>hadoop.tmp.dir</name>
<value>/app/hadoop/tmp</value>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://hadoopmaster:54310</value>
</property>
</configuration>


^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute
```

```
 GNU nano 6.2                  hdfs-site.xml
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>2</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>file:/usr/local/hadoop_store/hdfs/namenode</value>
</property>
</configuration>



                [ Wrote 28 lines ]
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute
```

```
 GNU nano 6.2                  mapred-site.xml
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, softwa>
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or i>
  See the License for the specific language governing permissions a>
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>mapred.job.tracker</name>
<value>hadoopmaster:54311</value>
</property>
</configuration>


^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute
```

```
 GNU nano 6.2                  yarn-site.xml
<value>hadoopmaster:8025</value>
</property>
<property>
<name>yarn.resourcemanager.scheduler.address</name>
<value>hadoopmaster:8030</value>
</property>
<property>
<name>yarn.resourcemanager.address</name>
<value>hadoopmaster:8050</value>
</property>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>

                [ Wrote 38 lines ]
```

Création du fichier masters :
On crée le fichier masters qui contient le hostname de la machine master :
Modification du fichier workers :
Modifiez le fichier workers qui contient le hostname de chaque machine slave dans le répertoire :

```
 GNU nano 6.2                    masters
hadoopmaster
```

```
 GNU nano 6.2                    workers
slave1
slave2
```

On clone deux machines slave1 et slave2 à partir de la machine hadoopmaster.
On modifie les hostnames des deux machines.
Configuration des adresses IP fixes des deux machines :

```
 GNU nano 6.2              01-network-manager-all.yaml
# Let NetworkManager manage all devices on this system
network:
  version: 2
  ethernets:
   enp0s3:
    dhcp4: false
    addresses: [192.168.0.2/24]
    routes:
      - to: default
        via: 192.168.0.254
```

```
 GNU nano 6.2              01-network-manager-all.yaml
# Let NetworkManager manage all devices on this system
network:
  version: 2
  ethernets:
   enp0s3:
    dhcp4: false
    addresses: [192.168.0.3/24]
    routes:
      - to: default
        via: 192.168.0.254
```

supprimer les fichiers du répertoire de stockage de données créés par l'installation single node de Hadoop pour les deux machines :

```
hdtalbibencherif@slave2:~$ cd /usr/local/hadoop_store/
hdtalbibencherif@slave2:/usr/local/hadoop_store$ rm -rf *
hdtalbibencherif@slave2:/usr/local/hadoop_store$ mkdir -p /usr/local/hadoop_sto
re/hdfs/datanode
hdtalbibencherif@slave2:/usr/local/hadoop_store$ chown -R hdtalbibencherif /usr
/local/hadoop_store/hdfs/datanode
hdtalbibencherif@slave2:/usr/local/hadoop_store$
```

On copie la clé ssh pour configurer un accès ssh sans mot de passe entre les machines du cluster

```
hdtalbibencherif@hadoopmaster:~$ ssh-copy-id -i /home/hdtalbibencherif/.ssh/id_
rsa.pub hdtalbibencherif@hadoopmaster
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hdtalbiben
cherif/.ssh/id_rsa.pub"
The authenticity of host 'hadoopmaster (192.168.0.1)' can't be established.
ED25519 key fingerprint is SHA256:oQld88scwQ6JgUdr37D73+eYRqx+UVWByd9osfySn7Y.
This host key is known by the following other names/addresses:
    ~/.ssh/known_hosts:1: [hashed name]
    ~/.ssh/known_hosts:4: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter
 out any that are already installed

/usr/bin/ssh-copy-id: WARNING: All keys were skipped because they already exist
 on the remote system.
                (if you think this is a mistake, you may want to use -f option)

hdtalbibencherif@hadoopmaster:~$
```

```
hdtalbibencherif@hadoopmaster:~$ ssh-copy-id -i /home/hdtalbibenche
rif/.ssh/id_rsa.pub hdtalbibencherif@slave1
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/hom
e/hdtalbibencherif/.ssh/id_rsa.pub"
The authenticity of host 'slave1 (192.168.0.2)' can't be establishe
d.
ED25519 key fingerprint is SHA256:oQld88scwQ6JgUdr37D73+eYRqx+UVWBy
d9osfySn7Y.
This host key is known by the following other names/addresses:
    ~/.ssh/known_hosts:1: [hashed name]
    ~/.ssh/known_hosts:4: [hashed name]
    ~/.ssh/known_hosts:5: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])
? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s
), to filter out any that are already installed

/usr/bin/ssh-copy-id: WARNING: All keys were skipped because they a
lready exist on the remote system.
                (if you think this is a mistake, you may want to us
e -f option)

hdtalbibencherif@hadoopmaster:~$
```

```
hdtalbibencherif@hadoopmaster:~$ ssh-copy-id -i /home/hdtalbibenche
rif/.ssh/id_rsa.pub hdtalbibencherif@slave2
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/hom
e/hdtalbibencherif/.ssh/id_rsa.pub"
The authenticity of host 'slave2 (192.168.0.3)' can't be establishe
d.
ED25519 key fingerprint is SHA256:oQld88scwQ6JgUdr37D73+eYRqx+UVWBy
d9osfySn7Y.
This host key is known by the following other names/addresses:
    ~/.ssh/known_hosts:1: [hashed name]
    ~/.ssh/known_hosts:4: [hashed name]
    ~/.ssh/known_hosts:5: [hashed name]
    ~/.ssh/known_hosts:6: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])
? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s
), to filter out any that are already installed

/usr/bin/ssh-copy-id: WARNING: All keys were skipped because they a
lready exist on the remote system.
                (if you think this is a mistake, you may want to us
e -f option)

hdtalbibencherif@hadoopmaster:~$
```

On teste la connexion ssh entre les différents machines du cluster : Dans la machine hadoopmaster

```
hdtalbibencherif@hadoopmaster:~$ ssh slave1
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.15.0-43-generic x86_64)

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

0 updates can be applied immediately.

Last login: Fri Nov 18 17:32:42 2022 from 127.0.0.1
hdtalbibencherif@slave1:~$
```

```
hdtalbibencherif@slave1:~$ ssh hadoopmaster
The authenticity of host 'hadoopmaster (192.168.0.1)' can't be establi
shed.
ED25519 key fingerprint is SHA256:oQld88scwQ6JgUdr37D73+eYRqx+UVWByd9o
sfySn7Y.
This host key is known by the following other names/addresses:
    ~/.ssh/known_hosts:1: [hashed name]
    ~/.ssh/known_hosts:4: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])? y
es
Warning: Permanently added 'hadoopmaster' (ED25519) to the list of kno
wn hosts.
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.15.0-43-generic x86_64)

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

0 updates can be applied immediately.

Last login: Fri Nov 18 17:32:42 2022 from 127.0.0.1
hdtalbibencherif@hadoopmaster:~$
```

On modifie le fichier hdfs-site.xml de la machine virtuelle slave1 et slave2 comme suit :

```
  GNU nano 6.2                    hdfs-site.xml
  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or imp>
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>2</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>file:/usr/local/hadoop_store/hdfs/datanode</value>
</property>
</configuration>

                        [ Wrote 28 lines ]
^G Help       ^O Write Out  ^W Where Is   ^K Cut        ^T Execute
^X Exit       ^R Read File  ^\ Replace    ^U Paste      ^J Justify
```

Formatage du Namenode

```
84 entries
2022-11-18 20:38:14,675 INFO namenode.FSImage: Allocated new Block
PoolId: BP-815823492-192.168.0.1-1668803894637
2022-11-18 20:38:14,750 INFO common.Storage: Storage directory /ap
p/hadoop/tmp/dfs/name has been successfully formatted.
2022-11-18 20:38:14,883 INFO namenode.FSImageFormatProtobuf: Savin
g image file /app/hadoop/tmp/dfs/name/current/fsimage.ckpt_0000000
000000000000 using no compression
2022-11-18 20:38:15,195 INFO namenode.FSImageFormatProtobuf: Image
 file /app/hadoop/tmp/dfs/name/current/fsimage.ckpt_00000000000000
00000 of size 411 bytes saved in 0 seconds .
2022-11-18 20:38:15,279 INFO namenode.NNStorageRetentionManager: G
oing to retain 1 images with txid >= 0
2022-11-18 20:38:15,376 INFO namenode.FSNamesystem: Stopping servi
ces started for active state
2022-11-18 20:38:15,376 INFO namenode.FSNamesystem: Stopping servi
ces started for standby state
2022-11-18 20:38:15,383 INFO namenode.FSImage: FSImageSaver clean
checkpoint: txid=0 when meet shutdown.
2022-11-18 20:38:15,384 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at hadoopmaster/192.168.0.1
************************************************************/
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$
```

Lancer les commandes : $ start-dfs.sh et $ start-yarn.sh

```
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ start-dfs.s
h
Starting namenodes on [hadoopmaster]
Starting datanodes
Starting secondary namenodes [hadoopmaster]
2022-11-18 20:39:31,033 WARN util.NativeCodeLoader: Unable to load
 native-hadoop library for your platform... using builtin-java cla
sses where applicable
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ start-yarn.
sh
Starting resourcemanager
Starting nodemanagers
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$
```

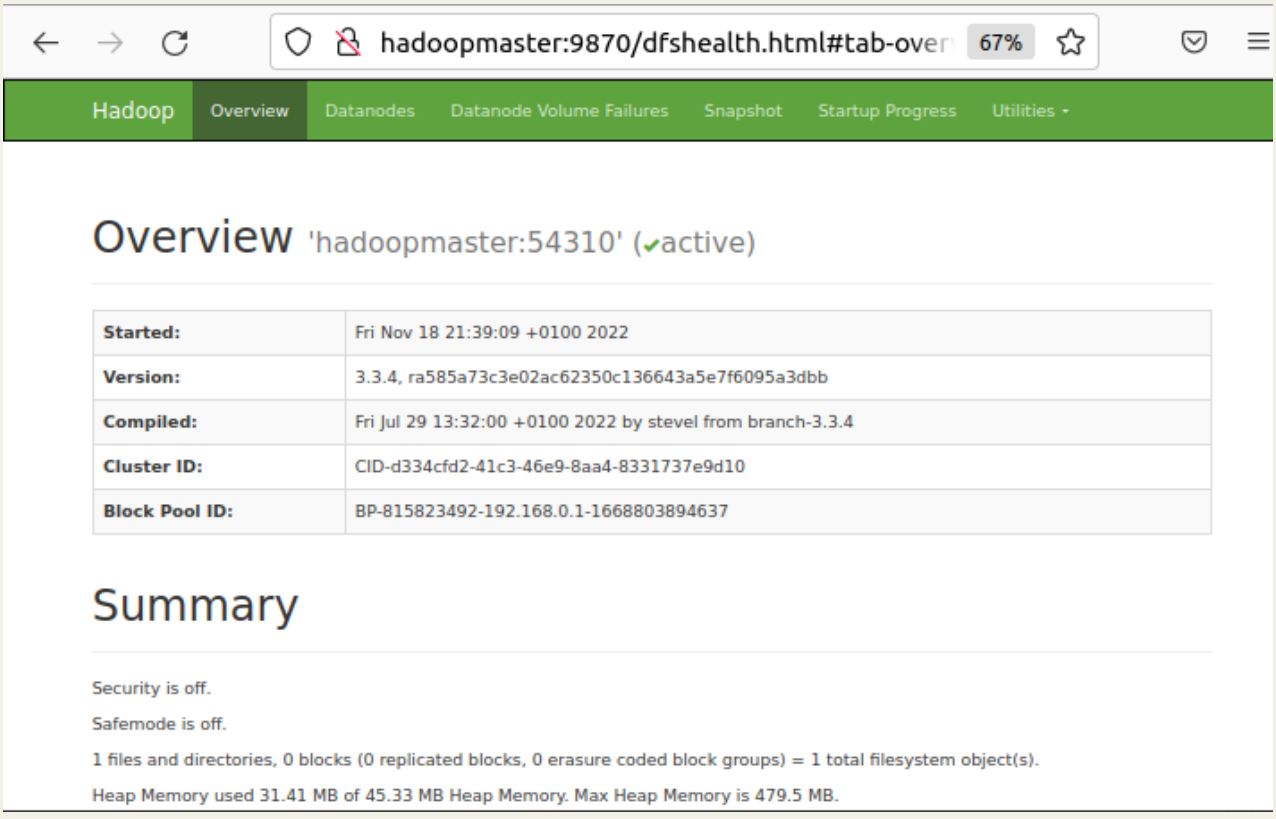
Lancer la commande jps dans chaque machine du cluster hadoop

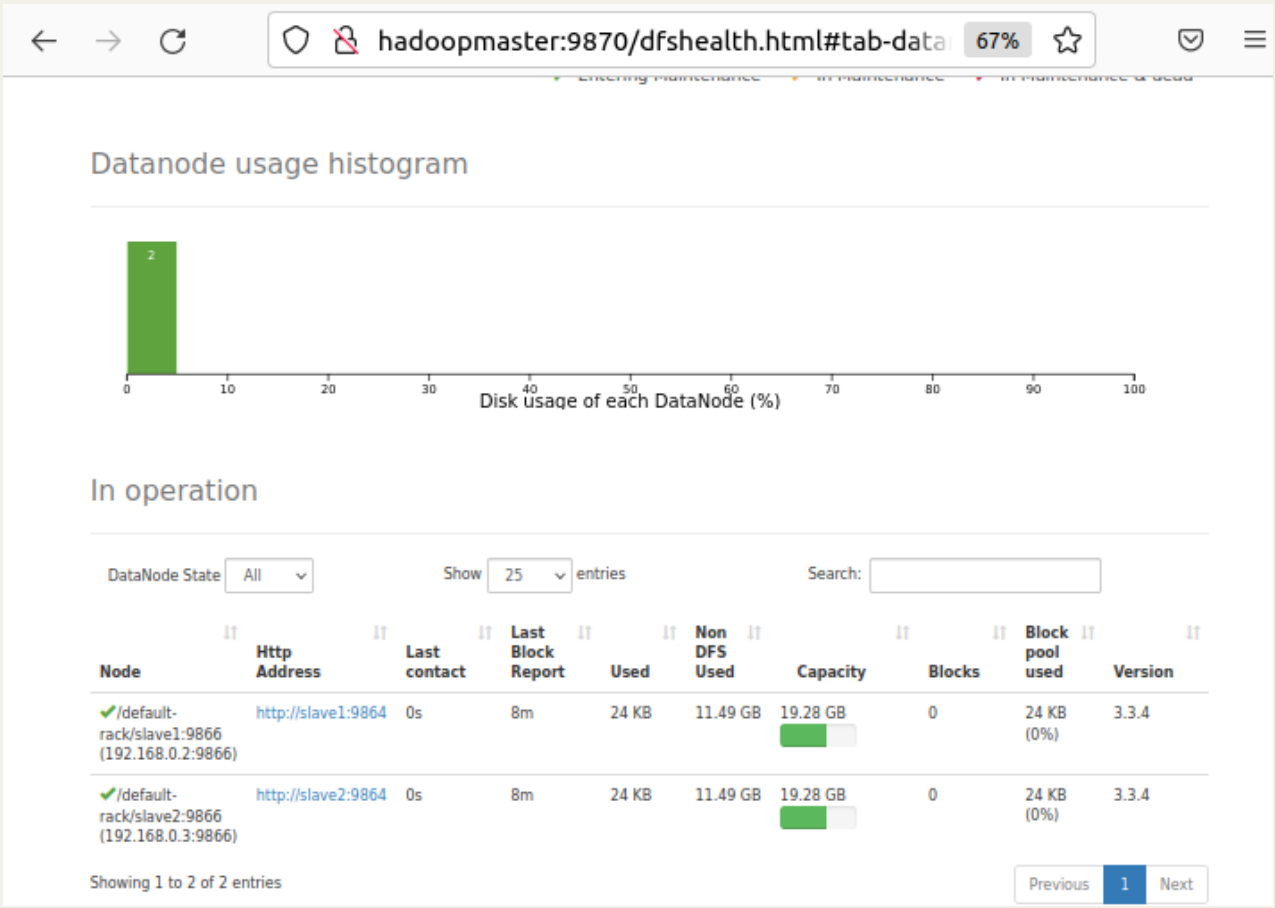
```
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ jps
2708 SecondaryNameNode
2903 ResourceManager
2535 NameNode
3193 Jps
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$
```

```
hdtalbibencherif@slave1:~$ jps
2546 Jps
2359 NodeManager
2234 DataNode
hdtalbibencherif@slave1:~$
```

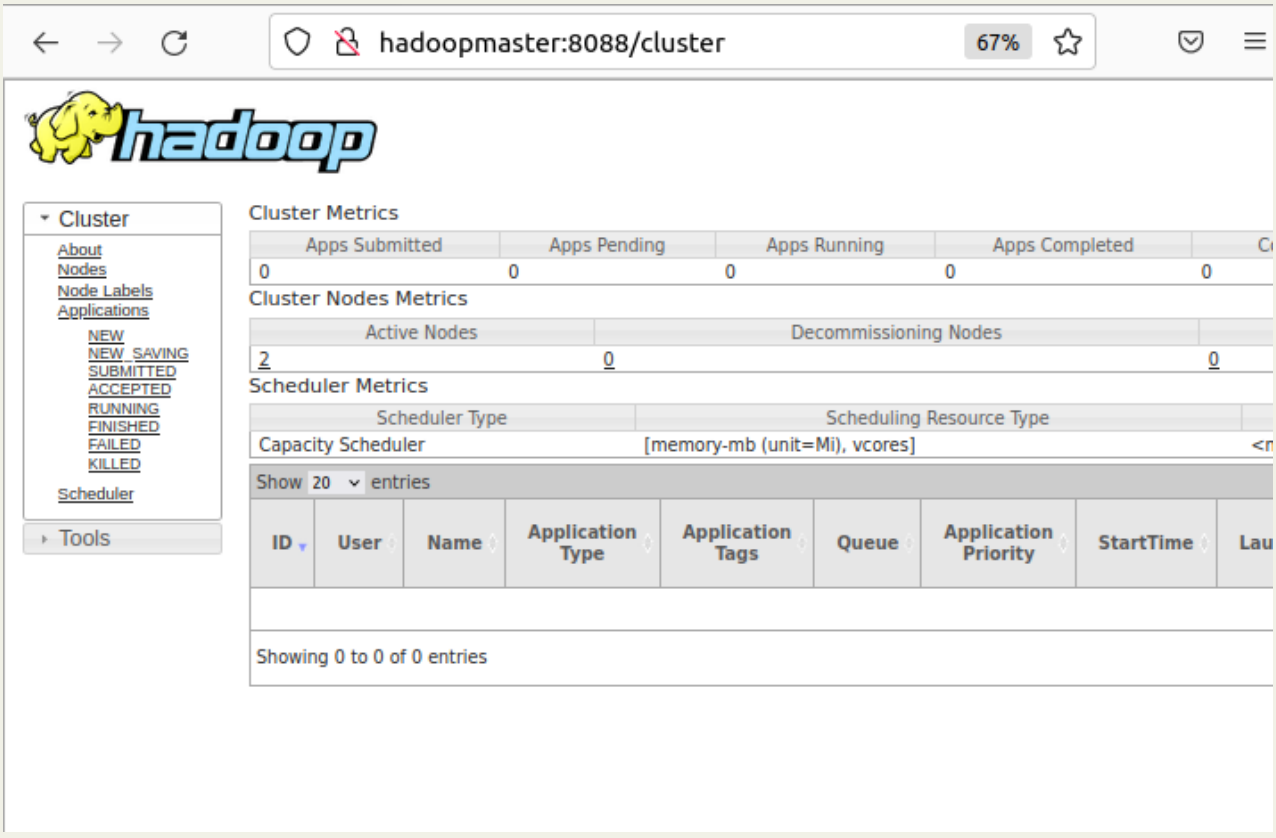
```
hdtalbibencherif@slave2:~$ jps
2353 NodeManager
2228 DataNode
2538 Jps
hdtalbibencherif@slave2:~$
```

Accéder aux services de Hadoop via le navigateur : Hadoop NameNode démarre sur le port 9870 par défaut. Accédez à **http://hadoopmaster : 9870 /** dans votre navigateur web :

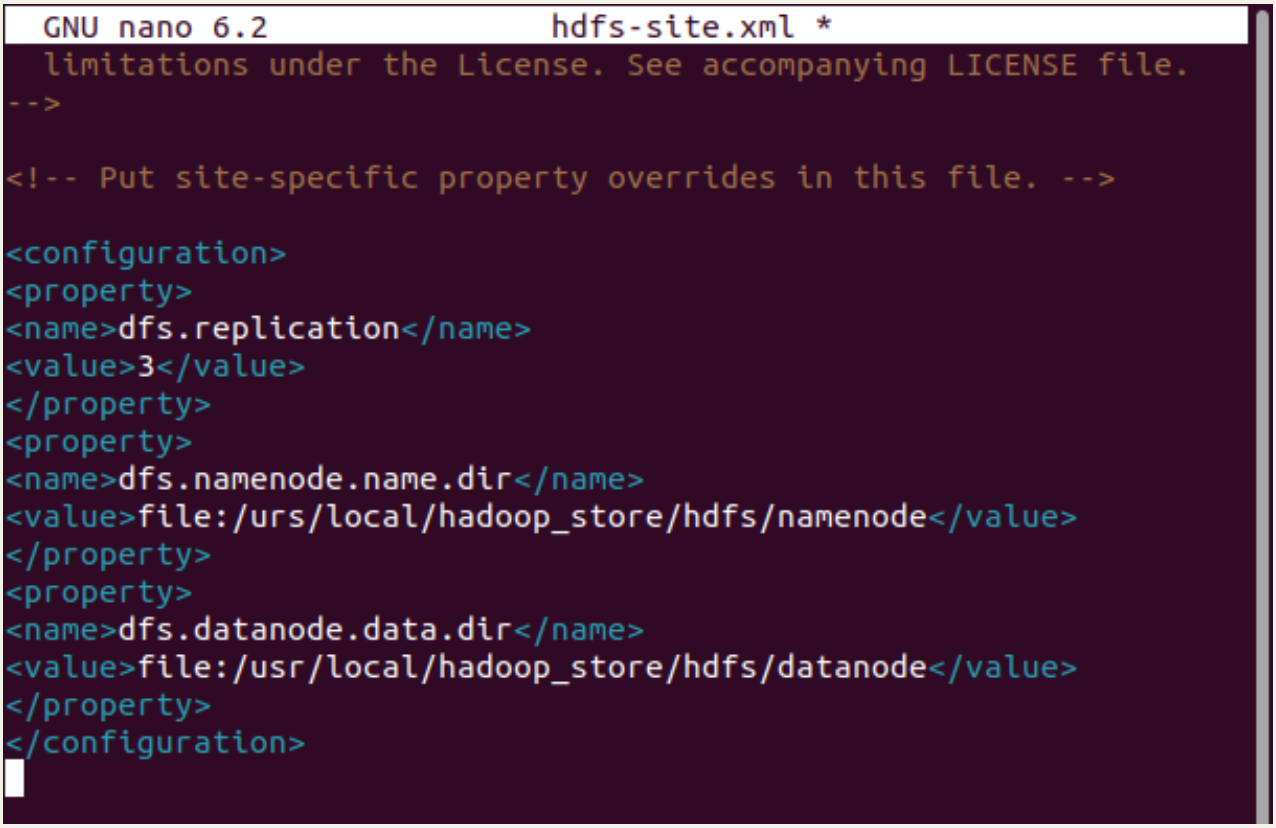




Pour accéder à l'interface du Ressource manager, tapez :
http://hadoopmaster : 8088 /



Vous pouvez ajouter le nœud hadoopmaster comme datanode aussi, pour cela les fichiers suivants doivent être modifiés comme suit :
Modifier le fichier hdfs-site.xml de hadoopmaster comme suit:



On modifie le fichier hdfs-site.xml de slave1 et slave2 aussi en mettant la valeur 3 pour replication.

Modification du fichier slaves dans hadoopmaster, slave1 et slave2

Ensuite videz le répertoire de stockage de hadoop «**/usr/local/hadoop_store/**» et créer les nouveaux répertoires Enfin, reformatez hadoopmaster et lancer les scripts : start-dfs.sh et yarn-dfs.sh.

```
hdtalbibencherif@hadoopmaster:~$ cd /usr/local/hadoop_store/
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ rm -rf *
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ mkdir -p /usr/local/hado
op_store/hdfs/namenode
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ mkdir -p /usr/local/hado
op_store/hdfs/datanode
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ chown -R hdtalbibencheri
f /usr/local/hadoop_store
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ hdfs namenode -format
2022-11-19 09:28:11,494 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = hadoopmaster/192.168.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.4
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share
/hadoop/common/lib/jetty-http-9.4.43.v20210629.jar:/usr/local/hadoop/share/hado
op/common/lib/commons-io-2.8.0.jar:/usr/local/hadoop/share/hadoop/common/lib/je
tty-io-9.4.43.v20210629.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-l
ogging-1.1.3.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-jaxrs-1.9.13
.jar:/usr/local/hadoop/share/hadoop/common/lib/jetty-security-9.4.43.v20210629.
jar:/usr/local/hadoop/share/hadoop/common/lib/jetty-xml-9.4.43.v20210629.jar:/u
sr/local/hadoop/share/hadoop/common/lib/kerb-common-1.0.1.jar:/usr/local/hadoop
/share/hadoop/common/lib/metrics-core-3.2.4.jar:/usr/local/hadoop/share/hadoop/
common/lib/snappy-java-1.1.8.2.jar:/usr/local/hadoop/share/hadoop/common/lib/dn
sjava-2.1.7.jar:/usr/local/hadoop/share/hadoop/common/lib/hadoop-annotations-3.
3.4.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-databind-2.12.7.jar:/
usr/local/hadoop/share/hadoop/common/lib/commons-daemon-1.0.13.jar:/usr/local/h
adoop/share/hadoop/common/lib/jackson-annotations-2.12.7.jar:/usr/local/hadoop/
```

```
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ start-dfs.sh
Starting namenodes on [hadoopmaster]
Starting datanodes
Starting secondary namenodes [hadoopmaster]
2022-11-19 09:30:02,576 WARN util.NativeCodeLoader: Unable to load native-hadoo
p library for your platform... using builtin-java classes where applicable
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hdtalbibencherif@hadoopmaster:/usr/local/hadoop_store$
```

4) Exécution d'un programme Map/Reduce dans un cluster multi-noeuds :

On lance la commande : « hdfs dfsadmin -report » pour vérifier le bon fonctionnement de tous
les nœuds du cluster.

```
Replicated Blocks:
        Under replicated blocks: 0
        Blocks with corrupt replicas: 0
        Missing blocks: 0
        Missing blocks (with replication factor 1): 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0
Erasure Coded Block Groups:
        Low redundancy block groups: 0
        Block groups with corrupt internal blocks: 0
        Missing block groups: 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0

-------------------------------------------------
Live datanodes (2):

Name: 192.168.0.2:9866 (slave1)
```

On suit les mêmes étapes décrites dans la section 1 du TP : Exécution d'un programme **Map/Reduce** dans un cluster à noeud unique

```
2022-10-21 12:56:05,850 INFO mapreduce.Job:  map 100% reduce 100%
2022-10-21 12:56:05,851 INFO mapreduce.Job: Job job_local1050561859_0001 comple
ted successfully
2022-10-21 12:56:05,906 INFO mapreduce.Job: Counters: 36
        File System Counters
                FILE: Number of bytes read=19688
                FILE: Number of bytes written=1302470
```

```
2022-10-21 13:00:12,178 WARN util.NativeCodeLoader: Unable
 to load native-hadoop library for your platform... using
builtin-java classes where applicable
a       6 occurences.
adoraient       1 occurences.
ailes   1 occurences.
aima    1 occurences.
amour   1 occurences.
au      11 occurences.
bas     1 occurences.
belle   1 occurences.
bles    1 occurences.
bras    1 occurences.
bretagne        1 occurences.
brula   1 occurences.
celle   1 occurences.
celui   20 occurences.
cette   1 occurences.
chancelle       1 occurences.
chapelle        1 occurences.
ciel    10 occurences.
```

FIN