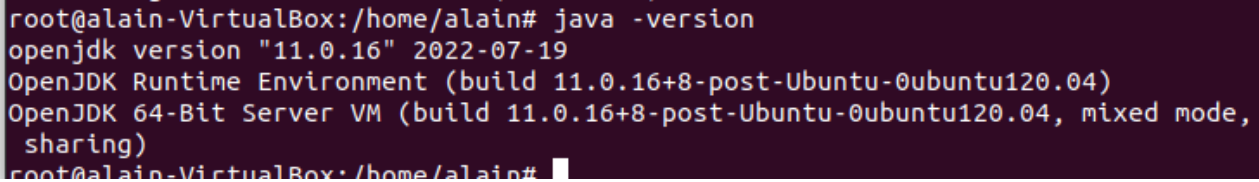# Mise en place d'une plateforme Hadoop

1- Installation de JVM

root@alain-VirtualBox:/home/alain# **apt install default-jdk default-jre -y**

Verifier la version de java

```
root@alain-VirtualBox:/home/alain# java -version
openjdk version "11.0.16" 2022-07-19
OpenJDK Runtime Environment (build 11.0.16+8-post-Ubuntu-0ubuntu120.04)
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Ubuntu-0ubuntu120.04, mixed mode,
 sharing)
root@alain-VirtualBox:/home/alain#
```

fig

2- creation d'un compte devant se connexter en utilisant de cle ssh

root@alain-VirtualBox:/home/alain# adduser hadoop
Ajout de l'utilisateur « hadoop »

ajout de l'utilisateur hadoop dans les groupes hadoop et sudo

root@alain-VirtualBox:/home/alain# usermod -aG sudo hadoop

3- installation de serveur et client ssh

root@alain-VirtualBox:/home/alain# su – hadoop

hadoop@alain-VirtualBox:~$ sudo apt install openssh-server openssh-client -y

4- generation de clé ssh

hadoop@alain-VirtualBox:~$ ssh-keygen -t rsa

```
hadoop@alain-VirtualBox:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa): passer
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in passer
Your public key has been saved in passer.pub
The key fingerprint is:
SHA256:O6ob5aRsHn/6AjfOejnivDSzFlCvJQ9aDKeX5MUAHdM hadoop@alain-VirtualBox
The key's randomart image is:
+---[RSA 3072]----+
|   .o+=          |
|   . =.E         |
|    O +          |
|   o O o         |
|    = *o S       |
|   ..+=+  .      |
|    X*.oo        |
|   =oBB...        |
|   oXB+*.        |
+----[SHA256]-----+
hadoop@alain-VirtualBox:~$
```

fig 2

hadoop@alain-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

hadoop@alain-VirtualBox:~$ chmod 640 ~/.ssh/authorized_keys

hadoop@alain-VirtualBox:~$ ssh localhost

4- installation de hadoop

hadoop@alain-VirtualBox:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.1/hadoop-3.3.1.tar.gz

hadoop@alain-VirtualBox:~$ tar -xvzf hadoop-3.3.1.tar.gz

adoop@alain-VirtualBox:~$ sudo mv hadoop-3.3.1 /usr/local/hadoop

hadoop@alain-VirtualBox:~$ sudo mkdir /usr/local/hadoop/logs

hadoop@alain-VirtualBox:~$ sudo chown -R hadoop:hadoop /usr/local/hadoop

5- configuration de hadoop

hadoop@alain-VirtualBox:~$ sudo nano ~/.bashrc

```
export HADOOP_HOME=/usr/local/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"
```

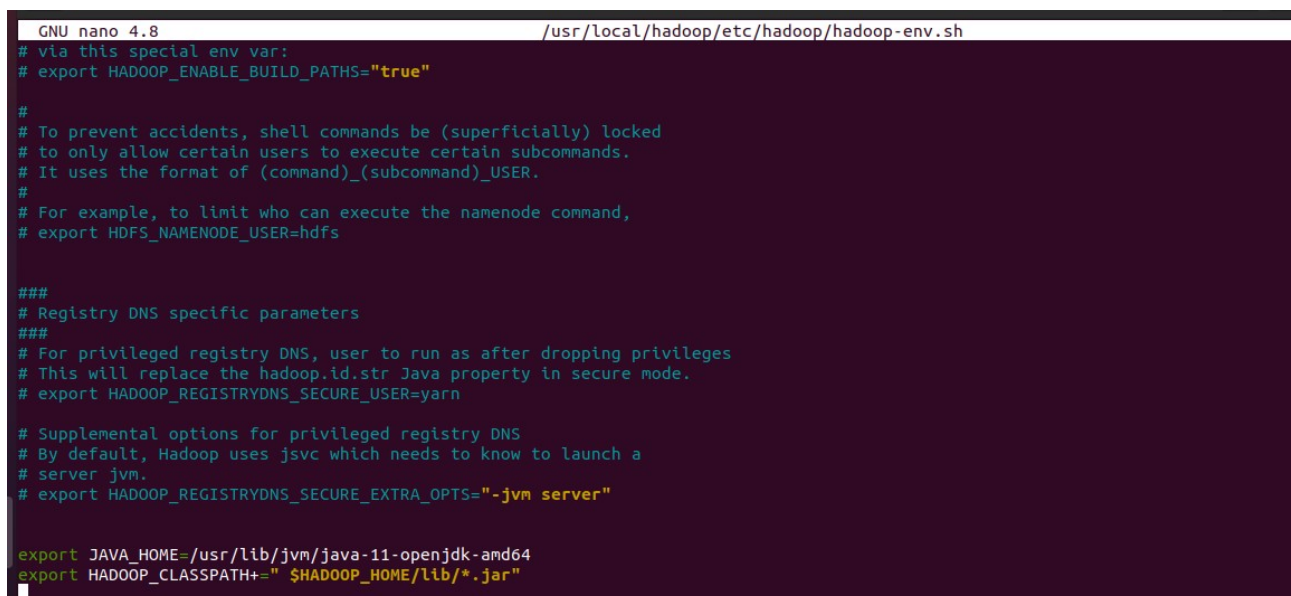**hadoop@alain-VirtualBox:~$ source ~/.bashrc**

6- configuration de l'environnement java

hadoop@alain-VirtualBox:~$ which javac
/usr/bin/javac

hadoop@alain-VirtualBox:~$ readlink -f /usr/bin/javac
/usr/lib/jvm/java-11-openjdk-amd64/bin/javac

hadoop@alain-VirtualBox:~$ sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_CLASSPATH+=" $HADOOP_HOME/lib/*.jar"
```


Ajout de variables environnement

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ sudo wget
https://jcenter.bintray.com/javax/activation/javax.activation-api/1.2.0/javax.activation-api-1.2.0.jar

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ hadoop version
Hadoop 3.3.1

Configuration de l'URL du nameqpace

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ sudo nano
$HADOOP_HOME/etc/hadoop/core-site.xml

```
<configuration>
    <property>
        <name>fs.default.name</name>
        <value>hdfs://0.0.0.0:9000</value>
        <description>The default file system URI</description>
    </property>
</configuration>
```
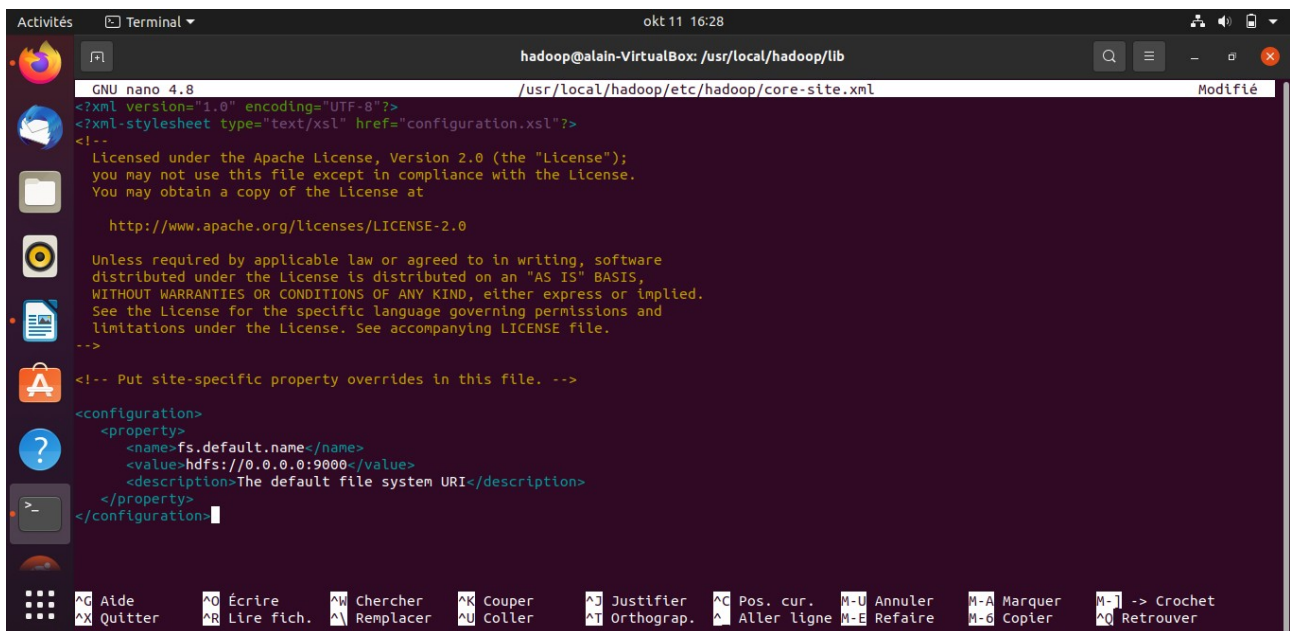

fig

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ sudo mkdir -p
/home/hadoop/hdfs/{namenode,datanode}

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ sudo chown -R hadoop:hadoop
/home/hadoop/hdfs

Configuration de la localisation des metadonnées
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ sudo nano
$HADOOP_HOME/etc/hadoop/hdfs-site.xml

```
<configuration>
    <property>
        <name>dfs.replication</name>
```

```
      <value>1</value>
   </property>

   <property>
      <name>dfs.name.dir</name>
      <value>file:///home/hadoop/hdfs/namenode</value>
   </property>

   <property>
      <name>dfs.data.dir</name>
      <value>file:///home/hadoop/hdfs/datanode</value>
   </property>
</configuration>
```

Configuration de mapreduce

```
 sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

```
<configuration>
   <property>
      <name>mapreduce.framework.name</name>
      <value>yarn</value>
   </property>
</configuration>
```
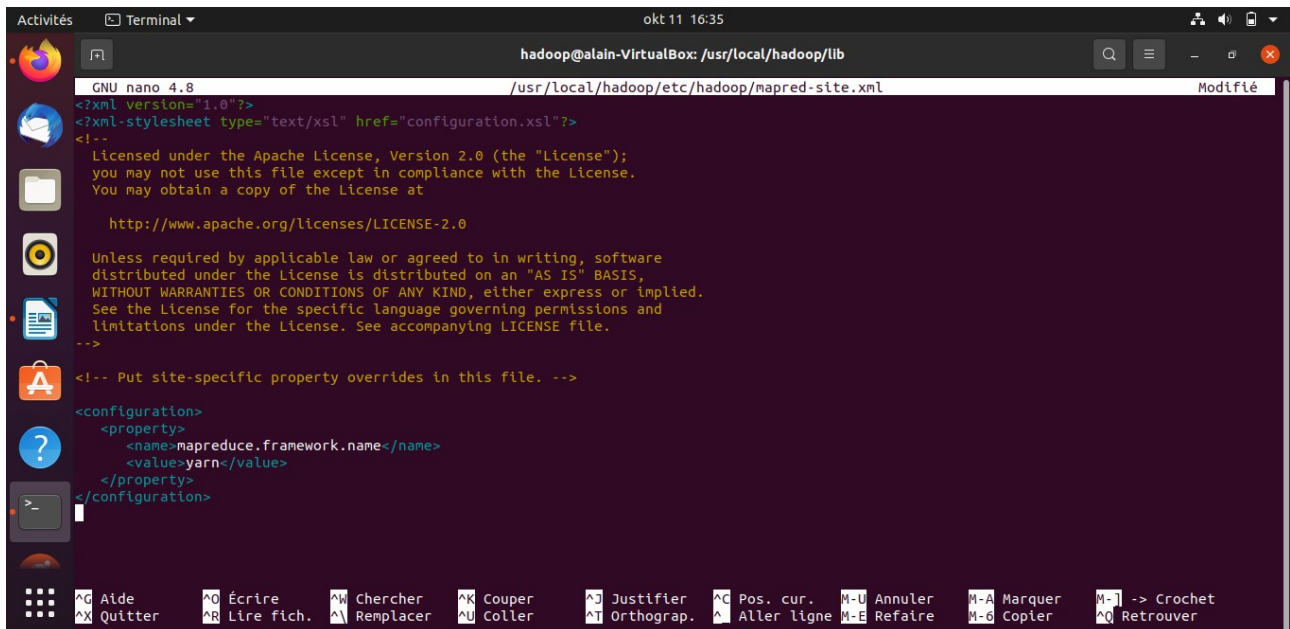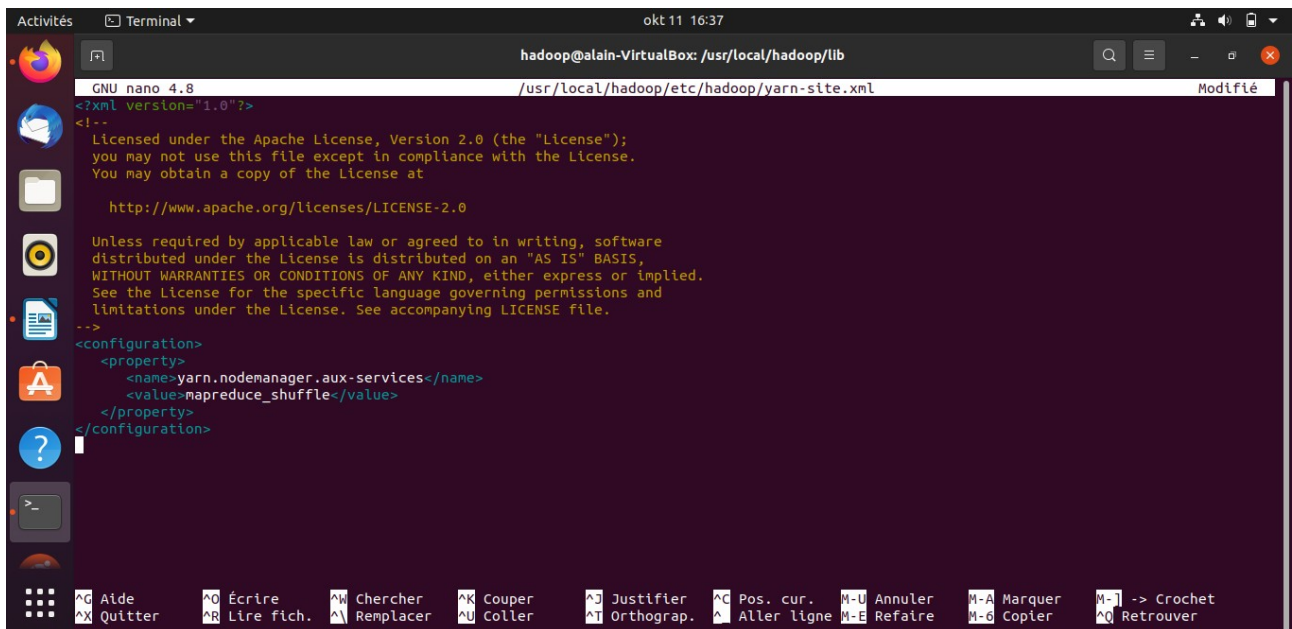

fig

Configuration de yarn
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ sudo nano
$HADOOP_HOME/etc/hadoop/yarn-site.xml

fig

Validate the Hadoop configuration and format the HDFS NameNode.

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ hdfs namenode -format

fig

8- demarrage du cluster hadoop

demarrage de namenode et de datanode

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ start-dfs.sh

fig

Demarrage de yarn ressource et de node manager

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ start-yarn.sh

fig

verification de demarrage des composants

hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ jps

verification demarrage

9- acces à l'interface de hadoop

interface hadoop

fig

NameNode Journal Status

Current transaction ID: 3

Journal Manager	State
FileJournalManager(root=/home/hadoop /hdfs/namenode)	EditLogFileOutputStream(/home/hadoop/hdfs/namenode/current /edits_inprogress_0000000000000000003)

NameNode Storage

Storage Directory	Type	State
/home/hadoop/hdfs/namenode	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service

fig