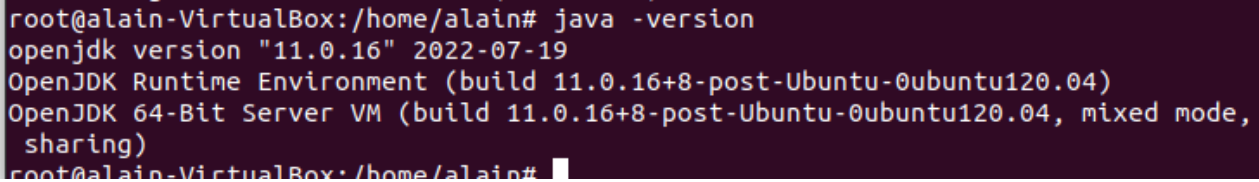


Mise en place d'une plateforme Hadoop

1- Installation de JVM

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

Verifier la version de java

A terminal window with a dark purple background. The text is white. It shows the command 'java -version' being executed. The output displays the OpenJDK version '11.0.16' from 2022-07-19, the runtime environment build '11.0.16+8-post-Ubuntu-0ubuntu120.04', and the 64-bit server VM build '11.0.16+8-post-Ubuntu-0ubuntu120.04' in mixed mode with sharing.

```
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 connecter 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:
SHA256:06ob5aRsHn/6Ajf0ejnivDSzFlCvJQ9aDKeX5MUAHdM 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
```

```
hadoop@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

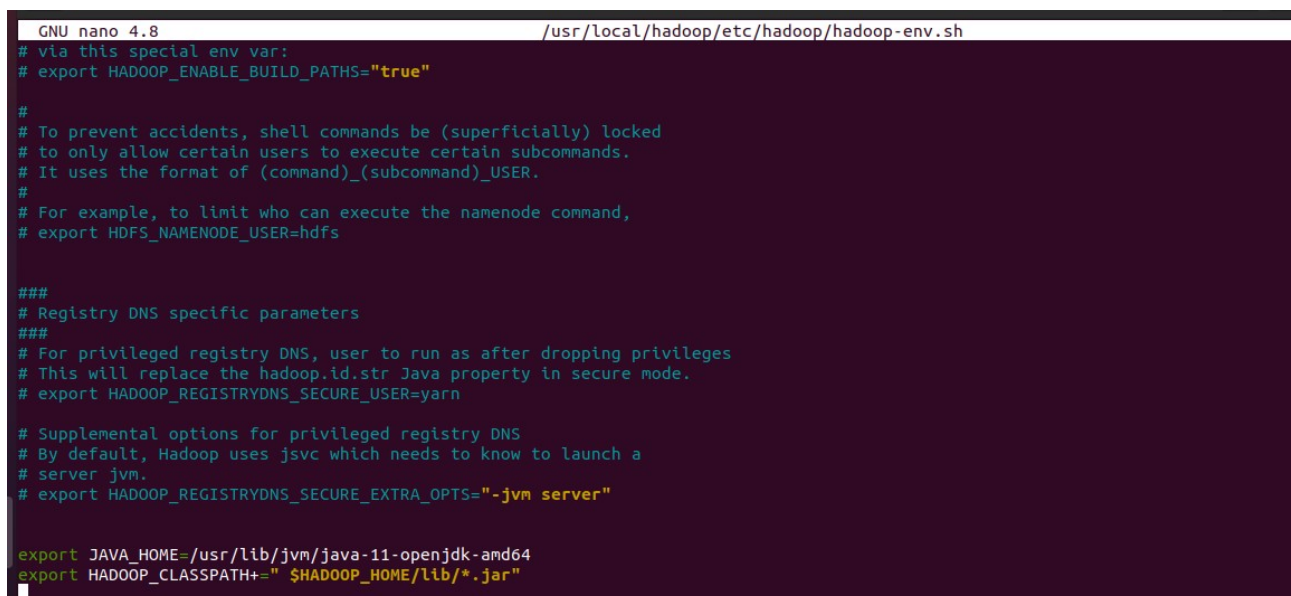
6- configuration de l'environnement java

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"
```



```
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hadoop-env.sh
# via this special env var:
# export HADOOP_ENABLE_BUILD_PATHS="true"

#
# To prevent accidents, shell commands be (superficially) locked
# to only allow certain users to execute certain subcommands.
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# export HDFS_NAMENODE_USER=hdfs

###
# Registry DNS specific parameters
###
# For privileged registry DNS, user to run as after dropping privileges
# This will replace the hadoop.id.str Java property in secure mode.
# export HADOOP_REGISTRYDNS_SECURE_USER=yarn

# Supplemental options for privileged registry DNS
# By default, Hadoop uses jsvc which needs to know to launch a
# server jvm.
# export HADOOP_REGISTRYDNS_SECURE_EXTRA_OPTS="-jvm server"

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

Ajout de variables environnement

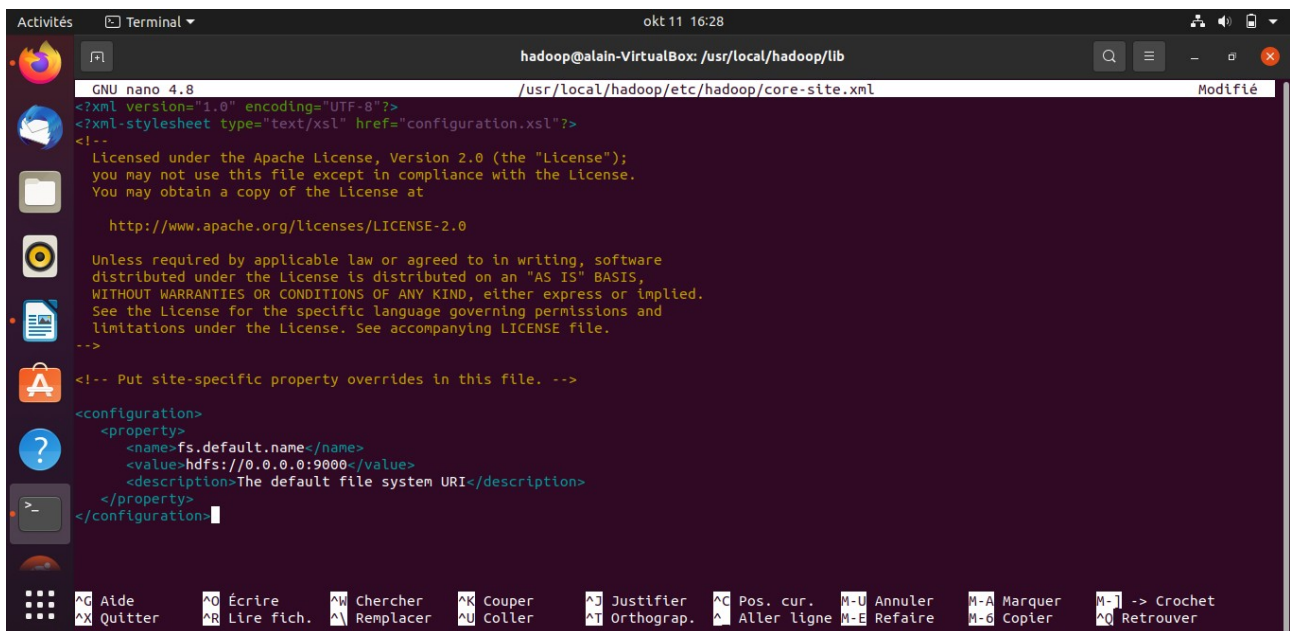
hadoop@alain-VirtualBox:/usr/local/hadoop/lib\$ sudo wget
<https://jcenter.bintray.com/javax/activation/javax.activation-api/1.2.0/javax.activation-api-1.2.0.jar>

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

Configuration de l'URL du namespace

```
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>

```

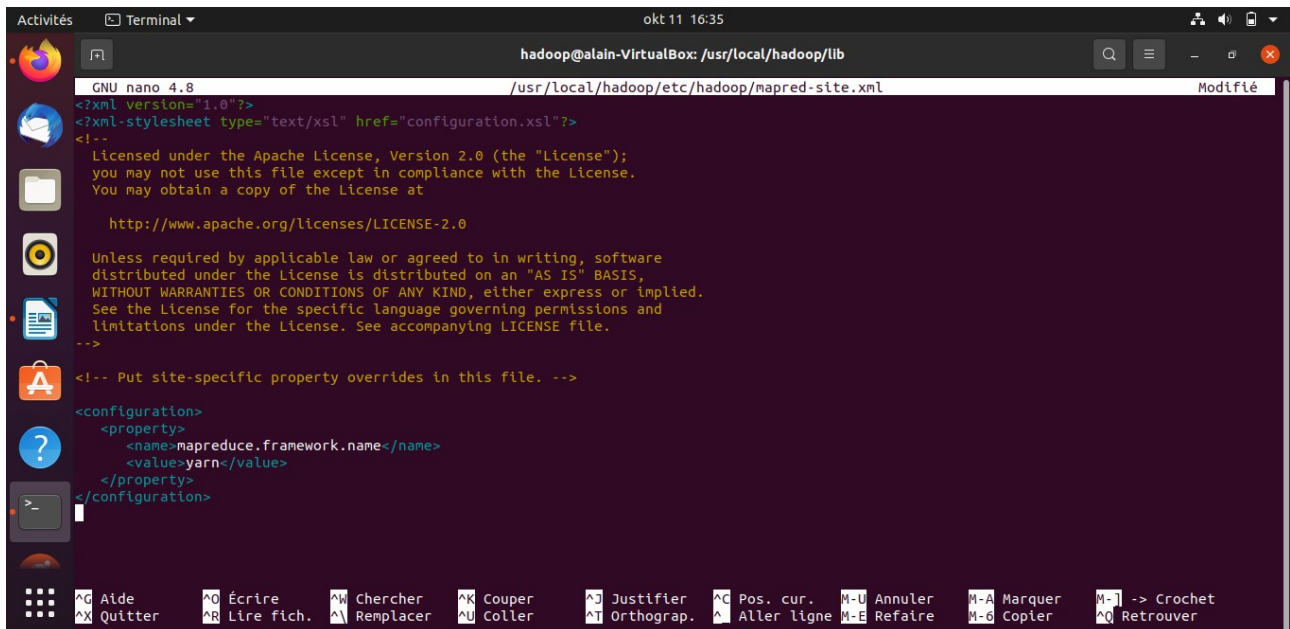
Configuration de mapreduce

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

```

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

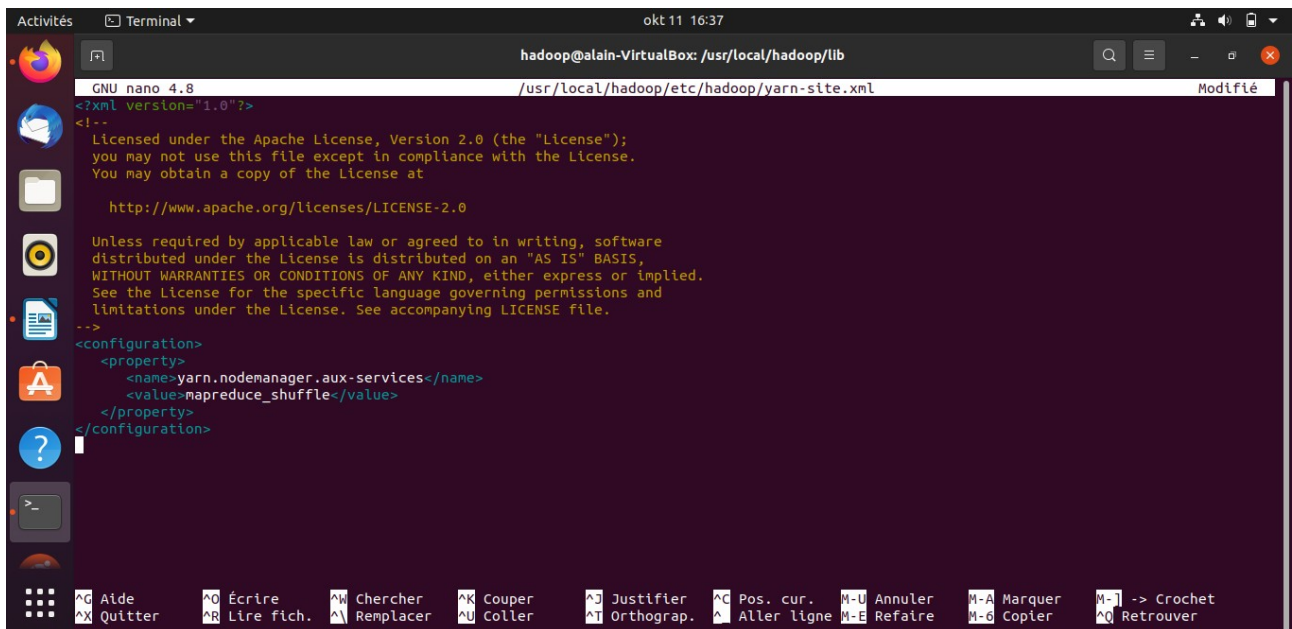
```



fig

Configuration de yarn

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



```
hadoop@alain-VirtualBox: /usr/local/hadoop/lib
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/yarn-site.xml
<?xml version="1.0"?>
<!--
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

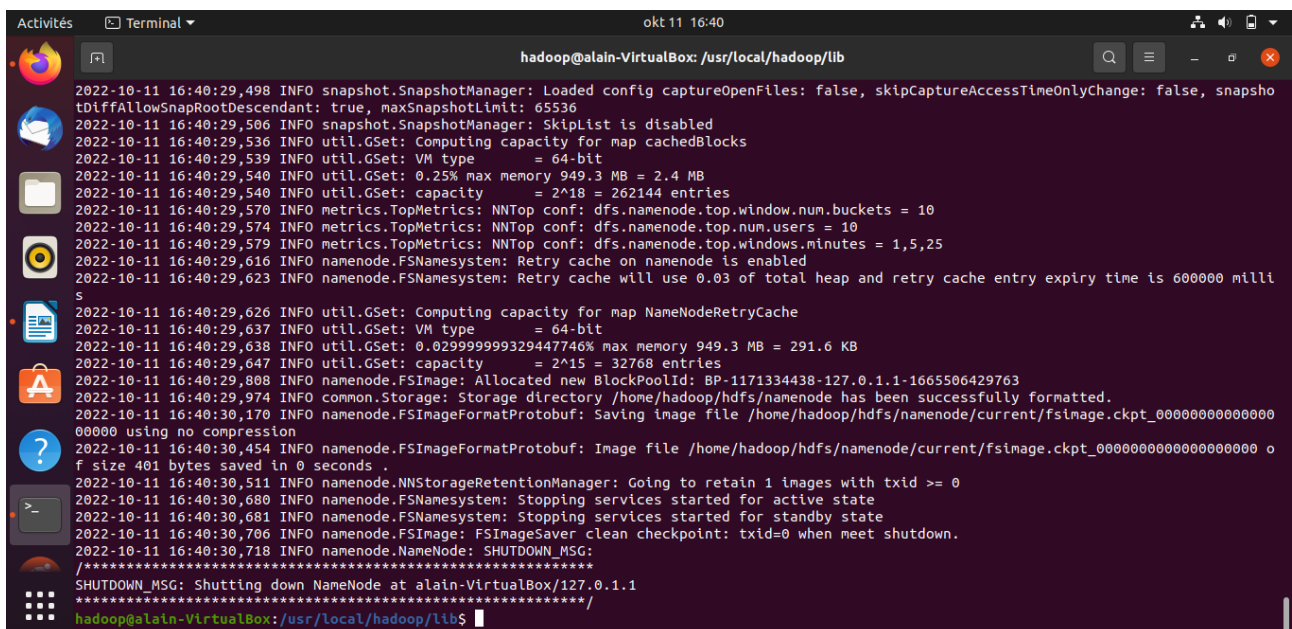
    http://www.apache.org/licenses/LICENSE-2.0

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 implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

fig

Validate the Hadoop configuration and format the HDFS NameNode.

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



```
hadoop@alain-VirtualBox: /usr/local/hadoop/lib
2022-10-11 16:40:29,498 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAccessTimeOnlyChange: false, snapshotDiffFollowSnapRootDescendant: true, maxSnapshotLimit: 65536
2022-10-11 16:40:29,506 INFO snapshot.SnapshotManager: SkipList is disabled
2022-10-11 16:40:29,536 INFO util.GSet: Computing capacity for map cachedBlocks
2022-10-11 16:40:29,539 INFO util.GSet: VM type = 64-bit
2022-10-11 16:40:29,540 INFO util.GSet: 0.25% max memory 949.3 MB = 2.4 MB
2022-10-11 16:40:29,540 INFO util.GSet: capacity = 2^18 = 262144 entries
2022-10-11 16:40:29,570 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-10-11 16:40:29,574 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-10-11 16:40:29,579 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-10-11 16:40:29,616 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-10-11 16:40:29,623 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 milliseconds
2022-10-11 16:40:29,626 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-10-11 16:40:29,637 INFO util.GSet: VM type = 64-bit
2022-10-11 16:40:29,638 INFO util.GSet: 0.0299999999329447746% max memory 949.3 MB = 291.6 KB
2022-10-11 16:40:29,647 INFO util.GSet: capacity = 2^15 = 32768 entries
2022-10-11 16:40:29,808 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1171334438-127.0.1.1-1665506429763
2022-10-11 16:40:29,974 INFO common.Storage: Storage directory /home/hadoop/hdfs/namenode has been successfully formatted.
2022-10-11 16:40:30,170 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/hdfs/namenode/current/fsimage.ckpt_00000000000000000000 using no compression
2022-10-11 16:40:30,454 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hdfs/namenode/current/fsimage.ckpt_00000000000000000000 of size 401 bytes saved in 0 seconds.
2022-10-11 16:40:30,511 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-10-11 16:40:30,680 INFO namenode.FSNamesystem: Stopping services started for active state
2022-10-11 16:40:30,681 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-10-11 16:40:30,706 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2022-10-11 16:40:30,718 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at alain-VirtualBox/127.0.1.1
*****/
hadoop@alain-VirtualBox: /usr/local/hadoop/lib$
```

fig

8- demarrage du cluster hadoop

demarrage de namenode et de datanode

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

```
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ start-dfs.sh
Starting namenodes on [0.0.0.0]
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts.
Starting datanodes
Starting secondary namenodes [alain-VirtualBox]
alain-VirtualBox: Warning: Permanently added 'alain-virtualbox' (ECDSA) to the list of known hosts.
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$
```

fig

Demarrage de yarn ressource et de node manager

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

```
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$
```

fig

verification de demarrage des composants

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

```
Starting nodemanagers
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$ jps
9361 Jps
8965 NodeManager
8310 DataNode
8839 ResourceManager
8187 NameNode
8492 SecondaryNameNode
hadoop@alain-VirtualBox:/usr/local/hadoop/lib$
```

verification demarrage

9- acces à l'interface de hadoop

Activités | Navigateur Web Firefox | okt 11 16:56

Install and Configure App | Firefox Politique de confi | Namenode information

192.168.1.110:9870/dfshealth.html#tab-overview

Hadoop | Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities

Overview '0.0.0.0:9000' (✓active)

Started:	Tue Oct 11 16:49:03 +0000 2022
Version:	3.3.1, ra3b9c37a397ad4188041dd80621bdeefc46885f2
Compiled:	Tue Jun 15 05:13:00 +0000 2021 by ubuntu from (HEAD detached at release-3.3.1-RC3)
Cluster ID:	CID-979c0a4d-2306-466f-ac56-9fdd7e1d16fe
Block Pool ID:	BP-1171334438-127.0.1.1-1665506429763

Summary

Security is off.
Safemode is off.

interface hadoop

Activités | Navigateur Web Firefox | okt 11 16:57

Install and Configure App | Firefox Politique de confi | Namenode information

192.168.1.110:9870/dfshealth.html#tab-overview

Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	10.7 GB
DFS Remaining:	7.33 GB (38.54%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Tue Oct 11 16:49:03 +0000 2022
Last Checkpoint Time	Tue Oct 11 16:40:30 +0000 2022

fig

Activités

Navigateur Web Firefox

okt 11 17:01

Install and Configure Apache Hadoop 3.2.1

Firefox Politique de confidentialité

Namenode information

192.168.1.110:9870/dfshealth.html#tab-overview

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
HDFS	10.00 GB	24 KB (0%)	7.99 GB (79.54%)	24 KB	1

fig