

Installation hadoop sur ubuntu

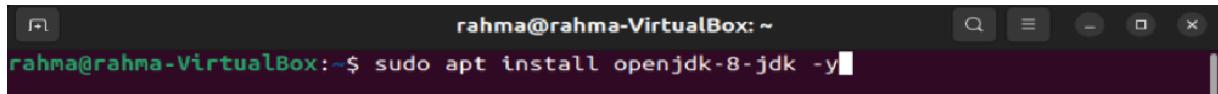
Phase 1 : Installation de la machine virtuelle ‘Oracle VirtualBox’

Phase 2 : Installation Ubuntu

Phase 3 : installation hadoop sur ubuntu

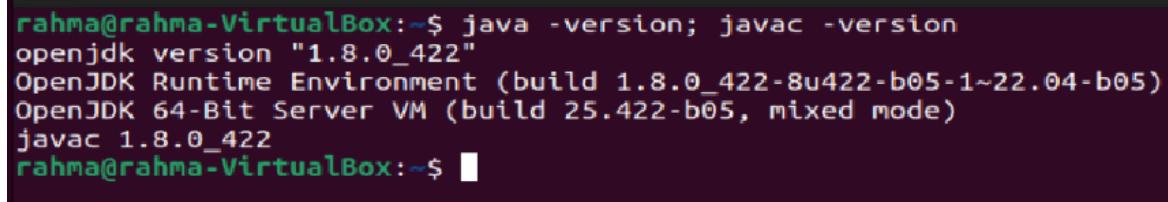
1/ installer version java 8

=➔ sudo apt-get install openjdk-8-jdk



```
rahma@rahma-VirtualBox:~$ sudo apt install openjdk-8-jdk -y
```

====➔ java -version; javac -version



```
rahma@rahma-VirtualBox:~$ java -version; javac -version
openjdk version "1.8.0_422"
OpenJDK Runtime Environment (build 1.8.0_422-8u422-b05-1~22.04-b05)
OpenJDK 64-Bit Server VM (build 25.422-b05, mixed mode)
javac 1.8.0_422
rahma@rahma-VirtualBox:~$
```

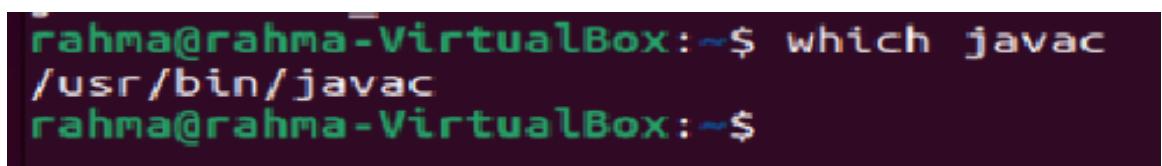
2/ installer les packages ssh et rsync en exécutant les commandes suivantes.

=➔ sudo apt-get install ssh -y

=➔ sudo apt-get install rsync -y

3/ Taper ‘ which javac ’ :

permet de déterminer le chemin d'accès de la commande javac (le compilateur Java) dans le système.



```
rahma@rahma-VirtualBox:~$ which javac
/usr/bin/javac
rahma@rahma-VirtualBox:~$
```

4/ taper ' readlink -f /usr/bin/javac '

⇒ permet d'obtenir le chemin réel (résolu) du fichier javac

```
rahma@rahma-VirtualBox:~$ readlink -f /usr/bin/javac
/usr/lib/jvm/java-8-openjdk-amd64/bin/javac
rahma@rahma-VirtualBox:~$
```

5/ taper 'sudo nano .bashrc'

```
[sudo] password for rahma:
rahma@rahma-VirtualBox:~$ sudo nano .bashrc
[sudo] password for rahma: [
```

Puis Ajouter 'export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64 '
A la fin

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

⇒ puis contrôle +x pour sortir

6/ Installer OpenSSH on Ubuntu

Tapper ' sudo apt install openssh-server openssh-client -y '

```
rahma@rahma-VirtualBox:~$ sudo apt install openssh-server openssh-client -y
```

7/ Activer SSH sans mot de passe pour l'utilisateur Hadoop

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

```
rahma@rahma-VirtualBox:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
```

8/ Taper ' cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys '

⇒ la commande cat pour stocker la clé publique sous le nom authorized_keys dans le répertoire ssh

```
rahma@rahma-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

9/Taper 'chmod 0600 ~/.ssh/authorized_keys'

=> commande chmod pour définir les autorisations de votre utilisateur

```
rahma@rahma-VirtualBox:~$ chmod 0600 ~/.ssh/authorized_keys
```

10/taper ' ssh localhost '

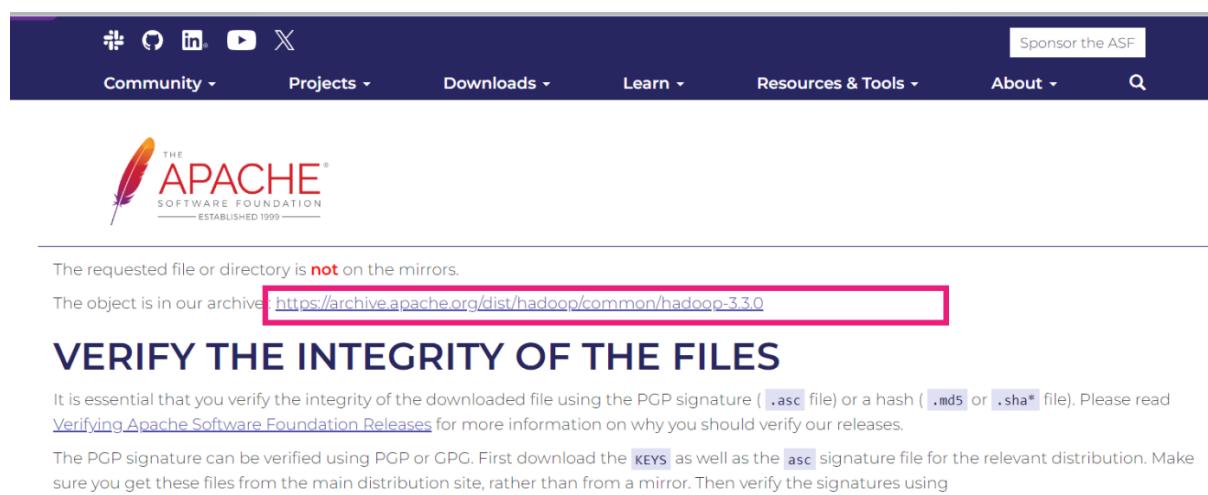
=> ssh localhost permet de se connecter en SSH à sa propre machine, ce qui est utile pour tester la configuration SSH ou pour exécuter des commandes sur la machine locale via une connexion sécurisée.

```
rahma@rahma-VirtualBox:~$ ssh localhost
```

11/Copie adresse du lien Hadoop version 3.3.0

⊕ Lien =>

<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.0>

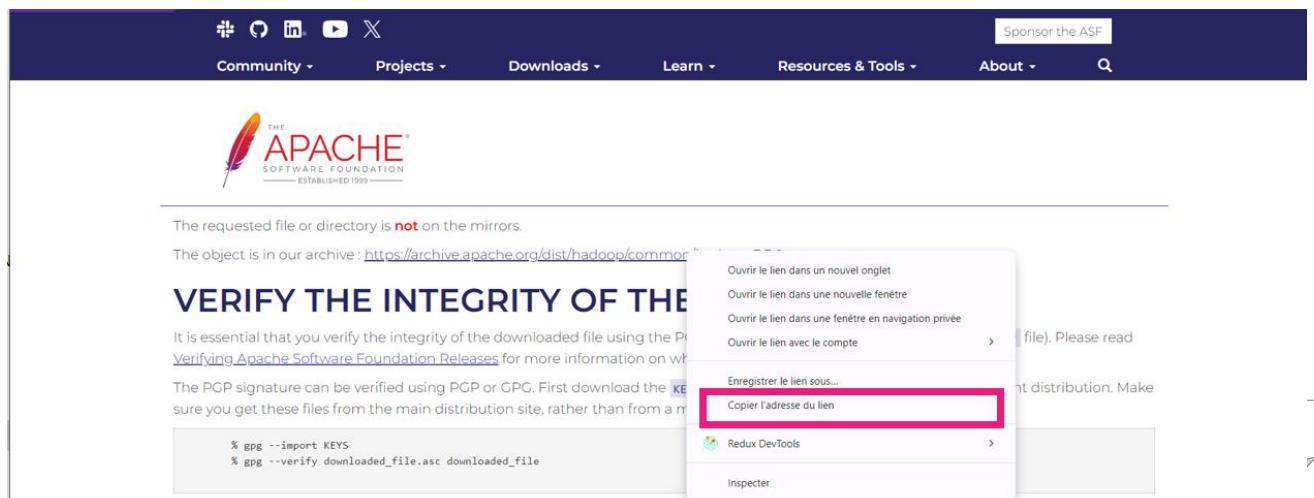


The requested file or directory is **not** on the mirrors.
 The object is in our archive: <https://archive.apache.org/dist/hadoop/common/hadoop-3.3.0>

VERIFY THE INTEGRITY OF THE FILES

It is essential that you verify the integrity of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file). Please read [Verifying Apache Software Foundation Releases](#) for more information on why you should verify our releases.

The PGP signature can be verified using PGP or GPG. First download the `KEYS` as well as the `.asc` signature file for the relevant distribution. Make sure you get these files from the main distribution site, rather than from a mirror. Then verify the signatures using



The requested file or directory is **not** on the mirrors.
 The object is in our archive: <https://archive.apache.org/dist/hadoop/common/hadoop-3.3.0>

VERIFY THE INTEGRITY OF THE FILES

It is essential that you verify the integrity of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file). Please read [Verifying Apache Software Foundation Releases](#) for more information on why you should verify our releases.

The PGP signature can be verified using PGP or GPG. First download the `KEYS` as well as the `.asc` signature file for the relevant distribution. Make sure you get these files from the main distribution site, rather than from a mirror.

```
% gpg --import KEYS
% gpg --verify downloaded_file.asc downloaded_file
```

Ouvrir le lien dans un nouvel onglet
 Ouvrir le lien dans une nouvelle fenêtre
 Ouvrir le lien dans une fenêtre en navigation privée
 Ouvrir le lien avec le compte
 Enregistrer le lien sous...
 Copier l'adresse du lien

12/ Télécharger et installer Hadoop sur Ubuntu on utilisant terminal

‘<https://archive.apache.org/dist/hadoop/common/hadoop-3.3.0>’

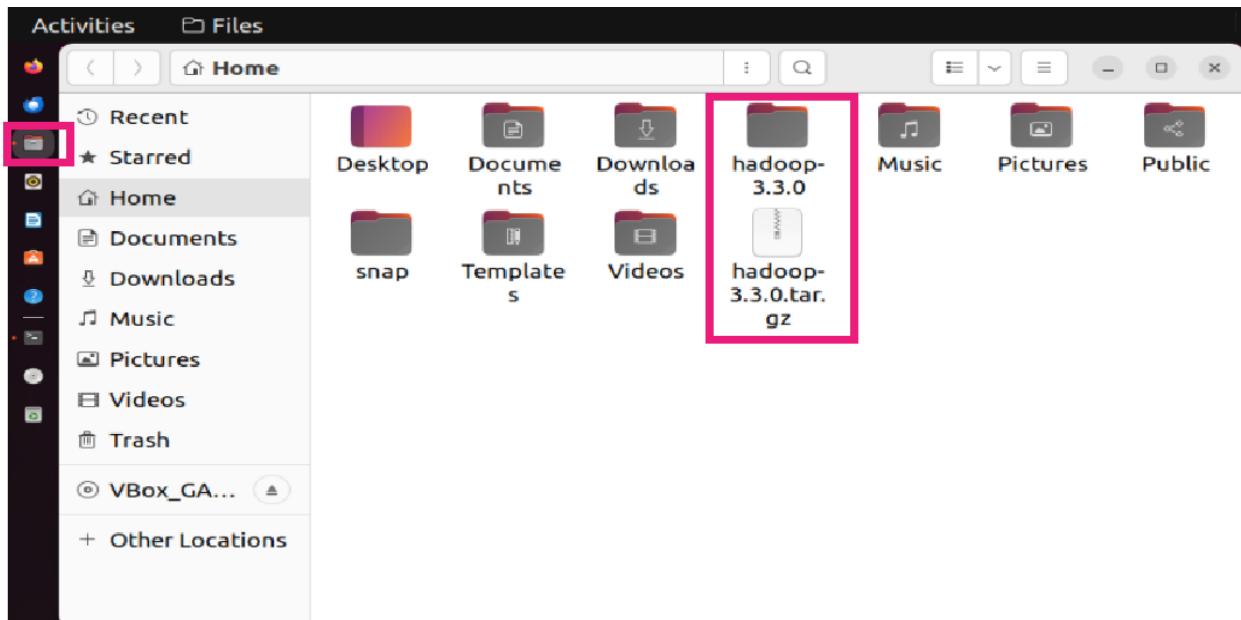
=→ c'est l'adresse du lien de notre hadoop

```
rahma@rahma-VirtualBox:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-3.3.0
```

13/ Extraire le fichier

Taper ‘tar xzf hadoop-3.3.0.tar.gz ’

====→ Donc le dossier est déjà téléchargé



14/on va modifiez les files suivantes

- ✚ bashrc
- ✚ hadoop-env.sh
- ✚ core-site.xml
- ✚ hdfs-site.xml
- ✚ mapred-site.xml
- ✚ yarn-site.xml

14/*1 bashrc

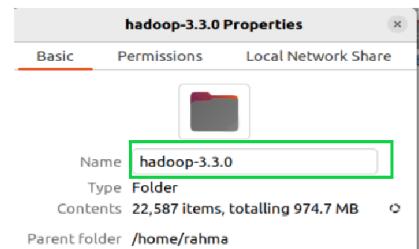
==→ ouvrir avec le terminal . Taper ‘ sudo nano .bashrc ‘
puis taper votre mot de passe

```
rahma@rahma-VirtualBox:~$ sudo nano .bashrc
[sudo] password for rahma: 
```

On ajoute

```
#Hadoop Related Options

export HADOOP_HOME=/home/rahma/hadoop-3.3.0
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"
```



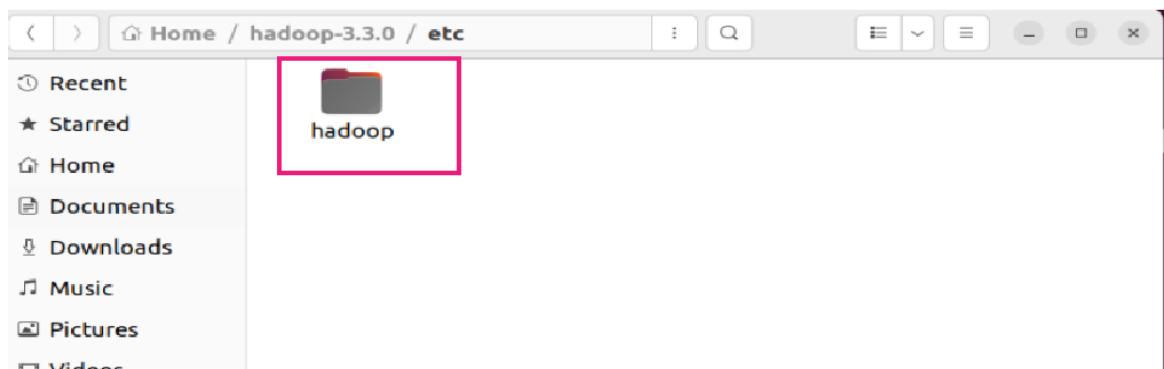
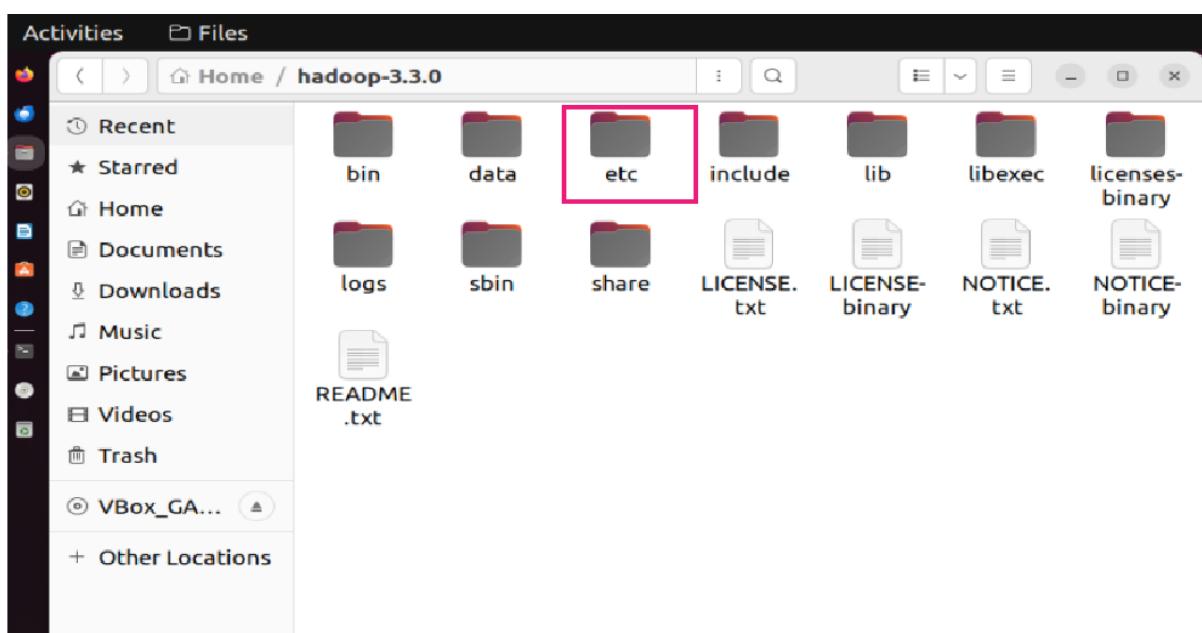
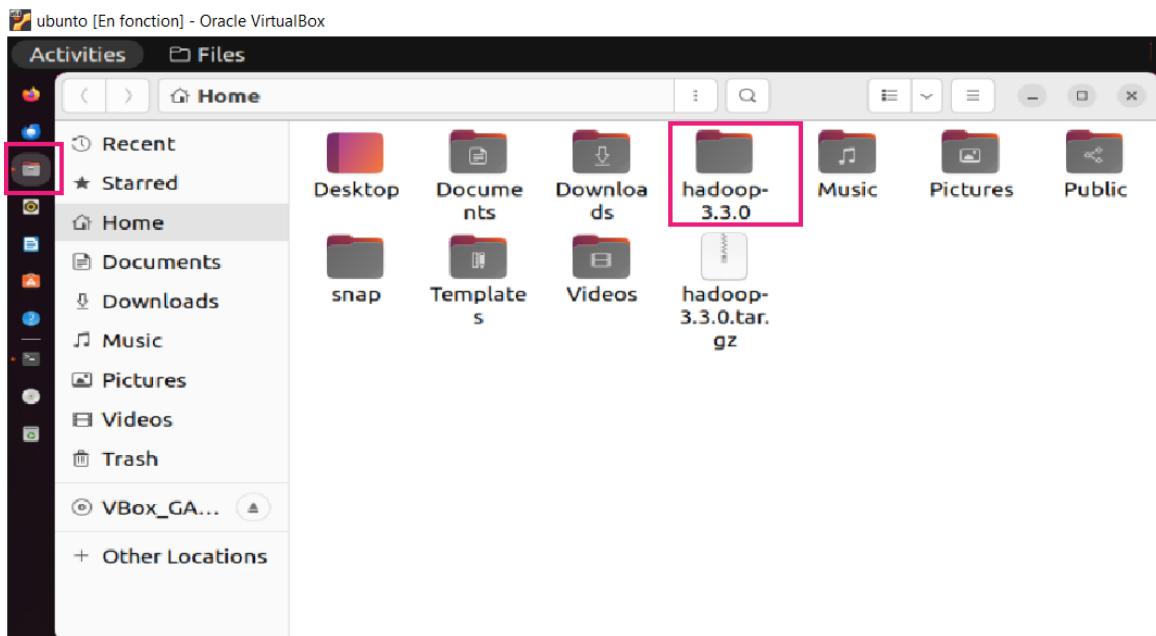
```
#Hadoop Related Options
export HADOOP_HOME=/home/rahma/hadoop-3.3.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

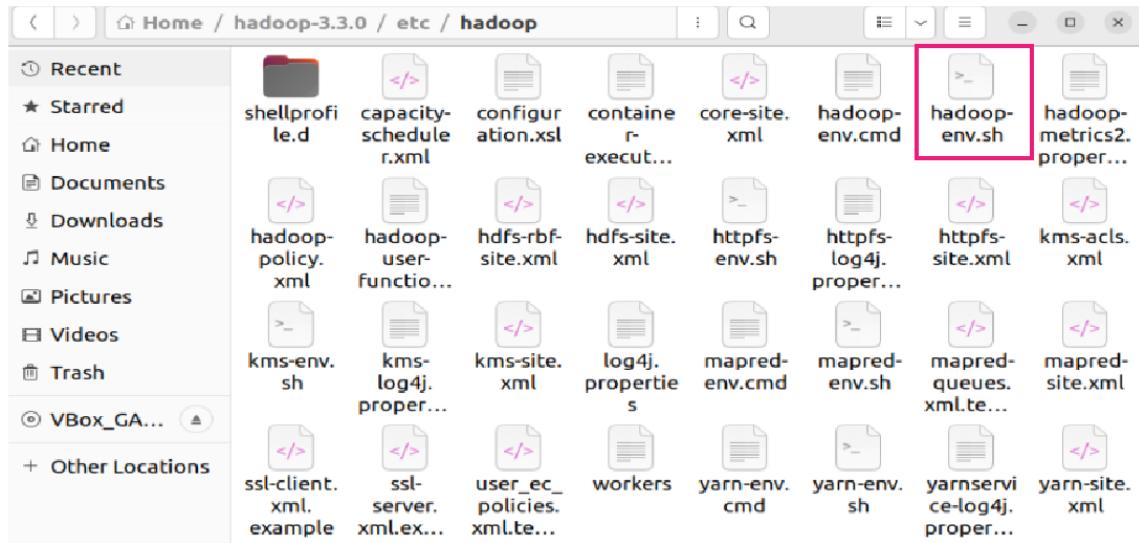
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

Puis contrôle+o → pour sauvegarder vos changements

Puis contrôle +x → pour quitter l'éditeur

14/*2 File : hadoop-env.sh





Décommentez JAVA_HOME et modifier selon :

JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64 ==> qui existe dans bashrc

```

29 ##
30 ## {YARN_xyz/HDFS_xyz} > HADOOP_xyz > hard-coded defaults
31 ##
32
33 # Many of the options here are built from the perspective that users
34 # may want to provide OVERWRITING values on the command line.
35 # For example:
36 #
37 # JAVA_HOME=/usr/java/testing hdfs dfs -ls
38
39 JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
40
41

```

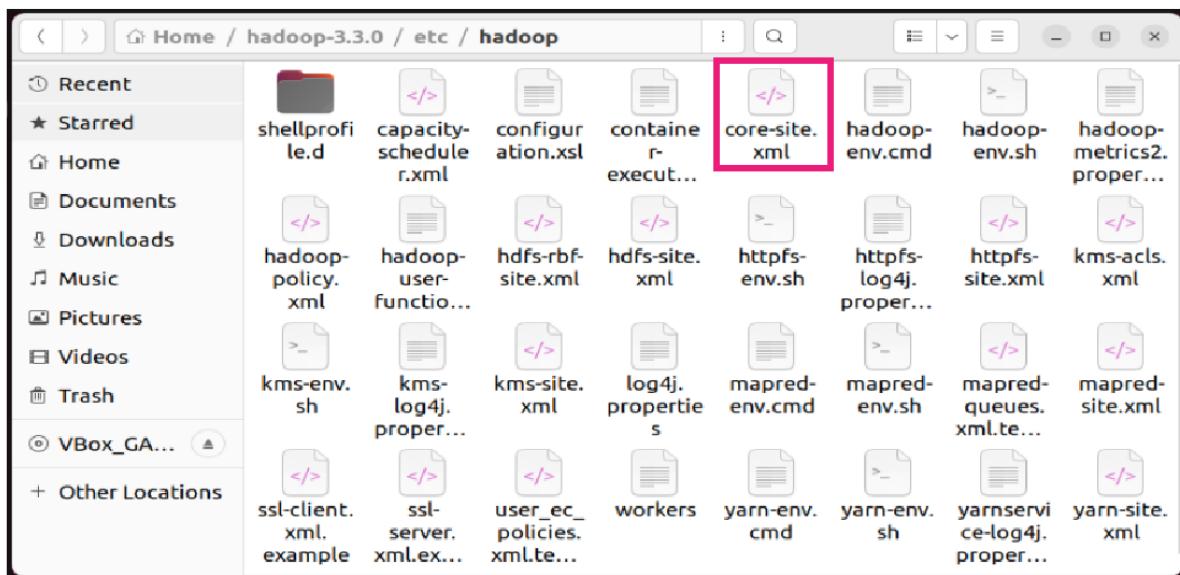
14/*3 File: core-site.xml

==> On ajoute

```

<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>

```

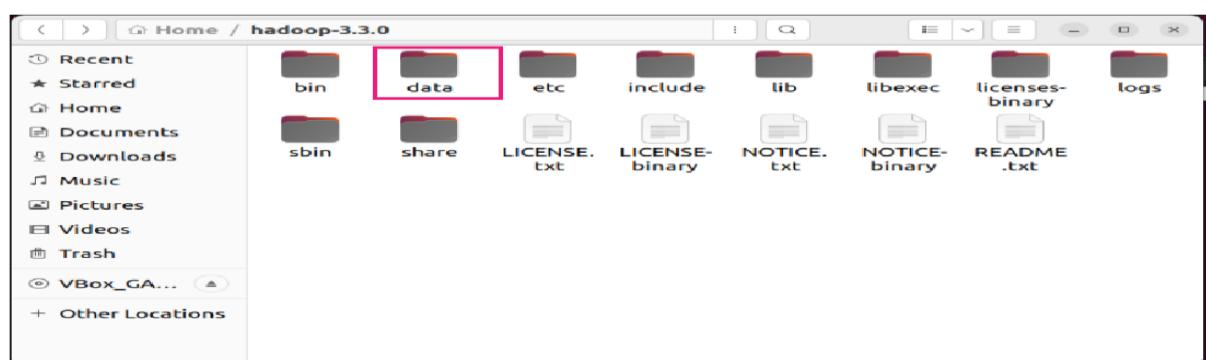


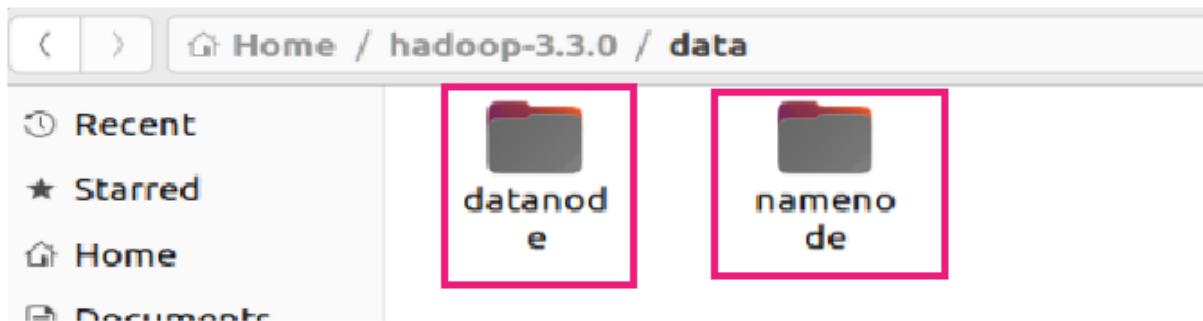
```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4   Licensed under the Apache License, Version 2.0 (the "License");
5   you may not use this file except in compliance with the License.
6   You may obtain a copy of the License at
7
8     http://www.apache.org/licenses/LICENSE-2.0
9
10  Unless required by applicable law or agreed to in writing, software
11  distributed under the License is distributed on an "AS IS" BASIS,
12  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13  See the License for the specific language governing permissions and
14  limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20 <property>
21   <name>fs.default.name</name>
22   <value>hdfs://localhost:9000</value>
23 </property>
24 </configuration>

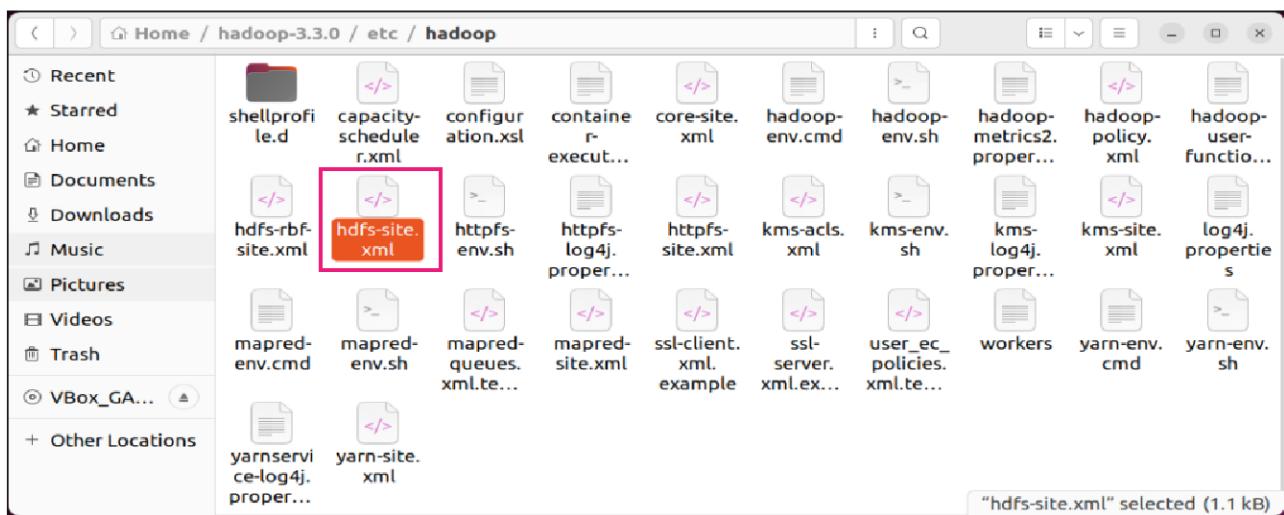
```

- Créez un dossier 'data' qui contient un dossier 'datanode' et un dossier 'namenodes'





✚ 14/*4 File : hdfs-site.xml

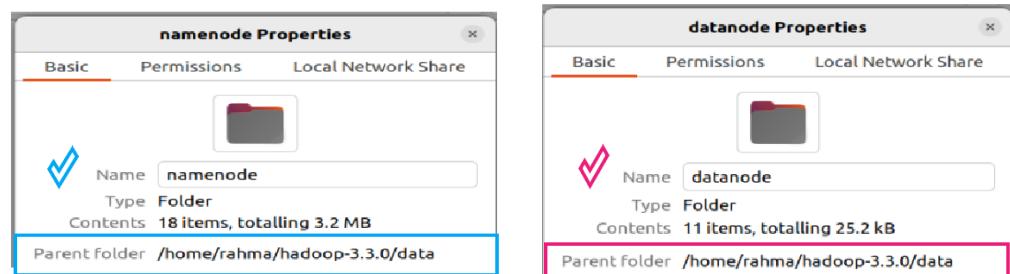


✚ On Ajout

```

<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/home/rahma/hadoop-3.3.0/data/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/home/rahma/hadoop-3.3.0/data/datanode</value>
</property>

```



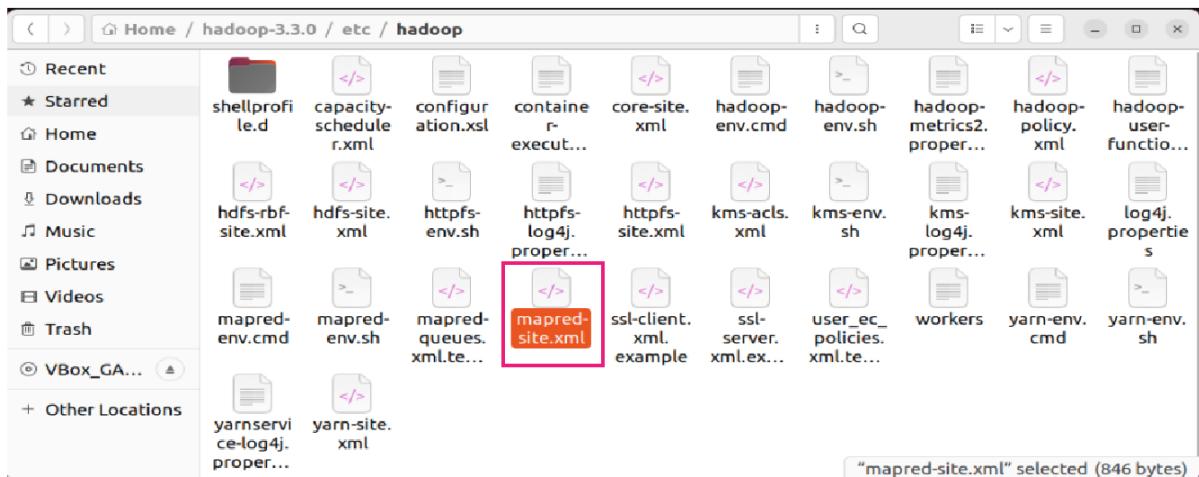


```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>dfs.replication</name>
22     <value>1</value>
23   </property>
24   <property>
25     <name>dfs.namenode.name.dir</name>
26     <value>/home/rahma/hadoop-3.3.0/data/namenode</value>
27   </property>
28   <property>
29     <name>dfs.datanode.data.dir</name>
30     <value>/home/rahma/hadoop-3.3.0/data/datanode</value>
31   </property>
32
33 </configuration>

```

14/*5 File : mapred-site-xml

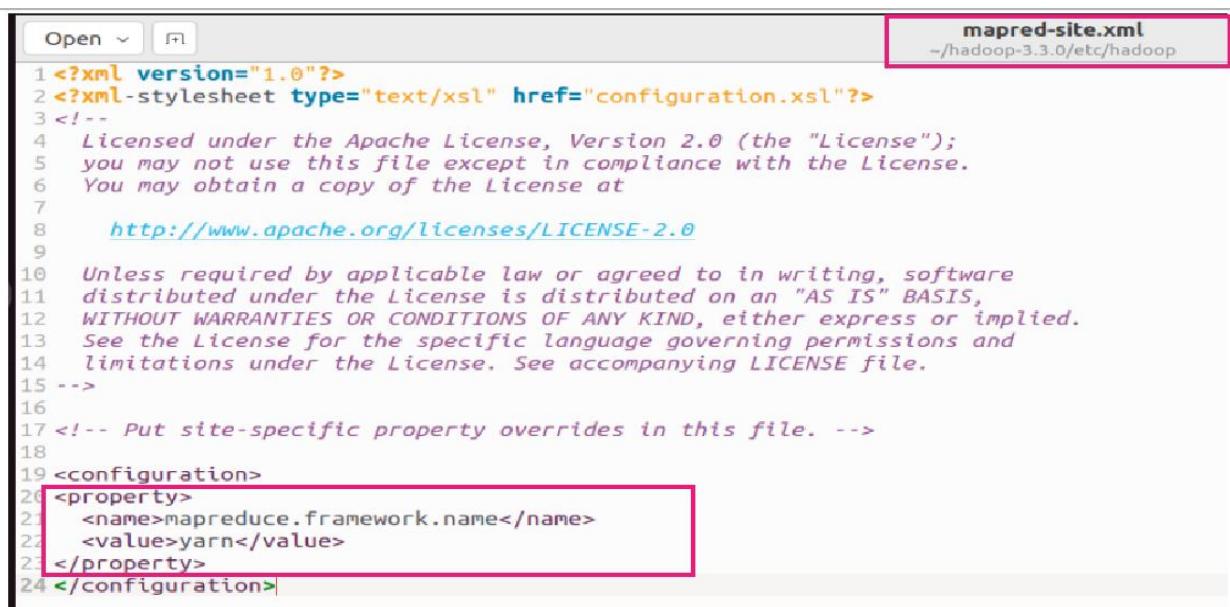


On Ajout :

```

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

```

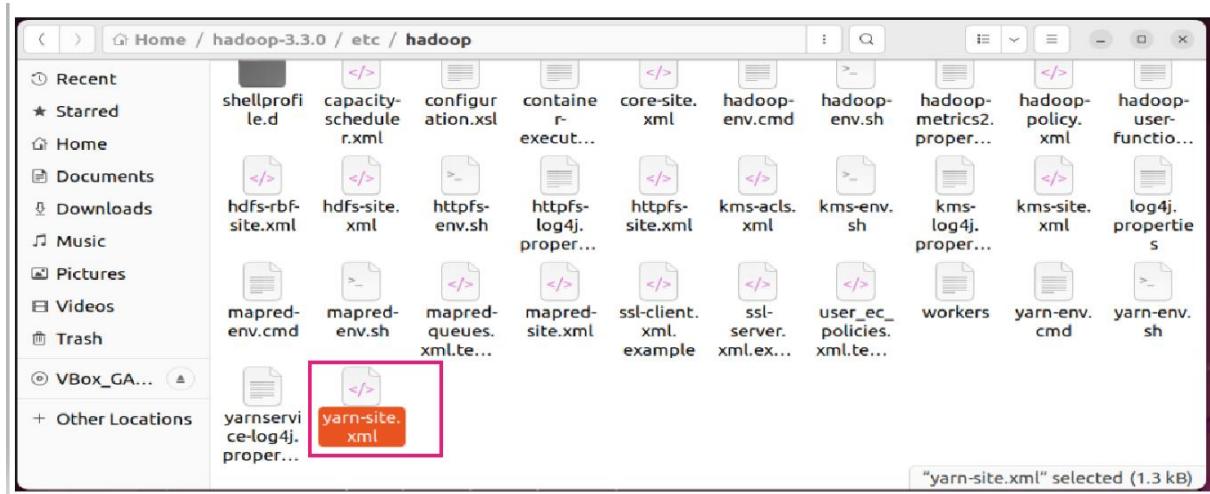


```

1 <?xml version="1.0"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20 <property>
21 <name>mapreduce.framework.name</name>
22 <value>yarn</value>
23 </property>
24 </configuration>

```

14/*6 File : yarn-site.xml



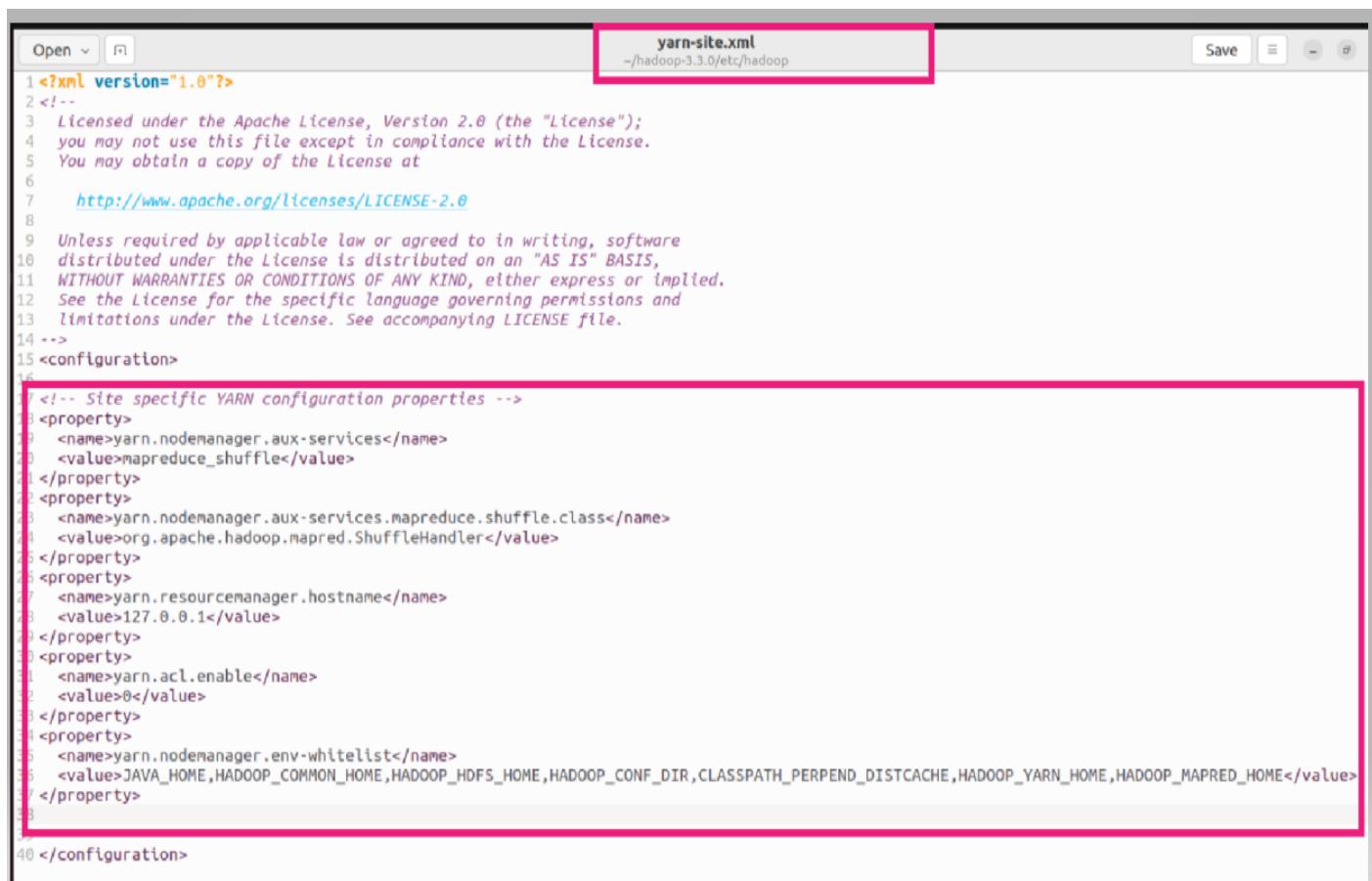
On Ajoute

```

<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>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>

```

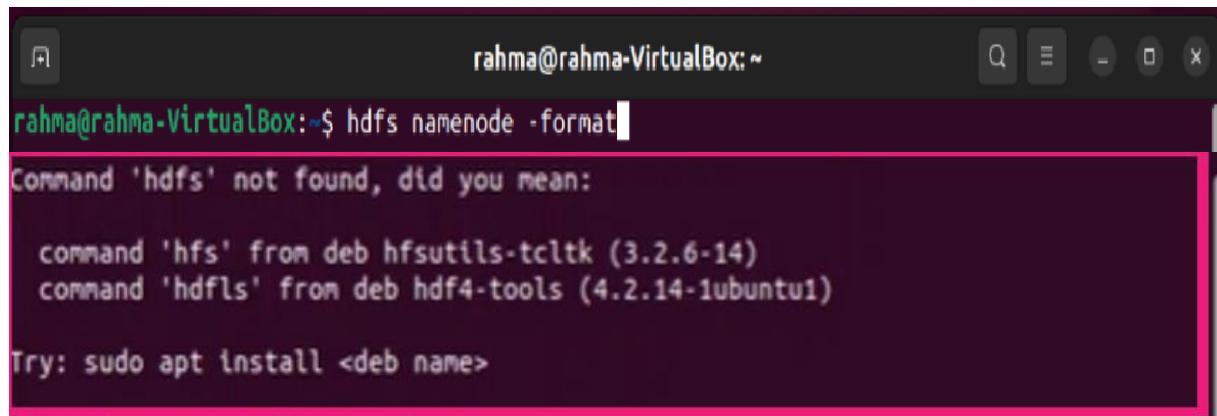
```
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>
  <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HAD
OOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,H
ADOOP_MAPRED_HOME</value>
</property>
```



```
Open Save yarn-site.xml ~/hadoop-3.3.0/etc/hadoop
1 <?xml version="1.0"?>
2 <!--
3   Licensed under the Apache License, Version 2.0 (the "License");
4   you may not use this file except in compliance with the License.
5   You may obtain a copy of the License at
6
7     http://www.apache.org/licenses/LICENSE-2.0
8
9   Unless required by applicable law or agreed to in writing, software
10  distributed under the License is distributed on an "AS IS" BASIS,
11  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
12  See the License for the specific language governing permissions and
13  limitations under the License. See accompanying LICENSE file.
14 -->
15 <configuration>
16
17 <!-- Site specific YARN configuration properties -->
18 <property>
19   <name>yarn.nodemanager.aux-services</name>
20   <value>mapreduce_shuffle</value>
21 </property>
22 <property>
23   <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
24   <value>org.apache.hadoop.mapred.ShuffleHandler</value>
25 </property>
26 <property>
27   <name>yarn.resourcemanager.hostname</name>
28   <value>127.0.0.1</value>
29 </property>
30 <property>
31   <name>yarn.acl.enable</name>
32   <value>0</value>
33 </property>
34 <property>
35   <name>yarn.nodemanager.env-whitelist</name>
36   <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HAD
OOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
37 </property>
38
39 </configuration>
```

15/ Formatez le NameNode HDFS

Taper ' hdfs namenode -format '



```
rahma@rahma-VirtualBox:~$ hdfs namenode -format
Command 'hdfs' not found, did you mean:
  command 'hfs' from deb hfsutils-tcltk (3.2.6-14)
  command 'hdfls' from deb hdf4-tools (4.2.14-1ubuntu1)
Try: sudo apt install <deb name>
```

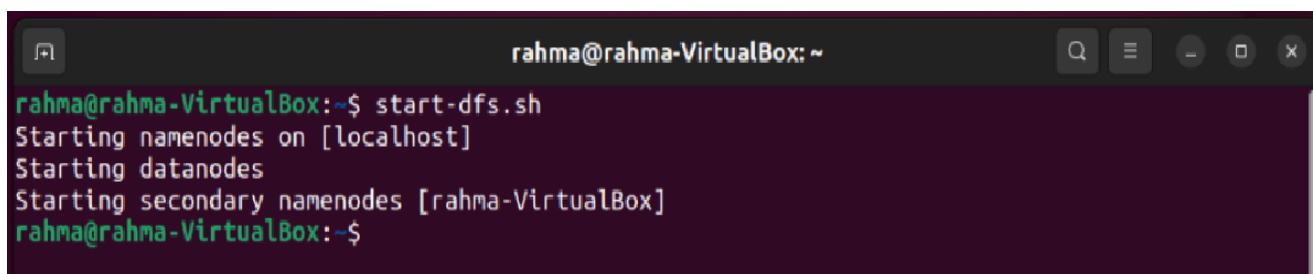
=====→ Pour résoudre ce problème :

=→ je fais cette commande ' sudo apt-get install Hadoop '

Cette commande installe Hadoop via les dépôts d'Ubuntu, ce qui permet d'avoir toutes les commandes, y compris hdfs, disponibles et fonctionnelles.

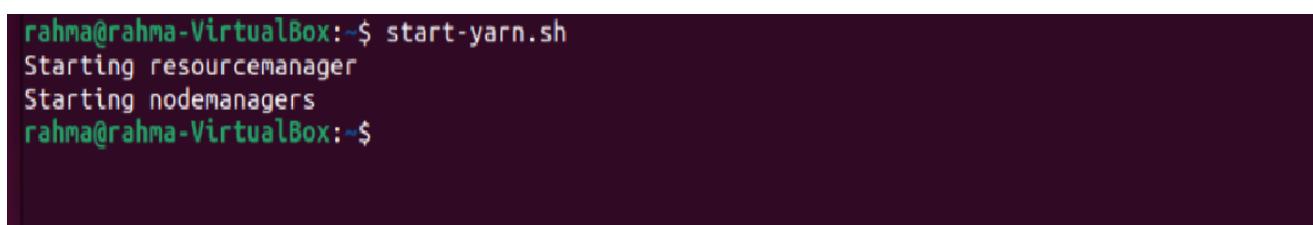
16/ Démarrer le cluster Hadoop

=→ Taper 'start-dfs.sh '



```
rahma@rahma-VirtualBox:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [rahma-VirtualBox]
rahma@rahma-VirtualBox:~$
```

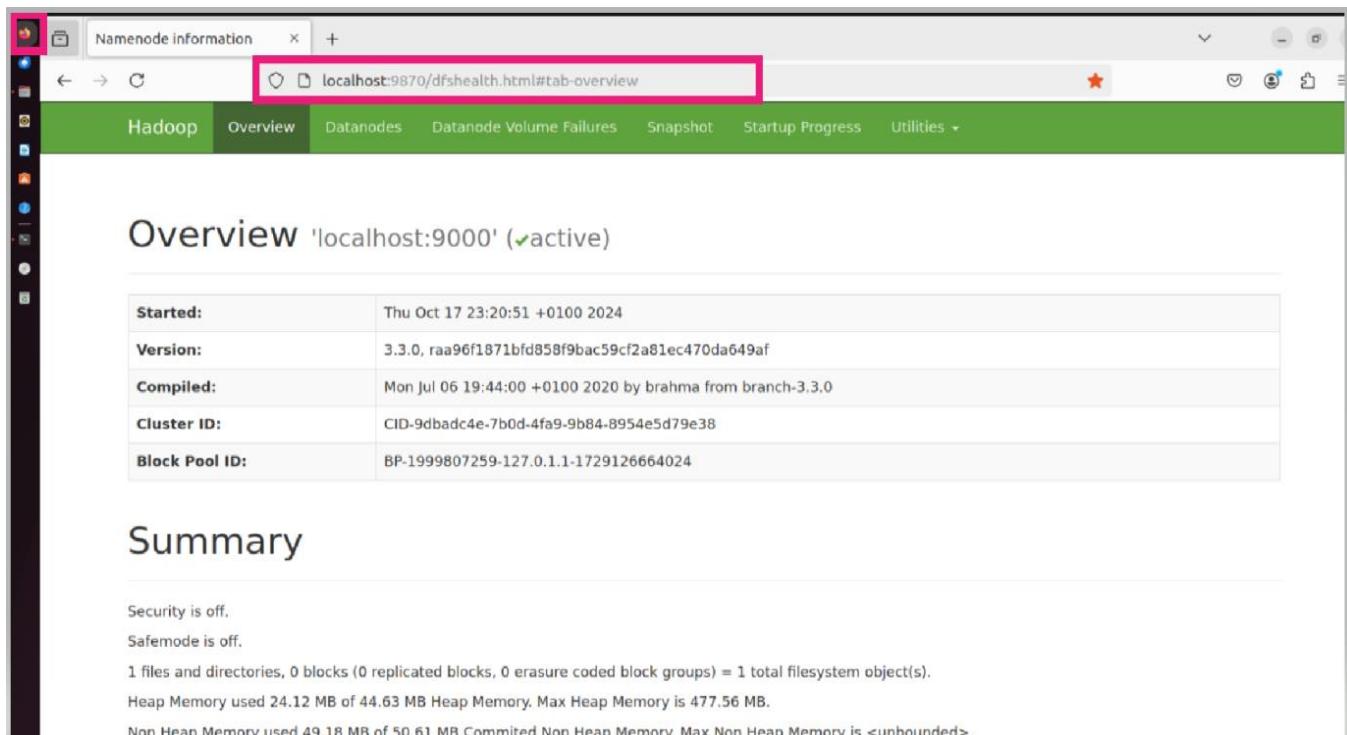
====→ Taper ' start-yarn.sh '



```
rahma@rahma-VirtualBox:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
rahma@rahma-VirtualBox:~$
```

17/ Accéder à l'interface utilisateur de Hadoop depuis un navigateur.

====→Interface utilisateur Hadoop NameNode : <http://localhost:9870>



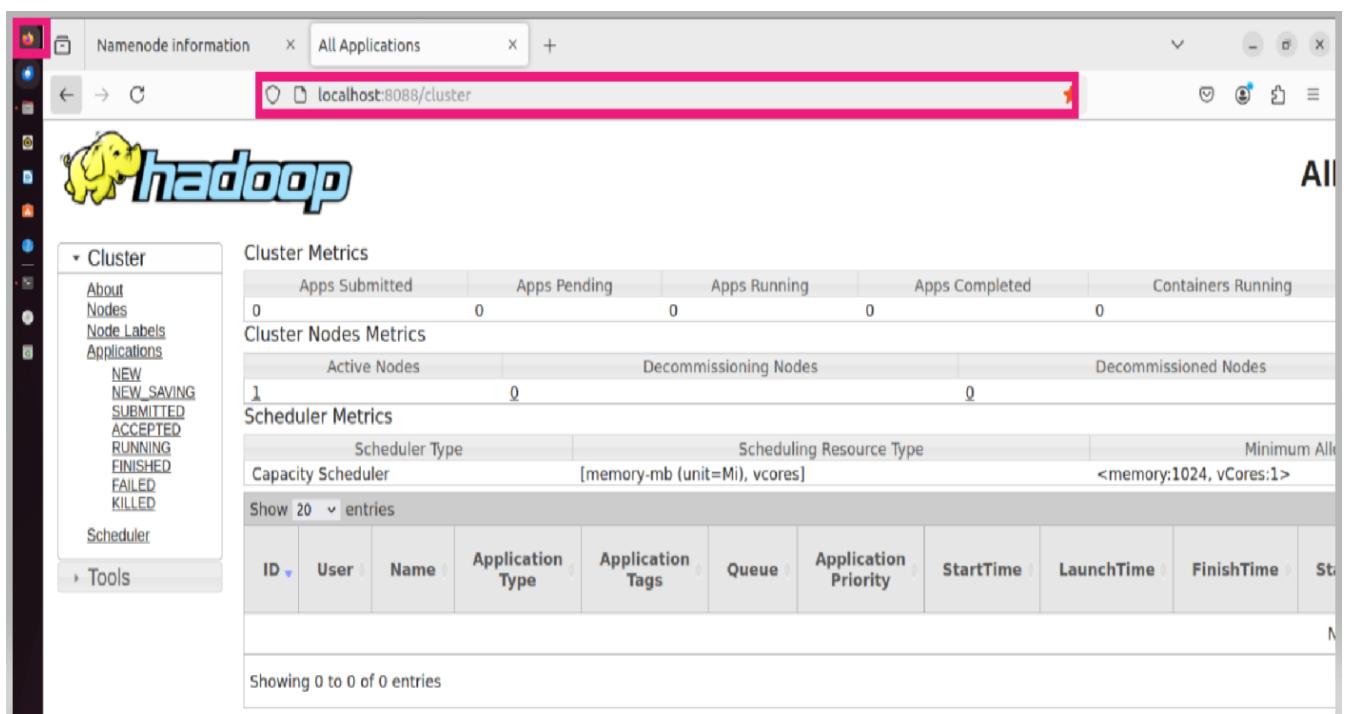
Overview 'localhost:9000' (✓active)

Started:	Thu Oct 17 23:20:51 +0100 2024
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Mon Jul 06 19:44:00 +0100 2020 by brahma from branch-3.3.0
Cluster ID:	CID-9dbadc4e-7b0d-4fa9-9b84-8954e5d79e38
Block Pool ID:	BP-1999807259-127.0.1.1-1729126664024

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 24.12 MB of 44.63 MB Heap Memory. Max Heap Memory is 477.56 MB.
Non Heap Memory used 49.18 MB of 50.61 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

====→Gestionnaire de ressources YARN : <http://localhost:8088>



hadoop

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
0	0	0	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocated
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State
Showing 0 to 0 of 0 entries										

Remarque :1

- Pour assurer la communication (copie, coller) entre la machine virtuelle et Windows, il est recommandé d'installer les "Guest Additions".

Les Etapes D'installation Les *Guest Additions*

1/taper ‘sudo -i ’

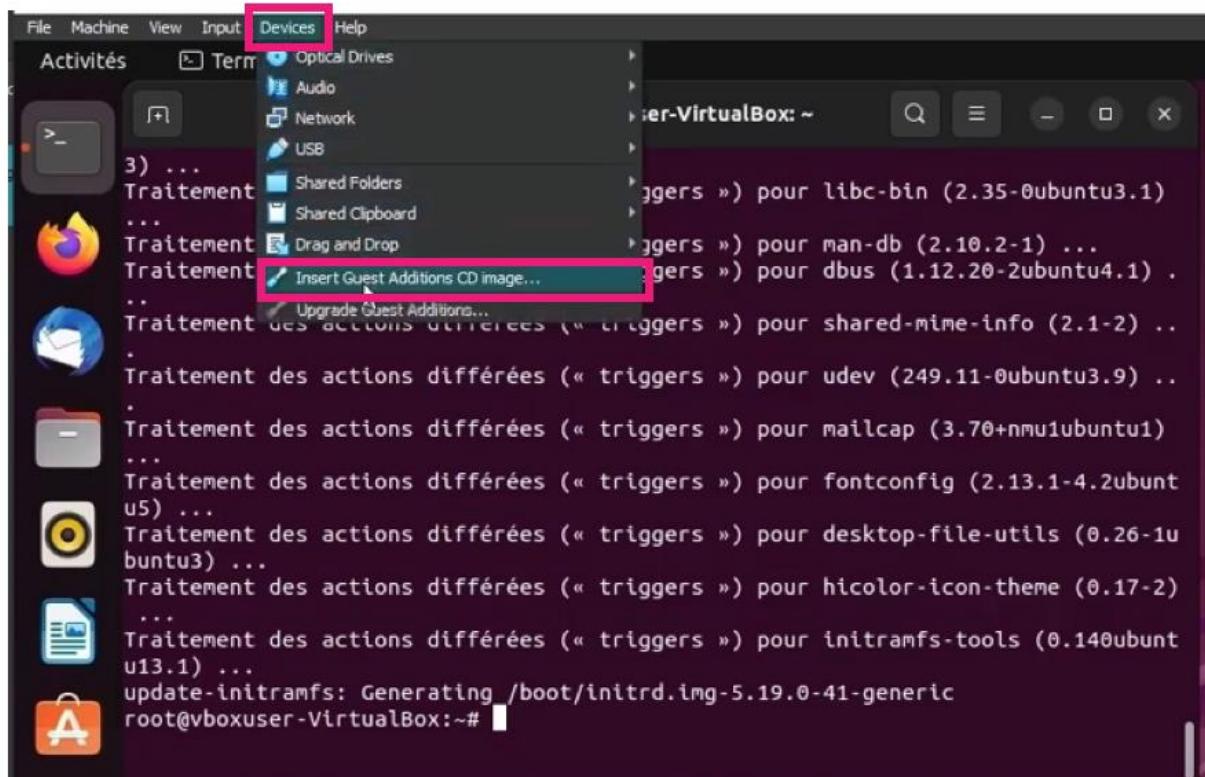


rahma@rahma-VirtualBox:~\$ sudo -i
[sudo] password for rahma:
root@rahma-VirtualBox:~#

2/taper 'apt update && apt upgrade'

```
rahma@rahma-VirtualBox:~$ sudo -i
[sudo] password for rahma:
root@rahma-VirtualBox:~# apt update && apt upgrade
```

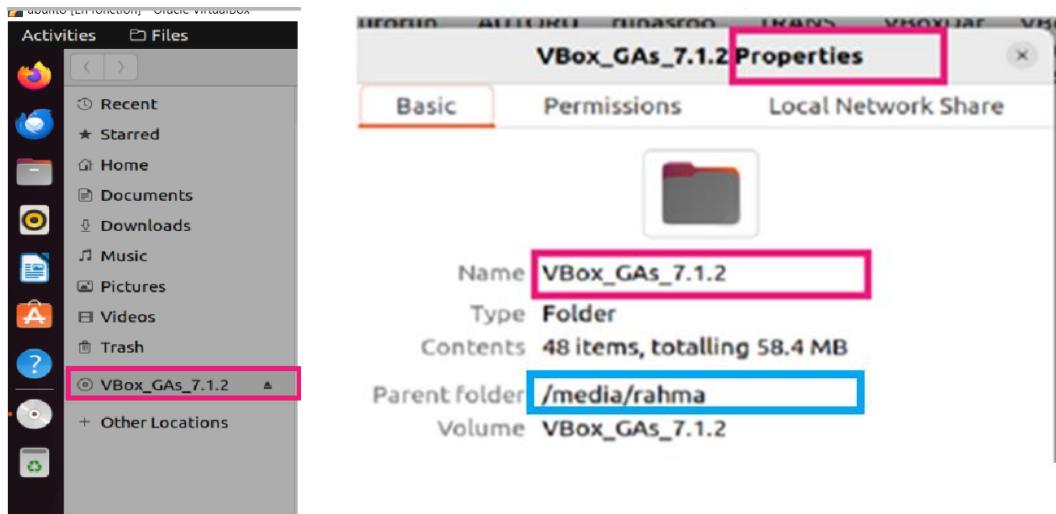
3/ Device =>insert Guest Additions cd image



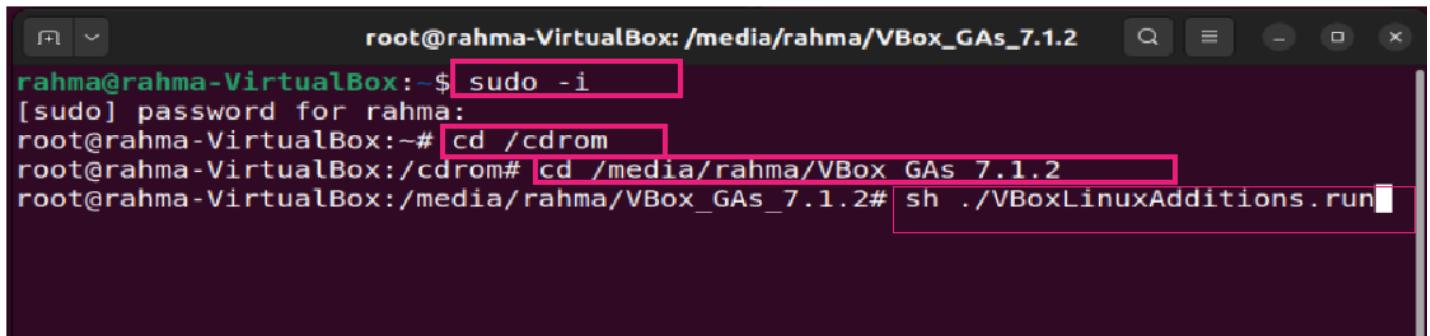
4/ taper 'cd /cdrom'

Puis taper 'cd /media/rahma/VBox_GAs_7.1.2'

Et taper 'sh ./VBoxLinuxAdditions.run'

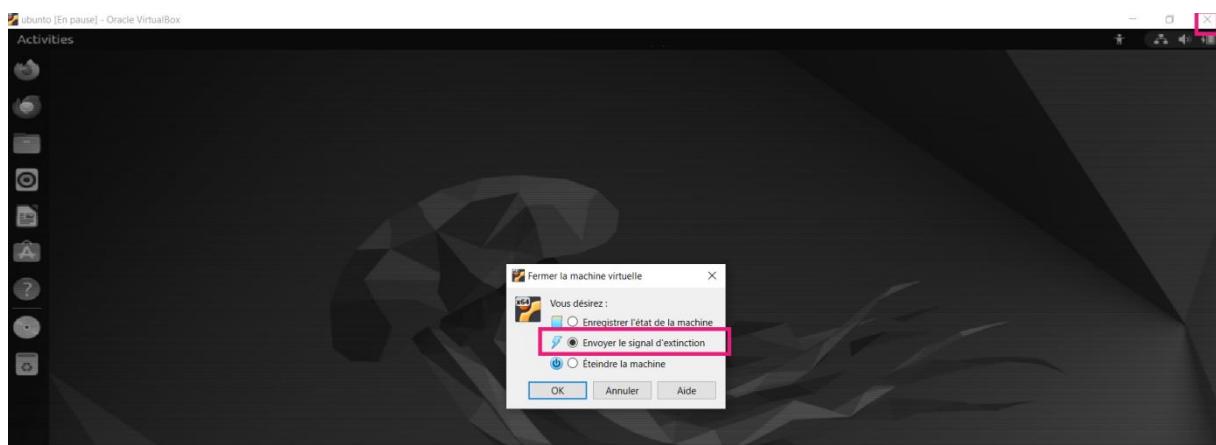


=====> En Resume



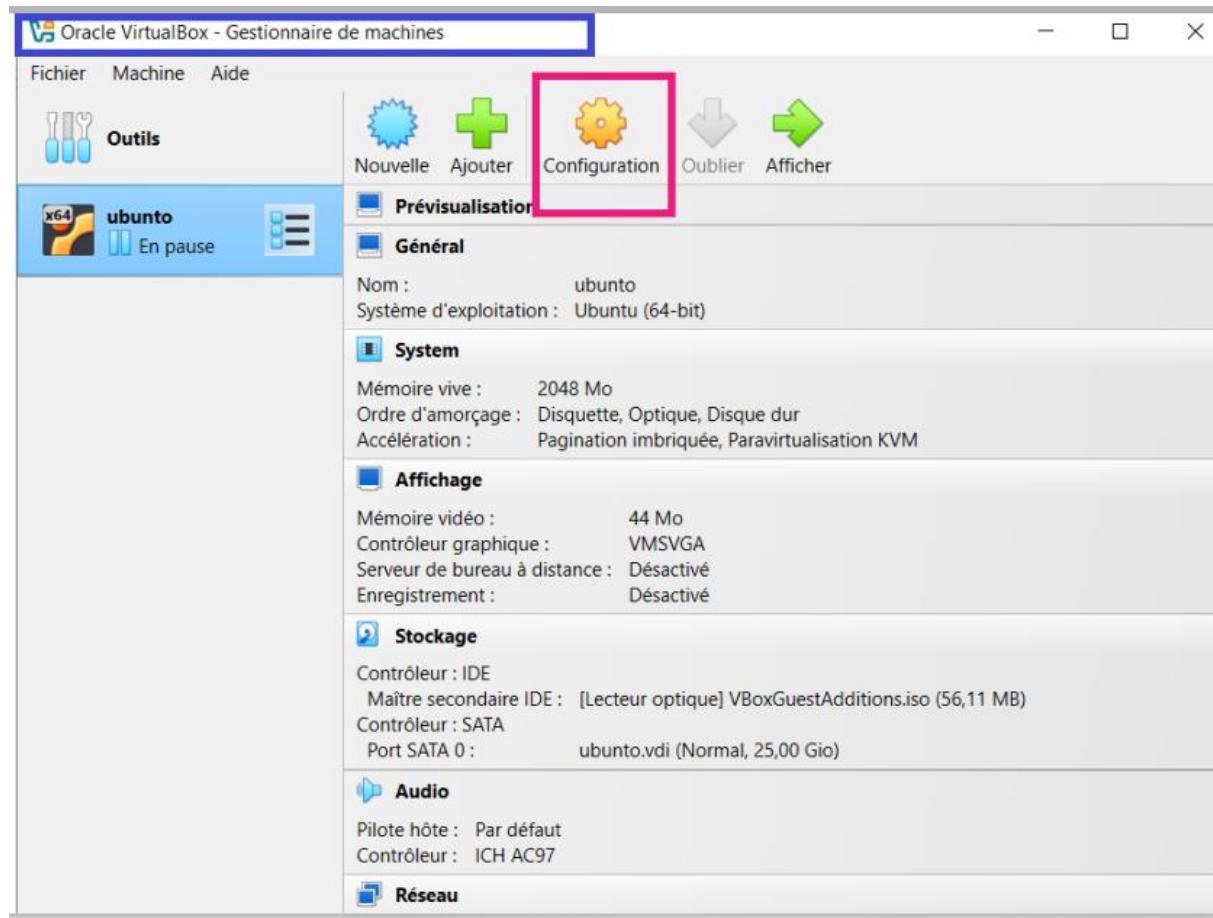
```
rahma@rahma-VirtualBox:~$ sudo -i
[sudo] password for rahma:
root@rahma-VirtualBox:~# cd /cdrom
root@rahma-VirtualBox:/cdrom# cd /media/rahma/VBox_GAs_7.1.2
root@rahma-VirtualBox:/media/rahma/VBox_GAs_7.1.2# sh ./VBoxLinuxAdditions.run
```

5/ redémarrer la machine



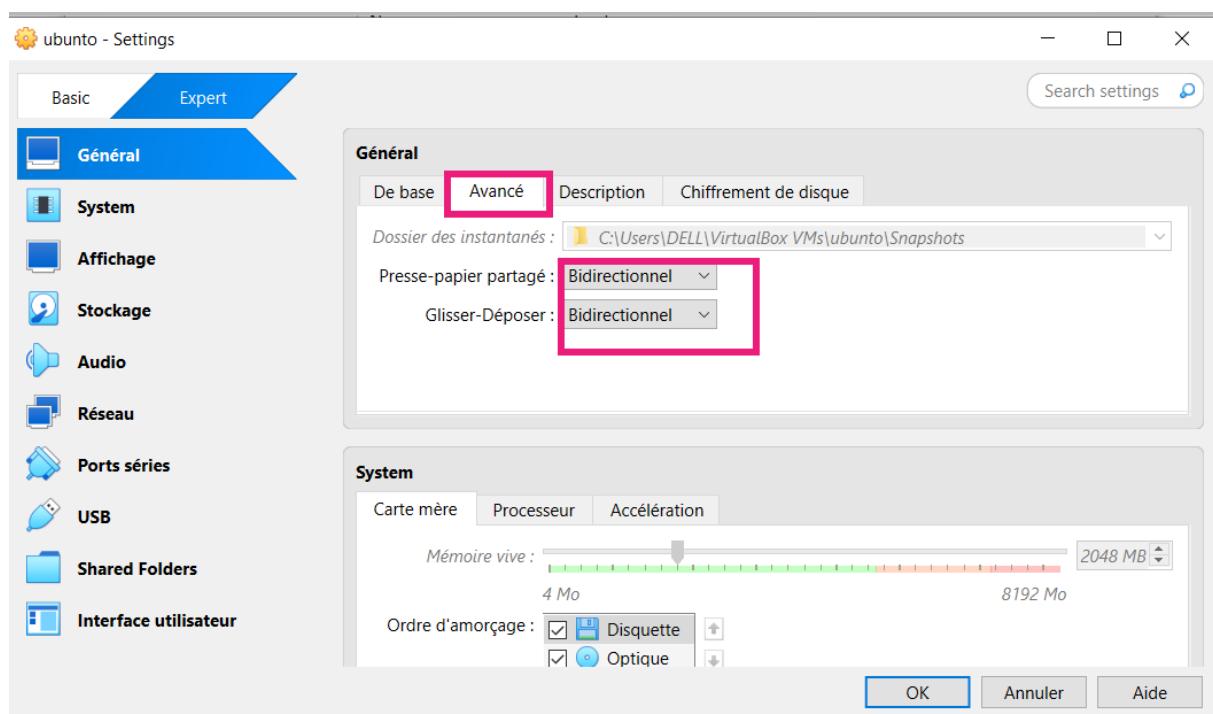
6/dans virtual machine ' oracle virtualBox'

-cliquer : Configuration



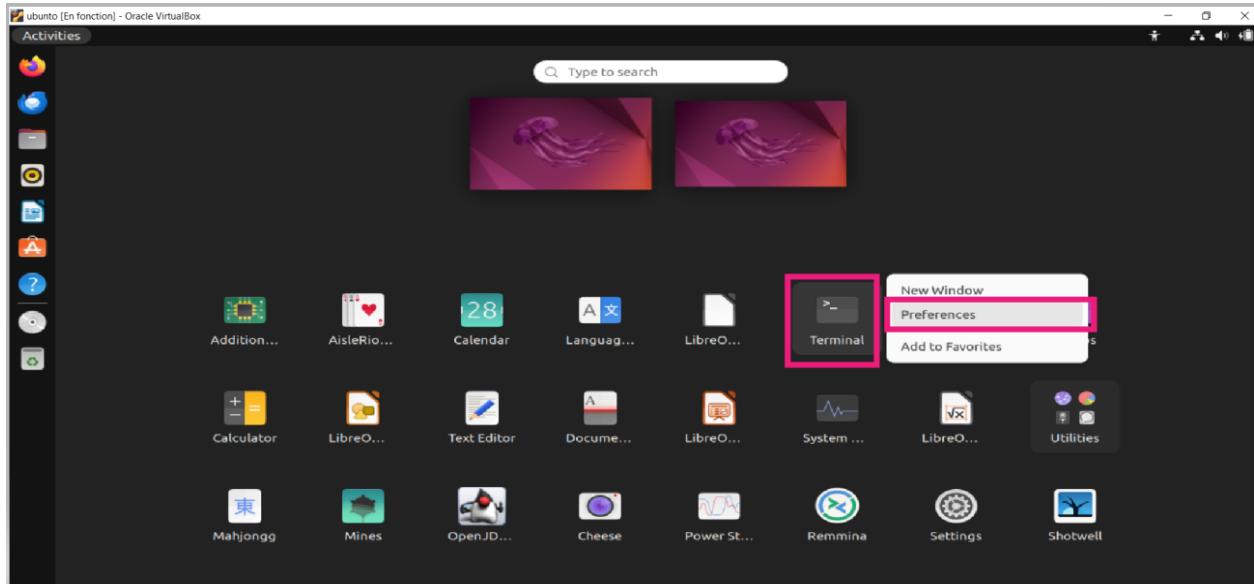
-General==> Avancé :

Choisir **bidirectionnel** pour presse-papier partager et Glisser-Déposer

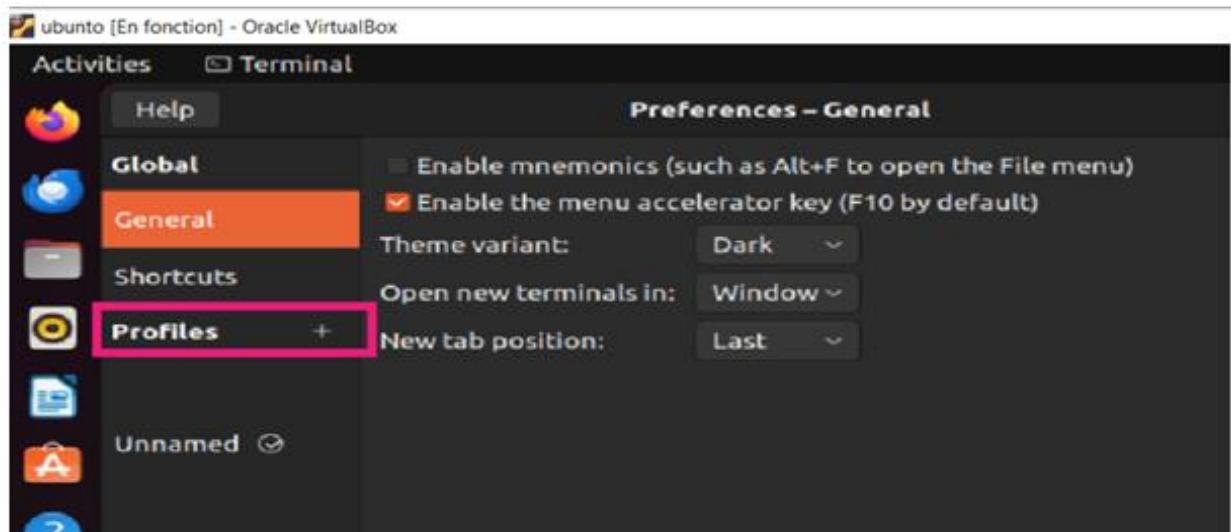


Remarque :2

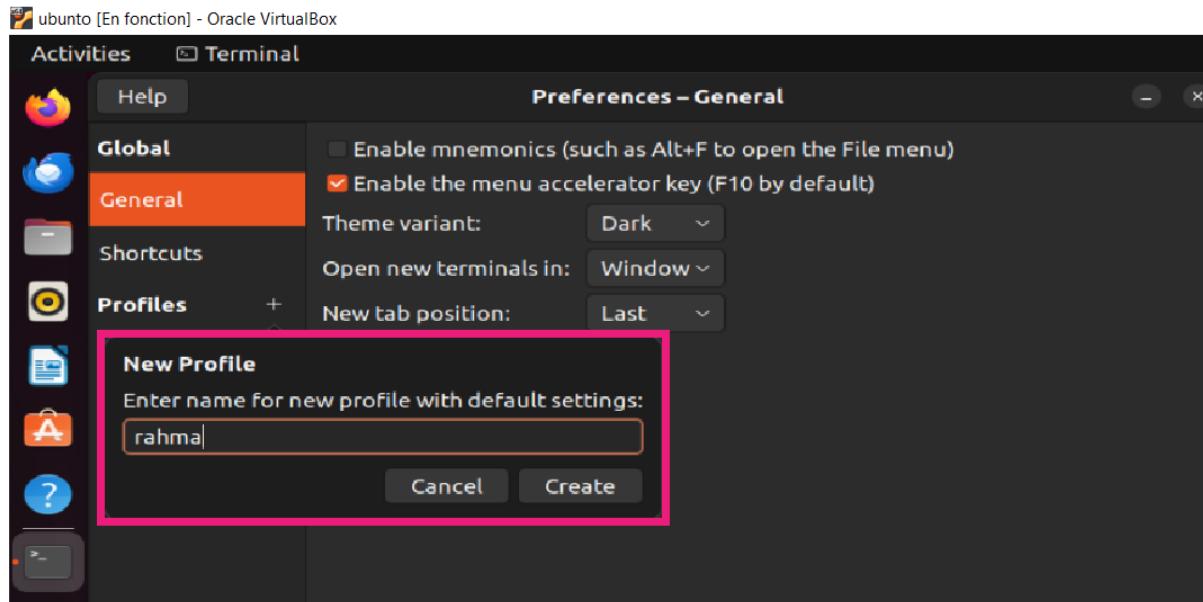
- ❖ Pour changer le format du texte dans le terminal de la machine virtuelle.
- Faites un clic droit sur le terminal et choisissez "Préférences"



- Cliquez sur le "+" de profiles.



- Créer un nouveau profil dans mon cas nomme 'rahma'



- Changer le Custom font =>DejaVu Sans Mono

